

AI Assisted Innovation for Predicting Traffic Accident on Addis -Adama Expressway



Name Chernet Debale Galmesa

**A Thesis Submitted to the Department of Technology and Innovation Management
School of Humanity and Social Science**

**Presented in Partial Fulfillment of the Requirement for the Degree of Master's in
Technology and Innovation Management**

**Office of Graduate Studies
Adama Science and Technology University**

**June, 2024
Adama, Ethiopia**

**AI Assisted Innovation for Predicting Traffic Accident on Addis -Adama
Expressway**

Name of Candidate Chernet Debele

Advisor(s) Dr. Ashenafi Teshome

A Thesis Submitted to the Department of Technology and innovation Management,
School of Humanity and Social Science

Presented in Partial Fulfillment of the Requirement for the Degree of Master's in
Technology and Innovation Management

**Office of Graduate Studies
Adama Science and Technology University**

**June, 2024
Adama, Ethiopia**

Declaration

I hereby declare that this Master Thesis entitled AI Assisted Innovation for Predicting Traffic Accident on Addis -Adama Expressway is my original work. That is, it has not been submitted for the award of any academic degree, diploma or certificate in any other university. All sources of materials that are used for this thesis have been duly acknowledged through citation

Name of student

Signature

Date

Recommendation of Advisors/ Supervisors

I, the major advisor of this research proposal, hereby certify that I have closely advised/supervised the student while developing this thesis and read the thesis I entitled Artificial Intelligence Assisted Innovation for Predicting Traffic Accident on Addis Adama Expressway prepared under my guidance by Chernet Debele Therefore, I recommend the submission of the Thesis to the department for further review and evaluation.

_____	_____	_____
Major Advisor/Supervisor	Signature	Date
_____	_____	_____
Co-advisor/Co-supervisor	Signature	Date

Approval Page of M.Sc. Thesis

I/we, the advisors of the thesis Artificial Intelligence Assisted Innovation for Predicting Traffic Accident on Addis Adama Expressway and developed by Chernet Debele here by certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

_____	_____	_____
Major Advisor	Signature	Date

_____	_____	_____
Co-advisor Signature	Signature	Date

We, the undersigned, members of the Board of Examiners of the thesis by Chernet Debele have read and evaluated the thesis entitled Artificial Intelligence Assisted Innovation for Predicting Traffic Accident on Addis Adama Expressway and examined the candidate during open defense. This is, therefore, to certify that the thesis is accepted for partial fulfillment of the requirement of the degree of Master of Science in Technology and Innovation Management

_____	_____	_____
Chairperson	Signature	Date

_____	_____	_____
Internal Examiner	Signature	Date

_____	_____	_____
External Examiner	Signature	Date

Final approval and acceptance of the thesis is contingent upon submission of its final copy to the Office of Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

_____	_____	_____
Department Head	Signature	Date

_____	_____	_____
School Dean	Signature	Date

_____	_____	_____
Office of Postgraduate Studies, Dean	Signature	Date.

ACKNOWLEDGEMENT

I extend my sincere gratitude to my advisors, Advisor(s) Dr. Ashenafi Teshome for their unwavering support, invaluable guidance, and mentorship throughout the entire journey of this Thesis. Their expertise, insightful feedback, and encouragement have played a pivotal role in shaping the direction and quality of this study. I am deeply appreciative of their dedication to fostering intellectual growth and their commitment to ensure the success of this research project.

ACKNOWLEDGEMENT	VI
ABSTRACT.....	XI
CHAPTER ONE	1
INTRODUCTION	1
1.1. Background of the Study.....	1
1.2. Statement of the Problem.....	3
1.3. Research Questions.....	4
1.4. Objective of the Study	5
1.4.1. General Objective.....	5
1.4.2. Specific Objectives	5
1.5. Scope of the Study.....	5
1.6. Significance of the Study	6
1.7. Ethical Considerations.....	7
1.8. Organization of the Study	7
CHAPTER TWO	8
LITERATURE REVIEW	8
2.1. Introduction to Traffic Accident Prediction and Artificial Intelligence.....	8
2.2. Traditional Methods of Traffic Accident Prediction.....	8
2.3. Challenges and Consequences of Road Accidents	9
2.4. Artificial Intelligence and Its Potential Applications in Road Safety	9
2.5. Machine Learning Methods.....	10
2.6. AI Techniques for Traffic Accident Prediction.....	12
2.7. State-of-the-Art AI Applications in Traffic Management	13
2.8. Case Studies of AI-Assisted Traffic Accident Prediction Systems	14
2.9. Challenges and Criticisms in AI-based Traffic Accident Prediction Models.....	15
2.10. Gap Analysis and Research Needs	16
CHAPTER THREE	18
RESEARCH METHODOLOGY.....	18
3.1. Experimental Research Design.....	18
3.1.1. Dataset	18
3.1.2. Data Preprocessing.....	18
3.1.3. Model Development.....	18
3.1.4. Training	18
3.1.5. Evaluation.....	19
3.2. Implementation Tools.....	19
CHAPTER FOUR.....	21

DATA ANALYSIS.....	21
4.1 Data Preparation.....	22
4.2 Exploratory Data Analysis.....	24
4.3. Model Description.....	49
4.3.1 Build a Model	50
4.3.2 Model Performance Comparison	51
4.4 Results and Discussion.....	53
CHAPTER FIVE	55
5.1 CONCLUSION.....	55
5.2. Recommendations	56
5.3. Limitation	57
5.4. Feature Works.....	57
REFERENCES	58
Appendix.....	60

List of Figures

Figure 1. Machine Learning Process Flow Diagram	21
Figure 2. Accident Types by Year	24
Figure 3. Accident Types by Month.	25
Figure 4 Accident Types by Date.	26
Figure 5 Distribution of Day of weeks by Accident Types	27
Figure 6. Distribution of Cause of Accident by Accident Types.....	28
Figure 7. Distribution of Vehicle Types Accident by accident Types	29
Figure 8. Distribution of Light condition by Accident Types.....	30
Figure 9 Distribution of Crash Types by Accident types.....	32
Figure 10. Distribution of Road Surface Condition Accident Types.....	33
Figure 11. Distribution of Geo-Location Accident Types	34
Figure 12 Distribution of Weather Condition by Accident Types	35
Figure 13. Distribution of Accident Types	37
Figure 14 Distribution of Accident Types by Weather Condition	38
Figure 15. Distribution of Accident Types by Weather Condition.....	39
Figure 16. Distribution of Accident Types by Vehicle Type.....	40
Figure 17. Distribution of Accident Types by Weather Condition.....	41
Figure 18. Distribution of Accident Types by Light Condition.....	42
Figure 19. Distribution of Accident Types by Crash Type	43
Figure 20. Distribution of Accident Types by Road Surface Condition.....	44
Figure 21. Distribution of Accident Types by Geo-Location	45
Figure 22. Distribution of Accident Types by Hour	46
Figure 23. Distribution of Accident Over Time.....	47
Figure 24. Distribution of Accident Types	48
Figure 25. Model Accuracy Comparison.....	52
Figure 26. Important Features derived from SVM model.....	51

ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Networks
CNN	Convolutional Neural Networks
KNN	K-Nearest Neighbors
ML	Machine Learning
SVM	Support Vector Machines
XGB	Extreme Gradient Boosting algorithm

ABSTRACT

Despite the critical importance of expressway safety, there is a notable absence of a dedicated and accurate predictive model specifically designed to forecast traffic accidents on the Addis-Adama Expressway. My research examines the application of artificial intelligence (AI) to solve the lack of a tailored predictive model and emergency services to proactively address safety concerns on Addis Adama Expressway by analyzing historical traffic accident datasets using various machine learning models, including Random Forest, Support Vector Machine, K-Nearest Neighbor, gradient boosting, and Decision Tree classifiers. My analysis identifies the major factors contributing to road Traffic accidents on the Addis Adama Expressway, such as Chainage, driver age, driver experience, day of the week, cause of accident, geo-location, vehicle types, crash type, weather condition, driver relationship, road surface condition, and sex. And also, I identify support vector machines as the most effective model, achieving an accuracy of 78.11%. Therefore, I developed a road traffic accident prediction model using support vector machines. Furthermore, I recommended that Continuous Awareness Training on Safe Driving to enhance road safety, implement a centralized traffic data system that continuously collects and analyzes information from sources such as GPS, traffic cameras, and sensors to identify accident patterns and high-risk areas, and also use predictive analytics to forecast potential hotspots and deploy preventive measures proactively.

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

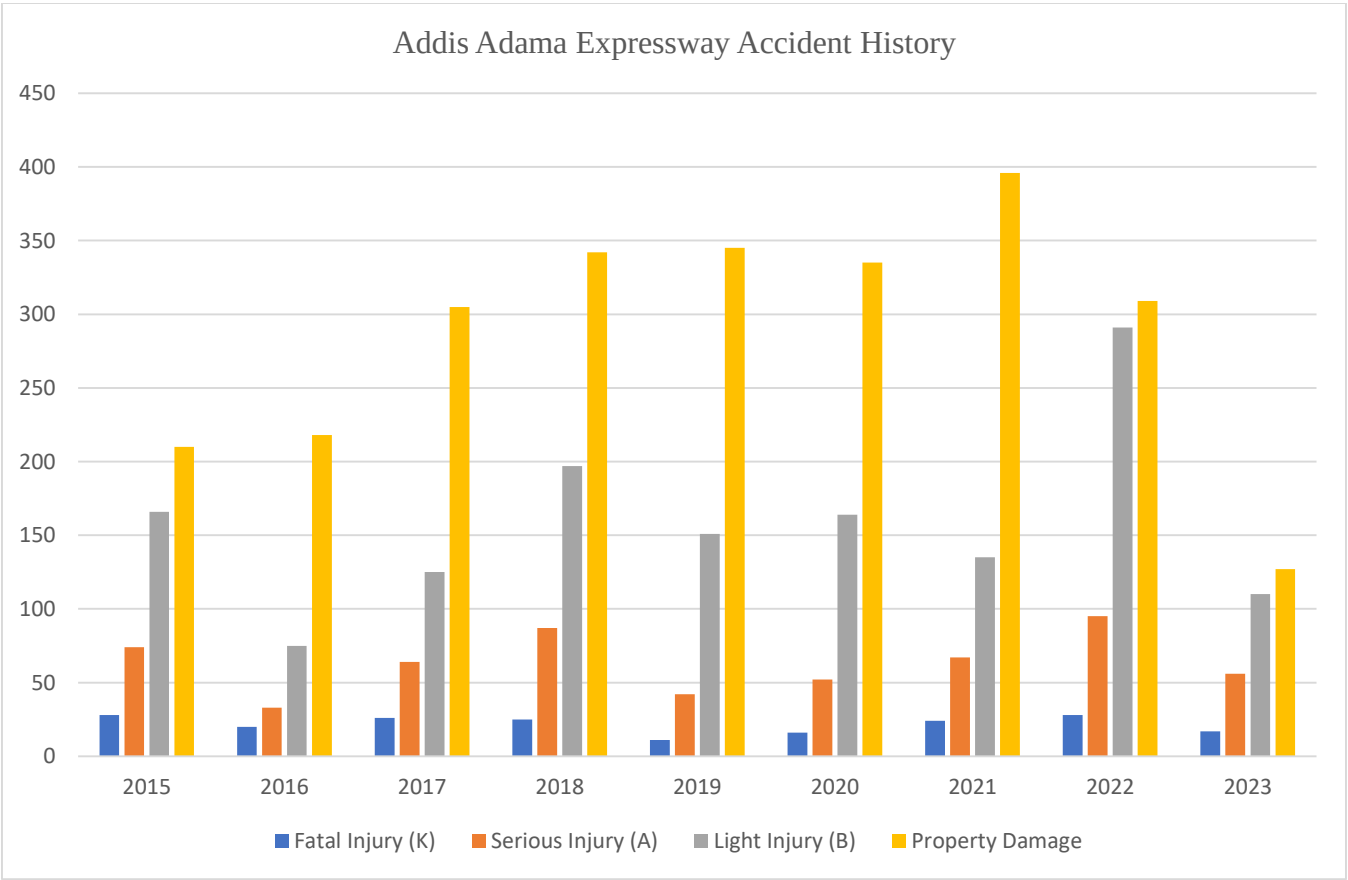
Road traffic accident is a major global concern. It is among the leading causes of death. Considerable number of people die around the world every year and several millions of people leave with non-fatal injuries as a result of traffic accident. Especially, the highest annual rate of traffic accident fatalities is recorded in Africa and it is even expected to rise with the growing economy of the region. Traffic accident crashes cost most countries 3 % off their gross domestic product. Despite the remarkable enhancements that has been made in public health status, traffic accident is a major public health problem in Ethiopia as one of the countries in Africa(Abegaz & Gebremedhin, 2019)

The road safety record of the Ethiopian Expressway is also alarming the severity of the situation and calling for an integral effort of all pertinent stakeholders to reverse the trend. The road has technologically advanced safety management system aiming to reduce traffic accident. However, traffic accidents occur frequently. Records of the Ethiopian Toll Roads Enterprise indicated that, about 3,213 traffic accidents were reported in nine years since the expressway was open for traffic. The data acquired from the enterprise, spanning from 2015 to 2023, reveals an upward trend in both traffic flows and accidents. Ethiopia's Addis-Adama Expressway, serving as a crucial transportation route, plays a pivotal role in connecting major urban hubs and facilitating economic activities. Nevertheless, the surge in traffic volume on this expressway has also resulted in an increase in traffic accidents, thus presenting significant challenges to public safety and transportation efficiency. Traditional methods of accident prediction often fall short in providing timely and accurate insights. These methods rely heavily on subjective experience and do not effectively capture the dynamic spatio-temporal correlation of traffic data (Hu et al., 2023).

Traditional methods of accident prediction often fall short in providing timely and accurate insights, necessitating a paradigm shift towards advanced technologies like artificial in Artificial intelligence (AI) plays a crucial role in traffic accident prediction and safety. AI techniques such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) have been used to predict and classify traffic accidents based on various factors (Gatarić et al., 2023). These models can accurately predict the number of accidents and the severity of their consequences, including fatalities, injuries, and property damage. By analyzing objective factors such as road length, terrain type, road width, average daily traffic volume, and speed limit, AI models can identify the most influential factors contributing to accidents, helping improve road safety and reduce the number of accidents. Additionally, AI models can predict accident

severity by considering features such as time of day, weather conditions, ages of drivers, and road surface conditions (Samia et al., 2022). These predictions aid in emergency response readiness and increase the chances of surviving accidents. Overall, AI technologies offer preventive and predictive solutions to enhance road safety and reduce the incidence of accidents. The adoption of AI in traffic management has shown promise globally, with predictive models capable of analyzing complex data patterns to forecast potential accidents. The proposed study recognizes the need for a tailored solution to address the unique challenges of the Addis-Adama Expressway. By harnessing machine learning algorithms, the research aims to develop a predictive model that considers a myriad of contributing factors, ranging from historical accident patterns to real-time weather and traffic conditions.

Moreover, the study acknowledges the ethical dimensions of deploying AI in public spaces. Ensuring fairness, transparency, and bias mitigation are integral aspects of the research, aligning with the broader objective of responsible AI development. The successful implementation of an AI-assisted innovation for traffic accident prediction on the Addis-Adama Expressway holds the potential to revolutionize how transportation authority’s approach safety, ultimately fostering a safer and more efficient commuting experience for all road users.



1.2. Statement of the Problem

Expressways play a crucial role in modern transportation, facilitating the rapid movement of vehicles over long distances. However, the inherent high speeds and complex dynamics of expressway traffic contribute to an increased risk of accidents, often resulting in severe consequences. The Addis-Adama Expressway, serving as a crucial transportation link in Ethiopia, has witnessed a concerning increase in traffic accidents, jeopardizing the safety of commuters and impeding the smooth flow of vehicular movement. Despite the critical importance of expressway safety, there is a notable absence of a dedicated and accurate predictive model specifically designed to forecast traffic accidents on the Addis-Adama Expressway. The existing methods for predicting and preventing accidents on this expressway have proven inadequate. Traditional statistical methods may be insufficient due to their inherent assumptions (Madushani et al., 2023). Traditional accident prediction methods have been mostly designed with statistical analysis that finds and analyzes the causal relationships between traffic accidents and various factors such as human, road geometry, and environmental factors.

The statement of the problem underscores the imperative for a more proactive and effective approach to address the escalating challenges of traffic accidents on the Addis-Adama Expressway. Current accident prediction models, highlighted as a primary issue, lack the capacity to handle the intricate and dynamic datasets encompassing traffic conditions, weather patterns, and road infrastructure, resulting in inaccurate and untimely predictions that impede the implementation of preventive measures. The limited integration of advanced technologies, particularly artificial intelligence (AI), exacerbates the problem, hindering the development of a sophisticated and adaptive system capable of learning from historical data and dynamically adapting to changing conditions for proactive risk mitigation. Ethical concerns arising from the deployment of AI in public spaces, such as biases and transparency issues, add another layer of complexity. The absence of tailored frameworks for the local context exacerbates these ethical considerations. Moreover, the unique challenges posed by the Addis-Adama Expressway, including specific traffic patterns, weather conditions, and road characteristics, demand solutions specifically designed for this critical transportation corridor.

In response to these multifaceted challenges, the research aims to develop an AI-assisted innovation tailored to the expressway's nuances, with the overarching goal of enhancing road safety, reducing accidents, and improving the overall commuting experience for all road users. The proposed research seeks to provide transportation authorities with a valuable tool to implement targeted safety measures, optimize resource allocation for emergency response, and ultimately enhance the overall safety of the

expressway users. The research process will involve data collection, analysis, and the application of advanced machine learning techniques to develop a reliable and efficient model. The successful implementation of a specialized expressway accident prediction model holds the potential to significantly reduce the frequency and impact of accidents on the expressway, fostering a safer and more efficient transportation system. By doing so, the study aims to contribute to the enhancement of road safety, the reduction of accidents, and the overall improvement of the commuting experience for all road users on this vital expressway.

1.3. Research Questions

This study addresses the following research questions.

RQ1: What are the key factors contributing to traffic accidents on the Addis-Adama Expressway?

This question aims to identify and analyze the various factors, including traffic accident information, weather conditions, road infrastructure, and historical accident patterns, that significantly contribute to the occurrence of accidents on the expressway. Understanding these factors is crucial for building a comprehensive and effective predictive model.

RQ2: How can artificial intelligence algorithms be leveraged to predict traffic accidents on the Addis-Adama Expressway?

This question focuses on the exploration of suitable machine learning algorithms and techniques that can process and analyze the complex datasets associated with the expressway. It encompasses the development of an AI-assisted predictive model capable of learning from historical data.

RQ3: What ethical considerations are essential for the responsible deployment of artificial intelligence in predicting traffic accidents on public roadways in Ethiopia?

Addressing the ethical dimensions of AI implementation is critical. This question explores the ethical considerations associated with the deployment of AI in the context of public safety on the Addis-Adama Expressway.

1.4. Objective of the Study

1.4.1. General Objective

The general objective of this research is to develop an artificial intelligence-assisted innovation for predicting traffic accidents on the Addis-Adama Expressway, with the aim of enhancing road safety and minimizing the occurrence of accidents on this critical transportation corridor.

1.4.2. Specific Objectives

To achieve the general objective of the study, the following specific objectives are set:

- To identify and analyze the key contributing factors to traffic accidents on the Addis-Adama Expressway
- To select and implement appropriate machine learning algorithms capable of processing and analyzing the identified factors.
- To develop a predictive model that can learn from historical data.

1.5. Scope of the Study

The primary focus of this study is on the Addis-Adama Expressway, a major transportation corridor in Ethiopia. The geographical scope is limited to the specific characteristics and conditions of this expressway, including its unique road infrastructure, traffic patterns, and accident history. The study will utilize historical accident data, traffic flow information, weather conditions, and road infrastructure details specific to the Addis-Adama Expressway. The data will be collected from relevant authorities and agencies responsible for transportation management.

The research will employ machine learning algorithms for the development of the predictive model. The selection of algorithms will be based on their suitability for processing and analyzing the identified factors contributing to traffic accidents on the expressway. The integration of the developed predictive model with existing traffic management systems will be explored within the context of the Addis-Adama Expressway. Strategies for seamless integration and potential enhancements to the current traffic management infrastructure will be proposed. The study will specifically address ethical considerations related to the deployment of artificial intelligence in predicting traffic accidents on the Addis-Adama Expressway. The development of guidelines and frameworks for responsible AI use will be tailored to the local context and the unique challenges of the study area.

The evaluation of the AI-assisted predictive model's performance will be conducted using relevant metrics such as accuracy, precision, recall, and F1 score. The assessment will focus on the model's ability

to provide timely and accurate predictions for traffic accidents on the expressway. The study acknowledges certain limitations, including the availability and quality of historical data. Additionally, external factors such as unforeseen changes in road infrastructure or policy decisions may impact the generalizability of the findings beyond the scope of the study.

The research will be conducted within a specified time frame, considering the data collection, model development, integration, and evaluation phases. The time frame ensures the timely completion of the study while maintaining the quality and reliability of the results. It is important to note that while the scope of the study is focused on the Addis-Adama Expressway, the methodologies and ethical considerations explored in this research may serve as a valuable reference for similar studies in different geographical locations or transportation contexts.

1.6. Significance of the Study

The primary significance of this study lies in its potential to significantly enhance road safety on the Addis-Adama Expressway. By leveraging artificial intelligence to predict and prevent traffic accidents, the research aims to reduce the frequency and severity of accidents, ultimately saving lives and minimizing injuries among commuters. The development of an AI-assisted predictive model contributes to more efficient traffic management on the expressway. By providing real-time insights and early warnings about potential accident-prone areas, traffic authorities can proactively implement measures to mitigate risks, optimize traffic flow, and reduce congestion. The predictive model, once integrated into existing traffic management systems, can optimize the allocation of resources such as emergency services, law enforcement, and maintenance crews. This ensures a more effective and targeted response to potential accidents, improving the overall emergency response and recovery processes.

The findings of this study can inform policy decision-makers in the transportation sector. Insights gained from the analysis of contributing factors to accidents can guide the development of evidence-based policies aimed at preventing accidents and improving overall road safety. The research contributes to the advancement of technological innovation in traffic management in Ethiopia. By embracing artificial intelligence, the transportation sector can position itself at the forefront of global efforts to leverage cutting-edge technologies for the improvement of public safety and infrastructure efficiency.

This study addresses the ethical considerations associated with deploying artificial intelligence in the context of public safety. Establishing guidelines and frameworks for responsible AI use ensures that technological advancements are aligned with ethical standards, promoting fairness, transparency, and

accountability. The developed AI-assisted predictive model can serve as a template for similar applications in other regions and transportation systems. Its adaptability and scalability make it a valuable model for future research and implementations, potentially influencing the adoption of similar innovations in diverse settings. By reducing the occurrence of traffic accidents and improving overall road safety, this research has the potential to positively impact the socio-economic landscape. Fewer accidents lead to lower healthcare costs, less damage to infrastructure, and increased productivity due to improved transportation efficiency.

In conclusion, the significance of this study extends beyond its immediate application on the Addis-Adama Expressway. It paves the way for advancements in road safety, technology integration, and ethical considerations in the deployment of artificial intelligence within the broader context of transportation management in Ethiopia.

1.7. Ethical Considerations

The proposed research will be conducted according to the ethical considerations stated by the Ministry of Science and Technology (2014). This research will meet the eight criteria of this national guideline that includes; Ethical justification, Science and social value, Favorable risk-benefit ratio to research participants and their communities, Fair selection and enrollment of human subjects, Privacy, Informed consent, and Community engagement.

1.8. Organization of the Study

This thesis begins with Chapter 1, introduces the background of the study, and subsequently followed by statement of the problem, research question, objective, scope, significance of the study, and organization of the paper. In Chapter 2, extensive literature reviews and related works explore different scientific papers on historical overview of traditional methods of traffic accident prediction and the recent advancements in artificial intelligence for traffic management. It identifies challenges and criticisms associated with the use of AI in traffic safety. Chapter 3 describes thoroughly the experimental setups and the proposed approach to be followed to answer the formulated research questions. Chapter 4 detail process of Data Analysis, exploratory data analysis, feature selection, build model, model comparison result and dissection and chapter 5 conclusion, recommendation, limitation and future works.

CHAPTER TWO

LITERATURE REVIEW

2.1. Introduction to Traffic Accident Prediction and Artificial Intelligence

Traffic accident prediction has become a hot topic in research due to the increasing number of accidents and the need to prevent and reduce them. Various techniques have been explored, including traditional statistical analysis methods, machine learning methods, neural network methods, time series analysis methods, and techniques based on spatio-temporal data mining(Wang, Changshun, et al., 2023). These methods analyze the influencing factors of traffic accidents, build prediction models, and improve prediction accuracy, providing support for road traffic accident prevention. The success of prediction models depends on the integration and correlation of data from different sources (Lazar & Jarir, 2022).

