4-22-2024

Driver's Accident Behavioral Analytics Using AI

Mohamad Amin Obaid
mo3658@mail.rit.edu

Driver's Accident Behavioral Analytics Using AI

by

Mohamad Amin Abdallah Obaid

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree

of Master of Science in Professional Studies:

Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

22/04/2024

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Thesis Approval

Student Name**: Mohamad Amin Abdallah Obaid**

Thesis Title**: Driver's Accident Behavioral Analytics Using AI**

Graduate Committee:

Name: Dr. Sanjay Modak **Date:**

Chair of committee

Name: Dr. Ehsan Warriach **Date:**

Member of committee

Acknowledgement

I express my deepest gratitude to those whose unwavering support and contributions have played a pivotal role in the successful completion of this research on "Driver Accident behavioral Analytics using AI." This exploration has been both challenging and enlightening, and the collaborative efforts of many have significantly enriched its outcome.

I am profoundly thankful to my research advisor. Dr. Ehsan Warrich, whose expertise and guidance have been indispensable throughout every stage of this study. Their insightful input, unwavering support, and encouragement have been crucial in shaping the trajectory of this research, and I am truly grateful for their mentorship.

I extend my appreciation to RIT Dubai for providing an academic environment conducive to undertaking this thorough investigation into road safety factors. The resources and facilities provided have been instrumental in conducting a comprehensive analysis.

To my friends and family, who have been pillars of strength, I am profoundly grateful. Your constant support, understanding, and patience have sustained me through the challenges and triumphs of this academic endeavor. Your encouragement has been a driving force behind the completion of this research.

I would like to acknowledge everyone who contributed, in various capacities, to this project. Whether through engaging discussions, valuable feedback, or simply lending a sympathetic ear, your collective efforts have left an indelible mark on this research. I deeply appreciate the collaborative spirit that has brought this in-depth analysis of traffic accident severity to fruition.

In conclusion, my sincere thanks go out to everyone who has been a part of this journey. Your support and collaboration have been instrumental in the successful completion of this research on factors influencing traffic accident severity.

Abstract:

This comprehensive dissertation constitutes a significant contribution to the ongoing global discourse on road safety. Through a judicious utilization of advanced data analysis techniques, with a particular emphasis on machine learning applications, this research endeavors to address and bridge crucial gaps in our comprehension of multifaceted aspects related to road safety. Specifically, the study aims to delve into the intricacies of accident severity factors, driver characteristics, vehicle attributes, and the complex dynamics of road conditions. By systematically exploring these dimensions, the research endeavors to unearth more nuanced and precise relationships that influence accident outcomes. Moreover, a particular focus is dedicated to unraveling the intricate interplay between driver demographics, such as age and gender, and their interactions with other pertinent variables. The dissertation also places a spotlight on the often-overlooked potential of advanced data analysis techniques, underscoring their capability to extract profound insights from extensive datasets pertaining to road accidents. As the research unfolds, due acknowledgment is given to the evolving landscape of vehicle technologies, and a thorough assessment is conducted to discern their impact on road safety. This nuanced analysis contributes significantly to the overarching goal of developing evidence-based safety measures and fostering informed policymaking. The ultimate aim is to mitigate the societal toll of road accidents and pave the way for a safer and more secure transportation ecosystem globally. The thesis is structured into six chapters: Introduction, Literature Review, Research Methodology, Findings and Data Analysis, Discussion, and Conclusions, each addressing specific aspects of the research process and outcomes.

Keywords: Road safety, accident severity, driver characteristics, vehicle attributes, road conditions, data analysis, machine learning, advanced statistical techniques, evolving vehicle technologies, evidence-based policymaking.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Road safety is a global concern of paramount importance, affecting lives, well-being, and economies worldwide. The pervasive impact of road accidents necessitates a profound understanding of their dynamics and causal factors for effective safety measures and policymaking. In today's data-rich environment, the opportunity to utilize data analysis and machine learning for a deeper exploration of road safety is compelling. While existing literature provides valuable insights, critical gaps hinder a comprehensive understanding of accident severity factors, driver characteristics, vehicle attributes, and road conditions. This dissertation aims to address these gaps by employing advanced data analysis techniques to contribute significantly to the development of evidence-based road safety measures and policymaking.

