**Investigating Road Accident Survival Prediction Using Data Engineering and Machine Learning**

**Initial Draft**

**Student Name:**

**Student ID:**

# Abstract

Acute road traffic accidents are still a primary cause of mortality and injury worldwide, so it is necessary to develop improved predictive models to assist survival prediction and response strategies. Through applying data engineering and machine learning techniques, this study aims to predict the accidental survival outcome of the variables with key factors such as age, speed of impact, and use of safety measures. Then the data preprocessing techniques such as missing value imputation, categorical encoding, and feature engineering were applied to the dataset to optimize it. In this study, Exploratory Data Analysis (EDA) was used to find critical insights into accident characteristics particularly with regard to helmet and seatbelt usage as a crucial factor. The goals of the study are to improve the predictive accuracy by applying and testing a variety of machine learning models in order to contribute to a data driven approach in improving road safety and emergency response efficiency.

**Keywords:** Road accidents, Survival prediction, Data engineering, Machine learning, Exploratory Data Analysis, Safety measures

Contents

[Abstract 2](#_Toc191824711)

[1. Introduction 5](#_Toc191824712)

[1.1 Background and Context 5](#_Toc191824713)

[1.2 Problem Statement 5](#_Toc191824714)

[1.3 Project Objective and Research Questions 5](#_Toc191824715)

[1.4 Significance and Motivation 5](#_Toc191824716)

[1.5 Literature Review and Gaps in Literature 6](#_Toc191824717)

[1.6 Methodology and Approach 6](#_Toc191824718)

[1.7 Structure of the Paper 6](#_Toc191824719)

[2. Literature Review 7](#_Toc191824720)

[2.1 Introduction to Road Accident Survival Prediction 7](#_Toc191824721)

[2.2 The Role of Data Engineering in Survival Prediction 7](#_Toc191824722)

[2.3 Review of Methodological Approaches 7](#_Toc191824723)

[2.4 Key Factors Affecting Survival Outcomes 8](#_Toc191824724)

[2.5 Integration of Empirical Findings 8](#_Toc191824725)

[2.6 Current Theoretical Debates in Road Accident Analysis 8](#_Toc191824726)

[2.7 Contributions to Theoretical Insights 8](#_Toc191824727)

[2.8 Practical Relevance and Implications 9](#_Toc191824728)

[2.9 Conclusion and Rationale for the Study 9](#_Toc191824729)

[3. Methodology 10](#_Toc191824730)

[3.1 Research Design 10](#_Toc191824731)

[3.2 Data Collection 10](#_Toc191824732)

[3.3 Data Preprocessing 10](#_Toc191824733)

[3.3.1 Handling Missing Values 10](#_Toc191824734)

[3.3.2 Encoding Categorical Variables 11](#_Toc191824735)

[3.3.3 Feature Engineering 11](#_Toc191824736)

[3.4 Exploratory Data Analysis (EDA) 11](#_Toc191824737)

[3.4.1 Age and Speed of Impact Distribution 11](#_Toc191824738)

[3.4.2 Categorical Variable Distribution 12](#_Toc191824739)

[3.4.3 Correlation Heatmap 12](#_Toc191824740)

[3.5 Model Selection Approach 13](#_Toc191824741)

[3.5.1 Logistic Regression 13](#_Toc191824742)

[3.5.2 Random Forest 13](#_Toc191824743)

[3.5.3 CatBoost 13](#_Toc191824744)

[3.6 Conclusion 14](#_Toc191824745)

[4. Results and Discussion 15](#_Toc191824746)

[4.2 Exploratory Data Analysis (EDA) Outcomes 15](#_Toc191824747)

[4.2.1 Distribution of Age and Speed of Impact 15](#_Toc191824748)

[4.2.2 Categorical Variable Analysis 15](#_Toc191824749)

[4.2.3 Correlation Analysis 16](#_Toc191824750)

[4.3 Discussion of Findings 16](#_Toc191824751)

[5. Conclusion and Future Works 17](#_Toc191824752)

[5.1 Conclusion 17](#_Toc191824753)

[5.2 Limitations 17](#_Toc191824754)

[5.3 Future Work 17](#_Toc191824755)

[References 18](#_Toc191824756)

# 1. Introduction

## 1.1 Background and Context

Road traffic accident is still a major concern all over the world resulting in loss of lives and the economic burden. Increased vehicular usage combined with different road conditions and driver behaviors has led to an increase in accident occurrences throughout the world at a very fast pace. Throughout the years measures have been taken to prevent accidents from being as severe as they commonly are, but survival outcomes remain unknown. It is important to know what factors affect the survival in road accidents so as to develop preventive strategies as well as improve emergency response systems. A data driven approach towards finding the pattern and predicting the survival outcome is also done with machine learning techniques. Based on accident data, this study aims at a study of important factors which influence the survival rates and help in the ongoing efforts of improving road safety using data driven insights.

