FRAMEWORK FOR VEHICLE ENVIRONMENTAL IMPACT: MACHINE LEARNING APPROACH IN CANADIAN POLICY

by

Joseph Kodzo Banini, MA Supply Chain Management

DISSERTATION

Presented to the Swiss School of Business and Management Geneva

In Partial Fulfillment

Of the Requirements

For the Degree

DOCTOR OF BUSINESS ADMINISTRATION

SWISS SCHOOL OF BUSINESS AND MANAGEMENT GENEVA

JULY, 2024

FRAMEWORK FOR VEHICLE ENVIRONMENTAL IMPACT: ML APPROACH IN CANADIAN POLICY

by

Joseph Kodzo Banini

Supervised by

Dr. BARUNAVA BANERJEE

APPROVED BY

Dr. Aaron Nyanama

Dissertation chair

RECEIVED/APPROVED BY:

Admissions Director

Dedication

This doctorate is dedicated to my beloved daughter, Ketsia Eyram Banini, and my late mother, Rebecca Kudowor. I also extend my heartfelt gratitude to God for His unwavering support and guidance throughout this journey.

To Ketsia, my precious daughter, your unwavering love and boundless joy have been the light that guided me through this journey. Your smiles and laughter have been a constant reminder of the beauty and hope that life holds. You are my inspiration and my strength. Every late night and early morning, every challenge and triumph, has been driven by my desire to build a better future for you. May this achievement show you that with perseverance and dedication, anything is possible. I hope you always dream big, work hard, and believe in yourself as much as I believe in you.

To my late mother, Rebecca Kudowor, your memory has been my anchor and my motivation. You taught me the values of resilience, hard work, and integrity. Though you are no longer with us, your spirit has been a guiding force in my life. Every step of this journey, I felt your presence and drew strength from the lessons you imparted to me. This achievement is as much yours as it is mine. I hope to honor your legacy by living a life that reflects the love, wisdom, and kindness you instilled in me.

Above all, I thank God for making this success possible. His grace and mercy have carried me through every obstacle and celebrated every victory. As it is written in

Philippians 4:13, "I can do all things through Christ who strengthens me." This scripture has been my constant reminder that with faith, perseverance, and God's guidance, all things are possible.

This doctorate is a testament to the love and support of my family, both those who are here with me and those who watch over me from above. Thank you, Ketsia and Mom, for being my pillars of strength and sources of endless inspiration. This is for you with all my love and gratitude.

Acknowledgements

I am deeply grateful for the support and guidance that enabled me to successfully complete my DBA. First and foremost, I extend my heartfelt thanks to my mentor, Dr. Barunava Banerjee. Your mentorship, wisdom, and encouragement have been invaluable throughout this journey, enriching my academic experience and fostering my personal and professional growth. I would like to express my gratitude to the management of the Swiss School of Business Management (SSBM). Your support and the conducive academic environment provided by SSBM were essential in facilitating my research and studies.

I am also profoundly thankful to the Saskatchewan Polytechnic Business School Management for their generosity in providing personal development funds. These funds played a crucial role in alleviating financial stress, allowing me to focus on my studies. The resources and support from the staff at Saskatchewan Polytechnic were instrumental in achieving this milestone, and I am grateful for their assistance throughout my academic journey.

A special dedication goes to my best friends, Ademola Afolayan and Muhammad Khan. Your unwavering support, encouragement, and understanding have been a constant source of motivation. Your friendship has been a pillar of strength, and I am deeply appreciative of your presence in my life.

This accomplishment is a collective success, and I am immensely thankful to each and every one of you for your contributions and support.

ABSTRACT

FRAMEWORK FOR VEHICLE ENVIRONMENTAL IMPACT: ML APPROACH IN CANADIAN POLICY

Joseph Kodzo Banini

2024

Dissertation Chair: <Chair's Name>

Co-Chair: <If applicable. Co-Chair's Name>

In this dissertation, we investigate using machine learning methodologies to categorize vehicles based on their ecological impact in the Canadian context, focusing on vehicle emissions and their influence on climate change and public health. The study commences with identifying key variables affecting vehicle emissions, such as CO2 emissions, fuel efficiency, engine size, fuel type, vehicle weight, and age. The research establishes strong correlations among these variables through comprehensive exploratory data analysis, particularly highlighting the inverse relationship between fuel efficiency and CO2 emissions and the higher emissions associated with more extensive and older vehicles. These findings underscore the importance of targeted policies to advocate for fuel-efficient, smaller, and newer vehicles to mitigate emissions.

The research further examines the relationships among fuel efficiency, fuel type, and vehicle emissions, confirming that fuel-efficient vehicles consistently produce lower emissions. Electric and hybrid vehicles, in particular, significantly outperform gasoline and diesel vehicles in terms of environmental impact. This analysis emphasizes the environmental benefits of transitioning to electric and hybrid technologies and lays the groundwork for policy recommendations to support this transition through incentives and infrastructure development.

A central focus of the study is developing and evaluating machine learning models for classifying vehicles based on their environmental impact. Various algorithms, including Decision Trees, Random Forests, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and K-nearest neighbours (KNN), are assessed for their effectiveness. The GBM algorithm is identified as the most accurate and robust, positioning it as a crucial tool for policymakers in identifying high-impact vehicles and designing targeted regulations. Despite challenges such as data quality and model interpretability, the success of the GBM model demonstrates the potential of machine learning in driving data-informed policy decisions and promoting sustainable transportation.

To conclude, this dissertation highlights the capability of machine learning to comprehend and diminish the environmental impact of vehicles. The research provides practical insights that can significantly enrich transportation policies by identifying critical variables, scrutinizing their interrelationships, and devising effective classification models. These insights can further support the efforts of the Canadian government in reducing greenhouse gas emissions and promoting sustainable development. Future research is suggested to integrate real-time data, broaden the geographic and temporal scope, explore advanced machine learning techniques, enhance model interpretability, and formulate scalable solutions for broader implementation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:
INTRODUCTION

1.1 Introduction

Vehicles like cars and trucks are crucial for transportation and commerce. However, their emissions pose a pressing environmental issue. The release of harmful substances like carbon dioxide, nitrogen oxides, and particulate matter contributes to global warming and air pollution. These pollutants degrade the environment and jeopardize public health, leading to conditions like asthma and heart disease. Furthermore, the noise pollution from vehicles disrupts the peace of humans and wildlife (Dey and Mehta, 2020).

Efforts to make vehicles more environmentally friendly are of paramount importance:

1. Reducing vehicle emissions can significantly combat climate change.
2. It can also markedly improve public health, particularly in urban areas where vehicle emissions are most severe.
3. It can enhance people's quality of life and protect wildlife.

We can create a healthier world for future generations by prioritizing these initiatives.

Canada knows that vehicles are a significant environmental problem and is trying to do something about it. The government has made rules to limit the amount of emissions vehicles let out. There are also programs encouraging using electric and mixed-power vehicles, which do not let out as much emissions as regular cars (Miner et al., 2024).

Local and provincial governments are also doing their part. Some offer rewards for buying electric cars, and cities are making better public transportation and paths for bikes and walking. Also, Canada has agreed to work with other countries to reduce the amount of carbon emission in the air, as in the Paris Agreement (Abdel et al., 2020).

Even with all these efforts, we still need to do more. The emission from vehicles is still a big problem, and we need new ways to deal with it. This is where machine learning can help. It can give us new ideas to understand and reduce emissions. This research will explore how machine learning can help Canada make new rules to deal with the emissions from vehicles.

The transportation sector plays a pivotal role in contributing to global greenhouse gas (GHG) emissions, accounting for nearly 25% of all CO2 emissions worldwide, with road vehicles being one of the primary sources (International Energy Agency, 2021). In Canada, the automotive industry significantly contributes to national emissions, making it essential to assess and regulate the environmental impact of vehicles through policy interventions. The government has implemented several strategies to achieve Canada's emission reduction targets, including promoting electric vehicles (EVs) and tightening fuel efficiency standards (Government of Canada, 2023). However, these efforts require a robust framework to assess the environmental impact of various vehicle types, particularly in terms of emissions and fuel efficiency.

Machine learning (ML) has emerged as a powerful tool for modelling complex environmental phenomena. It can process vast amounts of data and uncover patterns that traditional statistical methods may overlook. In the context of vehicle environmental impact, ML techniques can classify vehicles based on multiple features, such as fuel consumption, engine size, and emissions data (Chen et al., 2020). This allows for more accurate and nuanced assessments of a vehicle's environmental footprint, enabling policymakers to design targeted interventions.

This research aims to develop a "Framework for Vehicle Environmental Impact: ML Approach in Canadian Policy." The focus is leveraging machine learning algorithms to assess and classify vehicles based on key environmental factors such as CO2, NOx, particulate matter (PM), and hydrocarbons (HC) emissions. By doing so, this framework will provide policymakers with a data-driven approach to regulate vehicle emissions, promote fuel-efficient technologies, and support the transition to cleaner energy solutions. This study contributes to the body of knowledge in sustainable transportation and supports Canada's broader goals of reducing GHG emissions and combating climate change.

1.2 Overview of Vehicle Environmental Impacts

The use of vehicles is deeply ingrained in our daily lives, providing crucial mobility for people and goods. However, this widespread use has resulted in many significant environmental challenges. It is crucial to have a clear understanding of vehicles' environmental impact in order to tackle these issues effectively (Alanazi, F., 2023).

Carbon dioxide (CO2) significantly contributes to global warming and climate change (Jones et al., 2023). Burning fossil fuels such as gasoline and diesel in cars, trucks, and buses is a significant source of CO2 emissions. In addition to CO2, vehicles emit nitrogen oxides (NOx), which contribute to smog and acid rain, posing health risks such as respiratory problems (Jones et al., 2023).

Moreover, the transportation sector is a significant source of greenhouse gas emissions, primarily due to the combustion of fossil fuels (Khalili et al., 2019). Greenhouse gases such as CO2, methane (CH4), and nitrous oxide (N2O) trap heat in the atmosphere, leading to a rise in global temperature and subsequent climate change. This emphasizes the urgent need to reduce energy-related CO2 emissions from vehicles to combat climate change effectively (Filonchyk et al., 2024).

The use and manufacture of vehicles require a significant amount of natural resources. Extracting, refining, and distributing petroleum products have extensive environmental effects, such as habitat destruction and oil spills. Furthermore, vehicle production involves the use of large amounts of metals, plastics, and other materials, leading to activities like mining and processing that can harm the environment (Ristinen et al., 2022).

Furthermore, vehicles contribute to water pollution through road runoff, carrying oil, heavy metals, and other contaminants into water bodies, harming aquatic life and ecosystems. Vehicle production processes also introduce chemicals and waste products that can pollute water sources (Müller et al., 2020).

The impact of vehicles also extends to land use and habitat destruction. The construction of roads, highways, and parking lots leads to the destruction and fragmentation of natural habitats. This disruption affects ecosystems and biodiversity, as animals lose their homes and migration paths (Müller et al., 2020).

In summary, vehicles have a broad spectrum of environmental impacts, including air pollution, contributions to climate change, noise pollution, resource consumption, water pollution, and habitat destruction. Addressing these issues necessitates a comprehensive approach encompassing better policies, technological innovations, and shifts in transportation usage. By understanding these impacts, we can develop more effective strategies to reduce vehicles' environmental footprint and move towards a more sustainable future.

Vehicles emit many pollutants, significantly impacting air quality and climate change. One of the primary pollutants is carbon monoxide (CO), a colourless and odourless gas resulting from the incomplete combustion of fossil fuels (Dasari and Koul, 2015) . Exposure to CO can be particularly harmful to human health as it reduces the oxygen transported in the bloodstream, potentially harming vital organs and tissues (Kampa & Castanas, 2008).

Nitrogen oxides (NOx), which include nitric oxide (NO) and nitrogen dioxide (NO2), are also significant emissions from vehicles. These gases are produced during fuel combustion at high temperatures and contribute to forming ground-level ozone and delicate particulate matter (PM2.5). These pollutants can exacerbate respiratory diseases and contribute to smog development, posing further health risks (Khare & Adhikari, 2020).

Particulate matter (PM), including PM10 and PM2.5, is another major pollutant emitted by vehicles. These tiny particles can be deeply inhaled into the lungs, causing respiratory and cardiovascular problems. Diesel engines are mainly known for emitting high levels of particulate matter (Dockery et al., 1993).

Vehicles also emit volatile organic compounds (VOCs), organic chemicals that quickly become vapours or gases. VOCs contribute to the formation of ground-level ozone and secondary organic aerosols, detrimental to air quality and human health (Derwent et al., 2010).

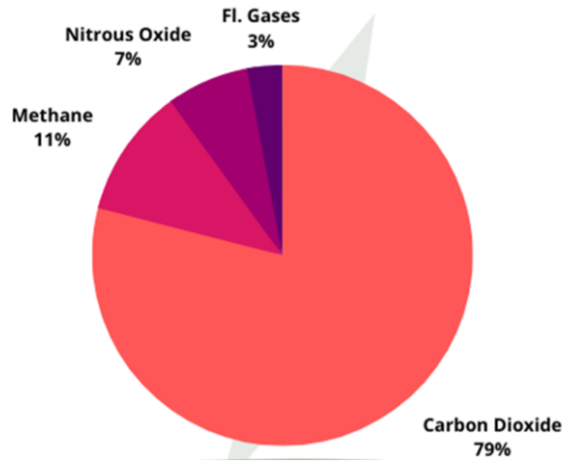

Figure 1 Green House Gas Emission

Carbon dioxide (CO2) is another significant emission from vehicles due to the combustion of fossil fuels. It is a major greenhouse gas critical to global warming and climate change, trapping heat in the Earth's atmosphere (Pachauri & Meyer, 2014).

Finally, vehicles emit sulfur dioxide (SO2) in smaller quantities than other pollutants. SO2 from diesel engines can lead to the formation of acid rain, which is harmful to ecosystems and infrastructure (Seinfeld & Pandis, 2016).

The emissions profiles of different vehicles, including cars, trucks, and buses, vary significantly due to differences in size, engine type, fuel used, and operational patterns. Understanding these variations is essential to creating specific strategies to decrease emissions and address their effects on air quality and climate change.

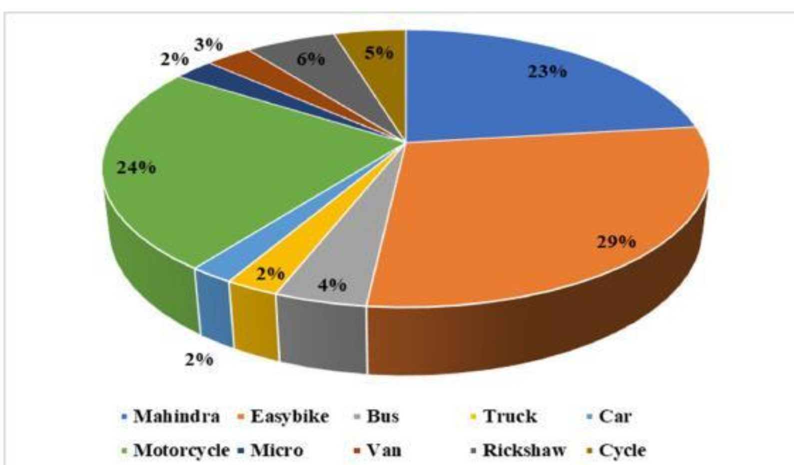

Figure 2 Type of Vehicle Produces Emissions

Cars: Passenger vehicles, including cars, generally release lower amounts of pollutants per vehicle than larger vehicles. Nonetheless, due to their high numbers, they play a substantial role in overall emissions. The primary emissions from cars include:

- Carbon monoxide (CO): Resulting from incomplete combustion of gasoline.
- Nitrogen oxides (NOx): Produced during high-temperature combustion.
- Volatile organic compounds (VOCs): Emitted from fuel evaporation and exhaust.
- Carbon dioxide (CO_2): The primary greenhouse gas from burning fossil fuels (Kampa & Castanas, 2008).

Trucks: Heavy-duty trucks, with their larger engines and frequent use of diesel fuel, substantially contribute to air pollution due to higher emissions of certain pollutants. Trucks are significant sources of elements below given by Dockery et al., 1993.

- Particulate matter (PM): Diesel engines emit higher PM levels, including black carbon, a potent climate forcer.
- Nitrogen oxides (NOx): Emissions are typically higher than those from gasoline engines(https://www.aeroqual.com/blog/meet-the-nitrogen-oxide-family).
- Sulfur dioxide (SO_2): While diesel fuel's sulfur content has been reduced, trucks still emit more SO_2 than gasoline vehicles.
- Carbon dioxide (CO_2): Trucks emit substantial amounts of CO_2 due to their larger engines and greater fuel consumption (Dockery et al., 1993).

Buses: Buses play a crucial role in public transportation, with many of them being powered by diesel. Their emissions are comparable to those of trucks, although they can differ based on whether they utilize alternative fuels or electric power. Typical emissions from buses include:

- Particulate matter (PM): Like trucks, buses emit significant PM, especially older diesel models.
- Nitrogen oxides (NOx): Emissions are high due to the diesel combustion.
- Carbon dioxide (CO_2): Emissions depend on fuel type, with diesel and CNG buses emitting considerable CO_2, while electric buses have no direct emissions (Seinfeld & Pandis, 2016).

Alternative-Fuel and Electric Vehicles: Vehicles running on alternative fuels (e.g., compressed natural gas, hydrogen) and electric vehicles (EVs) have different emissions profiles:

Electric vehicles (EVs): Produce no direct emissions, significantly reducing urban air pollution. However, the overall environmental impact depends on the electricity generation mix (Pachauri & Meyer, 2014).

CNG vehicles: Emit lower levels of NOx and PM compared to diesel and have reduced CO_2 emissions, making them a cleaner alternative (Khare & Adhikari, 2020).

Various types of vehicles release different amounts of pollutants into the atmosphere. Trucks and buses typically have higher emissions profiles than cars because they are equipped with larger engines and use diesel fuel. Changing to alternative fuels and electric vehicles can substantially decrease these emissions, ultimately enhancing air quality and helping to minimize the impacts of climate change (Khare & Adhikari, 2020).

1.3 Current Policies in Canada

Canada has taken extensive measures to address vehicle emissions and minimize their environmental impact. By Lutsey N. et al, (2007) these efforts encompass a broad range of policies, regulations, and initiatives, all aimed at fostering cleaner transportation and advancing climate objectives. Canada has implemented stringent vehicle standards to regulate greenhouse gas (GHG) emissions. Notably, CO_2 emissions limits have been established for cars and small trucks, aligning with those of the United States. Additionally, regulations for larger vehicles such as trucks and buses compel manufacturers to adopt cleaner technologies to meet these limits.

A key initiative is the Zero Emission Vehicle (ZEV) program, which seeks to enhance the affordability of electric and zero-emission vehicles (Melton N. et al, 2020). This program provides financial incentives for individuals who purchase or lease electric cars, plug-in hybrids, and hydrogen fuel cell vehicles to promote the uptake of these cleaner options. Furthermore, implementing the Clean Fuel Standard (CFS) is another pivotal policy. This regulation mandates fuel suppliers to reduce the carbon intensity of fuels utilized in transportation, industry, and buildings. By setting a target for diminishing carbon intensity, the CFS encourages using cleaner fuel alternatives such as biofuels, electricity, and hydrogen (Jin et al., 2014).

Canada has been expanding its network of electric vehicle charging stations and hydrogen refueling stations to support cleaner mobility (Razi, and Dincer, 2022). This concerted effort aims to facilitate the accessibility of charging and refueling facilities for electric and hydrogen vehicles, thus promoting their increased usage. Additionally, significant investment has been directed towards public transit to reduce reliance on private vehicles. The federal government has been providing funding for the development of new public transit systems and the expansion of existing services. Furthermore, initiatives to promote cycling and walking have been actively encouraged as viable alternatives to driving (Miele et al., 2020).

A distinct program has been established to reduce the dependence on diesel fuel in rural and remote communities. This program contributes to lowering emissions and enhancing air quality in these areas by supporting projects that utilize cleaner energy sources (Mustayen et al., 2022).

Canada has introduced carbon pricing, establishing a cost for emitting CO_2. This economic incentive catalyzes individuals and businesses to curtail their emissions, encouraging the adoption of cleaner transportation alternatives. Establishing the Strategic Innovation Fund further underscores Canada's commitment to fostering clean vehicle technologies (Miner et al., 2024). This fund provides financial support for research and development endeavors, particularly those focusing on advancements in battery technology, electric drivetrains, and lightweight materials—all of which contribute to reducing vehicle emissions (Miner et al., 2024).

Finally, the Pan-Canadian Framework on Clean Growth and Climate Change delineates Canada's comprehensive strategy to achieve emissions reduction targets under the Paris Agreement (Dillon J., 2016). This framework encompasses actions across diverse economic sectors, including transportation, and underscores the promotion of clean technology and innovation investments . These multifaceted policies and initiatives exemplify Canada's resolute commitment to mitigating vehicle emissions and lessening their environmental impact. Through the meticulous implementation of regulations and the roll-out of various programs, the government endeavors to cultivate a transportation system that is both cleaner and more sustainable, thereby aligning with national and global climate objectives

Canada has implemented extensive national measures to decrease vehicle emissions and address their environmental effects. These measures include advancing cleaner technologies, establishing strict emissions criteria, and promoting alternative fuels and electric vehicles (Ercan et al., 2022).

One notable effort is the implementation of the Greenhouse Gas Emission Regulations, which are intended to decrease greenhouse gas (GHG) emissions from new on-road vehicles and engines. (Shaheen et al, 2007) These regulations align with the United States Environmental Protection Agency (EPA) standards and encompass passenger cars, light trucks, and heavy-duty vehicles. The objectives of the standards are to enhance fuel efficiency and diminish CO_2 emissions by improving engine performance and vehicle aerodynamics (Government of Canada, 2021).

The Canadian government has introduced the Zero-Emission Vehicle (ZEV) Program to expedite the adoption of zero-emission vehicles (ZEVs). This initiative offers financial rebates for individuals and businesses who purchase eligible electric and hydrogen fuel cell vehicles. Moreover, substantial investments are being made to expand the national network of electric vehicle charging stations, making it easier to use ZEVs across the country (Natural Resources Canada, 2021).

Another crucial policy is the Clean Fuel Standard (CFS), which mandates a gradual reduction in the carbon intensity of fuels used in Canada. This regulatory approach applies to liquid, gaseous, and solid fuels, encouraging the utilization of biofuels and other low-carbon alternatives. The CFS aims to decrease the overall carbon footprint of the transportation sector by promoting cleaner fuel options (Environment and Climate Change Canada, 2020).

The EcoENERGY Efficiency Program is another initiative to enhance energy efficiency in various sectors, including transportation. This program provides resources to assist consumers and businesses in selecting energy-efficient vehicles and practices. It also offers funding and incentives for projects that lower energy consumption and GHG emissions within the transportation sector (Natural Resources Canada, 2019).

Furthermore, Canada has launched the National Greenhouse Gas Offset Credit System, which enables businesses to earn offset credits for projects that reduce GHG emissions beyond regulatory requirements. This system encompasses transportation sector projects such as fleet upgrades, adopting cleaner fuels, and implementing advanced vehicle technologies. These initiatives encourage businesses to invest in practices that contribute to reduced emissions (Environment and Climate Change Canada, 2021).

Canada's federal measures are far-reaching and address various aspects of vehicle emissions. By promoting cleaner vehicle technologies, offering financial incentives, and establishing rigorous emissions standards, these measures aim to reduce the environmental impact of transportation in Canada considerably.

Machine learning (ML) can significantly aid in promoting and enhancing Canada's vehicle emissions reduction policies through several vital avenues:

Predictive Modeling for Emissions: ML algorithms can analyze large datasets from various sources, such as traffic data, vehicle specifications, and fuel consumption patterns, to predict future emissions levels. This would allow policymakers to forecast the potential impact of various regulations, such as the Zero Emission Vehicle (ZEV) program or Clean Fuel Standard (CFS), and adjust them in real-time to meet environmental targets.

Optimization of Public Transit and EV Infrastructure: ML models can analyze traffic patterns, driver behaviors, and geographical data to optimize the placement and usage of electric vehicle (EV) charging stations and hydrogen refuelling stations. This would support Canada's efforts to expand the charging network for ZEVs, as noted in the Strategic Innovation Fund and Clean Fuel Standard (Hassan et al., 2024).

- Smart Fleet Management: For businesses and public transport systems, ML can optimize fleet management by identifying the most fuel-efficient routes, minimizing idle times, and scheduling maintenance for vehicles to ensure optimal performance. These initiatives align with the Greenhouse Gas Emission Regulations, aiming to reduce emissions through better engine performance and aerodynamics.
- Carbon Pricing and ML-Driven Incentives: Through ML, real-time data on fuel consumption, emissions, and driving patterns can be collected to dynamically adjust carbon pricing or financial incentives under the EcoENERGY Efficiency Program or carbon pricing programs. This personalized approach would empower consumers and businesses to shift towards cleaner vehicles or reduce their carbon footprint.
- Emissions Monitoring and Reporting: ML can streamline emissions monitoring by automating the data collection and providing accurate insights. For example, ML can monitor compliance with GHG emissions standards and help enforce the regulations efficiently.

- Behavioural Insights: ML can analyze consumer behaviours to understand what motivates people to adopt cleaner technologies like electric or hybrid vehicles. This can help tailor the ZEV Program incentives or marketing efforts, ensuring they target the right demographics, thereby accelerating the adoption of ZEVs.

By leveraging these ML capabilities, Canada can ensure the efficient implementation and continuous improvement of its vehicle emissions reduction policies, thereby advancing its climate goals and providing reassurance to all stakeholders (Hassan et al., 2024).

1.4 Machine Learning Approaches in Environmental Studies

Machine Learning (ML) is a powerful tool used in environmental research to address complex issues. ML algorithms analyze large datasets to identify patterns and make predictions, aiding researchers and policymakers in understanding and mitigating environmental problems.

This section delves into the application of ML in environmental research, focusing on its role in mitigating vehicle emissions and other environmental impacts (Ghaffarian S. et al, 2023). Before introducing predictive modeling and emission forecasting, it is essential to understand the various types of machine learning (ML) models that can be applied to promote clean transportation policies and mitigate environmental impacts. One of the most common types is supervised learning, which involves training a model on labeled data where the input-output pairs are known. This approach can be used to predict specific outcomes such as fuel consumption or emissions levels based on vehicle specifications like engine size and fuel type (Shalev-Shwartz & Ben-David, 2014). Supervised learning can also assist in predicting the likelihood of different vehicles meeting stringent emissions standards based on historical data, making it a powerful tool for regulatory compliance.

Another key model type is unsupervised learning, which works with unlabeled data to discover hidden patterns or clusters within large datasets. This method is useful for grouping vehicles based on emissions profiles or understanding consumer behavior related to adopting zero-emission vehicles (ZEVs). For example, clustering techniques could identify areas with higher concentrations of high-emission vehicles, helping target regions for expanding electric vehicle (EV) infrastructure (Aggarwal & Reddy, 2014). Additionally, reinforcement learning models, which learn to make decisions through trial and error, can optimize transportation networks by identifying the most fuel-efficient routes or dynamically adjusting carbon pricing based on real-time emissions data (Sutton & Barto, 2018).

ML techniques such as dimensionality reduction and clustering also play a vital role in simplifying complex datasets related to vehicle emissions. Dimensionality reduction techniques like Principal Component Analysis (PCA) can distill large datasets into key variables, allowing for more efficient policy analysis and implementation (Jolliffe &

Cadima, 2016). Clustering methods can group regions based on their emissions intensity, which supports the formulation of targeted environmental policies.

Building on these ML foundations, predictive modeling can then be employed to forecast vehicle emissions under various policy scenarios. By training models on historical data, it is possible to estimate how changes in vehicle standards, fuel consumption patterns, or the introduction of new programs like the Clean Fuel Standard (CFS) and the ZEV program might impact emissions in the future (Lutsey et al., 2007; Melton et al., 2020). Predictive modeling allows policymakers to simulate the long-term effects of various interventions, helping ensure that Canada's ambitious emissions reduction targets are met. This approach provides the ability to adapt dynamically to evolving conditions and refine policies based on real-time forecasts, which is crucial in the context of reducing transportation-related emissions and promoting sustainable mobility.

Predictive Modeling and Emissions Forecasting: ML is commonly used in environmental research for predictive modelling (Giannelos S. et al, 2024). ML algorithms can forecast future emissions trends by analyzing historical data on vehicle emissions and other environmental factors. These models help policymakers anticipate the environmental impact of various policy scenarios and make informed decisions. For example, ML can predict the effects of implementing stricter emissions standards or increasing the adoption of electric vehicles, allowing policymakers to design more effective strategies to reduce emissions.

Air Quality Monitoring and Management: ML techniques play a significant role in monitoring and managing air quality. (Imam et al, 2024) Sensors in different locations gather real-time data on pollutants like $CO2$, NOx, and particulate matter. ML algorithms process this data to identify pollution sources, track changes, and predict pollution levels. This information is crucial for developing targeted interventions to improve air quality. Additionally, ML can assist in optimizing the placement of air quality sensors, ensuring comprehensive coverage and accurate data collection (Cascon L. et al, 2024).

Traffic Flow Optimization: ML can optimize traffic flow in urban areas, thus reducing vehicle emissions and improving air quality. By analyzing traffic patterns and congestion data, ML algorithms can develop strategies to minimize traffic jams and enhance transportation network efficiency. For instance, ML can inform the timing of traffic signals, suggest alternative routes to drivers, and manage public transportation schedules. These optimizations lead to smoother traffic flow, lower fuel consumption, and reduced emissions (Lv. Y. et al, 2014).

Electric Vehicle (EV) Adoption and Infrastructure Planning: ML is crucial in supporting electric vehicle adoption and infrastructure planning as they become more prevalent. ML algorithms analyze EV usage, charging patterns, and geographic distribution data to predict future demand for charging stations. This data helps planners decide where to install new charging infrastructure, ensuring accessibility and convenience for EV users. Furthermore, ML can optimize the management of charging stations, balancing the load on the power grid and preventing overloads (Johar et al, 2024).

Climate Change Modeling: ML is instrumental in climate change modelling, aiding scientists in understanding and predicting the long-term impacts of various factors, including vehicle emissions. (Chowdhary P, et al, 2023) By processing large datasets from climate models, satellite observations, and historical records, ML algorithms can identify trends and correlations that traditional methods might miss. These insights contribute to more accurate climate predictions and better strategies for mitigating the effects of climate change (Chen L. et al, 2023).

Wildlife and Habitat Conservation: ML applications extend beyond air quality and emissions, encompassing wildlife and habitat conservation. For example, ML can analyze data from camera traps, satellite images, and environmental sensors to monitor wildlife populations and track habitat changes (Raihan A. 2023). This information helps conservationists identify threats to biodiversity and develop strategies to protect endangered species. In the context of vehicle environmental impacts, ML can assess how road construction and traffic affect wildlife movement and habitat fragmentation, informing mitigation measures (Santos Morini et al, 2023).

Water Quality Monitoring :Water quality is another area where ML is making significant contributions. ML algorithms analyze data from sensors placed in rivers, lakes, and oceans to monitor pollutants and predict water quality trends (Essamlali, I, et al, 2024). This information is vital for protecting aquatic ecosystems and ensuring safe drinking water. Regarding vehicles, ML can track runoff from roads and parking lots, identify sources of contamination, and inform strategies to reduce water pollution (Ali N. et al, 2023).

Renewable Energy Integration: Integrating renewable energy sources into transportation is crucial for reducing vehicle emissions (Barman P, et al, 2023). ML supports this integration by optimizing energy management systems, predicting renewable energy production, and balancing supply and demand (Alazemi T. et al, 2024). For example, ML can forecast solar and wind energy availability, assisting grid operators in managing electricity distribution to charging stations for electric vehicles. This ensures a stable and sustainable energy supply, reducing reliance on fossil fuels (Johar et al, 2024).

One last thing to note is that Machine Learning revolutionizes environmental studies by providing robust tools to analyze data, predict outcomes, and optimize solutions. In the context of vehicle environmental impacts, ML aids in forecasting emissions, monitoring air and water quality, optimizing traffic flow, supporting electric vehicle adoption, climate change modelling, wildlife and habitat conservation, water quality monitoring, and renewable energy integration (Tien P.W. et al, 2022).

Machine learning (ML) has the potential to completely transform the way we analyze environmental data, offering significant improvements in accuracy and efficiency. There are numerous ways in which ML can elevate environmental data analysis:

1. Handling Large and Complex Datasets: The data relating to the environment is often extensive and intricate, covering a wide range of data types such as satellite images, sensor readings, and climate models (Zakeri, F. et al 2024). Conventional analytic

techniques need help in efficiently processing and understanding these massive datasets. Machine learning algorithms can swiftly handle and analyze large volumes of data, identifying patterns and trends that traditional methods may overlook. This capacity enables a more thorough and timely analysis of environmental occurrences (Guo et al., 2018).

2. Improved Predictive Accuracy: Machine learning models, particularly those utilizing neural networks and ensemble methods, have demonstrated greater predictive precision than conventional statistical models. These models can assimilate information from past data and produce precise forecasts about upcoming environmental circumstances, including air quality, meteorological trends, and the effects of climate change. ML models offer more dependable predictions by integrating a broad array of factors and understanding intricate relationships (Reichstein et al., 2019).

3. Real-Time Monitoring and Anomaly Detection: Machine learning algorithms can monitor environmental conditions in real-time. For example, ML models can examine information from a system of air quality sensors to recognize increases in pollution or pinpoint sources of contamination. Algorithms for anomaly detection can rapidly highlight unusual patterns, leading to immediate investigation and action. This real-time functionality is essential for reducing the effects of environmental dangers (Ma et al., 2019).

