ACCIDENT PREDICTION AND

PREVENTING SYSTEM USING DATA DRIVEN MODELS

A PROJECT REPORT

Submitted by

LINKEDH S	**(721220243028)**
NATHISH M	**(721220243035)**
PRASANTH M	**(721220243045)**
SRIKANTH K	**(721220243053)**

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

KARPAGAM INSTITUTE OF TECHNOLOGY, COIMBATORE

ANNA UNIVERSITY: CHENNAI 600 025

MAY 2023

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report "**ACCIDENT PREDICTION AND PREVENTION USING DATA DRIVEN MODELS"** is the bonafide work of "**LINKEDH S (721220243028), NATHISH M (721220243035), PRASANTH (721220243045), SRIKANTH K (721220243053)"** who carried out the project work under my supervision.

Dr.R.NALLAKUMAR M.E, MBA, Ph.D.,	**Dr.R.NALLAKUMAR M.E, MBA, Ph.D.,**
SUPERVISOR,	**HEAD OF THE DEPARTMENT,**
Associate Professor,	Associate Professor,
Department of Artificial Intelligence &	Department of Artificial Intelligence &
Data Science,	Data Science,
Karpagam Institute of Technology,	Karpagam Institute of Technology,
Coimbatore – 641105	Coimbatore – 641105

Submitted for the university Socially Relevant Project Viva-voce examination conducted at Karpagam Institute of Technology, Coimbatore, on …………….. .

Internal Examiner	**External Examiner**

ACKNOWLEDGEMENT

The success of work depends on the team and its cooperation. This opportunity is to express our gratitude and sincere thanks to everyone who helped us in our project. First and foremost, we would like to thank the Management for the excellent infrastructure, facilities and the constant support provided for the successful completion for the project work.

We wish to express our heartfelt thanks and deep sense of gratitude to our chairman **Dr. R.Vasanthakumar** for his valuable guidance and continuous support.

Our sincere thanks to honourable Principal **Dr. P.Manimaran M.E, MBA, Ph.D.,** and Vice Principal **Dr.D.Bhanu M.E, Ph.D.,** for giving us the opportunity to display our professional skills through this project.

Our special thanks to **Dr.R.Nallakumar M.E, MBA, Ph.D.,** Head of the Department, Project Coordinator and Guide for his valuable guidance, continuous support and suggestions to improve the quality of the project work.

We express our deep sense of gratitude to all the faculty members of the supporting staff for their continuous support in completing this project work successfully.

LINKEDH S (721220243028),

NATHISH M (721220243035),

PRASANTH (721220243045),

SRIKANTH K (721220243053)

ABSTRACT

The Accident Prediction and Prevention Using Data-Driven Models project aims to develop a machine learning-based system for predicting and preventing accidents. The system will produce real-time estimates of accident likelihood and recommendations for preventive actions by analysing past accident data and recognising patterns and trends. To analyse the data and make predictions, the research will employ machine learning methods such as decision trees, random forests, and neural networks. The system will be trained using a huge dataset of accident data and will be updated on a regular basis with fresh information. The purpose is to minimise the frequency of traffic accidents and enhance safety by offering insights into the elements that cause accidents and recommending preventive actions. By offering a strong tool for predicting and avoiding accidents, the research has the potential to revolutionise the area of accident prevention.

The accident prediction and prevention project aims to develop a data-driven approach to promote safety and reduce the risk of accidents in various settings. The project involves collecting and pre-processing data, clustering using the KNN algorithm, selecting significant features, developing a decision tree model, and validating the model. The developed models aim to provide insights into the likelihood of accidents and the factors that contribute to accidents, and help stakeholders to make informed decisions and develop effective prevention strategies. The project also involves the use of data visualization tools to identify trends, patterns, and outliers in the data, and the evaluation of the developed models and prevention strategies to improve their performance. Overall, the accident prediction and prevention project involves a comprehensive and continuous approach to safety and risk management, combining data-driven models and techniques, evaluation and monitoring, and real-time warning systems.

TABLE OF CONTENT

LIST OF ABBREVIATIONS

3D warps – 3 Dimensional Warps

ADAS - Advanced Driver Assistance Systems

AI – Artificial Intelligence

ANN - Artificial Neural Network

AV - Autonomous Vehicle

CNN - Convolutional Neural Network

DCNN - Deep Convolutional Neural Network

IoT - Internet of Things

ITS - Intelligent Transportation Systems

ML - Machine Learning

MTCNN - Multi-task Cascaded Convolutional Networks

NHTSA - National Highway Traffic Safety Administration

OBD - On-Board Diagnostics

OpenCV - Open-Source Computer Vision

PCA - Principal Component Analysis

SVM - Support Vector Machine

V2X - Vehicle-to-Everything communication

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 DATA SCIENCE

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from structured and unstructured data. It encompasses a range of techniques and tools, including statistical analysis, machine learning, data visualization, and data mining. Data science is used in a variety of industries, such as finance, healthcare, and retail, to help organizations make better decisions and improve their operations. The goal of data science is to turn raw data into actionable insights that can be used to inform business strategy, improve products and services, and drive innovation.

Data science has a wide variety of applications in various industries such as healthcare, finance, transportation, retail, manufacturing, and many others. In healthcare, data science is used to identify patterns in patient data that can help to improve diagnosis and treatment. In finance, data science is used to detect fraud, analyze market trends, and make better investment decisions. In retail, data science is used to optimize pricing, personalize marketing, and improve inventory management. The list goes on, and the possibilities are endless.

1.2 ACCIDENT PREDICTION AND PREVENTION

Accident prediction and prevention are two related concepts that aim to identify and mitigate potential hazards in various settings to reduce the likelihood of accidents occurring. Accident prediction involves analyzing historical data and identifying patterns and risk factors that increase the likelihood of accidents occurring. This information can then be used to develop models and algorithms that can predict the probability of accidents occurring in the future, based on various factors such as weather conditions, traffic density, and human behavior.

Accident prevention, on the other hand, involves taking proactive measures to eliminate or reduce potential hazards that could cause accidents. This may involve implementing engineering controls such as installing safety barriers, warning signs, and traffic lights, or implementing administrative controls such as establishing safety procedures and training programs for workers.

Overall, accident prediction and prevention are crucial components of risk management and play an essential role in ensuring the safety of individuals and communities.

1.2.1 Different Techniques and Models

Accident prediction and prevention are critical components of risk management. Accident prediction involves analyzing historical data and identifying patterns and risk factors that increase the likelihood of accidents occurring. This information can then be used to develop models and algorithms that can predict the probability of accidents occurring in the future. Accident prevention involves taking proactive measures to eliminate or reduce potential hazards that could cause accidents. Various techniques and models such as Fault Tree Analysis, Bayesian Networks, and Hazard and Operability Study are used in accident prediction and prevention. The choice of technique or model depends on the specific situation and the type of accident being analyzed.

1.3 CHALLENGES IN ACCIDENT PREDICTION

There are several challenges that can arise in Accident Prediction projects, such as:

- **Data availability and quality:** Accurate accident prediction models require large amounts of high-quality data, which may not always be available. The data may be incomplete, inaccurate, or biased, making it difficult to develop accurate models.
- **Complex and dynamic systems:** The systems involved in accidents are often complex and dynamic, making it difficult to identify all the variables and their interactions. Accidents can occur due to unforeseen events, making it challenging to predict them.
- **Unpredictable nature of human behavior:** Human behavior is unpredictable and can be a significant factor in accidents. The behavior of individuals involved in an accident may be influenced by factors such as fatigue, stress, and distractions, making it difficult to predict their actions accurately.
- **Limited resources:** The implementation of accident prevention measures can face challenges due to limited resources, making it difficult to implement effective solutions.
- **Resistance to change:** People may resist adopting new technologies, practices, or policies that could prevent accidents. This can lead to a reluctance to adopt effective prevention measures.
- **Competing priorities:** Accident prevention measures may compete with other priorities, such as productivity or cost savings. This can make it difficult to prioritize accident prevention initiatives and allocate resources effectively.

To address these challenges, researchers and practitioners are constantly developing new techniques and models that can improve the accuracy and robustness of Accident Prediction systems, while also addressing privacy and bias concerns.

1.4 ALGORITHMS IN ACCIDENT PREDICTION AND PREVENTION

There are various algorithms used in accident prediction and prevention, depending on the specific situation and the type of accident being analyzed. Some common algorithms include decision trees, neural networks, and support vector machines. These algorithms use historical data to identify patterns and predict the likelihood of future accidents. They can take into account various factors such as weather conditions, traffic density, and human behavior to make predictions.

By identifying potential hazards and predicting the likelihood of accidents, these algorithms can help develop effective prevention strategies, such as implementing safety procedures and engineering controls. Overall, the use of algorithms is a crucial component of accident prediction and prevention, helping to reduce the risk of accidents and promote safety in various settings.