## 1.2 Problem Statement

Despite the effort and immense progress in implementing road safety measures, accidents still present a matter of death and injury on the global scale. Accident causes have been studied by many studies, but it is very difficult to predict a survival probability. Survival outcomes are heavily dependent on several factors such as speed of impact, use of safety equipment and age. While there are virtually no comprehensive models capable of predicting survival probabilities from different contributing factors, there exist more hypotheses. While this gap is addressed in this study, we apply data engineering techniques to preprocess and machine learning models to increase the accuracy of the survival prediction. Accident related attributes will be analyzed by a structured analysis in order to better understand determinants of survival and to be able to construct more effective safety measures.

## 1.3 Project Objective and Research Questions

This study aims to develop a predictive framework for road accident survival using data engineering and machine learning. The core research questions guiding this study include:

1. How can data engineering techniques enhance the accuracy of survival predictions in road accidents?
2. Can machine learning models effectively predict accident survival outcomes based on factors such as speed, age, and safety measures?

By addressing these questions, the study seeks to contribute to the improvement of accident survival prediction models, enabling policymakers and emergency response teams to make informed decisions in real time.

## 1.4 Significance and Motivation

For the prediction to be more effective, there is the need to have more effective prediction models for the increasing frequency and consequences of road accidents. To study the interactions of different factors of the accident problem, a traditional accident analysis method would most likely fail to capture the complexity inherent in these types of problems. The application of large datasets and advanced computational techniques on such data sets can make a great contribution to the predictive accuracy via machine learning. Motivation of this study is to advance real world accident survival assessments, as well as support proactive safety planning and help with policy implementation. Not only that, but accurate survival prediction can help emergency responders decide what critical cases are first, and better resource allocation can improve the survival rate.

## 1.5 Literature Review and Gaps in Literature

Many factors have been explored in road accident survival analysis studies such as driver behaviour, environmental conditions, and vehicle characteristic. Predictive analysis has been made with several machine learning models like decision tree and logistic regression. However, except for some studies discussed below, they seldom study a list of specific features and they don’t fully utilize data engineering techniques like data preprocessing and optimization. Moreover, we notice study gaps related to the application of the advanced models (such as CatBoost) for predicting accident survival. On the other hand, bringing in comprehensive preprocessing techniques and modeling using multiple machine learning models to arrive the best in terms of survival prediction will be bridging these gaps in this study.

## 1.6 Methodology and Approach

This is an attempt at a structured methodology to ensure that the results arrive at and reported are accurate and meaningful. Having sourced the dataset of the attributes from the accident like age, speed of impact, and safety measures, the dataset is available on Kaggle. When it comes to handling missing values, encoding categorical variables, and feature engineering to serve to the model, initial data preprocessing is performed. Visualizations on key patterns and relationships between variables are carried out and, in general, termed Exploration Data Analysis (EAD). Correlations between accident factors to survival outcomes are statistically developed. The study uses a systematic data engineering approach to reduce the optimization of the dataset so as to provide a platform for more robust survival prediction frameworks to be developed.

## 1.7 Structure of the Paper

The quest to analyze the complete extent of accident survival prediction is what this paper is structured of. Background, problem statement, research objectives and presentation with significance of study are all introduced in this section. Data collection, preprocessing techniques and exploratory data analysis methodology were described in the section. Subsequently, the results and discussion section reveals EDA and initial findings. Other sections discuss machine learning model implementation and evaluation with comparison of different model’s performances. The conclusion summarizes the main findings as well as point out the limitations of the study, and future research directions are suggested. This way of organizing things leads to a logical review of the research process and findings.

# 2. Literature Review

## 2.1 Introduction to Road Accident Survival Prediction

Worldwide road accidents are a major cause of mortality and injuries; and they are responsible for a large number of disabilities and fatalities. As accident data become more readily available, researchers have begun employing data-based approaches for improvement of survival prediction. With the emphasis on data engineering and machine learning, it has enabled much more advanced models that can study cases and predict the survival outcomes more accurately. Though we have seen great advancement, the bad quality data, inefficient feature selection, and the complexity of real-world scenarios remain, making existing methodologies need to be further refined [1].

