A PREDICTIVE MODEL OF ROAD TRAFFIC ACCIDENTS USING SURROGATE SAFETY MEASURES

By **Prithviraj – M22AI603**

Under the supervision of **Dr. Ranju Mohan**

in partial fulfilment of the requirements for the award of the degree of

MTech in Data and Computational Science

Indian Institute of Technology Jodhpur

Data and Computational Science

May, 2024.

Declaration

I hereby declare that the work presented in this Project Report titled **A Predictive Model of Road Traffic Accident Using Surrogate Safety Measures** submitted to the **Indian Institute of Technology Jodhpur** in partial fulfilment of the requirements for the award of the degree of **MTech in Data and Computational Science**, is a bonafide record of the research work carried out under the supervision of **Dr. Ranju Mohan**. The contents of this Project Report in full or in parts, have not been submitted to, and will not be submitted by me to, any other Institute or University in India or abroad for the award of any degree or diploma.

Prithviraj

M22AI603

Certificate

This is to certify that the Project Report titled **A Predictive Model of Road Traffic Accident Using Surrogate Safety Measures**, submitted by **Prithviraj (M22AI603)** to the **Indian Institute of Technology Jodhpur** for the award of the degree of **MTech in Data and Computational Science**, is a bonafide record of the research work done by him under my supervision. To the best of my knowledge, the contents of this report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Ranju Mohan

Abstract

This project presents a comprehensive approach to develop a predictive model for road traffic accidents using Surrogate Safety Measures (SSM). The primary objective is to estimate the likelihood of accidents at both micro and macro levels, encompassing intersections and mid-blocks, as well as identifying accident-prone areas within city networks. SSM, derived from various traffic-related variables including road geometry, traffic flow characteristics, and driver behaviour, serve as indirect indicators for potential crash occurrences. A predictive model is developed using machine learning algorithms, adapting to the complexity of the data and desired accuracy. The performance of the model is evaluated through metrics like accuracy, precision, recall, and F1 score, which gauges its reliability in estimating accident risks. Furthermore, the model is deployed into practical applications, potentially integrated into traffic management systems, providing real-time accident risk predictions to drivers and relevant stakeholders. This research offers a holistic framework for predicting road traffic accidents, contributing to improved road safety and accident prevention.

Contents

Abstract

1. Introduction 1
2. Background 2
3. Problem Statement 2
4. Existing System 3
5. Proposed System 3
6. Surrogate Safety Measures (SSM) 3
7. Methodology 10
8. Result 14
9. Conclusion 18
10. Future Enhancement 18

References

1. INTRODUCTION

Addressing road traffic accidents on a global scale is of paramount importance due to the substantial loss of life, injuries, and economic damages they cause. In pursuit of enhanced road safety, various predictive models have been developed to reduce the occurrence of accidents. This research project introduces an innovative approach to road safety, focusing around the creation of a predictive model that utilizes Surrogate Safety Measures (SSM) to gauge the likelihood of accidents.

Unlike conventional accident prediction models, which heavily depend on historical accident data, this project incorporates SSM, encompassing traffic flow characteristics, road geometry, and driver behaviour. The aim is to proactively identify accident-prone areas, both at micro-levels such as intersections and mid-blocks, and at a macro-level within city networks. This approach has the potential to facilitate timely interventions and pre-emptive measures, thereby reducing the frequency of accidents.

The significance of Surrogate Safety Measures (SSM) in traffic safety evaluation arises from the lack of reliable statistical safety models in many scenarios. This is particularly relevant for transportation facilities with complex site characteristics and/or nontraditional traffic safety treatments, where historical crash data may be limited or unavailable for developing safety predictive models [6]. Research on SSM dates back to the early 1970s [3], and significant progress has been achieved in this field since then.

Lee et al [8] developed a probabilistic model relating significant crash precursors to changes in crash potential. Abdel [9] built a previous crash prediction model with the matched case-control logistic regression technique. No specific approach available for the traffic police to predict which area is accident prone at a specific time. The traffic accident prediction plays an important role in the integrated planning and management of traffic, the reason which with much randomness about the traffic accident include some nonlinear elements, such as people, car, road, climate and so on. The traditional way of linear analyses cannot reveal the really situation since the noise pollution and amount of data are too little, cause the result of prediction cannot satisfactory.

The ultimate objective of this research is to contribute to road safety by providing a predictive model that offers real-time risk estimations for accidents. Such a transportation model can seamlessly integrate into traffic management systems, providing valuable insights to drivers, traffic authorities, and other stakeholders. This integration aims to facilitate safer road usage and prevent accidents. This research aspires to be a pioneering step towards establishing a safer and more efficient road system.

2. BACKGROUND

Road accidents, on a scale are a concern when it comes to public health and safety. According to the World Health Organization (WHO) [4] these accidents lead to millions of deaths and injuries every year. Not do they cause harm but they also have significant social and economic consequences for societies around the world. Fortunately, many road accidents can be. Identifying risks early on is crucial in minimizing their impact.

To understand the likelihood of an accident occurring researchers have turned to Surrogate Safety Measures (SSM). These measures act as indicators of crashes since collecting and analysing actual crash data can be challenging and time consuming. SSM consider factors such as traffic flow characteristics, road layout, weather conditions, driver behaviour, and other variables that influence the likelihood of road accidents.

Developing models for road accident prediction holds promise. It provides insights for traffic authorities, transportation planners and stakeholders in implementing targeted safety measures allocating resources efficiently and ultimately reducing the occurrence of road accidents. By embracing data analysis techniques predicting road accidents using safety measures has gained attention as an approach, to enhancing road safety and saving lives.

3. PROBLEM STATEMENT

Addressing road accidents is a significant concern for public health and safety, and predicting accident risks early on is essential for minimizing their impact. However, the conventional reliance on actual crash data, such as First Inspection Reports (FIR), for accident prediction is not only time-consuming but also inefficient. There is a growing necessity to create predictive models capable of estimating the likelihood of road accidents using surrogate safety measures, indirect indicators that signal potential crash occurrences.