Traffic accident prediction refers to the utilization of machine learning and data analysis methodologies to anticipate the likelihood and seriousness of road traffic incidents. Traffic accident prediction is significant because it can help in improving road safety, reducing the number of accidents, and saving lives. Predicting traffic accidents allows for better decision-making in transportation planning and can assist in identifying the most influential factors contributing to accidents. The importance of predicting traffic accidents lies in its capacity to enable managers to proactively monitor and govern pertinent regions, thereby diminishing the likelihood of accidents and augmenting traffic safety. Future research should focus on developing traffic accident prediction models using deep learning, exploring and using data sources such as driver data and light conditions, and addressing issues related to high dimensionality and information imbalance (Chen & Tao, 2022).

2.2. Traditional Methods of Traffic Accident Prediction

Traditional methods of traffic accident prediction include traditional statistical analysis methods, machine learning methods, neural network methods, time series analysis methods, and techniques based on spatio-temporal data mining. These methods analyze the influencing factors of traffic accidents, build prediction models, and improve prediction accuracy, providing strong support for road traffic accident prevention (Zhang et al., 2023).traditional methods and advanced techniques contribute to the scientific and accurate prediction of road traffic accidents, aiding in the improvement of traffic safety. These methods analyze the influencing factors of traffic accidents, build prediction models, and improve prediction accuracy, providing strong support for road traffic accident prevention.

However, traditional research has limitations in deducing the trajectories of accident-related objects and extracting the spatiotemporal relationships among objects. These methods analyze the influencing factors of traffic accidents, build prediction models, and improve prediction accuracy. Statistical analysis methods, such as traditional statistical analysis and time series analysis, have been used to analyze accident data and identify patterns and trends. Rule-based systems, such as association rule mining, have been employed to discover relationships between variables and identify common accident sites. These methods provide valuable insights into accident prevention and enhance traffic safety. Traditional statistical analysis methods rely on historical data and quantitative theory, while machine learning methods use algorithms to analyze historical data and build prediction models based on patterns and regularities in the data. Both approaches have their advantages and are used in the field of traffic accident prediction.

2.3. Challenges and Consequences of Road Accidents

Road accidents have various challenges and consequences. The causes of road accidents include poor infrastructure, overloading of vehicles, law impunity, inadequate laws, corruption, negligence of drivers, lack of road safety education, and absence of road traffic signs (Haulle & Kisiri, 2016). These accidents result in loss of lives, destruction of property, increase in poverty, injuries, and permanent disabilities. Psychological problems such as acute stress reactions, phobic anxiety, depression, and posttraumatic stress disorder are common consequences of road accidents. The economic impact includes increased medical costs, longer hospitalizations, decreased productivity, and help-seeking behavior. In terms of road safety management, challenges include a lack of police officers to enforce traffic laws, absence of a road safety lead agency, and inadequate monitoring and evaluation. Safer roads and mobility are hindered by poor road design and lack of organization among governmental bodies. Additionally, vehicle safety standards are lacking, and there is a lack of awareness among road users. The post-crash response is also affected by limited ambulance coverage and difficult access to trauma centers.

2.4. Artificial Intelligence and Its Potential Applications in Road Safety

Artificial intelligence (AI) constitutes a discipline within the realm of computer science with the primary objective of facilitating the execution of tasks by computers that are analogous to those performed by humans. It encompasses a wide range of technologies, including machine learning and data analytics, which empower computers to perceive, acquire knowledge, engage in logical deduction, resolve predicaments, and strategize. The capability of AI to replicate the cognitive faculties of human beings renders it proficient in tackling intricate and ambiguous quandaries in a purposeful, intelligent, and

adaptable manner. (Olugbade et al., 2022) Artificial intelligence (AI) and machine learning (ML) have the potential to improve road safety by addressing various challenges and reducing accidents. AI and ML can be used to develop automatic incident detection systems, optimize routes, forecast cargo volume, perform predictive fleet maintenance, track vehicles in real-time, and manage traffic. These technologies can also be applied to analyze accident data and identify factors that contribute to fatal crashes and injuries, helping in better road infrastructure design and safety improvements. Additionally, AI can be used to continuously monitor road weather conditions and provide timely warnings to drivers about potentially dangerous situations (Shi & Tang, 2021) .

In the field of computer vision, AI techniques such as deep learning and convolutional neural networks (CNN) can enhance the detection and recognition of traffic road signs, improving the accuracy of intelligent vehicles and autonomous. Artificial Intelligence (AI) has emerged as a crucial tool in traffic safety, addressing various challenges and improving road management systems. AI-based techniques can detect and track traffic violations, such as helmet detection, smartphone usage while driving, and illegal parking, leading to the issuance of tickets and stricter traffic restrictions(Bala Kishore et al., 2023). Additionally, AI is used to optimize traffic flow by adjusting traffic signals based on real-time data from cameras, sensors, and other devices, reducing congestion and improving efficiency (Reddy, 2023). AI models, such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM), are employed to predict and classify traffic accidents, providing valuable insights for decision-makers and improving road safety. Furthermore, AI-enabled systems can detect pedestrians and animals on the road, transmitting this information to other road users, thereby preventing accidents and saving lives. The integration of AI in traffic safety not only enhances safety measures but also contributes to the development of smart cities and efficient transportation systems. The utilization of machine learning techniques enables the identification of influential factors such as road length, terrain type, road width, average daily traffic volume, speed limit, and weather conditions(Pourroostaei Ardakani et al., 2023). These factors can assist in improving road safety and reducing the number of accidents.

2.5. Machine Learning Methods

Machine learning techniques involve discovering patterns and regularities between data by learning autonomously through computers. These methods are based on analyzing historical traffic data and using machine learning algorithms to build prediction models to predict the probability and impact of future traffic accidents. Algorithms such as support vector machines, decision trees, logistic regression, support vector regression (SVR), and Random Forest Regression (RFR) are used to build traffic accident

prediction models. These algorithms can handle large volumes of data and capture complex relationships between various factors affecting traffic safety. These AI techniques have been applied in combination with advanced data preprocessing, feature engineering, and model evaluation methods to develop accurate and reliable traffic accident prediction models. By leveraging these techniques, researchers and practitioners aim to enhance traffic safety analysis and develop proactive measures for accident prevention. Accident prediction models have benefited significantly from the application of various machine learning algorithms, including neural networks, support vector machines (SVMs), and random forests. These algorithms are employed to analyze historical data, identify patterns, and make predictions to improve safety measures and prevent accidents (Shatnawi et al., 2023) Some of these algorithms which have been used in accident prediction models are discussed below, along with relevant citations:

Neural Networks: Neural networks, especially deep learning models, have been widely used in accident prediction due to their ability to capture complex patterns and relationships within data. In a study by (Sameen & Pradhan, 2017), deep neural networks were employed for traffic accident prediction. The authors used a combination of traffic flow data, weather conditions, and road characteristics to predict accident occurrences. The deep learning model demonstrated improved accuracy compared to traditional methods.

Support Vector Machines (SVM): SVMs are powerful classifiers that have been applied to accident prediction by mapping input features into a high-dimensional space and finding the hyperplane that best separates different classes. In a study by (Farhat et al., 2020), SVMs were used for traffic accident prediction based on a dataset containing road and traffic parameters. The SVM model demonstrated good performance in identifying potential accident locations (Zhong & Du, 2023).

Random Forests: Random Forests are ensemble learning methods that combine the predictions of multiple decision trees. They have been employed in accident prediction models to enhance accuracy and robustness. In a study by (Yan & Shen, 2022), a random forest model was utilized for traffic accident prediction using a dataset containing various road and environmental factors (Jiang & Deng, 2020). The authors found that the random forest model outperformed traditional statistical models in predicting accident occurrences. These examples illustrate the diverse applications of machine learning algorithms in accident prediction models, emphasizing their capability to handle complex data patterns and improve the accuracy of predictions. As technology and data collection methods continue to advance, machine learning algorithms are likely to play an increasingly crucial role in enhancing the effectiveness of

accident prevention strategies. AI techniques offer several advantages in terms of accuracy and efficiency when applied to traffic accident prediction.

Enhanced Accuracy: AI techniques, such as machine learning and deep learning models, can effectively analyze large volumes of complex traffic data, leading to more accurate predictions of traffic accidents. These models can capture intricate patterns and relationships among various contributing factors, resulting in improved accuracy in accident prediction

Real-Time Prediction: AI techniques enable real-time analysis of traffic data, allowing for the continuous monitoring and prediction of potential accidents. This real-time capability enhances the efficiency of accident prediction systems by providing timely alerts and interventions to mitigate risks on the road.

Adaptive Learning: AI models can adapt and learn from new data, enabling them to continuously improve their predictive accuracy over time. This adaptive learning capability ensures that accident prediction systems remain effective in dynamically changing traffic environments.

Integration of Diverse Data Sources: AI techniques can integrate diverse sources of data, such as traffic flow, weather conditions, road infrastructure, and vehicle characteristics, to provide a comprehensive understanding of the factors influencing traffic accidents. This integration enhances the accuracy of accident prediction by considering multiple relevant variables. These advantages demonstrate the potential of AI techniques to significantly improve the accuracy and efficiency of traffic accident prediction systems, leading to better-informed decision-making and proactive safety measures on the road.

2.6. AI Techniques for Traffic Accident Prediction

AI techniques for traffic accident prediction have been explored in several studies. Barakat and Ghuzlan developed a prediction model using Artificial Neural Networks (ANNs) with input features such as time of day, weather conditions, and driver characteristics, achieving an accuracy of approximately 80% (Hamdan et al., 2023). (Pourroostaei Ardakani et al., 2023) investigated meaningful data features and used decision trees, random forest, multinomial logistic regression, and naïve Bayes for prediction, with acceptable levels of accuracy except for naïve Bayes. Alagarsamy et al. also focused on identifying key contributing factors for accidents and proposed a system for predicting accidents using machine learning algorithms, with the most accurate model determined through contrasting methods (Alagarsamy

et al., 2023). Zhang and Guo proposed a deep learning method called attention-based residual dilated network (ARDN) for accurate accident prediction, outperforming other machine learning methods in experimental evaluations. These studies highlight the potential of AI techniques in predicting traffic accidents and improving road safety.

2.7. State-of-the-Art AI Applications in Traffic Management

Artificial Intelligence (AI) has significant applications in traffic management. It can address challenges such as traffic congestion, safety concerns, and environmental effects. AI can optimize traffic flow by adjusting traffic signals based on real-time data from cameras, sensors, and other devices (Reddy, 2023). It can also forecast future traffic density using AI-based methods, helping to reduce congestion (Nechoska et al., 2022). Additionally, AI can prioritize emergency vehicles during traffic jams using RFIDs. AI in transportation can collect traffic data to alleviate congestion and improve public transport scheduling. It can also be used for transportation management, planning, control, and accident avoidance. AI technology can be applied to traffic signal control systems, creating an all-round, multi-level traffic control "brain" for more accurate and real-time control. Furthermore, AI has applications in driverless driving, intelligent dispatching, and route planning. However, the ethical and legal implications of AI in traffic management need to be considered.

Artificial Intelligence (AI) has significant applications in traffic management. It can address challenges such as traffic congestion, safety concerns, and environmental effects. AI can optimize traffic flow by adjusting traffic signals based on real-time data from cameras, sensors, and other devices (Reddy, 2023). It can also forecast future traffic density using AI-based methods, helping to reduce congestion (Nechoska et al., 2022). AI can Predictive Analytics: use historical data and machine learning models to predict traffic patterns and congestion. This information can be used to implement proactive measures, such as adjusting signal timings or suggesting alternative routes to drivers, to alleviate potential traffic issues. The paper discusses the application of Artificial Intelligence (AI) in traffic management and control to address transportation problems such as increased travel demand, safety concerns, congestion, and environmental effects.

Recent advancements in artificial intelligence (AI) for traffic management have shown promising results in improving various aspects of transportation systems. Machine learning and AI techniques have been utilized to analyze and interpret data in the context of smart cities, particularly in the field of intelligent transportation systems (Bharadiya). These technologies have been applied to tasks such as traffic prediction, congestion management, and traffic flow control. Additionally, AI techniques have been

employed in the context of Cloud Radio Access Networks to estimate and mitigate traffic congestion, leading to improved bandwidth utilization and reduced latency. The use of AI in air traffic management has also been explored, offering alternative approaches to conventional modeling techniques and providing efficient solutions to multivariable problems. Overall, recent advancements in AI have shown great potential in enhancing traffic management systems and addressing the challenges associated with traffic congestion.

2.8. Case Studies of AI-Assisted Traffic Accident Prediction Systems

These real-world implementations demonstrate the effectiveness of AI-assisted traffic accident prediction systems in enhancing road safety and enabling proactive interventions. By leveraging predictive analytics and real-time monitoring, these systems contribute to the reduction of accidents and the improvement of traffic management in urban and regional environments. Some case studies provide practical insights into the successful application of AI in traffic accident prediction. The subsequent paragraphs present the case studies of AI-assisted traffic accident prediction systems across various cities in the world.

The City of Bellevue implemented an AI-based traffic accident prediction system that integrates historical accident data, real-time traffic flow information, and environmental factors to identify accident-prone locations and predict potential high-risk scenarios. The system uses machine learning models to enable proactive interventions and targeted safety measures. (Zhang, H., Sadek, A. W., & Lee, D. H., 2018)

The Singapore Land Transport Authority deployed an AI-based traffic accident prediction and prevention system that utilizes machine learning models to analyze traffic flow patterns, weather conditions, and driver behavior data collected from connected vehicles. The system provides predictive alerts and interventions for accident-prone zones, contributing to enhanced road safety measures.(Cheng, 2019) .

Prediction of traffic accidents based on random forest in a highway network.

Wang et al. proposed a traffic accident prediction system based on the Random Forest classifier specifically tailored for highway networks. They collected a comprehensive dataset consisting of historical accident records, traffic flow data, and weather conditions. By extracting relevant features from the data and training a Random Forest classifier, the system was capable of accurately predicting traffic accidents on highways. The study demonstrated the applicability of machine learning techniques, such

as Random Forest, in improving the safety of highway transportation through proactive accident prediction and prevention strategies(Khanum et al., 2023).

Li et al. developed a traffic accident prediction model based on Long Short-Term Memory (LSTM) neural networks. They collected historical urban traffic accident data and combined it with various features such as time, weather conditions, and road geometry. The LSTM model was trained to learn temporal patterns from the data and accurately predict the occurrence of traffic accidents in urban areas. The study demonstrated the effectiveness of LSTM-based models in capturing complex temporal dependencies and improving the accuracy of traffic accident prediction(Wang, Yan, et al., 2023).

(Alqatawna et al., 2021)The paper discusses the analysis and prediction of road traffic accidents (RTAs) on high accident locations highways in Spain using Artificial Neural Networks (ANNs) .The ANN models were developed using 4 years of data for accident counts on Spain freeway roads from 2014 to 2017 .The best ANN model was selected, which considered variables such as highway sections, years, section length, annual average daily traffic, average horizontal curve radius, slope gradient, and traffic accidents involving heavy vehicles .The results show that the ANN model provides a potentially powerful tool for analyzing and predicting traffic accidents, with estimates close to actual traffic accidents in Spain .The performance of the ANN model was compared to a multivariate regression model, and the ANN model was found to be stronger in forecasting future highway traffic accidents in Spain .

2.9. Challenges and Criticisms in AI-based Traffic Accident Prediction Models

AI-based traffic accident prediction models face several challenges and criticisms. One major challenge is the data imbalance problem, where input data contains numerous zeros, negatively affecting risk estimation results(Lazar & Jarir, 2022). Another challenge is neglecting road environmental factors, which are essential in causing traffic accidents. Uncertainty in modeling traffic accidents is also a challenge, as it is difficult to account for all factors (Kim et al., 2023). Additionally, collecting proper samples for training the models can be a challenge Criticisms include the need for more objective and subjective factors to improve prediction accuracy. There is also a need for better interpretation of results and compatibility with existing road safety management systems. Overall, addressing these challenges and criticisms is crucial for the development and effectiveness of AI-based traffic accident prediction models. The use of AI technology in algorithmic decisions has raised issues of fairness, accountability, transparency, and ethics. The effectiveness of transparency methods in addressing these issues has been assessed, and it has been found that explanations provided by AI systems tend to be vague and less

informative. The growing use of AI in shaping information environments and generating content in the public arena has led to a strengthening of intermediary structures that can exercise control over the public arena (Hemment et al., 2023). The data-driven nature of AI applications also poses challenges to the political status quo and obstructs the assess ability of AI-enabled interventions (Gaud, 2023). There is a need for clarity on how decisions regarding ethical dilemmas are made in public administration implementations of AI and their impact on public values. While AI-based models offer substantial benefits, they are not without challenges. Concerns regarding data privacy, algorithmic bias, and the interpretability of complex machine learning models have been raised. The ethical implications of relying on AI for critical decision-making in public spaces, such as road safety management, require careful consideration. A comprehensive understanding of these challenges is essential for ensuring the responsible deployment of AI technology in traffic accident prediction.

2.10. Gap Analysis and Research Needs

Traditional traffic accident prediction techniques face several limitations, including reliance on small-scale datasets, dependence on extensive data, and a lack of real-time applicability. To address these gaps, recent research has proposed innovative solutions incorporating deep neural networks and leveraging easy-to-obtain but sparse data (Gatarić et al., 2023). These solutions integrate components like recurrent and fully connected networks to capture time-sensitive and time-insensitive data, respectively.

Existing literature on traffic accident prediction has explored various techniques, encompassing traditional statistical analysis methods, machine learning methods, neural network methods, time series analysis methods, and spatio-temporal data mining techniques (Cheng, 2019). Additionally, graph-based approaches like Graph Neural Networks (GNNs) have been proposed to incorporate graph structure information into accident prediction. However, the absence of suitable graph-structured traffic accident datasets has limited the application of GNNs. There is also a notable gap in exploring the relationships between different accident locations within road networks, highlighting the need for further research.

The existing literature on traffic accident prediction reveals significant gaps and limitations, including the absence of suitable graph-structured traffic accident datasets for applying GNNs (Huang et al., 2023) and limited consideration of text data related to traffic accidents, which can offer valuable insights for predicting accident duration. Small-scale datasets with limited coverage and a dependency on extensive data represent important shortcomings, constraining the real-time applicability of existing approaches and hindering the prediction of rare accident events (Chen & Tao, 2022). Addressing these limitations necessitates more comprehensive and diverse datasets, along with the integration of text data and real-

time information, to enhance the accuracy of traffic accident prediction models. Traffic accidents on expressways pose a distinct context requiring a tailored solution, yet existing literature often lacks a specific focus on expressway accidents, resulting in knowledge gaps. Factors such as the high-speed environment, specific infrastructure design and layout, and unique traffic dynamics on expressways contribute to the complexity of accident occurrences. The higher speed limits on expressways introduce unique risks, including reduced reaction time and an increased potential for severe collisions, emphasizing the need to understand specific contributing factors for effective preventive measures. Investigating the interaction between expressway infrastructure design, human behavior, and accident occurrence is crucial for informed safety improvements. Additionally, heavy traffic flow on expressways during peak hours can lead to congestion-related accidents, necessitating a deeper understanding of traffic dynamics to develop mitigation strategies.

Despite advancements in AI-based traffic accident prediction, there remains a notable gap in the literature regarding the specific conditions and challenges of the Addis-Adama Expressway. The unique traffic patterns, infrastructure, and contextual factors demand a tailored solution, and this research aims to fill this gap by developing an artificial intelligence-assisted innovation designed for predicting traffic accidents on the Addis-Adama Expressway. This literature review provides a structured overview, positioning the research within existing knowledge while emphasizing specific contributions and objectives.

CHAPTER THREE

RESEARCH METHODOLOGY

The study adopted a quantitative research approach, as it involves the collection and analysis of numerical data to generate predictive models (Lindholm et al., 2022). This approach is suitable for developing a predictive model for traffic accidents on the Addis-Adama Expressway.

3.1. Experimental Research Design

This study adopted experimental research design (Shukla, 2022), focusing on the development and evaluation of a machine learning model for the successful implementation of the proposed artificial intelligence-assisted innovation for predicting traffic accidents on the Addis-Adama Expressway.

3.1.1. Dataset

Historical accident data obtained from relevant traffic management authorities. Traffic accident information, weather conditions, and road infrastructure details will be collected from appropriate agencies and meteorological departments. The sample size will be determined based on the volume of historical accident data and the need for a representative dataset for model development. A balance between sufficiency and feasibility will be considered. Sample traffic Accident Data Type Received from Ethiopian Toll Roads Enterprise include date of accident occur, time of accident occur, day of the week, crash type, Cause of accident, fatal injuries, serious injuries, light injuries, chianage, geometrical condition, road surface condition, weather condition, light condition, Vehicle Types, Vehicle Ownership, age of driver, Sex and Driving Experience.

3.1.2. Data Preprocessing

The collected data undergo several preprocessing steps to ensure quality and consistency. Cleaning, normalization, and feature engineering will be performed on the collected data to enhance its quality and relevance for model training.

3.1.3. Model Development

Machine learning algorithms, including neural networks, decision trees, and ensemble methods, will be implemented for the development of the predictive model. The training set will be used to train the model, and the testing set will be employed to assess its predictive performance. Machine learning algorithm will be employed for model implementation.

3.1.4. Training

The dataset split into training, validation, and testing sets. The model will be trained on the preprocessed dataset using a suitable optimization algorithm and loss function. Hyper parameter tuning will be

performed to optimize the model's performance, and training progress will be monitored through appropriate metrics.

3.1.5. Evaluation

The developed model evaluated using standard metrics such as accuracy, precision, recall, and F1 score. Cross-validation techniques will be employed to ensure the robustness of the model's performance.

- Accuracy: The proportion of correctly classified instances.
- Precision: The ratio of true positives to the sum of true positives and false positives.
- Recall: The ratio of true positives to the sum of true positives and false negatives.
- F1-score: The harmonic means of precision and recall.

3.2. Implementation Tools

In this study, various tools and libraries will be used for implementation.

Python

Python programming language was used for implementing the machine learning model. Python provides extensive libraries and frameworks are widely used in the development of machine learning models.

NumPy

NumPy is a fundamental library for numerical operations in Python, essential for handling arrays and matrices. NumPy used for data manipulation and preprocessing.

Pandas

Pandas can be employed for structured data manipulation and analysis, facilitating tasks such as dataset organization and statistical analysis.

Scikit-learn

Scikit-learn provides a variety of tools for machine learning tasks, including model selection, evaluation, and preprocessing. It will be used for splitting the dataset and evaluating the model.

Jupyter Notebooks

Jupyter Notebooks offer an interactive and flexible environment, allowing seamless integration of code, visualizations, and documentation. They will be used for prototyping and sharing research findings.

Optuna

Optuna is an automatic hyper parameter optimization framework. It will be used to efficiently search for optimal hyper parameters and enhance the model's performance.

Google Colab or AWS Sagemaker

Cloud computing services utilized for training large models and conducting experiments that require substantial computational resources. Google Colab provides free access to GPU resources, while AWS Sage maker offers scalable cloud computing resources.

Matplotlib: -

Matplotlib is a python library that is used for creating static, animated, and interactive visualizations in Python. In the proposed study, it will be used for visualizing data and results by using a graph. The selection of appropriate implementation tools is crucial for the success of the study. The tools outlined above provide a comprehensive set of resources for data preprocessing, model development, training, and evaluation, ensuring a robust and efficient implementation of the proposed artificial intelligence-assisted innovation for predicting traffic accidents on the Addis-Adama Expressway.