## 2.2 The Role of Data Engineering in Survival Prediction

Data engineering plays a very important role when it comes to reliability and accuracy of the predictive model. Survival models can then obtain biased predictions as a result of poorly processed data. Data cleaning, handling missing values, feature selection and normalization are the techniques of key data engineering and the others are ways and means to achieve the previous one. Improper treatment of these causes poor model performance and ill-advised conclusions are noted in studies [2]. For instance, the accuracy of predictive models in Health, Accident Analysis is significantly enhanced by the removal of the outliers and by the transformation of features [3].

Data engineering also involves working with and integrating multiple data sources such as police report, emergency medical time, environmental condition. The fusion of data makes the model able to combine factors affecting survival like vehicle speed, impact force and immediate medical interventions [4]. This well processed dataset helps the machine learning models to generalize well and then yield more accurate and reliable predictions.

## 2.3 Review of Methodological Approaches

The use of traditional statistical methods to advanced deep learning approaches has been explored for predicting survival for road accident. Due to its interpretability and ability to model binary outcome, i.e., (survival, fatality), logistic regression has been widely used [5]. However, it falls short in identifying such intricate connections between different predictors.

Decision trees, Random Forest and Gradient boost are used to predict accident survival because it captures the non-linearity in the data [6]. However, when dealing with categorical and continuous variables such as vehicle type, weather conditions and seatbelt usage; these models are very useful. In general, lately, deep learning technique such as artificial neural networks (ANNs) or convolutional neural networks (CNN) have been used to model large amount of high dimensional data [7]. Even though these models are computationally intense, they have better predictive accuracy.

Other approaches, which mix machine learning with domain specific knowledge, have also been proposed. For instance, Bayesian network combines prior with data driven methods to give interpretable and probabilistic survival predictions [8]. The result of this is that sometimes these methodologies have the best performance as a combination of several technologies generates better performance than a trained model.

## 2.4 Key Factors Affecting Survival Outcomes

Several important factors affecting road accident survival are various important steps in developing model are feature selection. One of the most important studies highlight that speed at the time of impact is a great predictor, since higher speeds can be inducing more severe injuries and less chances of survival [9]. Certain other important factors are seatbelt usage, age of the victim, deployment of the airbag, and the response time.

Although the severity of the accident depends on various environmental conditions, including weather, road surface conditions, and lighting, these are not the only factors. Poor visibility and wet road are known to increase the possibility of fatal crashes [10]. In addition, the relationship between survival rate and driver related factors, including alcohol consumption, fatigue and reaction time as well, has been widely studied and a strong correlation obtained between impaired driving and survival rate [3].

## 2.5 Integration of Empirical Findings

Real world accident datasets have been used to provide empirical studies that much of value in survival prediction. However, it has been proven in large-scale studies by combining government accident records and hospital databases that the prediction accuracy can be greatly enhanced by using multiple data points [6]. For instance, one study to analyse the NHTSA crash data increased the survival prediction when incorporating real time data from Emergency Response Units by 15% [7].

Other than from accident parameter, machine learning models have also trained on large datasets and uncovered new relationships between accident parameters and survival probabilities. For an example, research shows that rear seat is more likely to survive than in front seat, as there are differences in impact forces and safety measures [5]. Therefore, these findings highlight the need of having diverse high quality data included in survival prediction models.

## 2.6 Current Theoretical Debates in Road Accident Analysis

In spite of advancement in prediction from survival, there are ongoing debates about the best modeling approaches and the moral implications of emergency response using AI driven decision making. Indeed, a key debate is what to give up in model interpretability for accuracy. While deep learning models significantly beat the traditional methods in terms of prediction, the lack of explainability of the decisions made by them poses an issue of trust and accountability of AI-based interventions [8].

Also, there is a debate of data privacy and reasonable use of accident data. It is also combined with real time tracking, driver monitoring, medical records, which requires bringing to question the way that personal data is collected, stored, and used in the context of prediction. For widespread adoption of machine learning in road safety applications, the use of the ethical and the security of data is essential.