Within the realm of mobility, road safety emerges as a critical aspect with the primary goal of preventing accidents and mitigating their consequences. The problem statement encompasses key elements related to road safety, including accident prediction, accident prevention, and data analysis for safety improvement. In essence, the problem statement aligns with the broader objective of enhancing road safety by developing predictive models that leverage surrogate safety measures. This proactive approach aims to prevent road accidents, optimize resource allocation, and facilitate evidence-based decision-making to improve road safety.

4. EXISTING SYSTEM

No specific approach available for the traffic police to predict which area is accident prone at a specific time. The traditional Back propagation network has defects. It has a 17% lower accuracy than the proposed model. We propose the use of a machine learning technique. Machine learning has the ability to model complex non-linear phenomenon.

5. PROPOSED SYSTEM

An ML powered web app which predicts accidents severity based on the current conditions. It is trained with huge volume of historical data. More data means greater accuracy. The purpose of such a model is to be able to predict which conditions will be more prone to accidents, and therefore take preventive measures. We will even try to locate more precisely future accidents in order to provide faster care and precaution service. According to the predicted severity, a message will be sent to the traffic police to take preventive measures.

6. SURROGATE SAFETY MEASURES (SSM)

The assessment of safety performance in designs, countermeasures, or systems often relies on crash frequency and severity, crucial indicators of their effectiveness. However, crashes are infrequent events, making it impractical and, to some extent, ethically questionable to rely solely on historical crash data for evaluating the performance of new safety strategies, such as the introduction of a new traffic sign. In response to this challenge, Surrogate Safety Measures (SSM) derived from traffic conflicts have gained popularity as an alternative.

Traffic conflicts represent observable non-crash events where interactions among multiple road users create a risk of collision unless their courses of movement are altered [1]. A conflict is deemed etiologically connected to a crash when a failure (e.g., human operator failure, road failure, vehicle failure) leading to the conflict cannot be adequately corrected [2] [7]. Due to this causal relationship, measures employed to identify traffic conflicts and assess their severities can be considered as SSM. Notably, traffic conflicts are more frequent compared to actual crashes.

It is important to highlight that numerous safety-related measures have been developed over time, but not all qualify as SSM. According to [6], two criteria for SSM qualification are:

- It must be derived from traffic conflicts directly linked to crashes.
- The relationship between traffic conflicts and the potential crash frequency and/or severity can be quantified using practical methods.

From this standpoint, traffic exposure/flow measurements such as Annual Average Daily Traffic (AADT), speed variation, and average operating speed do not meet the criteria for SSM, despite their proven associations with crash risk and occasional adoption as crash "surrogates." Here we consider the safety measures that satisfy the two qualifying criteria mentioned above, and the parameters considered are: Latitude, Longitude, Day of Week, Weather Conditions, Light Conditions, Road Surface Conditions, Age of Driver, Vehicle Type, Age of Vehicle, Engine Capacity in CC, Gender, Speed Limit, Number of Police Officers Present at the Scene.

Crash Potential Index (CPI), which incorporates the parameters listed to assess the likelihood of crashes at a given location or roadway segment. Here's how CPI could be formulated to include the listed parameters:

$$\text{CPI} = w1 \times \text{Road Factor} + w2 \times \text{Driver Factor} + w3 \times \text{Vehicle Factor} + w4 \times \text{Environmental Factor}$$

Where:

- $w1$, $w2$, $w3$, $w4$ are weights assigned to each factor based on their relative importance.
- **Road Factor** includes parameters such as road type, speed limit, road curvature, and road surface conditions.

- **Driver Factor** incorporates driver-related parameters such as age, gender, and compliance with traffic laws.
- **Vehicle Factor** considers vehicle-related parameters including vehicle type, age, engine capacity, and safety features.
- **Environmental Factor** accounts for environmental conditions such as weather, light conditions, and the presence of law enforcement officers.

Each factor can be quantified based on empirical data, expert judgment, or statistical analysis of historical crash data. The weights $w1$, $w2$, $w3$, $w4$ are determined to reflect the relative influence of each factor on crash potential.

For example:

- If a location has a higher speed limit (Road Factor), it might receive a higher weight in the CPI calculation.
- Drivers with a history of traffic violations or involvement in crashes (Driver Factor) might contribute more to the overall CPI.
- Older vehicles with fewer safety features (Vehicle Factor) could increase crash potential.
- Adverse weather conditions (Environmental Factor) might significantly elevate the CPI for a particular location.

By combining these factors into a single index, the Crash Potential Index provides a comprehensive measure of crash risk, allowing transportation agencies to prioritize safety interventions and allocate resources effectively to mitigate potential hazards. Additionally, ongoing data collection and analysis can refine the CPI model over time to enhance its accuracy and relevance.

Multivariate Hazard Index (MHI):

- **Formula**: MHI = $\sum_{i=1}^{n} \mathrm{wi} \times \mathrm{Zi}$
- **Description**: The Multivariate Hazard Index combines multiple parameters to assess crash risk comprehensively. Each parameter (Zi) represents a specific aspect of safety, and

weights (wi) are assigned based on their relative importance. Here's how the parameters listed could be incorporated:

1. **Location Factors**:
   - Latitude and Longitude: These geographic coordinates could be used to identify the location of interest.
2. **Temporal Factors**:
   - Day of Week: Represented as a categorical variable indicating different days of the week.
3. **Environmental Conditions**:
   - Weather Conditions: Categorized into clear, rainy, snowy, etc.
   - Light Conditions: Categorized into daytime, nighttime, dawn, dusk.
   - Road Surface Conditions: Categorized into dry, wet, icy, etc.
4. **Driver and Vehicle Characteristics**:
   - Age of Driver: Represented as categorical groups (e.g., young, middle-aged, elderly).
   - Gender: Categorized as male or female.
   - Vehicle Type: Categorized based on vehicle classification (e.g., passenger car, truck, motorcycle).
   - Age of Vehicle: Categorized into new, moderately aged, old vehicles.
   - Engine Capacity in CC: Represented as a continuous variable indicating vehicle power.
5. **Speed Limit and Enforcement**:
   - Speed Limit: Represented as a continuous variable indicating the legal speed limit for the roadway segment.
   - Number of Police Officers Present at the Scene: Represented as a binary variable indicating presence or absence of law enforcement.