CHAPTER FOUR

DATA ANALYSIS

Data analysis is the process of examining, cleaning, transforming, and modeling data with the goal of extracting meaningful insights, discovering patterns, and making data-driven decisions. It involves applying statistical and analytical techniques to raw data in order to uncover hidden relationships, trends, and anomalies. In this phase, the process of procuring and processing the dataset is performed so that the dataset is usable for EDA and modeling in the subsequent steps. Moreover, the dataset has been explored briefly to indicate the features that will be dealt with. Predicting traffic accidents using machine learning can be a vital task in ensuring road safety and minimizing the number of accidents. By analyzing various features such as weather conditions, road surface conditions, and other relevant factors, machine learning models can help in identifying the likelihood and severity of different types of accidents. In my scenario, the target variable is the Accident Types, which can be categorized as Fatal Accident, Serious Accident and Slight Accident. The goal is to build a predictive model that can accurately classify the type of accident based on the provided features. To tackle this problem, I will employ a supervised machine learning approach. Supervised learning involves training a model on labeled data, where the model learns the relationship between input features e.g., weather conditions, road surface conditions and the target variable. Once trained, the model can make predictions on new, unseen data. I utilize a classification algorithm to handle the multi-class nature of the target. The general workflow for building and evaluating the predictive model will involve several key

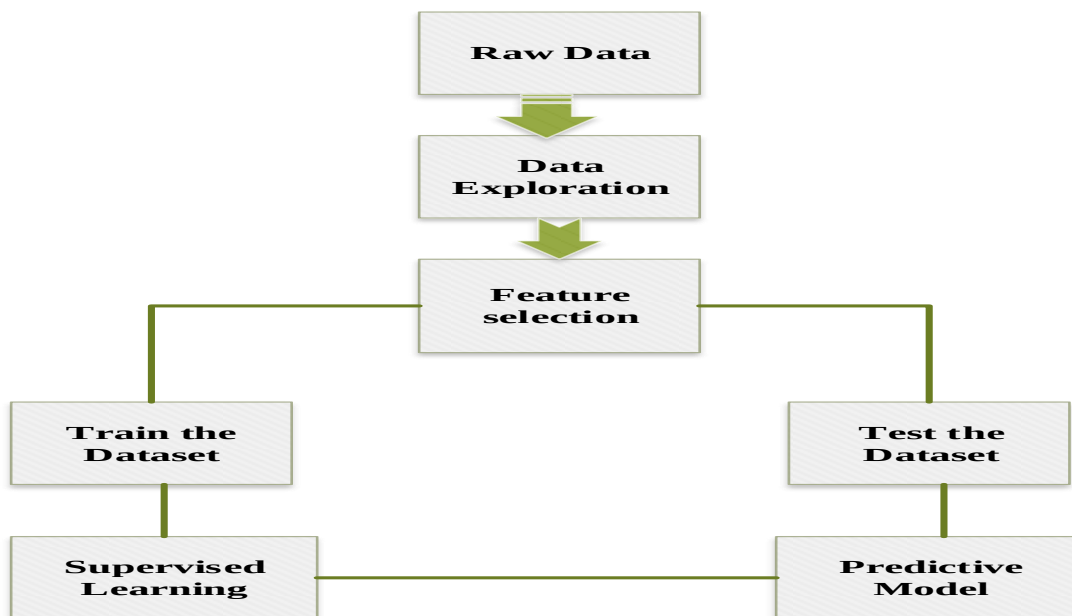


Figure 1. Machine Learning Process Flow Diagram

4.1 Data Preparation

Data preparation is the process of cleaning, transforming, and organizing raw data into a usable format for analysis and modeling. It involves removing errors, inconsistencies, and irrelevant data, as well as converting data into a suitable structure and format for subsequent processing. Data preparation is a crucial step in the data analysis process that involves transforming raw data into a clean and usable format for analysis. It ensures that data is accurate, consistent, and complete, thereby enhancing the quality and reliability of the analysis results. The main goal of data preparation is to make the data suitable for modeling and analysis by addressing issues such as missing values, errors, and inconsistencies. My target variable is Accident Types with the Following Value.

- **Fatal Accidents:** Fatal accidents are traffic accidents that result in the death of one or more individuals involved. These accidents often involve severe impacts, high speeds, or vulnerable road users such as pedestrians. (WHO).
- **Serious Accidents:** Serious accidents are those that cause major injuries but do not result in death. These injuries might include severe fractures, internal injuries, severe burns, or injuries that require extensive medical treatment and hospitalization(WHO).
- **Slight Accidents:** Slight accidents involve minor injuries and property damage. These might include minor cuts, bruises, or slight vehicle damage(WHO).

Historical traffic accident datasets provide a wealth of information about the causes, patterns, and consequences of traffic accidents. These datasets typically include data on a range of features, including: Time, Date of week, Crash types, Cause of accident, Chainage, Light condition Geolocation, Road surface condition, Weather condition, Accident types, Driver relationship Driver age, Driving experience Driver sex.

Day of week: This feature indicates the day of the week on which the accident occurred. It can be useful for identifying patterns and trends in accident occurrence based on weekdays and weekends.

Crash types: This feature describes the type of accident that occurred, such as single-vehicle, multi-vehicle, etc. This information can help in understanding the different types of accidents that occur and their frequency.

Cause of accident: This feature identifies the primary cause of the accident, such as driver error, vehicle malfunction, road conditions, etc. This information is crucial for developing targeted safety interventions and improving road safety.

Chainage: This feature refers to the distance along a road where the accident occurred. It can be useful for identifying high-risk locations and implementing appropriate safety measures.

Light condition: This feature indicates the lighting conditions at the time of the accident, such as daylight, dawn/dusk, or nighttime. This information can help in understanding the impact of visibility on accident occurrence.

Geolocation: This feature provides the precise location of the accident, including latitude and longitude coordinates. This information can be used for spatial analysis and identifying accident hotspots.

Road surface condition: This feature describes the condition of the road surface at the time of the accident, such as dry, wet, icy, or snowy. This information can help in understanding the role of road conditions in accident causation.

Weather condition: This feature indicates the weather conditions at the time of the accident, such as clear, rain, snow, fog, etc. This information can help in identifying the impact of weather on accident occurrence.

Accident types: This feature classifies the accident based on its severity, such as fatal, serious injury, minor injury, or property damage only. This information is important for understanding the consequences of accidents and prioritizing safety efforts.

Driver relationship: This feature describes the relationship between the driver and the other vehicle(s) involved in the accident, such as same direction, opposite direction, or turning. This information can help in understanding the dynamics of accident occurrence.

Driver age: This feature indicates the age of the driver involved in the accident. This information can be useful for identifying age-related risk factors and targeting safety programs.

Driving experience: This feature refers to the number of years of driving experience of the driver involved in the accident. This information can help in understanding the role of experience in accident causation.

Driver sex: This feature identifies the sex of the driver involved in the accident. This information can be useful for analyzing gender-related differences in accident occurrence and developing targeted safety Intervention.

4.2 Exploratory Data Analysis

Exploratory data analysis (EDA) is a crucial step in the data analysis process where researchers and analysts seek to gain a deeper understanding of their data by exploring, visualizing, and summarizing its key characteristics. EDA helps identify patterns, trends, and potential outliers that may guide further analysis and hypothesis formulation.

In this section of the report, the data has been leveraged and python was used to shape the dataset. After the data is processed and ready for use, various visualization techniques have been used such as univariate and bivariate analysis using different packages. This will help in understanding some of the core aspects of the dataset along with some understanding like data distribution, relationship between different features in the data as well as many other details. After determining an understanding of the dataset, the information and insights are used to move on to the next phase of the analysis and modeling.

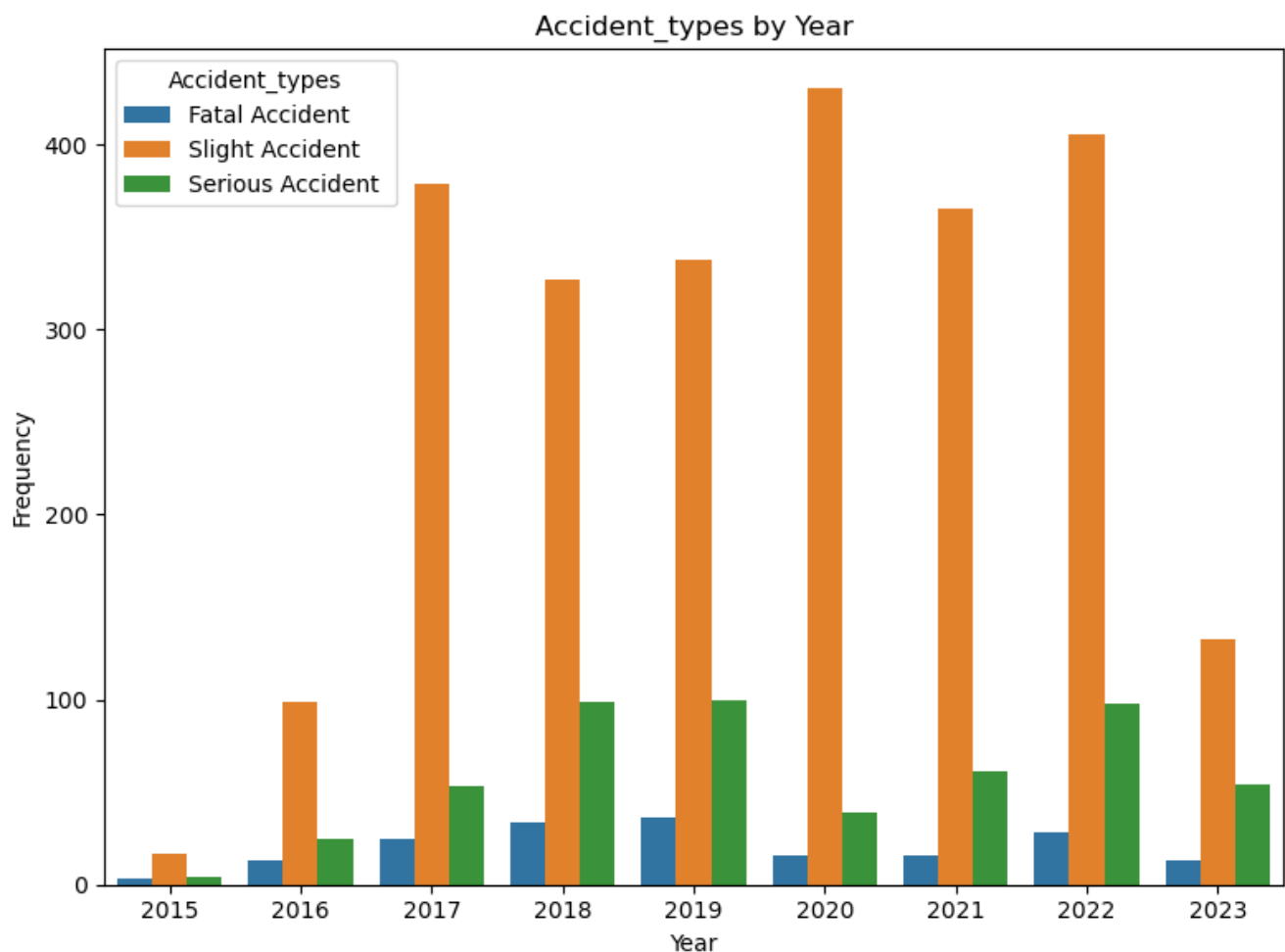


Figure 2. Accident Types by Year.

The graph shows the number of fatal accidents, serious accidents, and slight accidents that occurred each year. The x-axis of the graph represents the year, and the y-axis represents the number of accidents.

The graph shows that the number of fatal accidents has been increased over the years. For example, in 2022, there were approximately 28 fatal accidents, which is increase from the approximately 24 fatal accidents that occurred in 2021. The number of serious accidents has also been increased over the years. In 2022, there were approximately 97 serious accidents, which is increase from the approximately 67 serious accidents that occurred in 2021. The number of slight accidents also increased over the years. In 2021, there were approximately 291 slight accidents, which is increase from the approximately slight accidents that occurred in 2020. Overall, the graph shows a negative trend in terms of the number of traffic accidents. The number of fatal and serious and slight accidents has been increased. This suggests there must be improve road safety needs more improvement. Overall, this plot provides a clear visualization of the distribution of different accident types across different years. The graph shows the distribution of traffic accident Types in Addis Adama Expressway in the past Eight years. The trend shows up and dawn in the proportion of severe accidents, from 2.8% in 2015 to 16.5% in 2023.

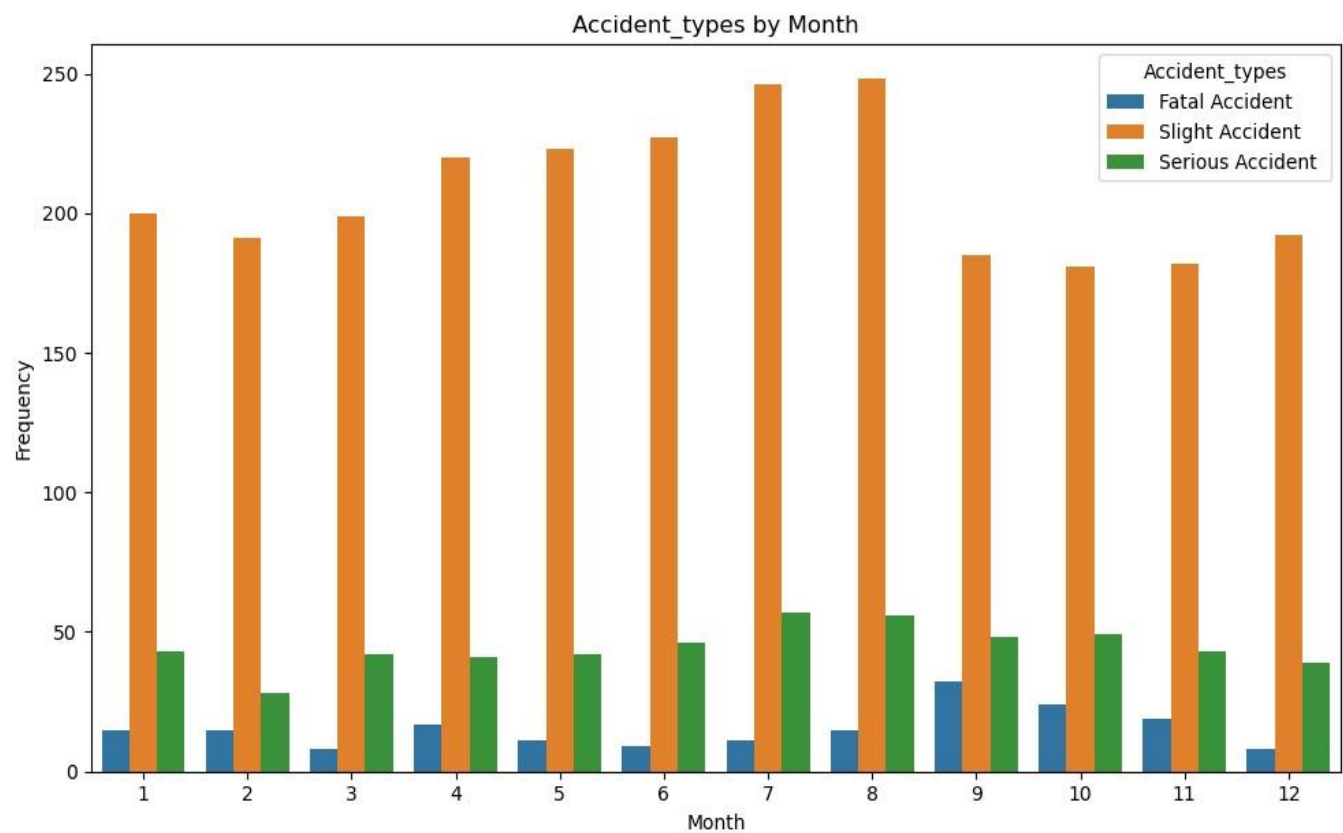


Figure 3. Accident Types by Month.

The graph shows the number of fatal accidents, serious accidents, and slight accidents that occurred each month. The x-axis of the graph represents the month, and the y-axis represents the number of accidents. The graph shows that the number of fatal accidents is highest in the September, October, June, July and August. This is likely due to the increased number Traffic flow and other factors. The number of serious accidents is also highest in such months. The number of slight accidents is relatively increase in each month. Overall, the graph shows that the risk of traffic accidents is highest during the summer months. This is likely due to the increased number traffic flow and the more favorable weather conditions for driving. Over all the distribution of accident Types in Addis Adama Expressway by month over the past Eight years. The overall pattern shows that the proportion of Accident Accidents Types is highest during the winter months and lowest during the summer months.

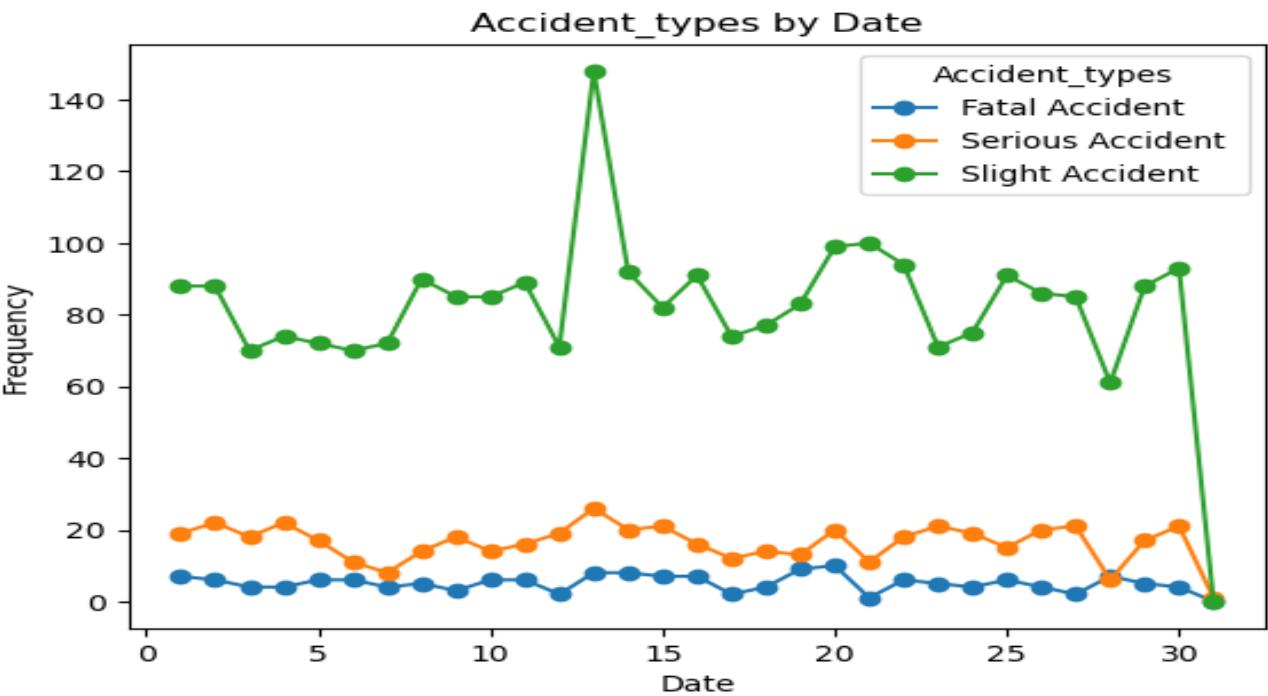


Figure 4. Accident Types by Date.

The graph shows the number of fatal accidents, serious accidents, and slight accidents that occurred on each Date. The x-axis of the graph represents the date within a five days’ intervals, and the y-axis represents the number of accidents. The graph shows that the number of fatal accidents is up and down within each days. This is likely due to the increased number of traffic flow and rest date of the weeks. The number of serious accidents is also highest on weekends date, but there is also a peak 15 to 20 dates. The number of slight accidents is relatively increase throughout the week. Overall, the graph shows that the risk of traffic accidents is highest on weekends. This is likely due to the increased number traffic flow and the more relaxed driving habits that are common on weekends. The graph shows the distribution of

accident Types in Addis Adama expressway by date over the past eight years. The overall pattern shows that the proportion of accidents Types is highest on weekends and holidays. This is likely due to increased traffic volume and congestion during these times, as well as potentially more impaired driving.

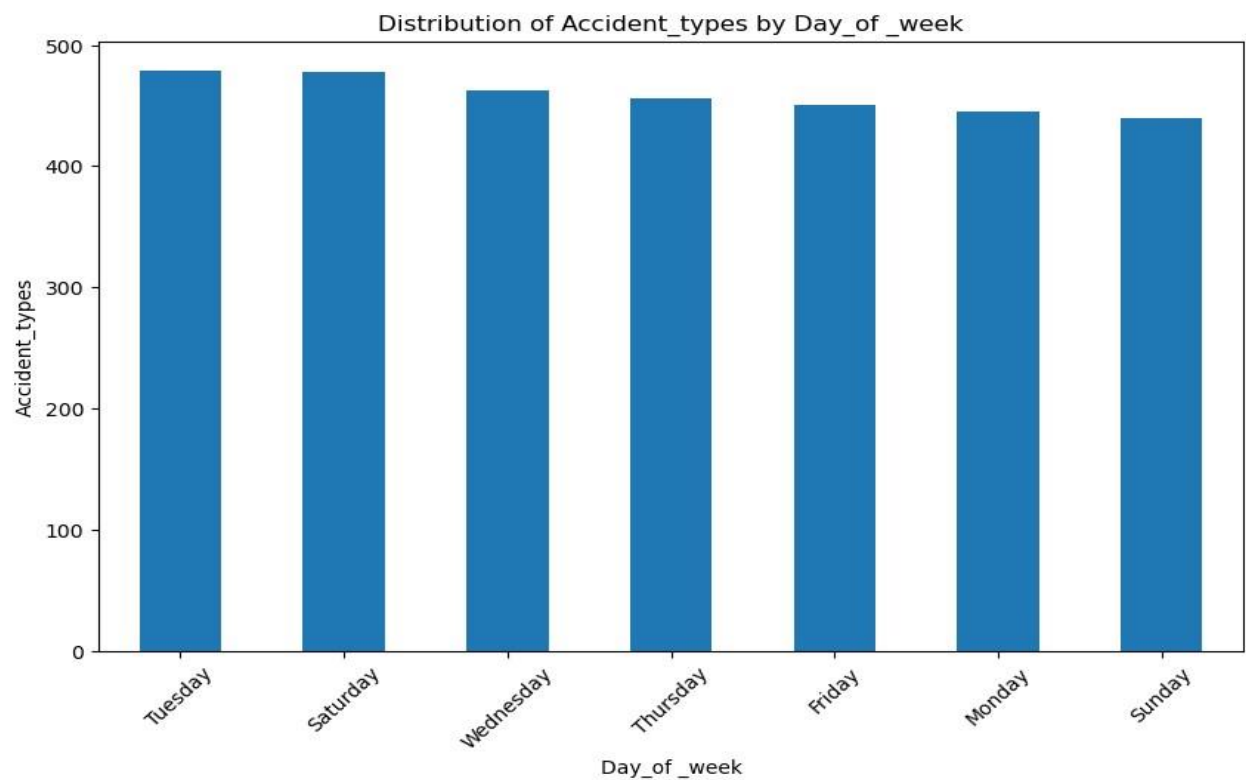


Figure 5. Distribution of Day of weeks by Accident Types

The graph depicting the distribution of traffic accidents by day of the week shows clear variations in accident frequency. The data indicates that traffic accidents are most prevalent on Tuesday and Saturdays, suggesting increased risk towards the end of the week, possibly due to higher traffic volumes and weekend activities. Conversely, Sundays and Mondays have the lowest number of accidents, pointing to a decrease in traffic incidents at the start of the week. Midweek days, including Tuesday, Wednesday, and Thursday, exhibit moderate and relatively stable accident rates. This distribution highlights a pattern where traffic accidents tend to rise as the week progresses, peaking on weekends and tapering off at the beginning of the week. Over all the graph shows the distribution of accident Types by day of the week in Addis Adama Expressway during the past eight years.

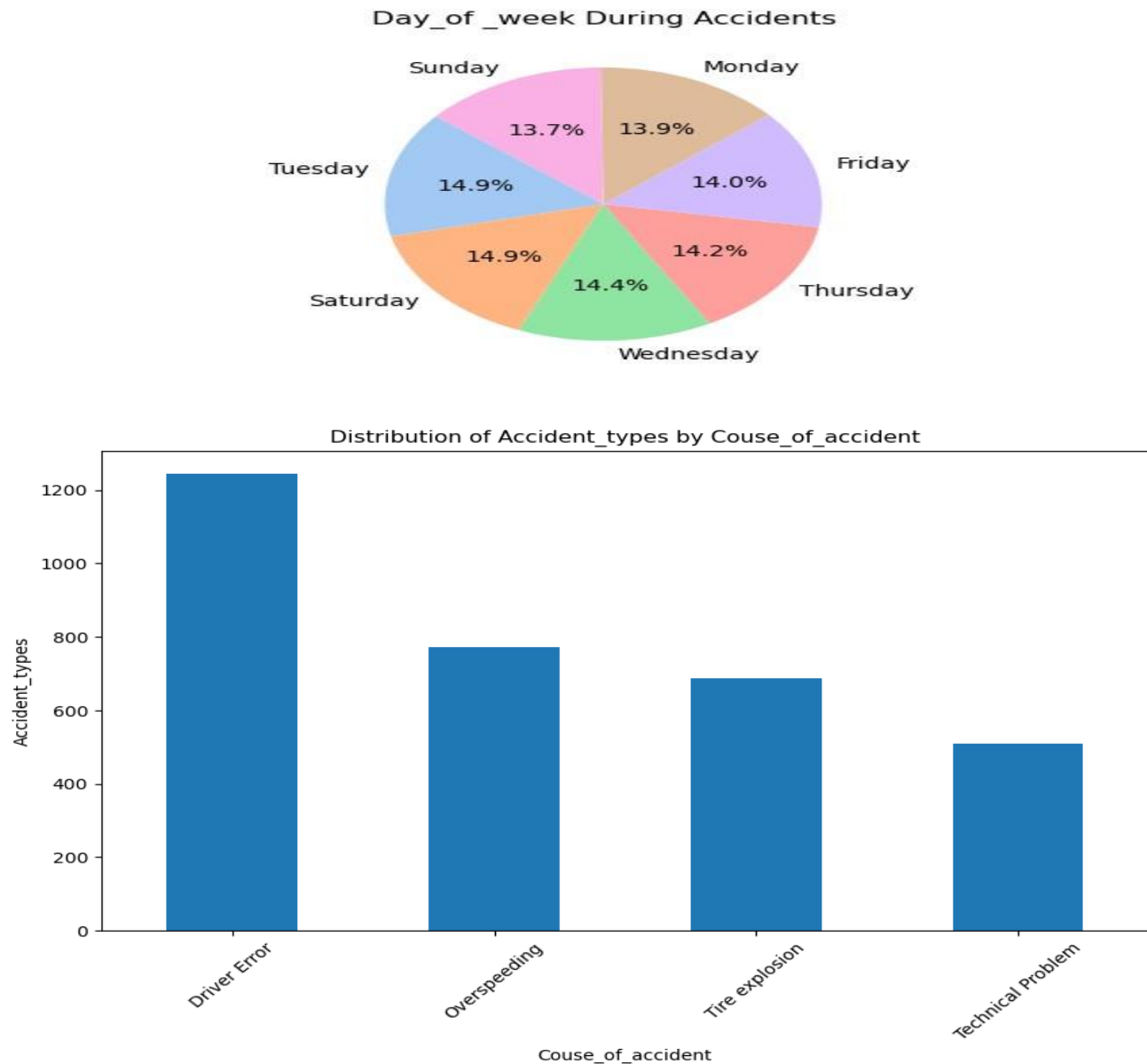


Figure 6. Distribution of Cause of Accident by Accident Types

The graph illustrating the distribution of causes of traffic accidents highlights four primary factors. Driver error is the leading cause, accounting for the majority of accidents and encompassing behaviors such as failure to yield, improper turns, and distracted driving. Over speeding is the second most prevalent cause, indicating a significant number of incidents where drivers exceed speed limits, often resulting in severe collisions. Tire explosions rank third, representing a notable proportion of accidents likely due to poor maintenance or defective tires. Technical problems, such as brake failures and engine issues, are the fourth cause, contributing to a smaller yet important segment of accidents. This distribution underscores the critical need for driver education on safe driving practices, adherence to speed limits, regular vehicle maintenance, and addressing technical issues promptly to enhance overall road safety. The graph shows

the distribution of accident causes in Addis Adama Expressway during the past eight year. The most common cause, accounting for 38.79 % of accidents, is Driver error. This is followed by Over speeding (24%), Tire explosion (21.4%) and Technical problem (15.9%).