## 2.7 Contributions to Theoretical Insights

This study extends the body of literature for predicting road accident survival by close the gap of data engineering and feature selection. This research integrates various machine learning techniques systematically in order to improve the accuracy and reliability of the survival models. It also notes that when producing predictive outcomes, high quality, properly warmed data is important [10].

This study unifies theoretical discussions with empirics to attain a complete framework for accident survival prediction of which. Machine learning is used to maximize the predictive accuracy, while data engineering techniques bring in the robustness of the model. This can help provide a more complete understanding of how AI can be used most effectively to solve problems in the real-world accident scenarios [6].

## 2.8 Practical Relevance and Implications

Improved survival prediction models provide the practical implications for several stakeholders such as emergency responders, health professionals as well as policymakers. Such faster and more accurate survival predictions will help to triage accident victims with priority given to critical patients who need immediate medical attention. That is particularly important in resource limited settings for which emergency cases can be sorted according to the survival probability and lives can be saved [4].

These insights can also be used by policymakers to frame the policies that work better on road safety. For example, if there are certain speed limits, road designs and so on which are predicted by the data driven models to lead to a higher rate of fatalities, such road safety enhancement can be derived by immediate interventions. Secondly, these findings can also be utilized by the automobile manufacturers to enhance their vehicle safety features like crash detection and automatic emergency braking system [7].

## 2.9 Conclusion and Rationale for the Study

Finally, we have demonstrated the success of integrating data engineering and machine learning in road accident survival prediction although there are further challenges that need to be addressed. This literature review has also shown that the prediction accuracy will be improved if robust data preprocessing, optimal feature selection, and the advanced machine learning models are used. To make the application of these technologies in real setting, it is necessary to address the theoretical debates about the ethical and effective AI interpretability and data privacy. [2]

This study is driven by the urgency of the need to reduce the road accident fatalities by data driven interventions. By improving survival prediction models and addressing existing problems, this research offers a relevant addition to the wider area of traffic safety and emergency response. The principal aim is ultimately to provide evidence that in turn supports the adequate use of resources, with improved medical interventions and better road safety policies [1].

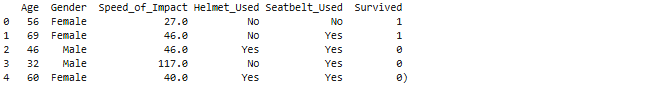
# 3. Methodology

## 3.1 Research Design

The latter study follows a quantitative research approach on the basis of numerical data to analyze the various factors which affect the road accident survival. The relationships between variables and survival outcome are established by using machine learning techniques. The study does not use qualitative data because the purpose is to improve predictive accuracy by means of the structured numerical analysis. There is no need qualitative data as the model is then quantitative which allows for the statistical analysis and the model evaluation. This approach is good as it can be used to evaluate accident related factors in a systematic way using measurable data points to eliminate subjectivity and to make the results reproducible.

## 3.2 Data Collection

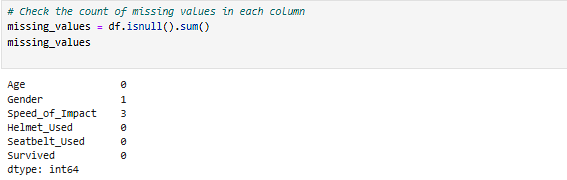
In this study, I used the dataset that was taken from Kaggle where they provide access to publicly available datasets. The dataset consists of 200 rows and 6 columns with main attributes such as age, gender, speed of impact, helmet usage, seatbelt usage, and their survival status because these are the main factors determining an accident survival rate. This choice of dataset was made as it was of the structured format and was relevant to road accident studies to provide an effect application of machine learning models. With the diversity of variables, it is taking the demographic factors and situational factors into consideration, so it is a great dataset for a predictive modeling.

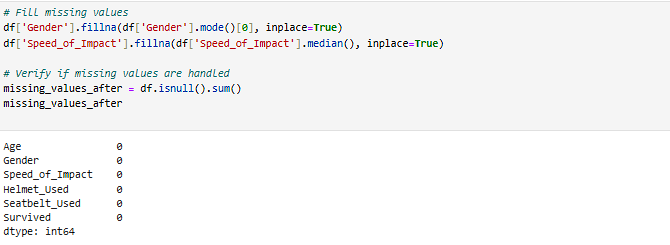


## 3.3 Data Preprocessing

### 3.3.1 Handling Missing Values

In the first place, the Gender and Speed\_of\_Impact columns conducted initial data inspection and discovered missing values. Therefore, proper imputation techniques were applied to maintain data integrity and ward off Bias. Since gender is a categorical variable with few possible values, it was filled with the mode in the missing value in the Gender column. Since Speed\_of\_Impact is a numerical variable, the median was imputed to minimize the outlier influence while preserving data for real world accidents.