The Multivariate Hazard Index provides a holistic measure of crash risk by considering multiple factors simultaneously. It allows transportation agencies to prioritize safety interventions based on the combined influence of various parameters rather than considering them in isolation. The weights assigned to each parameter can be determined through expert judgment, statistical analysis, or optimization techniques to reflect their relative importance in contributing to crash risk.

Driver Risk Index (DRI): The DRI aims to assess the risk of crashes by considering various parameters related to driver behaviour, vehicle characteristics, and environmental conditions.

$$\text{DRI} = w1 \times P(\text{Speeding}) + w2 \times P(\text{Weather}) + w3 \times P(\text{Light}) + w4 \times P(\text{Road}) + w5 \times P(\text{Driver}) + w6 \times P(\text{Vehicle})$$

Where,

- $P(\text{Speeding})$: Probability of Speeding
- $P(\text{Weather})$: Probability of Adverse Weather Conditions
- $P(\text{Light})$: Probability of Low Light Conditions
- $P(\text{Road})$: Probability of Poor Road Surface Conditions
- $P(\text{Driver})$: Probability of Risky Driver Behaviour (e.g., based on age, gender)
- $P(\text{Vehicle})$: Probability of Vehicle-related Factors (e.g., vehicle age, engine capacity)

Parameters Incorporation:

- **Probability of Speeding**: Based on historical data, traffic studies, or enforcement data, the likelihood of speeding violations could be estimated for the specific location or roadway segment.
- **Probability of Adverse Weather Conditions**: Using weather data, such as precipitation and temperature, the probability of adverse weather conditions (e.g., rain, snow) affecting driving conditions can be determined.
- **Probability of Low Light Conditions**: Light conditions, such as daytime, nighttime, dawn, or dusk, could be categorized, and their impact on crash risk assessed based on visibility and driver reaction times.

- **Probability of Poor Road Surface Conditions**: Road surface conditions, including wet, icy, or uneven surfaces, can be quantified based on maintenance records or observational data.
- **Probability of Risky Driver Behaviour**: Demographic factors such as age and gender, along with behavioural factors like distracted driving or impaired driving, can be incorporated to estimate the likelihood of risky driver behaviour.
- **Probability of Vehicle-related Factors**: Vehicle characteristics such as age, type, and engine capacity can influence crash risk and can be quantified based on vehicle registration data or surveys.

Weighting Factors ($w1, w2, \ldots, w6$): The weighting factors represent the relative importance of each parameter in contributing to crash risk and can be determined through expert judgment or statistical analysis of crash data.

The Driver Risk Index provides a comprehensive measure of crash risk by considering multiple factors related to driver behaviour, vehicle characteristics, and environmental conditions. It can help transportation agencies identify high-risk areas and prioritize interventions to improve road safety.

Speeding Risk Index (SRI): This measure will integrate parameters such as latitude, longitude, day of the week, weather conditions, light conditions, road surface conditions, speed limit, and driver demographics.

$\text{SRI} = w1 \times \text{Weather} + w2 \times \text{Light} + w3 \times \text{Road Condition} + w4 \times \text{Day of Week} + w5 \times \text{Driver Age} + w6 \times \text{Speed Limit}$

Where:

- Weather: A categorical variable representing weather conditions (e.g., clear, rainy, snowy).
- Light: A categorical variable representing light conditions (e.g., daytime, nighttime, dawn/dusk).
- Road Condition: A categorical variable representing road surface conditions (e.g., dry, wet, icy).

- Day of Week: A categorical variable representing the day of the week (e.g., Monday, Tuesday, ..., Sunday).
- Driver Age: A continuous variable representing the age of the driver.
- Speed Limit: A continuous variable representing the posted speed limit for the roadway segment.

Description:

- Each parameter in the formula contributes to the overall Speeding Risk Index (SRI) based on its influence on the likelihood of speeding-related crashes.
- Weather, light conditions, and road surface conditions are categorical variables that reflect environmental factors affecting driving conditions and speed management.
- Day of the week captures variations in traffic patterns and enforcement levels that may influence speeding behaviour.
- Driver age accounts for differences in risk-taking behaviour and driving experience.
- The speed limit of the roadway segment serves as a reference point for evaluating speeding behaviour relative to legal limits.

Application:

- Transportation agencies can use the SRI to identify high-risk roadway segments prone to speeding-related crashes.
- Targeted interventions such as speed limit adjustments, enhanced enforcement efforts, road surface improvements, and driver education campaigns can be prioritized based on SRI values.
- Longitudinal analysis of SRI values over time can track the effectiveness of interventions and inform ongoing safety improvement efforts.

This example demonstrates how a surrogate safety measure like the Speeding Risk Index can integrate multiple parameters to assess the risk of specific types of crashes and inform targeted safety interventions.

7. METHODOLOGY

This outlines the systematic process of developing a predictive model for road traffic accidents using Surrogate Safety Measures, starting from data collection and pre-processing, feature selection, model development, and evaluation, and finally, deployment and validation. It ensures that the model is accurate, reliable, and effective in preventing road accidents.

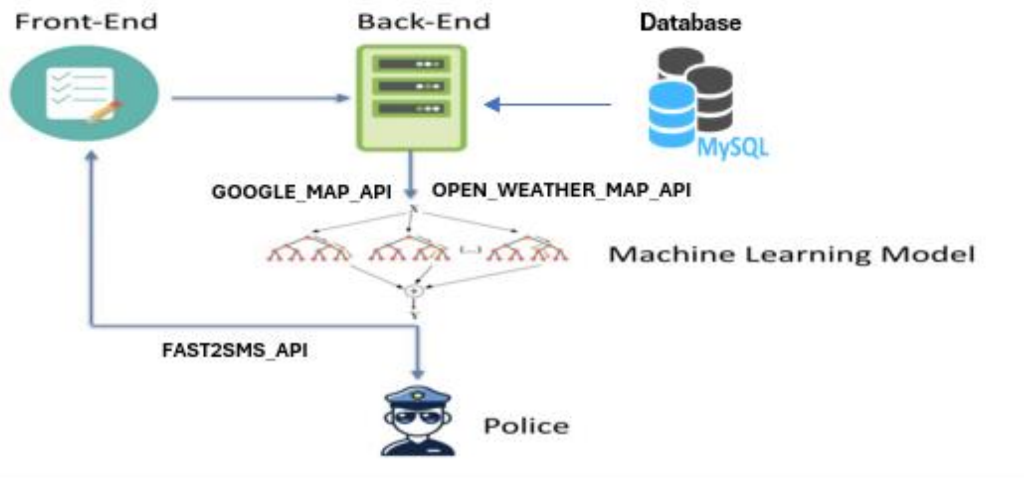

Figure 1: System Design

Front-End: Users input for the prediction factors are taken and sent to the backend server.