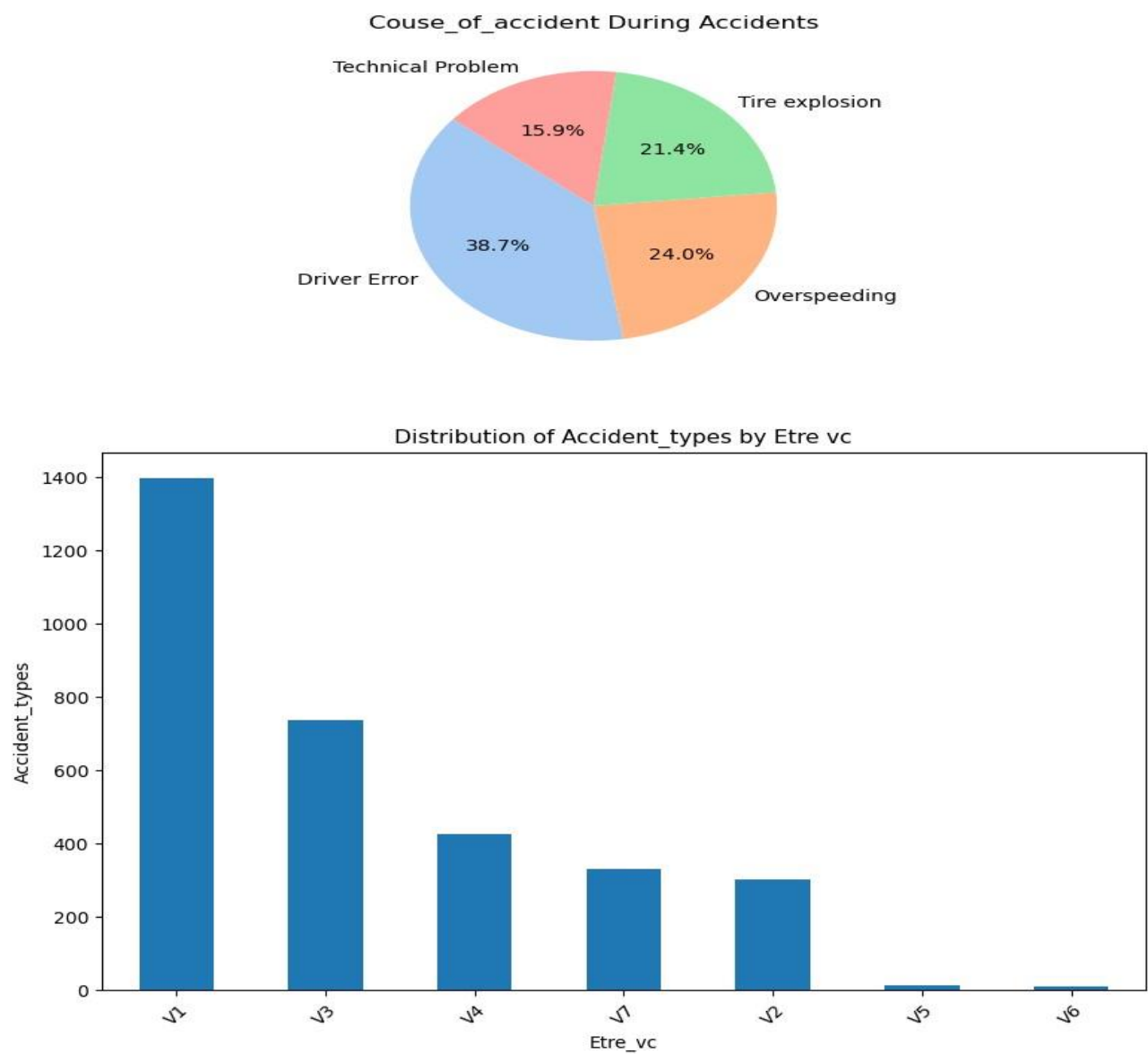


Figure 7. Distribution of Vehicle Types Accident by accident Types

The graph showing the distribution of vehicle types involved in traffic accidents reveals several key trends. Vehicle Type One is the most frequently involved in accidents, indicating it may be the most common or highest-risk vehicle on the road. Vehicle Type Three follows as the second most prevalent, suggesting a significant presence in traffic incidents. Vehicle Type Four ranks third, still contributing notably to the overall accident data. Vehicle Type Seven comes next, with a slightly lower but considerable share. In the subsequent categories, Vehicle Types Two, Five, and Six all show similar and

relatively lower frequencies of involvement in accidents, indicating they are either less common on the road or generally less prone to accidents. This distribution emphasizes the need to focus on the most frequently involved vehicle types when designing traffic safety measures and interventions. Over all the graph shows the distribution of accident Types by vehicle type in Addis Adama Expressway during the past eight years. The highest proportion of accidents Types (43.5%) involves Vehicle one(V1). This is followed by Vehicle Three (V3) (22.9%), Vehicle Four (13.2%), and Vehicle Seven (10.3%). The remaining 10.1% of severe accidents involve other vehicle types such as vehicle type two, vehicle Type five and vehicle type six.

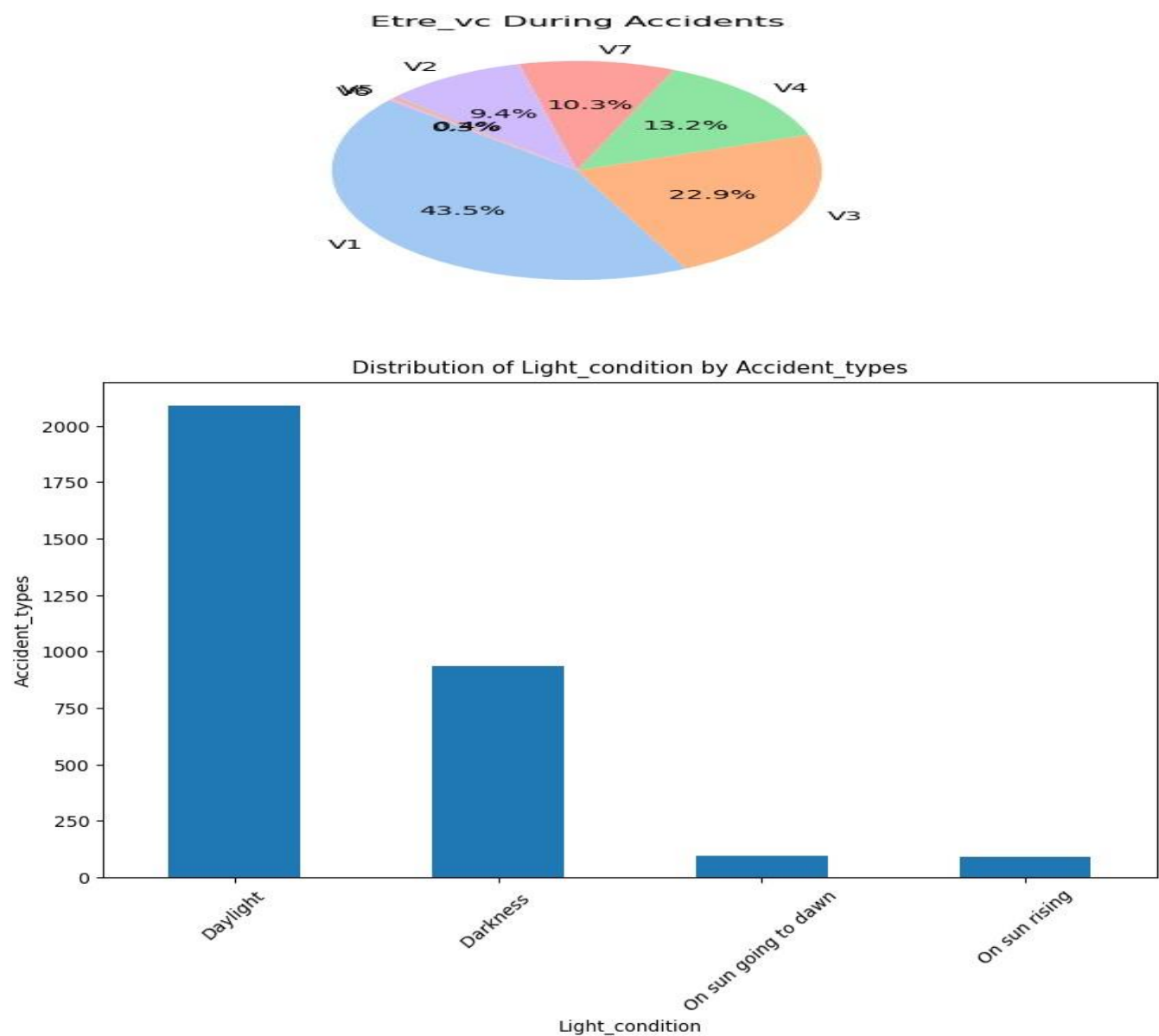
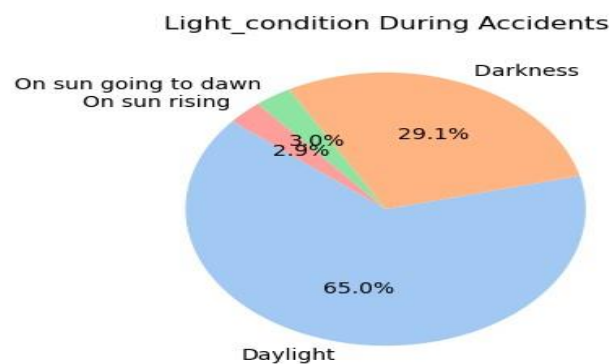


Figure 8. Distribution of Light condition by Accident Types

The graph depicting the distribution of light conditions in traffic accidents reveals distinct patterns. Daylight conditions account for the majority of accidents, suggesting that even during well-lit hours, high traffic volumes and other factors contribute significantly to incidents. Darkness ranks second, highlighting the increased risks associated with nighttime driving, such as reduced visibility and driver fatigue. Accidents occurring at dawn show a notable presence, indicating that the transition from darkness to light can pose challenges for drivers, possibly due to changing light conditions and lower driver alertness. Finally, a smaller yet significant portion of accidents occurs during sunrise, when the glare from the rising sun can impair visibility. This distribution underscores the importance of addressing safety measures across different light conditions, emphasizing the need for heightened caution and appropriate lighting on roadways during night and transitional periods. Over all graph shows the distribution of accident severity by light condition in Addis Adama Expressway during the past eight years. The highest proportion of severe accidents (65%) occurs during the Daylight. This is followed by darkness (29.1%), on sun going to dawn (3%), and on sun rising (2.9%).



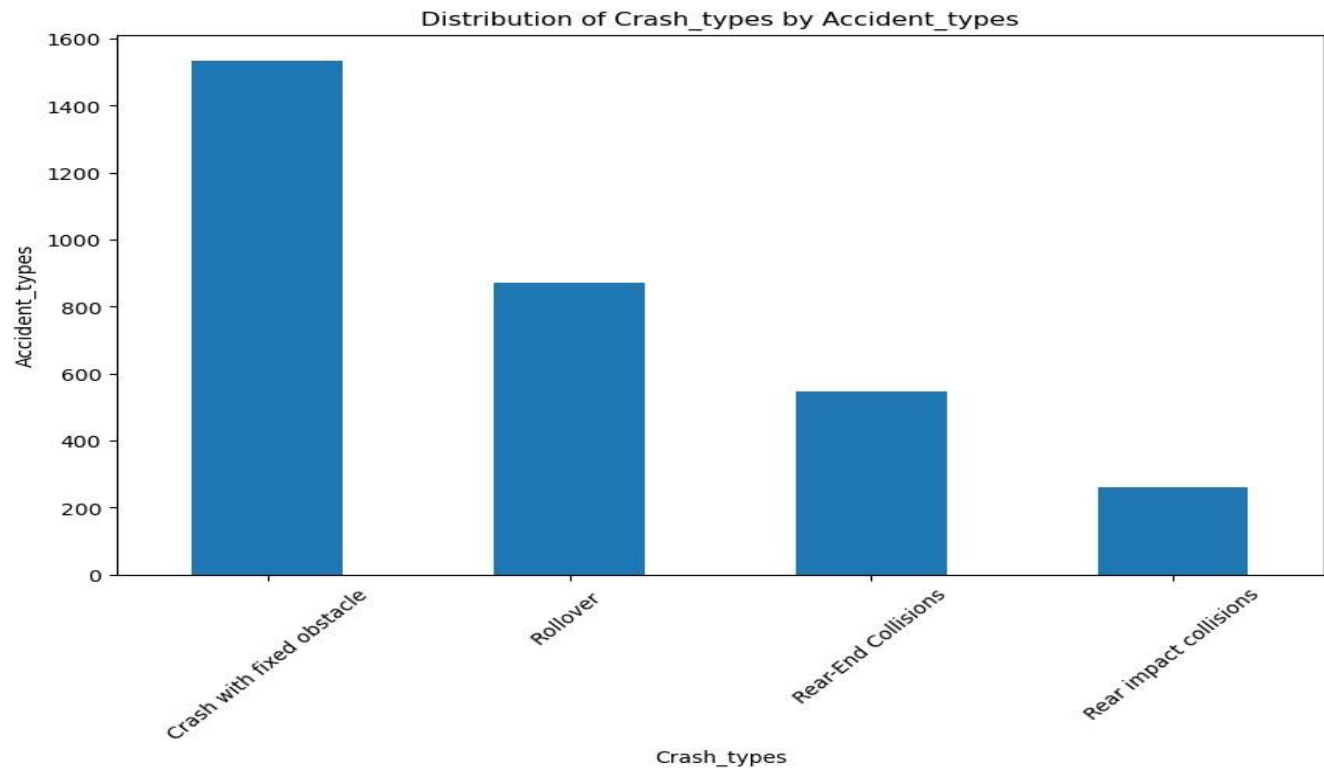


Figure 9. Distribution of Crash Types by Accident types

The graph illustrating the distribution of crash types in the traffic accident dataset reveals several key insights. Crashes with fixed obstacles, such as hitting guardrails, trees, or other stationary objects, are the most common, indicating frequent loss of vehicle control or navigation errors. Rollover accidents come second, highlighting the severity of incidents where vehicles flip over, often due to high speeds, sharp turns, or collisions. Rear-end collisions rank third, showing a significant number of incidents where one vehicle hits the back of another, typically caused by following too closely or sudden braking. Rear impact collisions, where a vehicle is struck from behind, are the fourth most common type, emphasizing similar risk factors as rear-end collisions but with the vehicle being stationary or moving slowly. This distribution underscores the need for targeted safety measures to address the prevalent types of crashes, particularly those involving fixed obstacles and rear-end impacts, through improved road design, driver education, and enforcement of safe driving practices. over all, shows the distribution of accident severity by crash type in Addis Adama Expressway during the past Eight years. The highest proportion of severe accidents (47.8%) involves Crash with fixed obstacles. This is followed by Rollover (27.1%), Rear-end collision (17%), and Rear impact collision (8.1%).

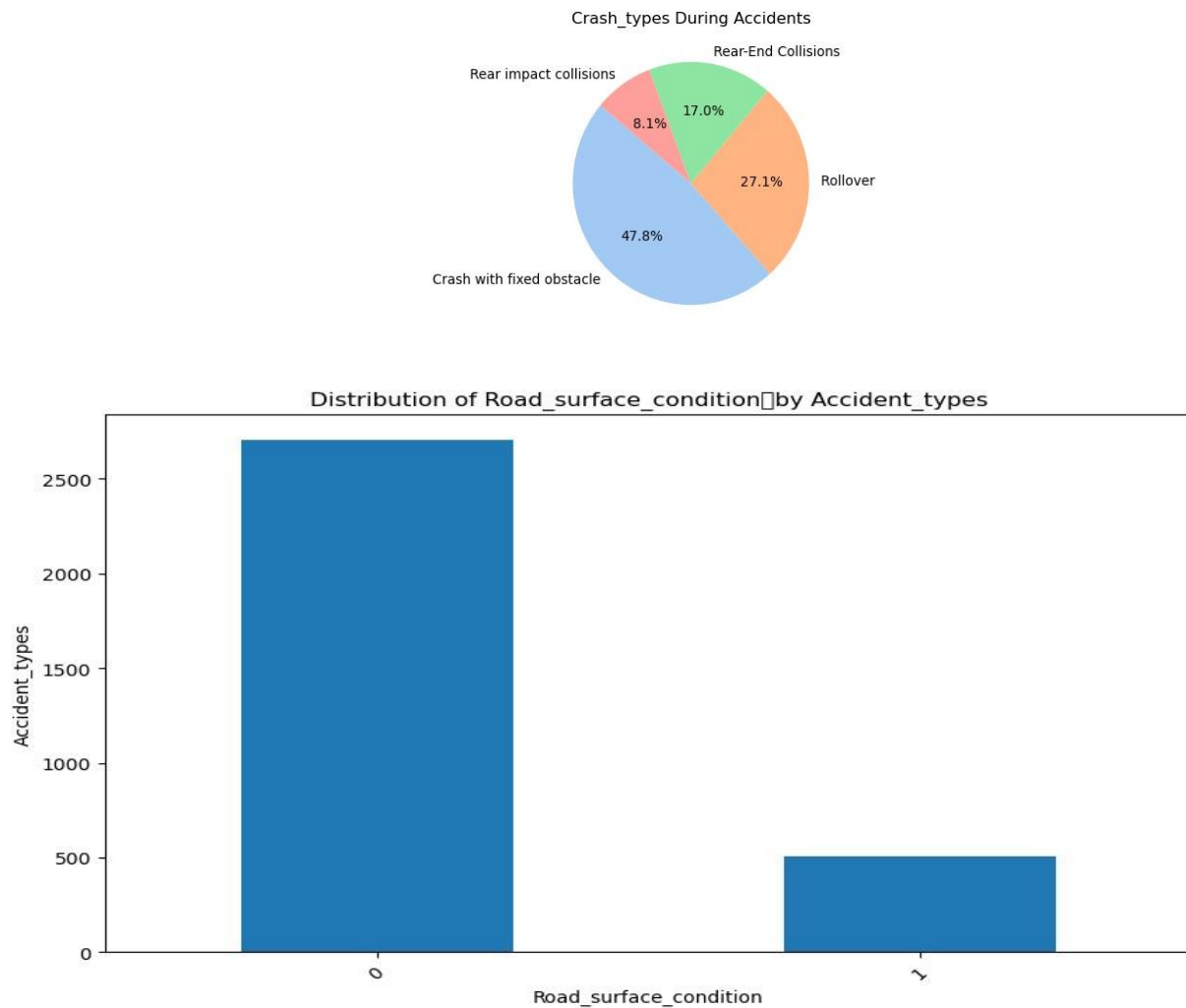


Figure 10. Distribution of Road Surface Condition Accident Types

The graph illustrating the distribution of road surface conditions in traffic accidents shows a clear distinction between dry and wet surfaces. The majority of accidents occur on dry road surfaces, indicating that even in optimal driving conditions, factors such as driver error, vehicle malfunction, or traffic density contribute significantly to accidents. Wet road surfaces, while less frequent, still account for a substantial portion of the accidents, underscoring the increased risks posed by reduced traction and visibility during rain or after precipitation. This distribution suggests that while dry conditions see more accidents due to their commonality, wet conditions present a heightened danger that necessitates careful driving and robust road maintenance to mitigate risks. The above graph shows the distribution of accident severity by Road surface condition in Addis Adama Expressway during the past Eight years. The highest proportion of severe accidents (47.8%) involves on dry road surface condition. This is followed by Wet road surface condition (27.1%),

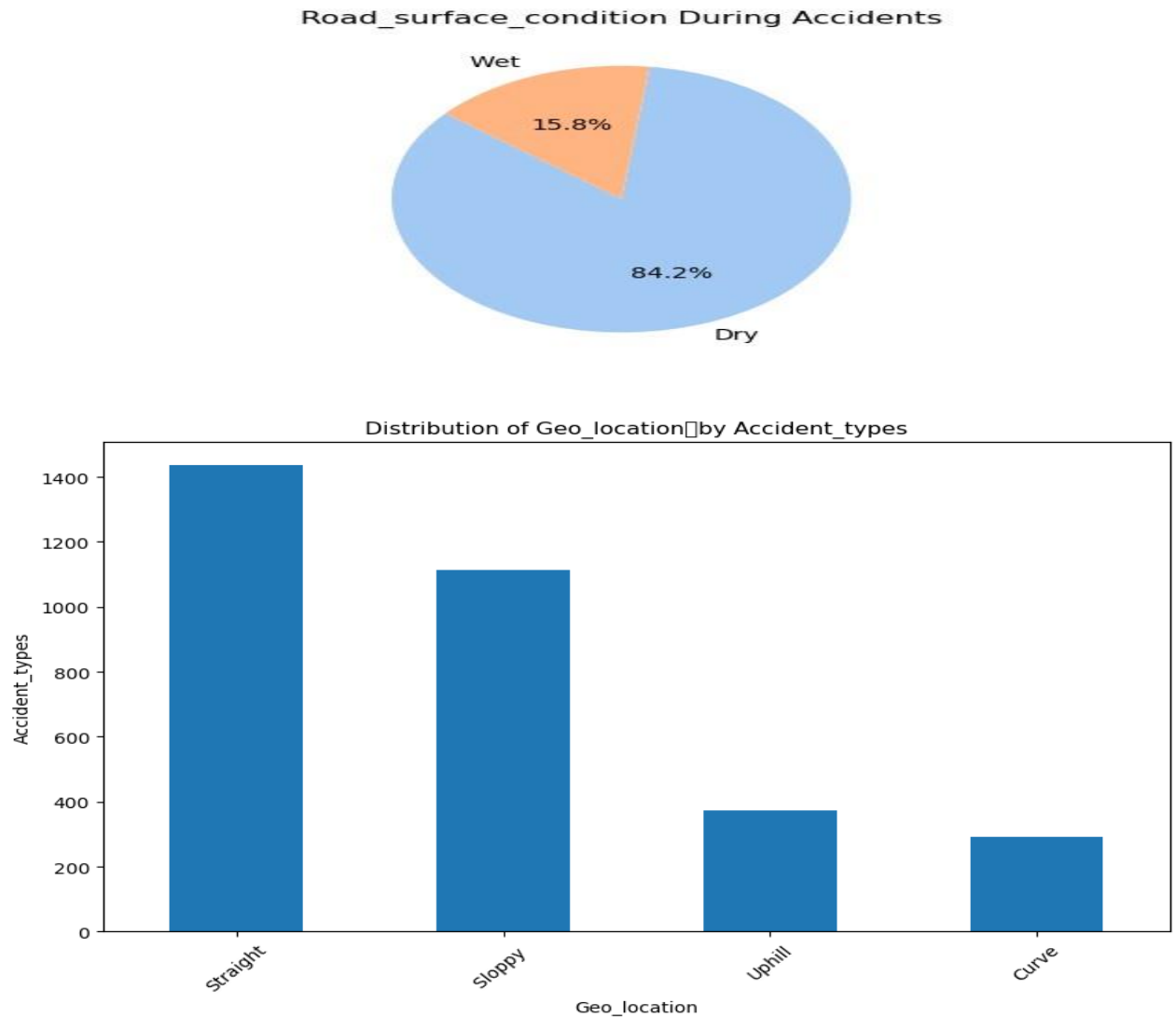


Figure 11. Distribution of Geo-Location Accident Types

The graph depicting the distribution of geometric locations in traffic accidents reveals distinct patterns across different types of road configurations. Straight sections of roads are the most common location for accidents, indicating that a significant proportion of incidents occur on relatively straightforward stretches of road. Slopes or hilly areas follow as the second most prevalent location, suggesting that changes in elevation may contribute to accident risk, potentially due to reduced visibility or driver error. Accidents on uphill sections rank third, indicating that ascending gradients pose challenges for drivers, especially in terms of vehicle control and navigation. Curves in the road come fourth, indicating that sharp bends or turns present increased risks for accidents, potentially due to issues such as excessive speed or improper lane positioning. This distribution underscores the need for targeted safety measures across various geometric locations, emphasizing the importance of road design, signage, and driver awareness to mitigate the risks associated with different types of road configurations. The above graph

shows the distribution of accident severity by Geometric Location in Addis Adama Expressway during the past Eight years. The highest proportion of severe accidents (44.7%) involves Straight Geometric Location. This is followed by Sloppy (34.6%), Uphill (11.6%), and Rear impact collision (9.1%).