1.1 Background

Road safety is a critical concern of paramount importance, impacting the lives and well-being of people worldwide. Road accidents result in substantial human suffering, economic losses, and strain on healthcare systems, law enforcement, and infrastructure. Understanding the dynamics of road accidents and their causal factors is imperative for the development of effective road safety measures and informed policymaking.

In today's data-rich environment, there is a compelling opportunity to harness the power of data analysis and machine learning to delve deeper into the intricacies of road safety. While existing literature has provided valuable insights into various facets of road accidents, there are critical gaps that hinder a comprehensive understanding of this complex issue.

1. Accident Severity Factors: Research to date has identified numerous variables that contribute to accident severity, including weather conditions, road type, and driver behavior. High-speed roads and adverse weather conditions have been associated with more severe accidents. These insights have been crucial for designing specific interventions but require further investigation to establish more precise relationships.

2. Driver Characteristics: Driver demographics, such as age and gender, have been shown to influence accident outcomes. Young and inexperienced drivers often face higher risks of accidents. Yet, the role of these factors, and their interplay with other variables, remains a subject that needs further exploration.

3. Vehicle Characteristics: The type, age, and engine capacity of vehicles have also been linked to accident outcomes. Older vehicles may be more prone to accidents, and specific vehicle types, such as motorcycles, have higher accident rates. Understanding the role of vehicle characteristics and how they interact with other variables is essential for tailored safety measures.

4. Location and Road Conditions: The road environment, including road type, lighting conditions, and the presence of hazards, significantly influences the likelihood and severity of accidents. However, there is a need to gain a more nuanced understanding of these interactions and their impact on accident causality.

5. Advanced Data Analysis Techniques: While conventional statistical approaches have yielded valuable insights, limited studies have explored the application of advanced data analysis techniques, such as machine learning, in this domain. The potential for extracting deeper insights from large accident datasets remains underutilized.

6. Evolving Vehicle Technologies: As the automotive landscape evolves with technological advancements, it is essential to assess the impact of changing vehicle technologies on road safety. In particular, understanding the relationship between vehicle age and accident severity in the context of evolving technologies is becoming increasingly relevant.

Considering these gaps and opportunities, this project endeavors to provide a comprehensive and data-driven approach to analyzing road accidents and vehicle information. By addressing these challenges, we aim to contribute significantly to the development of improved road safety measures and informed policymaking, ultimately reducing the toll of road accidents on society.

Problem Statement:

The pervasive and enduring threat posed by road accidents to public safety and societal well-being constitutes a substantial global concern, resulting in profound human and economic consequences. In an era characterized by an abundance of data, the imperative to meticulously analyse road accidents and vehicle information datasets has become more pronounced than ever before. Despite the wealth of valuable insights garnered from previous research efforts, there persist critical gaps in our comprehension of the multifaceted factors that contribute to accident severity. Additionally, the intricate interplay between driver characteristics, vehicle attributes, and road conditions remains a complex terrain, warranting further exploration.

Moreover, the adoption of advanced data analysis techniques, particularly within the realm of machine learning, has been relatively limited in the field of road safety research. This limitation represents a significant opportunity for innovation and enhancement in understanding accident dynamics and predicting outcomes. In the face of a continually evolving automotive landscape marked by emerging technologies, it becomes imperative to delve into the nuanced impact of factors such as vehicle age and technological advancements on accident outcomes. The absence of a comprehensive understanding of these elements hampers the development of effective road safety measures and evidence-based policymaking.

The overarching objective is to bridge these knowledge gaps by undertaking a thorough and comprehensive analysis of road accidents and vehicle information datasets. By employing advanced data preparation techniques, conducting rigorous exploratory data analysis, and leveraging state-of-the-art machine learning methodologies, this research seeks to unravel intricate patterns and correlations within the data. The ultimate aim is to derive meaningful insights into the multifaceted determinants of accident severity and the intricate relationships between driver characteristics, vehicle attributes, and road conditions.