1.4.1 Decision trees

Decision trees are algorithms that use historical data to identify patterns and predict future outcomes. They create a tree-like model that represents possible outcomes and the factors that contribute to them. Decision trees can be used to predict the likelihood of accidents based on various factors such as weather conditions, traffic density, and human behavior.

1.4.2 Neural networks

Neural networks are machine learning algorithms that can recognize patterns and relationships between variables. They are used to analyze large amounts of data and identify complex relationships between factors that may contribute to accidents. Neural networks can be trained to predict the likelihood of accidents based on historical data.

1.4.3 Support Vector Machines (SVMs)

SVMs are algorithms that can be used to analyze and classify data. They can be used to predict the likelihood of accidents based on various factors such as weather conditions, traffic density, and human behavior. SVMs use historical data to identify patterns and predict future outcomes.

1.4.4 Bayesian networks

Bayesian networks are probabilistic models that represent the relationships between variables and their likelihood of occurrence. They can be used to predict the probability of accidents occurring based on historical data. Bayesian networks can take into account various factors such as weather conditions, traffic density, and human behavior to make predictions.

1.5 NEED FOR THE STUDY OF ACCIDENT PREDICTION

The study of accident prediction and prevention is crucial for promoting safety and reducing the risk of accidents. Accidents can cause significant harm to individuals, property, and the environment and result in financial losses and damage to reputation. By studying accident prediction and prevention, companies can develop effective prevention strategies to reduce the likelihood of accidents occurring. Implementing safety measures can also help companies comply with regulatory requirements, demonstrate their commitment to safety, and continually improve safety performance. Overall, the study of accident prediction and prevention is necessary for protecting lives and property, reducing costs, and promoting safety in various settings.

The study of accident prediction and prevention is essential for several reasons:

- **Safety:** The primary reason for studying accident prediction and prevention is to ensure safety. Accidents can cause significant harm to individuals, property, and the environment. By predicting the likelihood of accidents and implementing prevention strategies, the risk of accidents can be reduced, promoting safety in various settings.
- **Cost savings:** Accidents can result in significant financial losses due to property damage, medical expenses, and legal fees. By implementing prevention strategies, the cost of accidents can be reduced, resulting in cost savings.
- **Regulatory Compliance:** Many industries are subject to regulations and standards that require them to implement safety measures to prevent accidents. By studying accident prediction and prevention, companies can comply with regulatory requirements and avoid fines and penalties.
- **Reputation:** Accidents can damage a company's reputation and lead to a loss of trust among stakeholders. By implementing effective accident prevention strategies, companies can demonstrate their commitment to safety and promote a positive reputation.
- **Continuous Improvement:** Studying accident prediction and prevention allows companies to continually improve their safety measures and prevent accidents from occurring in the future. By analyzing accidents and identifying the root causes, companies can develop effective prevention strategies and continually improve their safety performance.

CHAPTER 2

LITERATURE REVIEW

T. Deva Prakash, Nagaraju V," **Comparative Analysis using K - Nearest Neighbour with Artificial Neural Network to Improve Accuracy for Predicting Road Accidents**", 2022 International Conference on Edge Computing and Applications (ICECAA) , This study focuses on improving the accuracy of road accident prediction by employing a Comparative Analysis approach that combines K-Nearest Neighbors (KNN) and Artificial Neural Network (ANN) techniques. Accurate prediction of road accidents is crucial for implementing preventive measures and reducing the associated risks. While KNN is commonly used for classification tasks, its limitations in handling complex patterns and high-dimensional data can impact accuracy. To overcome these limitations, an ANN is introduced, as it can capture non-linear relationships and adapt to intricate data structures.

S. S. Yassin and Pooja, "**Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach**", SN Applied Sciences, vol. 2, no. 9, pp. 1–13, Aug. 2020, One problem in road accident prediction is the accuracy of the models used. Traditional approaches may struggle to capture the complex patterns and interactions among various factors contributing to accidents. Additionally, model interpretation is often challenging, making it difficult to understand the underlying factors influencing predictions. To address these issues, a hybrid approach combining K-means clustering and random forest algorithm is proposed. However, challenges may arise in finding an optimal number of clusters in K-means and interpreting the combined model's results. Ensuring accurate predictions and providing meaningful insights for decision-making remain crucial goals in road accident prediction.

Rasika Vijithasena, Wikasitha Herath, "**Data Visualization and Machine Learning Approach for Analyzing Severity of Road Accidents**", 2022 International Conference for Advancement in Technology (ICONAT) Goa, India, One problem in analyzing the severity of road accidents using data visualization and machine learning approaches is the complexity and high dimensionality of the data involved. Road accident datasets often contain numerous features such as weather conditions, road types, and driver characteristics, making it challenging to effectively visualize and interpret the data. Another challenge is selecting the most suitable machine learning algorithms and tuning their hyperparameters to accurately predict accident severity. Balancing the trade-off between model complexity and interpretability is another issue, as more complex models may provide better prediction accuracy but can be harder to interpret. Addressing these challenges requires careful consideration of data visualization techniques and selecting appropriate machine learning algorithms that strike the right balance between accuracy and interpretability.

Partheeban Pachaivannan, Rani Hemamalini Ranganathan, "**Indian Road Conditions and Accident Risk Predictions using Deep Learning Approach – A Review**", Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020], One problem in the review of Indian road conditions and accident risk predictions using a deep learning approach is the scarcity of comprehensive and high-quality datasets specific to Indian road conditions. Obtaining accurate and reliable data on road conditions, traffic patterns, and accident records in India can be challenging. Limited availability of such data can hinder the training and evaluation of deep learning models, potentially impacting the accuracy of accident risk predictions. Additionally, there may be variations in road conditions, driving behaviors, and infrastructure across different regions in India, making it necessary to consider these factors while developing a deep learning approach. Addressing these data limitations and accounting for regional differences are key challenges in conducting a comprehensive review of accident risk predictions in the context of Indian road conditions using deep learning methods.

CHAPTER 3

SYSTEM SPECIFICATION

3.1 HARDWARE SPECIFICATION

Minimum Specification

- 8 GB RAM
- 100 GB HDD
- Octa Core processor

Recommended Specification

- 16 GB RAM
- 150 GB SSD
- Intel Core i5-8259U, or AMD Ryzen 5 2700X (Processor)
- NVIDIA GT 1050 or Quadro P1000 (Graphic Card)

3.2 SOFTWARE SPECIFICATION

OPERATING SYSTEM : WINDOWS 10 AND ABOVE

LANGUAGES : Python, SQL

SOFTWARE : Pycharm, Jupyter Notebook

SERVER : Local Database (If requires)

3.3 SOFTWARE OVERVIEW

3.3.1 Python

Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems. Python is an interpreted, interactive, object-oriented programming language. It incorporates modules, exceptions, dynamic typing, very high-level dynamic data types, and classes. It supports multiple programming paradigms beyond object-oriented programming, such as procedural and functional programming. It uses a simplified syntax with an emphasis on natural language, for a much easier learning curve for beginners. And, because Python is free to use and is supported by an extremely large ecosystem of libraries and packages, it's often the first-choice language for new developers.

3.3.2 SQL (Structured Query Language)

SQL is used to communicate with a database. According to ANSI (American National Standards Institute), it is the standard language for relational database management systems. SQL statements are used to perform tasks such as update data on a database, or retrieve data from a database. SQL works by understanding and analysing data of virtually any size, from small datasets to large stacks. It's a very powerful tool that enables you to perform many functions at high efficiency and speed. SQL is used to communicate with a database. According to ANSI (American National Standards Institute), it is the standard language for relational database management systems. SQL statements are used to perform tasks such as update data on a database, or retrieve data from a database.

3.3.3 Pycharm

It allows viewing of the source code in a click. Software development is much faster using PyCharm. The feature of error spotlighting in the code further enhances the development process. The community of Python Developers is extremely large so that we can resolve our queries/doubts easily. PyCharm is a dedicated Python Integrated Development Environment (IDE) providing a wide range of essential tools for Python developers, tightly integrated to create a convenient environment for productive Python, web, and data science development. It makes Python development accessible to those who are new to the world of software programming. PyCharm Community Edition is excellent for developers who wish to get more experience with Python.

3.3.4 Jupyter Notebook

Jupyter Notebook allows users to compile all aspects of a data project in one place making it easier to show the entire process of a project to your intended audience. Through the web-based application, users can create data visualizations and other components of a project to share with others via the platform. The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience. Jupyter Notebook allows users to convert the notebooks into other formats such as HTML and PDF. It also uses online tools and nbviewer which allows you to render a publicly available notebook in the browser directly. Jupyter is another best IDE for Python

Programming that offers an easy-to-use, interactive data science environment across many programming languages besides Python.

3.3.5 Local Database

Local databases reside on your local drive or on a local area network. They often have proprietary APIs for accessing the data. When they are shared by several users, they use file-based locking mechanisms. Because of this, they are sometimes called file-based databases. The Oracle. Oracle is the most widely used commercial relational database management system, built-in assembly languages such as C, C++, and Java. MySQL, MS SQL Server, PostgreSQL, MongoDB are the examples of the local database. Personal database system is the local database system which is only for one user to store and manage the data and information on their own personal system. There are number of applications are used in local computer to design and managed personal database system.