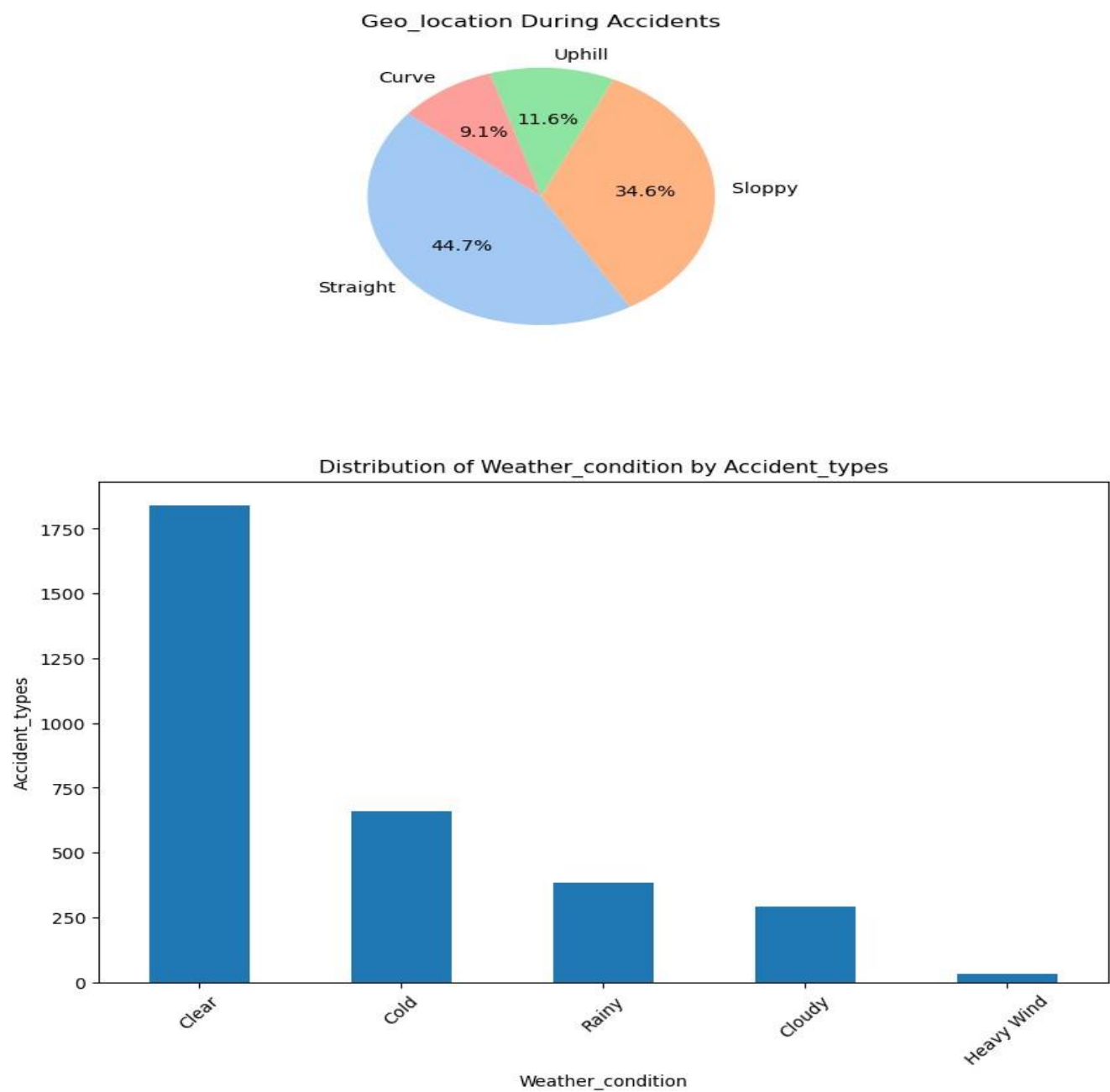
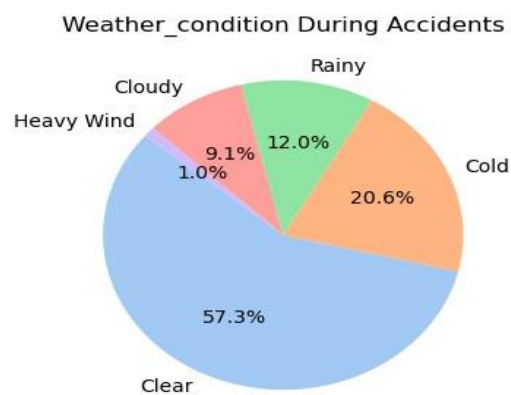


Figure 12. Distribution of Weather Condition by Accident Types

The graph illustrates the distribution of weather conditions by accident types, presented in the following order: clear, cold, rainy, cloudy, and heavy wind. Clear weather accounts for the highest proportion of accidents, indicating that a significant number of accidents occur under normal, unobstructed conditions. Cold weather conditions come next, suggesting a notable increase in accidents possibly due to ice or

reduced vehicle performance. Rainy conditions follow, highlighting the risk posed by wet and slippery roads. Cloudy weather sees a moderate number of accidents, potentially due to reduced visibility and varying light conditions. Lastly, heavy wind conditions contribute the least to the total number of accidents, though this still underscores the danger posed by high wind speeds affecting vehicle control. Overall, the graph underscores the significant impact of adverse weather conditions on accident rates, with clear weather surprisingly having the highest accident occurrence.



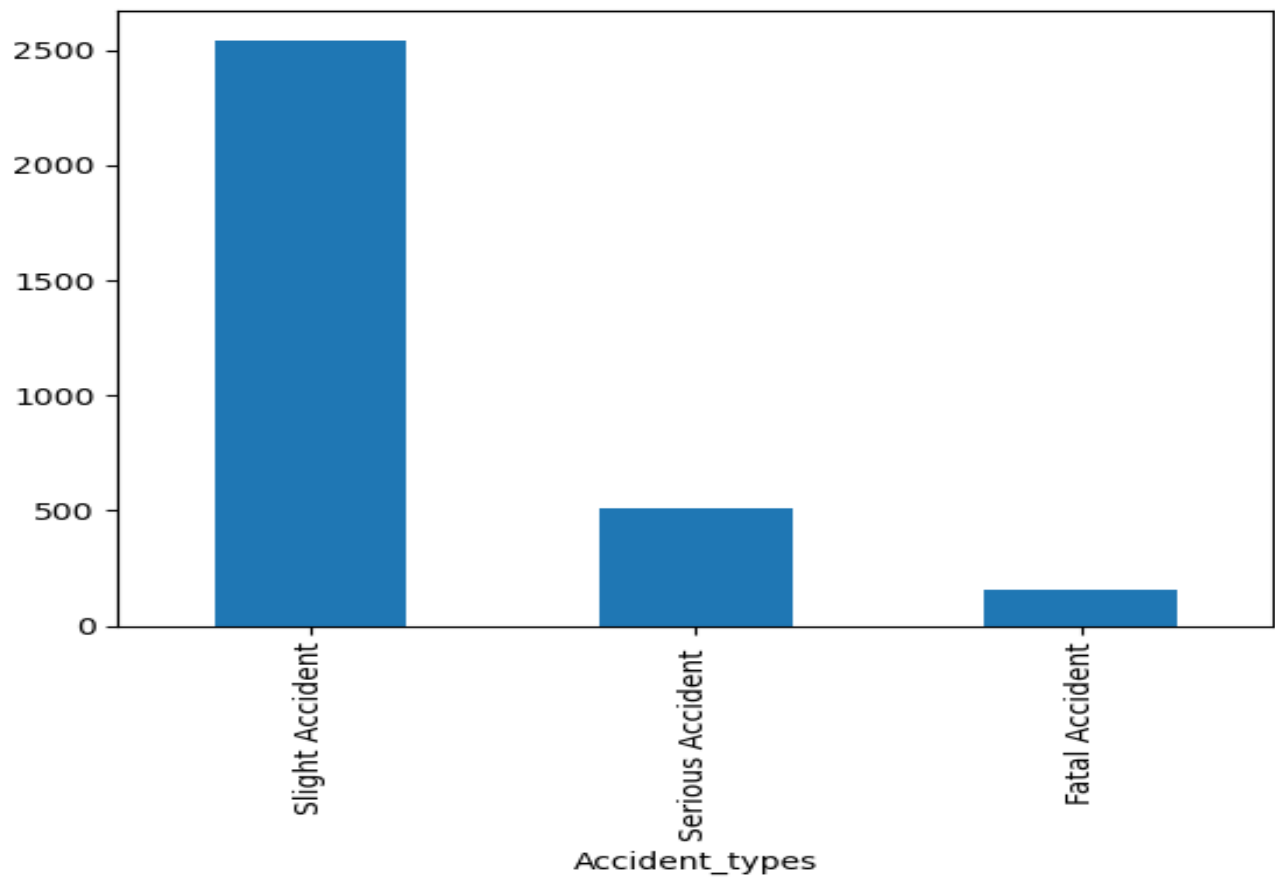
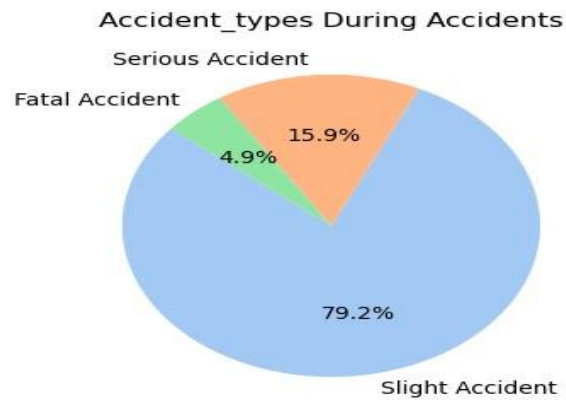


Figure 13. Distribution of Accident Types



The graph illustrates the distribution of traffic accident types, categorizing them into fatal accidents, serious accidents, and slight accidents. It reveals critical insights into the prevalence and severity of traffic incidents. Fatal accidents, representing the most severe category, are relatively infrequent but underscore the grave consequences of the most dangerous collisions, often involving high speeds or severe impacts. Serious accidents, while more common than fatal accidents, result in significant injuries and highlight

the need for improved road safety measures and emergency response. Slight accidents, the most frequent category, involve minor injuries or property damage, reflecting the everyday risks faced by drivers and pedestrians. This distribution emphasizes the spectrum of traffic accident severities and the importance of targeted safety interventions to reduce both the frequency and severity of traffic accidents. Understanding this breakdown can help policymakers, traffic safety expert’s strategies to enhance road safety and mitigate the impacts of traffic incidents across all severity levels. The graph shows the distribution of accident severity in Addis Adama Expressway during the past eight years. The majority of accidents (79.2%) Slight accident, this is followed by Serious Accident (25%) and Fatal accident (10%).

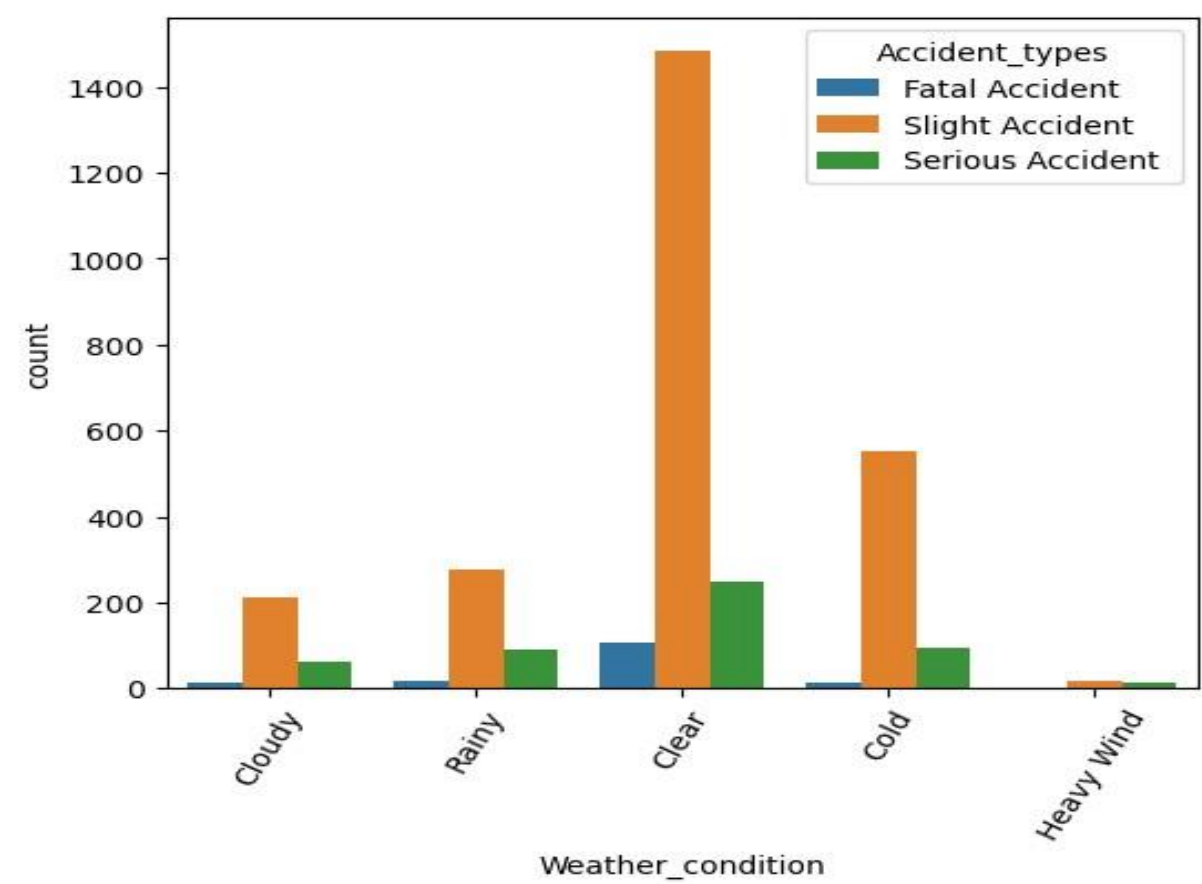


Figure 14. Distribution of Accident Types by Weather Condition

The graph illustrates the distribution of traffic accident types fatal, serious, and slight across various weather conditions. The data reveals significant correlations between weather conditions and the accidents Types. Fatal accidents are notably more prevalent during adverse weather conditions, such as heavy rain and clear, which likely contribute to reduced visibility and over speed, increasing the risk of severe collisions. Serious accidents also show higher frequencies under these challenging weather

conditions, indicating the heightened dangers drivers face when weather impacts road safety. Slight accidents, while occurring under all weather conditions, are more frequent during clear weather, possibly due to the higher volume of traffic and driving activity. This distribution underscores the critical impact of weather on road safety, emphasizing the need for improved driving precautions and safety measures during adverse weather conditions. By understanding these patterns, traffic safety authorities can implement targeted interventions, such as public awareness campaigns and enhanced road maintenance, to mitigate the risks associated with different weather conditions and improve overall road safety. Overall the graph shows the relationship between weather condition and accident severity in Addis Adama expressway during the past Eight years. The highest proportion of accidents Types occurs during Clear conditions. This is followed by Rainy conditions, cold conditions, and cloudy.

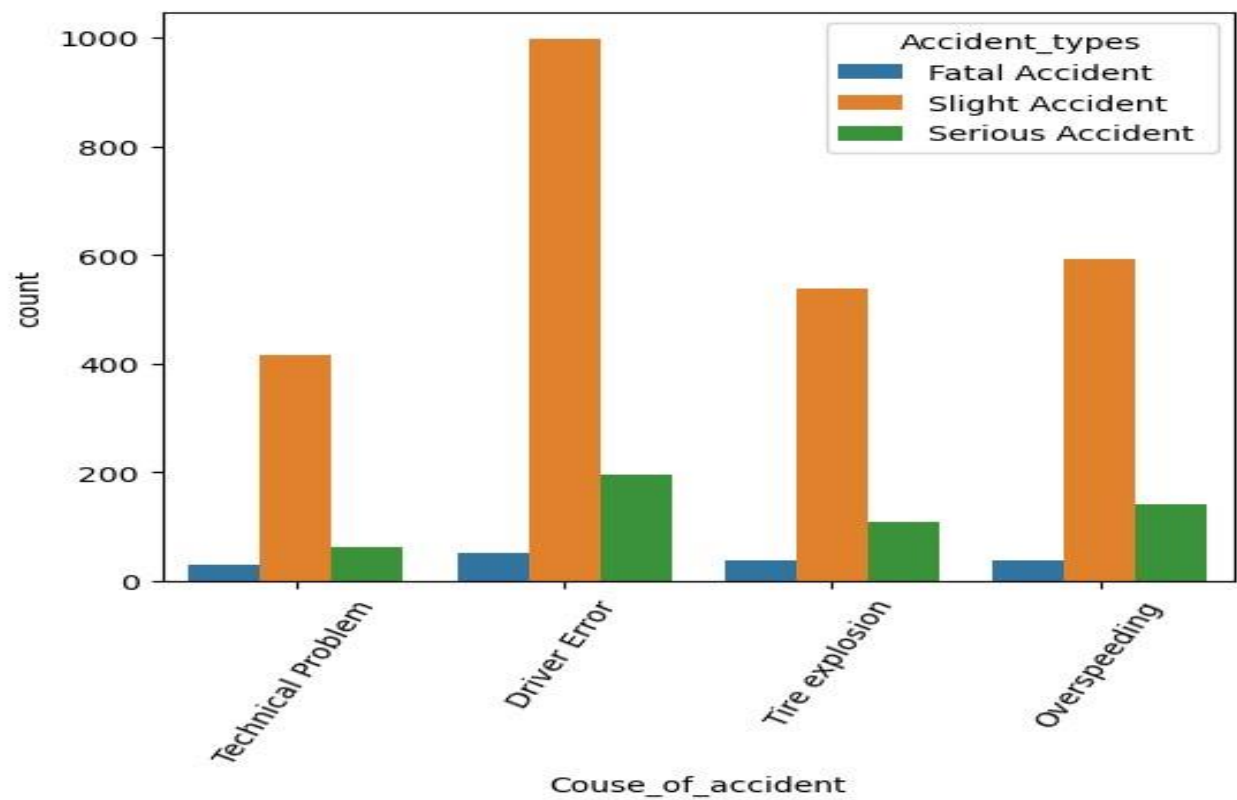


Figure 15. Distribution of Accident Types by Weather Condition

The graph shows the number of fatal accidents, serious accidents, and slight accidents that occurred due to each cause of accident. The x-axis of the graph represents the cause of accident, and the y-axis represents the number of accidents. The graph shows that the most common cause of fatal accidents is driver error, followed by speeding and Tire explosion. The most common cause of serious accidents is also driver error, followed by speeding and Technical problem. The most common cause of slight

accidents is driver error, followed by speeding and Tire explosion. Overall, the graph shows that driver error is the leading cause of traffic accidents of all traffic accident Types. This highlights the importance of safe driving practices and the need for continued efforts to reduce the number of accidents caused by driver error.

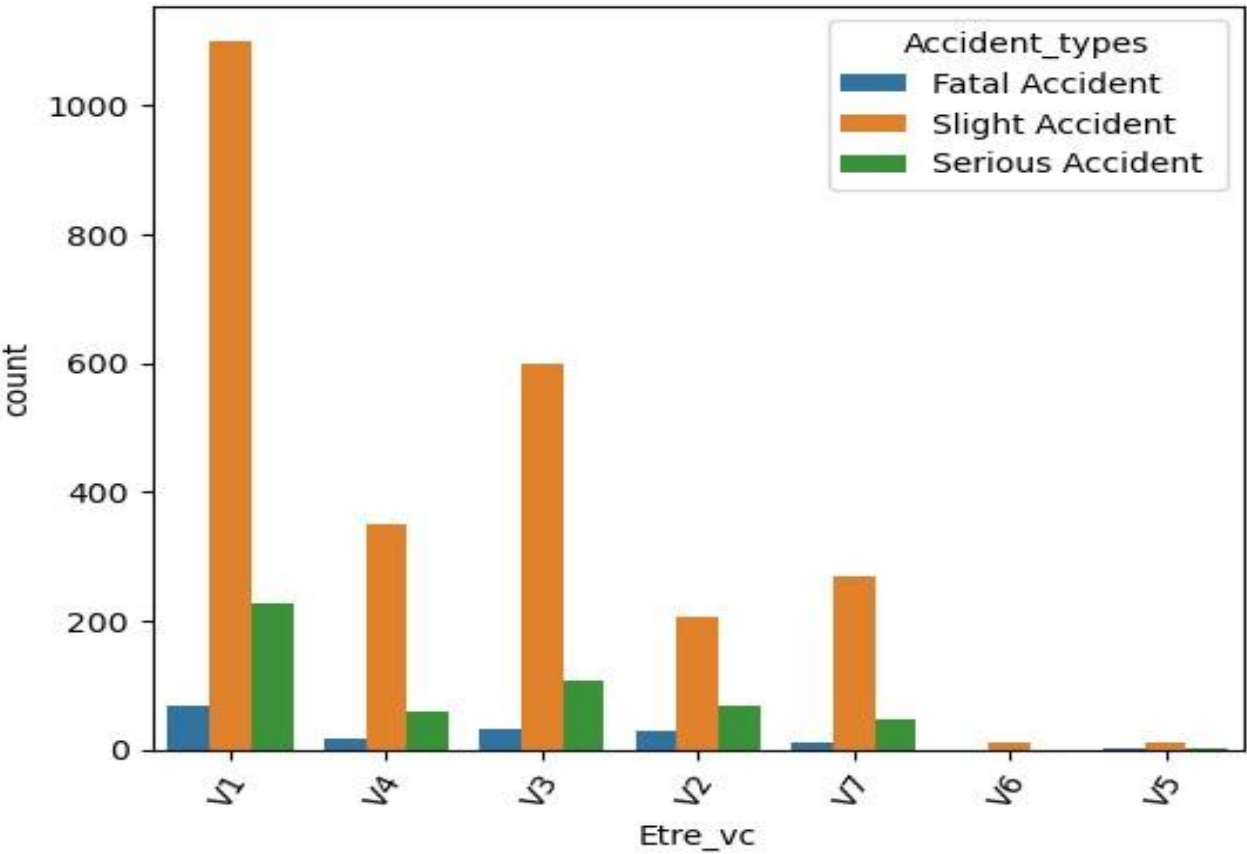


Figure 16. Distribution of Accident Types by Vehicle Type

The graph provides a visual representation of traffic accident types fatal, serious, and slight categorized by vehicle types. It clearly illustrates the distribution and frequency of each accident type across different vehicle categories. For instance, fatal accidents are predominantly associated with small vehicles such as automobile, pick up a higher risk of severe outcomes in collisions involving these vehicles and also serious and slight accidents are more evenly distributed among smaller vehicles, including cars and heavy truck indicating that while these vehicles are involved in numerous incidents, the severity is generally lower. The graph highlights the critical need for targeted safety measures and regulations for different vehicle types to reduce the occurrence and severity of traffic accidents. The above graph shows the relationship between Ethiopian Toll roads vehicle Classification and accident Types in Addis Adama expressway during the past Eight years. The highest proportion of severe accidents occurs by vehicle

one(V1). This is followed by Vehicle type three(V3), Vehicle type four(V4), Vehicle type two(V2), Vehicle type Six(V6) and Vehicle Type five(V7).