Addressing these critical gaps in understanding promises not only to contribute significantly to the scholarly discourse on road safety but also to inform the development of more effective and targeted measures for enhancing road safety. The insights garnered from this research endeavour are anticipated to serve as a valuable foundation for evidence-based policymaking, thereby fostering a safer and more secure transportation environment in the face of evolving technological landscapes and the persistent challenges posed by road accidents.

Research Aim:

The overarching aim of this research is to conduct a comprehensive and data-driven analysis of road accidents and vehicle information to deepen our understanding of the multifaceted factors influencing accident severity. This aim aligns with the imperative to contribute significantly to the development of evidence-based road safety measures and policymaking.

Objectives:

1. To identify and analyze the precise relationships between accident severity and factors such as weather conditions, road type, and driver behavior.
2. To explore how driver demographics, including age and gender, interact with other variables to influence accident outcomes.
3. To examine the role of vehicle characteristics, such as type, age, and engine capacity, in determining accident severity and understand their interactions with other factors.
4. To assess the nuanced impact of road environment variables, including road type, lighting conditions, and hazards, on the likelihood and severity of accidents.
5. To evaluate the extent to which advanced data analysis techniques, particularly machine learning, can provide deeper insights into road safety beyond conventional statistical approaches.
6. To investigate how evolving vehicle technologies, including changes in vehicle age, impact road safety outcomes and understand their relationship with accident severity.

These objectives collectively form a structured framework aimed at fulfilling the broader goal of advancing our understanding of road safety dynamics.

Research Questions:

1. **Hypothesis 1:**
- Research Question: Which gender, male or female, is more associated with driving accidents of fatal severity?
- Sub-question: What is the probability of an accident being fatal given the gender of the driver?

2. **Hypothesis 2:**
- Research Question: Is speeding a primary cause of car crashes, and does it significantly contribute to fatal accidents?
- Sub-question: What is the probability of a crash being fatal given that speeding is the cause?

3. **Hypothesis 3:**
- Research Question: Are accidents more severe when they occur at night compared to daytime?
- Sub-question: What is the probability of a nighttime accident being fatal?

4. **Hypothesis 4:**
- Research Question: Are accidents on asphalt roads more severe than those on earth roads?
- Sub-question: What is the probability of an accident on asphalt roads being fatal?

5. **Hypothesis 5:**
- Research Question: Do rainy days contribute to more severe accidents compared to normal days?
- Sub-question: What is the probability of an accident on a rainy day being fatal?

6. **Hypothesis 6:**
- Research Question: Are accidents involving collisions with pedestrians less severe than those involving vehicle-to-vehicle collisions?
- Sub-question: What is the probability of a pedestrian-involved accident being fatal?

7. **Hypothesis 7:**
- Research Question: Do drunk driving and overturning significantly contribute to high severity accidents?
- Sub-question: What is the conditional probability of high severity given drunk driving or overturning?

8. **Hypothesis 8:**
- Research Question: Is there a correlation between the education level of drivers and the rate and severity of accidents?
- Sub-question: What is the mean severity of accidents for different education levels?

9. **Hypothesis 9:**
- Research Question: Do younger drivers commit more accidents than older drivers, and are these accidents more severe?
- Sub-question: What is the probability of an accident being fatal given the age group of the driver?

These research questions form the core of the investigative process, guiding the exploration of critical facets within the domain of road safety.

Limitations of the Study:

Despite the comprehensive nature of this research, it is essential to acknowledge certain limitations that may impact the scope and generalizability of the findings. These limitations include:

1. **Data Limitations:** The study relies on the available datasets ("accident_information.csv" and "vehicle_information.csv"), and any inherent biases or inaccuracies in the data may influence the research outcomes.
2. **Temporal Constraints:** The study is conducted within a specific timeframe, and changes in road safety dynamics or advancements in technology occurring after this period may not be fully captured.
3. **Geographical Scope:** The research focuses on a specific geographical area, and variations in road safety practices and conditions across different regions may not be fully addressed.
4. **Variable Complexity:** The intricate interplay of various variables influencing road accidents introduces a level of complexity. While efforts are made to analyze these factors comprehensively, the study may not capture all potential variables.
5. **Technological Limitations:** The effectiveness of machine learning techniques may be influenced by technological constraints, such as computational resources and algorithmic limitations.