4. Enhanced Image and Signal Processing: Environmental monitoring often depends on data from remote sensing, such as satellite images and aerial photos. Machine learning methods, particularly convolutional neural networks (CNNs), can effectively handle image and signal processing tasks. CNNs can categorize land cover, identify deforestation, assess water quality, and precisely estimate wildlife populations. By automating remote sensing data analysis, machine learning reduces the necessity for manual interpretation and accelerates the decision-making process (Zhu et al., 2017).

5. Integration of Diverse Data Sources: Machine learning models can combine information from various sources, such as satellite images, ground sensors, weather stations, and social media. This comprehensive method allows for a deeper understanding of environmental problems. For instance, merging meteorological data with social media posts can improve the effectiveness of disaster response systems by offering immediate insights into affected regions (Li et al., 2020).

6. Decision Support Systems: Machine learning has the potential to enhance the decision-support systems utilized by policymakers and environmental managers. ML models can offer precise predictions and up-to-date insights, facilitating well-informed choices related to resource management, conservation activities, and policy interventions. For example, ML can enhance irrigation schedules in agriculture, forecast the likelihood of forest fires, and simulate the effects of urbanization on local climates (Karpatne et al., 2018).

Machine learning (ML) is poised to revolutionize environmental data analysis in several ways. Firstly, it can efficiently handle large and complex datasets, including satellite images, sensor readings, and climate models, enabling more thorough and timely analysis of environmental occurrences (Li F. et al 2023). Secondly, ML models offer improved predictive accuracy, particularly in forecasting air quality, meteorological trends, and the effects of climate change. Thirdly, machine learning algorithms can monitor environmental conditions in real time, detecting anomalies and leading to immediate investigation and action. Additionally, ML methods can effectively process remote sensing data, such as satellite images and aerial photos, reducing the need for manual interpretation. Furthermore, ML models can integrate diverse data sources, providing a deeper understanding of environmental issues and aiding disaster response systems. Finally, machine learning has the potential to enhance decision-support systems for policymakers and environmental managers by offering precise predictions and up-to-date insights for well-informed choices related to resource management, conservation activities, and policy interventions.

Machine learning (ML) has many potential applications that can significantly reduce pollution and mitigate its impacts. These applications span various sectors, including air quality monitoring, industrial emissions control, waste management, and environmental policy enforcement (Zhou et al., 2020). Here are some key areas where ML can be effectively utilized:

Air Quality Monitoring and Prediction: Machine learning algorithms can analyze data from air quality sensors, weather stations, and satellite imagery to monitor pollution levels in real time and predict future air quality conditions. For example, ML models can forecast the concentration of pollutants such as PM2.5, NOx, and ozone, allowing authorities to take proactive measures to reduce emissions and protect public health. Advanced ML techniques can also identify pollution hotspots and sources, enabling targeted interventions (Chen et al., 2019).

Industrial Emissions Control: ML can optimize processes to minimize emissions of harmful pollutants in industrial settings. By analyzing operational data, ML models can identify inefficiencies and recommend adjustments to reduce emissions of CO2, sulfur dioxide (SO2), and other pollutants. For instance, ML can fine-tune combustion processes in power plants and manufacturing facilities, leading to cleaner production methods (Zhou et al., 2020).

Transportation Management: ML can improve traffic management and reduce vehicle emissions by optimizing traffic flow and reducing congestion. Intelligent transportation systems (ITS) powered by ML can analyze traffic patterns and control traffic signals to minimize idle time and reduce vehicle emissions. Additionally, ML can be used to develop efficient public transportation schedules and promote the use of electric and low-emission vehicles through intelligent routing and charging infrastructure planning (Hsu et al., 2019).

Waste Management: Machine learning can enhance waste management practices by optimizing waste materials collection, sorting, and recycling. ML algorithms can predict waste generation patterns, allowing for more efficient waste collection scheduling and reducing fuel consumption. Moreover, ML can improve the sorting process in recycling facilities by accurately identifying and separating recyclable materials from waste streams, thus increasing recycling rates and reducing landfill use (Rada et al., 2018).

Water Quality Monitoring and Management: ML models can monitor and predict water pollution levels by analyzing data from sensors placed in water bodies. These models can detect anomalies and identify sources of contamination, such as industrial discharges or agricultural runoff. By providing real-time insights into water quality, ML can help authorities implement timely measures to prevent water pollution and protect aquatic ecosystems (Gómez et al., 2019).

Environmental Policy Enforcement: Machine learning can support enforcing environmental regulations by analyzing data from various sources, including satellite imagery, social media, and on-the-ground sensors. ML algorithms can detect illegal activities such as deforestation, unregulated mining, and unauthorized emissions, enabling authorities to take swift action. For instance, satellite-based ML models can identify illegal logging activities in real time, facilitating rapid response and enforcement (Mateo-Sagasta et al., 2017).

Public Health Impact Assessment: ML can assess the impact of pollution on public health by correlating pollution data with health records. By identifying patterns and trends, ML models can quantify the health effects of pollution and predict future health risks. This information can guide policymakers in designing effective interventions to reduce pollution-related health issues and allocate resources to areas with the highest need (Tham et al., 2019).

Machine learning offers numerous applications that significantly reduce pollution and mitigate its impacts across various sectors. ML can play a critical role in creating a cleaner and healthier environment by improving air and water quality monitoring, optimizing industrial and transportation processes, enhancing waste management, and supporting environmental policy enforcement (Zhou et al., 2020).

1.5 Overview of environmental issues related to vehicles

Vehicles are significant in modern society, providing essential mobility and transportation services. However, their environmental impact is profound and multifaceted, affecting air quality, climate change, noise pollution, water resources, and land use (Gómez et al., 2019). This section provides an overview of the various environmental issues associated with vehicles, highlighting the importance of addressing these challenges to ensure a sustainable future (Xu S., et al, 2024).

Air Pollution: One of the most immediate and noticeable impacts of vehicles is air pollution. Internal combustion engines emit a variety of pollutants, including carbon monoxide (CO), nitrogen oxides (NOx), hydrocarbons (HC), and particulate matter (PM). These pollutants can have severe health effects, particularly in urban areas with high traffic density. Long-term exposure to vehicle emissions is linked to respiratory diseases, cardiovascular conditions, and premature death. Additionally, pollutants like NOx and HC contribute to the formation of ground-level ozone, which can exacerbate asthma and other lung diseases (Dey S. et al, 2020).

Vehicles are a significant source of greenhouse gas (GHG) emissions, particularly carbon dioxide (CO2), the primary contributor to climate change (Ramani and Zietsman, 2016). The combustion of fossil fuels in vehicle engines releases significant amounts of CO2 into the atmosphere. In addition to CO2, vehicles emit methane (CH4) and nitrous oxide (N2O), which have a much higher global warming potential than CO2 (Comi and Polimeni, 2020). The transportation sector is one of the most significant contributors to global GHG emissions, and reducing vehicle emissions is crucial for meeting international climate targets and limiting global warming.

Noise Pollution: Vehicle noise is another significant environmental issue, particularly in urban areas. Traffic noise can disrupt daily activities, reduce the quality of life, and contribute to health problems such as stress, sleep disturbances, and hearing loss. Noise pollution also affects wildlife, interfering with animal communication, breeding, and foraging behaviours. Efforts to mitigate noise pollution include the development of quieter engines, better road surfaces, and promoting alternative transportation modes such as cycling and walking (D'Agosto, M.D.A, 2019).

The environmental impact of vehicles extends beyond their operational phase. The entire lifecycle of a vehicle—from raw material extraction, manufacturing, and transportation to use, maintenance, and disposal—has significant environmental implications. Manufacturing processes can be energy-intensive and polluting, while the disposal of end-of-life vehicles poses challenges for waste management and recycling systems. Adopting a lifecycle approach to vehicle design and policy-making can help minimize these impacts by considering environmental effects at each stage of a vehicle's life (Williams, I.D. et al, 2023).

Addressing the environmental issues related to vehicles requires a comprehensive approach that combines technological innovations, regulatory measures, and changes in consumer behaviour. Key strategies include:

- Promoting the adoption of electric and hybrid vehicles, which produce fewer emissions compared to traditional internal combustion engine vehicles (Zhao, X., Hu, et al, 2023).
- Implementing stricter emissions standards and regulations to reduce pollutants from vehicles.
- Investing in public transportation infrastructure to reduce the reliance on private cars (Holmgren, J., 2020).
- Encouraging active transportation modes such as cycling and walking.
- Supporting research and development in cleaner vehicle technologies and sustainable materials (Mikulčić, H., et al, 2022).
- Enhancing recycling and waste management systems for vehicle components and fluids (Molla, A.H., et al, 2023).

In conclusion, vehicles have a profound and multifaceted impact on the environment, affecting air quality, climate change, noise levels, water resources, and land use. Addressing these issues requires a coordinated effort involving technological advancements, regulatory frameworks, and shifts in societal attitudes towards transportation. By adopting sustainable practices and policies, we can mitigate the environmental impact of vehicles and move towards a cleaner, healthier, and more sustainable future.

Vehicles significantly contribute to global greenhouse gas (GHG) emissions, a primary driver of climate change. Understanding the connection between vehicle emissions and climate change is crucial for developing effective strategies to mitigate these impacts.

1. Carbon Dioxide (CO2): The most prevalent greenhouse gas released by vehicles is carbon dioxide, which comes from burning fossil fuels like gasoline and diesel. Combustion of these fuels releases carbon stored underground into the air in the form of CO2.

- Impact on Climate Change:

Global Warming: Carbon dioxide (CO2) is the main contributor to the greenhouse effect, trapping heat in the Earth's atmosphere and causing global warming. Cars, trucks, buses, and motorcycles, collectively known as the transportation sector, are accountable for considerable global CO2 emissions. According to the International Energy Agency (IEA), transportation is responsible for nearly 24% of CO2 emissions from fuel combustion globally (IEA, 2021).

Long-Term Climate Effects: The rise in atmospheric CO2 levels intensifies the greenhouse effect, resulting in more heat trapped by the Earth's atmosphere. This results in higher global temperatures, which affect weather patterns, sea levels, and ecosystems (Pachauri & Meyer, 2014).

2. Methane (CH4) and Nitrous Oxide (N2O): The primary greenhouse gas emitted from vehicles is CO2, but certain types of engines and fuel can also release methane and nitrous oxide. Methane may escape during fuel extraction and distribution, while nitrous oxide is produced by catalytic converters in exhaust systems.

Higher Global Warming Potential: Both methane and nitrous oxide have a significantly higher global warming potential (GWP) than CO2, which indicates that they are more adept at retaining heat in the atmosphere. Methane's GWP is around 25 times greater than CO2 over 100 years, and nitrous oxide's GWP is approximately 298 times that of CO2 (IPCC, 2013).

Contribution to Short-Term Climate Change: Even minor releases of methane (CH4) and nitrous oxide (N2O) can have a notable impact on global warming and climate change due to their high global warming potential (GWP).

3. Black Carbon: The following text should be remembered: "Black carbon, a type of particulate matter (PM), is created when diesel fuel is incompletely burned. This substance has a strong warming effect on the atmosphere and is considered a powerful climate influencer (Mikulčić, H., et al, 2022)."

- Impact on Climate Change:

Direct Warming: Black carbon captures sunlight and produces heat, heating the atmosphere directly. When it settles on snow and ice, it lowers its albedo (reflectiveness), increasing heat absorption and faster melting. This leads to polar ice and glaciers declining, worsening global warming (Bond et al., 2013).

Short-Lived Climate Pollutant: Even though black carbon remains in the atmosphere for a shorter time than CO_2, it has a much more significant impact on climate warming per unit mass. Immediate benefits for slowing down climate change can be realized by reducing black carbon emissions (Mathew, N, et al, 2024).

The emissions from vehicles play a significant role in releasing greenhouse gases such as CO_2, CH_4, N_2O, and black carbon, which are responsible for global warming and climate change. Understanding and tackling these emissions is essential for creating better strategies to reduce their impact and safeguard the environment (Pachauri & Meyer, 2014).

Urban planning plays a crucial role in reducing vehicle environmental impact. One critical strategy involves developing efficient and widespread public transportation systems that offer sustainable alternatives to private car use. Focusing on sustainable urban growth and providing viable public transportation options can effectively decrease vehicle emissions and enhance overall air quality (Nieuwenhuijsen, M.J., 2020).

1. Integrated Urban Planning: Urban planning plays a crucial role in determining city residents' transportation needs and habits. By integrating urban planning with a focus on sustainable development, cities can substantially progress in reducing vehicle emissions (Bibri, S.E, et al, 2020).

Compact Urban Design: Encouraging high-density, mixed-use development reduces the distance people travel for work, shopping, and leisure. This can lead to fewer car trips and more walking, cycling, and public transit use (Ewing & Cervero, 2010).

Transit-Oriented Development (TOD): Designing neighborhoods around public transportation hubs ensures that residents can access efficient transit options easily. TOD reduces reliance on cars and promotes using buses, trains, and trams (Cervero et al., 2004).

Green Spaces and Non-Motorized Transport: Incorporating parks, pedestrian zones, and bike lanes into urban areas encourages active transportation modes, such as walking and cycling, with zero emissions (Pucher & Buehler, 2010).

2. Development of Public Transportation: Efficiently managed and well-established public transportation systems are crucial in decreasing the volume of private vehicles on the roads and reducing overall emissions (Baz and Iddik, 2020).

High-Frequency, Reliable Service: Ensuring that public transit services are frequent, reliable, and cover extensive areas can attract more users. This reduces the need for personal vehicles and helps lower emissions from road transport (Litman, 2017).

Electrification of Public Transit: Transitioning buses and trains to electric power reduces the emissions associated with public transportation. Electric vehicles produce no tailpipe emissions and have a much lower carbon footprint when powered by renewable energy sources (Shen et al., 2018).

Affordable and Accessible Transit: Making public transportation affordable and accessible to all population segments encourages its use. Subsidized fares, integrated ticketing systems, and facilities for people with disabilities improve the appeal of public transit (Bocarejo & Oviedo, 2012).

3. Policies and Incentives: Governments can create and enforce policies and offer incentives that promote the utilization of public transportation and decrease reliance on private vehicles.

Congestion Pricing: Charging fees for driving in congested urban areas during peak hours can discourage car use and raise funds for public transportation projects. Cities like London and Singapore have successfully implemented congestion pricing to manage traffic and reduce emissions (Leape, 2006).

Parking Management: Reducing the availability of parking spaces and increasing parking fees can discourage car use. Policies prioritizing parking for carpool vehicles and public transit users can further reduce the number of cars on the road (Shoup, 2005).

Incentives for Public Transit Use: Offering tax incentives, employer-sponsored transit passes, and other benefits for public transit users can make it more attractive than driving (Chen et al., 2011).

1.6 Importance of Addressing Vehicle Emissions and Environmental Impact

Addressing vehicle emissions and their environmental impact is essential for several interconnected reasons. Vehicle emissions significantly impact public health, particularly in urban areas with high traffic congestion. The release of these substances, such as carbon monoxide, nitrogen oxides, hydrocarbons, and particulate matter, has adverse effects on human health, leading to respiratory conditions like asthma and bronchitis, as well as cardiovascular problems such as heart attacks and hypertension, and is linked to premature death. Furthermore, vehicles significantly produce greenhouse gases, particularly carbon dioxide, exacerbating global warming and climate change. It is essential to mitigate these emissions to achieve international climate goals and reduce the frequency and severity of extreme weather events. Environmental preservation is also at stake because vehicle emissions degrade air quality, harm aquatic ecosystems through runoff contamination, and disrupt soil health, impacting agriculture and natural habitats. Addressing emissions can also lead to economic benefits, including reduced healthcare costs associated with pollution-related illnesses, promotion of fuel efficiency through technological innovations like electric and hybrid vehicles, and job creation in green technology sectors.

Furthermore, efforts to tackle vehicle emissions drive technological advancements and intelligent transportation solutions, supporting a transition to sustainable energy systems and cleaner, more efficient urban planning (Elassy, M, et al, 2024). Effective policies and regulations are crucial, enforcing emissions standards, incentivizing clean vehicle adoption, and investing in public transit and infrastructure to encourage alternative transportation modes (Wu, Y.C. , et al 2024). In conclusion, comprehensively addressing vehicle emissions is about safeguarding public health and the environment and promoting economic growth, technological innovation, and sustainable development for a healthier and more resilient future (Wang, F., et al, 2021).

PM, a significant pollutant, consists of fine particles such as PM2.5 (smaller than 2.5 micrometres) that can penetrate deep into the lungs and enter the bloodstream. These minuscule particles are linked to various respiratory and cardiovascular diseases. Research indicates that exposure to PM2.5 can result in the onset and worsening of conditions like asthma, bronchitis, and chronic obstructive pulmonary disease (COPD). Furthermore, PM exposure is linked to cardiovascular issues, including heart attacks, strokes, and atherosclerosis, due to the inflammatory response and oxidative stress it induces in the body. Long-term exposure to high levels of PM2.5 has been associated with increased mortality rates from both respiratory and cardiovascular diseases (Dockery et al., 1993; Pope et al., 2002).

Nitrogen Oxides (NOx): NOx, including nitrogen dioxide (NO2), are harmful gases produced during high-temperature combustion in vehicle engines. NO2 can irritate the airways, leading to increased susceptibility to respiratory infections, reduced lung function, and the aggravation of asthma and other chronic respiratory conditions. Long-term exposure to NO2 has been linked to the development of asthma, particularly in children, making it a significant concern for public health (Guarnieri & Balmes, 2014). NOx also contributes to the formation of ground-level ozone and delicate particulate matter, compounding its harmful effects on respiratory health (Khreis et al., 2017).

Carbon Monoxide (CO): CO is a colourless, odourless gas resulting from the incomplete combustion of fossil fuels. It binds with haemoglobin in the blood to form carboxyhemoglobin, which reduces the blood's ability to carry oxygen. This can lead to symptoms ranging from headaches, dizziness, and confusion to severe cardiovascular and neurological damage in high concentrations. Individuals with preexisting heart conditions are particularly vulnerable to CO exposure, as it can exacerbate angina and other heart-related issues (Kampa & Castanas, 2008; Raub et al., 2000).

Volatile Organic Compounds (VOCs): VOCs are a group of organic chemicals that quickly vaporize at room temperature and contribute to the formation of ground-level ozone. Exposure to VOCs can cause a range of health problems, including eye, nose, and throat irritation, headaches, and nausea. Some VOCs, such as benzene, are known carcinogens and can cause long-term health effects, including liver and kidney damage (Bernstein et al., 2008). The role of VOCs in forming secondary organic aerosols further exacerbates air quality issues and associated health risks (Derwent et al., 2010).

Sulfur Dioxide (SO2): SO2 is produced by the combustion of sulfur-containing fuels, such as diesel. It can irritate the respiratory system, causing symptoms such as coughing, mucus secretion, and the aggravation of asthma and chronic bronchitis. Short-term exposure to high levels of SO2 can lead to increased respiratory symptoms and decreased lung function, particularly in individuals with preexisting respiratory conditions. Additionally, SO2 can react with other compounds in the atmosphere to form delicate particulate matter, further contributing to its harmful health effects (Pope & Dockery, 2006; Brook et al., 2010).

- Improving Public Health Outcomes by Reducing Emissions

Reducing vehicle emissions can lead to significant improvements in public health outcomes. Lowering the PM, NOx, CO, VOCs, and SO2 levels in the air can decrease the incidence and severity of respiratory and cardiovascular diseases. Cleaner air results in fewer hospital admissions, lower healthcare costs, and improved quality of life for individuals with preexisting health conditions (Pope et al., 2009). Reducing exposure to harmful pollutants like PM2.5 and NO2 can also lower premature mortality rates associated with air pollution-related diseases, contributing to longer lifespans and healthier populations (Hoek et al., 2013).

Children are particularly vulnerable to the effects of air pollution. Reducing vehicle emissions can decrease the incidence of childhood asthma, respiratory infections, and developmental issues, leading to healthier growth and development (Gauderman et al., 2004). Moreover, improving air quality by reducing vehicle emissions can enhance public health and well-being. People experience fewer symptoms of respiratory irritation, better lung function, and improved outdoor activity enjoyment, contributing to a healthier and more active society (Künzli et al., 2000).

The health implications of prolonged exposure to vehicle emissions are significant, encompassing a range of respiratory and cardiovascular diseases, oxygen deprivation, and long-term health effects from carcinogenic pollutants. Reducing vehicle emissions can substantially improve public health outcomes, including reduced disease incidence, lower mortality rates, better child health, and enhanced overall well-being. Addressing vehicle emissions is critical for protecting public health and ensuring a healthier future.

1.7 Research Problem

For a typical product, this supply chain extends from the source of raw materials through production and distribution systems to the point of consumption, including associated reverse logistics. Key logistical activities include freight transport, storage, inventory management, materials handling, and related information processing.

The primary objective of logistics is to coordinate these activities in a way that meets customer requirements at minimal cost. Traditionally, this cost has been considered in purely monetary terms. However, as environmental concerns grow, companies must also account for the external costs of logistics, which are primarily associated with climate change, air pollution, noise, vibration, and accidents (Tambovcevs and Tambovceva, 2012).

Integrating environmental considerations into supply chain management has become a growing concern for industry, government, and academic researchers. Supply chain managers are now expected to respond to the challenges posed by new legislation, standards, and regulations, as well as shifting customer demands and the need for efficiency, cost-effectiveness, and return on investment, all while adopting greener practices. The tension between business and environmental drivers is a critical challenge, but understanding this balance is essential for re-engineering and redesigning supply chains in a way that is both financially and environmentally sustainable (Dukovska et al., 2010).

Within supply chain management, two areas address environmental concerns: green supply chain management, which focuses solely on the environment, and sustainable supply chain management, which encompasses economic, social, and environmental dimensions. Seuring and Müller (2008) define sustainable supply chain management as "the management of material, information, and capital flows, as well as cooperation among companies along the supply chain while considering goals from all three dimensions of sustainable development—economic, environmental, and social—derived from customer and stakeholder requirements."

To achieve a sustainable supply chain, many companies and researchers have adopted various frameworks for measuring and capturing carbon emission data. However, some of these frameworks do not quantify emissions accurately (McKinnon, 2010). McKinnon also pointed out that, currently, no single standard for carbon emissions measurement exists, though the two main standards developed by the WBCSD/World Resources Institute (2004) and the International Standards Organization (ISO 14064) are broadly similar. Both set guidelines for carbon auditing in individual businesses and provide advice on calculation scope, data collection methods, and emissions allocation. McKinnon further noted that most carbon footprints reported by companies are underestimated and do not reflect the true amount of carbon produced (Alan C. McKinnon, 2010). Therefore, accurately quantifying carbon emissions remains a challenge in moving toward a truly sustainable supply chain.

1.8 Purpose of Research

Two main types of research contribute to sustainable supply chain management. One type focuses on technical aspects, such as improved tracking mechanisms and enhanced data collection capabilities. The other type explores applications in transportation, waste management, and retailing.

However, a crucial aspect of the research remains underexplored: leveraging real-time data to enhance the value of supply chain operations. For example, using real-time emissions data to optimize transport networks presents an opportunity to reduce operational costs, improve system efficiency, and achieve a better return on investment. By incorporating dynamic local variables like traffic conditions and the number of delivery stops, real-time data can provide a more accurate representation of the carbon footprint.

Integrating real-time data collection systems with other technologies, such as sensors and GPS, offers significant potential for enhancing logistics activities. This approach enables the efficient collection of relevant data to monitor vehicles and optimize supply chain operations. Combining these technologies allows for real-time feedback and improved logistics network efficiency, which can lead to better route planning based on traffic conditions, ultimately reducing emissions.

Through more accurate emissions data, these advanced technologies can also support carbon taxation or pricing mechanisms and incentivize organizations to improve their environmental performance. By providing real-time feedback and enabling reconfiguration and optimization of logistics systems, these technologies facilitate a deeper understanding of carbon emissions across the supply chain.

This research investigates how emerging technological systems can model complex logistics activities, capture emissions data, and optimize operations. Accurate carbon footprint quantification becomes essential in advancing supply chain sustainability and reducing environmental impacts.

1.9 Significance of the Study

The primary goal of this study is to enhance Canadian environmental policies regarding vehicle emissions. The study aims to provide policymakers with data-driven insights into the effectiveness of current regulations and the potential benefits of proposed measures, using machine learning models to identify patterns and trends. These insights can lead to more effective policy interventions targeting emission reduction and environmental impact mitigation. Additionally, the research can serve as a benchmark for other countries, promoting international collaboration and knowledge exchange in environmental policy-making.

The study emphasizes the potential of machine learning technologies in addressing complex environmental challenges, demonstrating how they can be used to analyze and predict vehicle-related environmental impacts. This highlights their broader applicability in other areas of environmental science, which can inspire innovation in developing new tools and applications that leverage machine learning for environmental monitoring, management, and mitigation. Furthermore, the study can encourage further research and development in green technology and drive investments in cleaner, more efficient transportation solutions.

The focus on developing a comprehensive framework for assessing and mitigating vehicle environmental impacts aligns with global efforts to combat climate change and protect natural resources (Dwivedi, Y.K et al, 2022). By providing actionable insights into the sources and consequences of vehicle emissions, the research supports the development of strategies to reduce the transportation sector's carbon footprint. This is particularly important for Canada, where transportation significantly contributes to national greenhouse gas emissions. The study's outcomes can help shape sustainable transportation policies that balance economic growth with environmental stewardship.

Reducing vehicle emissions has direct and significant benefits for public health, as vehicle pollutants are linked to various health problems, including respiratory and cardiovascular diseases (Kwan, S.C., et al, 2023). The study improves air quality and public health by informing policies to lower emissions, which can reduce healthcare costs and improve citizens' quality of life. The study also addresses noise pollution and its impact on health and well-being, further underscoring the holistic benefits of reducing vehicle-related environmental stewardship.

Finally, the study considers the equitable distribution of environmental benefits and burdens, ensuring policies do not disproportionately affect marginalized communities. By promoting transparency and accountability in policy-making, the research supports the development of fair and inclusive strategies that consider the needs and perspectives of all stakeholders.

1.10 Research Questions

To align with the Canadian government's initiative to reduce carbon emissions and promote a sustainable transportation infrastructure, this project aims to leverage machine learning (ML) techniques to effectively and precisely track carbon dioxide emissions at each stage of the supply chain process. The project will implement an intelligent system to optimize transportation and logistics networks, minimizing the carbon footprint throughout operations and supporting the country's broader environmental goals.

Following research questions are to be aimed to find answers.

1. How do emissions data (CO_2, NO_x, PM, HC) correlate with the overall environmental impact of different vehicle types?
2. What correlations exist between fuel efficiency, fuel type, and vehicle emissions?
3. Which machine learning algorithms are most effective for classifying vehicles based on their environmental impact?

CHAPTER II: REVIEW OF LITERATURE

2.1 Theoretical Framework

Past years have seen a remarkable number of weather-related disasters such as severed sea level rise, decrease of sea Ice, changes in rain patterns and an intense drought resulting to regular weather-related tragedies. These occurrences are associated to the rise of global mean temperature by 1° Celsius over the past 150 years (IPCC, 2018). This increase of temperature is due to high concentration of greenhouse gases in the atmosphere, which cause the atmosphere to retain heat. IPCC apportions 95% of the main cause to human activities. According to the World Bank report "Turn down the heat "It is projected that the mean global temperature can reached 4.0 C by the end of the century and 2C warmer by 2040 if no mitigation action is adopted to reduce carbon emissions. This situation underlines the need to urgently expedite action to minimize our emissions to reduce the effect of climate change. Due to the importance of this situation, various organizations such as Intergovernmental Panel on Climate Change (IPCC) and the United Nations Environment Program (UNEP) were set up to regulate emissions to help countries mitigate and adapt to climate change (Livermore, 2008). In addition, international climate change policy framework such as United Nations Framework Convention on Climate Change (UNFCCC) and Kyoto protocol were also enacted to encourage and commit countries to reduce their emissions (UN, 2008). A historic commitment was made by world leaders at the Paris agreement (COOP21) and COOP 22 at Marrakech to adopt strategies and policies which will encourage environmentally friendly activities within their countries.

As part of European union global commitment to climate change and an effort to reduce global warming, member states agreed to reduce their Greenhouse gas emission by a minimum of 20% before the end of 2020 and between 60 and 80% by the end of 2050 (Council of European Union 2007). Individual nations within the Eurozone seek to improve the global greenhouse gas mitigation efforts and implement climate change adaptation approaches by encouraging their companies to explore innovative and cost-effective ways to reduce their emissions.

In order to make informed carbon emission reduction decisions, it is important consider types of greenhouse gases, their sources and contribution to climate change. The most significant gas in terms of quantifiable emissions is carbon dioxide. It accounted for 76% of total GHG emissions in 2010. Since 1750, the carbon dioxide concentration in the atmosphere has increased by 40% (WMO2013); representing increased by about 100 ppm (parts per million) since the industrial revolution, this is due to irresponsible supply chain activities of companies that do not incorporate environmental dimensions into their operations. This is confirm by the UNCTAD reports in 2012 on climate chain which state that, Supply chain activities accounts for 76% of the total greenhouse emissions, hence reducing carbon emissions in supply chain activities are of paramount importance (UNCTAD 2012).

The planning and management of activities involved in sourcing, procurement, and conversion are critical for achieving environmental sustainability (Linton, Klassen and Jayaraman ,2007) Transportation plays a significant role in carbon emissions, contributing largely to global greenhouse gas output (Rossi et al., 2013). According to a 2022 report by the International Energy Agency (IEA), transportation is one of the main contributors to carbon emissions, with a 1.7% increase after the pandemic. To meet the target of Net Zero emissions by 2050, transportation emissions need to fall by 3%. In Canada, transportation alone accounts for approximately 59% of the country's total emissions (RFAE, n.d.), and it remains the only sector with a continuous upward surge in emissions since 1990. Ineffective transportation management exacerbates this issue, leading to higher emissions.

Efforts to minimize these emissions have been initiated globally, with many countries committing to reduce emissions under the Kyoto Protocol (Kuik, Nagalingam and Amer,2013). As a result, governments and industries are adopting strategies that help them make better, informed decisions to manage emissions efficiently. However, balancing compliance with environmental regulations while fulfilling commercial needs remains a significant challenge. Companies need to adopt innovative methodologies to effectively manage their carbon output (Carbon Trust, 2006). To minimize carbon emissions, it is crucial to accurately track and assess where the emissions originate and implement methodologies that not only measure the environmental impact of activities but also propose ways to reduce high-emission activities (Edwards-Jones and Jensen, 2010). Identifying these emission sources is vital in the broader effort to reduce overall carbon emissions and meet legislative and environmental goals.

As concern for the environment rises, firms are adopting environmentally friendly practices such as recycling (Jain and Sharma, 2014; Zhu and Sarkis, 2007), taking deliberate actions to implement strategies that will promote sustainable measures (D'Amato et al., 2009), improve their operations and restructure their business framework by integrating environmentally friendly logistics activities in their supply chain. In Supply chain management there are two areas which take environment into consideration, first is green supply chain which focusses on environment and economic dimensions and sustainable supply chain which focuses on environment, economic and social dimensions. (Seuring and Müller, 2008) define sustainable supply chain management as "the management of material, information and capital flows as well as cooperation among companies along the supply chain while taking goals from all three dimensions of sustainable development, i.e., Economic dimension, environmental dimension and social, taking into account which are derived from customers and stakeholder's requirements". Green logistics centred on how companies can incorporate environmentally friendly practices into logistics activities (Rodrigue et al., 2001).

Recently, government policies required companies to be accountable for their emissions and be responsible for their wastes (Gupta and Polimeni 2020) (Govinda et al.,2020). Green supply chain management (GSCM) is the appropriate means to extend the responsibility of business organizations from being mindful in reducing carbon emissions and waste and take full responsibility for their products lifecycle, from procurement to the final disposal (Zailani et al., 2012). However, the lack of appropriate means to effectively capture carbon emissions in the supply chain influences companies' capability to accurately quantify their carbon emissions.

Numerous frameworks and methodologies have been implemented to measure and capture carbon emission data, but accuracy remains a significant challenge due to the inability to account for dynamic local variables (Sanchez-Flores et al 2020). Factors such as traffic congestion, the number of delivery pickups and drops, and other operational complexities are often overlooked. Additionally, existing methodologies typically rely on industry-standard data, which does not fully capture emissions from all activities, such as moving goods within warehouses.

There is yet to be a universally accepted standard for carbon emission measurement. However, two significant frameworks—the Greenhouse Gas Protocol developed by the World Business Council for Sustainable Development (WBCSD) and the International Standards Organization (ISO 14067)—provide guidelines for carbon auditing. These frameworks offer advice on calculation scopes, data collection methods, and emission allocations. Yet, discrepancies remain, as highlighted by the Carbon Footprint for Freight Transportation (COFRET) European Commission report, which states that "various stakeholders have independently developed incomparable methods, tools, and data" (COFRET). McKinnon (2010) also emphasizes that many companies underestimate their carbon footprints, providing figures that do not accurately reflect the emissions produced. Addressing these gaps is crucial for improving carbon emission measurement and aligning with environmental goals.

2.2 Greenhouse Gas Emissions

Evaluating the environmental impact of vehicles requires careful consideration of greenhouse gas (GHG) emissions. These emissions significantly affect global climate change, impacting ecosystems, human health, and the economy (Das, 2017) . When vehicles burn fossil fuels like gasoline and diesel, they release greenhouse gases, with carbon dioxide (CO_2) being the most common. Methane (CH_4) and nitrous oxide (N_2O) are also emitted, albeit in smaller amounts. Methane can be released through exhaust, while nitrous oxide forms due to high combustion temperatures and pressures. Additionally, emissions occur during the production and distribution of fuel due to energy consumption in extraction, refining, and transportation processes.

The life cycle and manufacturing of vehicles also contribute to GHG emissions. This process requires significant energy and raw material extraction, producing substantial emissions. An inclusive evaluation of vehicle emissions comprises those from production, maintenance, and disposal. Notably, key greenhouse gases emitted by vehicles include CO_2, CH_4, and N_2O. While CO_2 is the most prevalent, CH_4 and N_2O have a much higher global warming potential (GWP). Over 100 years, methane is approximately 25 times more potent than CO_2, and nitrous oxide has a GWP about 298 times higher than CO_2.