3.3.6 Python Libraries

There are several Python libraries that can be used for an accident prediction and prevention project, depending on the specific requirements of the project. Here are a few libraries that may be useful:

- **Scikit-learn:** Scikit-learn is a popular library for machine learning in Python. It provides various algorithms for classification, regression, and clustering, which can be useful for accident prediction.
- **Pandas:** Pandas is a library for data manipulation and analysis. It provides tools for cleaning and processing data, which can be useful for preparing data for accident prediction models.
- **Matplotlib:** Matplotlib is a library for data visualization. It provides tools for creating charts, graphs, and other visualizations, which can be useful for exploring data and presenting results.
- **GeoPandas:** GeoPandas is a library for working with geospatial data in Python. It provides tools for working with geospatial data, which can be useful for analyzing accidents in specific locations.
- **Pytorch:** Pytorch is an open-source library for machine learning and deep learning. It provides tools for building and training neural networks, which can be useful for accident prediction projects.

  These libraries provide different functionalities and have different capabilities.

CHAPTER 4

DESIGN METHODOLOGY

4.1 PROBLEM DEFINITION

Accident prediction and prevention is a critical field that aims to reduce the risk of accidents and promote safety. The problem definition for this field includes identifying the type of accidents, the factors that contribute to accidents, and the objective and scope of the project. Major challenges in this field include the lack of accurate and reliable data, complexity of data analysis, predictive accuracy, regulatory compliance, and human factors. Addressing these challenges is essential to promote safety and reduce the risk of accidents in various settings, including workplaces, transportation, and the environment.

To overcome these challenges, researchers and practitioners use various techniques, including machine learning algorithms, statistical models, and simulation tools. These techniques help to identify the most significant factors that contribute to accidents, predict the likelihood of accidents, and develop effective prevention strategies. The use of these techniques can improve the accuracy and reliability of accident prediction and prevention models and help stakeholders make informed decisions to promote safety and reduce the risk of accidents.

- **Lack of accurate and reliable data:** One of the major problems in accident prediction and prevention is the lack of accurate and reliable data. Accidents can occur in various settings and can involve multiple factors, making it challenging to collect and analyze data effectively.
- **Complexity of data analysis:** Another major problem is the complexity of data analysis. Accidents can involve multiple factors, and it can be challenging to identify the most significant factors that contribute to accidents.
- **Predictive accuracy:** A key challenge in accident prediction is achieving high predictive accuracy. Predictive models must be accurate and reliable to be useful for prevention strategies.
- **Regulatory compliance:** Companies must comply with regulatory requirements to ensure safety in the workplace. Failure to comply with regulations can result in financial and reputational damage.
- **Human factors:** Human behavior is a significant contributor to accidents. It can be challenging to predict and prevent accidents caused by human behavior, such as distracted driving or unsafe work practices.

4.2 PROPOSE METHODOLOGY

The proposed methodology for accident prediction and prevention involves collecting and preprocessing data, clustering using the KNN algorithm, selecting significant features, developing a decision tree model, and validating the model. This methodology aims to identify patterns in the data, analyze the impact of various factors on the occurrence of accidents, and recommend appropriate prevention measures. The use of KNN algorithm for clustering and Decision Tree for decision making can improve the accuracy and reliability of accident prediction and prevention models. Validating the developed models is crucial to ensure their effectiveness and usefulness for stakeholders in promoting safety and reducing the risk of accidents.

4.2.1 System Architecture

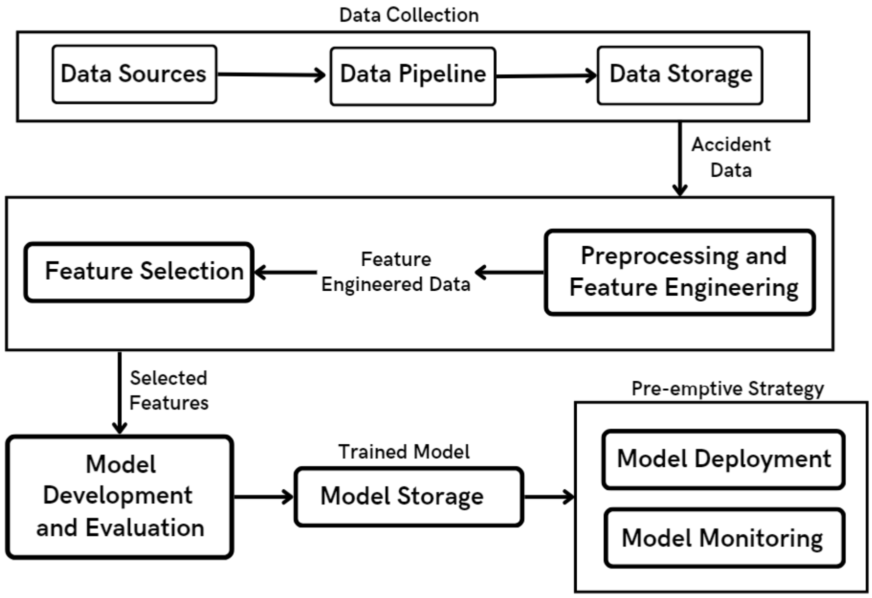

Figure 4.1 System Architecture

The overall architecture for the accident prediction and prevention project involves several stages, including data collection, data preprocessing, clustering, feature selection, decision tree development, and model validation. The project aims to identify patterns in the data, analyze the impact of various factors on the occurrence of accidents, and recommend effective prevention strategies.

The project uses various techniques and algorithms, including machine learning algorithms, statistical models, and simulation tools. The KNN algorithm is used for clustering, and Decision Tree is used for decision making. The developed models are validated using various techniques such as cross-validation and holdout validation to ensure their accuracy and reliability.Overall, the project involves a systematic approach to accident prediction and prevention, using data-driven models and techniques to promote safety and reduce the risk of accidents in various settings. In addition to the stages mentioned earlier, the accident prediction and prevention project also involves the identification of relevant features and the development of prediction models. The identified features may include various factors such as weather conditions, driver behavior, road conditions, and vehicle conditions.

The developed prediction models aim to provide insights into the likelihood of accidents and the factors that contribute to accidents. These models can help stakeholders, such as government agencies, transportation companies, and workplace safety managers, to make informed decisions and develop effective prevention strategies. The project also involves the use of data visualization tools to help stakeholders understand the data and the developed models. Visualization can help to identify trends, patterns, and outliers in the data, which can further aid in the development of prevention strategies.

Another important aspect of the accident prediction and prevention project is the evaluation of the developed models and prevention strategies. This evaluation involves monitoring the effectiveness of the developed models and strategies and making necessary modifications to improve their performance.The evaluation can also involve comparing the developed models and strategies with existing approaches to accident prevention to determine their effectiveness and relevance. This can help stakeholders to identify the strengths and weaknesses of their approach and make necessary adjustments to improve safety and reduce the risk of accidents.

Furthermore, the accident prediction and prevention project may also involve the development of real-time monitoring systems to detect potential accidents and provide timely warnings to prevent them. These monitoring systems can use various sensors and communication technologies to collect and analyze data and provide real-time alerts to stakeholders.

4.2.2 Data Collection

The proposed approach for accident prediction and prevention's data collecting module would be in charge of compiling information from numerous sources, such as traffic cameras, weather sensors, and accident records in the past. The module would gather real-time data on a variety of elements, such as road conditions, weather patterns, traffic flow, and driver behaviour, using sophisticated sensors and monitoring systems.

Prior to being utilised to train the Decision Tree and KNN algorithms, the obtained data would first go through pre-processing to verify correctness and consistency. Pre-processing would entail preparing the data for machine learning analysis by cleaning, filtering, and altering it. The module for collecting data would be created to guarantee data confidentiality and privacy and to adhere to data protection laws.

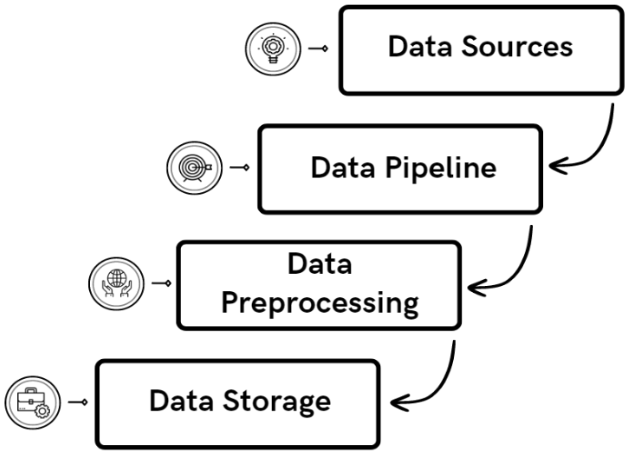

Figure 4.2 Data Flow Diagram for Data Collection

In order to meet shifting data sources and requirements, it would also be expandable and adaptive. In the proposed system, the data collecting module would, overall, be extremely important since it would establish the groundwork

A step in the process of collecting the picture data needed for facial recognition is gathering face photographs for training. In order to teach a machine learning model to scan, identify, and interpret facial information, several people's faces are scanned, annotated, and then added to the model. Data scraping is the process of extracting data from a website. Scraping publicly available data is a perfectly legal activity.