Acknowledging these limitations is crucial for interpreting the findings accurately and providing insights into potential areas for future research and refinement of methodologies.

Chapter – 2 Literature Review

Road safety, a perennial concern, has garnered considerable scholarly attention over the years, with an array of studies contributing substantially to our understanding of the multifaceted factors influencing accident severity and the intricate web of road safety measures. This literature review meticulously synthesizes key insights from seminal papers, presenting a comprehensive overview of the road safety knowledge domain, with an emphasis on the nuanced understanding fostered by each study.

Zou and colleagues in [1] undertook a seminal endeavor in visualizing the intricate landscape of the road safety knowledge domain. By identifying key topics and exploring the evolution of research interests, this study provides a panoramic view that not only aids in recognizing current trends but also lays the groundwork for predicting future priorities within the realm of road safety research [1]. Cheng et al.'s meticulous exploration of highway roadside safety represents a significant contribution to the field. The study delves into specific concerns and safety measures, offering profound insights that extend beyond the obvious challenges to underscore potential opportunities for enhancing safety on high-speed roadways. This in-depth analysis contributes valuable perspectives to the ongoing discourse on road safety [2].

Schlögl and Stütz embark on an exploration of a pivotal challenge within the realm of road safety research: data uncertainty. Their study stands out as a pivotal reference, shedding light on the methodological hurdles and complexities that come with handling uncertain data. Through their diligent examination of data analysis intricacies, Schlögl and Stütz significantly deepen our comprehension of the nuanced aspects of road safety data. This contribution is particularly valuable for researchers who are navigating the complexities inherent in road safety data, offering them crucial insights and guidance.[3]

Hagenzieker and colleagues undertake a comprehensive journey through the annals of road safety research, employing a quantitative approach to map its historical trajectory. Their analysis uncovers clear trends and notable shifts, offering a window into the research priorities that have historically shaped the field of road safety. By applying quantitative methods to dissect past perspectives, this study furnishes a critical context for appreciating how research priorities in road safety have evolved over time. This meticulous exploration into the history of road safety research enriches the current understanding and lays a foundation for future investigations.[4]

Abou Elassad, Mousannif, Al Moatassime, and Karkouch's (2020) research presents a thorough investigation into driving behavior (DB) analysis through a multi-faceted lens that encompasses the Driver-Vehicle-Environment (DVE) system. This all-encompassing approach allows for a richer comprehension of the complex elements influencing DB. Their pioneering application of Machine Learning (ML) techniques introduces an innovative methodology for DB analysis, adept at navigating the intricate, non-linear datasets derived from diverse sources. Their critical review of empirical studies, especially on the use of ML in DB analysis over the recent decade, sheds light on the progressive development of this research area. The work of Cheng et al. not only underscores the efficacy of ML models in refining DB assessments but also points out the existing challenges and shortcomings within the current research practices. By offering thoughtful

recommendations for future inquiries, their study signifies a substantial advancement in the domain, paving the way for enhanced exploration and application of ML in the analysis of driving behavior. [5]

Zou and colleagues delve into the burgeoning field of driving behavior analysis (DBA), propelled by technological advancements in in-vehicle networks, sensors, and communication systems. Their comprehensive study systematically categorizes DBA techniques according to the types of data utilized, the objectives of the analysis, and the modeling approaches employed. By examining a wide array of DBA data sources and datasets, their research underscores the extensive scope of the discipline. Furthermore, the study explores the application of DBA across several critical areas, including traffic safety, the development of automated vehicles, energy and fuel efficiency, and driver profiling. Highlighting the potential of DBA advancements to significantly mitigate fatal car accidents through the detection of driver inattention or impairment, Zou and colleagues' research not only maps the current landscape of DBA but also pinpoints key challenges and delineates avenues for future investigation. Their insights offer a valuable perspective on the evolving research priorities within road safety and automotive technology. [6]