Back-End: The model is deployed here and the input data is fed into the Machine Learning model.

Database: MySQL DB is used for the user management to implement the user registration and login functionality. ML model will directly fetch the user data from the database, no need of manual entry.

Machine Learning Model: ML models used are decision tree, random forest and logistic regression and also applied hyperparameter tuning to increase its efficiency. Random Forest algorithm showed the highest accuracy of 89.50% and hence chosen for the model. The model runs and predicts the severity. The severity metrics are 1= Fatal, 2= Serious, 3= Slight. The output is sent back to the front-end and displayed to the user. A SMS containing the location coordinates and the severity of accident is sent to the police so that it can take preventive measures at the location.

7.1 Data Collection

Gathering relevant data related to road accidents, surrogate safety measures, and other relevant variables such as road characteristics, weather conditions, and traffic patterns. This data may come from various sources, including traffic cameras, accident databases, and road surveys. The dataset is taken from Kaggle The dataset contains 3 files - AccidentsBig.csv, CasualtiesBig.csv, and VehiclesBig.csv

The dataset is very huge and detailed based on the various parameters including location, weather conditions, road type, police force, number of vehicles, type of vehicle, speed limit, date, light conditions, accident severity, etc.

7.2 Data Preprocessing and Data Visualization

Cleaning and preparing the collected data for analysis, which may involve data validation, outlier detection, and data normalization. This step is crucial to ensure that the data used for modelling is accurate and reliable. vaex python library is used to load the data as pd dataframe. Visualizing the data based on total number of accidents on the day of the week, time of the day/night, age of people involved in the accidents, accident percentage in speed zone, co-relation between variables.

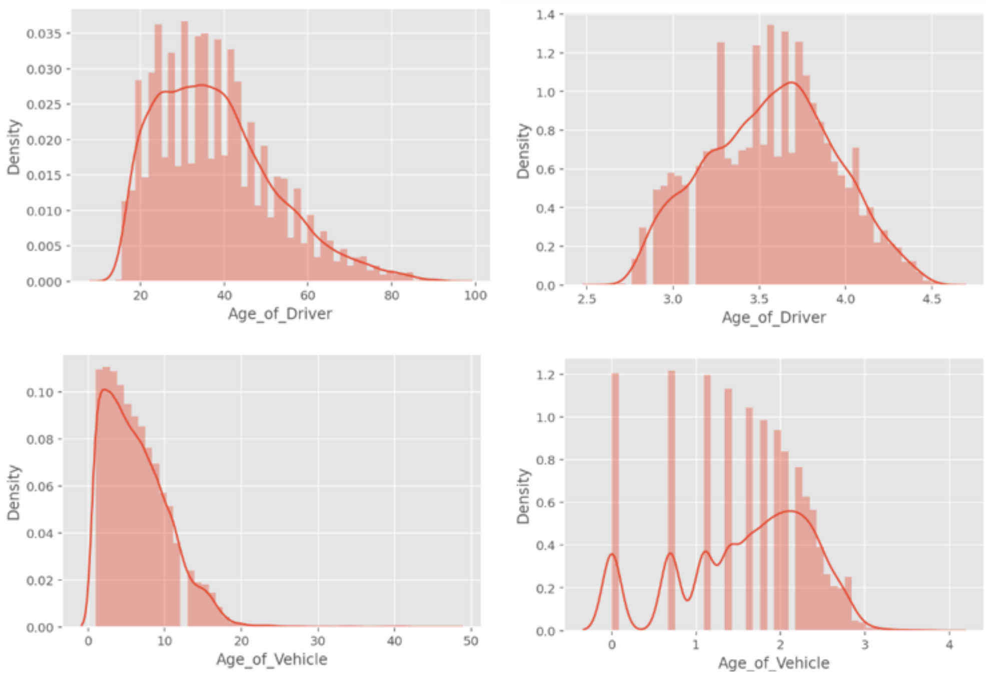

Figure 2: Data Normalization

7.3 Feature Selection

Identifying the most relevant surrogate safety measures that have a significant impact on road accidents. This may involve statistical analysis, feature ranking, and feature engineering techniques to select the most informative variables for the predictive model. Identify the most relevant surrogate safety measures (SSM) for accident prediction.

Potential SSM may include:

- Crash Potential Index (CPI)
- Multivariate Hazard Index (MHI)
- Driver Risk Index (DRI)
- Speeding Risk Index (SRI)

Using statistical analysis and feature ranking techniques to prioritize SSM and Conduct feature engineering to create new features if necessary.

7.4 Model Development

Building a predictive model that can estimate the likelihood of road accidents based on the selected surrogate safety measures. This may involve machine learning techniques such as:

- Decision tree
- Random forest
- Logistic regression

Then the data can be split into training and testing sets for model training and evaluation. Train the model using the training dataset and optimize hyperparameters. Finally, Validate the model's accuracy and performance through cross-validation techniques.

Here decision tree, random forest and logistic regression is used for ML model and also applied hyperparameter tuning to increase its efficiency. Random Forest algorithm showed the highest accuracy of 89.09% and hence chosen for the model. The model runs and predicts the severity. The severity metrics are 1= Fatal, 2= Serious, 3= Slight.

7.5 Model Evaluation

Assessing the performance of the developed model using appropriate evaluation metrics, such as:

- Accuracy: The proportion of correct predictions.
- Precision: The ability to make accurate positive predictions.
- Recall: The ability to identify all relevant instances.
- F1 Score: A combined metric of precision and recall.