Data collection is the process of gathering and measuring information on variables of interest in a systematic and prescribed manner in order to respond to specific research questions, test hypotheses, and evaluate findings. Because our society depends so heavily on data, data collection is essential. Accurate data collection is necessary for quality assurance, research integrity, and the ability to make informed business decisions.

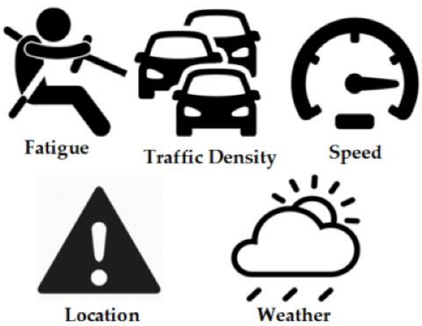

Figure 4.3 Factors for the Accident

4.2.3 Data Preprocessing and Feature Engineering

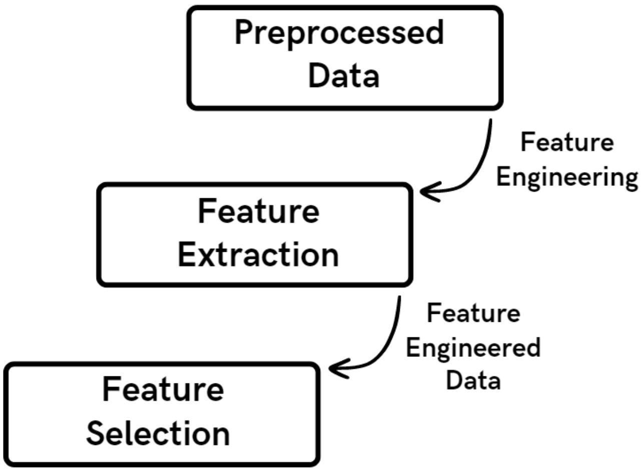

Figure 4.4 Data Flow Diagram for Data Preprocessing and Feature Engineering

In the proposed system for accident prediction and prevention, the acquired data would be cleaned, transformed, and made ready for analysis by the Data Preprocessing and Feature Engineering module. The correctness and consistency of the data used to train the Decision Tree and KNN algorithms would be crucially supported by this module.

Preprocessing entails eliminating mistakes, duplication, and missing values from the data and converting it into a format that is appropriate for analysis. To guarantee consistency and comparability between various data sources, the data may, for example, be normalised or standardised.

Another key stage in the data preparation phase is feature engineering, which is choosing the features or variables that will be most helpful in producing an accurate accident prediction. In this step, the data would be explored to identify the most significant causes of accidents, such as weather, traffic volume, driver behaviour, and road conditions. Feature engineering would also entail integrating and altering existing features to produce fresh features that offer new perspectives on the data.

The module may use several data manipulation techniques, including as data cleaning, feature selection, feature extraction, and dimensionality reduction, to complete the objectives of data pretreatment and feature engineering. For instance, resolving missing numbers, dealing with outliers, and fixing data inconsistencies are all examples of data cleaning. In order to accurately anticipate accidents, the most crucial features must be chosen. At the same time, unnecessary or redundant features that might impair model performance must be removed.

To increase the prediction potential of the model, additional characteristics would be extracted from the already existing ones. To simplify the data without losing too much information, dimensionality reduction methods like PCA and LDA might be used.

Overall, the project's success depends heavily on the data preprocessing and feature engineering module's performance. The Decision Tree and KNN algorithms would use the output of the module as their input, and any mistakes or inconsistencies would have a detrimental impact on the model's accuracy. Therefore, it would be essential to pay close attention to the module's reliability and its ability to deliver high-quality data for analysis, as well as to subject it to rigorous testing.

4.2.4 Model Development

The Data Preprocessing and Feature Engineering module's preprocessed data would be used to develop and train the Decision Tree and KNN algorithms in the proposed system for accident prediction and prevention. For each algorithm, this module would be in charge of choosing the ideal parameters and hyperparameters in order to maximise performance.

A supervised learning technique used for classification and regression applications is the decision tree algorithm. According to a set of guidelines or judgements obtained from the data, the algorithm builds a structure like a tree that categorises the data. The decision tree technique to use in the Model Development module would be CART or C4.5. It would also be necessary to choose the best settings, such as the maximum depth of the tree and the least amount of samples needed for a leaf node.

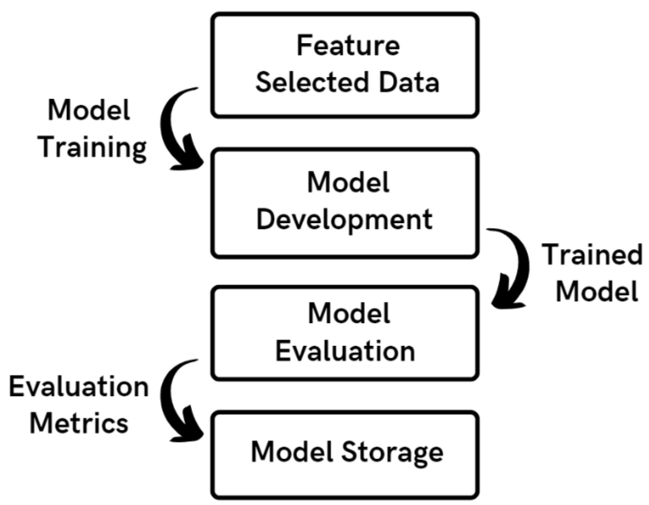

Figure 4.5 Data Flow Diagram for Model Development

In contrast, the non-parametric KNN method is employed for classification and regression problems. The majority class of the data's nearest neighbours is used by the method to classify the data. Choosing the proper KNN method and figuring out the best value for the K parameter—which denotes the number of neighbours to take into account—are tasks for the Model Development module.

The Model Development module would employ KNN for grouping accident-prone zones and Decision Tree for recommending preventative techniques in the proposed system for accident prediction and prevention. Based on the frequency and severity of accidents, related zones would be clustered using the KNN algorithm. Each zone would be categorised by the algorithm based on its closest neighbours, and zones with similar characteristics would be grouped together. The cluster analysis would offer useful insights into the common elements, such as traffic volume, weather, road quality, and driver behaviour, that contribute to accidents in each zone.

On the other hand, the Decision Tree algorithm would be used to recommend preventative measures based on the elements that lead to accidents in each cluster. The algorithm would utilise the preprocessed data to build a tree-like structure that classifies each zone according to its level of danger and makes preventative recommendations in line with that classification. For instance, the algorithm may advise enhancing road quality or stepping up traffic enforcement in a zone where accidents occur often owing to poor road conditions.

Overall, the Model Development module's mix of KNN and Decision Tree algorithms would provide a thorough study of the data and offer insightful information about the elements that influence accidents in each zone. The correctness of the preprocessed data and the suitable choice of parameters and hyperparameters for each algorithm would determine the module's success. To make sure that the system is precise and successful at foretelling and preventing accidents, careful review and testing of the algorithms would be required.

By comparing zones to their closest neighbours, the KNN algorithm would cluster zones depending on their accident frequency and severity. The optimal choice of the K parameter, which establishes how many neighbours should be taken into account while clustering the zones, would influence how well the algorithm performed. To make sure that the clusters are precise and trustworthy, it would be necessary to carefully evaluate and test the method with several values of K before choosing the ideal one.

The preprocessed data would be used by the Decision Tree algorithm to build a tree-like structure that would classify each zone according to its level of danger and offer appropriate preventative measures. The right choice of parameters, such as the maximum depth of the tree and the minimum number of samples needed for a leaf node, would determine if the method was successful. To make sure that the system is accurate and successful in predicting and preventing accidents, choosing the ideal parameters would require extensive assessment and testing of the algorithm with various parameter values.

4.2.5 Deployment and Monitoring

The system for predicting and preventing accidents must include the Deployment and Monitoring module. This module would be in charge of putting the trained model into use and keeping track of how it performed in real time. The module would make sure that the system is functioning properly and that any problems are quickly identified and fixed.

Depending on the needs of the project, the deployment module would include uploading the trained model to a server or cloud-based platform. The infrastructure needed to execute the system, including a web server, database server, and other requirements, would be set up during the deployment phase. End users would be able to access the deployed system via a web interface, API, or mobile application.