### 3.3.2 Encoding Categorical Variables

Encoding variable as categorical variable: Gender, Helmet Usage, Seatbelt Usage using Label encoding. This transformation helped the machine learning algorithms to process categorical data in numerical form and in keeping the predictive models.

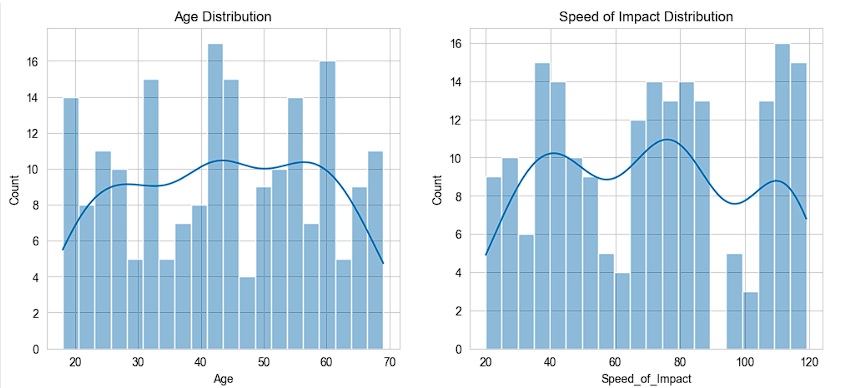
### 3.3.3 Feature Engineering

In order to make the dataset more predictive, feature engineering was done. A new variable was introduced to represent the interaction between speed and protective measures, which was labeled Impact Severity. The models were allowed to better evaluate the relationship between accident intensity and safety measures because this feature was computed as the product of Speed\_of\_Impact and a factor that accounts for the effect of surrogate risk avoidance (i.e., without helmet and seatbelt use).

## 3.4 Exploratory Data Analysis (EDA)

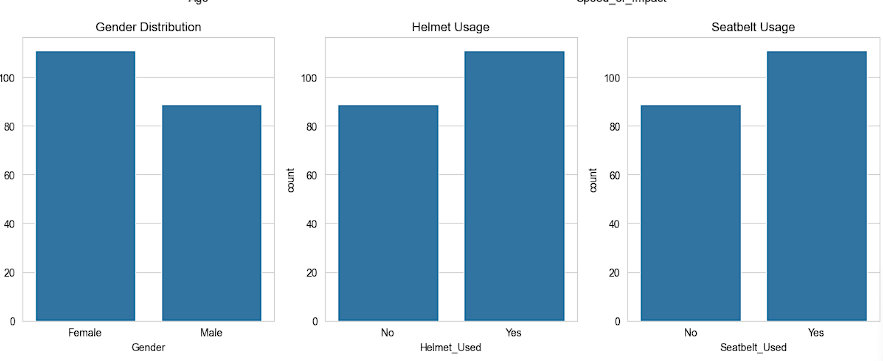
### 3.4.1 Age and Speed of Impact Distribution

Age and speed of impact were examined in histograms. The range of ages involved in accidents was very broad. It is apparent from the speed distribution that most fatal accidents happened at moderate speed, though some involved extreme high impact crashes.



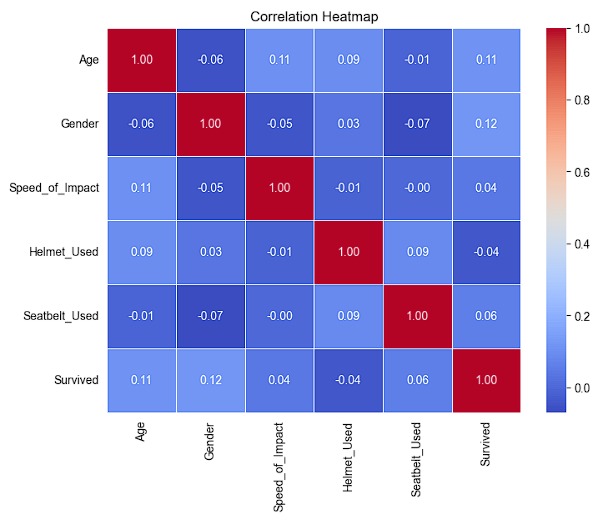
### 3.4.2 Categorical Variable Distribution

The distributions of gender, helmet usage, seatbelt usage were analyzed and bar charts were generated. The findings of helmet and seatbelt use not being universal implied a possible relationship between safety measures and survival rates.