Below data shows the accuracy level of each ML model which are used:

i. Random Forest: 89.50%
ii. Logistic Regression: 89.36%
iii. Decision Tree: 77.28%
iv. Logistic Regression with Hyperparameter tuning: 89.45%
v. Decision Tree with Hyperparameters tuning: 88.95%
vi. Random Forest with Hyperparameter tuning: 89.45%

7.6 Model Deployment

Implementing the developed model into a practical application, such as a traffic management system, which can provide real-time predictions of road accident risks to drivers, traffic authorities, or other relevant stakeholders. The model will be deployed in AWS Cloud.

This involves, Integration with Traffic Management System, Realtime Predictions, Alert Generation, Visualization and User Interface, and Response Mechanism. The model runs and predicts the severity. The severity metrics are 1= Fatal, 2= Serious, 3= Slight. The output is sent back to the front-end and displayed to the user. A SMS containing the location coordinates and the severity of accident is sent to the police so that it can take preventive measures at the location.

7.7 Model Validation

Validating the predictive model using real-world data to ensure its accuracy and reliability in real-world scenarios. This may involve comparing the model’s predictions with actual road accidents data to assess its performance and make necessary refinements.

8. RESULT

The result of this project includes the development of a heat-map and a user interface for traffic authorities. These outcomes are crucial for providing actionable insights to enhance road safety.