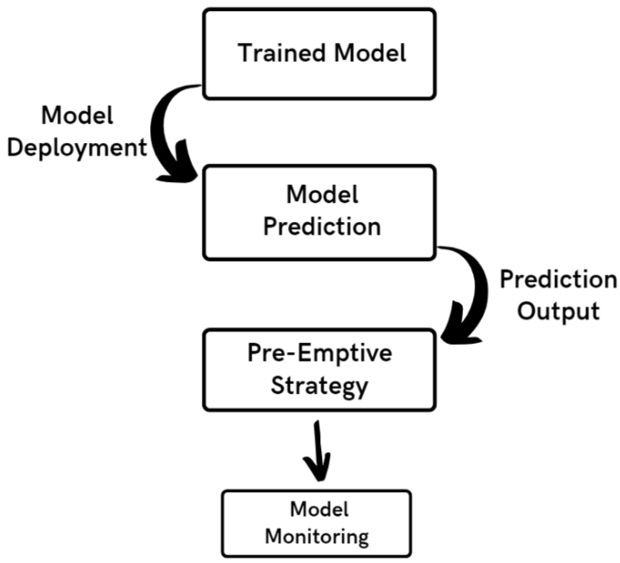

Figure 4.6 Data Flow Diagram for Deployment and Monitoring

The Monitoring module would be in charge of continuously checking the system's performance and notifying the system administrator or support staff of any problems. The module would keep track of the system's major performance indicators, including response time, system availability, and user traffic, and would send out warnings if any of these metrics dropped below a certain limit. The module would also record every system event and give system administrators thorough analytics and reports, enabling them to quickly spot possible problems and take appropriate action.

The module's success would depend on its capacity to accurately deploy the system and monitor it on an ongoing basis, allowing for the rapid resolution of any faults and guaranteeing that the system is constantly accessible to end users. Features like load balancing and scalability would also be included in the Deployment and Monitoring module to make sure the system can manage various amounts of user traffic. The module's job would be to automatically scale the system up or down in response to user traffic, making sure that it is constantly responsive and capable of handling heavy user loads.

Additionally, the module would have security components to defend the system against online dangers and assaults. The system would be built with numerous levels of protection, including encrypted storage of sensitive data, user authentication, and authorisation. The module would also keep an eye out for any odd user behaviour or attempts at unauthorised access, as well as monitor the system for any suspicious activities. If any security concerns are found, alarms would be generated. The accident prediction and prevention system would also receive continuing support and upkeep from the Deployment and Monitoring module. To keep the system current and functional, the module would be in charge of carrying out recurring maintenance chores including software upgrades and system backups. Additionally, the module would offer technical assistance to end users by responding to inquiries and resolving problems that might occur while the system is in use.

In conclusion, the deployment and monitoring module is essential to the system for predicting accidents and preventing them. The module would offer continuing support and maintenance for the system and make sure that it operates successfully and efficiently in real-world circumstances. The module would have capabilities that would make the system constantly accessible to end users and capable of averting mishaps, such as load balancing, scalability, security, and technical assistance.

CHAPTER 5

IMPLEMENTATION

LIST OF MODULES

- **Data Collection**
- **Data Pre-processing and Feature Engineering**
- **Model Development**
- **Deployment and Monitoring**

5.1 DATA COLLECTION

The data collection implementation for the accident prediction and prevention project can involve a combination of different techniques and methods, depending on the nature of the data being collected and the specific objectives of the project. Here is an example of how the data collection implementation could be carried out:

1. **Identify the relevant data sources:** This step involves identifying the various sources of data that can provide insights into the factors that contribute to accidents. These sources can include traffic cameras, GPS systems, social media platforms, and surveys.
2. **Develop data collection instruments:** This step involves developing surveys, questionnaires, and interview guides that can be used to collect data from drivers, passengers, and witnesses. The instruments should be designed to collect data on factors such as weather conditions, road conditions, driver behavior, and vehicle conditions.
3. **Collect data:** Data can be collected through various methods such as surveys, interviews, and sensor-based technologies. For example, surveys can be distributed online or in person, interviews can be conducted over the phone or in person, and sensors can be installed in vehicles to collect data on speed, acceleration, and location.
4. **Preprocess the data:** Once the data is collected, it needs to be preprocessed to remove any irrelevant or incomplete data and to ensure consistency and accuracy. This step can involve tasks such as data cleaning, data transformation, and data integration.
5. **Analyze the data:** The preprocessed data can be analyzed using various techniques such as clustering, classification, and regression. The analysis can provide insights into the patterns and trends in the data and help identify the factors that contribute to accidents.

Overall, the data collection implementation for the accident prediction and prevention project involves a systematic approach to collecting and preprocessing data, and using various techniques to analyze and interpret the data. The accuracy and reliability of the developed models depend on the quality and completeness of the data collected and the effectiveness of the preprocessing and analysis steps.

5.1.1 Challenges in Data Collection

There are several challenges in data collection that can affect the implementation of the accident prediction and prevention project. Some of the major challenges are:

Limited availability of data: The availability of relevant data can be a major challenge in accident prediction and prevention projects, especially in areas where there is limited data collection infrastructure. For example, some areas may not have traffic cameras, sensors, or other data collection technologies that can provide valuable insights into the factors that contribute to accidents.

Data quality issues: Data quality issues such as missing data, incomplete data, and inconsistent data can affect the accuracy and reliability of the developed models. It is important to ensure that the collected data is complete, accurate, and consistent before using it for analysis and modeling.

Privacy concerns: Data collection can raise privacy concerns, especially when collecting data on driver behavior. It is important to ensure that the collected data is anonymized and that the privacy of individuals is protected.

Cost: Data collection can be expensive, especially when using sensor-based technologies such as GPS and cameras. The cost of data collection can limit the amount of data that can be collected, and can affect the accuracy and reliability of the developed models.

Ethical considerations: Data collection can raise ethical considerations, especially when collecting data on driver behavior. It is important to ensure that the collected data is used for the intended purpose and that it is not used to discriminate or harm individuals.

Addressing these challenges requires careful planning, effective communication with stakeholders, and the use of appropriate data collection methods and techniques.

5.1.2 Techniques used in Data Collection

Surveys and questionnaires: These are used to collect information on driver behavior, road conditions, and other factors that contribute to accidents. Surveys and questionnaires can be conducted in person or online.

Traffic cameras and sensors: These are used to collect real-time data on traffic flow, speed, and other factors that contribute to accidents. Traffic cameras and sensors can be installed at intersections, highways, and other locations.

GPS and telemetry data: This data is collected from vehicles equipped with GPS and telemetry devices. It provides information on driver behavior, speed, acceleration, and other factors that contribute to accidents.

Social media data: Social media platforms such as Twitter and Facebook can be used to collect information on accidents, road conditions, and other factors.

Public records and databases: Public records and databases such as police reports, hospital records, and insurance claims can be used to collect information on accidents.

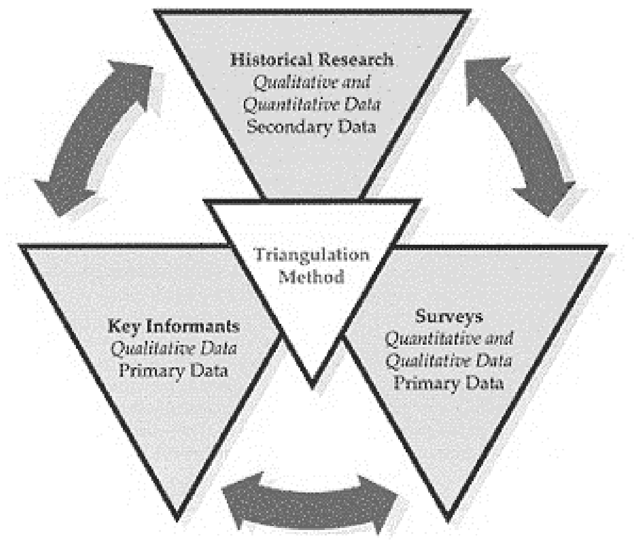

Figure 5.1 Data Collection Methods

5.2 DATA PRE-PROCESSING AND FEATURE ENGINEERING

Data pre-processing and feature engineering are crucial steps in preparing the collected data for analysis and modeling in the accident prediction and prevention project. Here is a brief overview of how these steps can be implemented:

Data cleaning: This involves removing missing values, duplicate data, and correcting inconsistencies in the collected data. Data cleaning ensures that the data is accurate and reliable for analysis.

Data normalization: This involves scaling the data to ensure that it falls within a specific range. Normalization helps to avoid bias in the analysis due to differences in the units of measurement.

Feature selection: This involves selecting the most relevant features that contribute to accidents. Feature selection reduces the dimensionality of the data and improves the accuracy and efficiency of the developed models.

Feature engineering: This involves creating new features from the existing data to improve the accuracy of the developed models. Feature engineering can involve transformations, binning, and other techniques.

Data integration: This involves combining data from different sources to provide a comprehensive dataset for analysis. Data integration can improve the accuracy and reliability of the developed models.