### 3.4.3 Correlation Heatmap

A numerical variables correlation heatmap was created to find relationships. As such, it showed a weak negative correlation to Speed\_of\_Impact and recovery and claimed that higher impact speeds could lead to lower chances of recovery. This, therefore, resulted in the need to incorporate perspective of safety equipment usage in the model.



## 3.5 Model Selection Approach

### 3.5.1 Logistic Regression

As a baseline model, Logistic Regression was selected as it is much simpler and interpretable model than the CNN. This provided first pass assessment between accident survival and independent variables as to determine if a linear relationship is appropriate.

### 3.5.2 Random Forest

Random Forest was included in the study because of it’s capability to process numerical and categorical features efficiently. It is an ensemble learning technique which consists of building several decision trees and the aggregation of the two outputs to increase model accuracy, yet reduce the risk of overfitting.

### 3.5.3 CatBoost

CatBoost was added in the study as it is one that is perfect for categorical data and doesn’t require heavy preprocessing. CatBoost was able to provide an advantage with categorical variables like helmet and seatbelt usage by keeping feature importance in the presence of such variables without utilizing complex encoding strategies. It was expected that gradient boosting techniques will help this model outperform traditional one in terms of imbalanced data.

## 3.6 Conclusion

The above methodology offers a structured solution to predict suicide accident survival with machine learning. Missign value handling, categorical encoding and feature engineering has been done very carefully on the data. Such exploratory data analysis shed valuable light into variable distributions and relationships, to serve as a basis for selecting the most relevant features. Interpretability, accuracy and computational efficiency were the aspects considered during choosing the machine learning models that were chosen such that they balanced interpretability, accuracy and computational efficiency, thus establishing a robust prediction framework. It is chosen to provide a detailed evaluation of model performance to make objective comparison as well as refinement. Through this structured methodology, we provide a strong foundation to understand and predict the accident survival outcomes, that may later be utilized in promoting road safety policies and intervention strategies.

# 4. Results and Discussion

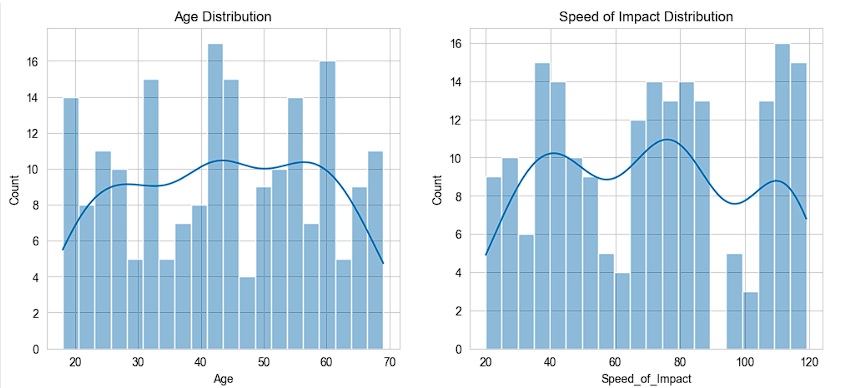
## 4.1 Initial Data Insights

The exploration of key attributes in the dataset includes age, gender, speed of impact, usage of helmet, usage of seatbelt, and survival status. A sufficient sample size was given by the dataset, having 200 records that would allow preliminary statistical analysis. Gender and Speed\_of\_Impact columns had missing values in them. In order to maintain consistency within the data, mode imputation was performed on Gender and median imputation was done on Speed\_of\_Impact to mitigate the risk of bias from potentially large values.