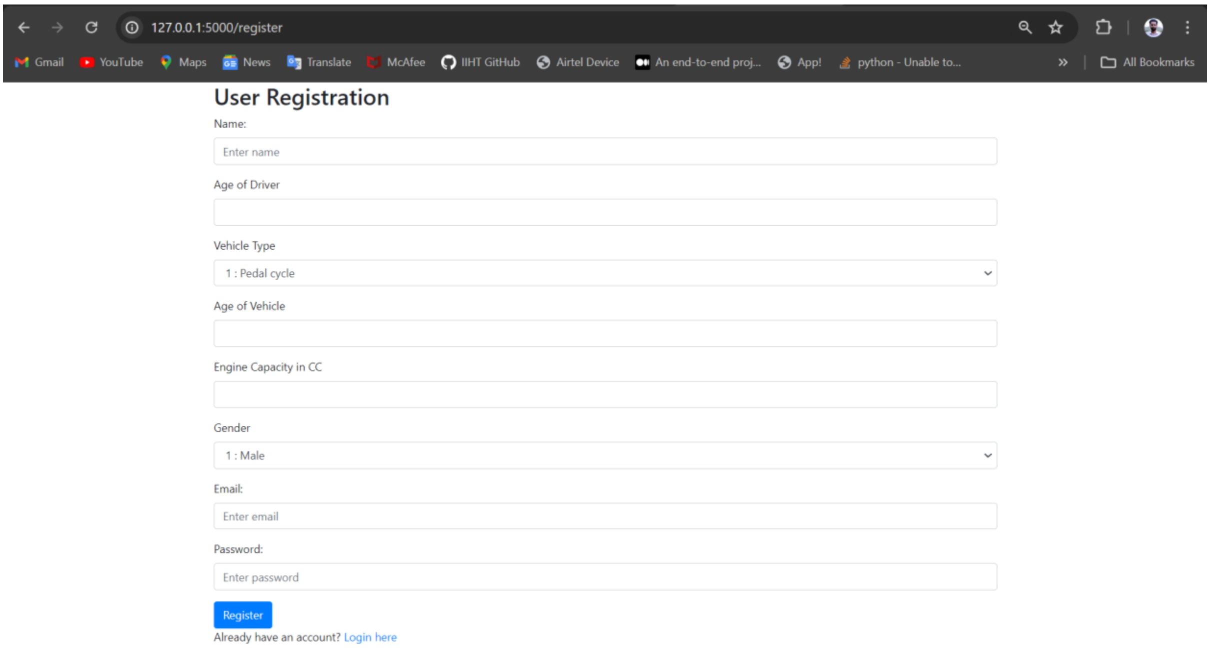

Figure 3: User Registration

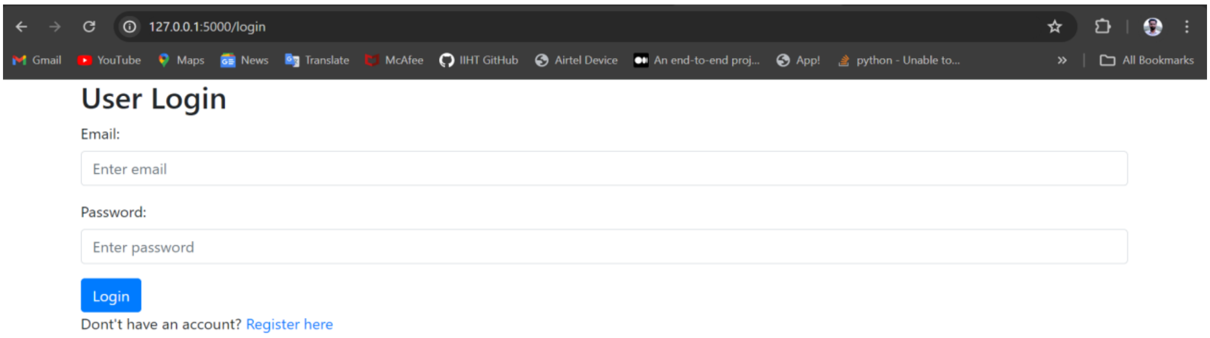

Figure 4: User Login

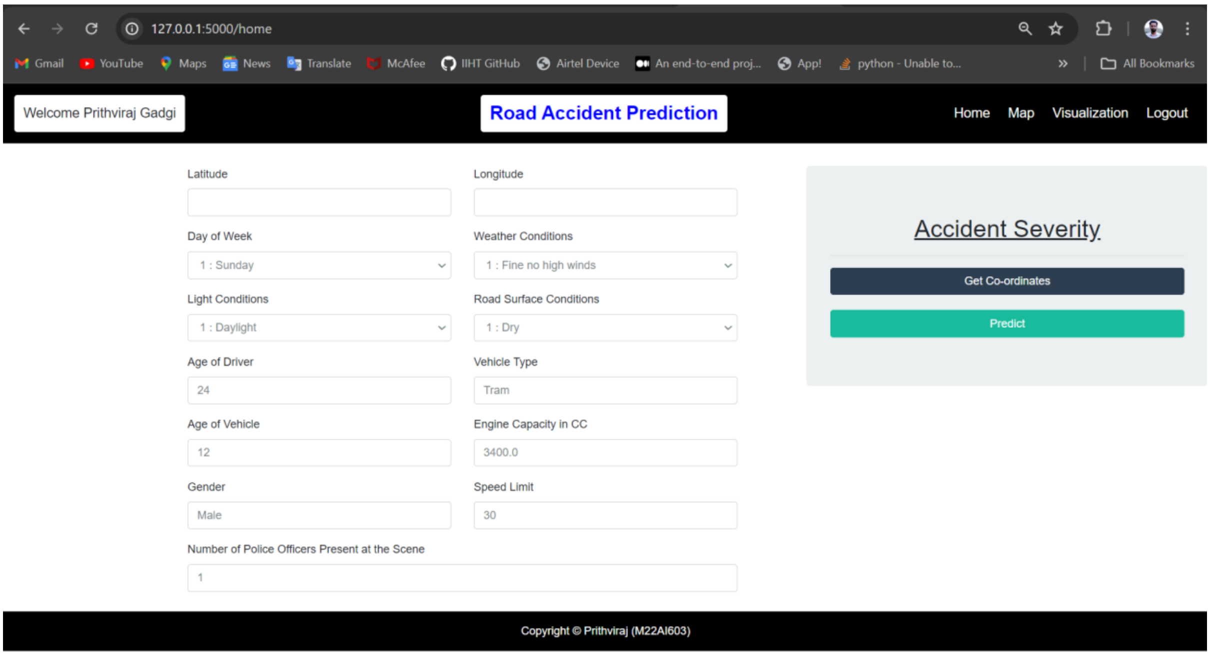

Figure 5: Home Page

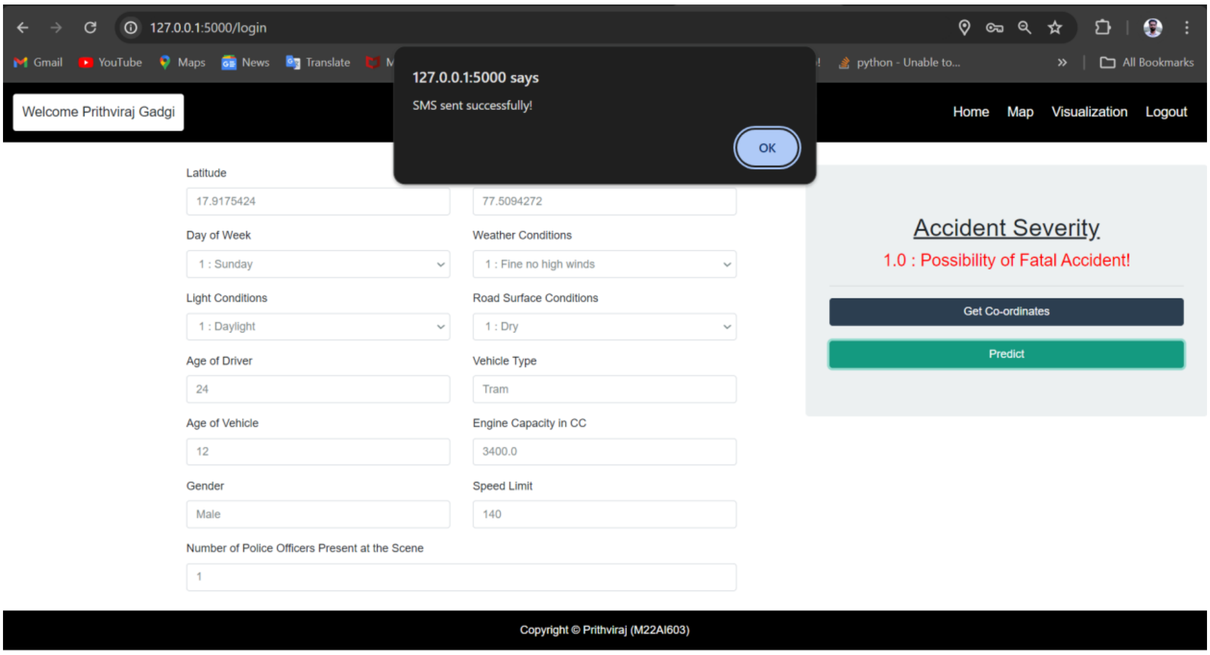

Figure 6: Accident Prediction

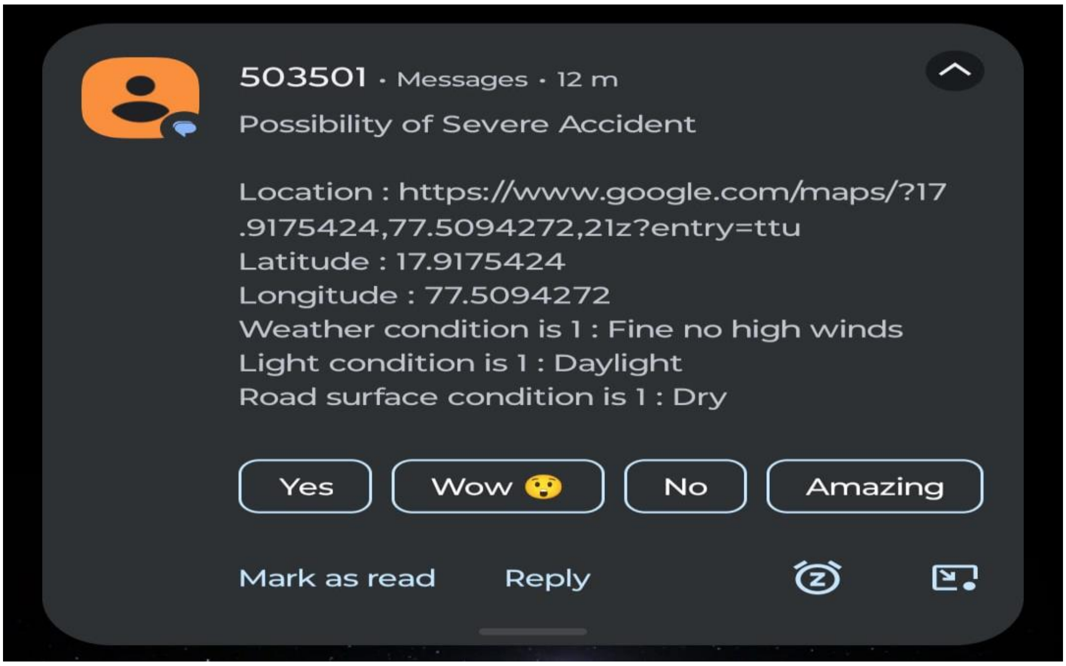

Figure 7: SMS received with location and other details

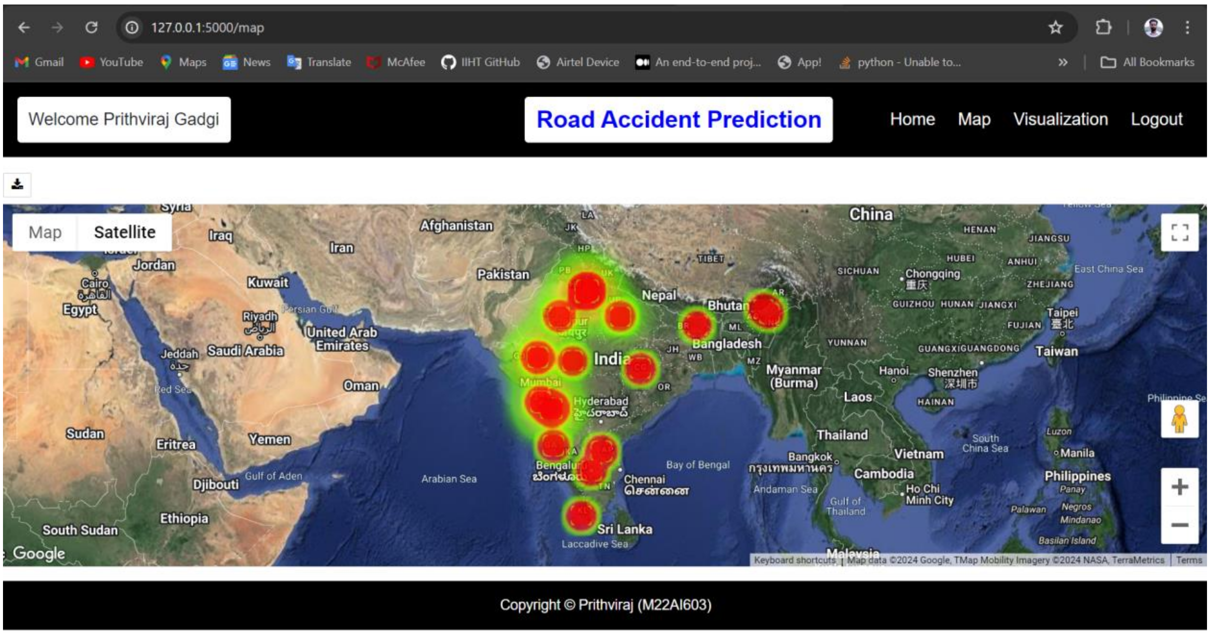

Figure 8: Heat Map

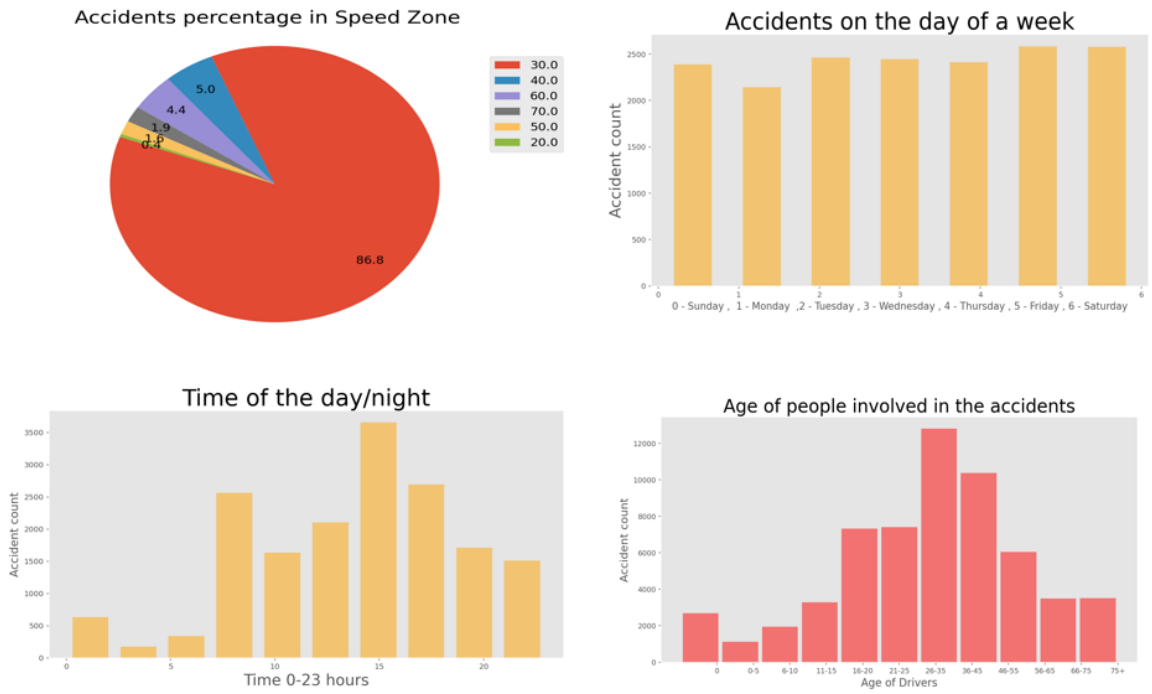

Figure 9: Data visualization

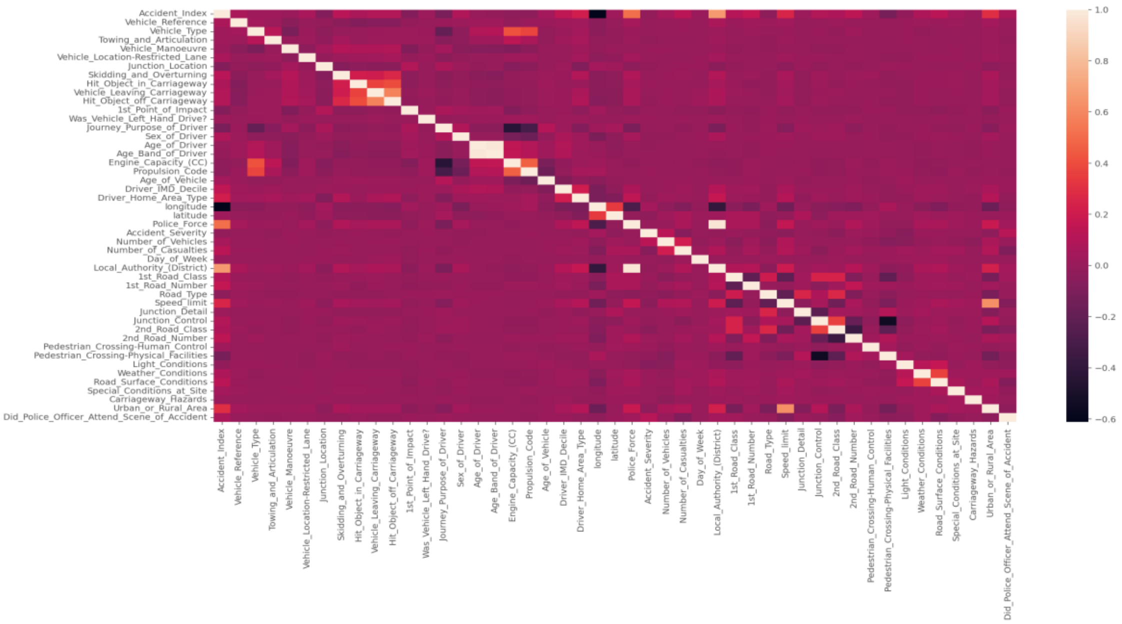

Figure 10: Co-relation between variables

The user needs to register by entering the necessary information in the registration form, the user data will be stored in MySQL database, and the password will base64 encrypted. In the login page the user can login by entering his/her email and password. Upon login the user will see home page of the web app where all the necessary information will automatically populated form the database, to update the co-ordinates user needs to click on Update Co-ordinates button, once the co-ordinates are updated the weather condition, light condition, day of week, road surface condition will be automatically fetched form the Open Whether Map API of that particular location. The user needs to click on Predict button, the ML model will predict the accident severity, and the application will send an SMS with all the necessary details using FAST2SMS API.

CONCLUSION

In the face of escalating road traffic accidents and their devastating consequences, this project aimed to pioneer a proactive approach to road safety through the development of a predictive model utilizing Surrogate Safety Measures (SSM).