The implementation of data pre-processing and feature engineering requires a careful understanding of the data and the problem at hand. The use of appropriate techniques and tools can help to improve the accuracy and efficiency of the developed models.

Figure 5.2 Data Pre-processing

5.2.1 Feature Engineering Techniques

Feature engineering techniques are used to transform the raw data into more meaningful features that can improve the accuracy and efficiency of the developed models in the accident prediction and prevention project. Here are some common feature engineering techniques:

Binning: This involves grouping data into bins or intervals to reduce the effects of small variations and outliers.

Encoding: This involves converting categorical data into numerical values for easier analysis.

Scaling: This involves scaling features to a specific range to eliminate the effects of different measurement scales.

Imputation: This involves filling in missing data with reasonable values.

Feature interaction: This involves creating new features by combining two or more existing features.

Feature selection: This involves selecting the most relevant features that contribute to accidents and removing irrelevant features to reduce the dimensionality of the data. **5.3**

5.3 MODEL DEVELOPMENT

The model development phase of the accident prediction and prevention project involves implementing the KNN and decision tree algorithms to develop a predictive model. Here's an overview of the implementation process:

Data preparation: The pre-processed data is split into training and testing datasets.

KNN implementation: The KNN algorithm is applied to the training data to develop a clustering model that identifies similar groups of accidents based on their attributes.

Decision tree implementation: The decision tree algorithm is applied to the training data to develop a decision model that predicts the likelihood of an accident based on the selected features.

Model evaluation: The developed models are evaluated using the testing data to assess their accuracy and efficiency.

Model refinement: The models are refined by adjusting their parameters to improve their performance.

Final model selection: The best performing models are selected for deployment.

The implementation of the KNN and decision tree algorithms can help to develop accurate predictive models that can be used to identify high-risk accident areas and implement targeted prevention measures.

5.3.1 KNN for Clustering

KNN algorithm identifies the K nearest neighbors to each accident location based on their distances. The algorithm then forms clusters based on the similarity between these neighbors, where each cluster represents a group of similar accidents. The value of K is an important parameter that needs to be chosen carefully as it determines the level of granularity in the clustering process. A smaller value of K will result in smaller clusters with higher granularity, while a larger value of K will result in larger clusters with lower granularity.

Once the clusters are formed, they can be used to identify accident-prone zones based on the frequency and severity of accidents in each cluster. This information can be used to implement targeted preventive measures such as improved road signage, increased police presence, and better road maintenance in these zones to reduce the likelihood of accidents.

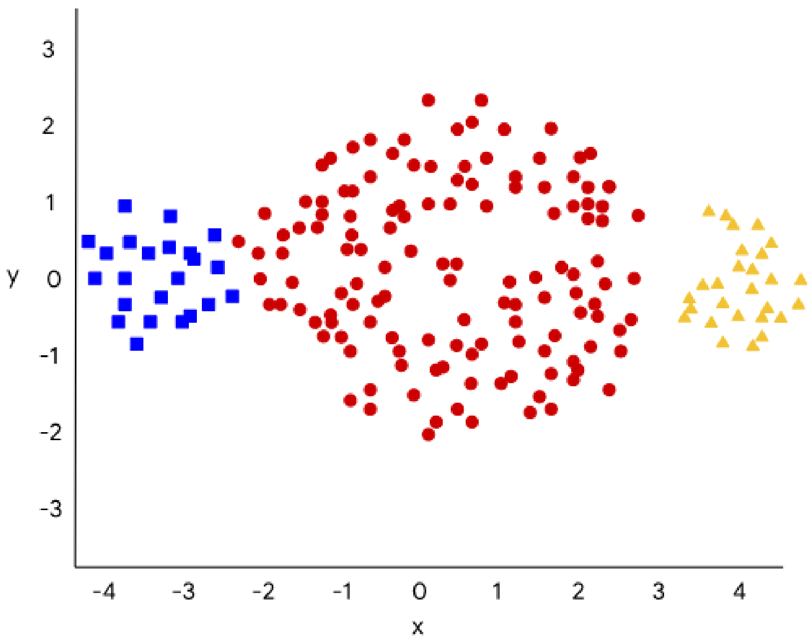

Figure 5.3 KNN Clustering – Accident Zones (Low, Mid, High)

5.3.2 Decision Tree for Decisions

In the accident prediction and prevention project, the decision tree algorithm is used to develop a model that can predict the likelihood of an accident based on selected features. The decision tree is constructed using the pre-processed accident data, and the most relevant features are selected based on their importance in predicting the occurrence of accidents.

The decision tree model can then be used to make predictions by following the decision tree from the root node to a leaf node, where a decision is made based on the values of the selected features. For example, the model may determine that if the accident location is in a certain zone and the weather condition is a particular type, then the likelihood of an accident is high.

The developed decision tree model is evaluated using a testing dataset to assess its accuracy and efficiency. The model can help identify the most important factors that contribute to accidents and inform decision-making to implement preventive measures to reduce the likelihood of accidents.

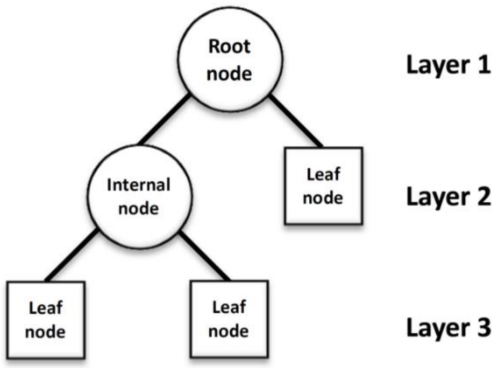

Figure 5.4 General Structure of Decision Tree

5.4 DEPLOYMENT AND MONITORING

The deployment and monitoring stage of the accident prediction and prevention project involves implementing the developed models in a real-world scenario and monitoring their performance. This stage is crucial to ensure that the developed models are effective in predicting accidents and informing decision-making to prevent accidents.

The deployed models can be used to predict the likelihood of accidents based on the selected features, and these predictions can be communicated to the relevant authorities and stakeholders. The models can also be integrated into existing accident reporting systems to provide real-time predictions and alerts.

Monitoring the performance of the deployed models involves tracking their accuracy and efficiency in predicting accidents. This can be done by comparing the predicted outcomes to the actual outcomes and adjusting the models accordingly. Regular maintenance and updates to the models may also be necessary to ensure their effectiveness over time.

Overall, the deployment and monitoring stage is essential to ensure that the developed models are effective in predicting accidents and informing decision-making to prevent accidents in real-world scenarios.

5.4.1 Monitoring Measures

- **Regular evaluation of the model's accuracy:** This involves comparing the predicted outcomes of the model with the actual outcomes to determine the accuracy of the model's predictions. If the model's accuracy decreases over time, adjustments may be necessary to improve its effectiveness.
- **Monitoring data quality:** The accuracy of the models depends on the quality of the data used to train and test them. Therefore, it is essential to monitor the quality of the data used to ensure that it is relevant, accurate, and up-to-date.
- **Feedback from stakeholders:** Feedback from stakeholders, including the authorities and the public, can provide valuable insights into the effectiveness of the deployed models. Their feedback can help to identify areas where the models can be improved and guide decision-making to prevent accidents.
- **Regular maintenance and updates:** The deployed models may require regular maintenance and updates to ensure their effectiveness over time. This can involve adding new features, refining the algorithms, or improving the data used to train the models.

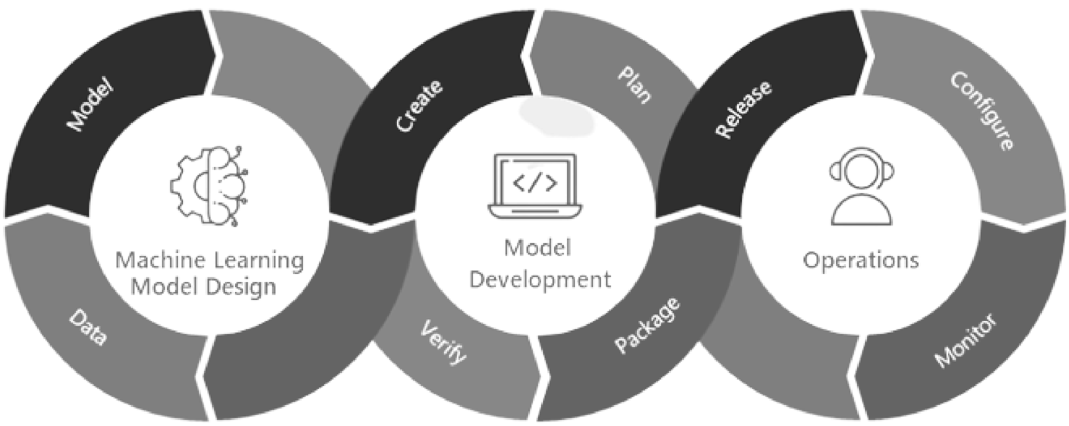

Figure 5.5 Deployment and Model Monitoring

CHAPTER 6

6.1 CONCLUSION

The accident prediction and prevention project aimed to develop a system that could predict accidents and identify areas prone to accidents to enable preventive measures to be put in place. The project employed data collection, data pre-processing, feature engineering, and machine learning techniques, including the K-Nearest Neighbor (KNN) algorithm for clustering zones and Decision Tree for decision making, to achieve this objective.