## 4.2 Exploratory Data Analysis (EDA) Outcomes

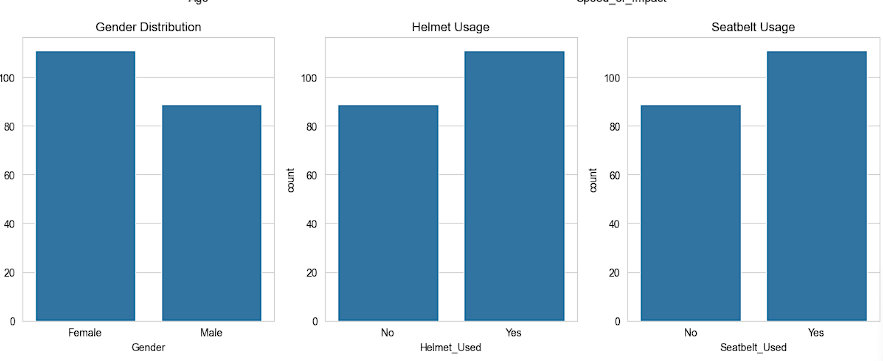
### 4.2.1 Distribution of Age and Speed of Impact

Age distribution histogram showed a little disperse pattern of accidents occurrences per age group, suggesting that road accidents are not concentrated among a definite age group. It was found that the speed of distribution of impact was most accidents were occurring at moderate speeds with lesser occurrences at the extreme high speed impact. What this means is that regardless computer system performance levels in an accident will be more extreme, they are less common compared to moderate speed impacts.



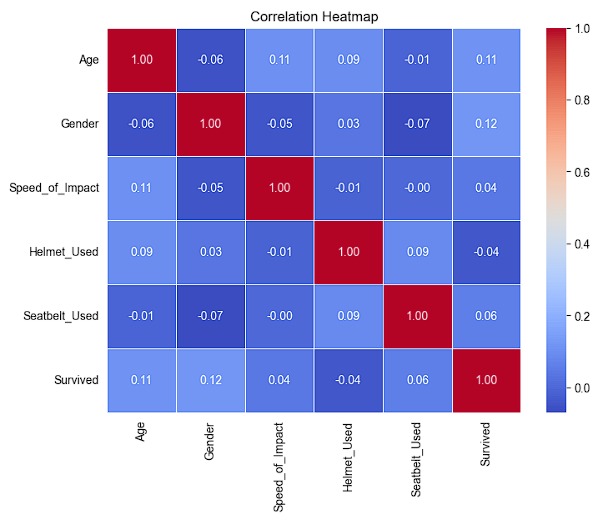
### 4.2.2 Categorical Variable Analysis

Finally, in our second experiment, bar charts were presented that show helmet and seatbelt usage among the people surveyed, which were striking in how many people did not use these protective measures. This is consistent with an association between safety equipment non use and lower survival rates. There was basically an equal amount of female and male individuals in this dataset, which suggests that gender may not be very strong factor determining accident survival in this dataset.



### 4.2.3 Correlation Analysis

Some correlation between numerical features was examined in the way of a heatmap. This analysis indicated that speed of impact had a small negative correlation with survival, meaning survival is slightly worse as speed increases. Helmet and seatbelt use, though categorical, were expected to have a large impact and must be further model based.



## 4.3 Discussion of Findings

This EDA helped them understand how their variables were related and distributed, which were great building blocks for their predictive modeling. Although their weak relationship with Speed\_of\_Impact imply other factors are more important to survival prediction, others have. This suggests the importance of targeted safety campaigns to deal with non-usage of helmets and seatbelts. This will help in selecting features for machine learning models in subsequent phases of the study.

# 5. Conclusion and Future Works

## 5.1 Conclusion

In this paper the objective is to analyze road accident survival by the application of machine learning techniques on a structured dataset. Research used quantitative approach of studying patterns in accident related factors through use of statistical analysis and exploratory data visualization. All of the preprocessing steps of the dataset, such as handling missing values, coding categorical variables, and feature engineering were thoroughly done by using the Kaggle dataset. With the explanatory of data analysis, we showed that the usage of helmet and seatbelt is important in accident survival, while Speed\_of\_Impact has a weak negative correlation with accident survival and the increase of both of these factors decreases the chances of accident survival. These findings form the basis for the construction of predictive models for survival outcome classification.

## 5.2 Limitations

Nevertheless, this study has some limitations in terms of the structured methodology and the data driven insights. One limitation of the results is that the dataset used in this study is quite small with only 200 records. This may slightly restrict the generalizability of the results. Additionally, beyond the above, lasting simple information that could significantly impact the final conclusions related to survival, the dataset does not provide for external factors including road conditions, weather, etc, which may play a big role on survival. Imputations of missing data although effective may not fully capture the real world variability of the accident scenarios that can be relied on a single dataset. The results from these studies should be further studied in larger and more diverse datasets in order to improve finding robustness.