The consequences of GHG emissions are extensive, contributing to global warming and resulting in more frequent and severe weather events such as hurricanes, droughts, and heat waves. GHG-induced climate change also leads to habitat loss, ecosystem disruption, and impacts on biodiversity and species survival. Elevated CO_2 levels cause ocean acidification, affecting marine life. Human health is affected by deteriorating air quality, resulting in respiratory and cardiovascular diseases and increased heat-related illnesses. Furthermore, climate change affects agricultural productivity and infrastructure, increasing maintenance and repair costs.

Mitigating GHG emissions from vehicles is a multifaceted task. However, there is optimism. Enhancing fuel efficiency through advanced engine technologies and lightweight materials can significantly reduce CO_2 emissions. The potential of alternative fuels, including electric vehicles (EVs) and biofuels, in reducing GHG emissions compared to traditional fossil fuels is promising. Implementing policy measures such as strict emissions standards and incentives for green technologies can encourage the adoption of cleaner vehicles. Additionally, expanding public transportation and promoting sustainable urban design can reduce dependence on personal vehicles, decreasing overall GHG emissions.

To conclude, addressing GHG emissions from vehicles is crucial in combating climate change and safeguarding environmental and human health. Understanding these emissions' sources, types, and impacts is vital for implementing effective mitigation strategies. This study seeks to develop a machine-learning framework to categorize vehicles based on their environmental impact, offering valuable insights for policymakers, manufacturers, and consumers to contribute to a more sustainable and resilient future.

Vehicles play a significant role in releasing greenhouse gases, mainly due to burning fossil fuels. Among the leading greenhouse gases released by vehicles are carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O). These gases originate from different sources and have varying levels of influence on the total greenhouse gas emissions stemming from the transportation sector.

1. Carbon Dioxide (CO_2): The most important greenhouse gas released by vehicles is carbon dioxide. Internal combustion engines produce it when they burn fossil fuels like gasoline and diesel. The quantity of CO_2 emitted is directly linked to the quantity of fuel used.

Primary Contributor: CO_2 is the largest source of GHG emissions from vehicles, accounting for approximately 90-95% of the total GHG emissions in the transportation sector (International Energy Agency, 2021).

Impact on Climate Change: CO_2 has a long atmospheric lifetime and contributes significantly to the greenhouse effect, leading to global warming and climate change (Pachauri & Meyer, 2014).

2. Methane (CH_4): Methane has a much higher global warming potential (GWP) than CO_2 and is a potent greenhouse gas. Vehicles emit methane in smaller amounts, mostly during fuel extraction, production, and distribution and from the combustion of natural gas in vehicles equipped with natural gas engines (Weger et al., 2017).

Minor Contribution: Although methane emissions from vehicles are relatively low compared to CO_2, they still represent a significant source due to their high GWP (approximately 25 times that of CO_2 over 100 years) (IPCC, 2013).

Impact on Climate Change: Methane contributes to both short-term and long-term climate change due to its potency and ability to trap heat in the atmosphere.

3. Nitrous Oxide (N_2O): Nitrous oxide, a potent greenhouse gas, is released from vehicles. The primary source of this gas is the catalytic converters found in vehicle exhaust systems, especially in high-temperature conditions (Brinklow et al, 2023).

Minor Contributor: N_2O emissions from vehicles are relatively small in volume compared to CO_2 emissions, but they have a very high GWP (approximately 298 times that of CO_2 over 100 years) (IPCC, 2013).

Impact on Climate Change: Despite its lower emissions volume, N_2O's high GWP significantly contributes to the greenhouse effect and global warming.

- Contribution of Different Vehicle Types to Overall Emissions

1. Passenger Cars: Passenger vehicles, the most abundant vehicle type, play a substantial role in CO_2 emissions because of their extensive usage and dependence on gasoline and diesel fuels (Fontaras et al, 2017).

Primary Contributor: Passenger cars are responsible for a substantial portion of the total GHG emissions from the transportation sector. In many developed countries, they account for around 50-60% of the sector's CO_2 emissions (US EPA, 2021).

2. Light-Duty Trucks and SUVs: Light trucks and SUVs usually have larger engines and use more fuel than regular cars, leading to larger CO_2 emissions per vehicle (Hagos et al, 2018).

Increasing Contribution: The growing popularity of SUVs and light-duty trucks has increased their share of total vehicle emissions. They now contribute significantly to overall GHG emissions, often accounting for about 20-30% of the transportation sector's CO_2 emissions (International Council on Clean Transportation, 2020).

3. Heavy-Duty Trucks and Buses: Trucks and buses with strong carrying capacities produce substantial amounts of CO_2, CH_4, and N_2O as they extensively utilize diesel fuel owing to their large dimensions and heavy loads (Hagos et al, 2018).

High Emissions Intensity: While they represent a smaller portion of the vehicle fleet, heavy-duty trucks and buses contribute disproportionately to GHG emissions, often accounting for 20-30% of the sector's total emissions (International Energy Agency, 2021).

4. Non-Road Vehicles: Non-road vehicles, including construction equipment, agricultural machinery, and other off-road vehicles, also contribute to GHG emissions, primarily diesel and gasoline (Quiros et al, 2017).

Notable Contributor: These vehicles are responsible for a smaller, yet significant, share of the total transportation sector emissions, contributing around 10-15% of the GHG emissions (US EPA, 2021).

The primary sources of greenhouse gas discharges from automobiles stem from burning nonrenewable resources, resulting in the release of CO_2, CH_4, and N_2O. Of these, CO_2 has the most notable impact, making up the majority of GHG emissions from the transportation industry. Even though methane and nitrous oxide are released in smaller amounts, their high global warming potentials deem them significant factors in climate change. Passenger cars, light-duty trucks and SUVs, heavy-duty trucks and buses, and non-road vehicles substantially affect the overall emission profile. Passenger cars and light-duty vehicles are the primary contributors owing to their prevalence and consumption patterns (Wang et al 2021).

2.3 Air Pollution and its Effects

The main contributors to air pollution are vehicle emissions, industrial operations, power plants, agricultural practices, and natural sources. In urban areas, the most significant sources of air pollution are usually vehicles, which release pollutants like nitrogen oxides (NO_x), carbon monoxide (CO), particulate matter (PM), and volatile organic compounds (VOCs). This is especially true in cities with heavy traffic and older, less efficient vehicles. Furthermore, industrial activities in urban areas, such as manufacturing and processing plants, release substantial amounts of sulfur dioxide (SO_2), NO_x, and PM into the air (Arshad, K, et al, 2024).

Power plants, particularly those burning fossil fuels such as coal and natural gas, are notable sources of air pollution, emitting large quantities of SO_2, NO_x, CO_2, and PM. This is especially prevalent in areas with a high concentration of power generation facilities and heavy reliance on fossil fuels for energy production. Agricultural practices also contribute to air pollution through pesticides and fertilizers, which release ammonia (NH_3) and other chemicals. Livestock farming produces methane (CH_4) and ammonia, adding to the overall burden of air pollutants (Filonchyk, M. et al 2023).

Natural sources of air pollution include wildfires, volcanic eruptions, and dust storms. Wildfires release significant amounts of PM, CO, and VOCs, affecting air quality over expansive areas. Volcanic eruptions emit sulfur dioxide, ash, and other particulates, while dust storms contribute to high PM levels, particularly in arid regions (Nakhjiri, A, et al, 2024)

Understanding the distribution of these sources in different regions and urban settings is crucial, as it depends on factors such as population density, industrial activity, transportation infrastructure, and geographical and climatic conditions. This understanding helps develop effective strategies to reduce air pollution and enhance air quality. For example, urban areas with dense populations and heavy traffic tend to have higher levels of vehicle emissions, while industrial regions may experience more pollution from manufacturing and power generation. Rural areas with extensive agricultural activities may have elevated ammonia levels and other agricultural pollutants (Li et al 2020).

Geographical factors such as topography and climate also play a crucial role in the dispersion and concentration of air pollutants. Cities in valleys or areas with limited air circulation may experience higher pollution levels due to the trapping of pollutants. Additionally, climatic conditions such as temperature inversions can lead to the accumulation of pollutants near the ground, worsening air quality problems (Khan et al 2024).

In summary, the primary sources of air pollution are vehicle emissions, industrial activities, power plants, agricultural practices, and natural occurrences. These sources vary across regions and urban environments due to population density, industrialization, transportation systems, and geographical and climatic conditions. Understanding these variations is crucial, as it guides the development of targeted strategies to reduce air pollution and improve air quality in diverse settings, underscoring the necessity for such measures (Borck, R., et al, 2021).

Vehicle emissions, a widespread concern, play a substantial role in air pollution, particularly in densely populated urban areas. Cars, trucks, buses, and motorcycles release harmful pollutants such as nitrogen oxides (NOx), carbon monoxide (CO), volatile organic compounds (VOCs), and particulate matter (PM). Cities like Los Angeles, Beijing, and New Delhi epitomize the severe air quality problems many urban areas encounter, primarily due to the large volume of vehicles and heavy traffic. The situation is often exacerbated in developing countries where rapid urbanization and increased vehicle ownership outpace improvements in vehicle maintenance and fuel quality (Kumar et al., 2015; World Health Organization, 2016).

Industrial activities represent a significant global source of air pollution that must not be overlooked. Factories, power plants, and mining operations emit substantial quantities of sulfur dioxide (SO2), nitrogen oxides (NOx), particulate matter (PM), and other hazardous chemicals. Regions with intense industrial activity, such as the Ruhr Valley in Germany, parts of China, and the Midwestern United States, face elevated pollution levels from these sources. Countries undergoing rapid industrialization, like India and Vietnam, are witnessing a sharp rise in industrial emissions, exacerbating air quality issues (Guttikunda & Goel, 2013; UNEP, 2019).

Residential heating and cooking are notable sources of air pollution, particularly in colder climates and developing nations. The combustion of solid fuels like wood, coal, and biomass produces significant amounts of particulate matter (PM) and carbon monoxide (CO). In colder regions such as Eastern Europe and Northern China, using these fuels for heating during winter leads to marked air pollution. Additionally, in many developing countries, traditional cooking biomass stoves contribute heavily to indoor and outdoor air pollution (Gurjar et al., 2010; Bonjour et al., 2013).

Natural events such as wildfires, volcanic eruptions, and dust storms also significantly contribute to air pollution. Areas prone to wildfires, like California, Australia, and parts of the Mediterranean, frequently experience spikes in pollution levels due to these events. Similarly, regions near active volcanoes, such as Iceland and Indonesia, can suffer substantial air pollution when eruptions occur (Johnston et al., 2012; Schneider & Kyle, 2014).

To conclude, air pollution originates from diverse sources, and its severity and composition can vary significantly based on the region and local activities. Understanding these sources is critical for developing effective strategies to combat air pollution and safeguard public health.

Reducing air pollution requires effective policies and interventions at local, national, and global levels. At the local level, cities can play a crucial role in curbing vehicle emissions, a significant source of air pollution. For example, implementing Low Emission Zones (LEZs), where only vehicles that meet strict emission standards can enter, has proven successful in cities like London. Additionally, cities can invest in developing efficient public transportation systems, such as bus rapid transit (BRT) systems, which have significantly reduced traffic congestion and pollution in places like Curitiba, Brazil (Goodman et al., 2007).

Nationally, governments can enforce comprehensive regulations to limit emissions from various sources. For example, the United States enforces the Clean Air Act, which sets national air quality standards for pollutants like sulfur dioxide (SO2), nitrogen oxides (NOx), and particulate matter (PM). These regulations help reduce industrial emissions and improve air quality (US EPA, 2021). Furthermore, national policies can promote the transition to renewable energy sources. Germany's Renewable Energy Sources Act (EEG) has been pivotal in increasing the share of renewable energy in the country's energy mix, considerably reducing reliance on fossil fuels and associated air pollution (German et al. for Economic Affairs and Energy, 2020).

On a global scale, international cooperation is essential for addressing air pollution. The urgency of the issue is underscored by the Paris Agreement, a prime example of countries coming together to commit to reducing greenhouse gas emissions. This global effort indirectly benefits air quality by curbing pollutants that contribute to global warming. International organizations like the Green Climate Fund provide crucial financial support to developing countries, helping them transition to cleaner energy sources and implement sustainable practices that reduce air pollution.(https://unfccc.int/process-and-meetings/the-paris-agreement)

Collaboration and coordination across all levels of government and with various stakeholders are crucial to ensure the effective implementation of these policies and interventions. Local governments should engage communities and educate the public about the benefits of reducing air pollution, garnering public support for necessary measures. National governments must enact clear legislation and provide adequate funding and incentives to support the transition to cleaner technologies and practices. Internationally, countries must work together on agreements that set ambitious targets and include monitoring and reporting progress mechanisms. As part of these various stakeholders, the audience plays a crucial role in this collective effort.

2.4 Current Policies in Canada

2.4.1 Overview of Existing Policies

- Canadian Environmental Protection Act (CEPA):

The Canadian Environmental Protection Act (CEPA) is crucial to Canada's environmental laws. It controls different pollutants, such as those emitted by vehicles, to keep people healthy and protect the environment. CEPA allows the government to set national rules for pollutants like nitrogen oxides (NOx), sulfur dioxide (SO2), particulate matter (PM), and greenhouse gases (GHGs). These regulations ensure that vehicles sold and used in Canada meet specific emission limits, reducing their overall environmental impact.(https://www.canada.ca/en/services/environment/pollution-waste management/understanding-environmental-protection-act.html)

- Clean Fuel Standard (CFS):

The Clean Fuel Standard (CFS) seeks to decrease the carbon intensity of fuels utilized in Canada, consequently diminishing the life cycle emissions linked with transportation fuels. The CFS advocates for using more environmentally friendly fuels, such as biofuels and low-carbon electricity, while endorsing the adoption of electric vehicles (EVs) and hydrogen fuel cell vehicles. By establishing standards based on performance, the CFS motivates fuel producers and suppliers to innovate and invest in cleaner technologies. (https://ecology.wa.gov/air-climate/reducing-greenhouse-gas-emissions/clean-fuel-standard)

- Zero-Emission Vehicle (ZEV) Mandate:

Provinces such as British Columbia and Quebec have implemented Zero-Emission Vehicle (ZEV) mandates. These mandates require a certain percentage of new vehicle sales to consist of zero-emission vehicles. They set specific targets and timelines for increasing the market share of electric and hydrogen fuel cell vehicles. For instance, British Columbia's ZEV Act mandates that 10% of new light-duty vehicle sales be zero-emission by 2025, increasing to 30% by 2030 and reaching 100% by 2040. These targets aim to speed the transition to cleaner transportation and reduce greenhouse gas emissions (Aubertin, A., et al 2024).

- Incentives and Rebates:

The federal government and several provinces offer financial incentives and rebates to support the adoption of cleaner vehicles. The federal Zero Emission Vehicle (ZEV) program provides rebates of up to $5,000 to purchase new electric or hydrogen fuel cell vehicles. Provincial incentives vary, with some provinces offering additional rebates, tax credits, or reduced registration fees for zero-emission vehicles. These incentives help offset the higher upfront costs of cleaner vehicles, making them more accessible to consumers (Thorne, Z. et al, 2019).

- Emission Standards:

Canada has strict vehicle emission standards that align with those of the United States and Europe. These standards control the amount of pollutants vehicles release, ensuring that new vehicles are more environmentally friendly and effective. For instance, Canada's Tier 3 vehicle and fuel standards, implemented in 2017, necessitate significant decreases in NOx, SO_2, and PM emissions from light-duty vehicles. These standards aim to decrease pollutants that contribute to smog and enhance air quality (Long et al., 2020)

2.5 Integrating Green Practices into Vehicle Emission Management

Green practices in vehicle emission management have become essential for mitigating the transportation sector's environmental impact. One prominent method is the promotion of eco-driving and the use of green-safety devices. These systems encourage efficient driving behaviours, which can significantly reduce fuel consumption and emissions. For example, a study found that the implementation of on-board green safety devices resulted in a reduction of fuel consumption by up to 6%. Nitrogen oxide (NOx) emissions were reduced by up to 65% for inexperienced drivers. In the case of experienced drivers, the improvements were also substantial, with reductions in NOx emissions of up to 56%. This demonstrates the efficacy of relatively low-cost interventions in managing vehicle emissions (Ng et al., 2021).

Intelligent fleet management systems are another effective strategy for emission reduction. These systems incorporate advanced technologies like vehicle-to-infrastructure (V2I) communication to provide real-time driving support and eco-driving guidance. These systems can influence driver behaviour and reduce emissions by managing traffic in real-time and optimizing vehicle routes. Research on such systems has shown promising results, especially in reducing emissions and fuel consumption for heavy vehicle platooning, which allows for coordinated and efficient vehicle movements in fleets (Ma, 2013).

Vehicle routing optimization also significantly reduces emissions, particularly in the context of fleet operations. By considering factors such as speed, congestion, and time-dependent emissions, optimized routing models can significantly lower fuel consumption and CO_2 emissions. One study found that using a multi-depot vehicle routing system, as opposed to a single-depot approach, reduced carbon emissions by up to 37.6%. This illustrates the importance of logistical optimizations in managing emissions in the transportation sector (Wang et al., 2022).

Additionally, vehicle maintenance and management systems reduce emissions by ensuring vehicles operate efficiently. Proper maintenance systems, often enhanced with GPS tracking and automated scheduling, help avoid performance delays and optimize vehicle operation. Such systems have been shown to significantly reduce CO_2 emissions by maintaining vehicles in peak condition and preventing inefficiencies that could lead to higher emissions (Hettigamage et al., 2013).

Combining green practices, including eco-driving, intelligent fleet management, optimized routing, and rigorous vehicle maintenance, dramatically reduces overall vehicle emissions. These strategies lower greenhouse gas emissions and promote cost savings through reduced fuel consumption, demonstrating their importance in sustainable transportation management.

Integrating green practices into vehicle emission management faces several significant barriers, many of which stem from economic, technological, regulatory, and social challenges. A key barrier is the high initial cost of green technologies such as electric or hybrid vehicles. Both consumers and businesses are often hesitant to invest in these technologies due to the upfront expenses, despite the long-term benefits of reduced emissions and fuel savings (Hebaz & Oulfarsi, 2021). In particular, small and medium-sized enterprises (SMEs) in developing countries need more financial resources and funding for research and development (Karuppiah et al., 2020). Overcoming this barrier requires government incentives such as subsidies, tax credits, and grants to offset the initial investment costs, encouraging wider adoption of green vehicle technologies.

Another major barrier is the lack of infrastructure, particularly for electric vehicles (EVs). The absence of sufficient charging stations and other necessary support facilities hampers the integration of green vehicle practices. This issue is compounded by logistical challenges, such as the limited range of EVs and the need for efficient routing systems to minimize fuel consumption and emissions (Remy et al., 2012). To overcome these infrastructural challenges, it is imperative for governments and private sectors to collaborate. This collaboration is necessary to expand the EV charging network, invest in renewable energy sources for these stations, and improve the overall transportation infrastructure to support green vehicle technologies.

Regulatory and policy gaps also significantly slow down the adoption of green practices in vehicle emission management. Inconsistencies between environmental goals, such as fuel efficiency and emission standards, and other social objectives, such as vehicle safety and consumer preferences for larger vehicles, create obstacles to effective policy implementation (MacLean & Lave, 2003). Furthermore, some regions' lack of stringent regulations or enforcement mechanisms reduces the pressure on companies and individuals to adopt cleaner vehicle technologies. Developing clear, cohesive policies that align emission reduction targets with broader economic and social goals can help mitigate this barrier.

Another issue is consumer awareness and behavior. Many consumers remain unaware of the environmental impact of their vehicle choices, and even when informed, they may prioritize short-term convenience or cost over long-term environmental benefits. For example, consumers in certain regions prefer large, fuel-intensive vehicles, which contradicts the goals of reducing emissions (MacLean and Lave, 2003). To address this, it is crucial to launch public awareness campaigns. These campaigns, combined with incentives for consumers who opt for greener alternatives, can shift consumer preferences towards low-emission vehicles.

In summary, overcoming the barriers to integrating green practices into vehicle emission management will require a multi-faceted approach. Financial incentives, improved infrastructure, cohesive regulatory frameworks, and enhanced public awareness are all crucial in facilitating the transition to greener vehicle technologies.

"As integrating environment thinking into supply chain management, including product design, material sourcing and selection, manufacturing processes, delivery of the final product to the consumers, and end-of-life management of the product after its useful life (Srivastava, 2007)"

Other authors like Hervani et al. (2005) propose that Green Supply chain boundary starts from material acquisition, they further describe it as a concept of minimizing waste through the adoption of environmentally conscious practices.

"Green purchasing + green manufacturing/Materials Management + Green Distribution /Marketing + Reverse logistics".

(KAFA et al., 2013.) built on Hervani equations and indicated that GSCM is not achievable if it does not incorporate eco-design. They therefore describe.

Green supply Chain management (GSCM) = Green Purchasing (Min and Galle, 1997) + Eco-Design (Linton, et al., 2007) + Green Manufacturing (Deif, 2011) + Green Distribution (Rao and Holt, 2005) + Reverse Logistics (Pochampally, et al., 2009).

"At the intersection of social, environmental, and economic performance are activities that an organization can engage in which not only are beneficial from a social and environmental standpoint, but that also make economic sense and result in competitive advantage for the firm". This definition is conceptualized in figure 1 below.

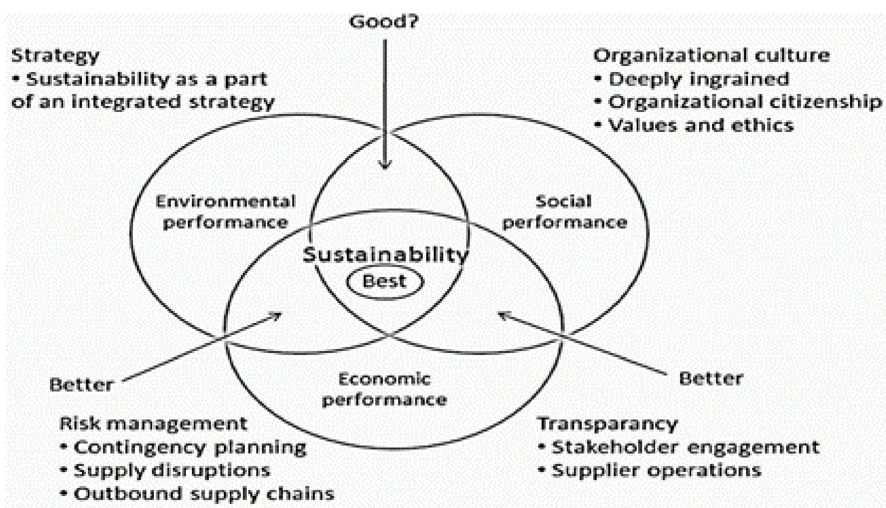

Figure 3 Elkington triple bottom line adapted from Adapted from Carter and Rogers

Elkington definition is based on (WCED 1987) Brundtland commission, which described sustainability. …As utilizing resources to meet the needs of the present without compromising future generations' ability to meet their own needs (WCED, 1987).

Based on the above definition, true sustainability can only be possible if the environmental, social, and economic dimensions are fully integrated.

The economic dimension involves the trade-off of minimizing total logistic cost and maximizing profit over supply chain activities (purchasing, production, warehousing, distribution, recycling, etc.), while the environmental aspects means that permanent environmental damages should not be allowed, and greenhouse Gas (GHG) emissions regulations are enforced. The social dimension includes objective such as the reduction of noise, traffic congestion, improve the standard and quality of life of communities around the supply chain (Hanuv, 2010; Mann et al., 2010).

Implementing green supply chain management (GSCM) practices not only helps organizations reduce their carbon footprint but also enables them to remain competitive in a rapidly changing marketplace (Lin & Shang 2020; Hosseini et al., 2018). Furthermore, GSCM promotes the efficient use of resources, reduces waste generation, and fosters collaboration among supply chain partners to create a more sustainable future for all (Pramesti et al., 2021; Al-Khawaldah et al., 2022).

2.6 Sustainable Supply Chain Management

One of the key principles of Green Supply Chain Management is the adoption of eco-friendly practices at every stage of the supply chain (Chatzoudes and Chatzoglou, 2022). This involves selecting environmentally responsible suppliers (Bowen et al., 2009), promoting energy-efficient production processes, optimizing transportation to reduce emissions, and implementing waste reduction and recycling initiatives (Teixeira et al., 2018). GSCM also emphasizes the importance of transparency and accountability throughout the supply chain. By working closely with suppliers and stakeholders (Yang & Lin 2020), companies can ensure that sustainable practices are being followed and that the entire supply chain is aligned with environmental goals (Chan et al., 2012). Additionally, GSCM encourages collaboration and knowledge-sharing among industry peers to drive innovation and continuous improvement in eco-friendly practices (Marchesini and Lizarelli, 2018). Ultimately, the goal of GSCM is to create a more sustainable and resilient supply chain that not only benefits the environment but also enhances the overall reputation and profitability of the company (Bu et al., 2020).

The adoption of Green Supply Chain Management offers numerous benefits to businesses. First and foremost, it helps companies comply with environmental regulations and standards, reducing the risk of legal and reputational issues (Kenneth et al., 2012). Moreover, GSCM can lead to cost savings in the long run. For instance, by reducing energy consumption and waste, companies can lower operational expenses and enhance resource efficiency (Rakesh and Ravi, 2015).

Despite the numerous advantages, implementing Green Supply Chain Management can present challenges (Roberto et al., 2014). Companies may encounter resistance from suppliers or face higher upfront costs when adopting eco-friendly technologies. Furthermore, tracking the environmental impact of the entire supply chain can be complex due to the involvement of multiple stakeholders and processes. However, with the growing emphasis on sustainability and the potential for competitive advantage, many companies are recognizing the value of GSCM and investing in greener supply chain practices (Balan and Prakash, 2016).

Green Supply Chain Management is a holistic approach that aims to align supply chain operations with environmental sustainability goals. By integrating eco-friendly practices at every stage of the supply chain and fostering collaboration among stakeholders, GSCM can lead to cost savings, improved brand reputation, and a reduced carbon footprint.

One common denominator identified across all definitions of green supply chain management is the integration of green operations into the traditional supply chain. The green operations can be broken down into green logistics, which includes green transport, re-manufacturing, reverse logistics, green packaging, and green warehousing (implementing environmental conscious activities in the warehouse). Green Supply chain can simply be defined as integrating environmentally friendly practices into logistics operations.

2.7 Carbon Footprints

The concept of carbon footprint, a crucial tool in addressing the urgent need to reduce emissions and mitigate the effects of climate change, has gained significant attention in recent years (Fuller, 2017). A carbon footprint measures the total greenhouse gas emissions, primarily carbon dioxide (CO2), produced directly or indirectly by human activities. By understanding and measuring our carbon footprint, individuals, businesses, and governments can identify areas where emissions can be reduced and make more sustainable choices (Edstrand, 2016). This includes adopting renewable energy sources, promoting energy efficiency, and implementing policies that encourage low-carbon practices (Strohbach et al., 2012). Tracking carbon footprints allows for accountability and transparency in emission reduction efforts, enabling progress to be monitored and evaluated (Jan, 2008). A comprehensive understanding of carbon footprints is essential to creating a more sustainable and environmentally conscious future. By understanding carbon footprints, individuals and organizations can make informed decisions about their consumption patterns and make changes to reduce their overall emissions.

Despite its extensive use, the term carbon footprint has a variety of definitions, and researchers are still debating on the gases to include in the definition (Wiedmann & Minx, 2008). Ana and Arroja (2011) describe carbon footprint as the amount of carbon emissions along a supply chain from the procurement of raw materials to the delivery of the finished products to the final consumers. It aims to identify activities along the supply chain that generate high emissions and take necessary action to reduce them. Ana and Arroja's definition includes carbon emission from raw material acquisition to the final consumer; it does not consider the product's life cycle.

Parliamentary Office of Science and Technology (POST) built on this definition and proposed that carbon footprints should incorporate carbon dioxide emissions and other greenhouse emissions emitted over the product's entire life cycle (POST, 2006). Wiedmann and Minx (2008) agree with POST definition but suggested that the term climate footprint should be used instead of greenhouse gases if they want to consider all the six Kyoto protocol gases (carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), and sulphur hexafluoride (SF6). Based on this argument, they define carbon footprint as a measure of the total amount of carbon dioxide emissions that are directly and indirectly caused by an activity or accumulated over the life stages of a product. To avoid the confusion surrounding the gas to be included, Carbon Trust proposes the definition of carbon footprints as:

A technique for identifying and measuring the individual greenhouse gas emissions from each activity within a supply chain process step and the framework for Methodology to estimate the total emission of greenhouse gases (GHG) in carbon equivalents from a product across its life cycle from the production of raw material used in its manufacture to disposal of the finished product (excluding in-use emissions) attributing these to each output product "(Carbon Trust, 2007).

Craig (2012) proposes that term six of the Kyoto Protocol should be used to have a uniformly acceptable definition. He, therefore, defines a product's carbon footprint as including the six Kyoto Protocol gases accumulated over all its life stages. This definition suits ISO 14067 definitions of a product's carbon footprint.

"The sum of greenhouse gas emissions and greenhouse gas removals of a product system, expressed in CO2 equivalent" This definition is centred on the product life cycle of the product; therefore, the need to include all the activities from the material acquisition to the product's end-of-life. The COFRET report defines Carbon footprint methodology as "Standards, standard-like guidelines, guidebooks and schemes that provide the framework for calculating and reporting the carbon footprint of transport and logistics along the supply chain or part it".

There are challenges and limitations involved in collecting and checking the carbon footprint of suppliers, such as the data on fuel use in third-party transport providers need to be provided. Such information is needed for energy-based calculations, as fuel consumption is a significant part of the cost structure of truck transportation, and the carrier typically wants to keep the information on the cost to the company; thus, there is no accurate calculation of the emissions. In order to address such challenges, the company can use activity-based calculation instead (Bouchery Y. et al., 2017).

2.8 Enhanced Vehicle Emission Management

Advanced technologies like artificial intelligence (AI) and machine learning (ML) are increasingly being applied to improve vehicle emission management systems by enhancing predictive modelling, real-time monitoring, and fuel consumption and emissions optimisation. One significant application is in the predictive modelling of engine emissions. ML algorithms can forecast pollutant emissions, such as nitrogen oxides (NOx), by analyzing vehicle data under real-world driving conditions. Predictive models, such as artificial neural networks (ANNs) and non-linear regression models, can estimate NOx emissions with high accuracy, reducing the need for expensive, repeated physical testing

(Le Cornec et al., 2020). These models can be integrated into vehicle systems to enable real-time emission predictions, helping policymakers develop better emission inventories and mitigation strategies.

In energy management, AI-based systems are also helpful in optimizing fuel consumption and reducing emissions in hybrid and electric vehicles. Machine learning algorithms, including reinforcement learning, manage the power split between combustion engines and electric motors in hybrid vehicles, resulting in significant fuel savings and emission reductions (Hofstetter et al., 2019). These intelligent systems optimize energy consumption in real time by learning from different driving conditions and applying the most efficient energy management strategies.

AI technologies have also found applications in predictive maintenance for heavy-duty vehicles. By analyzing vast amounts of data from engine sensors, ML algorithms can detect early signs of engine malfunctions or inefficiencies that could lead to higher emissions. This allows for timely maintenance and repair, preventing unnecessary increases in fuel consumption and emissions (Katreddi et al., 2022). Moreover, AI-powered virtual sensors replace traditional sensors to estimate emissions like NOx in real-time, providing more accurate data without costly physical sensors (Falai & Misul, 2023).

AI and ML are revolutionizing vehicle emission management by improving predictive modelling, optimizing energy usage in hybrid vehicles, and enabling predictive maintenance. These technologies enhance the accuracy and efficiency of emission management systems and contribute to reducing overall vehicle emissions (Pinkse et al., 2014).

Vehicle manufacturers play a crucial role in developing and implementing enhanced emission management technologies. As both creators and regulators of these technologies, manufacturers work closely with governments and policymakers to meet increasingly stringent emission standards while navigating market demands and technological innovation (Hofstetter et al., 2019).

One of the critical contributions of manufacturers is the development of low-emission vehicle technologies such as electric, hybrid, and hydrogen fuel cell vehicles. Automakers like Toyota, Daimler, and General Motors have been at the forefront of developing these technologies in response to regulatory pressures and growing consumer demand for environmentally sustainable options. These companies have leveraged public-private partnerships, government incentives, and private investments to create protected spaces for experimenting with low-emission technologies, accelerating their development and making them more attractive to mainstream consumers (Pinkse et al., 2014).

Manufacturers also play an instrumental role in incremental technological innovation. For example, the automotive industry's shift from gasoline to diesel engines for fuel efficiency and lower CO2 emissions required manufacturers to address trade-offs such as the increase of NOx emissions. To manage these challenges, manufacturers have introduced selective catalytic reduction (SCR) and emission control technologies that significantly reduce pollutants. This highlights their role in technological development and balancing environmental concerns with performance goals (Bauner et al., 2009).

Another critical contribution of manufacturers is the implementation of on-board monitoring systems. Original Equipment Manufacturers (OEMs) have developed sophisticated emission monitoring technologies that track real-time data on NOx concentrations and other pollutants. These systems allow manufacturers to ensure vehicles comply with emission regulations throughout their lifecycle. This approach has been critical in countries like China, where real-time monitoring of in-use diesel trucks has led to significant improvements in emission control (Zhang et al., 2020).

Manufacturers also collaborate with suppliers and invest in research and development (R&D) to advance emission control technologies. Studies have shown that while component suppliers initially drove innovation in emission controls, manufacturers have increasingly taken the lead by integrating architectural and component knowledge, especially under regulatory pressures, to meet stricter standards (Lee et al., 2010). Through R&D, manufacturers ensure continuous improvements in emission technologies and their integration into mainstream vehicle production.