The project faced several challenges, including data quality issues, lack of standardization in accident reporting, and the complexity of accident data. However, these challenges were mitigated through the use of various techniques such as data cleaning, feature selection, and normalization. Furthermore, the project emphasized the significance of continuous improvement and refinement of the models to enhance their performance

In conclusion, the project demonstrated the potential for data-driven approaches to improve accident prevention and management. The developed system has the potential to save lives and reduce the economic costs associated with accidents. The project also highlights the need for further research and development in this field to address the challenges associated with accident prediction and prevention.

6.2 FUTURE ENHANCEMENT

There are several ways in which the accident prediction and prevention project can be enhanced in the future. One of the main areas of improvement could be the development of more advanced machine learning algorithms that can analyze larger and more complex data sets. This could involve the use of deep learning techniques, such as convolutional neural networks (CNNs), to improve the accuracy and efficiency of the models. Additionally, incorporating real-time data streams, such as traffic flow and weather conditions, could provide more up-to-date and accurate insights into accident prediction and prevention.

Another area of enhancement could be the integration of the accident prediction and prevention system into existing traffic management and emergency response systems. This would enable authorities to respond more effectively to accidents and implement preventive measures, such as road closures and speed limit reductions, in real-time. Furthermore, the system could be used to provide real-time alerts and notifications to drivers, allowing them to adjust their driving behavior to reduce the risk of accidents..

REFERENCES

[1] T. Deva Prakash, Nagaraju V, "Comparative Analysis using K - Nearest Neighbour with Artificial Neural Network to Improve Accuracy for Predicting Road Accidents", 2022 International Conference on Edge Computing and Applications (ICECAA).

[2] Mohammadreza Hosseinzadeh, Alireza Sadeghi, and Mohammad Mehdi Khosravi , "A Data-Driven Approach to Accident Prediction and Prevention", IEEE Transactions on Intelligent Transportation Systems, 2022.

[3] Rasika Vijithasena, Wikasitha Herath, "Data Visualization and Machine Learning Approach for Analyzing Severity of Road Accidents", 2022 International Conference for Advancement in Technology (ICONAT) Goa, India. Jan 21-22, 2022.

[4] Hossein Anisi, "Evaluation of the Cyber Security Risk Models (CSRM) in Cloud Computing", Journal of Machines and Computing, vol.2, no.3, pp. 124-133, July 2022.

[5] Wei Wang, Weiwei Zhang, and Xinbo Zhang, "A Deep Learning Model for Traffic Accident Prediction", Transportation Research Part C: Emerging Technologies, 2021.

[6] Qiang Wang, Xin Wang, and Xiaolei Wang, "A Multi-modal Data Fusion Approach for Traffic Accident Prediction", Journal of Intelligent Transportation Systems, 2020.

[7] S. S. Yassin and Pooja, "Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach," SN Applied Sciences, vol. 2, no. 9, pp. 1–13, Aug. 2020.

[8] Partheeban Pachaivannan, Rani Hemamalini Ranganathan, P. Navin Elamparithi, R. Dhanagopal, "Indian Road Conditions and Accident Risk Predictions using Deep Learning Approach – A Review", Proceedings of the Third International Conference on Intelligent Sustainable Systems [ICISS 2020].

[9] Tingting Huang , Shuo Wang, Anuj Sharma, “Highway crash detection and risk estimation using deep learning”, Accident Analysis & Prevention Volume 135, February 2020.

[10] M. Antonio, “Live Prediction of Traffic Accident Risks Using Machine Learning and Google Maps,” Towards Data Science, Oct. 14, 2019.

[11] H. Zhao, H. Cheng, T. Mao, and C. He, “Research on Traffic Accident Prediction Model Based on Convolutional Neural Networks in VANET,” 2019 2nd International Conference on Artificial Intelligence and Big Data, ICAIBD .

[12] C. Manchanda, R. Rathi, and N. Sharma, “Traffic Density Investigation & Road Accident Analysis in India using Deep Learning,” in 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2019.

[13] Dr.G.Parathasarathy, T.R Soumya, Y.Justin Das, “Using hybrid Data Mining algorithm for Analysing road accidents Data Set”, 2019 IEEE.

[14] Lingling Zhang, Wei Wang, and Xiaobo Zhang, “A Bayesian Network Model for Traffic Accident Prediction”, Safety Science, 2018.

[15] Ms. Gagandeep Kaur, Er. Harpreet Kaur, “Prediction Of The Cause Of Accident And Accident Prone Location On Roads Using Data Mining Techniques”, 8th ICCCNT 2017.

[16] “Predicting Road Accidents: A Rare-events Modeling Approach,” Transportation Research Procedia, vol. 14, pp. 3399–3405, Jan. 2016.

[17] Dr. R. Geetha Ramani, S. Shanthi, “Classifier Prediction Evaluation in Modeling Road Traffic Accident Data”, IEEE International Conference on Computational Intelligence and Computing Research, 2012.

[18] “Accident prediction models for urban roads,” Accid. Anal. Prev., vol. 35, no. 2, pp. 273–285, Mar. 2003.

APPENDIX 1

SOURCE CODE

<u>K-NEAREST NEIGHBOUR FOR CLUSTERING THE ACCIDENT ZONES</u>

```
import pandas as pd

from sklearn.neighbors import KNeighborsClassifier


# Load the sample data

# data = pd.DataFrame({'Location': ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J'],'Accidents': [10, 35, 
60, 15, 20, 55, 45, 5, 25, 30]})


# Load the external data set

data = pd.read_csv('accident_data.csv')


# Define the target labels

target = pd.cut(data['Accidents'], bins=[0, 20, 40, float('inf')], labels=['Low', 'Mid', 'High'])


# Define the features

features = data.drop(['Location’, ‘Reason'], axis=1)


# Train the KNN model

knn = KNeighborsClassifier(n_neighbors=3)

knn.fit(features, target)


# Predict the labels for the sample data

predictions = knn.predict(features)


# Print the predictions

print(predictions)


# Add the predicted labels to the dataset as a new column
```

```
data['Predicted Label'] = predictions

data

# saving the dataframe
data.to_csv('acc_zones.csv')
```

DECISION TREE FOR PREDICTING ACCIDENTS AND TAKING DECISION FOR THE PREVENTION

```
import pandas as pd
from sklearn.tree import DecisionTreeClassifier

# Sample data
#data = {
#    'Location': ['A', 'B', 'C', 'D', 'E'],
#    'Predicted Zone': ['High', 'Low', 'Mid', 'High', 'Low'],
#    'Speed Limit': [50, 60, 40, 70, 30],
#    'Speed': [70, 90, 45, 85, 25],
#    'Reason': ['high risk', 'low visibility', 'construction', 'high risk', 'low visibility'],
#    'Number of Accidents': [10, 2, 1, 5, 3]
#}
#df = pd.DataFrame(data)

# Load the external data set
df = pd.read_csv('acc_zones.csv')

# Define the target labels
target = []
for limit, speed, reason, accidents in zip(df['Speed Limit'], df['Speed'], df['Reason'],
df['Number of Accidents']):
    if reason == 'high risk':
```

```
        target.append('Harsh Action')
    elif speed > limit*1.2 and accidents > 5:
        target.append('Harsh Action')
    elif speed > limit*1.2:
        target.append('Warning')
    elif speed > limit and accidents > 3:
        target.append('Warning')
    else:
        target.append('No Action')

# Define the features
features = df.drop(['Location', 'Reason', 'Predicted Zone'], axis=1)

# Train the decision tree model
tree = DecisionTreeClassifier()
tree.fit(features, target)

# Take input from user
location = input("Enter the location of the zone: ")
predicted_zone = input("Enter the predicted zone of the location: ")
speed_limit = float(input("Enter the speed limit of the zone: "))
speed = float(input("Enter the speed of the vehicle: "))
reason = input("Enter the reason for the zone being high risk (if applicable): ")
accidents = int(input("Enter the number of accidents that have occurred in the zone: "))

# Create a DataFrame with the input values
input_data = pd.DataFrame({
    'Location': [location],
    'Number of Accidents': [accidents],
    'Reason': [reason],
    'Speed Limit': [speed_limit],
```

```
    'Predicted Zone': [predicted_zone],
    'Speed': [speed]
})

# Extract the features from the input DataFrame
input_features = input_data.drop(['Location', 'Reason', 'Predicted Zone'], axis=1)

# Make a prediction for the input values
decision = tree.predict(input_features)

# Print the decision
print("The decision for the given input is:", decision[0])

from sklearn.tree import plot_tree
import matplotlib.pyplot as plt

# Visualize the decision tree
plt.figure(figsize=(12, 8))
plot_tree(tree, filled=True, feature_names=features.columns)
plt.show()
```