The study under discussion delves into the critical role of human factors—such as fatigue, distraction, alcohol influence, and reckless behavior—in contributing to traffic accidents. Highlighting the potential of modern smartphones to detect signs of tiredness and inappropriate driving behaviors, this research stands out for its focused review on smartphone-based methodologies. It comprehensively examines various sensing technologies, detection algorithms, their effectiveness, and the limitations identified by researchers. Furthermore, the paper navigates through challenges related to integrating smartphone-based driver behavior classification systems with context-aware technologies, mobile crowdsensing, and active steering control systems. The exploration includes detailed discussions on model training and real-time updates executed both on smartphones and via cloud computing. By offering an exhaustive overview of smartphone-based strategies for monitoring and analyzing driver behavior, this work aims to contribute significantly to enhancing road safety measures. [7]

The study presents a nuanced analysis of how the advent of automotive computerization, advanced sensor technologies, and communication devices has transformed vehicles into repositories of rich data. The unique aspect of this research lies in its approach to harnessing this data deluge to enhance driver safety and comfort. While recognizing the wealth of analytical tools available, the research focuses on exploring less-examined driving safety issues and mathematical approaches within driving behavior studies. The authors propose an innovative methodology for processing and analyzing vehicle-generated data, aimed at automating traffic rule compliance verification and offering a comparative analysis of driver behaviors. This method employs numerical domain abstraction to reduce data volume, integrates probabilistic graphical models with machine learning techniques for constructing a formal model of driver behavior, and utilizes model testing and graph matching for comprehensive analysis. Initial experiments highlighted by the paper indicate that the design of the numerical domain significantly influences the outcome of the analysis, offering valuable insights that could shape future research in driving behavior analysis and vehicle safety enhancements. [8]

The naturalistic driving study (NDS) methodology, valued for its ability to capture driver behavior in real-world contexts devoid of the biases inherent in controlled experiments, is at the forefront of the authors' critical examination. The report delves into the role of NDS in shedding light on the contributory factors of crashes attributed to distracted or drowsy driving each year, with a particular focus on how NDS-derived video records can reveal driver errors. It undertakes a meticulous review of studies that leverage computer vision technologies to autonomously evaluate video data and classify driving behaviors, thereby bridging NDS with advancements in computer vision research. This comprehensive review distinguishes between research efforts based on their focus on low-level (e.g., head orientation) versus high-level (e.g., distraction detection) driver information, as well as their reliance on public versus various proprietary datasets, discussing the design of data collection and model performance across these studies. By scrutinizing the methodologies and effectiveness of these research endeavors, the report establishes a reference point in the field, critically comparing tools for analyzing NDS video data to pinpoint existing research gaps. This exhaustive approach not only provides the computer vision community with detailed technical insights but also equips NDS researchers with a deeper understanding of driver behavior, paving the path for future explorations in both spheres. [9]

In their research, the authors unveil a cutting-edge strategy for the identification and modeling of risky driving behaviors, leveraging the capabilities of advanced sensors alongside sophisticated machine learning algorithms. Centering their analysis on data obtained from accelerometers and gyroscopes, they employ various algorithms—including the C4.5 Decision Tree, Random Forest, and K-Star—to forge precise driver profiles. Remarkably, the application of the K-Star algorithm culminates in a flawless accuracy rate of 100%. This methodology paves the way for the creation of an efficient and cost-effective solution, reminiscent of a driver's black box, which holds substantial potential for augmenting road safety. Furthermore, this system offers insurance companies a novel avenue to encourage safer driving practices through the adoption of usage-based insurance policies. [10]

The authors introduce an innovative methodology aimed at enhancing road safety within mixed-traffic contexts by utilizing data-driven models that forecast driving risks specific to individual drivers. By implementing sophisticated regression techniques and harnessing data from the SHRP 2 study, their research significantly improves the efficacy of driver assistance systems. This approach yields encouraging outcomes and opens avenues for future innovations in traffic safety technologies. [11]

The paper by the authors delves into the dynamic landscape of driver behavior modeling (DBM), spotlighting the recent breakthroughs in wireless communication, mobile computing, and context-aware services. Providing a detailed survey of the newest innovations in in-vehicle and smartphone sensing technologies and their utilization in DBM, the work underscores the principal research hurdles and prospective pathways in this swiftly evolving field. [12]

In their paper, the authors explore the rapidly advancing field of driver behavior modeling (DBM), emphasizing the latest progressions in wireless communication, mobile computing, and context-aware services. The paper provides an extensive overview of current innovations in both in-vehicle and smartphone sensing technologies and their applications within DBM. It highlights the significant research challenges and anticipates future trends in this swiftly evolving area. [13]