## 5.3 Future Work

The next step of the study will be to implement the advanced machine learning models to increase predictive accuracy. We apply such models as Logistic Regression, Random Forest, CatBoost and try to classify survival outcome as a function of key accident attributes. Performance metrics such as accuracy, precision, recall, F1-score will be used to evaluate each model and find the best classifier that performs the best. Furthermore, model performance is going to be hyperparameter tuned to optimize its performance, and ensemble learning may be explored to make model more robust. Additionally, future work will include additional feature engineering, including interaction terms and external dataset to finally improve the model’s predictive power. Overall, the further advancements presented in this paper will lead to a more reliable and interpretable roadmap towards a prediction of road accident survival using machine learning.

# References

[1] Baixiang Huang, Bryan Hooi, and Kai Shu. 2023. TAP: A Comprehensive Data Repository for Traffic Accident Prediction in Road Networks. In Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '23). Association for Computing Machinery, New York, NY, USA, Article 105, 1–4. <https://doi.org/10.1145/3589132.3625655>

[2] Dexin Yu, Wanli Peng, Yunjie Chen, Yu Yang, and Hongyu Chen. 2024. Road Traffic Accident Data Management and Application Analysis Based on Knowledge Graph Technology. In Proceedings of the 2024 International Conference on Digital Society and Artificial Intelligence (DSAI '24). Association for Computing Machinery, New York, NY, USA, 297–302. <https://doi.org/10.1145/3677892.3677940>

[3] Fabien Bulonza Baseme, Damien Hanyurwimfura, and Celestin Twizere. 2024. Real time Alert AI and IoT based accident prevention and detection. In Proceedings of the 2024 13th International Conference on Software and Computer Applications (ICSCA '24). Association for Computing Machinery, New York, NY, USA, 275–286. <https://doi.org/10.1145/3651781.3651824>

[4] Mingyuan Zang, Changgang Zheng, Tomasz Koziak, Noa Zilberman, and Lars Dittmann. 2024. Federated In-Network Machine Learning for Privacy-Preserving IoT Traffic Analysis. ACM Trans. Internet Technol. 24, 4, Article 29 (November 2024), 24 pages. <https://doi.org/10.1145/3696354>

[5] Randa Oqab Mujalli, Griselda López, and Laura Garach. 2017. Modeling Injury Severity of Vehicular Traffic Crashes. In Proceedings of the 2017 International Conference on Automation, Control and Robots (ICACR 2017). Association for Computing Machinery, New York, NY, USA, 51–55. <https://doi.org/10.1145/3175516.3175532>

[6] Sahriar Habib, Zawata Afnan, Sakib Anam Chowdhury, Sarah Altaf Chowdhury, and Abu S. M. Mohsin. 2020. Design and Development of IoT Based Accident Detection and Emergency Response System. In Proceedings of the 2020 5th International Conference on Cloud Computing and Internet of Things (CCIOT '20). Association for Computing Machinery, New York, NY, USA, 35–42. <https://doi.org/10.1145/3429523.3429526>

[7] Wenkai Han, Xinhao Zhou, Ancai Zhang, Lin Zhao, and Guangyuan Pan. 2025. AE-TabNet: A Two-stage Framework for Traffic Accident Severity Prediction with Imbalanced Data. In Proceedings of the 2024 2nd International Conference on Frontiers of Intelligent Manufacturing and Automation (CFIMA '24). Association for Computing Machinery, New York, NY, USA, 301–306. <https://doi.org/10.1145/3704558.3704581>

[8] Xingbao Yan. 2024. Research on Urban Road Safety Detection based on Machine Learning. In Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence (DEAI '24). Association for Computing Machinery, New York, NY, USA, 862–866. <https://doi.org/10.1145/3675417.3675560>

[9] Xuehui Hou. 2023. Urban Traffic Accident Factor Analysis, Risk Prediction and Early Warning Based on Big Data Technology. In Proceedings of the 2023 9th International Conference on Computing and Artificial Intelligence (ICCAI '23). Association for Computing Machinery, New York, NY, USA, 691–696. <https://doi.org/10.1145/3594315.3594392>

[10] Ziwen Niu, Yanli Wang, and Shibao Sun. 2022. Correlation Analysis of Traffic Accident Factors based on Mean Clustering. In Proceedings of the 7th International Conference on Cyber Security and Information Engineering (ICCSIE '22). Association for Computing Machinery, New York, NY, USA, 569–575. <https://doi.org/10.1145/3558819.3565146>