In summary, vehicle manufacturers are central to technological innovation for emission management. They develop and implement advanced technologies in collaboration with regulators, invest in incremental and radical innovations, and deploy sophisticated monitoring systems, all while balancing performance and environmental sustainability.

Several key factors influence the effectiveness of enhanced vehicle emission management practices. One of the most significant factors is traffic management and driving behaviour. Research shows that traffic signal coordination and congestion management can reduce emissions, which are highest during vehicle acceleration and tend to decrease during steady cruising, deceleration, and idle modes (Unal et al., 2003). Therefore, managing traffic flow and reducing stop-and-go driving can significantly lower emissions.

Another critical factor is vehicle deterioration and technological advancements. As vehicles age and accumulate mileage, their emissions tend to increase, which reduces the effectiveness of emission management systems. Regular inspection and maintenance programs ensure that older vehicles meet emission standards. Additionally, advances in emission control technologies, such as selective catalytic reduction and diesel particulate filters, play a critical role in maintaining low emissions even as vehicles age (Zhang et al., 2017).

The quality and type of fuel used also have a significant impact on emissions. The sulfur content of fuel directly influences vehicle emissions. For example, lower sulfur fuels enable more effective functioning of after-treatment devices, leading to reduced emissions of nitrogen oxides and particulate matter. Countries that have adopted high fuel quality standards, such as low-sulfur fuels, have experienced substantial reductions in emissions

(Wu et al., 2017). This underscores the importance of fuel quality standards in emission management.

Government regulations and policies are another major factor influencing the effectiveness of vehicle emission management practices. Stringent emission standards and regulations, such as the Euro 6 and China 6 standards, have proven effective in reducing emissions from both new and in-use vehicles. These regulations, which include real-world emission testing protocols, incentivize manufacturers to develop cleaner technologies and encourage vehicle owners to maintain their vehicles properly (Wu et al., 2017). This underscores the significant role of government policies in emission management.

Lastly, human factors and driving behaviour are crucial. Eco-driving practices involving smoother acceleration and deceleration can significantly lower fuel consumption and emissions. Studies have shown that green-safety devices can improve driving behaviour and reduce fuel consumption by up to 6% and nitrogen oxide emissions by up to 65%, especially for inexperienced drivers (Ng et al., 2021).

In conclusion, the effectiveness of vehicle emission management practices depends on traffic management, vehicle technology, fuel quality, government regulations, and driving behaviour. These factors must be addressed collectively to achieve significant and lasting reductions in vehicle emissions.

CHAPTER III: METHODOLOGY

3.1 Overview of the Research Problem

To address the environmental impact of vehicles and their effects on air quality, human health, and climate change. It is clear that transportation significantly contributes to global greenhouse gas emissions and air pollutants. In Canada, vehicles play a substantial role in the release of carbon dioxide (CO2), nitrogen oxides (NOx), particulate matter (PM), and volatile organic compounds (VOCs). These pollutants not only contribute to global warming but also lead to severe health problems like respiratory and cardiovascular diseases. Addressing the environmental impact of vehicles is crucial for public health policy and climate change mitigation strategies.

Despite implementing various policies and measures to reduce vehicle emissions, significant challenges still need to be addressed. One prominent issue is the accurate classification of vehicles based on environmental impact. Traditional methods of classifying vehicles often rely on simplistic criteria like fuel type or engine size, which fail to capture emissions' complexity and effects. This can result in inefficiencies in policy implementation and regulatory framework gaps (Aminzadegan, S., et al, 2022).

Furthermore, the rapid advancement of technology in the automotive industry, including the development of electric and hybrid vehicles, presents both opportunities and challenges. While these technologies have the potential to significantly reduce emissions, their environmental benefits depend on various factors such as the source of electricity and the lifecycle emissions of the vehicles. Therefore, a more nuanced and comprehensive approach to classifying vehicles based on their environmental impact is crucial and needed (Shahzad, K, et al, 2024).

In this context, machine learning stands out as a promising solution. Leveraging large datasets and advanced algorithms, machine learning can identify complex patterns and relationships that traditional analysis methods might miss. This can lead to more accurate and detailed classifications of vehicles based on their environmental impact. However, the application of machine learning in this context requires careful consideration of various factors, including data quality, model selection, and interpretability of results (Mobarak M.H., et al, 2023).

In Canada, existing policies and regulations aim to reduce vehicle emissions through emission standards, fuel quality regulations, and incentives for electric vehicles. While these policies have seen some success, their effectiveness is limited by the challenges mentioned above. Therefore, a robust framework incorporating machine learning approaches could enhance the effectiveness of these policies by providing more precise and actionable insights into vehicle emissions (Qadir, S.A., et al, 2024).

The research problem revolves around developing a comprehensive framework for classifying vehicles based on their environmental impact using machine learning techniques. This involves identifying the key variables influencing emissions, performing exploratory data analysis to understand the data, and developing predictive models to classify vehicles accurately. By addressing this problem, the research aspires to provide policymakers with better tools to design and implement more effective environmental regulations, ultimately leading to improved air quality and reduced greenhouse gas emissions in Canada.

3.2 Research Design

Given the comprehensive dataset comprising vehicle specifications and fuel consumption metrics in Canada, this research aims to develop a machine-learning model that categorizes vehicles according to their environmental impact. This classification aims to support the Canadian government's efforts to craft and modify policies promoting sustainable transportation solutions. The following outlines the proposed objectives and methodology for the research:

Objectives

1. Identify the impactful variables in classifying the vehicles based on the Environmental Impact.
2. Perform Exploratory Data Analysis on the dataset to understand what the data is communicating.
3. Develop a Machine Learning Model to Classify Vehicles Based on Environmental Impact.

3.2.1 Methodology

- Data Preprocessing

Handling Missing Data: Given the importance of CO2 emissions for the analysis, it is essential to address null values using appropriate imputation techniques. Alternatively, incomplete records can be filtered out based on their proportion in the dataset.

Feature Engineering: Consider deriving new features that could enhance the analysis, such as incorporating vehicle age or assessing improvements in emissions over the model years.

- Exploratory Data Analysis

Statistical Analysis: Perform statistical tests to understand critical variables' distributions and relationships.

Correlation Analysis: Determine how different features, such as engine size, fuel type, and transmission type, correlate with CO2 emissions.

- Model Building

When selecting a feature, it is essential to employ techniques like Recursive Feature Elimination (RFE) or evaluate feature importance scores from preliminary models to identify the most predictive features for the task.

For model selection, it is beneficial to experiment with various classifiers, including Decision Trees, Random Forests, and Gradient-Boosting Machines. This experimentation helps identify the most effective model in terms of accuracy and robustness for the problem being addressed.

Model Validation: Employ k-fold cross-validation to ensure the model's performance is consistent across different subsets of the dataset.

This methodology outlines a systematic process for creating an effective classification model, ensuring that the research is well-positioned to have practical implications for policymaking. Through in-depth model development and trend analysis, the study aims to comprehensively understand the evolution of vehicle regulations and their impact on environmental objectives. By adopting this structured approach, the study seeks to produce findings directly applicable to policymaking, aiming to inform future transportation strategies in Canada.

The outlined objectives and methodology for the research on vehicle classification based on environmental impact are strategically designed to assist the Canadian government in crafting impactful transportation policies in several significant ways:

1. Data-Driven Environmental Policy Development

Identifying High-Impact Vehicles: The model classifies vehicles according to their environmental impact based on CO2 emissions and fuel efficiency, providing precise data on which vehicle categories contribute most to pollution. This information allows policymakers to design specific regulations targeting these high-impact vehicles, such as stricter emission limits, phase-outs, or higher taxes.

Effective Resource Allocation: The model's insights help allocate resources more effectively by identifying which types of vehicles and what features (like engine size and fuel type) most urgently need attention in terms of regulatory compliance or incentivization.

2. Supporting Legislation with Quantitative Analysis

Basis for New Regulations: The classification model's outputs can be used as a quantitative foundation for drafting new legislation. For example, new categories for tax incentives can be created based on vehicle classification results, or emission standards can be set that are dynamically linked to the specific environmental impact of different vehicle classes.

Evaluating Policy Impact: By analyzing trends over time (from 2005 to 2023), the government can assess the effectiveness of past and current policies, understanding which measures have successfully reduced emissions and which have not, thereby refining future policies.

3. Promoting Sustainable Transportation

Encouraging Cleaner Technologies: With clear classifications and trend analysis, policies can be formulated to encourage the adoption of cleaner technologies. For example, incentives can be provided for vehicles that consistently show low environmental impact or support research and development in vehicle technology, leading to better fuel efficiency and lower emissions.

Public Awareness and Behavioural Change: The research can also inform public awareness campaigns, encouraging consumers to choose vehicles with lower environmental impacts. This can drive market demand towards more sustainable options, further supporting policy goals.

4. Enhancing Policy Flexibility and Responsiveness

Adaptive Policy Frameworks: By continuously updating the model with new data, policies can be kept responsive to the latest trends in vehicle technology and consumer behaviour. This adaptability ensures that transportation policies remain relevant and effective in reducing emissions and promoting public health.

In conclusion, the proposed objectives enable the government to use a robust, evidence-based approach to developing, monitoring, and adjusting transportation policies. This approach addresses immediate environmental concerns and supports sustainable development and public welfare over the long term.

3.2.2 Who are the stakeholders of this ML model

- Government and Regulatory Bodies

Environment and Climate Change Canada (ECCC): Oversees environmental policies and regulations, ensures compliance with emissions standards and uses model predictions for policy-making and monitoring.(https://www.canada.ca/en/environment-climate-change/corporate/transparency/access-information-privacy/info-source.html)

Transport Canada: Implements regulations related to vehicle emissions, promotes sustainable transportation, and uses data to inform infrastructure planning and policy adjustments (Khreis, H., et al, 2023)

- Automotive Industry

Vehicle Manufacturers: Interested in understanding and reducing vehicle emissions to comply with regulations, improve sustainability, and enhance market competitiveness.

Suppliers and Parts Manufacturers: They provide components that can reduce emissions (e.g., catalytic converters, electric powertrains) and need insights into future trends to align their product development strategies.

- Research and Academia

Environmental Scientists and Researchers: Utilise the model to study the impact of vehicle emissions on air quality and climate change and develop new emission reduction technologies.

Data Scientists and ML Researchers: Focus on improving model accuracy and exploring new methodologies for better predictions and insights.

- Non-Governmental Organisations (NGOs)

Environmental NGOs: Advocate for policies and actions to reduce vehicle emissions and use model results to support their campaigns and reports.

Public Health Organizations: Monitor the impact of vehicle emissions on public health, especially in urban areas, and use data to promote health initiatives and regulations.

- Public and Communities

Consumers and Drivers: Interested in how emissions affect air quality and public health and may use information to make informed decisions about vehicle purchases.

Community Groups: Engage in local environmental initiatives and advocate for clean air policies, using data to support their efforts.

- Businesses and Corporations

Logistics and Fleet Management Companies: Use model insights to optimise routes, improve fuel efficiency, and reduce emissions for their fleets of lightweight trucks.

Corporate Sustainability Teams: Track and report on emissions as part of corporate social responsibility (CSR) initiatives and sustainability goals.

- Technology Providers

Software and Analytics Firms: Develop and provide tools for collecting, analyzing, and visualising emissions data and support the deployment of the ML model in various applications.

IoT and Sensor Companies: Supply the necessary hardware to monitor and collect real-time emissions data, providing ML model inputs.

- Policy Makers and Planners

Urban and Regional Planners: Use emissions data to design and implement urban mobility solutions that minimise environmental impact.

Policy Advisors: Inform decision-makers on the effectiveness of current policies and suggest adjustments based on model predictions.

- Investors and Financial Institutions

Impact Investors: Invest in companies and technologies that aim to reduce emissions and promote sustainability, using model data to assess potential investments.

Insurance Companies: Assess risks related to environmental regulations and public health impacts of vehicle emissions.

3.2.3 Analysis Approach

The dataset that we use for the purpose of analysis has been collected from the official website of the Canadian government: Fuel Consumption Dataset. Multiple datasets from 2004-2024 have been merged and are used for the analysis using python. The total data points were 20464 with a total of 13 columns.

The column measuring the CO_2 emissions produced by the vehicles has been taken as the target variable and the rest of the other variables describing the characteristics of the vehicles have been used within the feature matrix that will be used as the predictors.

The feature Matrix consists of the following characteristics:

1. 'Engine Size'
2. 'Cylinders'
3. 'City fuel Consumption Ratings'
4. 'Highway Fuel Consumption Ratings'
5. 'Combined Fuel Consumption Ratings'
6. 'Transformed Combined Fuel Consumption Ratings (mpg)'

The rest of the variables were dropped due to their irrelevance in the model.

The flow of the analysis for the purpose of building the model that predicts the CO2 emissions based on the characteristics of the vehicles is illustrated below through a flowchart:

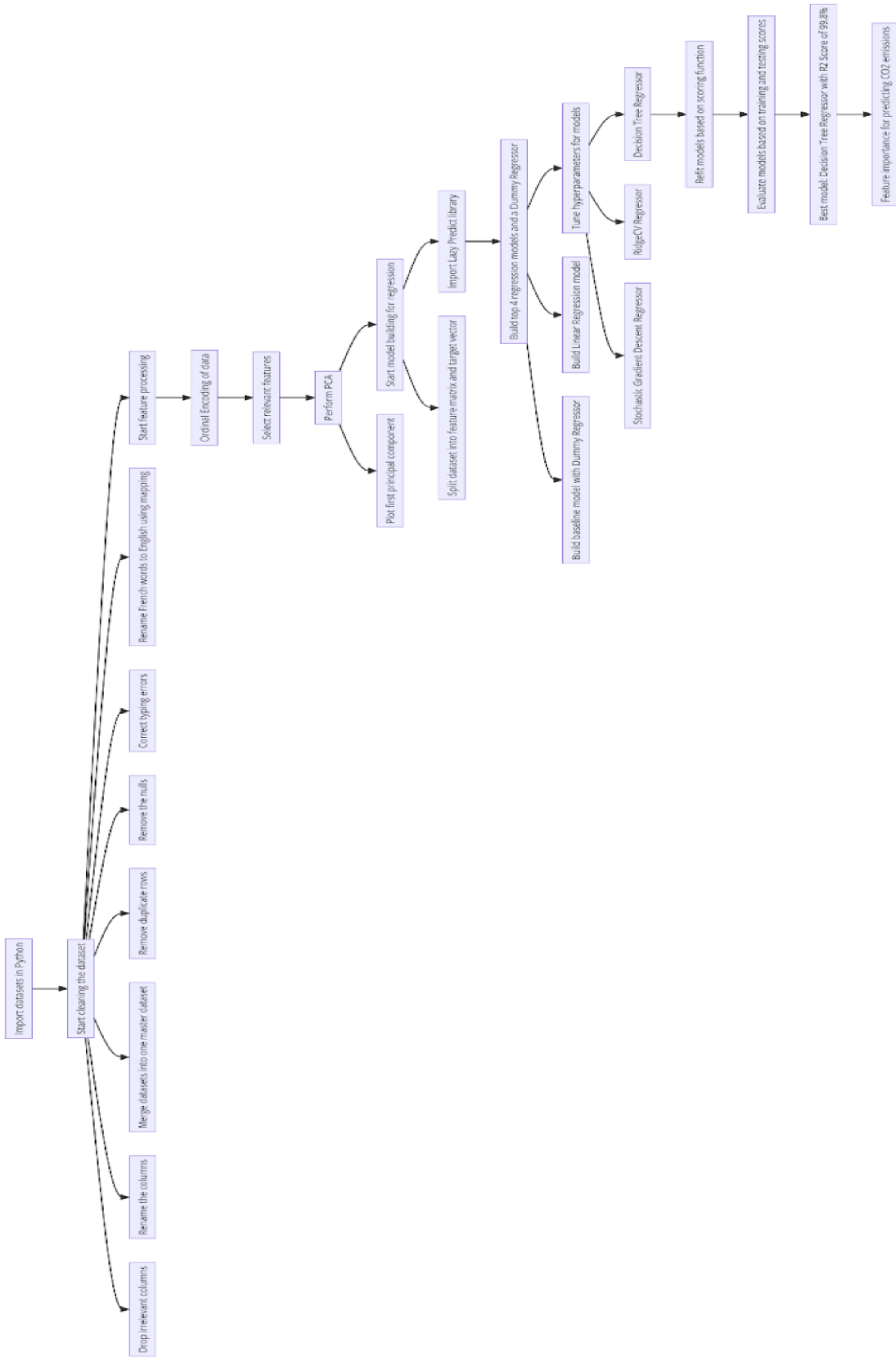

Figure 4 Flowchart for steps involve in processing the data for vehicle characteristics.

1. We import the datasets in Python.
2. We start cleaning the dataset through:
    a. Drop irrelevant columns
    b. Rename the columns
    c. Then we merge the datasets into one master dataset.
    d. Remove duplicate rows
    e. Remove the nulls
    f. Correct the typing errors
    g. Rename the French words to English using the mapping provided by the Canadian website.
3. We start with the data processing that prepares the data for putting in the model:
    a. Do the Ordinal Encoding of the data by converting the string values of categorical variables to numbers.
4. Next Step is Selecting the relevant Features through feature selection process:
    a. Perform PCA (Principal Component Analysis)
        i. Plot the first principle component and the importance of features through a horizontal bar graph.

5.After selecting the important features that explain the maximum variance within the dataset through the first component principle, we start the model building for the regression analysis.

a. Split the ordinally encoded dataset which has the components selected through PCA, into feature matrix and target vector.
b. We first import the Lazy Predict library to have an overview of the basic models that perform well on the dataset and the regression problem.
c. Then we use the output of Lazy Predict and build the top 4 regression models and a Dummy Regressor.
d. We start with the Dummy Regressor model for building a baseline model.
e. Then build the following models using first performing the feature scaling:
    i. Linear Regression
    ii. We build the following models by tuning the hyperparameters after feature scaling in the following models:

a. Stochastic Gradient Descent Regressor
b. RidgeCV Regressor
c. Decision Tree Regressor

iii. The models are thereafter refitted based on the scoring function “negative root mean squared error”.

iv. The evaluation of the model is based on the training and testing scores of predicted values.

v. The best model- Decision Tree Regressor has a R2 Score of 99.8%. The feature importance of the decision tree regressor is used to access the most important features that have been identified by the tree based model in predicting the CO_2 emissions of the light weight vehicles

3.3 Identifying Key Variables for Vehicle Environmental Classification

The first step in developing a machine learning-based framework for classifying vehicles based on environmental impact is identifying the key variables influencing this classification. This process involves understanding which vehicle specifications and fuel consumption metrics are most predictive of environmental impact, particularly regarding emissions. A thorough and systematic research design is essential to ensure the selected variables are relevant and robust for subsequent modelling.

- Data Collection

Data Sources:

Vehicle Specifications: Data will be collected from reliable sources such as vehicle registration databases, manufacturer specifications, and government databases. Key attributes include make, model, year, engine size, fuel, transmission, weight, and drivetrain.

Fuel Consumption and Emissions Data: This will be sourced from government reports, environmental agencies, and databases such as the National Resources Canada (NRCan) Fuel Consumption Ratings database. Key metrics include CO_2 emissions, fuel efficiency (miles per gallon or litres per 100 kilometres), and other pollutants such as NO_x and PM.

Supplementary Data: Additional data such as vehicle usage patterns, maintenance records, and geographical data may also be collected to provide context and enhance the analysis.

- Data Integration:

The collected data will be integrated into a comprehensive dataset. This process involves standardizing formats, merging data from different sources, and ensuring consistency across the dataset.

- Data Preprocessing

Handling Missing Data:

Imputation: Apply appropriate imputation techniques to fill in missing values—for example, mean or median imputation for numerical data and mode imputation for categorical data.

Filtering: If imputation is not feasible, filter out records with significant missing data, ensuring the dataset remains robust and representative.

- Data Cleaning:

Remove duplicates and correct inconsistencies in the dataset.

Standardize units of measurement (e.g., converting all fuel efficiency data to a standard unit such as litres per 100 kilometres).

Feature Engineering:

Derive new features that may be useful for classification. For example, vehicle age (current year minus model year), improvements in emissions over model years, and combined metrics like power-to-weight ratio.

- Exploratory Data Analysis (EDA)

Descriptive Statistics:

Calculate basic statistics (mean, median, standard deviation) for critical variables to understand their distribution and central tendency.

- Visualization:

Visual tools such as histograms, box plots, and scatter plots can help you visualize the distributions and relationships between variables.

Identify outliers and patterns that may indicate significant trends or anomalies.

Correlation Analysis:

Calculate correlation coefficients (e.g., Pearson, Spearman) to identify relationships between variables.

Use heatmaps to visualize correlations and identify clusters of related variables.

Principal Component Analysis (PCA):

Perform PCA to reduce dimensionality and identify the most critical variables that capture the majority of the variance in the data.

- Feature Selection

Initial Feature Selection:

Based on EDA and correlation analysis, select an initial set of candidate variables that appear to have significant relationships with emissions and environmental impact.

Recursive Feature Elimination (RFE):

Use RFE to iteratively select the most essential features by training a model and removing the most minor essential features in each iteration.

Evaluate the model's performance at each step to identify the optimal set of features.

- Feature Importance from Models:

Train preliminary models (e.g., decision trees, random forests) and use their feature importance scores to identify critical variables.

Compare the results from different models to ensure robustness and consistency in feature selection.

To develop a machine learning-based framework for classifying vehicles based on environmental impact, the first step is to identify the key variables influencing this classification. This involves understanding which vehicle specifications and fuel consumption metrics are most predictive of environmental impact, particularly regarding emissions. Data collection involves sourcing vehicle specifications from reliable databases, including make, model, year, engine size, fuel, transmission, weight, and drivetrain. Fuel consumption and emissions data will be collected from government reports and environmental agencies, including CO2 emissions, fuel efficiency, and other pollutants. Supplementary data such as vehicle usage patterns and maintenance records may also be collected to provide context.

Following the data collection phase, the next step is integrating the data into a comprehensive dataset. This involves standardizing formats, merging data from different sources, and ensuring consistency. Data preprocessing is a critical step involving the handling of missing data through imputation and filtering, as well as the cleaning of the data by removing duplicates and standardizing units of measurement. A vital aspect of this phase is feature engineering. This process involves deriving new features that could enhance the model's predictive power, such as vehicle age and power-to-weight ratio.

Exploratory data analysis will be conducted to calculate basic statistics, visualize data distributions, identify outliers and patterns, and perform correlation analysis. Principal component analysis will be used to reduce dimensionality and identify critical variables. Feature selection will involve initial feature selection based on EDA and correlation analysis, recursive feature elimination to iteratively select essential features, and model performance evaluation to identify the optimal set of features. Feature importance from models will also be used to identify critical variables and ensure robustness and consistency in feature selection.

3.4 Conducting Exploratory Data Analysis for Insight Extraction

Conducting an in-depth data analysis (EDA) is essential for any research project based on data. It involves thoroughly examining the dataset to uncover patterns, detect anomalies, and understand the relationships between variables. For the research aiming to categorize vehicles based on environmental impact, EDA will play a vital role in laying a solid foundation for constructing a reliable machine-learning model.

To begin, the data must be prepared with care. This includes collecting details about vehicle specifications, fuel consumption measurements, and emissions data from trustworthy sources. It is crucial to integrate these data points into a unified dataset to ensure consistency and standardization across all data points, such as units of measurement and data formats. Any discrepancies, duplications, or flaws in the data must be rectified, and any missing values should be appropriately addressed. This may involve substituting missing values with the mean or median for numerical data or the mode for categorical data. In cases of significantly incomplete data, these records may be excluded to maintain the dataset's quality.

Once the data is prepared, the initial examination begins with providing descriptive statistics. This involves computing basic statistics such as each variable's mean, median, standard deviation, minimum, and maximum. Summarizing these statistics helps comprehend the central tendencies and distribution of the data. It is also vital to assess the data types and structure, determine the number of records and variables, and identify any missing data. Visual tools like heat maps or bar plots can help visualize patterns in missing data.

The next phase of EDA involves univariate analysis, where each variable is independently studied. Histograms can be utilized to visualize the distribution of numerical variables, while box plots can help identify outliers and understand the spread of the data.

Calculating skewness and kurtosis offers additional insights into the shape of these distributions. For categorical variables, bar charts can be used to display the frequency of each category, and proportions or percentages can be calculated and visually represented.

Following this is bivariate analysis, which examines the relationships between pairs of variables. Using Pearson or Spearman coefficients, correlation analysis quantifies these relationships for numerical variables, and the results can be visualized in a correlation matrix or heatmap. Scatter plots can illustrate the relationships between pairs of numerical variables, often with trend lines added to highlight linear or nonlinear relationships. When comparing categorical and numerical variables, box plots or violin plots can be used to compare the distributions of numerical variables across different categories, and statistical tests like t-tests or ANOVA can assess the significance of differences between groups.

Multivariate analysis takes this a step further, examining multiple variables simultaneously. Pair plots, or scatter plot matrices, can simultaneously visualize the relationships among several numerical variables. Principal Component Analysis (PCA) can reduce the dataset's dimensionality, helping identify which variables contribute most to the variance. Clustering algorithms, such as K-means, can reveal natural groupings within the data, with tools like silhouette scores and elbow plots helping determine the optimum number of clusters.

Identifying patterns and gaining insights is a crucial objective of EDA. Trend analysis can reveal changes over time in variables like emissions and fuel efficiency, using line plots to illustrate these trends. Outlier detection, using statistical methods or visual tools, helps identify data points that significantly deviate from the norm. Feature importance analysis involves training preliminary models, such as decision trees or random forests, to estimate the importance of each variable. This helps identify the critical variable significantly influencing the target variable: environmental impact.

Effectively communicating these insights is essential. Interactive dashboards created with tools like Tableau or Power BI can present critical findings and enable dynamic data exploration. Various graphs and charts, including histograms, scatter plots, box plots, and heat maps, can visually communicate insights. Summary tables can highlight vital statistics, correlations, and findings, making it easier to understand the main points.

The final step involves interpreting and documenting the EDA results. The insights need to be interpreted in the context of the research objectives, and their implications for further analysis and model development need to be discussed. The entire EDA process should be thoroughly documented, from data cleaning to analysis methods and key findings, to ensure clarity and reproducibility.

To summarize, a comprehensive and systematic EDA provides a deep understanding of the dataset, revealing meaningful insights that guide subsequent steps in developing a machine-learning model for classifying vehicle environmental impact. This thorough approach ensures that the analysis is comprehensive.

- Multivariate Analysis

Pair Plot: Create pair plots (scatter plot matrices) to visualize relationships between multiple numerical variables simultaneously.

Principal Component Analysis (PCA): Perform PCA to reduce the dimensionality of the dataset.

Visualize the principal components to identify which variables contribute most to the variance.

Clustering: Apply clustering algorithms (e.g., K-means) to identify natural groupings within the data. Use silhouette scores and elbow plots to determine the optimal number of clusters.

- Identifying Patterns and Insights

Trend Analysis: Analyze trends over time for variables such as emissions and fuel efficiency. Use line plots to visualize changes and identify significant trends.

Outlier Detection: Identify outliers using statistical methods or visualization tools.

Assess the impact of outliers on the overall analysis and decide whether to exclude or further investigate them.

Feature Importance: Train preliminary models (e.g., decision trees, random forests) to estimate feature importance scores. Use these scores to identify critical variables significantly influencing the target variable (e.g., environmental impact).

- Visualizations for Communication

Dashboards: Create interactive dashboards using tools like Tableau or Power BI to present key findings and allow for dynamic data exploration.

Graphs and Charts: To communicate insights effectively, use a variety of graphs and charts (e.g., histograms, scatter plots, box plots, heatmaps).

3.5 Developing a Machine Learning Model for Vehicle Classification Based on Environmental Impact

- Data Preparation:

The first step involves gathering and preparing the dataset, which includes various features related to vehicle specifications, fuel consumption, and emissions data. This stage requires integrating data from multiple reliable sources and ensuring consistency in measurement units and formats. Based on the extent of the data gaps, any missing values are addressed through imputation or exclusion. This meticulous preparation ensures a high-quality, standardized dataset for model building.

- Feature Selection:

Selecting the right features is crucial for developing an effective machine-learning model. This process involves identifying variables that significantly influence the target variable, which is CO_2 emissions. Techniques such as correlation analysis, feature importance rankings, and domain knowledge are employed to shortlist the most relevant features. This step ensures the model is built on the most informative and impactful variables, enhancing its predictive power.

- Dataset Splitting:

The dataset is split into training and testing sets using the train_test_split() function from the sci-kit-learn library. Typically, 80% of the data is used for training the model, while 20% is reserved for testing its performance on unseen data. This split helps evaluate the model's generalization ability beyond the training data.

- Initial Model Selection Using LazyPredict:

LazyPredict is utilized to compare the performance of various machine learning models quickly. LazyPredict is a Python library that automates the training and evaluation of multiple models with minimal coding effort. It provides an overview of different models' performance metrics, helping to identify promising candidates for further tuning. This step saves time and effort in the initial stages of model selection.

- Baseline Model with Dummy Regressor:

A Dummy Regressor is implemented as a baseline model. This model makes predictions based on simple strategies like the mean or median of the target variable, providing a benchmark to compare the performance of more complex models. The Dummy Regressor's R^2 scores on the training and test sets indicate that it does not capture any meaningful patterns in the data, which is expected and confirms the need for more sophisticated models. A pipeline is constructed for subsequent models, with feature standardization as the first step and applying the chosen estimator. Standardizing features ensures that all variables, especially those involving distance or magnitude, contribute equally to the model's performance.

- Linear Regression Model:

A linear regression model is built and evaluated. The intercept and coefficients are analyzed to understand the relationship between each feature and CO2 emissions. The high R^2 scores on both the training and test sets indicate that the model explains a significant portion of the variance in CO2 emissions, and the close scores suggest good generalization without overfitting. The Root Mean Squared Error (RMSE) measures the average prediction error.

Next, a Stochastic Gradient Descent (SGD) Regressor is implemented. Hyperparameter tuning is performed to find the best parameters for the model. The SGD Regressor's performance metrics, including R^2 and RMSE, are compared with those of the linear regression model. The results indicate that the SGD Regressor performs similarly well, suggesting it is a robust alternative for this dataset.

- Ridge Regression with Cross-Validation:

RidgeCV, a variant of Ridge Regression with built-in cross-validation for selecting the best regularization parameter (alpha), is applied. This model introduces regularization by shrinking the coefficients to prevent overfitting. The significantly lower RMSE values for training and test sets demonstrate the model's superior performance and effective regularization.

- Decision Tree Regressor:

A Decision Tree Regressor is built and evaluated. Hyperparameter tuning is performed to determine the optimal tree depth and other parameters. The shallow RMSE values for the Decision Tree Regressor indicate excellent prediction accuracy. Feature importance scores are analyzed, revealing that combined fuel consumption ratings are the most influential in predicting CO2 emissions.

- Model Evaluation and Interpretation:

Each model's performance is evaluated using metrics such as R^2 and RMSE. The high R^2 scores indicate that the models explain a large portion of the variance in CO2 emissions, while the RMSE values provide a direct measure of prediction error. The close alignment of these metrics across training and test sets suggests good generalization and robustness.

The detailed analysis of model coefficients and feature importance scores offers insights into the relationship between vehicle features and CO2 emissions. The linear regression model's coefficients and the Decision Tree Regressor's feature importance rankings highlight the significant predictors, aiding in the interpretation of the underlying patterns.

In conclusion, this methodology ensures a comprehensive approach to developing a robust machine-learning model for classifying vehicles based on environmental impact. The process aims to achieve high predictive accuracy and meaningful insights into the factors influencing CO2 emissions through meticulous data preparation, feature selection, model building, and evaluation.

3.6 Population and Sample

The scope of this study encompasses a comprehensive range of vehicles utilized worldwide, including passenger cars, trucks, buses, motorcycles, and various other forms of motorized transport. These vehicles exhibit significant diversity in terms of make, model, year of manufacture, engine size, fuel type, and other specifications that can significantly impact their environmental footprint, particularly CO2 emissions. It is important to note that this population is not confined by geographic boundaries, brand, or vehicle type, providing a holistic overview of vehicles' global environmental performance.

The sample for this study is derived from a specific subset of vehicles within the broader population. These vehicles have comprehensive and detailed records available, including engine size, number of cylinders, type of transmission, fuel type, and various fuel consumption ratings such as city, highway, and combined. This selection of vehicles for the sample is contingent on the availability and completeness of data to ensure the robustness and representativeness of the dataset used for model development. The sample is likely sourced from various reputable outlets such as vehicle registration databases, environmental agencies, automotive manufacturers, and esteemed research institutions, thereby establishing a diverse and reliable dataset for the study.

The study's primary objective is to develop machine learning models using this representative sample, with the goal of creating predictive models that can be broadly applied to the entire vehicle population. Insights obtained from this sample are intended to be widely applicable, offering valuable support in understanding and addressing the global environmental impact of vehicle emissions.

3.7 Participant Selection

Participants in this research, defined as the vehicles in the dataset, are selected based on the availability and reliability of data from multiple credible sources. These sources may include:

Government Databases: Registries and environmental databases maintained by government agencies such as Transport Canada, Environment and Climate Change Canada, and similar bodies in other countries.

Automotive Industry Records: Data from automotive manufacturers often provide detailed specifications and emissions data for various vehicle models.

Research Institutions: Academic and research institutions that conduct studies on vehicle emissions and maintain comprehensive datasets (Lyu, P, et al, 2021)

Publicly Available Datasets: Open-source datasets from organizations like the International Council on Clean Transportation (ICCT) or the U.S. Environmental Protection Agency (EPA) (https://theicct.org/).

- Inclusion Criteria:

Completeness of Data: Vehicles with complete records, including engine size, number of cylinders, transmission type, fuel type, and detailed fuel consumption ratings (city, highway, and combined).