In conclusion, this project serves as a pioneering step toward safer and more efficient road transportation systems. By equipping traffic authorities and stakeholders with advanced tools and insights, it reinforces the mission to reduce road traffic accidents and create safer roadways for all. The project underscores the significance of data-driven innovation in addressing one of the most pressing challenges of our time — Road Safety.

FUTURE ENHANCEMENT

- **Advanced Feature Engineering:** Continued advancements in feature engineering techniques. Exploring feature selection methods by considering temporal and spatial dependencies. Incorporating domain-specific knowledge to capture the complex dynamics of traffic accidents.
- **Multi-modal and Multi-source Data Fusion**: Integrating data from multiple sources and modes of transportation, such as pedestrian and cyclist data. By fusing data from various transportation modes and sources, predictive models can capture the interactions and interdependencies that contribute to accidents across different road user groups

- **Integration of IoT Technologies:** Integration of IoT can help in real-time data collection, real-time alerts and interventions. It can improve communication capabilities by using sensor networks and infrastructure such as connected vehicles and smart traffic signals.
- **Incorporation of External Factors:** Weather conditions, road construction, and special events, can enhance the predictive capabilities of the model. By integrating these external factors into the modelling process, the models can provide more accurate and context-aware predictions.

REFERENCES

[1] Finn H Amundsen and Guro Ranes. 2000. Studies on traffic accidents in Norwegian road tunnels. Tunnelling and underground space technology 15, 1 (2000), 3–11.