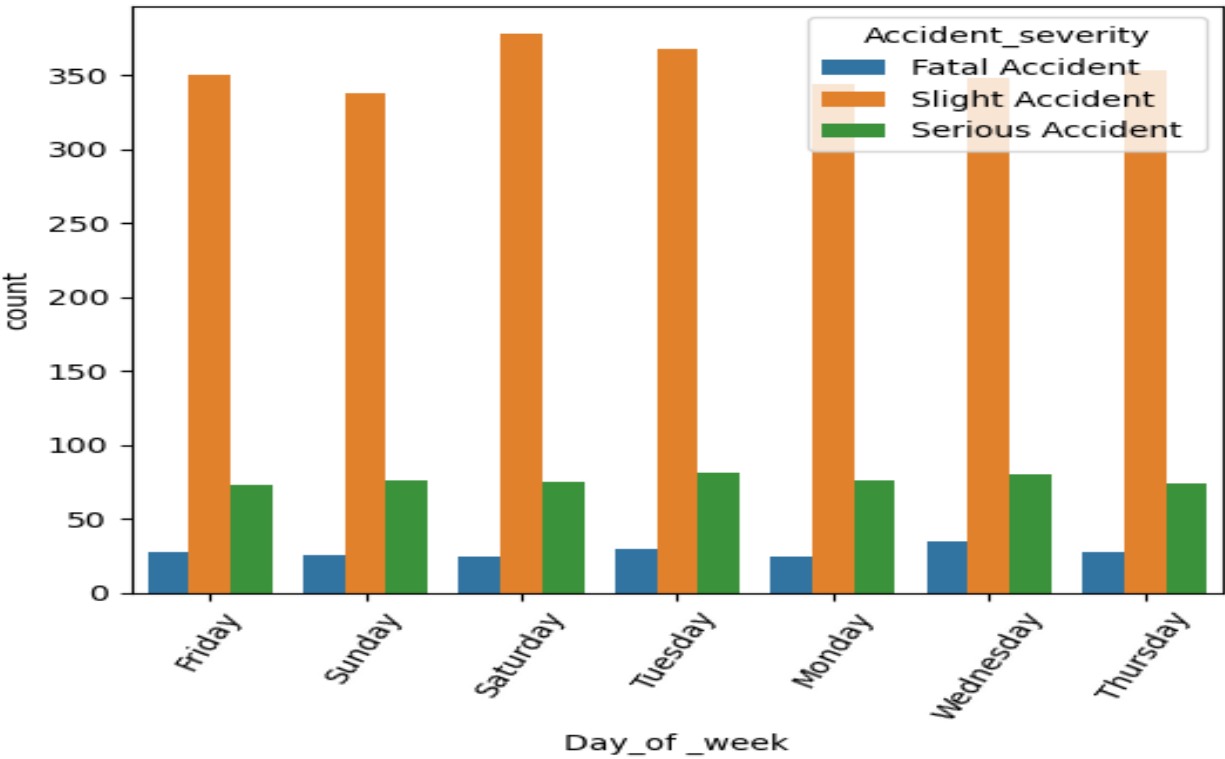


Figure 17. Distribution of Accident Types by Weather Condition

The graph depicting the distribution of accident types by day of the week highlights notable trends in the severity of incidents. Slight accidents consistently dominate across all days of the week, indicating a higher frequency of minor incidents such as minor collisions. Serious accidents come next, displaying a significant but lesser occurrence compared to slight accidents, suggesting a notable presence of incidents requiring medical attention. Fatal accidents represent the smallest proportion, occurring infrequently across the week, indicating the least common yet most severe incidents resulting in loss of life. This distribution underscores the importance of addressing road safety measures to mitigate the occurrence of all accident types, with a particular focus on reducing the severity of incidents and preventing fatal accidents through effective enforcement, education, and infrastructure improvements and the distribution of accident Types by day of the week in Addis Adama Expressway during the past eight years. The highest proportion of severe accidents occurs on Tuesday, followed by Saturday and The lowest proportion of severe accidents occurs on Sunday. Over all the graph shows the relationship between Days of the weeks and accident Types in Addis Adama expressway during the past Eight years. The highest proportion of accidents Types occurs on Saturday. This is followed by Tuesday, Wednesday, Thursday, Friday, Monday and Sunday.

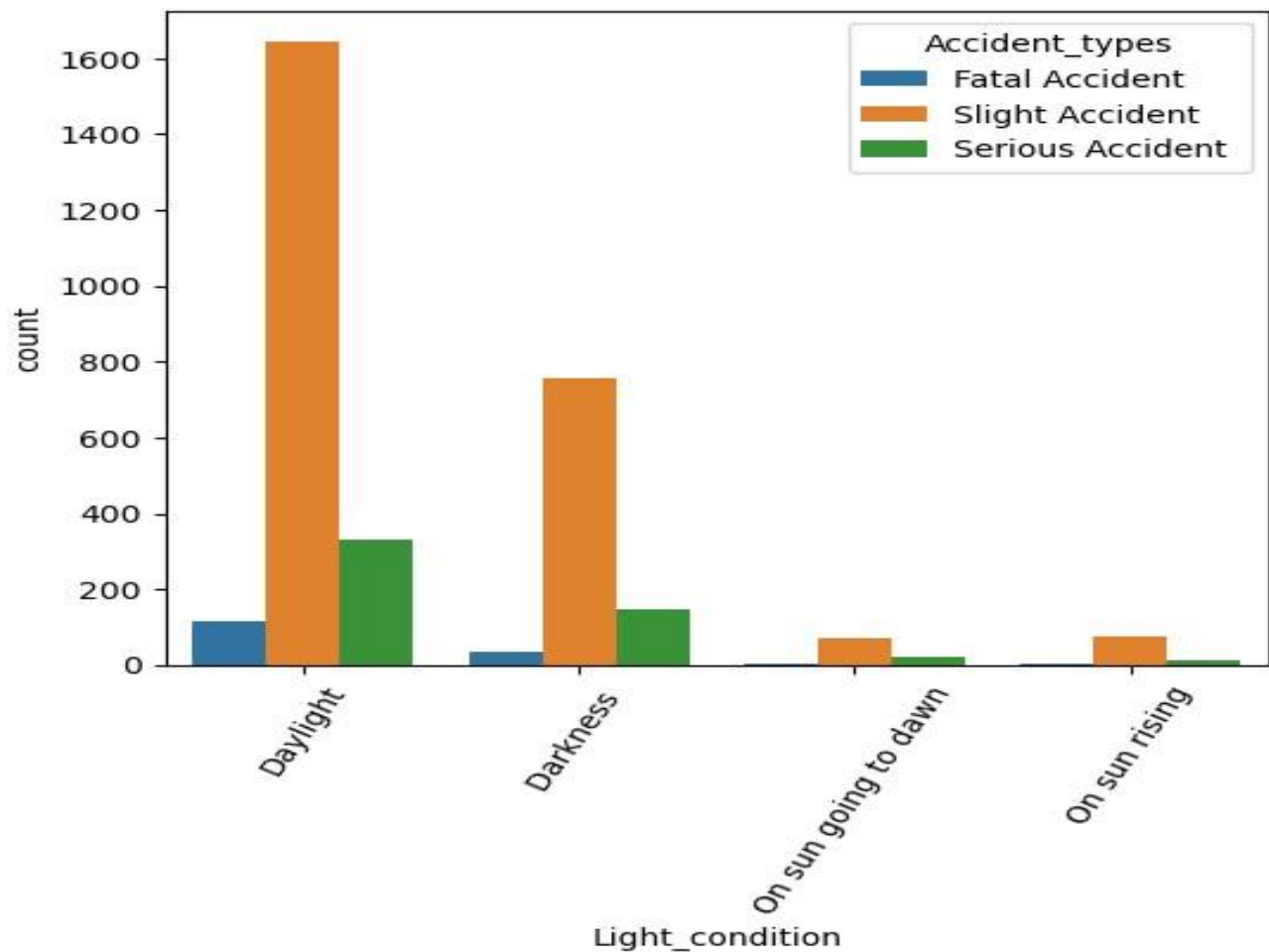


Figure 18. Distribution of Accident Types by Light Condition

The graph illustrating the distribution of accident types by light conditions provides valuable insights into the severity of incidents under varying levels of illumination. Across all light conditions, daylight stands out as the period with the highest occurrence of accidents, regardless of severity. Slight accidents, serious accidents, and fatal accidents are all more prevalent during daylight hours, indicating that a significant portion of incidents, regardless of severity, occur when visibility is optimal. Similarly, accidents are also relatively common during darkness, with slight, serious, and fatal accidents all showing notable frequencies. However, accidents during periods of sunrise and dawn have the lowest occurrence across all severity levels, suggesting that the transitioning light conditions may contribute to heightened driver awareness or reduced traffic volumes. This distribution underscores the need for continued attention to road safety measures during both day and night, with a particular emphasis on addressing

factors that contribute to accidents during periods of low visibility. Over all the graph shows the relationship between Light condition and accident types in Addis Adama expressway during the past Eight years. The highest proportion of accidents types occurs on Daylight. This is followed by Darkness, on sun going down and on sun Going dawn.

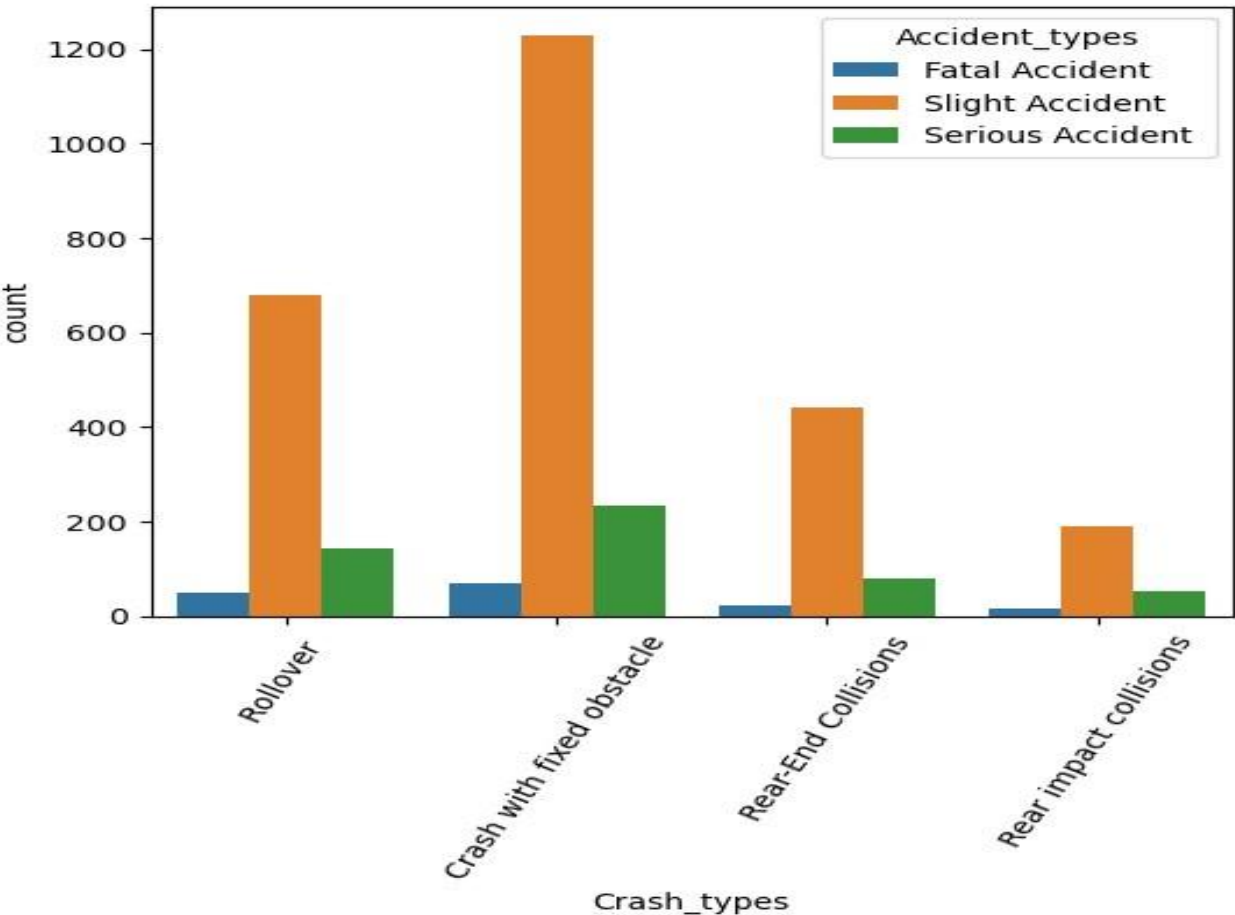


Figure 19. Distribution of Accident Types by Crash Type

The graph depicts the distribution of traffic accident types fatal, serious, and slight across various crash types. It highlights how different kinds of collisions contribute to the severity of accidents. For example, crash with fixed obstacle and rollover crashes are shown to result in a higher proportion of fatal and serious accidents compared to other crash types, indicating the severe impact and danger associated with these types of crashes. Conversely, rear-end collisions and Rear impact crashes tend to result in a smaller number of slight accidents, suggesting that while these crashes are frequent, they generally result in less severe injuries. Intersection-related crashes, another prominent category, exhibit a mixed distribution, with a significant number of both serious and slight accidents. This visualization underscores the

importance of tailored safety measures and interventions for specific crash types to reduce the severity and frequency of traffic accidents. By understanding the nuances of how different crash types affect accident outcomes, road safety authorities can better target their efforts to enhance overall traffic safety

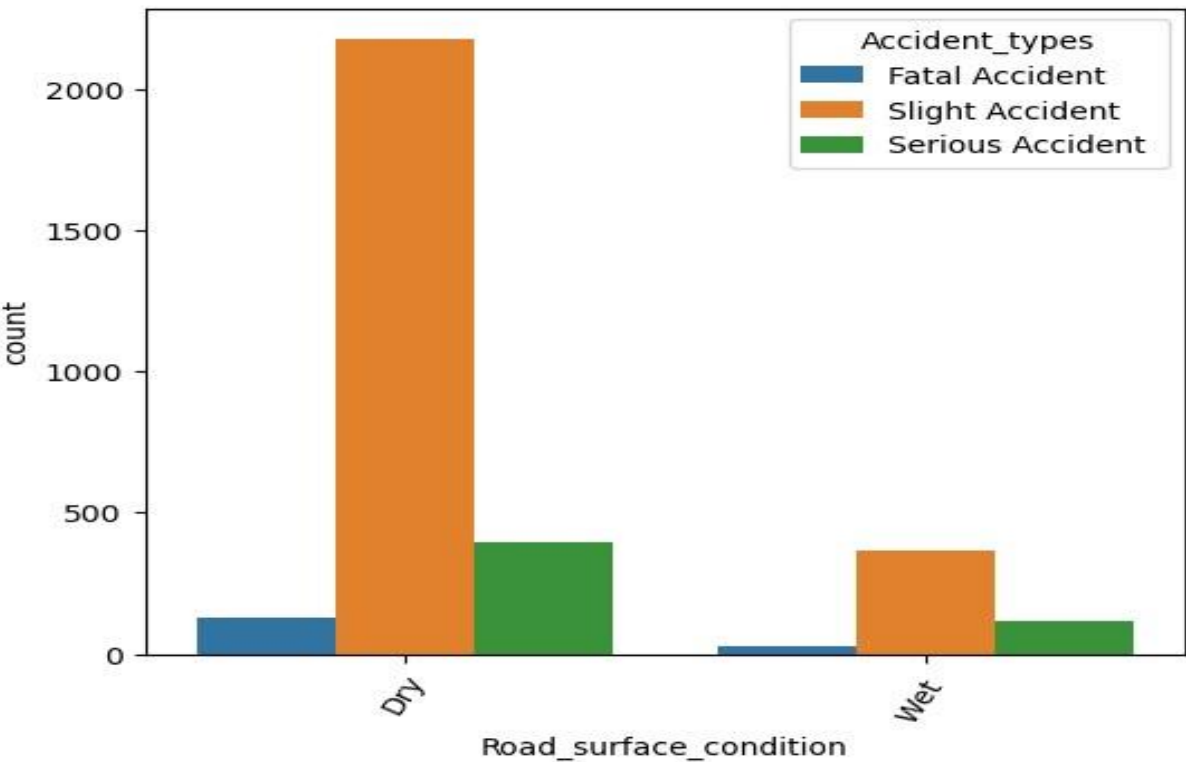


Figure 20. Distribution of Accident Types by Road Surface Condition

The graph illustrates the distribution of traffic accident types fatal, serious, and slight across various road surface conditions. The data reveals a clear relationship between the condition of the road surface and the severity of accidents. The graph depicting the distribution of accident types by road surface condition reveals significant disparities in the severity of incidents between wet and dry conditions. Across both conditions, accidents are more prevalent on dry surfaces, with slight, serious, and fatal accidents all exhibiting higher frequencies compared to wet conditions. This suggests that a substantial proportion of accidents occur when road surfaces are dry, potentially due to factors such as increased speeds or decreased stopping distances. Conversely, while accidents still occur on wet surfaces, the frequencies of serious and fatal accidents are notably lower, indicating a reduced severity of incidents under these conditions. This disparity underscores the importance of adjusting driving behavior and maintaining heightened caution, particularly during dry conditions, to mitigate the risks associated with accidents of varying severity on wet surfaces.

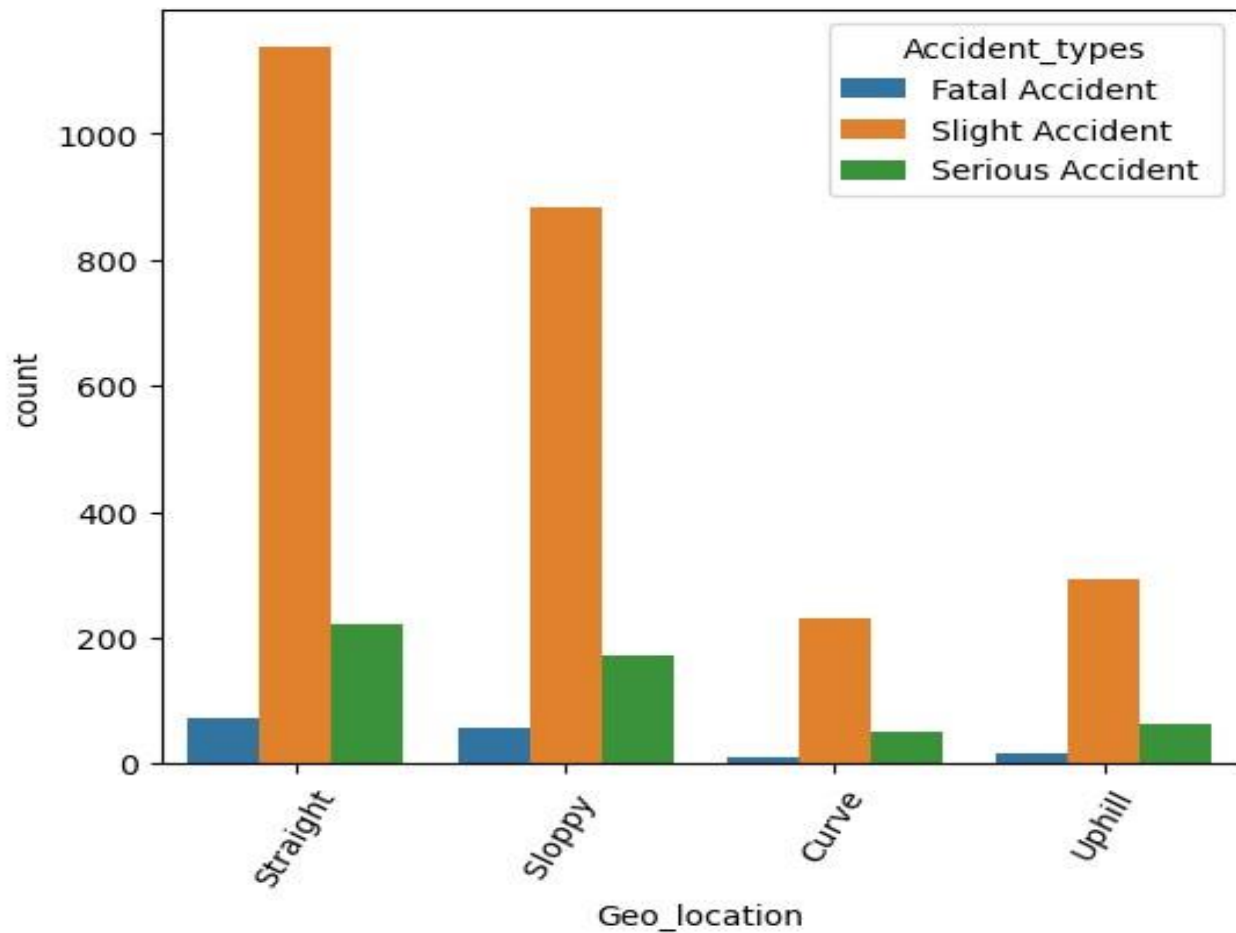


Figure 21. Distribution of Accident Types by Geo-Location

The graph illustrating the distribution of accident types by geometric location provides valuable insights into the severity of incidents across different road configurations. Accidents occur most frequently on straight and sloppy sections of roads, with higher frequencies of slight, serious, and fatal accidents observed in these locations. This suggests that accidents are more prevalent on relatively flat and straightforward stretches of road, possibly due to factors such as higher speeds or driver inattention. Conversely, accidents are less common on uphill sections and curves, with lower frequencies of all accident types recorded in these areas. This disparity indicates that road features such as inclines and bends may contribute to a reduced likelihood of accidents or lessen their severity. Understanding these patterns can inform targeted safety measures and infrastructure improvements aimed at mitigating the occurrence of accidents, particularly on straight and sloppy sections where incidents are more frequent.

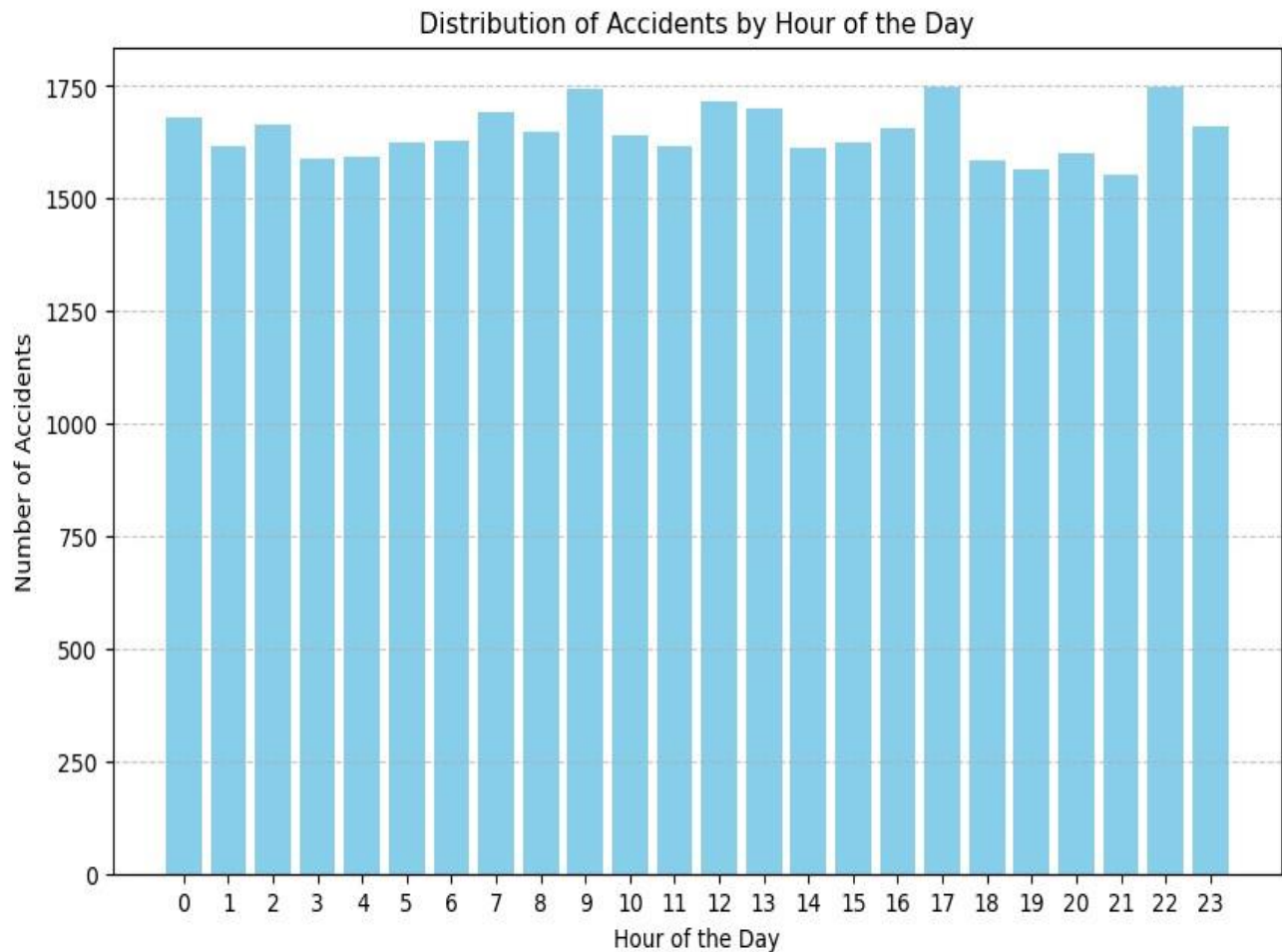


Figure 22. Distribution of Accident Types by Hour

The graph shows the distribution of traffic accidents by the hour of the day, providing insights into when accidents are most likely to occur. The data reveals distinct peaks and troughs in accident frequency throughout the 24-hour period. There is a noticeable increase in accidents during the daylight hour, typically between 1 and 6. A second significant peak is observed in the late afternoon and early evening, from approximately 8 to 12 AM. Additionally, there is a marked rise in accidents during the late night hours, particularly around 17 to 22 and 23 which may be attributed to factors such as reduced visibility, fatigue, and impaired driving. The lowest accident frequencies occur during the early morning hours, between 11 and 1 when traffic volumes are generally at their lowest. This distribution highlights the importance of targeted road safety measures during peak traffic times and late night hours to reduce the occurrence of traffic accidents and enhance overall road safety. Over all the graph shows visual representation of the distribution of accident types across different hours of the day, providing insights into when accidents are most likely to occur and their severity during different times of the day.