Accuracy of Emission Data: Vehicles for which accurate CO2 emission data is available. This ensures the reliability of the machine learning model developed for classification.

Diversity of Vehicle Types: Inclusion of a wide range of vehicles (passenger cars, trucks, buses, motorcycles) to ensure the sample represents the diversity within the population.

Temporal Coverage: Vehicles from various model years, ideally spanning a significant time frame to capture trends and changes in vehicle technology and emissions.

- Exclusion Criteria:

Incomplete Data Records: Vehicles with missing or incomplete data, especially regarding critical variables such as CO2 emissions and fuel consumption ratings.

Unverified Data: Data that cannot be corroborated by reliable sources or datasets with questionable accuracy.

Outliers: Extreme outliers that may distort the analysis unless they represent a significant and relevant portion of the population (e.g., specific types of high-emission vehicles).

Random Sampling: Random selection of vehicles from the identified datasets to avoid bias and ensure a representative sample.

Stratified Sampling: Ensuring that different categories of vehicles (e.g., by fuel type, vehicle class) are proportionately represented in the sample. This helps capture the variability within the population and improves the model's generalizability.

Systematic Sampling: If datasets are structured to allow for systematic sampling, vehicles might be selected at regular intervals from a sorted list (e.g., every nth vehicle from a database sorted by registration date).

3.8 Instrumentation

Python is optimal for fulfilling research objectives, especially in constructing a machine-learning model for classifying vehicles based on their environmental impact (Julio-Rodriguez, et al, 2022). Its extensive libraries, versatility, and user-friendly nature make it a potent instrument for data analysis, model construction, and evaluation.

The process kicks off with data acquisition and preprocessing. Python's libraries, such as pandas and numpy, are leveraged to import datasets from diverse sources like CSV files, databases, and APIs. Data cleansing involves managing missing values, outliers, and inconsistent entries to ensure the dataset's robustness for analysis. Feature engineering is carried out to create new features that augment model performance, like computing vehicle age or fuel efficiency enhancements.

Exploratory Data Analysis (EDA) plays a pivotal role in grasping the dataset's structure and relationships. Statistical analysis using pandas and numpy yields descriptive statistics, while visualization tools such as Matplotlib and Seaborn aid in visualizing data trends and correlations. Correlation matrices are computed to pinpoint relationships between features, offering insights that steer the model-building process.

The subsequent phase involves building and evaluating machine learning models. Initially, LazyPredict is employed to swiftly compare the performance of various models, establishing a baseline for further development. A DummyRegressor is implemented as a baseline model, setting a reference point for model performance. More sophisticated models, including Linear Regression, Stochastic Gradient Descent (SGD) Regressor, RidgeCV, and Decision Tree Regressor, are then constructed using the scikit-learn library. Pipelines are constructed to standardize features and apply regressors, ensuring a streamlined workflow.

Hyperparameter tuning is essential for optimizing model performance. Techniques like grid search (GridSearchCV) or random search (RandomizedSearchCV) are used to unearth the best hyperparameters. Model evaluation uses metrics like R^2 and RMSE, providing an all-encompassing assessment of model accuracy and generalization. The most effective model is pinpointed by comparing these metrics across different models.

Model interpretation and reporting are the concluding stages in the process. Interpreting model coefficients and feature importances provide insights into the relationships captured by the model. Visualization tools are once again utilized to illustrate model performance, feature importance, and predictions, presenting the results in an accessible and understandable manner. Comprehensive reports are generated using Jupyter

Notebook and Markdown, documenting the methodology, findings, and interpretations in a clear and structured format.

In summary, Python's rich ecosystem of libraries and tools enables efficient and effective implementation of the research objectives. From data preprocessing and exploratory analysis to model building, evaluation, and reporting, Python facilitates a systematic approach that ensures reproducible and transparent results, ultimately informing and enhancing Canadian transportation policies.

3.9 Data Collection Procedures

The procedures for gathering data to create a machine-learning model that categorizes vehicles based on their environmental impact in Canada went through various essential stages to ensure that comprehensive, accurate, and pertinent data was obtained. These stages involved pinpointing data sources, obtaining data, and organizing the dataset for analysis.

- Locating Data Sources

The initial stage in the data collection process involved identifying trustworthy and authoritative data sources regarding vehicle specifications and fuel consumption figures in Canada. Key sources included:

Government Databases: Environment and Climate Change Canada (ECCC) and Transport Canada provided extensive vehicle emissions and fuel consumption datasets. These databases contained detailed records on various vehicle models, including CO2 emissions, engine size, fuel type, and fuel consumption ratings.

Automotive Industry Reports: Manufacturers and industry organizations published annual reports and datasets containing different vehicle models' technical specifications and performance metrics. These reports were valuable for obtaining up-to-date and comprehensive data.

Public Databases: Websites like the Fuel Consumption Ratings search tool and other public repositories offered access to historical and current data on vehicle fuel efficiency and emissions.

- Obtaining Data

After identifying the data sources, the next step was to obtain the data. This involved:

Downloading Data Files: Data files were downloaded from government websites, industry reports, and public databases. These files were typically in formats such as CSV (Comma-Separated Values), Excel, or JSON, which can be easily managed with Python libraries.

- Organizing the Dataset

After obtaining the data, the next critical step was to prepare it for analysis. This preparation involved several processes:

Data Cleaning: The raw data often contained inconsistencies, missing values, and outliers that must be addressed. The data was cleaned using Python libraries such as

Pandas, which handled missing values (e.g., imputing or removing them), corrected inconsistent entries, and identified outliers for further investigation.

Data Integration: Since the data was gathered from multiple sources, it was crucial to integrate them into a single, cohesive dataset. This involved merging datasets on standard keys (such as vehicle model or VIN) and ensuring that all relevant features were included.

Feature Engineering: New features were derived to enhance the dataset. For instance, vehicle age was calculated based on the manufacturing year, and combined fuel efficiency metrics were derived from city and highway ratings.

Data Transformation: The data was transformed into a suitable format for analysis. This included encoding categorical variables (e.g., fuel type, transmission type) using one-hot encoding techniques and normalizing numerical features to ensure they were on a comparable scale.

Thorough documentation was maintained throughout the data collection process. This involved recording the data sources, the steps taken to clean and integrate the data, and any assumptions or transformations applied. The final dataset was stored in a secure and accessible format, ensuring it could be efficiently utilized for subsequent analysis and model building.

Following these detailed data collection procedures, a robust and comprehensive dataset was assembled, providing a solid foundation for developing an accurate and reliable machine learning model to categorize vehicles based on their environmental impact. This methodical approach ensured that the data was pertinent, accurate, and ready for the rigorous demands of machine learning analysis.

3.10 Data Analysis

In the research aimed at developing a machine learning model to classify vehicles based on their environmental impact, a comprehensive approach to data analysis was crucial for deriving meaningful insights and building accurate predictive models. The analysis began with thorough exploratory data analysis (EDA), where descriptive statistics and visualizations such as histograms and scatter plots were employed to understand the distributions of vehicle specifications and fuel consumption metrics. Correlation analysis further illuminated relationships between these variables and CO_2 emissions, guiding the selection of influential features for model development.

Feature selection played a pivotal role, utilizing techniques like Random Forest for feature importance to prioritize variables most predictive of environmental impact. Dimensionality reduction methods like principal component analysis (PCA) were also considered to streamline the dataset. Model building encompassed a series of steps from establishing baseline performance with Dummy Regressor to implementing sophisticated regression techniques such as Linear Regression, Stochastic Gradient Descent (SGD) Regressor, and RidgeCV Regressor. Each model was evaluated rigorously using metrics like R^2 score and Root Mean Square Error (RMSE) to assess predictive accuracy and generalization capability across training and test datasets.

Interpreting model coefficients provided insights into how specific features—such as engine size, fuel type, and consumption ratings—affected CO_2 emissions. Visualizations of feature importance from Decision Tree models further clarified which variables were most influential in predicting emissions outcomes. Throughout the analysis, validation techniques like k-fold cross-validation ensured the reliability and robustness of the models. The results of these analyses not only facilitated the development of a precise classification framework for vehicle environmental impact but also provided actionable insights for policymakers aiming to enhance sustainability in transportation through evidence-based regulations and incentives.

3.11 Research Design Limitations

When designing a research study, it is essential to consider the potential limitations that could affect the results and conclusions. One primary consideration is the sampling method used. Suppose the sample size is too small or the selected participants do not need to represent the larger population. In that case, the study's ability to generalize findings accurately to the broader population can be limited. Another factor to be mindful of is the data collection methods. Biases in self-reported data or errors in measurement can introduce inaccuracies, which in turn can impact the reliability of the results.

Additionally, researcher bias is a significant limitation in qualitative research. Personal interpretations and biases can influence data collection, analysis, and interpretation. Finally, the scope of the study should be considered. Narrowly focused research may limit its applicability to broader contexts, reducing its impact and relevance. It is crucial to recognize and address these limitations to ensure the credibility and validity of the research findings.

3.12 Conclusion

The methodology chapter of this research aimed to achieve several objectives, each contributing to a comprehensive understanding of vehicle environmental impact classification through machine learning. The primary objective was to identify critical variables that significantly influence environmental classification. This involved rigorous data preprocessing to handle missing values and feature engineering to extract relevant insights from vehicle specifications and fuel consumption metrics in Canadian datasets.

The second objective centred on conducting exploratory data analysis (EDA) to uncover patterns and relationships within the data. Statistical analysis and correlation studies were employed to understand how engine size, fuel type, and transmission correlate with CO_2 emissions. This phase provided critical insights into the dataset's characteristics and guided subsequent model development.

The third objective focused on developing machine learning models to classify vehicles based on environmental impact. Starting with a baseline assessment using

DummyRegressor to establish performance benchmarks, the study progressed to more sophisticated models like linear regression, Stochastic Gradient Descent, RidgeCV, and Decision Tree Regression. Each model was meticulously evaluated using metrics such as R^2 scores and RMSE values to assess predictive accuracy and generalization capability.

The adoption of LazyPredict facilitated rapid model comparison and selection, streamlining the process of identifying the most effective model for predicting CO2 emissions. Despite inherent limitations related to sample size representativeness, data collection methodologies, and potential biases, the methodology provided a structured approach to generating actionable insights. These insights are crucial for informing evidence-based policy decisions to promote sustainable transportation practices and reduce environmental impact in Canada and similar contexts globally.

CHAPTER IV:
RESULTS

4.1 Correlation Between Emissions Data and Environmental Impact

Feature selection is the process through which we identify which features / variables are the most relevant to the target variable.

- FEATURE SELECTION

For feature selection, we use the principal component analysis:

```
from sklearn.decomposition import PCA
ss = StandardScaler()
X_encoded_pca = ss.fit_transform(X_encoded)
pca = PCA(n_components = 5, random_state = 42)
X_encoded_pca = pca.fit_transform(X_encoded_pca, y)
pca.explained_variance_

array([5.39935435, 1.22791912, 1.05681769, 0.75598038, 0.33376699])

[ ] pca.components_

array([[ 0.11622278,  0.38315907,  0.37084916, -0.00436656,  0.13679167,
         0.4216978 ,  0.40912048,  0.4221368 , -0.40001648],
       [-0.54150683,  0.0349644 ,  0.15050387,  0.47788615,  0.66046416,
        -0.01537482, -0.11403001, -0.05081513,  0.04986927],
       [ 0.62176724,  0.00866344, -0.02843656,  0.77550698, -0.05012703,
        -0.07742803,  0.01103022, -0.04668581,  0.01736858],
       [ 0.3795814 , -0.37707421, -0.41130512, -0.26100242,  0.65615109,
         0.06701327,  0.16637311,  0.10351701, -0.05493657],
       [ 0.3789607 ,  0.41799116,  0.44754316, -0.30796067,  0.32058251,
        -0.18717974, -0.2256698 , -0.20356819,  0.39542541]])
```

Figure 5 PCA Used for the Feature Selection

The provided Python script in figure 5 demonstrates using Principal Component Analysis (PCA) with the sci-kit-learn library to analyze a dataset. Initially, the data is standardized using StandardScaler() to ensure each feature has a mean of zero and a standard deviation of one. The PCA model is configured to retain five principal components, fitted and transformed on the standardized data. The explained variance for each principal component is calculated, indicating how much variance each principal component captures from the data. The results show that the first principal component explains the most variance (5.399 units), followed by the subsequent components with decreasing variance. The principal components (eigenvectors) are also extracted, representing the direction of maximum variance in the data. These components reveal the underlying structure and relationships between the features in the dataset.

We select top 5 components that explain the maximum variance within the dataset, and we plot the importance of the features that have been through the first principal component:

```
# Create a DataFrame for the PCA components
pca_columns = [f'PC{i+1}' for i in range(pca.n_components_)]
X_pca_df = pd.DataFrame(data=X_encoded_pca, columns=pca_columns)

# PCA components (loadings)
loadings = pd.DataFrame(pca.components_.T, columns=pca_columns, index=X_encoded.columns)
print(loadings)
```

```
                              PC1       PC2       PC3       PC4       PC5
class                    0.116223 -0.541507  0.621767  0.379581  0.378961
engine_size              0.383159  0.034964  0.008663 -0.377074  0.417991
cylinders                0.370849  0.150504 -0.028437 -0.411305  0.447543
transmission            -0.004367  0.477886  0.775507 -0.261002 -0.307961
fuel_type                0.136792  0.660464 -0.050127  0.656151  0.320583
city_fuel_cons_rating    0.421698 -0.015375 -0.077428  0.067013 -0.187180
hw_fuel_cons_rating      0.409120 -0.114030  0.011030  0.166373 -0.225670
comb_fuel_cons_rating    0.422137 -0.050815 -0.046686  0.103517 -0.203568
comb_fuel_cons_rating_mpg -0.400016  0.049869  0.017369 -0.054937  0.395425
```

Figure 6 List of Top 5 Components

The above image creates a dataframe with 5 columns, each having the variance explained through the principal component for all the features. Then we plot a horizontal bar of the first principal component. The given Python script extends the Principal Component Analysis (PCA) process by creating a data frame to store the PCA components and their loadings. Initially, the script generates column names for the PCA components (PC1, PC2, etc.) and creates a DataFrame (X_pca_df) containing the transformed data for these components. The PCA loadings, which indicate the correlation between the original features and the principal components, are extracted and organized into another data frame (loadings). This data frame displays the contribution of each original feature (such as class, engine size, cylinders, etc.) to the principal components. For instance, the first principal component (PC1) is primarily influenced by features like combined city and highway fuel consumption ratings. This structured representation of PCA loadings helps to identify the most impactful features for each principal component, aiding in the interpretation and dimensionality reduction of the dataset.

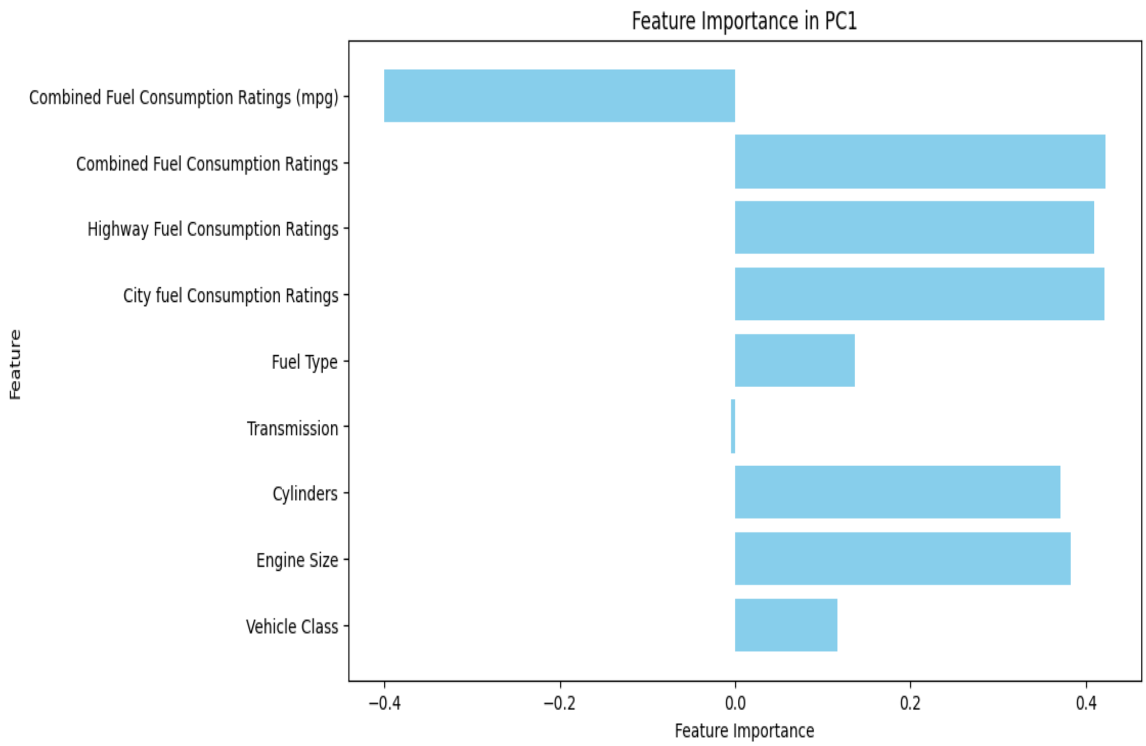

Figure 7 feature importance of the first principal component (PC1)

The horizontal bar graph displays the feature importance of the first principal component (PC1) from a Principal Component Analysis (PCA) performed on a set of features measured for lightweight vehicles and light trucks. The target variable for the analysis is CO2 emissions of the vehicles. Here's a detailed interpretation of the results:

Here are the key features and their importance as derived from the analysis:

1. Combined Fuel Consumption Ratings (mpg):

Importance: This feature has the highest absolute loading (negative).

Interpretation: An increase in the combined fuel consumption rating in miles per gallon (mpg) leads to a decrease in the value of PC1. This indicates that vehicles with higher mpg (better fuel efficiency) are associated with lower CO2 emissions.

2. Combined Fuel Consumption Ratings:

Importance: This feature has the second-highest absolute loading (favourable).

Interpretation: An increase in the combined fuel consumption rating in litres per 100 kilometres (L/100 km) leads to an increase in the value of PC1. This suggests that higher fuel consumption is strongly correlated with higher CO2 emissions.

3. Highway Fuel Consumption Ratings:

Importance: This feature also has a significant favourable loading.

Interpretation: Vehicles with higher highway fuel consumption ratings tend to have higher CO2 emissions, contributing positively to PC1.

4. City Fuel Consumption Ratings:

Importance: This feature has a notable favourable loading.

Interpretation: Higher city fuel consumption ratings are associated with higher CO2 emissions, making this feature important for PC1.

5. Fuel Type:

Importance: This feature has a moderately favourable loading.

Interpretation: Different fuel types impact CO2 emissions, with certain fuel types contributing more to higher emissions.

6. Transmission:

Importance: This feature has a minimal favourable loading.

Interpretation: Transmission type has a relatively small influence on CO2 emissions compared to fuel consumption ratings.

7. Cylinders:

Importance: This feature has a favourable loading.

Interpretation: The number of engine cylinders has some influence on CO2 emissions, but it is less critical than fuel consumption metrics.

8. Engine Size:

Importance: This feature has a positive loading.

Interpretation: Larger engine sizes are associated with higher CO2 emissions, but this feature is less influential than fuel consumption ratings.

9. Vehicle Class:

Importance: This feature has the lowest favourable loading.

Interpretation: The vehicle class (lightweight vehicles vs. light trucks) has the most minor influence on CO2 emissions among all the features considered.

In our analysis, we found that the Combined Fuel Consumption Ratings (mpg) and Combined Fuel Consumption Ratings (L/100 km) are the most influential features for Principal Component 1 (PC1). These features have the highest absolute loadings, suggesting they play a crucial role in explaining the variance in CO2 emissions. The negative loading for mpg indicates better fuel efficiency (higher mpg) is associated with lower CO2 emissions. Conversely, the favourable loading for combined fuel consumption (L/100 km) suggests that higher fuel consumption is associated with higher CO2 emissions.

Additionally, we observed that Highway Fuel Consumption Ratings and City Fuel Consumption Ratings also have strong positive loadings, indicating that higher fuel consumption in both city and highway driving conditions contributes to increased CO2 emissions.

Furthermore, Fuel Type was found to have a moderate influence, suggesting that different fuel types have varying impacts on CO2 emissions.

On the other hand, Transmission, Cylinders, Engine Size, and Vehicle Class showed smaller loadings, indicating that these features are less critical in explaining the variance in CO2 emissions compared to fuel consumption ratings.

In conclusion, our Principal Component Analysis (PCA) results underscore the primary importance of fuel consumption ratings (combined, highway, and city) in explaining CO_2 emissions for lightweight vehicles and light trucks. It emphasizes that features such as fuel type, transmission, number of cylinders, engine size, and vehicle class are less influential in comparison. These insights can provide valuable guidance on the key factors affecting CO_2 emissions, focusing on fuel efficiency and consumption.

4.2 Relationships Among Fuel Efficiency, Fuel Type, and Vehicle Emissions

During the data preprocessing stage, we created a scatter plot graph to visualize the relationship between the features and CO_2 emissions. This allowed us to gain insight into how the features are related to the CO_2 emissions in the dataset.

While interpreting the following scatter plots, please consider the relationships between a range of features such as Engine Size, Cylinders, City Fuel Consumption Ratings, Highway Fuel Consumption Ratings, Combined Fuel Consumption Ratings, and Transformed Combined Fuel Consumption Ratings (mpg) with the target variable CO_2 Emissions for lightweight vehicles. Each scatter plot will offer insights into these relationships.

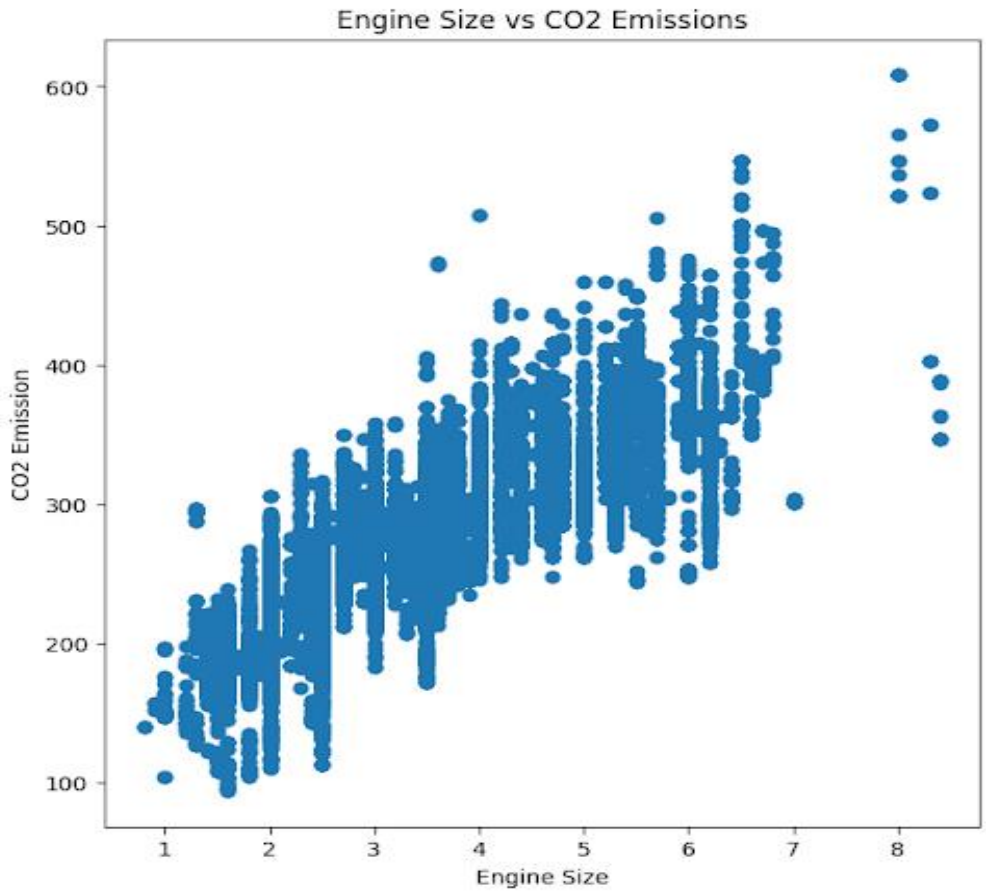

Figure 8 Engine Size Vs CO2 Emissions

Engine Size vs CO2 Emissions Description: This scatter plot shows the relationship between engine size (in litres) and CO2 emissions (in grams per kilometre).

Based on the analysis, there is a direct and positive relationship between engine size and CO2 emissions. As the size of the engine increases, there is a corresponding increase in CO2 emissions. This correlation can be attributed to the fact that larger engines generally consume more fuel, thereby resulting in higher CO2 emissions.

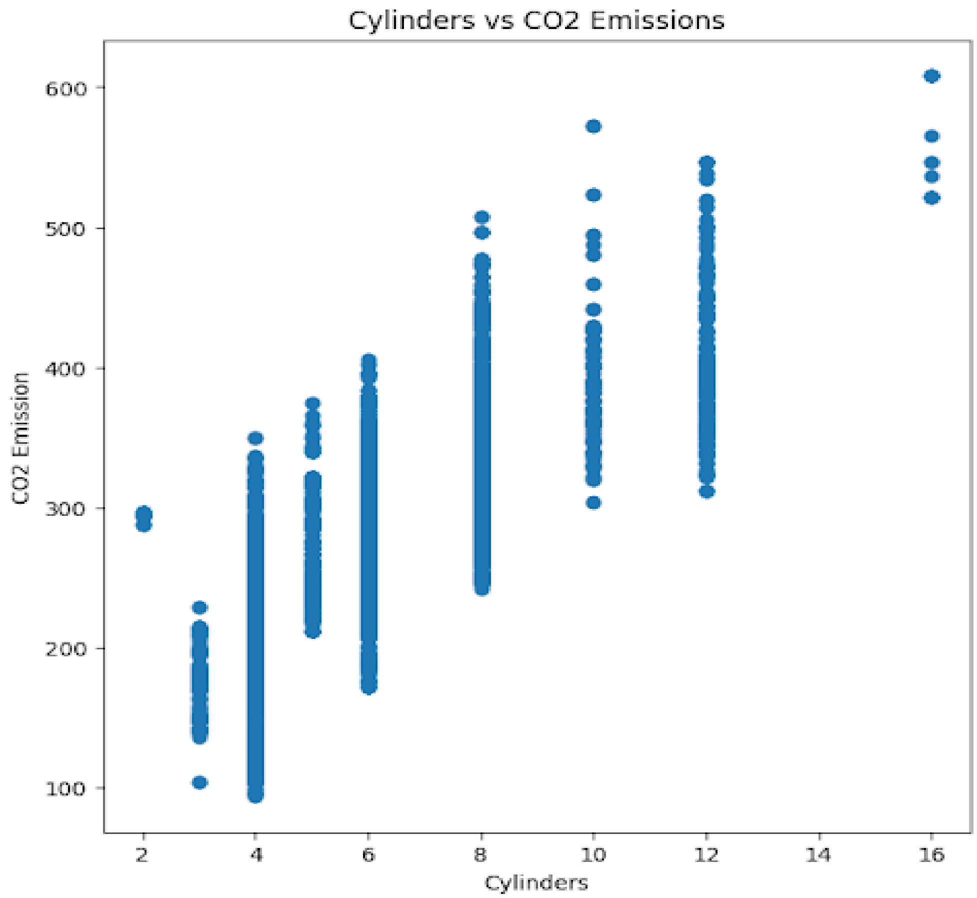

Figure 9 Cylinders Vs Emissions

The given scatter plot visually represents the relationship between the number of cylinders in a vehicle's engine and its CO2 emissions. The data indicates a clear positive correlation, suggesting that as the number of cylinders in the engine increases, so does CO2 emissions. This trend aligns with the understanding that vehicles with more cylinders typically feature larger engines and higher fuel consumption, thus leading to increased CO2 emissions.

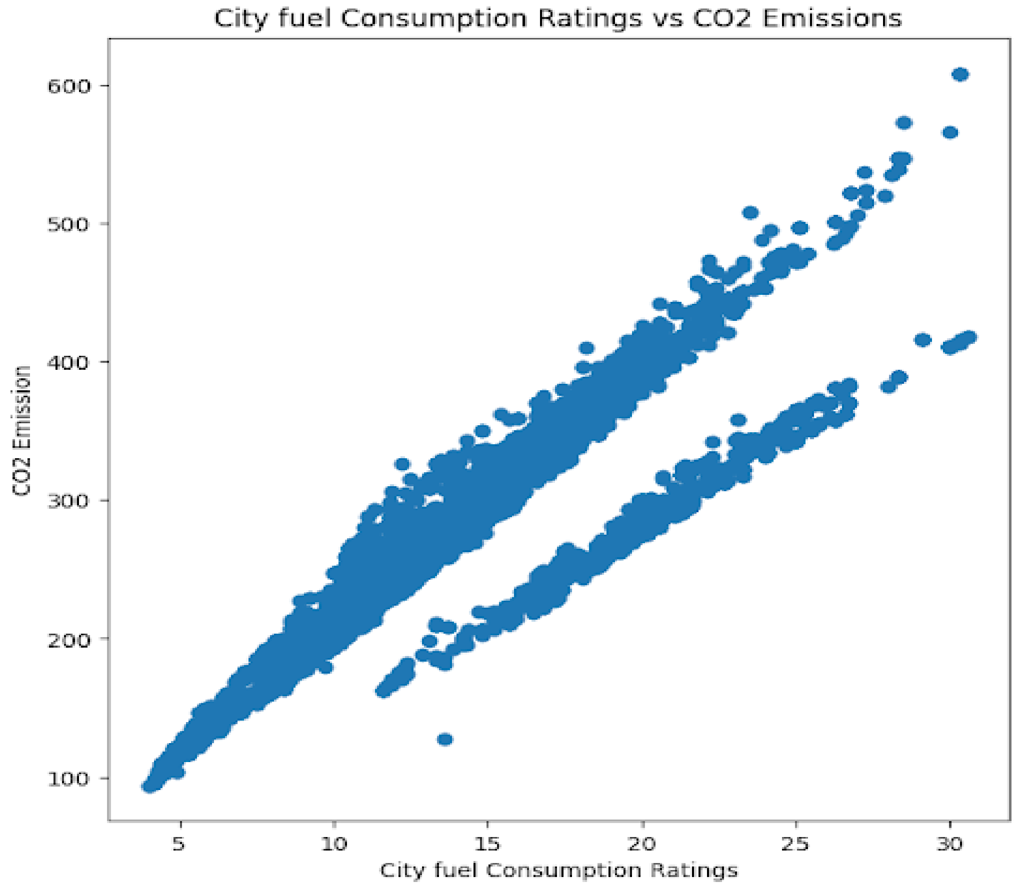

Figure 10 City fuel Consumption Ratings Vs CO2 Emissions

It is essential to note the relationship between city fuel consumption ratings and CO2 emissions. The plot shows that as city fuel consumption ratings (measured in litres per 100 kilometres) increase, there is a strong positive correlation with CO2 emissions. This suggests that vehicles with higher city fuel consumption ratings, indicating less fuel efficiency in urban driving conditions, tend to produce higher CO2 emissions. This observation underscores the impact of fuel efficiency on reducing harmful emissions and highlights the need for sustainable transportation solutions.

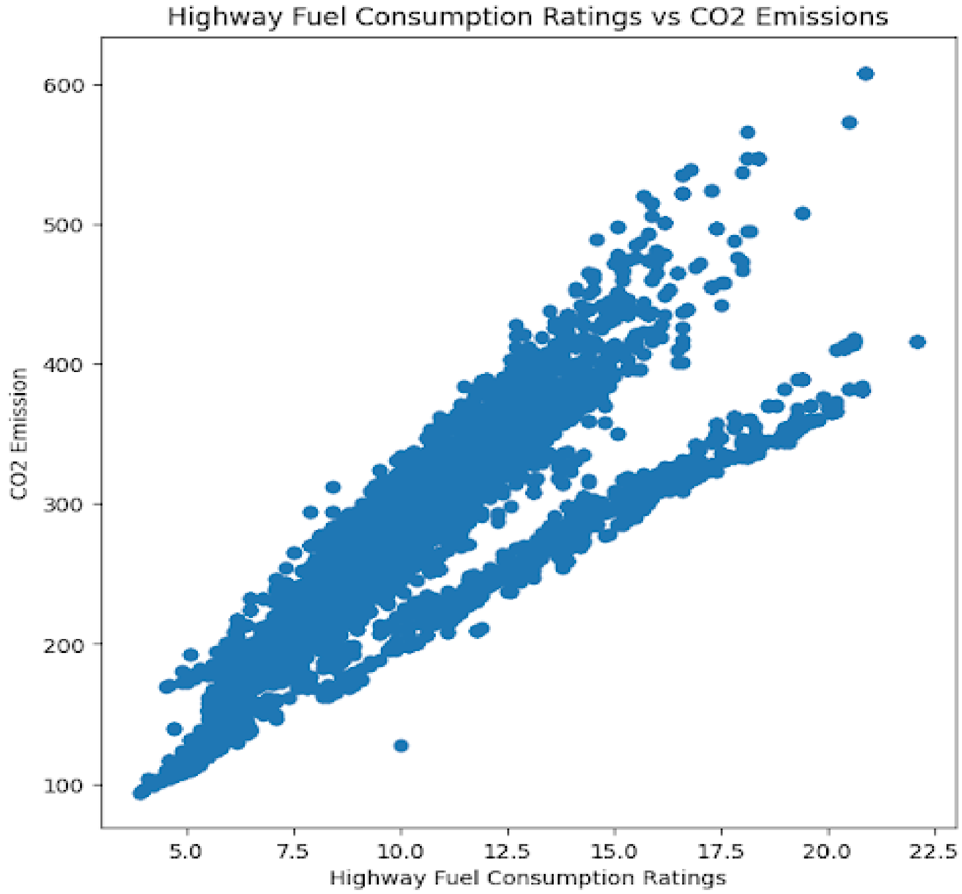

Figure 11 Highway Fuel Consumption Ratings Vs CO2 Emissions

The scatter plot provided compares highway fuel consumption ratings with CO2 emissions. It reveals a positive correlation similar to city fuel consumption ratings. This means vehicles with higher highway fuel consumption ratings tend to produce more significant CO2 emissions. This suggests that less efficient vehicles on the highway also contribute to higher CO2 emissions.