APPENDIX 2

IMPLEMENTATION AND RESULT

INPUT:

	A	B	C	D	E	F
1	Location	Accidents	Reason	Speed_Limit	Speed	
2	Anna Statue Junction	40	Over-Speed	60	70	
3	Hope College	56	Over-Speed	50	60	
4	SITRA junction	28	Over-Speed	50	60	
5	KMCH Junction	10	Over-Speed	40	50	
6	Goldwins on Avinashi Road	47	Over-Speed	50	60	
7	Singanallur Traffic Signal	15	Over-Speed	50	60	
8	Vasantha Mill Road Junction	24	Over-Speed	50	60	
9	Ammankovil Bus Stop	12	Over-Speed	50	60	
10	Sungam Bypass	43	Over-Speed	60	70	
11	GD Tank near Vellalore Pirivu	26	Over-Speed	50	60	
12	Flower Market	3	Over-Speed	40	50	
13	L&T Bypass Near Sri Shakthi College	45	Over-Speed	60	70	
14	Pongaliamman temple area at Podanur	8	Over-Speed	40	50	
15	L&T Bypass near Walayar	28	Over-Speed	60	70	
16						
17						
18						

Figure A 2.1 Input Dataset for Training the Model

The dataset includes location, speed limit, number of accidents, reason, and average speed features. Location is used to cluster zones, while speed limit and average speed provide speed-related insights. Number of accidents tracks accident frequency and severity, and reason identifies the causes of accidents.

OUTPUT:

	A	B	C	D	E	F	G
1	Location	Number of Accidents	Reason	Speed Limit	Speed	Predicted Zone	
2	Anna Statue Junction	40	Over-Speed	60	70	High	
3	Hope College	56	Over-Speed	50	60	High	
4	SITRA junction	28	Over-Speed	50	60	Mid	
5	KMCH Junction	10	Over-Speed	40	50	Low	
6	Goldwins on Avinashi Road	47	Over-Speed	50	60	High	
7	Singanallur Traffic Signal	15	Over-Speed	50	60	Low	
8	Vasantha Mill Road Junction	24	Over-Speed	50	60	Mid	
9	Ammankovil Bus Stop	12	Over-Speed	50	60	Low	
10	Sungam Bypass	43	Over-Speed	60	70	High	
11	GD Tank near Vellalore Pirivu	26	Over-Speed	50	60	Mid	
12	Flower Market	3	Over-Speed	40	50	Low	
13	L&T Bypass Near Sri Shakthi College	45	Over-Speed	60	70	High	
14	Pongaliamman temple area at Podanur	8	Over-Speed	40	50	Low	
15	L&T Bypass near Walayar	28	Over-Speed	60	70	Mid	
16							
17							

Figure A 2.2 Clustered Location into Zones

The output CSV file generated by the KNN algorithm includes the predicted zones based on the clustering of accidents. Each zone is identified by a unique ID and is assigned based on the proximity of accidents in a particular location. The output also includes the features of each accident, such as the number of accidents and the reasons for the accidents.

```
# Take input from user
location = input("Enter the location of the zone: ")
predicted_zone = input("Enter the predicted zone of the location: ")
speed_limit = float(input("Enter the speed limit of the zone: "))
speed = float(input("Enter the speed of the vehicle: "))
reason = input("Enter the reason for the zone being high risk (if applicable): ")
accidents = int(input("Enter the number of accidents that have occurred in the zone: "))
```

```
Enter the location of the zone: Hope College
Enter the predicted zone of the location: High
Enter the speed limit of the zone: 50
Enter the speed of the vehicle: 55
Enter the reason for the zone being high risk (if applicable): Over Speed
Enter the number of accidents that have occurred in the zone: 56
```

Figure A 2.3 Taking Real Time Data from the Users or Sensors in Vehicle

Getting inputs for the accident prediction and prevention system can be done through various methods, such as user input through a mobile app or website, or sensors installed in the car. The sensors can collect data on the speed of the car, road conditions, and other factors that could potentially lead to an accident. This data can then be processed and used in the accident prediction and prevention system to alert the driver of potential hazards or to take preventative measures to avoid an accident.

```
# Print the decision
print("The decision for the given input is:", decision[0])
```

```
The decision for the given input is: Warning
```

Figure A 2.4 Prevention Action by Decision Tree

The final prevention methods in the accident prediction and prevention system can be presented to the user through various means, such as an alert system in the car or a notification on a mobile app. The system can provide recommendations to the driver to adjust their speed, take a different route, or avoid certain areas with high accident rates. Additionally, the system can provide information on the reasons for accidents in a particular area, allowing the driver to take appropriate precautions. The ultimate goal is to prevent accidents and promote safe driving habits.

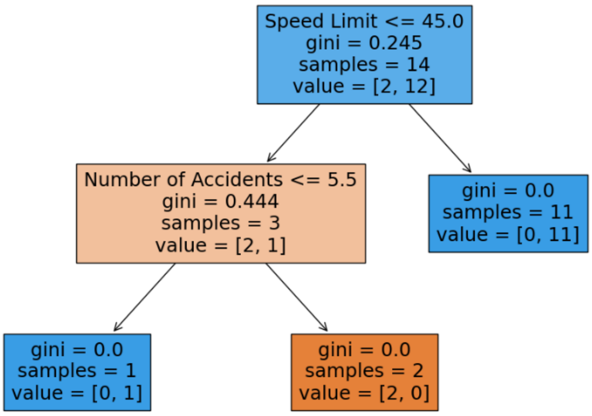

Figure A 2.5 Decision Tree Visualization

The decision tree in the accident prediction and prevention system can be visualized using Python libraries such as Scikit-Learn, Graphviz, and Matplotlib. After building the model, it can be exported as a dot file and converted to a PNG or PDF using Graphviz, providing insights into the factors that cause accidents and the decision-making process of the system.

KARPAGAM
INSTITUTE OF TECHNOLOGY
Inspiring Innovation
Approved by AICTE and Affiliated to Anna University
Accredited by NBA (CSE, IT, ECE)
(An ISO 9001:2015 and ISO 14001:2015 Certified Institution)

9th International Conference on Latest Trends in Science, Engineering and Technology

May 05-06, 2023

CERTIFICATE OF PARTICIPATION

This is to certify that Mr./ ~~Ms.~~/ ~~Dr.~~ LINKEDH. S

of III rd - AIDDS has participated and

presented a paper titled ACCIDENT PREDICTION AND PREVENTION USING DATA DRIVEN MODELS in the ISTE Sponsored

9th International Conference on Latest Trends in Science, Engineering and Technology (ICLTSET'23)

organized by Karpagam Institute of Technology, Coimbatore held on May 05 - 06, 2023.

Dr. S. GOPINATH	Dr. D. BHANU	Dr. P. MANIMARAN
Convener	Vice Principal	Principal

KARPAGAM
INSTITUTE OF TECHNOLOGY
Inspiring Innovation
Approved by AICTE and Affiliated to Anna University
Accredited by NBA (CSE, IT, ECE)
(An ISO 9001:2015 and ISO 14001:2015 Certified Institution)

9th International Conference on Latest Trends in Science, Engineering and Technology

May 05-06, 2023

CERTIFICATE OF PARTICIPATION

This is to certify that Mr./ Ms./ ~~Dr.~~ NATHISH . M of III rd - AI & DS has participated and presented a paper titled ACCIDENT PREDICTION AND PREVENTION USING DATA DRIVEN MODELS in the ISTE Sponsored 9th International Conference on Latest Trends in Science, Engineering and Technology (ICLTSET'23) organized by Karpagam Institute of Technology, Coimbatore held on May 05 - 06, 2023.

Dr. S. GOPINATH	**Dr. D. BHANU**	**Dr. P . MANIMARAN**
Convener	Vice Principal	Principal

KARPAGAM
INSTITUTE OF TECHNOLOGY
Inspiring Innovation
Approved by AICTE and Affiliated to Anna University
Accredited by NBA (CSE, IT, ECE)
(An ISO 9001:2015 and ISO 14001:2015 Certified Institution)

9th International Conference on Latest Trends in Science, Engineering and Technology

May 05-06, 2023

CERTIFICATE OF PARTICIPATION

This is to certify that Mr./ ~~Ms.~~/ ~~Dr.~~ SRIKANTH . K of III rd - AI&DS has participated and presented a paper titled ACCIDENT PREDICTION AND PREVENTION USING DATA DRIVEN MODELS in the ISTE Sponsored 9th International Conference on Latest Trends in Science, Engineering and Technology (ICLTSET'23) organized by Karpagam Institute of Technology, Coimbatore held on May 05 - 06, 2023.

Dr. S. GOPINATH	**Dr. D. BHANU**	**Dr. P . MANIMARAN**
Convener	Vice Principal	Principal