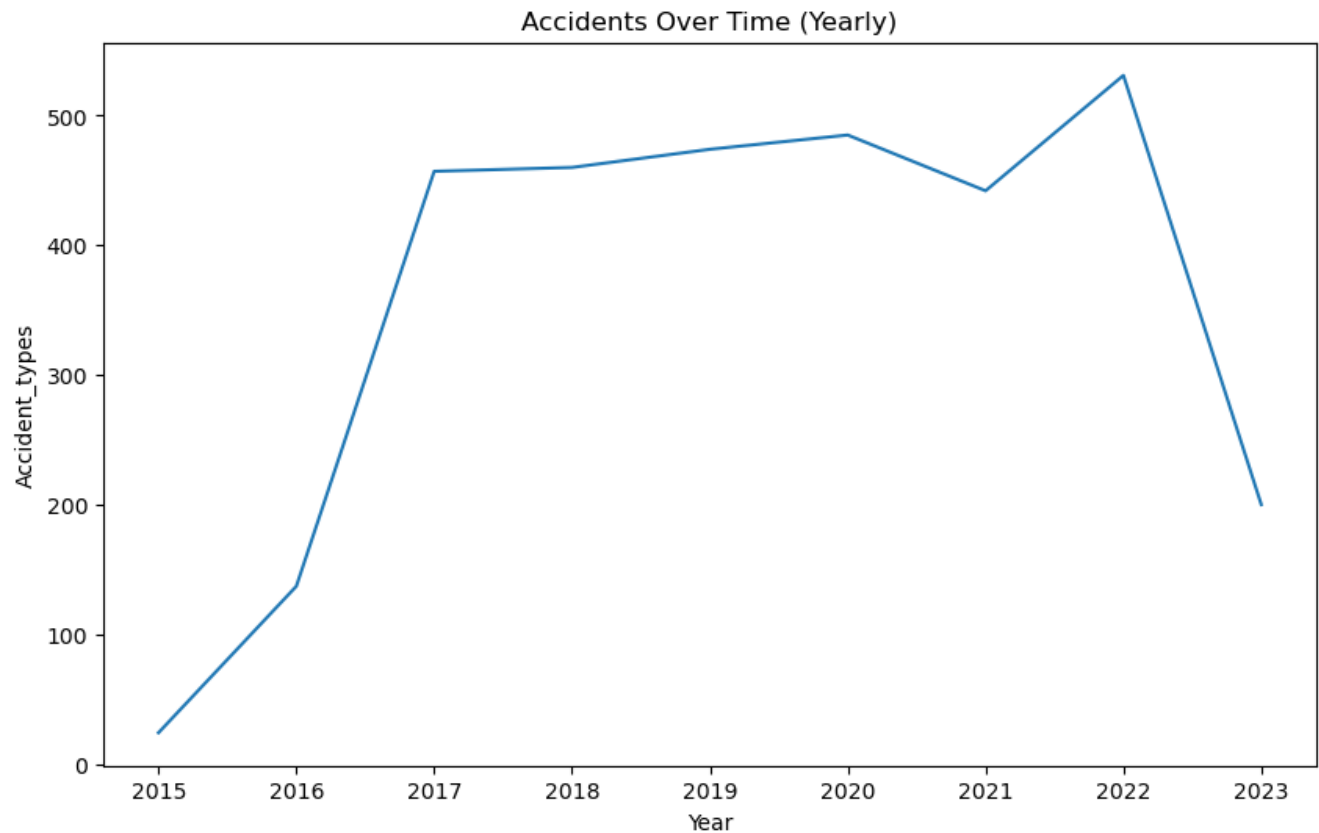


Figure 23. Distribution of Accident Over Time

The line plot explains the trend of traffic accidents over a span of Eight years, providing a visual representation of how accident frequencies have changed over time. The data reveals both short-term fluctuations and long-term trends in the occurrence of traffic accidents. In the earlier years, the graph shows a relatively increasing trend, indicating a consistent number of accidents annually. However, there is a noticeable peak at certain points, suggesting years with significantly higher accident rates, potentially due to factors such as increased traffic flow and weather anomalies. In more recent years, the line plot may show a decline off in the number of accidents, possibly reflecting the impact of improved road safety measures, advancements in vehicle technology, and stricter traffic regulations. This temporal analysis highlights the dynamic nature of traffic accidents and underscores the importance of continuous monitoring and evaluation of road safety initiatives to understand their effectiveness and adapt to emerging challenges. Over all the line plot shows the number of traffic accidents in Addis Adama Expressway over a period of eighth years The overall trend in traffic accidents over the eight-years period is slightly increasing. The number of traffic accidents in 2021 is significantly lower than in other years.

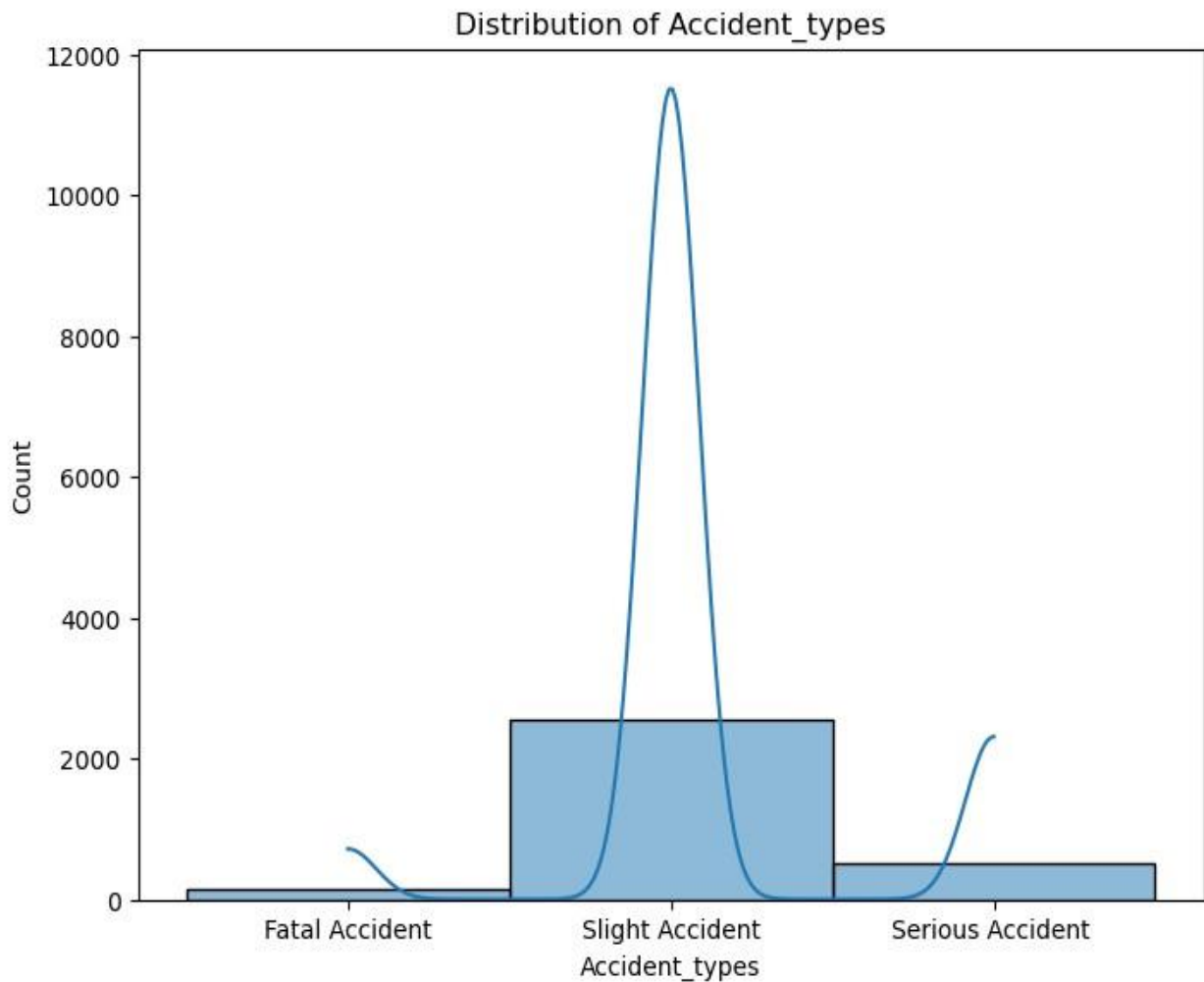


Figure 24. Distribution of Accident Types

The histogram plot displays the distribution of traffic accidents categorized into three types: fatal accidents, serious accidents, and slight accidents. Each bar represents the frequency of accidents falling within a specific range or category of Accident types. The data reveals distinct patterns in the distribution of accident types, with slight accidents exhibiting the highest frequency, followed by serious accidents, and finally, fatal accidents being the least frequent. This distribution reflects the severity gradient of traffic accidents, with slight accidents representing the majority of incidents involving minor injuries or property damage, serious accidents indicating more severe outcomes requiring medical attention, and fatal accidents representing the most severe cases resulting in loss of life. The histogram provides valuable insights into the relative prevalence of different accident types, highlighting the need for targeted safety interventions to mitigate the occurrence and severity of traffic accidents across all severity levels.

4.3. Model Description

The different models are used to train and predict on a test set of the data. This will provide accurate scores with which are used to determine the best fit model for the exercise. The purpose is to Develop Traffic Accident predictive model the accident Types through predictive analytics. Gradient Boosting. This is a machine learning technique which involves the use of regression and classification and forms collections of decision trees which is an ensemble. From these collections, the best tree is picked to derive the results in turn. Random Forest Classifier. Random trees are an ensemble of many decision trees and this is a classification and prediction method that is developed on both classification and regression tree methods. (Pierre, 2006) This training method of the model uses partitioning which is recursive in manner and splits the records with similar output values into segments Decision Tree Classifier. A decision tree model works by recursive partitioning to classify whether the accident type was highly severe, moderate or mild based on past trends. The model can be trained with different cross validations, complexity parameter controls how deep the tree grows, a small value allows for the splitting of even smaller nodes which doesn't improve prediction fit by a significant amount. Support Vector Machine (SVM). a supervised machine learning algorithm used for classification and regression tasks. It is particularly well-suited for binary classification problems, but can also be extended to handle multi-class classification and regression tasks. k-Nearest Neighbors (KNN). algorithm is a simple and intuitive supervised machine learning algorithm used for classification and regression tasks. In KNN, the prediction of a new data point is based on the majority vote (for classification) or averaging (for regression) of the labels or values of its k nearest neighbors in the feature space. In my endeavor to construct a reliable predictive model for traffic accidents, i meticulously followed a standardized process across each machine learning algorithm Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting, and Decision Tree. Initially, I meticulously preprocessed my dataset, addressing missing values, encoding categorical variables, and ensuring data consistency. Subsequently, we partitioned the dataset into training and testing sets to facilitate unbiased model evaluation. Utilizing the training data, I trained each model using appropriate hyper parameters, carefully selected through iterative experimentation and cross-validation techniques to prevent overfitting and maximize predictive performance. Once trained, I evaluated the models' predictive accuracy by employing the testing dataset, comparing their predictions against actual accident occurrences. This rigorous evaluation allowed us to quantitatively assess each model's effectiveness in accurately predicting traffic accidents.

4.3.1 Build a Model

Building a predictive model for traffic accidents using SVM model involves several key steps (Cortes, 1995). Initially, comprehensive data on various factors influencing accidents, such as weather, road conditions, and vehicle types, is collected and preprocessed to ensure its quality and consistency. Relevant features are then selected, and the data is split into training and testing sets. Subsequently, a SVM model is initialized and trained using the training data, with hyper parameters fine-tuned for optimal performance. The model's accuracy and effectiveness are evaluated using the testing data, employing metrics like accuracy and precision. If the model meets satisfactory performance criteria, it can be saved for future use using Python libraries like joblib. Ultimately, the saved model can be deployed for real-time prediction of traffic accidents, aiding in accident prevention and mitigation efforts. Feature importance derived from a SVM model measures the contribution of each feature to the model's predictive power. This importance is primarily calculated using metrics like Gini impurity reduction or information gain, summed across all trees in the forest where the feature is used for splitting, and then normalized. Alternatively, permutation importance can be used by shuffling feature values and observing the drop in model accuracy. SVM feature importance is advantageous as it handles multicollinearity, captures non-linear relationships, and offers intuitive insights into feature relevance.

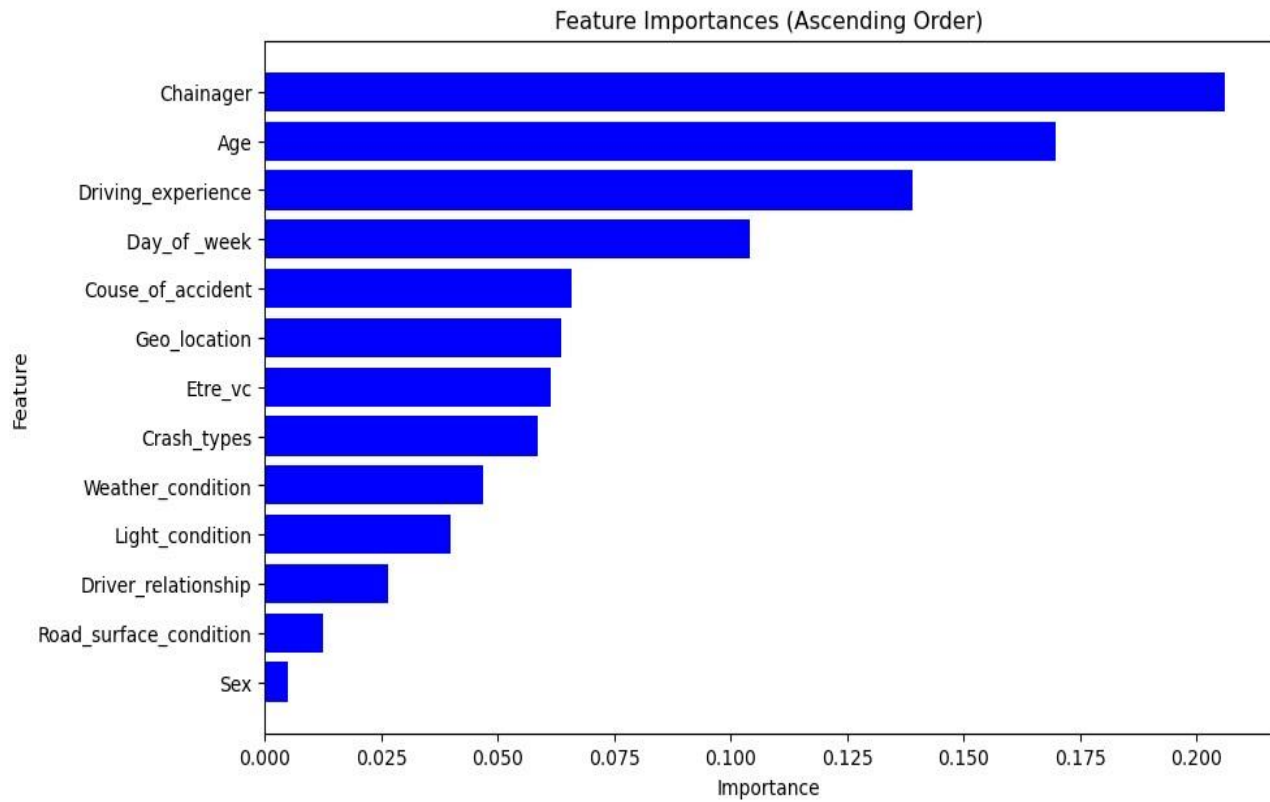


Figure 25. Important Features derived from SVM model

4.3.2 Model Performance Comparison

The performance of various models tells significant differences in their accuracies and classification metrics (Powers, 2011). The Random Forest model achieved a perfect training accuracy of 1.0 but had a test accuracy of 0.7769. Its classification report indicates high precision and recall for class 2, but poor performance for classes 0 and 1, reflected in a macro average f1-score of 0.30. The SVM model showed a training accuracy of 0.7967 and a test accuracy of 0.7811, with excellent performance for class 2 but no correct predictions for classes 0 and 1, resulting in a macro average f1-score of 0.29. The k-NN model had training and test accuracies of 0.8038 and 0.7324, respectively, performing well for class 2 but poorly for classes 0 and 1, with a macro average f1-score of 0.31. The Gradient Boosting model achieved a training accuracy of 0.8136 and a test accuracy of 0.7780, showing strong performance for class 2 but weak for classes 0 and 1, with a macro average f1-score of 0.30. Finally, the Decision Tree model had perfect training accuracy but a lower test accuracy of 0.6483, with a macro average f1-score of 0.36, indicating better but still insufficient performance for classes 0 and 1 compared to the other models. The performance of various models on the dataset reveals that the SVM model achieved the highest test accuracy at 0.7811, indicating it performed the best on unseen data. Following closely, the Gradient Boosting model had a test accuracy of 0.7780, demonstrating strong predictive capabilities. The Random

Forest model also performed well with a test accuracy of 0.7769. The k-NN model had a moderate test accuracy of 0.7324, showing it was less effective than the top three models but still reasonably accurate. The Decision Tree model had the lowest test accuracy at 0.6483, suggesting it struggled the most with generalizing from the training data to new data. These results highlight that while tree-based ensemble methods and SVM provided high accuracy, simpler models like the k-NN and Decision Tree were less effective in this scenario. The training accuracies of the models showcase their performance during the training phase, with the Random Forest and Decision Tree models achieving perfect accuracies of 1.0, implying they were able to precisely fit the training data. However, such high accuracies may also indicate overfitting, where the models memorize the training data instead of learning generalizable patterns, potentially leading to poorer performance on unseen data. Following closely behind, the Gradient Boosting model exhibited a training accuracy of 0.8136, indicating strong learning capabilities. The k-NN and SVM models had slightly lower training accuracies of 0.8038 and 0.7967, respectively, suggesting they were also effective in capturing the underlying patterns in the training data but with some level of error.

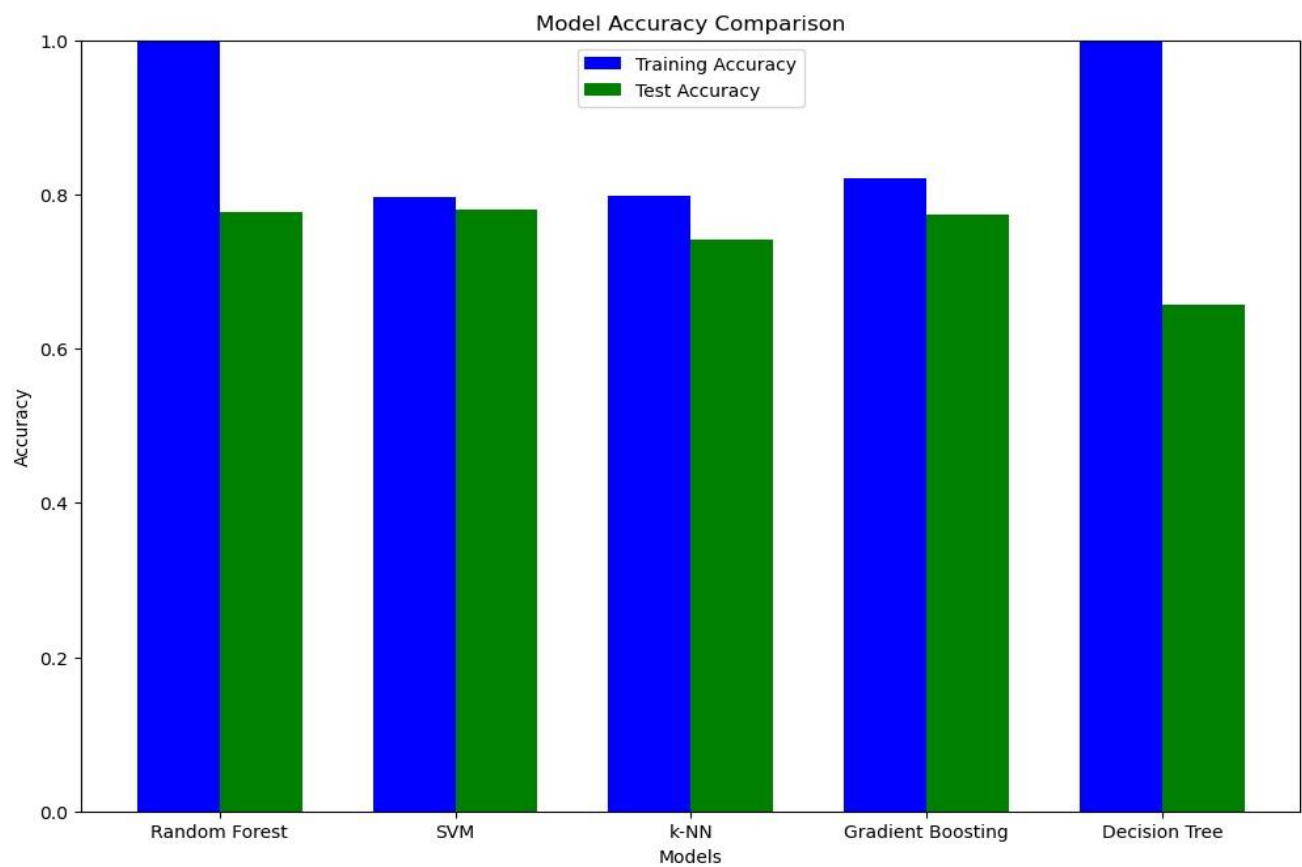


Figure 26. Model Accuracy Comparison

4.4 Results and Discussion

The dataset was initially examined for missing values, outliers, and inconsistencies. Categorical variables were encoded, and feature scaling was applied to normalize numerical features and also irrelevant features drops from the dataset. Features such as Chainage, Driver Age weather conditions, road surface condition, light condition, driver experience, driver age, Cause of Accident and time of day were identified as potential predictors of traffic accidents.

4.4.1 Discussion

In all Leveraging AI for predicting traffic accidents on the Addis-Adama Expressway involves a systematic approach encompassing data collection, preprocessing, feature selection, model selection, training, evaluation, deployment, and continuous improvement (R. Vanitha, 2023). Initially, comprehensive data related to accidents, encompassing factors like Date, Time, Day of the week, Geo-location, weather condition, road surface conditions, and vehicle types, Light condition, Chainage and etc. is gathered from sources Ethiopian Toll Roads Enterprise. Following this, meticulous preprocessing steps, including cleaning, normalization, and feature engineering, are undertaken to ensure data quality. Relevant features, such as Chainage, Driver Age weather conditions, road surface condition, light condition, driver experience, driver age are then selected based on their correlation with accident occurrence. Next, appropriate AI algorithms, such as Random Forest, SVM, decision trees, Gradient Boosting Machines, and K-NN are chosen for building predictive models (Swati, 2023). These models are trained on preprocessed data, learning patterns and relationships between input features and accident likelihood. Evaluation metrics like accuracy and precision are employed to assess model performance on testing datasets. Through this iterative process, AI algorithms play an essential role in enhancing road safety on the Addis-Adama Expressway by accurately predicting and preempting traffic accidents, ultimately contributing to the reduction of accidents and ensuring smoother traffic flow. Overall, Support Vector machine model as the top-performing model for predicting traffic accidents, offering a balance between accuracy and interpretability. The importance of features such as Features such as Chainage, Driver Age, Driver experience, Day of the Weeks, Cause of accident, Geo-location, Vehicle types, Crash Type, weather condition, Driver Relationship, road surface condition and sex underscores the significance factors in influencing accident occurrence. Deploying artificial intelligence for predicting traffic accidents on public roadways in Ethiopia necessitates a thorough consideration of various ethical

dimensions to ensure responsible and equitable utilization of the technology (Royakkers, 2018). Among the paramount ethical considerations are data privacy and security. Predicting traffic accidents on the Addis-Adama Expressway using artificial intelligence algorithms can be an effective approach to enhance road safety.