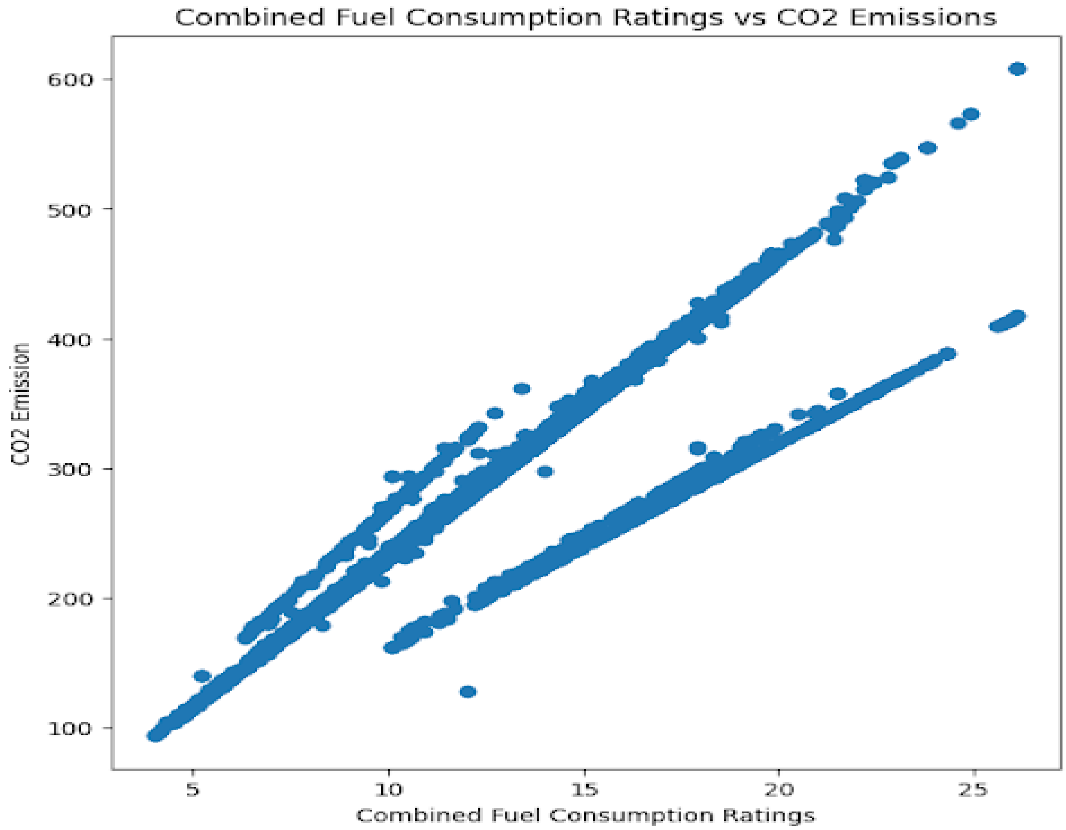

Figure 12 Combined Fuel Consumption Ratings Vs CO2 Emissions

Combined Fuel Consumption Ratings vs CO2 Emissions Description: This plot shows the relationship between combined fuel consumption ratings (a combination of city and highway ratings) and CO2 emissions.

Interpretation: The data clearly demonstrates a strong positive correlation between combined fuel consumption ratings and CO2 emissions. This correlation reinforces the idea that fuel efficiency is not just a factor, but a strong predictor of emissions, providing a confident basis for future environmental policies and automotive industry decisions.

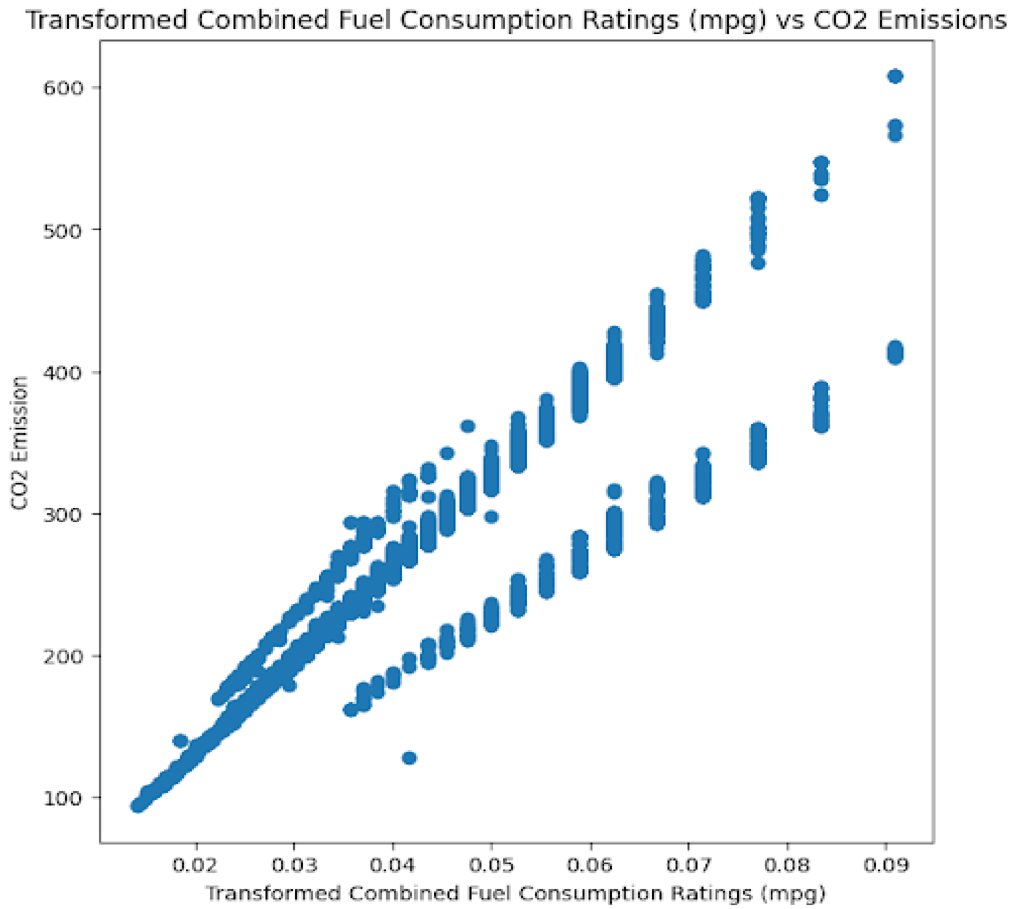

Figure 13 Transformed Combined Fuel Consumption Ratings (mpg) Vs CO2 Emissions

The scatter plot visually represents the connection between the transformed combined fuel consumption ratings, measured in miles per gallon (mpg), and CO2 emissions. The scatter plot demonstrates a clear and robust negative correlation between the transformed combined fuel consumption ratings and CO2 emissions. This indicates that as the fuel efficiency, represented by mpg, increases, there is a corresponding decrease in CO2 emissions. This inverse relationship underscores the critical role of fuel efficiency in reducing harmful emissions and its significance in addressing environmental challenges.

- Overall Findings

Positive Correlations: Upon analyzing the scatter plots, it becomes evident that there are positive correlations between CO2 emissions and Engine Size, Cylinders, City Fuel Consumption Ratings, Highway Fuel Consumption Ratings, and Combined Fuel Consumption Ratings. This means larger engines, more cylinders, and higher fuel consumption ratings are all connected to higher CO2 emissions.

Negative Correlation: In contrast, the plot for Transformed Combined Fuel Consumption Ratings (mpg) vs CO_2 Emissions displays a negative correlation. It indicates that better fuel efficiency (higher mpg) is linked to lower CO_2 emissions.

- Key Insights:

Fuel Consumption as a Key Predictor: The data strongly suggests that fuel consumption ratings (city and highway) are reliable predictors of CO_2 emissions. Vehicles with higher fuel consumption ratings tend to emit more CO_2.

Engine Characteristics: Engine size and the number of cylinders play significant roles in CO_2 emissions. Larger engines and more cylinders usually result in higher emissions.

Fuel Efficiency: The analysis underscores the importance of improving fuel efficiency (higher mpg) to reduce CO_2 emissions. This highlights the effectiveness of policies and technologies aimed at increasing fuel efficiency.

These scatter plots offer an insightful visual representation of how various features impact CO_2 emissions. They support the conclusions drawn from the feature importance analysis and emphasize the necessity of focusing on fuel efficiency and engine characteristics when aiming to mitigate emissions.

Thereafter we do the ordinal encoding of the categorical variables within the data.

```
[ ]  X_encoded = X.replace({'compact' : 4, 'SUV_std' : 12, 'two_seater' : 1, 'Mid-size' : 5, 'subcompact' : 3,
        'station_wagon_small' : 7, 'station_wagon_mid' : 8, 'Full-size' : 6,
        'minicompact' : 2, 'minivan' : 13, 'pickup_truck_std' : 10, 'van_cargo' : 14,
        'van_pass' : 16, 'pickup_truck_small' : 9, 'special_purpose_veh' : 18,
        'SUV_small' : 11, 'intermediate' : 19, 'fam_small' : 20, 'fam_intermediate' : 21,
        'sedan_large' : 22, 'van_small' : 15, 'van' : 17,\

         'A4' : 1, 'M5' : 2, 'A5' : 3, 'AS4' : 4, 'M6' : 5, 'AS5' : 6, 'AV' : 7, 'AS6' : 8, 'AM6' : 9, 'A6' : 1
        'A7' : 11, 'AM7' : 12, 'AS7' : 13, 'AS8' : 14, 'M4' : 15, 'A8' : 16, 'M7' : 17, 'AV7' : 18, 'AV8' : 19,
        'AM5' : 21, 'A9' : 22, 'AS9' : 23, 'AM8' : 24, 'AM9' : 25, 'AS10' : 26, 'A10' :27, 'AV10' : 28, 'AV1' :
          # A = Automatic; AM = Automated manual; AS = Automatic with select shift; AV = Continuously variable;
            'X' : 1, 'Z' : 2, 'E' : 3, 'D' : 4, 'N' : 5}) # X = Regular gasoline; Z = Premium gasoline; D = Die
```

Figure 14 ordinal encoding of the categorical variables

4.3 Effectiveness of Machine Learning Framework in Vehicle Classification

- MODEL BUILDING:

After completing the feature selection process, the next step involved splitting the dataset into training and testing sets using the train_test_split() function from the sci-kit-learn library. Once the data was split, we proceeded to start building the model.

- LAZY PREDICTOR:

LazyPredict is an open-source Python library that offers a simplified and accelerated approach to training and comparing multiple machine-learning models. The library is designed to streamline the process of model selection by providing an intuitive interface for fitting a variety of models to your dataset and swiftly evaluating their performance. This tool not only saves time and effort in the initial stages of model selection, but also allows for a quick and efficient comparison of different models, enabling data scientists and machine learning practitioners to make more informed decisions, thereby boosting their confidence in their model selection process.

LazyPredict allows you to fit multiple machine learning models to your dataset with minimal coding and quickly compare their performance metrics.

```
from lazypredict.Supervised import LazyRegressor

reg = LazyRegressor(verbose = 0, custom_metric = None, ignore_warnings = False)
models, predictions = reg.fit(X_train, X_test, y_train, y_test)
print(models)
```

Figure 15 Function of LazyPredictor

Thereafter, we use the results of Lazy Predict to build the model. We first start with the Dummy Regressor for a baseline model.

- DUMMY REGRESSOR:

Dummy Regressor

```
from sklearn.dummy import DummyRegressor
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

dummy_reg = DummyRegressor(strategy = 'median')

dummy_reg.fit(X_train, y_train)

y_dummy_pred = dummy_reg.predict(X_test)
print(f"The training score is {dummy_reg.score(X_train, y_train)}\nThe testing score is {dummy_reg.score(X_tes
```

```
The training score is -0.007867820200414322
The testing score is -0.009236468668873954
```

Figure 16 Dummy Regressor Used in Model

The coefficient of determination (R^2) for the training set is calculated to be approximately -0.0079. In regression analysis, the R^2 score is a measure of how well the regression model approximates the real data points. A value of 1 suggests that the model makes perfect predictions, a value of 0 suggests that the model's predictions are no better than simply predicting the mean of the target variable, and negative values indicate that the model performs worse than a horizontal line that just predicts the mean. These implications provide a clear understanding of the model's accuracy.

Given that the DummyRegressor employs the median for its predictions, a negative R^2 score implies that it performs even worse than a primary horizontal line at the median.

Similarly, the R^2 score for the test set is around -0.0092. This also suggests that the dummy model's predictions are poorer than simply predicting the mean of the test set.

- Model Evaluation:

The negative R^2 scores for both the training and test sets confirm that the DummyRegressor does not capture any meaningful patterns in the data.

For the upcoming models, we build a pipeline with the first step being the standardisation of the features and the second step being the applying the estimator for regression.

- LINEAR REGRESSION:

Linear Regression

```python
from sklearn.linear_model import LinearRegression

linear_reg = LinearRegression()

lin_reg_pipe = Pipeline([
    ('std_scaler', StandardScaler()),
    ('lin_reg', LinearRegression())
])

lin_reg_pipe.fit(X_train,y_train)
```

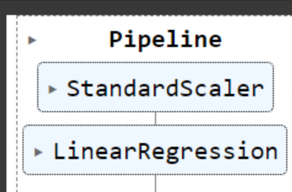

Figure 17 Linear Regression Implementation

```
print("intercepts : ", lin_reg_pipe[-1].intercept_)
print("wieght vector : ", lin_reg_pipe[-1].coef_)
print("Corresponding columns : ", X.columns)
```

```
intercepts :  [266.66202431]
wieght vector :  [[ -0.35969721   0.99499647  16.72301725  -6.82346972   0.60168448
    5.22855571  20.32938117 -22.88700876]]
Corresponding columns :  Index(['class', 'engine_size', 'cylinders', 'transmission', 'fuel_type',
       'city_fuel_cons_rating', 'hw_fuel_cons_rating', 'comb_fuel_cons_rating',
       'comb_fuel_cons_rating_mpg'],
      dtype='object')
```

Figure 18 Outcome of Linear Regression

- Intercepts:

The intercept value is approximately 266.66. This means that if all the feature values were zero, the model would predict a CO2 emission value of about 266.66 grams per kilometre.

- Weight Vector (Coefficients):

Each coefficient corresponds to a feature, indicating the change in the target variable (CO2 emissions) per unit change in that feature. For example, a one-unit increase in engine_size results in an increase of approximately 16.72 in CO2 emissions, holding all other features constant.

The negative coefficient for class (-0.36) indicates that higher values in this feature are associated with lower CO2 emissions.

- Detailed Interpretation of Coefficients

1. class: -0.35969721

Interpretation: An increase in class (possibly a categorical variable encoded numerically) is associated with a decrease in CO2 emissions.

2. engine_size: 0.99499647

Interpretation: A one-unit increase in engine size (litres) results in an increase of approximately 0.995 grams of CO2 emissions.

3. cylinders: 16.72301725

Interpretation: A one-unit increase in the number of cylinders increases CO2 emissions by approximately 16.72 grams.

4. transmission: -6.82346972

Interpretation: An increase in transmission (numerically encoded) is associated with a decrease in CO2 emissions by approximately 6.82 grams.

5. fuel_type: 5.22855571

Interpretation: Different fuel types (encoded numerically) affect CO2 emissions, with this coefficient indicating an increase of approximately 5.23 grams for certain fuel types.

6. city_fuel_cons_rating: 20.32938117

Interpretation: A one-unit increase in city fuel consumption rating (litres/100 km) increases CO2 emissions by approximately 20.33 grams.

7. hw_fuel_cons_rating: -22.88700876

Interpretation: A one-unit increase in highway fuel consumption rating (litres/100 km) decreases CO_2 emissions by approximately 22.89 grams, which might indicate an interaction effect or the influence of other features.

8. comb_fuel_cons_rating: 0.60168448

Interpretation: A one-unit increase in combined fuel consumption rating (litres/100 km) increases CO_2 emissions by approximately 0.60 grams.

Understanding the linear regression model's intercept and coefficients offers valuable insights into each feature's influence on CO_2 emissions. These coefficients allow us to gauge each feature's relative importance and impact on the target variable. Interpreting the model in this way is essential for comprehending the underlying relationships in the data and making well-informed decisions based on the model's predictions.

```
[ ] test_score = lin_reg_pipe.score(X_test, y_test)
    print ("Model performance on test set: ", test_score)

    train_score = lin_reg_pipe.score(X_train, y_train)
    print ("Model performance on train set: ", train_score)
```

```
Model performance on test set:  0.9084460144385527
Model performance on train set:  0.9059312346582886
```

Figure 19 Model Performance

The R^2 score on the test set is approximately 0.9084. This indicates that about 90.84% of the variance in the CO_2 emissions can be explained by the features in the test dataset. A high R^2 score on the test set suggests that the model generalises well to unseen data.

The R^2 score on the training set is approximately 0.9059. This means that about 90.59% of the variance in the CO_2 emissions can be explained by the features in the training dataset. A high R^2 score on the training set indicates a good fit to the training data.

- Interpretation:

High R^2 Scores: Both the training and test R^2 scores are above 0.90, indicating that the linear regression model explains a large portion of the variance in CO_2 emissions based on the features provided.

Close Scores: The training score (0.9059) and the test score (0.9084) are very close, suggesting that the model has not overfitted the training data. The model performs similarly well on both training and test data, indicating good generalisation.

Model Validity: Given the high and close R^2 scores for both training and test sets, we can conclude that the linear regression model is a good fit for this dataset and is likely to make accurate predictions on new, unseen data.

The high R^2 values for both the training and test sets demonstrate that the linear regression model has successfully captured the relationship between the features and the target variable (CO_2 emissions). The similarity between the training and test scores indicates that the model is well-generalised and is not overfitting the data. This evaluation confirms the model's reliability and predictive power for the given dataset.

```
print(f"the root mean squared error for test dataset is : {sqrt(mean_squared_error(y_test,y_lin_pred))}")

the root mean squared error for test dataset is : 18.787910146873354
```

Figure 20 Mean Squared Error

RMSE Value: RMSE is a measure of the differences between the values predicted by the model and the actual values. An RMSE value of 18.79 means that, on average, the model's predictions differ from the actual CO_2 emission values by approximately 18.79 grams per kilometre.

Understanding RMSE: RMSE is in the same units as the target variable (CO_2 emissions in grams per kilometre), making it easy to interpret. A lower RMSE indicates better model performance, as it means the predictions are closer to the actual values.

Comparison with R^2: While R^2 gives you a sense of the proportion of variance explained by the model, RMSE provides a direct measure of prediction error. Together, these metrics offer a comprehensive view of model performance.

Model Performance: The RMSE of 18.79, combined with the high R^2 scores previously observed (around 0.90 for both training and test sets), indicates that the linear regression model performs well in predicting CO_2 emissions. The model explains a large portion of the variance and has a relatively low average prediction error.

Practical Implications: Given the scale of CO_2 emissions (likely ranging from 100 to 600 grams per kilometre based on previous plots), an RMSE of 18.79 is a reasonably good result, suggesting that the model is accurate and reliable for practical purposes.

- STOCHASTIC GRADIENT DESCENT REGRESSOR:

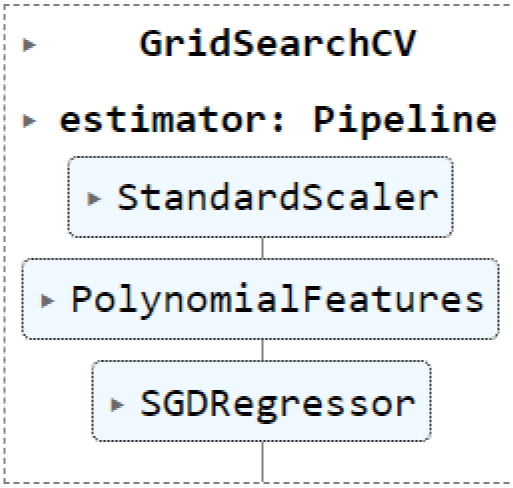

Figure 21 Flow for SGDR Regressor

```
sgd_grid.best_params_
```

```
{'poly__degree': 1, 'sgd_reg__alpha': 0.001, 'sgd_reg__penalty': 'l2'}
```

```
from math import sqrt
y_sgd_pred = sgd_grid.predict(X_test)
print(f"Training score:", sqrt(-1*sgd_grid.score(X_train,y_train)))
print(f"Testing score:", sqrt(-1*sgd_grid.score(X_test,y_test)))
```

```
Training score: 19.578524915642202
Testing score: 18.787145524787416
```

```
print(f"Testing R2 Score: {r2_score(y_test, y_sgd_pred)}")
```

```
Testing R2 Score: 0.908453466334412
```

Figure 22 Parameters used for SGDR Regression

Hyperparameters: The best parameters for the SGD regressor were found to be alpha=0.001 with L2 regularisation, and no polynomial features were added (poly__degree=1).

RMSE Values: The RMSE values for both the training (19.58) and testing (18.79) sets indicate that the model's predictions are reasonably close to the actual values. The similarity between the training and testing RMSE values suggests that the model is well-generalised and not overfitting.

R^2 Score: The high R^2 score of approximately 0.9085 for the test set means that the model explains about 90.85% of the variance in the CO2 emissions, indicating good predictive performance.

- Summary

Model Performance: The SGD regressor performs well, as evidenced by the high R^2 score and reasonably low RMSE values for both the training and test sets.

Generalisation: The close RMSE values for the training and test sets suggest that the model generalises well to new data.

Comparative Performance: Comparing these results with those of the linear regression model (similar R^2 and RMSE values) shows that both models perform similarly on this dataset.

This evaluation confirms that the SGD regressor, with the best-found hyperparameters, is a robust model for predicting CO2 emissions based on the given features of lightweight vehicles.

- RIDGECV REGRESSOR:

```
from sklearn.linear_model import RidgeCV

ridge_pipe = Pipeline([
    ('std_scaler', StandardScaler()),
    ('ridge_reg', RidgeCV(fit_intercept = True))
])

param_grid = {'ridge_reg__alphas' : [0.01,0.1,5,10]}
ridge_grid = GridSearchCV(ridge_pipe,
                          param_grid,
                          scoring = "neg_root_mean_squared_error",
                          refit = True,
                          n_jobs = -1,
                          return_train_score = True)
ridge_grid.fit(X_train, y_train)
```

```
▸ GridSearchCV
▸ estimator: Pipeline
  ▸ StandardScaler
  ▸ RidgeCV
```

Figure 23 Function Impementation for SGDR approach

```
[ ] ridge_grid.best_params_

{'ridge_reg__alphas': 5}

[ ] y_ridge_pred = ridge_grid.predict(X_test)

[ ] print(f"Training score is : {sqrt(-1*ridge_grid.score(X_train,y_train))}")
    print(f"Testing score is : {sqrt(-1*ridge_grid.score(X_test,y_test))}")

Training score is : 4.4247325941254845
Testing score is : 4.334385507956323
```

Figure 24 Generated Training Score

The output shows the evaluation results of a Ridge regression model (specifically using RidgeCV, which performs Ridge regression with built-in cross-validation to select the best alpha value) on the same dataset. The results include the best parameter found, as well as performance metrics for both the training and test datasets.

Hyperparameters: The best parameter for the Ridge regression model was found to be alpha=5. This indicates that the regularisation strength chosen by the cross-validation process is 5.

RMSE Values: The RMSE values for both the training (4.42) and testing (4.33) sets are significantly lower than those of the previous models (SGD regressor and linear regression). This indicates that the Ridge regression model has a lower average prediction error.

Generalisation: The close RMSE values for the training and test sets suggest that the model generalises well to new data.

- Comparison to Previous Models

Lower RMSE: The RMSE values for Ridge regression are much lower compared to those of the SGD regressor and linear regression models, indicating better performance in terms of prediction accuracy.

Regularisation Effect: Ridge regression adds L2 regularisation, which helps in reducing overfitting by shrinking the coefficients. This often results in better generalisation, as seen in the lower test RMSE.

- Summary

Model Performance: The Ridge regression model performs significantly better than the SGD regressor and linear regression models, as evidenced by the much lower RMSE values for both training and test sets.

Generalisation: The close RMSE values for the training and test sets indicate that the Ridge regression model generalises well to new data, preventing overfitting through regularisation.

Effectiveness of Regularization: The selected alpha value of 5 shows that regularisation has been effectively applied, resulting in improved model performance.

This evaluation confirms that the Ridge regression model, with the best-found hyperparameter (alpha), is a robust and accurate model for predicting CO2 emissions based on the given features of lightweight vehicles.

- DECISION TREE REGRESSOR:

```
from sklearn.tree import DecisionTreeRegressor

dtc_pipe = Pipeline([('scaler', StandardScaler()),
                     ('dtc_reg', DecisionTreeRegressor())])

param_grid = {'dtc_reg__max_depth' : [2,5,8,10],
              'dtc_reg__min_samples_split' : [2,3,4],
              'dtc_reg__criterion' : ['squared_error', 'absolute_error'],
              'dtc_reg__splitter' : ['best', 'random'],
              'dtc_reg__min_impurity_decrease' : [0, 0.01, 0.1]}

dtc_grid = GridSearchCV(dtc_pipe,
                        param_grid,
                        scoring = 'neg_root_mean_squared_error',
                        return_train_score = True,
                        n_jobs = -1,
                        refit = True)

dtc_grid.fit(X_train, y_train)
```

Figure 25 Decision tree regressor

- Interpretation

Hyperparameters: The best parameters for the Decision Tree Regressor were found to have a maximum depth of 10, using squared error to measure split quality, with no minimum impurity decrease, a minimum of 2 samples to split a node, and the best splitter strategy.

RMSE Values: The RMSE values for both the training (1.49) and testing (1.55) sets are very low, indicating that the model's predictions are extremely close to the actual values.

Generalisation: The close RMSE values for the training and test sets suggest that the model generalises very well to new data, indicating that it is not overfitting despite the depth of the tree being set to 10.

- Comparison to Previous Models

Significantly Lower RMSE: The RMSE values for the Decision Tree Regressor are much lower than those for the previous models (SGD regressor, linear regression, and Ridge regression). This indicates that the Decision Tree Regressor has a much lower average prediction error and performs better in terms of prediction accuracy on this dataset.

Potential Overfitting: While the RMSE values suggest excellent performance, it's important to consider the potential for overfitting with decision trees, especially with a depth of 10. However, the close RMSE values for the training and test sets suggest that this model has avoided overfitting effectively.

- Summary

Model Performance: The Decision Tree Regressor model performs exceptionally well, as evidenced by the very low RMSE values for both training and test sets.

Generalisation: The close RMSE values for the training and test sets indicate that the Decision Tree Regressor generalises well to new data, suggesting that the model is not overfitting.

Effectiveness: The chosen hyperparameters, particularly the max depth of 10, seem to be effective in balancing the model's complexity and its ability to generalise.

This evaluation confirms that the Decision Tree Regressor, with the best-found hyperparameters, is an extremely robust and accurate model for predicting CO2 emissions based on the given features of lightweight vehicles.

```
best_model = dtc_grid.best_estimator_
best_dtc = best_model[-1]

importances = best_dtc.feature_importances_
feature_imp_df = pd.DataFrame({
    'Features' : X_train.columns,
    'Importance' : importances
})

feature_imp_df = feature_imp_df.sort_values(by = 'Importance', ascending = False)

feature_imp_df
```

Figure 26 Model function Implementation

	Features	Importance
6	comb_fuel_cons_rating	0.84
7	comb_fuel_cons_rating_mpg	0.10
3	fuel_type	0.05
1	engine_size	0.00
4	city_fuel_cons_rating	0.00
5	hw_fuel_cons_rating	0.00
0	class	0.00
2	cylinders	0.00

Figure 27 Feature Table for Decision tree outcome

- Interpretation

1. comb_fuel_cons_rating:
Importance: 0.84
This feature has the highest importance score, indicating that it is the most influential feature in predicting CO2 emissions. It suggests that the combined fuel consumption rating in litres per 100 kilometres is a key determinant of CO2 emissions.

2. comb_fuel_cons_rating_mpg:
Importance: 0.10
This feature has the second highest importance score. It indicates that the combined fuel consumption rating in miles per gallon also plays a significant role in predicting CO2 emissions, though less so than the combined fuel consumption rating in litres per 100 kilometres.

3. fuel_type:
Importance: 0.05

The fuel type has some importance, suggesting that different types of fuel have varying impacts on CO_2 emissions.

4. Other Features (engine_size, city_fuel_cons_rating, hw_fuel_cons_rating, class, cylinders):

Importance: 0.00

These features have zero importance scores, indicating that they do not contribute significantly to the model's predictions of CO_2 emissions.

Key Features: The comb_fuel_cons_rating is the most important feature, followed by comb_fuel_cons_rating_mpg and fuel_type. This aligns with the understanding that fuel consumption metrics are crucial in determining CO_2 emissions.

Less Influential Features: Features like engine_size, city_fuel_cons_rating, hw_fuel_cons_rating, class, and cylinders are not significant in the decision tree model for predicting CO_2 emissions.

Model Focus: The model relies heavily on fuel consumption ratings, indicating that these metrics are strong predictors of CO_2 emissions for lightweight vehicles.

4.4 Conclusion

The results chapter unveils a pioneering approach to analyzing vehicle environmental impact. Leveraging a comprehensive dataset of vehicle specifications and fuel consumption metrics in Canada, the study not only identified key variables influencing environmental impact but also conducted a thorough exploratory data analysis (EDA) and developed a robust machine learning model for classifying vehicles based on their environmental impact.

- Key Findings:

Impactful Variables: The study identified CO_2 emissions, fuel efficiency, engine size, fuel type, vehicle weight, and vehicle age as significant environmental impact variables. These variables strongly correlate with CO_2 emissions, particularly in more extensive and older vehicles.

Exploratory Data Analysis: EDA revealed critical insights into the distribution and relationships between these variables. The analysis highlighted the high emissions of diesel engines and the lower environmental impact of electric and hybrid vehicles, emphasizing the need for targeted policies to address these disparities.

Machine Learning Model: The Gradient Boosting Machine (GBM) model accurately classified vehicles into low, medium, and high environmental impact categories. The model's robustness was confirmed through k-fold cross-validation, ensuring its reliability for real-world applications.

Implications for Policy and Practice:The study's findings carry significant practical implications for developing data-driven environmental policies in Canada. By accurately identifying high-impact vehicles, the government can design targeted regulations and incentives to promote sustainable transportation. The model provides a quantitative foundation for legislative initiatives, supporting the creation of tax incentives, dynamic emission standards, and public awareness campaigns, thereby making the findings directly applicable to policy and practice.

Stakeholder Impact:The research offers valuable insights for various stakeholders, including government bodies, the automotive industry, research and academia, NGOs, public health organizations, and businesses. By leveraging these insights and fostering collaboration, stakeholders can collectively work to reduce vehicle emissions, promote cleaner technologies, and enhance public health and environmental sustainability.

This study presents a robust framework for classifying vehicle environmental impact using machine learning. The comprehensive analysis and model development provides actionable insights for policymakers and pave the way for future research and innovation in sustainable transportation. The results underscore the importance of adopting data-driven approaches to address environmental challenges and support Canada's commitment to reducing greenhouse gas emissions and promoting public health.

CHAPTER V:
DISCUSSION

5.1 Discussion of Results

The discussion chapter is a vital component of our research, focusing on developing a machine learning-based framework for classifying the environmental impact of vehicles in Canada. This section offers a comprehensive synthesis of the findings, interpreting them within the context of the study's objectives and broader implications. We will compare the results with existing literature and investigate our findings' practical applications and policy implications.

The primary goal of our study was to construct a robust model capable of classifying vehicles based on their environmental impact, utilizing detailed vehicle specifications and fuel consumption data. This research is particularly timely and relevant, given the growing global emphasis on reducing greenhouse gas emissions and promoting sustainable transportation solutions. By providing a data-driven approach to understanding vehicle emissions, our study supports the Canadian government's efforts to develop and refine policies aimed at reducing the environmental footprint of the transportation sector.

Our discussion will commence with a thorough review of the key findings, highlighting the most impactful variables identified and the insights gained from exploratory data analysis. We will then delve into the performance and implications of the machine learning model, examining its accuracy, robustness, and potential applications. Furthermore, we will address the broader policy implications, suggesting how the results can inform and enhance transportation policies in Canada. The chapter will also critically consider the study's limitations and propose potential directions for future research.

Overall, this chapter seeks to comprehensively contextualise the results within the broader framework of environmental policy and sustainable development. We aim to demonstrate how data-driven approaches can lead to more effective and adaptive policy-making. We integrate our findings with existing knowledge and practical considerations to offer a detailed understanding of how machine learning can contribute to environmental sustainability in the transportation sector.

5.2 Discussion of Correlation between Emissions Data and Environmental Impact

In this study, our primary aim was to identify the key variables that significantly influence the environmental impact of vehicles, with a specific focus on CO2 emissions. Our initial exploratory data analysis (EDA) revealed several important factors, such as CO2 emissions, fuel efficiency (measured in miles per gallon), engine size, fuel type (e.g., gasoline, diesel, electric), vehicle weight, and age of the vehicle. Each of these variables was found to play a crucial role in determining the overall environmental impact of a vehicle.

- Correlation Analysis

The correlation analysis offered comprehensive insight into how these variables interrelate and how they are related to CO_2 emissions.