[2] Gary A. Davis, John Hourdos, Hui Xiong, and Indrajit Chatterjee. 2011. Outline for a causal model of traffic conflicts and crashes. Accident Analysis Prevention 43, 6 (2011), 1907–1919. https://doi.org/10.1016/j.aap.2011.05.001

[3] John Hayward. 1971. Near misses as a measure of safety at urban intersections . Pennsylvania Transportation and Traffic Safety Center.

[4] Chinmoy Pal, Shigeru Hirayama, Sangolla Narahari, Manoharan Jeyabharath, Gopinath Prakash, and Vimalathithan Kulothungan. 2018. An insight of World Health Organization (WHO) accident database by cluster analysis with selforganizing map (SOM). Traffic injury prevention 19, sup1 (2018), S15–S20.

[5] Steven G Shelby et al . 2011. Delta-V as a measure of traffic conflict severity. In 3rd International Conference on Road Safety and Simulati. September. 14–16.

[6] Andrew P Tarko. 2018. Surrogate measures of safety. In Safe mobility: challenges, methodology and solutions. Vol. 11. Emerald Publishing Limited, 383–405.

[7] Andrew P. Tarko. 2020. Chapter 3 - Traffic conflicts as crash surrogates. In Measuring Road Safety Using Surrogate Events, Andrew P. Tarko (Ed.). Elsevier, 31–45. https://doi.org/10.1016/B978-0-12-810504-7.00003-3

[8] Lu Wenqi, Luo Dongyu & Yan Menghua, “A Model of Traffic Accident Prediction” INSPEC Accession Number: 17239218 DOI: 10.1109/ICITE.2017.8056908

[9] Abdel-Aty, M., N. Uddin, and A. Pande. Split Models for Predicting Multivehicle Collisions during High-Speed and Low-Speed Operating Conditions on Freeways. In Transportation Research Record: Journal of the Transportation Research Board, No. 1908, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 51–58.