CHAPTER FIVE

5.1 CONCLUSION

Finally, there is a great deal of potential in the AI-assisted innovation for traffic accident prediction using machine learning models for improving road safety and reducing the enormous amount of traffic accidents. The performance of machine learning models in predicting traffic accidents varies based on the dataset characteristics and model implementation. Support Vector Machines excel in high-dimensional spaces but can be computationally demanding. Random Forests are robust and handle large datasets well, often outperforming other models due to their ability to manage complex interactions and reduce overfitting. k-Nearest Neighbors is simple and interpretable but struggles with large datasets and noise. Gradient Boosting methods, such as Boost, typically provide the highest accuracy and robustness in predictive tasks. Decision Trees are easy to interpret and useful as baseline models but generally less accurate than ensemble methods like RF and GB. By the use of powerful machine learning techniques like Random Forest, SVM, KNN, Gradient Boosting and Decision Tree along with an analysis of historical traffic accident data, the proposed method can accurately predict the probability of traffic accidents and identify Chainage, Driver Age, Driver experience are the most three relevant risk factors from one to three. The SVM model achieved the highest test accuracy at 78.11%, indicating the best performance on unseen data. Close behind, the Gradient Boosting model had a test accuracy of 77.80%, demonstrating strong predictive capabilities. The Random Forest model followed with a test accuracy of 77.69%, showing it also performed well on the test data. The k-NN model had a moderate test accuracy of 73.24%, reflecting its reasonable effectiveness but lower accuracy compared to the top three models. The Decision Tree model had the lowest test accuracy at 64.83%, suggesting it faced the most challenges in generalizing from the training data to new data. These percentages reflect the relative strengths of the models, with SVM leading in test performance, followed closely by Gradient Boosting and Random Forest, while k-NN and Decision Tree showed more moderate results. In all Machine learning models are increasingly leveraged to reduce traffic accidents through data analysis by identifying patterns and predicting high-risk scenarios. Over all by leveraging Machine learning models I identifies the major factors contributing to road Traffic accidents on the Addis Adama Expressway, such as chainage, driver age, driver experience, day of the week, cause of accident, geo-location, vehicle types, crash type, weather condition, driver relationship, road surface condition, and sex respectively.

5.2. Recommendations

Based on the research mentioned above, SVM model were shown to be the most accurate in identifying the types of accidents from the provided dataset. From the data research, it was also found that the days and times with the highest number of accidents occur on Tuesdays and Saturdays. As a result, these locations should be closely watched, and emergency response should be provided near high-accident areas. I recommended that Continuous Awareness Training on Safe Driving to enhance road safety, implement a centralized traffic data system that continuously collects and analyzes information from sources such as GPS, traffic cameras, and sensors to identify accident patterns and high-risk areas, and also use predictive analytics to forecast potential hotspots and deploy preventive measures proactively. To improving the factors that influence accident occurrence in Addis Adama Expressway the following targeted strategies addressing each aspect (NHTSA, 2021).

Chainage:-Implementing better road infrastructure and maintenance at specific chainages can reduce accidents. Regular inspections, timely repairs, and installation of safety features like guardrails and signage can mitigate risks.

Driver Age: - Tailored education programs focusing on younger and older drivers can enhance road safety. For young drivers, introducing more comprehensive driver training and graduated licensing systems can be beneficial. For older drivers, periodic medical assessments and refresher courses can help maintain their driving skills.

Driver Experience: -Encouraging continuous driver education and offering advanced driving courses can help drivers improve their skills over time. Companies can also implement mentorship programs where experienced drivers guide novices.

Day of the Week: -Analyzing accident data to identify high-risk days can lead to better resource allocation for law enforcement and emergency services. Public awareness campaigns can highlight the increased risk on certain days, encouraging safer driving behaviors.

Cause of Accident: -Conducting thorough investigations to understand the root causes of accidents can lead to more effective prevention strategies. Implementing targeted measures like speed control, DUI checkpoints, and distracted driving laws can address common causes.

Geo-location: Utilizing geographic information systems to map accident hotspots can guide infrastructure improvements and resource deployment. Enhanced lighting, road markings, and traffic calming measures in high-risk areas can significantly reduce accidents.

Vehicle Types: Promoting the use of safer vehicle technologies, such as automatic emergency braking and lane-keeping assist, can reduce accident rates. Encouraging regular vehicle maintenance and safety inspections ensures vehicles remain roadworthy.

Crash Type: Understanding the dynamics of different crash types allows for targeted interventions. For instance, to prevent rear-end collisions, improving traffic flow and installing advanced traffic management systems can be effective.

Weather Condition: Implementing advanced weather forecasting

systems and real-time road condition monitoring can inform drivers of hazardous conditions. Road authorities can proactively treat roads with de-icing agents or close dangerous sections during severe weather. Driver Relationship: Addressing interpersonal issues that may distract drivers, such as family or work-related stress, through public awareness and support programs can enhance focus and decision-making behind the wheel. Road Surface Condition: Regular maintenance and resurfacing of roads ensure better traction and fewer potholes, reducing accident risks. Implementing high-quality materials and construction standards can prolong the lifespan of road surfaces. Sex: Gender-specific educational programs can address the unique driving challenges faced by different genders.

5.3. Limitation

Data size and data quality are two major constraints when developing a predictive model for traffic accidents. The insufficiency of available data samples for AI model training is the root cause of the data size limitation. Inadequate sample sizes have the potential to produce biased or incomplete models by failing to capture the complexity and diversity of actual accident occurrences. The confusion matrices across all models highlight the challenge of accurately predicting minority classes.

5.4. Feature Works

Upcoming research may focus on improving data size and quality and also using another machine learning model and Deep Learning Algorithm to get better Accuracy with better system performance by combining different types of real-time data to provide current risk rankings and the combination of data from many sources, including as traffic flow and driver behavior data. Additionally, by tailoring risk rankings according to driver characteristics, the system may be enhanced and potentially promote safer driving practices.

REFERENCES

- Abegaz, T., & Gebremedhin, S. (2019). Magnitude of road traffic accident-related injuries and fatalities in Ethiopia. *PLOS ONE*, 14(1), e0202240. <https://doi.org/10.1371/journal.pone.0202240>
- Alagarsamy, S., Nagaraj, P., Srikanth, B., Krishna, Ch. V., Bharath, G., & Kalyan, Sv. S. (2023). A Novel Machine Learning Technique for Predicting Road Accidents. 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), 1547–1551. <https://doi.org/10.1109/ICAIS56108.2023.10073778>
- Alqatawna, A., Rivas Alvarez, A. M., & Garcia-Moreno, S. S. C. (2021). Comparison of Multivariate Regression Models and Artificial Neural Networks for Prediction Highway Traffic Accidents in Spain: A Case Study. *Transportation Research Procedia*, 58, 277–284. <https://doi.org/10.1016/j.trpro.2021.11.038>
- Bala Kishore, M., Srinivasa Reddy, I., Sujihelen, L., Sivasangari, A., & Christy, A. (2023). Artificial Intelligence (AI) based Detection of Traffic Violations by Two-Wheeler Vehicles. 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), 1788–1792. <https://doi.org/10.1109/ICICCS56967.2023.10142605>
- Bharadiya, J. P. (n.d.). Artificial Intelligence in Transportation Systems A Critical Review. www.ajpojournals.org
- Chen, J., & Tao, W. (2022). Traffic accident duration prediction using text mining and ensemble learning on expressways. *Scientific Reports*, 12(1), 21478. <https://doi.org/10.1038/s41598-022-25988-4>
- Cheng, Z. , L. J. , & G. E. (2019). Urban Road Traffic Accident Prediction Using Spatio-Temporal Data. 2019 IEEE International Conference on Data Mining (ICDM).
- Farhat, Z., Karouni, A., Daya, B., Chauvet, P., & Hmadeh, N. (2020). International Journal of Innovative Technology and Exploring Engineering (IJITEE). International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2020(7), 2278–3075. <https://doi.org/10.35940/ijitee.F4393.059720i>
- Gatarić, D., Ruškić, N., Aleksić, B., Đurić, T., Pezo, L., Lončar, B., & Pezo, M. (2023a). Predicting Road Traffic Accidents—Artificial Neural Network Approach. *Algorithms*, 16(5), 257. <https://doi.org/10.3390/a16050257>
- Gaud, D. (2023). Ethical Considerations for the Use of AI Language Model. *International Journal for Research in Applied Science and Engineering Technology*, 11(7), 6–14. <https://doi.org/10.22214/ijraset.2023.54513>
- Hamdan, S. M. S., Barakat, S., Mahfouz, K. H., & Ghuzlan, K. A. (2023). Traffic Accident Severity Prediction Model using AI. 2023 Advances in Science and Engineering Technology International Conferences (ASET), 1–5. <https://doi.org/10.1109/ASET56582.2023.10180861>
- Haulle, E., & Kisiri, M. (2016). The Impact of Road Accidents to the Community of Iringa Municipality: Challenges in Reducing Risks. *International and Multidisciplinary Journal of Social Sciences*, 5(3), 253. <https://doi.org/10.17583/rimcis.2016.1880>
- Hemment, D., Currie, M., Bennett, S., Elwes, J., Ridler, A., Sindors, C., Vidmar, M., Hill, R., & Warner, H. (2023). AI in the Public Eye: Investigating Public AI Literacy Through AI Art. 2023 ACM Conference on Fairness, Accountability, and Transparency, 931–942. <https://doi.org/10.1145/3593013.3594052>
- Hu, Z., Zhou, J., & Zhang, E. (2023a). Improving Traffic Safety through Traffic Accident Risk Assessment. *Sustainability*, 15(4), 3748. <https://doi.org/10.3390/su15043748>
- Huang, B., Hooi, B., & Shu, K. (2023). TAP: A Comprehensive Data Repository for Traffic Accident Prediction in Road Networks. <http://arxiv.org/abs/2304.08640>
- Jiang, H., & Deng, H. (2020). Traffic Incident Detection Method Based on Factor Analysis and Weighted Random Forest. *IEEE Access*, 8, 168394–168404. <https://doi.org/10.1109/ACCESS.2020.3023961>
- Khanum, H., Garg, A., & Faheem, M. I. (2023). Accident severity prediction modeling for road safety using random forest algorithm: an analysis of Indian highways. *F1000Research*, 12, 494. <https://doi.org/10.12688/f1000research.133594.1>
- Kim, Y. W., Lee, D. W., & Hwang, A. (2023). Development of a City-wide Traffic Accident Prediction Model Using Hybrid Machine-based Learning Approaches. *Crisis and Emergency Management: Theory and Praxis*, 13(3), 49–55. <https://doi.org/10.14251/jscm.2023.3.57>
- Lazar, H., & Jarir, Z. (2022). Road traffic accident prediction: a driving behavior approach. 2022 8th International Conference on Optimization and Applications (ICOA), 1–4. <https://doi.org/10.1109/ICOA55659.2022.9934000>
- Lindholm, A., Wahlström, N., Lindsten, F., & Schön, T. B. (2022). *Machine Learning*. Cambridge University Press. <https://doi.org/10.1017/9781108919371>
- Madushani, J. P. S. S., Sandamal, R. M. K., Meddage, D. P. P., Pasindu, H. R., & Gomes, P. I. A. (2023). Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers. *Transportation Engineering*, 13, 100190. <https://doi.org/10.1016/j.treng.2023.100190>
- Marcillo, P., Valdivieso Caraguay, Á. L., & Hernández-Álvarez, M. (2022). A Systematic Literature Review of Learning-Based Traffic Accident Prediction Models Based on Heterogeneous Sources. *Applied Sciences*, 12(9), 4529. <https://doi.org/10.3390/app12094529>
- Nechoska, D. K., Nechoska, R. P., & Duma, R. (2022a). Application of Artificial Intelligence for Traffic Data Analysis, Simulations and Adaptation. 2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), 1–4. <https://doi.org/10.1109/ICEST55168.2022.9828690>

- Olugbade, S., Ojo, S., Imoize, A. L., Isabona, J., & Alaba, M. O. (2022). A Review of Artificial Intelligence and Machine Learning for Incident Detectors in Road Transport Systems. *Mathematical and Computational Applications*, 27(5), 77. <https://doi.org/10.3390/mca27050077>
- Pourroostaei Ardakani, S., Liang, X., Mengistu, K. T., So, R. S., Wei, X., He, B., & Cheshmehzangi, A. (2023a). Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis. *Sustainability (Switzerland)*, 15(7). <https://doi.org/10.3390/su15075939>
- Reddy, Asst. Prof. J. D. (2023a). AI Based Smart Traffic Management. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 07(07). <https://doi.org/10.55041/IJSREM24517>
- Sameen, M. I., & Pradhan, B. (2017). Severity prediction of traffic accidents with recurrent neural networks. *Applied Sciences (Switzerland)*, 7(6). <https://doi.org/10.3390/app7060476>
- Samia, Z., NourElhouda, Z., & Warda, Z. (2022). AI Technologies for Predicting and Preventing Road Accidents - A Review. 2022 International Symposium on INnovative Informatics of Biskra (ISNIB), 1–6. <https://doi.org/10.1109/ISNIB57382.2022.10075776>
- Shatnawi, N., Al-Omari, A. A., & Alkhateeb, S. (2023). Prediction of Risk Factors Influencing Severity Level of Traffic Accidents Using Artificial Intelligence. *International Review of Civil Engineering (IRECE)*, 14(1), 1. <https://doi.org/10.15866/irece.v14i1.20534>
- Shi, D., & Tang, H. (2021). Research on Safe Driving Evaluation Method Based on Machine Vision and Long Short-Term Memory Network. *Journal of Electrical and Computer Engineering*, 2021. <https://doi.org/10.1155/2021/9955079>
- Shukla, M. (2022). Research Methods in Machine Learning: A Content Analysis. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 06(05). <https://doi.org/10.55041/IJSREM12736>
- Wang, S., Changshun, Y., & Yong, S. (2023). A Review of Road Traffic Accident Prediction Methods. *American Journal of Management Science and Engineering*. <https://doi.org/10.11648/j.ajmse.20230803.12>
- Wang, S., Yan, C., & Shao, Y. (2023). LSTM Road Traffic Accident Prediction Model Based on Attention Mechanism. 2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 215–219. <https://doi.org/10.1109/ICCCBDA56900.2023.10154750>
- Yan, M., & Shen, Y. (2022). Traffic Accident Severity Prediction Based on Random Forest. *Sustainability (Switzerland)*, 14(3). <https://doi.org/10.3390/su14031729>
- Zhang, S., Liu, H., Yang, Y., Zhang, S., Zhang, Z., Wang, C., & Wang, M. (2023). Prediction of traffic accident impact range based on CatBoost ensemble algorithm. In A. Palanisamy Muthuramalingam & K. Subramaniam (Eds.), *Second International Conference on Algorithms, Microchips, and Network Applications (AMNA 2023)* (p. 54). SPIE. <https://doi.org/10.1117/12.2679147>
- Zhong, W., & Du, L. (2023). Predicting Traffic Casualties Using Support Vector Machines with Heuristic Algorithms: A Study Based on Collision Data of Urban Roads. *Sustainability*, 15(4), 2944

Appendix

A. Sample Traffic Accident Dataset

DateTime	Date	TimeT	Time corrected	Day of week	Crash types	Couse of accident
1/11/2008	2008-01-11	12:05AM	12:00:00	Friday	Rollover	Technical Problem
11/1/2008	2008-11-01	11:00AM	11:00:00	Sunday	Crash with fixed obstacle	Driver Error
2/11/2008	2008-02-11	1:55PM	12:00:00	Saturday	Crash with fixed obstacle	Technical Problem
3/11/2008	2008-03-11	10:00PM	10:00:00	Sunday	Crash with fixed obstacle	Tire explosion
3/11/2008	2008-03-11	5:00PM	5:00:00	Sunday	Crash with fixed obstacle	Driver Error
5/11/2008	2008-05-11	1:00AM	1:00:00	Tuesday	Rollover	Over speeding
11/5/2008	2008-11-05	6:00PM	6:00:00	Monday	Rollover	Over speeding
11/6/2008	2008-11-06	2:25AM	2:02:00	Wednesday	Crash with fixed obstacle	Over speeding
11/6/2008	2008-11-06	11:30AM	11:03:00	Wednesday	Crash with fixed obstacle	Tire explosion
11/7/2008	2008-11-07	12:05AM	12:00:00	Thursday	Crash with fixed obstacle	Over speeding
11/7/2008	2008-11-07	2:40AM	2:04:00	Thursday	Crash with fixed obstacle	Technical Problem
11/13/2008	2008-11-13	5:00AM	5:00:00	Wednesday	Crash with fixed obstacle	Over speeding

Changer	1km in meter	Chainage	Light condition	Geo location	Road surface condition	Weather condition
37450	1000	37.45	Daylight	Straight	Dry	Cloudy
7340	1000	7.34	Daylight	Sloppy	Wet	Rainy
6730	1000	6.73	Darkness	Sloppy	Dry	Clear
38290	1000	38.29	Darkness	Straight	Wet	Rainy
23640	1000	23.64	Darkness	Straight	Wet	Rainy
24020	1000	24.02	Daylight	Straight	Wet	Rainy
2000	1000	2	Darkness	Straight	Dry	Clear
2000	1000	2	Daylight	Sloppy	Dry	Clear
2000	1000	2	Daylight	Sloppy	Dry	Clear
52000	1000	52	On sun going to dawn	Curve	Wet	Rainy

46590	1000	46.59	Daylight	Sloppy	Wet	Rainy
23000	1000	23	Darkness	Straight	Wet	Rainy
7340	1000	7.34	Darkness	Sloppy	Wet	Rainy
7960	1000	7.96	On sun going to dawn	Sloppy	Dry	Clear

Fatal injury	Serious injury	Slight injury	Accident severity	Etre vc	Driver relationship	Sex	Age	Driving experience
1	1	1	Fatal Accident	V1	Employer	M	26	3 Years
0	0	1	Slight Accident	V1	Employer	M	27	2 Years
0	1	1	Fatal Accident	V4	Employer	M	43	10 Years
1	1		Serious Accident	V3	Employer	M	28	5 Years
0	2	1	Serious Accident	V1	Employer	M	30	9 Years
2	2		Fatal Accident	V1	Employer	M	35	3 Years
0	0	1	Slight Accident	V4	Employer	M	23	3 Years
0	0	1	Slight Accident	V2	Employer	M	22	3 Years
0	0	1	Slight Accident	V3	Employer	M	60	30 Years
1	1	1	Serious Accident	V1	Owner	M	38	12 Years

B. Data Preparation

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3212 entries, 0 to 3211
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Date                  3212 non-null  object
1   Time                  3212 non-null  object
2   Day_of_week           3212 non-null  object
3   Accident_types        3212 non-null  object
4   Cause_of_accident     3212 non-null  object
5   Chainager             3212 non-null  int64
6   Chainage              3212 non-null  float64
7   Light_condition       3212 non-null  object
8   Geometric_condition   3212 non-null  object
9   Road_surface_condition 3212 non-null  object
10  Weather_condition     3212 non-null  object
11  Fatal_injury          3212 non-null  object
12  Serious_injury        3212 non-null  object
13  Slight_injury         3212 non-null  object
14  Accident_severity     3212 non-null  object
15  Etre_vc               3212 non-null  object
16  Driver_relationship   3212 non-null  object
17  Sex                   3212 non-null  object
18  Age                   3212 non-null  int64
19  Driving_experience     3212 non-null  object
```

```
print(df.isnull().sum())
```

```
Date          0
Time          0
Day_of_week    0
Accident_types 0
Cause_of_accident 0
Chainager      0
Chainage       0
Light_condition 0
Geometric_condition 0
Road_surface_condition 0
Weather_condition 0
Fatal_injury   0
Serious_injury 0
Slight_injury  0
Accident_severity 0
Etre_vc        0
Driver_relationship 0
Sex            0
Age           0
Driving_experience 0
dtype: int64
```

```
[15]:
```

	Date	Time	Day_of_week	Accident_types	Cause_of_accident	Chainager	Chainage	Light_condition	Geometric_condition	Road_surface_condition	V
0	2008-01-11	12:00:00	Friday	Rollover	Technical Problem	37450	37.45	Daylight	Straight		Dry
1	2008-01-11	11:00:00	Sunday	Crash with fixed obstacle	Driver Error	7340	7.34	Daylight	Sloppy		Wet
2	2008-02-11	12:00:00	Saturday	Crash with fixed obstacle	Technical Problem	6730	6.73	Darkness	Sloppy		Dry
3	2008-03-11	10:00:00	Sunday	Crash with fixed obstacle	Tire explosion	38290	38.29	Darkness	Straight		Wet
4	2008-03-11	5:00:00	Sunday	Crash with fixed obstacle	Driver Error	23640	23.64	Darkness	Straight		Wet

	Date	Time	Day_of_week	Crash_types	Couse_of_accident	Chainager	Chainage	Light_condition	Geo_location	I
0	2008-01-11	12:00:00	Friday	Rollover	Technical Problem	37450	37.45	Daylight	Straight	
1	2008-11-01	11:00:00	Sunday	Crash with fixed obstacle	Driver Error	7340	7.34	Daylight	Sloppy	
2	2008-02-11	12:00:00	Saturday	Crash with fixed obstacle	Technical Problem	6730	6.73	Darkness	Sloppy	
3	2008-03-11	10:00:00	Sunday	Crash with fixed obstacle	Tire explosion	38290	38.29	Darkness	Straight	
4	2008-03-11	5:00:00	Sunday	Crash with fixed obstacle	Driver Error	23640	23.64	Darkness	Straight	

Road_surface_condition	Weather_condition	Fatal_injury	Serious_injury	Slight_injury	Accident_severity	Etre_vc	Driver_relationship	Sex	Age	Driving_experience	
	Dry	Cloudy	1	1	1	Fatal Accident	V1	Employer	M	26	3 Years
	Wet	Rainy	0	0	0	Slight Accident	V1	Employer	M	27	2 Years
	Dry	Clear	0	1	0	Fatal Accident	V4	Employer	M	43	10 Years
	Wet	Rainy	1	1	0	Serious Accident	V3	Employer	M	28	5 Years
	Wet	Rainy	0	2	0	Serious Accident	V1	Employer	M	30	9 Years

```
df.columns
```

```
Index(['Day_of_week', 'Crash_types', 'Couse_of_accident', 'Chainager',
      'Light_condition', 'Geo_location', 'Road_surface_condition',
      'Weather_condition', 'Accident_types', 'Etre_vc', 'Driver_relationship',
      'Sex', 'Age', 'Driving_experience'],
      dtype='object')
```

C. Sample code to build Traffic accident predictive Model

- **Data Collection**

```
import pandas as pd
# Load the dataset
data = pd.read_csv('traffic_accidents.csv')
```

- **Data Preprocessing**

```
# Handling missing values
data.fillna(method='ffill', inplace=True)
# Encoding categorical variables
data = pd.get_dummies(data, drop_first=True)
```

- **Exploratory Data Analysis (EDA)**

```
import matplotlib.pyplot as plt
import seaborn as sns
```

- **Feature Engineering**

```
# Example: Creating a new feature combining date and time
data['Datetime'] = pd.to_datetime(data['Date'] + ' ' + data['Time'])
data['Hour'] = data['Datetime'].dt.hour
```

- **splitting the Data**

```
from sklearn.model_selection import train_test_split
# Define features and target variable
X = data.drop([Accident_type], axis=1)
y = data['Accident_type']
# Splitting the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

- **model build**

```
from sklearn.svm import SVC
# Create the SVM model
svm_model = SVC(kernel='linear')
# Train the model
svm_model.fit(X_train, y_train)
```

- **Model Evaluation**

```
from sklearn.metrics import classification_report, confusion_matrix
```

```
# Predict on the test set
y_pred = svm_model.predict(X_test)
# Evaluation metrics
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred))
```

- **Hyperparameter Tuning**

```
from sklearn.model_selection import GridSearchCV
# Define parameter grid
param_grid = {'C': [0.1, 1, 10, 100], 'kernel': ['linear', 'rbf', 'poly', 'sigmoid']}
# Grid search
grid = GridSearchCV(SVC(), param_grid, refit=True, verbose=2)
grid.fit(X_train, y_train)
# Best parameters
print(grid.best_params_)
# Best estimator
best_model = grid.best_estimator_
```

- **Model Deployment**

```
import joblib
# Save the model
joblib.dump(best_model, 'svm_traffic_accident_model.pkl')
```