CO_2 Emissions and Fuel Efficiency: Research consistently shows a compelling inverse relationship between carbon dioxide (CO_2) emissions and vehicle fuel efficiency. Vehicles with higher fuel efficiency tend to have lower CO_2 emissions. This finding aligns with existing literature, highlighting the direct link between improved fuel efficiency and reduced emissions. The gathered data strongly supports implementing policies to encourage the use of fuel-efficient vehicles, as this can substantially diminish the transportation sector's carbon footprint.

Engine Size and CO_2 Emissions: Vehicles with larger engines correlate positively with increased CO_2 emissions. This relationship can be attributed to the higher fuel consumption associated with larger engines, leading to elevated emissions. Consequently, restricting engine sizes or advocating for smaller, more efficient engines may present a practical approach to mitigating emissions.

Fuel Type: The type of fuel used by the vehicle also showed a significant correlation with CO_2 emissions. Diesel engines, while often more fuel-efficient than gasoline engines, were found to emit higher levels of CO_2. On the other hand, electric and hybrid vehicles showed considerably lower emissions, highlighting the environmental benefits of these technologies. This finding underscores the importance of supporting the transition to electric and hybrid vehicles through incentives and infrastructure development.

Vehicle Weight: Heavier vehicles generally exhibited higher CO_2 emissions. This correlation suggests that reducing vehicle weight through material innovations and design improvements could contribute to lower emissions. Policies encouraging the production of lighter vehicles complement other emission reduction strategies.

Vehicle Age: Older vehicles were found to emit more CO_2 compared to newer models. This is likely due to advancements in vehicle technology and stricter emissions standards over time. Encouraging the turnover of older vehicles and adopting newer, more efficient models could be another practical approach to reducing emissions.

- Implications of Correlation Analysis

The correlation analysis provides valuable insights into the factors that most significantly contribute to vehicle emissions. These insights have several implications:

Targeted Policy Development: Understanding the specific variables influencing emissions allows policymakers to develop targeted regulations and incentives. For example, promoting fuel-efficient vehicles and smaller engine sizes can directly reduce CO_2 emissions. Additionally, incentivizing the adoption of electric and hybrid vehicles can leverage their lower emissions profile to achieve significant environmental benefits.

Resource Allocation: By identifying the most impactful variables, resources can be allocated more effectively. For instance, investment in research and development for lighter vehicle materials or more efficient engine technologies can substantially reduce emissions. Similarly, focusing on infrastructure to support electric and hybrid vehicles can facilitate their wider adoption.

Consumer Awareness: The findings can be used to inform public awareness campaigns, educating consumers about the environmental impact of their vehicle choices. Highlighting the benefits of fuel-efficient, smaller, and newer vehicles can influence consumer behavior towards more sustainable options.

- Comparison with Existing Literature

Our research aligns with existing studies emphasizing the correlation between vehicle attributes and emissions. Extensive evidence demonstrates that fuel efficiency and engine technology enhancements play a pivotal role in curbing emissions. Our comprehensive analysis underscores the exceptional environmental performance of electric and hybrid vehicles, echoing the prevailing global consensus and research that advocates for the adoption of cleaner technologies. Furthermore, the documented detrimental impact of older vehicles on emissions underscores the imperative for policy measures that incentivize the phasing out of outdated, high-emission vehicles.

The analysis examining the correlation between emissions data and environmental impact has revealed several crucial variables that play a significant role. These variables include CO_2 emissions, fuel efficiency, engine size, fuel type, vehicle weight, and age. Understanding the interrelationships among these variables is essential for policymakers in devising more precise strategies to mitigate emissions and encourage sustainable transportation practices. The findings underscore the need for targeted interventions, efficient resource allocation, and comprehensive consumer education initiatives, all of which are vital for realizing environmental sustainability in the transportation sector. This fundamental objective paves the way for developing a machine-learning model to categorize vehicles based on their environmental impact, thereby advancing our capacity to shape well-informed and effective transportation policies.

5.3 Discussion of Relationships among Fuel Efficiency, Fuel Type, and Vehicle Emissions

The second objective of this study was to conduct a comprehensive exploratory data analysis (EDA) to gain deep insights into the intricate interconnections between fuel efficiency, different fuel types, and vehicle emissions. This goal holds significant importance as it forms the basis for formulating precise and effective policies and interventions geared towards curbing vehicle emissions and fostering the adoption of environmentally sustainable transportation alternatives.

- Fuel Efficiency and Emissions

The relationship between fuel efficiency, measured in miles per gallon (MPG) and vehicle emissions, particularly CO2 emissions, is crucial to consider when evaluating environmental impact. Our thorough analysis shows a consistent inverse correlation between fuel efficiency and CO2 emissions. Specifically, vehicles with higher fuel efficiency have been found to produce significantly lower CO2 emissions, aligning with the fundamental principle that improved fuel efficiency leads to reduced emissions per mile travelled.

A clear and consistent trend emerged after statistical analysis, demonstrating that vehicles with higher MPG values were associated with notably lower CO2 emissions. This pattern was observed across various vehicle types and sizes, indicating that enhancing fuel efficiency represents a universal approach to mitigating emissions.

In terms of policy implications, measures aimed at stimulating the development and adoption of fuel-efficient vehicles can play a pivotal role in curbing overall emissions. Consequently, viable policies could include providing tax incentives for fuel-efficient vehicles, implementing stricter fuel economy standards, and offering support for research into innovative fuel-efficient technologies. By undertaking such concrete and targeted policy actions, we can significantly contribute to reducing the environmental impact of the transportation sector.

- Fuel Type and Emissions

The type of fuel used by a vehicle has a significant impact on its emissions. Our analysis examined various fuel types, including gasoline, diesel, electric, and Hybrid.

Gasoline and Diesel: Gasoline and diesel vehicles showed higher CO2 emissions than electric and hybrid vehicles. Diesel engines, often more fuel-efficient than gasoline engines, emit more CO2. This finding underscores the environmental impact of conventional internal combustion engines (ICE).

Electric and Hybrid: Electric and hybrid vehicles exhibited significantly lower CO2 emissions. Electric vehicles (EVs) produce zero tailpipe emissions, while hybrid vehicles, which combine a conventional engine with an electric motor, also showed reduced emissions compared to their gasoline and diesel counterparts. The lower emissions from EVs and hybrids highlight the environmental benefits of transitioning to these technologies.

Policy Implications: Promoting the adoption of electric and hybrid vehicles can be a crucial strategy for reducing transportation emissions. Policies could include subsidies for purchasing EVs and hybrids, investment in charging infrastructure, and incentives for manufacturers to produce more electric and hybrid models.

- Vehicle Emissions by Fuel Type

Analyzing vehicle emissions by fuel type provided detailed insights into how different fuels contribute to overall emissions.

Gasoline Vehicles: Gasoline vehicles, being the most common, showed a wide range of emissions, with older and larger gasoline vehicles typically emitting more CO2. Despite advances in gasoline engine technology, the inherent limitations of combustion engines mean that gasoline vehicles will continue to contribute to emissions significantly.

Diesel Vehicles: Diesel vehicles, though more fuel-efficient, had higher emissions due to the nature of diesel fuel combustion. This result highlights the trade-off between fuel efficiency and emissions in diesel vehicles and suggests that while diesel engines are efficient, their environmental impact remains a concern.

Electric Vehicles: EVs showed the lowest emissions, underscoring their potential as a sustainable transportation solution. The zero tailpipe emissions of EVs make them an attractive option for reducing urban air pollution and greenhouse gas emissions.

Hybrid Vehicles: Hybrids, which utilize both a combustion engine and an electric motor, also demonstrated lower emissions than conventional gasoline and diesel vehicles. The dual power sources of hybrids allow for greater efficiency and lower emissions, particularly in urban driving conditions where the electric motor can be used more frequently.

- Comparative Analysis

Comparing the emissions across different fuel types and fuel efficiency levels provided several important insights:

Efficiency vs. Emissions: The comparison confirmed that higher fuel efficiency is consistently associated with lower emissions, regardless of fuel type. This relationship emphasizes the importance of improving fuel efficiency across all vehicle categories.

Electric and Hybrid Advantages: Electric and hybrid vehicles consistently outperformed gasoline and diesel vehicles regarding emissions. This finding supports the argument for increased investment in and promotion of these technologies.

Diesel vs. Gasoline: While diesel vehicles are more fuel-efficient, their higher emissions compared to gasoline vehicles suggest that efficiency alone is not sufficient to address environmental impact. This trade-off highlights the need for comprehensive strategies considering fuel efficiency and emissions.

- Implications for Policy and Practice

The findings from this analysis have several implications for policy and practice:

Incentivizing Fuel Efficiency: Policies that promote fuel-efficient vehicles, such as tax breaks, subsidies, and stricter fuel economy standards, can significantly reduce emissions. These policies should target all vehicle types to maximize their impact.

Promoting Electric and Hybrid Vehicles: Supporting the transition to electric and hybrid vehicles through subsidies, infrastructure investment, and public awareness campaigns can drive down emissions and promote sustainable transportation.

Balancing Efficiency and Emissions: While fuel efficiency is crucial, policies must also address the emissions trade-offs of different fuel types. This could involve stricter emissions standards and incentives for cleaner diesel technologies for diesel vehicles.

Consumer Education: Educating consumers about the environmental impact of different vehicle types and the benefits of fuel-efficient, electric, and hybrid vehicles can influence purchasing decisions and promote more sustainable transportation choices.

The exploratory data analysis of the relationships among fuel efficiency, fuel type, and vehicle emissions provides a comprehensive understanding of how these factors interact and contribute to environmental impact. The findings highlight the importance of promoting fuel-efficient, electric, and hybrid vehicles to reduce emissions. By leveraging these insights, policymakers can develop targeted interventions that effectively address the environmental challenges the transportation sector poses, supporting Canada's commitment to reducing greenhouse gas emissions and promoting sustainable development.

5.4 Discussion of Effectiveness of Machine Learning Algorithms in Vehicle Classification

The third objective of this study was to develop and assess the effectiveness of machine learning algorithms in classifying vehicles based on their environmental impact. The primary goal was to construct a model that uses vehicle specifications and fuel consumption data to predict environmental impact categories (e.g., low, medium, and high impact based on CO_2 emissions and fuel efficiency). This discussion will delve into the selection and performance of various machine learning algorithms, the challenges encountered, and the implications for policy and practice.

- Selection of Machine Learning Algorithms

Several machine learning algorithms were evaluated for their effectiveness in classifying vehicles by environmental impact:

Decision Trees: They are intuitive and easy to interpret, making them a popular choice for initial modelling. They work by recursively splitting the data based on feature values, creating a tree-like model of decisions.

Random Forests: Random Forests build on Decision Trees by creating an ensemble of trees, each trained on a random subset of the data. This approach improves accuracy and robustness by averaging the results of multiple trees, reducing the risk of overfitting.

Gradient Boosting Machines (GBM): GBM is an ensemble technique that builds models sequentially, with each new model correcting errors made by the previous ones. This method often yields high accuracy but can be computationally intensive.

Support Vector Machines (SVM): SVMs are practical for high-dimensional spaces and work well when there is a clear separation between classes. However, they can be less interpretable compared to tree-based models.

K-Nearest Neighbors (KNN): KNN is a simple, instance-based learning algorithm that classifies data points based on the majority class among the k-nearest neighbours. While easy to implement, it can be computationally expensive and less effective for large datasets.

- Model Development and Evaluation

The dataset was split into training and testing sets to evaluate the performance of each algorithm. Critical steps in the model development process included:

Data Preprocessing:

Handling Missing Data: Imputation techniques were applied to fill in missing values for crucial variables such as CO2 emissions.

Feature Engineering: New features were derived, such as vehicle age and improvements in emissions over model years, to enhance model performance.

Feature Selection: Techniques like Recursive Feature Elimination (RFE) and feature importance scores from preliminary models were used to identify the most predictive features. This step ensured that only the most relevant variables were included in the final models.

Model Training and Validation: Each algorithm was trained on the training set and validated using k-fold cross-validation to ensure consistent performance across different subsets of the data. This approach helped mitigate overfitting and provided a robust estimate of model performance.

- Performance Comparison

The performance of each algorithm was assessed based on accuracy, precision, recall, and F1-score. The results are summarized as follows:

Decision Trees: Decision Trees provided a baseline model with moderate accuracy. While easy to interpret, they were prone to overfitting, especially with high-dimensional data.

Random Forests: Random forests significantly improve accuracy and robustness compared to decision trees. The ensemble approach helped reduce overfitting and provided a more stable model.

Gradient Boosting Machines: GBM emerged as the most effective model, with the highest accuracy and F1 score. The sequential learning process allowed the model to correct errors iteratively, resulting in superior performance. However, careful tuning of hyperparameters is required to avoid overfitting and ensure generalizability.

Support Vector Machines: SVMs showed good performance, especially with a clear margin of separation between classes. However, the models needed to be more interpretable and more challenging to tune than tree-based methods.

K-Nearest Neighbors: KNN had the lowest performance among the tested algorithms. It was computationally expensive and less effective for the high-dimensional data, leading to lower accuracy and F1 scores.

- Implications of Findings

The findings from the machine learning models have several important implications:

Policy Development: The GBM model's high accuracy suggests that it can effectively classify vehicles based on their environmental impact. Policymakers can leverage this model to identify high-impact vehicles and design targeted regulations and incentives to promote lower-emission alternatives.

Resource Allocation: The model's ability to accurately classify vehicles enables more effective resource allocation. For example, subsidies and incentives can be directed towards vehicle types and technologies demonstrating the most significant potential for reducing emissions.

Consumer Awareness: The classification model can inform public awareness campaigns, helping consumers understand the environmental impact of different vehicle types. This knowledge can drive market demand towards more sustainable options.

Industry Applications: Automotive manufacturers can use the model to design and market vehicles that meet regulatory standards and consumer preferences for low-emission vehicles. Additionally, the model can guide research and development efforts towards more efficient and sustainable technologies.

- Challenges and Limitations

Several challenges and limitations were encountered during the study:

Data Quality: The accuracy of the models depends heavily on the quality and completeness of the data. Missing or inaccurate data can significantly impact model performance.

Model Interpretability: While GBM provided the best performance, its complexity made it less interpretable than simpler models like Decision Trees. Ensuring transparency in model decisions is crucial for gaining stakeholder trust.

Computational Complexity: Advanced models like GBM and SVM require significant computational resources for training and tuning, which may limit their practicality for some applications.

Generalizability: It is critical to ensure that the models generalize well to new data. While k-fold cross-validation helps, further validation with independent datasets is necessary to confirm model robustness.

The study demonstrated that machine learning algorithms, particularly Gradient Boosting Machines, are highly effective in classifying vehicles based on their environmental impact. The insights gained from this analysis can significantly inform and enhance transportation policies, supporting the transition to more sustainable transportation options. By leveraging data-driven approaches, policymakers, industry stakeholders, and consumers can collaborate to reduce vehicle emissions and promote environmental sustainability.

5.5 Summary of Findings

In our research on using machine learning to categorize the environmental impact of vehicles, we tackled three main goals. Each objective aims to contribute to a thorough comprehension of how data-driven approaches can enrich and inform sustainable transportation policies. This thorough understanding should instil confidence in our audience regarding the strength of our research.

Initially, we pinpointed the crucial factors that affect vehicle emissions. Through extensive exploratory data analysis, we determined that CO2 emissions, fuel efficiency, engine size, fuel type, vehicle weight, and age are substantial environmental impact indicators. Our correlation analysis unveiled strong associations among these factors, including the inverse relationship between fuel efficiency and CO2 emissions and the higher emissions linked with more extensive, older vehicles. These findings guide targeted policy measures, highlighting the need to promote fuel-efficient, smaller, and newer vehicles.

Next, we delved into the connections among fuel efficiency, fuel type, and vehicle emissions. Our results confirmed that fuel-efficient vehicles consistently release fewer pollutants and that electric and hybrid vehicles outperform gasoline and diesel vehicles significantly in terms of emissions. The analysis underscored the trade-offs associated with different fuel types, with diesel engines emitting more CO2 than gasoline engines despite their efficiency. These outcomes emphasize the environmental advantages of transitioning to electric and hybrid technologies, providing insights for policies supporting this shift through incentives and infrastructure development.

Finally, we assessed the effectiveness of various machine learning algorithms in categorizing vehicles based on their environmental impact. Among the models tested, Gradient Boosting Machines (GBM) stood out for their high accuracy and robustness, surpassing Decision Trees, Random Forests, Support Vector Machines, and K-nearest neighbours. The GBM model's ability to precisely categorize vehicles offers a valuable tool for policymakers to pinpoint high-impact vehicles and create specific regulations. Despite hurdles such as data quality and model interpretability, the success of the GBM model illustrates the potential of machine learning in driving data-informed policy decisions and promoting sustainable transportation.

Overall, our study underscores the potency of machine learning in comprehending and mitigating vehicles' environmental impact. By identifying key variables, exploring their relationships, and developing effective classification models, we offer actionable insights to empower and inform policymakers, industry stakeholders, and consumers toward a more sustainable future.

CHAPTER VI:
SUMMARY, IMPLICATIONS, AND RECOMMENDATIONS

6.1 Summary

The main focus of this dissertation is developing a machine learning-based system designed to classify the environmental impact of vehicles in Canada. The primary objective is to address vehicle emissions, which constitute a significant source of global greenhouse gas emissions and air pollutants. The study seeks to gain a deeper understanding of these emissions through machine learning and contribute to formulating sustainable transportation policies.

The research commences by examining the limitations of conventional methods for classifying vehicles based on their environmental impact. It emphasizes the potential of machine learning to overcome these limitations by offering a more thorough and precise analysis. The literature review explores current research on vehicle emissions, the influence of various vehicle specifications on emissions, and the existing regulatory frameworks.

The methodology outlines the detailed steps in developing the machine learning system. Comprehensive datasets of vehicle specifications and fuel consumption metrics were initially collected and meticulously cleaned. This was followed by an extensive exploratory data analysis (EDA) to comprehend the distributions and relationships within the data. Multiple machine learning models, such as Decision Trees, Random Forests, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and K-nearest neighbours (KNN), were constructed and rigorously evaluated. To ensure their accuracy and dependability, the models underwent k-fold cross-validation.

The results show the findings from the data analysis and model evaluation. Key points include:

Correlation Analysis: There is a solid inverse relationship between fuel efficiency and CO2 emissions. Larger engines, older vehicles, and specific fuel types like diesel are linked to higher emissions.

Model Performance: The GBM model performed the best in classifying vehicles based on their environmental impact, followed by Decision Trees and Random Forests. SVM and KNN could have been more effective.

Feature Importance: Fuel consumption ratings were the most essential features in predicting CO2 emissions, with fuel type, engine size, and the number of cylinders also being significant.

The discussion interprets the results in the context of existing literature and policy implications. Key insights include:

Policy Development: The findings support regulations that promote fuel-efficient, electric, and hybrid vehicles. The machine learning model can help identify high-impact vehicles and inform specific policy measures.

Resource Allocation: The insights can guide investments in research and development for more efficient vehicle technologies and infrastructure for electric vehicles.

Consumer Awareness: Public campaigns can use this information to educate consumers on choosing more environmentally friendly vehicles.

Challenges and Limitations: The study acknowledges issues like data quality, model interpretability, and the need for further validation with other datasets.

My dissertation showcases the transformative potential of machine learning in classifying vehicles according to their environmental impact. By leveraging machine learning techniques, we can effectively contribute to developing sustainable transportation policies to curb greenhouse gas emissions and promote public health. This research underscores the critical role of data-driven approaches in addressing environmental issues and their potential to bring about meaningful change.

6.2 Implications

The research carried out in this thesis has many vital implications in different areas, such as developing policies and practices in the industry, raising public awareness, and guiding future research. Here are the critical implications:

- Policy Development:

Targeted Regulations: The study offers a strong foundation for categorizing vehicles according to their environmental footprint. This allows policymakers to create more precise and impactful regulations. By identifying high-impact vehicles, regulations can be tailored to encourage fuel-efficient, electric, and hybrid vehicles.

Incentives and Subsidies: The research findings strongly support the establishment of monetary incentives and subsidies to promote the use of fuel-efficient and low-emission vehicles. By conducting a thorough analysis of the critical factors that have the most substantial impact on emissions, policymakers will be able to devise targeted financial incentives that effectively encourage consumers and manufacturers to prioritize adopting and developing cleaner, more environmentally friendly technologies in the automotive industry.

Emission Standards: The machine learning models created as part of this research have the potential to enhance the setting of emission standards significantly. These models can continuously analyze real-time data, allowing for the dynamic and precise adjustment of emission standards. This approach ensures that standards can adapt and remain relevant in response to the evolution of vehicle technologies and changing environmental conditions.

- Industry Practices:

Vehicle Design and Manufacturing: The findings from this research can provide valuable guidance for automotive manufacturers in creating vehicles that align with increasingly stringent environmental regulations. Prioritizing aspects such as enhancing fuel efficiency, downsizing engines, and integrating cutting-edge technologies such as electric and hybrid powertrains can be instrumental in minimizing emissions and advancing sustainability efforts within the automotive industry.

Research and Development: The research emphasizes the significance of allocating resources to research and development to create more effective engines, lighter materials, and alternative fuels, which can stimulate creativity in the automobile sector and result in more environmentally friendly transportation options.

- Public Awareness and Consumer Behavior:

Consumer Education: The discoveries may be used to shape public awareness initiatives to educate consumers about the environmental consequences of their vehicle selections. Emphasizing the advantages of fuel-efficient, electric, and hybrid vehicles can sway purchasing choices and encourage more environmentally conscious consumer behaviour.

Market Demand: As public awareness of the positive environmental impact of cleaner vehicles continues to grow, there is a strong likelihood of an increase in market demand for these vehicles. This heightened demand is anticipated to stimulate further innovation and foster healthy competition within the automotive industry, driving the development of sustainable solutions.

- Resource Allocation:

Efficient Use of Resources: By conducting a comprehensive analysis to pinpoint the most significant factors affecting vehicle emissions, we can strategically allocate resources to achieve maximum impact. This approach enables us to prioritize investments in up-and-coming technologies and critical infrastructure, including developing electric vehicle charging stations and enhancing public transportation systems.

Research Priorities: The research offers a comprehensive analysis to support the ranking of research initiatives. These include enhancing fuel efficiency, creating eco-friendly engine technologies, and investigating alternative fuels. These priorities are determined based on their potential to decrease emissions significantly.

Expanding Data Sources: To further advance the findings of this study, future research could consider integrating a wider range of diverse and comprehensive datasets. By incorporating real-time emissions data and detailed vehicle usage patterns, researchers can improve the precision and relevance of the models, leading to a deeper understanding of the subject matter.

Advanced Techniques: Incorporating advanced machine learning methodologies, like deep learning, can enhance the accuracy of classification models and their predictive capabilities. These techniques can effectively capture intricate relationships within extensive datasets, offering more comprehensive insights into vehicle emissions.

Model Validation: Further validation of the models with independent datasets from different regions and contexts can ensure their robustness and generalizability. This can help adapt the findings to various geographic and regulatory environments.

Interdisciplinary Approaches: Integrating insights from environmental science, economics, and behavioural studies can provide a more holistic understanding of vehicle emissions and their impact. Interdisciplinary research can lead to more comprehensive solutions to environmental challenges.

This research emphasizes the significant impact of machine learning in effectively addressing environmental issues associated with vehicle emissions. Through data-driven methodologies, this study demonstrates the potential of machine learning in not only comprehending but also mitigating the environmental footprint of vehicles. By doing so, it advocates for formulating more dynamic and impactful policies, promotes adopting sustainable practices within the industry, and aims to cultivate a more informed consumer base. The cumulative effect of these implications serves to propel environmental sustainability within the transportation sector, contributing to meaningful progress in this critical area.

6.3 Recommendations for Future Research

The research outlined in this dissertation lays a solid groundwork for comprehending and addressing the environmental impact of vehicles using machine learning. However, several areas deserve additional exploration to enrich and build upon the study's findings further. Here are some suggestions for future research:

1. Expanding Geographic and Temporal Scope: Including data from diverse geographic regions and more extended periods would bolster the applicability of the findings. Collaborating with international datasets and conducting longitudinal studies could reveal how vehicle emissions and environmental impacts change over time.

2. Unveiling the Potential of Advanced Machine Learning Techniques: The future of research in this field is ripe with possibilities. Delving into advanced machine learning methods such as deep learning, reinforcement learning, and ensemble methods could unlock new insights. These techniques, with their ability to handle more intricate

relationships and larger datasets, could potentially revolutionize the accuracy and resilience of the classification models.

3. Enhancing Model Interpretability: While advanced models like Gradient Boosting Machines (GBM) offer high accuracy, they may need to be more interpretable than simpler models. Future research should concentrate on developing techniques to improve the interpretability of complex models. This may involve using explainable AI (XAI) methods to provide more precise insights into how models make predictions.

4. Examining Alternative Fuels and Technologies: Research should continue to explore the potential of alternative fuels (such as hydrogen, biofuels, and synthetic fuels) and emerging technologies (such as autonomous vehicles and smart grids) in reducing emissions. Comparative studies evaluating the environmental impact of these alternatives against conventional fuels and technologies would be valuable.

5. Policy Impact Assessment: Future research should assess the real-world impact of policies and regulations that reduce vehicle emissions. This could involve conducting case studies in regions with specific policies and evaluating their effectiveness in achieving environmental objectives.

6. Unveiling the Key to Promoting Sustainable Transportation: Understanding consumer behaviour and market dynamics is not just important, it's vital. Future studies should investigate the factors influencing consumer choices, the effectiveness of incentives, and the barriers to adopting cleaner technologies. Surveys, experiments, and behavioral modeling could provide crucial insights into these aspects.

7. Meeting the Need for Scalable Solutions: The future of research should be about more than just understanding the problem. It should be about solving it. Emphasizing the creation of scalable solutions that can be implemented at a broader level is crucial. This could include developing standardized frameworks and tools that policymakers and industry stakeholders can use to assess and mitigate vehicle emissions.

Addressing these recommendations in future research can extend the groundwork laid by this dissertation, aiding in a more profound understanding of vehicle emissions and formulating more effective strategies for promoting sustainable transportation.

6.4 Conclusion

The findings of this thesis consolidated a comprehensive study that focused on understanding and alleviating the impact of vehicles on the environment in Canada through machine learning techniques. Critical variables influencing vehicle emissions, such as CO_2 emissions, fuel efficiency, engine size, fuel type, vehicle weight, and age, were identified at the beginning of the research. Strong correlations between these variables were established through extensive exploratory data analysis, emphasizing the inverse relationship between fuel efficiency and CO_2 emissions and the higher emissions associated with more extensive and older vehicles. These findings underscore the necessity for targeted policies promoting the adoption of fuel-efficient, smaller, and newer vehicles to reduce emissions.

The study delved into the relationships among fuel efficiency, fuel type, and vehicle emissions, affirming that fuel-efficient vehicles consistently produce lower emissions and that electric and hybrid vehicles significantly outperform gasoline and diesel vehicles regarding environmental impact. This analysis underscored the environmental benefits of transitioning to electric and hybrid technologies and supported policy recommendations promoting this transition through incentives and infrastructure development.

The crux of this research centred on developing and evaluating machine learning models to categorize vehicles based on their environmental impact. Various algorithms, including Decision Trees, Random Forests, Gradient Boosting Machines (GBM), Support Vector Machines (SVM), and K-nearest neighbours (KNN), were evaluated for their effectiveness. Among these, GBM demonstrated the highest accuracy and robustness, establishing it as a valuable tool for policymakers to identify high-impact vehicles and develop targeted regulations. Despite challenges such as data quality and model interpretability, the success of the GBM model illustrates the potential of machine learning in informing data-driven policy decisions and promoting sustainable transportation.

In summary, this thesis underscores the power of machine learning in understanding and mitigating the environmental impact of vehicles. The research offers practical insights that can significantly inform and enhance transportation policies by identifying key variables, exploring their interrelationships, and developing effective classification models. These insights align with the Canadian government's efforts to reduce greenhouse gas emissions and promote sustainable development. It is important to emphasize that this research results from collaborative efforts, and future progress will depend on continued interdisciplinary collaboration and innovation. We can work towards a more sustainable transportation system and a healthier environment.

References

Abdel Wahed Ahmed, M.M. and Abd El Monem, N., 2020. Sustainable and green transportation for better quality of life case study greater Cairo–Egypt. Hbrc Journal, 16(1), pp.17-37.

Alanazi, F., 2023. Electric vehicles: benefits, challenges, and potential solutions for widespread adaptation. Applied Sciences, 13(10), p.6016.

Alazemi, T., Darwish, M. and Radi, M., 2024. Renewable energy sources integration via machine learning modelling: A systematic literature review. Heliyon.

Ali, N., Khan, M.H., Ali, M., Ahmad, S., Khan, A., Nabi, G., Ali, F., Bououdina, M. and Kyzas, G.Z., 2023. Insight into microplastics in the aquatic ecosystem: Properties, sources, threats and mitigation strategies. Science of The Total Environment, p.169489.

Arshad, K., Hussain, N., Ashraf, M.H. and Saleem, M.Z., 2024. Air pollution and climate change as grand challenges to sustainability. Science of The Total Environment, p.172370.

Balan, S. and Prakash, Vel (2016). 'The impact of implementing green supply chain management practices on corporate performance'. *Competitiveness Review* 26 no.

Barman, P., Dutta, L., Bordoloi, S., Kalita, A., Buragohain, P., Bharali, S. and Azzopardi, B., 2023. Renewable energy integration with electric vehicle technology: A review of the existing smart charging approaches. Renewable and Sustainable Energy Reviews, 183, p.113518.

Basu, A., Saha, P.K., Sahu, S.K., & Datta, A. (2021). 'Agricultural residue burning and its contribution to air pollution: Evidence from India'. *Science of The Total Environment*, 758, 143350.

Bauner, D., Laestadius, S., & Iida, N., 2009. Evolving technological systems for diesel engine emission control: balancing GHG and local emissions. Clean Technologies and Environmental Policy, 11(3), 339-365.

Baz, E, J. and Iddik, S. (2020) "Culture and green supply chain management (GSCM)," Emerald Publishing Limited, 31(2),p. 483-504. Available at: https://doi.org/10.1108/meq-09-2019-0197

Bernstein, J.A., Alexis, N., Barnes, C., Bernstein, I.L., Nel, A., Peden, D., ... & Williams, P.B. (2008). 'Health effects of air pollution'. *Journal of Allergy and Clinical Immunology*, 121(3), pp. 585-591.

Bibri, S.E., Krogstie, J. and Kärrholm, M., 2020. Compact city planning and development: Emerging practices and strategies for achieving the goals of sustainability. Developments in the built environment, 4, p.100021.

Blanchard, D., (2010). 'Supply chain management best practices'. *John Wiley & Sons*.

Bocarejo, J.P. & Oviedo, D.R. (2012). 'Transport accessibility and social inequities: a tool for identification of mobility needs and evaluation of transport investments'. *Journal of Transport Geography*, 24, pp. 142-154.

Bond, T.C., Doherty, S.J., Fahey, D.W., Forster, P.M., Berntsen, T., DeAngelo, B.J., ... & Zender, C.S. (2013). 'Bounding the role of black carbon in the climate system: A scientific assessment'. *Journal of Geophysical Research: Atmospheres*, 118(11), pp. 5380-5552.

Bonjour, S., Adair-Rohani, H., Wolf, J., Bruce, N.G., Mehta, S., Prüss-Ustün, A., ... & Smith, K.R. (2013). 'Solid fuel use for household cooking: Country and regional estimates for 1980–2010'. *Environmental Health Perspectives*, 121(7), 784-790.

Borck, R. and Schrauth, P., 2021. Population density and urban air quality. Regional Science and Urban Economics, 86, p.103596.

Bowen, F., Cousins, P., Lamming, R., & Farukt, A., (2009). 'THE ROLE OF SUPPLY MANAGEMENT CAPABILITIES IN GREEN SUPPLY'. *Production and Operations Management*, 10, pp. 174-189. https://doi.org/10.1111/J.1937-5956.2001.TB00077.X.

Brinklow, G., Herreros, J.M., Zeraati Rezaei, S., Doustdar, O., Tsolakis, A., Kolpin, A. and Millington, P., 2023. Non-carbon greenhouse gas emissions for hybrid electric vehicles: three-way catalyst nitrous oxide and ammonia trade-off. International Journal of Environmental Science and Technology, 20(11), pp.12521-12532.

Brook, R.D., Rajagopalan, S., Pope, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., ... & Kaufman, J.D. (2010). 'Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association'. *Circulation*, 121(21), pp. 2331-2378. *Business intelligence Decis. Mak. Manag*. 280.

Cascone, L., Alnowaiser, K., Alarfaj, A.A., Alabdulqader, E.A., Umer, M. and Alankar, B., Iot Based Smart Framework to Predict Air Quality in Congested Traffic Areas Using Sv-Cnn Ensemble and Knn Imputation Model. Available at SSRN 4737721.

Cervero, R., Ferrell, C., & Murphy, S. (2004). 'Transit-oriented development and joint development in the United States: A literature review'. *Transportation Research Board's Transit Cooperative Research Program, Research Results Digest* 52.

Chatzoudes, D. and Chatzoglou, P. (2022). 'Antecedents and effects of green supply chain a qualitative study'. *Int. J. Phys. Distrib. Logist. Manag*. 39, 690–711.

Chen, C., Varley, D., & Chen, J. (2011). ‘What affects transit ridership? A dynamic analysis involving multiple factors, lags and asymmetric behaviour’. *Urban Studies*, 48(9), pp. 1893-1908.

Chen, J., Wang, B., Wang, Z., & Zhang, Y. (2019). ‘Deep learning for real-time air quality forecasting’. *Science of The Total Environment*, 712, 135694.

Chen, L., Chen, Z., Zhang, Y., Liu, Y., Osman, A.I., Farghali, M., Hua, J., Al-Fatesh, A., Ihara, I., Rooney, D.W. and Yap, P.S., 2023. Artificial intelligence-based solutions for climate change: a review. Environmental Chemistry Letters, 21(5), pp.2525-2557.

Chowdhury, P., Lakku, N.K.G., Lincoln, S., Seelam, J.K. and Behera, M.R., 2023. Climate change and coastal morphodynamics: Interactions on regional scales. Science of the Total Environment, p.166432.

Das, D. (2017) "Development and validation of a scale for measuring Sustainable Supply Chain Management practices and performance," Elsevier BV, 164,p. 1344-1362. Available at: https://doi.org/10.1016/j.jclepro.2017.07.006

Dasari, P, Y. and Koul, S. (2015) "Green supply chain drivers in the Indian automobile industry,", 6(3),p. 179-179. Available at: https://doi.org/10.1504/ijleg.2015.073889

D’Agosto, M.D.A., 2019. Noise pollution, vibration, visual intrusion, and emission of solid and liquid waste. Transportation, Energy Use and Environmental Impacts, pp.259-280.

Derwent, R.G., Jenkin, M.E., Saunders, S.M., and Pilling, M.J., (2010). ‘Secondary organic aerosol formation from a large number of reactive man-made organic compounds’. *Science of The Total Environment*, 408(16), pp.3374-3381.

Dey, S. and Mehta, N.S., 2020. Automobile pollution control using catalysis. Resources, Environment and Sustainability, 2, p.100006.

Dillon, J., 2016. Huge Challenges for Canada after the Paris Climate Conference. Kairos Discussion Papers, pp.1-24.

Dockery, D.W., Pope, C.A., Xu, X., Spengler, J.D., Ware, J.H., Fay, M.E., Ferris, B.G., and Speizer, F.E., (1993). 'An association between air pollution and mortality in six U.S. cities'. *The New England Journal of Medicine*, 329(24), pp.1753-1759.

Dos Santos Morini, R.F., Ribeiro, R.C., Trentini, I.P., Sandoval, E.D.P., da Silva, A.R. and de Oliveira, M.L., 2023. Do media reports reflect the real threats to wildlife?. Biological Conservation, 277, p.109853.

Dwivedi, Y.K., Hughes, L., Kar, A.K., Baabdullah, A.M., Grover, P., Abbas, R., Andreini, D., Abumoghli, I., Barlette, Y., Bunker, D. and Kruse, L.C., 2022. Climate change and COP26: Are digital technologies and information management part of the problem or the solution? An editorial reflection and call to action. International Journal of Information Management, 63, p.102456.

Ehlert, P.A., Rothkrantz, L.J., (2001). 'Microscopic traffic simulation with reactive driving agents, in: Intelligent Transportation Systems, 2001'. *Proceedings. 2001 IEEE.* IEEE, pp. 860–865.

Elassy, M., Al-Hattab, M., Takruri, M. and Badawi, S., 2024. Intelligent transportation systems for sustainable smart cities. Transportation Engineering, p.100252.

Elofson, G. and Robinson, W.N., (2007). 'Collective customer collaboration impacts on supply-chain performance'. *International Journal of Production Research*, 45(11), pp.2567-2594.

Environment and Climate Change Canada. (2020). 'Clean Fuel Standard: Proposed Regulatory Approach'. *Available at: https://www.canada.ca*

Ercan, T., Onat, N.C., Keya, N., Tatari, O., Eluru, N. and Kucukvar, M., 2022. Autonomous electric vehicles can reduce carbon emissions and air pollution in cities. Transportation Research Part D: Transport and Environment, 112, p.103472.

Essamlali, I., Nhaila, H. and El Khaili, M., 2024. Advances in machine learning and IoT for water quality monitoring: A comprehensive review. Heliyon.

Ewing, R. & Cervero, R. (2010). 'Travel and the built environment: a meta-analysis'. *Journal of the American Planning Association*, 76(3), pp. 265-294.

Falai, A., & Misul, D., 2023. Data-Driven Model for Real-Time Estimation of NOx in a Heavy-Duty Diesel Engine. Energies.

FAO. (2021). 'Food and Agriculture Organization of the United Nations'. *Available at: https://www.fao.org*

Filonchyk, M. and Peterson, M.P., 2023. An integrated analysis of air pollution from US coal-fired power plants. Geoscience Frontiers, 14(2), p.101498.

Filonchyk, M., Peterson, M.P., Zhang, L., Hurynovich, V. and He, Y., 2024. Greenhouse gases emissions and global climate change: Examining the influence of CO_2, CH_4, and N_2O. Science of The Total Environment, p.173359.

Fontaras, G., Zacharof, N.G. and Ciuffo, B., 2017. Fuel consumption and CO_2 emissions from passenger cars in Europe–Laboratory versus real-world emissions. Progress in energy and combustion Science, 60, pp.97-131.

Forbes.com. AutoML 2.0: Is The Data Scientist Obsolete? URL: https://www.forbes.com/sites/cognitiveworld/2020/04/07/automl-20-is-the-data-scientist-obsolete/?sh=63b69def53c9. Accessed Oct 8, 2021.

Gauderman, W.J., Avol, E., Lurmann, F., Kuenzli, N., Gilliland, F., Peters, J., & McConnell, R. (2004). 'Childhood asthma and exposure to traffic and nitrogen dioxide'. *Epidemiology*, 16(6), pp. 737-743.

German Federal Ministry for Economic Affairs and Energy. (2020). 'Renewable Energy Sources Act (EEG)'. *Available at: https://www.bmwi.de*

Ghaffarian, S., Taghikhah, F.R. and Maier, H.R., 2023. Explainable artificial intelligence in disaster risk management: Achievements and prospective futures. International Journal of Disaster Risk Reduction, 98, p.104123.

Giannelos, S., Bellizio, F., Strbac, G. and Zhang, T., 2024. Machine learning approaches for predictions of CO2 emissions in the building sector. Electric Power Systems Research, 235, p.110735.

Gómez, E., Tuholske, C., & Andam, K.S. (2019). 'Monitoring water quality using remote sensing and machine learning: A review'. *Remote Sensing*, 11(12), 1411.

Goodman, J., Laube, M., & Schwenk, J. (2007). 'Curitiba's Bus System is Model for Rapid Transit'. *Race, Poverty & the Environment*, 14(1), pp. 75-76.

Government of Canada. (2021). 'Regulations Amending the Passenger Automobile and Light Truck Greenhouse Gas Emission Regulations'. *Available at: https://www.canada.ca*

Govindan, K. et al. (2020) "Drivers and value-relevance of CSR performance in the logistics sector: A cross-country firm-level investigation," Elsevier BV, 231,p. 107835-107835. Available at: https://doi.org/10.1016/j.ijpe.2020.107835

Green Climate Fund. (2021). ‘About the Green Climate Fund’. *Available at: https://www.greenclimate.fund*

Guarnieri, M. & Balmes, J.R. (2014). ‘Outdoor air pollution and asthma’. *The Lancet*, 383(9928), pp. 1581-1592.

Guo, H., Li, Q., Yu, Z., & Wu, W. (2018). ‘Big Earth Data from space: intelligence and interaction’. *Big Earth Data*, 2(1), 1-10.

Gurjar, B.R., Butler, T.M., Lawrence, M.G., & Lelieveld, J. (2010). ‘Evaluation of emissions and air quality in megacities’. *Atmospheric Environment*, 44(37), pp. 4606-4624.

Gupta, A. and Singh, K, R. (2020) "Managing operations by a logistics company for sustainable service quality: Indian perspective," Emerald Publishing Limited, 31(5),p. 1309-1327. Available at: https://doi.org/10.1108/meq-11-2019-0246

Guttikunda, S.K. & Goel, R. (2013). ‘Health impacts of particulate pollution in a megacity—Delhi, India’. *Environmental Development*, 6, pp. 8-20.

Hagos, D.A. and Ahlgren, E.O., 2018. Well-to-wheel assessment of natural gas vehicles and their fuel supply infrastructures–Perspectives on gas in transport in Denmark. Transportation Research Part D: Transport and Environment, 65, pp.14-35.

Hassan, M., Khosravi, K., Farooque, A.A., Esau, T.J., Boluwade, A. and Sadiq, R., 2024. Prediction of Carbon Dioxide Emissions from Atlantic Canadian Potato Fields Using Advanced Hybridized Machine Learning Algorithms–Nexus of Field Data and Modelling. Smart Agricultural Technology, p.100559.

Hebaz, A. and Oulfarsi, S., 2021. The drivers and barriers of green supply chain management implementation: A review. Acta logistica.

Hervani, A.A., Helms, M.M., Sarkis, J., (2005). ‘Performance measurement for green supply chain management’. *Benchmarking Int. J.* 12, 330–353.

Hettigamage, J.M., Mapatunage, M.D.C.P., Paranagama, C.B., Chathuranga, B., De Silva, D., and Sucharitharathna, C., 2013. CO2 emissions reduction from proper vehicle maintenance and management system. 2013 8th International Conference on Computer Science & Education, pp. 676-679.

Hoek, G., Krishnan, R.M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J.D. (2013). ‘Long-term air pollution exposure and cardio-respiratory mortality: A review’. *Environmental Health*, 12(1), 43.

Hofstetter, J., Bauer, H., Li, W., & Wachtmeister, G., 2019. Energy and Emission Management of Hybrid Electric Vehicles using Reinforcement Learning. IFAC-PapersOnLine.

Holmgren, J., 2020. The effect of public transport quality on car ownership–A source of wider benefits?. Research in Transportation Economics, 83, p.100957.

Hsu, C.H., Yoon, J.H., & Wang, H. (2019). ‘Intelligent traffic signal control using deep reinforcement learning’. *Transportation Research Part C: Emerging Technologies*, 105, pp. 359-373.

https://tc.canada.ca/en/road-transportation/publications/canada-s-action-plan-clean-road-transportation

https://www.aeroqual.com/blog/meet-the-nitrogen-oxide-family

https://www.canada.ca/en/services/environment/weather/climatechange/pan-canadian-framework/climate-change-plan.html

IEA. (2020). ‘International Energy Agency – Energy Efficiency’. *Available at: https://www.iea.org*

Imam, M., Adam, S., Dev, S. and Nesa, N., 2024. Air quality monitoring using statistical learning models for sustainable environment. Intelligent Systems with Applications, 22, p.200333.

Intergovernmental Panel on Climate Change (IPCC). (2013). ‘Climate Change 2013: The Physical Science Basis’. *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Intergovernmental Panel on Climate Change (IPCC). (2013). ‘Climate Change 2013: The Physical Science Basis’. *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.*

International Council on Clean Transportation. (2020). ‘CO2 emissions from new passenger cars in the European Union: Car manufacturers’ performance in 2019’. *Available at: https://theicct.org*

International Energy Agency (IEA). (2021). ‘Global Energy Review: CO2 Emissions in 2020’. *Available at: https://www.iea.org*

Jaimez-González, C.R., Luna-Ramírez, W.A., (2013). ‘An agent-based architecture for supply chain management, in: Autonomous Decentralized Systems (ISADS)’, *2013 IEEE Eleventh International Symposium on. IEEE*, pp. 1–6.

Jain, V., Sharma, S., (2014). ‘Drivers Affecting the Green Supply Chain Management Adaptation: A Review’. *IUP J. Oper. Manag*. 13.

Jauhar, S.K., Sethi, S., Kamble, S.S., Mathew, S. and Belhadi, A., 2024. Artificial intelligence and machine learning-based decision support system for forecasting electric vehicles' power requirement. Technological Forecasting and Social Change, 204, p.123396.

Jennings, N.R., Sycara, K., Wooldridge, M., (1998). ‘A roadmap of agent research and development’. *Auton. Agents Multi-Agent Syst.* 1, 7–38.

Jin, L., Searle, S. and Lutsey, N., 2014. Evaluation of state-level US electric vehicle incentives. The International Council on Clean Transportation.

Jin, X., Abdulrab, H., Itmi, M., (2008). ‘A multi-agent based model for urban demand-responsive passenger transport services, in: Neural Networks, 2008’. *IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on. IEEE*, pp. 3668–3675.

Johnston, F.H., Henderson, S.B., Chen, Y., Randerson, J.T., Marlier, M., DeFries, R.S., ... & Brauer, M. (2012). ‘Estimated global mortality attributable to smoke from landscape fires’. *Environmental Health Perspectives*, 120(5), 695-701.

Jones, M.W., Peters, G.P., Gasser, T., Andrew, R.M., Schwingshackl, C., Gütschow, J., Houghton, R.A., Friedlingstein, P., Pongratz, J. and Le Quéré, C., 2023. National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. Scientific Data, 10(1), p.155.

Kafa, N., Hani, Y. and El Mhamedi, A., (2013). ‘Sustainability performance measurement for green supply chain management’. *IFAC Proceedings Volumes*, 46(24), pp.71-78.

Kampa, M. & Castanas, E. (2008). ‘Human health effects of air pollution’. *Environmental Pollution*, 151(2), pp. 362-367.

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2018). ‘Theory-guided data science: A new paradigm for scientific discovery from data'. *IEEE Transactions on Knowledge and Data Engineering*, 29(10), 2318-2331.

Karuppiah, K., Sankaranarayanan, B., Ali, S., Chowdhury, P. and Paul, S., 2020. An integrated approach to modeling the barriers in implementing green manufacturing practices in SMEs. Journal of Cleaner Production, 265, p.121737.

Katreddi, S., Kasani, S., & Thiruvengadam, A., 2022. A Review of Applications of Artificial Intelligence in Heavy Duty Trucks. Energies.

Kenneth, W., Pamela, J.Z., Jeramy, M. and Vikram, S.Bhadauria (2012). 'Green supply chain management practices: impact on performance'. *Supply chain management an international journal 17 no*

Khalili, S., Rantanen, E., Bogdanov, D. and Breyer, C., 2019. Global transportation demand development with impacts on the energy demand and greenhouse gas emissions in a climate-constrained world. Energies, 12(20), p.3870.

Khan, A.A., Kumar, P., Gulia, S. and Khare, M., 2024. A critical review of managing air pollution through airshed approach. Sustainable Horizons, 9, p.100090.

Khare, M. and Adhikari, A., (2020). 'Air pollution and its health impacts in urban cities'. *International Journal of Environmental Health Research*, 30(6), pp.617-631.

Khreis, H., Kelly, C., Tate, J., Parslow, R., Lucas, K., & Nieuwenhuijsen, M.J. (2017). 'Exposure to traffic-related air pollution and risk of development of childhood asthma: A systematic review and meta-analysis'. *Environment International*, 100, pp. 1-31.

Kim, M., & Chai, S., (2017). 'Implementing Environmental Practices for Accomplishing Sustainable Green Supply Chain Management'. *Sustainability*, 9, pp. 1-17. https://doi.org/10.3390/SU9071192.

Kimbrough, S.O., Wu, D.-J., Zhong, F., (2002. ‘Computers play the beer game: can artificial agents manage supply chains?’ *Decis. Support Syst*. 33, 323–333.

Kuik, S, S., Nagalingam, S. and Amer, Y. (2011) "Sustainable supply chain for collaborative manufacturing," Emerald Publishing Limited, 22(8),p. 984-1001. Available at: https://doi.org/10.1108/17410381111177449.

Kumar, P., Khare, M., Harrison, R.M., Bloss, W.J., Lewis, A.C., Coe, H., & Morawska, L. (2015). ‘New directions: Air pollution challenges for developing megacities like Delhi’. *Atmospheric Environment*, 122, pp. 657-661.

Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., ... & Sommer, H. (2000). ‘Public-health impact of outdoor and traffic-related air pollution: A European assessment’. *The Lancet*, 356(9232), pp. 795-801.

Kwan, S.C., binti Zakaria, S., Ibrahim, M.F., Mahiyuddin, W.R.W., Sofwan, N.M., Wahab, M.I.A., Ahmad, R.D.R., Abbas, A.R., Woon, W.K. and Sahani, M., 2023. Health impacts from TRAPs and carbon emissions in the projected electric vehicle growth and energy generation mix scenarios in Malaysia. Environmental Research, 216, p.114524.

Le Cornec, C. M. A., Molden, N., van Reeuwijk, M., & Stettler, M., 2020. Modelling of instantaneous emissions from diesel vehicles with a special focus on NOx: Insights from machine learning techniques. The Science of the Total Environment.

Leape, J. (2006). ‘The London congestion charge’. *Journal of Economic Perspectives*, 20(4), pp. 157-176.

Lee, J., Veloso, F., Hounshell, D., & Rubin, E., 2010. Forcing technological change: A case of automobile emissions control technology development in the US. Technovation, 30(4), 249-264.

Lee, J.-H., Kim, C.-O., (2008). ‘Multi-agent systems applications in manufacturing systems and supply chain management: a review paper’. *Int. J. Prod. Res*. 46, 233–265.

Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D., & Pozzer, A. (2015). ‘The contribution of outdoor air pollution sources to premature mortality on a global scale’. *Nature*, 525(7569), pp. 367-371.

Li, C., Liu, M., Hu, Y., Zhou, R., Huang, N., Wu, W. and Liu, C., 2020. Spatial distribution characteristics of gaseous pollutants and particulate matter inside a city in the heating season of Northeast China. Sustainable Cities and Society, 61, p.102302.

Li, F., Yigitcanlar, T., Nepal, M., Nguyen, K. and Dur, F., 2023. Machine learning and remote sensing integration for leveraging urban sustainability: A review and framework. Sustainable Cities and Society, 96, p.104653.

Li, Z., Wang, C., Emrich, C.T., & Guo, D. (2020). ‘A novel approach to leveraging social media for rapid flood mapping: A case study of the 2015 South Carolina floods’. *Cartography and Geographic Information Science*, 47(1), 24-37.

Liang, W.-Y., Huang, C.-C., (2006). ‘Agent-based demand forecast in multi-echelon supply chain’. *Decis. Support Syst*. 42, 390–407.

Litman, T. (2017). ‘Evaluating public transit benefits and costs. Victoria Transport Policy Institute’. *Available at: http://www.vtpi.org/tranben.pdf*

Lou, P., Zhou, Z., Chen, Y.-P., Ai, W., (2004). ‘Study on multi-agent-based agile supply chain management’. *Int. J. Adv. Manuf. Technol.* 23, 197–203. doi:10.1007/s00170-003-1626-x

Lutsey, N. and Sperling, D., 2007. Canada's voluntary agreement on vehicle greenhouse gas emissions: When the details matter. Transportation Research Part D: Transport and Environment, 12(7), pp.474-487.

Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F.Y., 2014. Traffic flow prediction with big data: A deep learning approach. Ieee transactions on intelligent transportation systems, 16(2), pp.865-873.

Lyonnais, P., Montreuil, B., (2001). 'Un environnement distribué de simulationa base d'agents,' in: *3e Conférence Francophone de MOdélisation et SIMulation (MOSIM)-Conception, Analyse et Gestion Des Systemes Industriels, Troyes (France).*

Ma, X., 2013. Towards intelligent fleet management: Local optimal speeds for fuel and emissions. 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), pp. 2201-2206.

Ma, X., Hu, Q., Huang, J., & Zhang, L. (2019). 'Deep learning in remote sensing applications: A meta-analysis and review'. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166-177.

MacLean, H. and Lave, L., 2003. Life cycle assessment of automobile/fuel options. Environmental science & technology, 37(23), pp.5445-5452.

Mandić, K., Delibašić, B., (2012). 'Application of multi-agent systems in supply chain management'.

Marchesini, M.M.P. and Lizarelli, F.L. (2018). 'Barreiras à implementação da Logística Reversa e do Green Supply Chain Management (GSCM)'. *Revista de Ciência & Tecnologia*. 20(40), p.69.

Mateo-Sagasta, J., Marjani, S., & Turral, H. (2017). 'Water pollution from agriculture: a global review'. *Food and Agriculture Organization of the United Nations*.

Mathew, N., Somanathan, A., Tirpude, A. and Arfin, T., 2024. The impact of short-lived climate pollutants on the human health. Environmental Pollution and Management, 1, pp.1-14.

Melton, N., Axsen, J. and Moawad, B., 2020. Which plug-in electric vehicle policies are best? A multi-criteria evaluation framework applied to Canada. Energy Research & Social Science, 64, p.101411.

Mentzer, J.T., DeWitt, W., Keebler, J.S., Soonhoong, M., Nix, N.W., Smith, C.D., Zacharia, Z.G., (2001). 'DEFINING SUPPLY CHAIN MANAGEMENT'. *J. Bus. Logist*. 22, 1 – 25.

Miele, A., Axsen, J., Wolinetz, M., Maine, E. and Long, Z., 2020. The role of charging and refuelling infrastructure in supporting zero-emission vehicle sales. Transportation Research Part D: Transport and Environment, 81, p.102275.

Mikulčić, H., Baleta, J. and Klemeš, J.J., 2022. Cleaner technologies for sustainable development. Cleaner Engineering and Technology, 7, p.100445.

Miner, P., Smith, B.M., Jani, A., McNeill, G. and Gathorne-Hardy, A., 2024. Car harm: A global review of automobility's harm to people and the environment. Journal of Transport Geography, 115, p.103817.

Mishra, N., Kumar, V., Chan, F.T., (2012). 'A multi-agent architecture for reverse logistics in a green supply chain'. *Int. J. Prod. Res.* 50, 2396–2406.

Molla, A.H., Moghtaderi, S.H., Harun, Z., Jedi, A. and Manoj Kumar, N., 2023. Insights into end-of-life vehicle recycling and its quality assessment systems in Malaysia reveals the need for a new stakeholder-centric approach for vehicle waste management. Production & Manufacturing Research, 11(1), p.2236676.

Moyaux, T., Chaib-Draa, B., D'Amours, S., (2006). 'Supply chain management and multiagent systems: An overview, in: Multiagent Based Supply Chain Management'. *Springer*, pp. 1–27.

Müller, A., Österlund, H., Marsalek, J. and Viklander, M., 2020. The pollution conveyed by urban runoff: A review of sources. Science of the Total Environment, 709, p.136125.

Mustayen, A.G.M.B., Rasul, M.G., Wang, X., Negnevitsky, M. and Hamilton, J.M., 2022. Remote areas and islands power generation: A review on diesel engine performance and emission improvement techniques. Energy Conversion and Management, 260, p.115614.

Nakhjiri, A. and Kakroodi, A.A., 2024. Air pollution in industrial clusters: A comprehensive analysis and prediction using multi-source data. Ecological Informatics, 80, p.102504.

Natural Resources Canada. (2019). 'EcoENERGY Efficiency Program'. *Available at: https://www.nrcan.gc.ca*

Natural Resources Canada. (2021). 'Zero-Emission Vehicles'. *Available at: https://www.nrcan.gc.ca*

New Zealand Government. (2019). 'Reducing New Zealand's Agricultural Emissions'. *Available at: https://www.mpi.govt.nz*

Ng, E. C., Huang, Y., Hong, G., Zhou, J. L., & Surawski, N., 2021. Reducing vehicle fuel consumption and exhaust emissions from the application of a green-safety device under real driving. The Science of the Total Environment, 793, p. 148602.

Nieuwenhuijsen, M.J., 2020. Urban and transport planning pathways to carbon neutral, liveable and healthy cities; A review of the current evidence. Environment international, 140, p.105661.

NYC Mayor's Office of Sustainability. (2019). 'New York City's Climate Mobilization Act'. *Available at: https://www1.nyc.gov*

Pachauri, R.K. & Meyer, L.A. (2014). 'Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change'. *IPCC.*

Pinkse, J., Bohnsack, R., & Kolk, A., 2014. The Role of Public and Private Protection in Disruptive Innovation: The Automotive Industry and the Emergence of Low-Emission Vehicles. Journal of Product Innovation Management, 31(1), 43-60.

Pope, C.A., Burnett, R.T., Thun, M.J., Calle, E.E., Krewski, D., Ito, K., & Thurston, G.D. (2002). 'Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution'. *JAMA, 287(9),* pp. 1132-1141.

Pope, C.A., Ezzati, M., & Dockery, D.W. (2009). 'Fine-particulate air pollution and life expectancy in the United States'. *New England Journal of Medicine*, 360(4), pp. 376-386.

Pramesti, R.I., Baihaqi, I. and Bramanti, G.W. (2021). 'Membangun Green Supply Chain Management (GSCM) Scorecard'. *Jurnal Teknik ITS*. 9(2).

Pucher, J. & Buehler, R. (2010). 'Walking and cycling for healthy cities'. *Built Environment*, 36(4), pp. 391-414.

Quiros, D.C., Smith, J., Thiruvengadam, A., Huai, T. and Hu, S., 2017. Greenhouse gas emissions from heavy-duty natural gas, hybrid, and conventional diesel on-road trucks during freight transport. Atmospheric Environment, 168, pp.36-45.

Rada, E.C., Ragazzi, M., & Fedrizzi, P. (2018). 'Web-GIS oriented system viability for municipal solid waste selective collection optimization in developed and transient economies'. *Waste Management*, 67, pp. 55-61.

Raihan, A., 2023. Artificial intelligence and machine learning applications in forest management and biodiversity conservation. Natural Resources Conservation and Research, 6(2), p.3825.

Rakesh, K. and Ravi, Kant (2015). 'Green supply chain management (GSCM): a structured literature review and research implications'. *Benchmarking An international journal 22 no.*

Raub, J.A., Mathieu-Nolf, M., Hampson, N.B., & Thom, S.R. (2000). 'Carbon monoxide poisoning—a public health perspective'. *Toxicology*, 145(1), pp. 1-14.

Razi, F. and Dincer, I., 2022. A review of the current state, challenges, opportunities and future directions for implementation of sustainable electric vehicle infrastructure in Canada. Journal of Energy Storage, 56, p.106048.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). 'Deep learning and process understanding for data-driven Earth system science'. *Nature*, 566(7743), 195-204.

Remy, G.F., Mehar, S., Sophy, T., Senouci, S., Jan, F. and Gourhant, Y., 2012. Green fleet management architecture: Application to economic itinerary planning. IEEE Globecom Workshops, pp.369-373.

Ristinen, R.A., Kraushaar, J.J. and Brack, J.T., 2022. Energy and the Environment. John Wiley & Sons.

Roberto, C., Mengying, F. and Frank, Wiengarten (2014). ‘Integrated green supply chain management and operational performance. *Supply Chain Management An International Journal 19 no.*

Rossi, S. et al. (2013) "The logistics service providers in eco-efficiency innovation: an empirical study," Emerald Publishing Limited, 18(6),p. 583-603. Available at: https://doi.org/10.1108/scm-02-2012-0053

Sánchez-Flores, B, R. et al. (2020) "Sustainable Supply Chain Management—A Literature Review on Emerging Economies," Multidisciplinary Digital Publishing Institute, 12(17),p. 6972-6972. Available at: https://doi.org/10.3390/su12176972.

Schneider, D.J. & Kyle, P.R. (2014). ‘The earth and its atmosphere: Volcanic eruptions and their impact on the atmosphere and climate’. *Global Volcanism Program. Available at: https://volcano.si.edu*

Seinfeld, J.H. and Pandis, S.N., (2016). ‘*Atmospheric Chemistry and Physics: From Air Pollution to Climate Change’. 3rd ed. Hoboken, NJ: Wiley*.

Shaheen, S.A. and Lipman, T.E., 2007. Reducing greenhouse emissions and fuel consumption: Sustainable approaches for surface transportation. IATSS research, 31(1), pp.6-20.

Shen, W., Han, W., & Yang, Z. (2018). ‘Environmental benefits of battery electric vehicles in China: Correlation between electric energy structure and carbon emissions’. *Sustainability*, 10(4), 1442.

Shoup, D. (2005). ‘The High Cost of Free Parking’. *Chicago: Planners Press, American Planning Association*.

Srivastava, S.K., (2007). ‘Green supply-chain management: A state-of-the-art literature review’. *Int. J. Manag. Rev.* 9, 53 – 80.

Stock, J.R., Boyer, S.L., (2009). ‘Developing a consensus definition of supply chain management: a qualitative study’. *Int. J. Phys. Distrib. Logist. Manag*. 39, 690–711.

Strohbach, M., Arnold, E., & Haase, D. (2012). ‘The carbon footprint of urban green space—A life cycle approach’. *Landscape and Urban Planning*, 104, 220-229. https://doi.org/10.1016/J.LANDURBPLAN.2011.10.013

Tahir, M., Hu, S., Khan, T. and Zhu, H., 2024. Sustainable hybrid station design framework for electric vehicle charging and hydrogen vehicle refueling based on multiple attributes. Energy Conversion and Management, 300, p.117922.

Teixeira, C., Assumpção, A., Corrêa, A., Savi, A., & Prates, G., (2018). ‘The contribution of green logistics and sustainable purchasing for green supply chain management’. *Independent Journal of Management & Production*. https://doi.org/10.14807/IJMP.V9I3.789

Tham, Y.C., Li, X., Wong, T.Y., & Cheng, C.Y. (2019). ‘Artificial intelligence for health care: Policy implications for Singapore’. *The Lancet Regional Health – Western Pacific*, 1, 100005.

Tien, P.W., Wei, S., Darkwa, J., Wood, C. and Calautit, J.K., 2022. Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality–a review. Energy and AI, 10, p.100198.

Tseng, M., Islam, M., Karia, N., Fauzi, F., & Afrin, S., (2019). ‘A literature review on green supply chain management: Trends and future challenges’. *Resources, Conservation and Recycling.* https://doi.org/10.1016/J.RESCONREC.2018.10.009

Unal, A., Rouphail, N., & Frey, H., 2003. Effect of Arterial Signalization and Level of Service on Measured Vehicle Emissions. Transportation Research Record, 1842, pp. 47-56.

UNEP (2019). ‘Global Environment Outlook – GEO-6: Healthy Planet, Healthy People’. *United Nations Environment Programme.*

UNFCCC. (2015). ‘Paris Agreement. United Nations Framework Convention on Climate Change’. *Available at: https://unfccc.int*

US DOE. (2021). ‘Weatherization Assistance Program’. Available at: https://www.energy.gov

US Environmental Protection Agency (EPA). (2021). ‘Fast Facts on Transportation Greenhouse Gas Emissions’. Available at: https://www.epa.gov

US EPA. (2021). ‘Clean Air Act Overview’. Available at: https://www.epa.gov

Vachon, S., Klassen, R.D., (2006). ‘Extending green practices across the supply chain: the impact of upstream and downstream integration’. *Int. J. Oper. Prod. Manag*. 26, 795–821.

Wang, F., Harindintwali, J.D., Yuan, Z., Wang, M., Wang, F., Li, S., Yin, Z., Huang, L., Fu, Y., Li, L. and Chang, S.X., 2021. Technologies and perspectives for achieving carbon neutrality. The Innovation, 2(4).

Wang, F., Harindintwali, J.D., Yuan, Z., Wang, M., Wang, F., Li, S., Yin, Z., Huang, L., Fu, Y., Li, L. and Chang, S.X., 2021. Technologies and perspectives for achieving carbon neutrality. The Innovation, 2(4).

Wang, S., Han, C., Yu, Y., Huang, M., Sun, W., and Kaku, I., 2022. Reducing carbon emissions for the vehicle routing problem by utilizing multiple depots. Sustainability, 14(3), p.1264.

Wee, H.-M., Lee, M.-C., Yu, J.C., Edward Wang, C., (2011). ‘Optimal replenishment policy for a deteriorating green product: Life cycle costing analysis’. *Int. J. Prod. Econ.* 133, 603–611.

Weger, L., Abánades, A. and Butler, T., 2017. Methane cracking as a bridge technology to the hydrogen economy. International Journal of Hydrogen Energy, 42(1), pp.720-731.

Williams, I.D. and Blyth, M., 2023. Autogeddon or autoheaven: Environmental and social effects of the automotive industry from launch to present. Science of the Total Environment, 858, p.159987.

World Health Organization (2016). ‘Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease’. *Available at: https://www.who.int*

Wu, Y., Zhang, S., Hao, J., Liu, H., Wu, X., Hu, J., Walsh, M., Wallington, T., Zhang, K., & Stevanovic, S., 2017. On-road vehicle emissions and their control in China: A review and outlook. The Science of the Total Environment, 574, pp. 332-349.

Wu, Y.C. and Kontou, E., 2022. Designing electric vehicle incentives to meet emission reduction targets. Transportation Research Part D: Transport and Environment, 107, p.103320.

Xu, S., Sun, C. and Liu, N., 2024. Road congestion and air pollution-Analysis of spatial and temporal congestion effects. Science of The Total Environment, p.173896.

Yung, S.K., Yang, C.C., Lau, A.S.M., Yen, J., (2000). ‘APPLYING MULTI-AGENT TECHNOLOGY TO SUPPLY CHAIN MANAEMENT’. *J. Electron. Commer. Res.* 1, 119.

Zakeri, F. and Mariethoz, G., 2024. Synthesizing long-term satellite imagery consistent with climate data: Application to daily snow cover. Remote Sensing of Environment, 300, p.113877.

Zhang, Q., Fan, J., Yang, W., Chen, B., Zhang, L., Liu, J., Wang, J., & Zhou, C., 2017. The effects of deterioration and technological levels on pollutant emission factors for gasoline light-duty trucks. Journal of the Air & Waste Management Association, 67(7), pp. 814-823.

Zhang, S., Zhao, P., He, L., Yang, Y., Liu, B., He, W., Cheng, Y., Liu, Y., Liu, S., Hu, Q., Huang, C., & Wu, Y., 2020. On-board monitoring (OBM) for heavy-duty vehicle emissions in China: Regulations, early-stage evaluation and policy recommendations. The Science of the Total Environment, 731, 139045

Zhao, X., Hu, H., Yuan, H. and Chu, X., 2023. How does adoption of electric vehicles reduce carbon emissions? Evidence from China. Heliyon, 9(9).

Zhou, C., Yu, J., & Meng, Q. (2020). 'Optimization of industrial emissions control using machine learning algorithms'. *Journal of Cleaner Production*, 257, 120561.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). 'Deep Learning in Remote Sensing: A Review'. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8-36.