The study centers on enhancing road safety within the framework of the Internet of connected vehicles (IoCV) through the examination and response to driver behavior. By applying sophisticated clustering and neural network methodologies for the effective analysis of vehicles and the rapid dissemination of emergency alerts, it showcases the significant capabilities of IoCV in improving driver assistance systems and overall road safety. [14]

The authors advance the capabilities of the Advanced Driver Assistance System (ADAS) through the innovative integration of an AdaBoost Multi-class Support Vector Machine (MSVM) with a Cat Mouse Optimizer (CMO) algorithm. This methodology adeptly addresses the challenges associated with high-dimensional and noisy data prevalent in ADAS, thereby enhancing the system's overall performance. Validated with data from the Jiangxi bus company, their approach evidences notable enhancements in accuracy and other critical metrics, representing a significant leap forward in ADAS technology and vehicle safety. [15]

This study addresses the challenge of formalizing the entire driving task within the context of a sensing-acting robotics system to achieve complete vehicle autonomy. Given the intricacies of real-world driving conditions and the presence of edge cases, it underscores the necessity for human oversight of autonomous driving systems. The MIT Autonomous Vehicle Technology (MIT-AVT) initiative plays a crucial role in this field by gathering extensive real-world driving data, including high-definition video, to craft deep learning-based perception systems. It also assesses how humans interact with vehicle automation by correlating video data with a variety of parameters, aiming to enhance road safety through the adoption of technology and automation. The study's methodology includes equipping diverse car models for both long-term and medium-term data collection, accumulating a vast dataset encompassing IMU, GPS, CAN communications, and high-definition video feeds. It details the study's design, the hardware utilized, data processing techniques, and the computer vision algorithms employed to extract actionable insights from the collected data. [16]

The study aims at enhancing transportation safety and mobility by delving into driving behaviors. While simulator-based and naturalistic driving studies offer insights into the interplay between demographics, road conditions, and safety, their high cost and extensive time requirements limit their accessibility. Capitalizing on the ubiquity of cellphones, this research leverages GPS, accelerometer, gyroscope, and camera data to augment traditional methodologies. It constructs statistical models using data from tens of thousands of drivers in the San Francisco metro area to dissect driving behavior. Analyzing traffic speed and driver movements with mobile sensor data from 500 drivers, the study categorizes drivers according to their driving behaviors, thereby uncovering street-level norms and deviations in driving practices. [17]

This study marries modern computer vision and AI to scrutinize traffic conflicts and their characteristics at signalized intersections in real-time. Acknowledging the limitations of extreme value theory in forecasting future crash risks due to its neglect of temporal associations, the research innovatively integrates extreme value theory with autoregressive integrated moving average (ARIMA) models. Utilizing video data from Queensland intersections, it employs a non-stationary generalized extreme value model to calculate real-time risks of rear-end crashes at the

signal cycle level. This approach captures the dynamic influence of conflict extremes over time, factoring in traffic flow, speed, shockwave areas, and platoon ratios. Crash risks identified at the signal cycle level are treated as a univariate time series, which ARIMA models then exploit to forecast future crashes. The findings reveal that the model, when augmented with exogenous variables, can accurately forecast crash probabilities for durations of 30–35 minutes, thus facilitating a proactive safety evaluation and the identification of temporal and spatial periods where safety conditions are deteriorating. [18]

This study tackles the pervasive issue of inattentive driving, a major cause of traffic accidents annually, by analyzing extensive vehicle trajectory data to identify signs of driver inattention and its impact on the behavior of commercial vehicle drivers. Focusing on key inattentive behaviors such as smoking, phone usage, turning around, and yawning, the research employs a sophisticated approach using a deep convolutional neural network (CNN), specifically Inception v3, complemented by data augmentation techniques like Mixup and SMOTE to enhance the training data distribution and the generalizability of the classification model. Furthermore, a model based on long short-term memory (LSTM), incorporating point of interest (POI) and climate data, is developed to forecast inattention-related driving anomalies, potentially leading to hazardous conditions such as abrupt acceleration or deceleration and aggressive lane changes. Tested on over 120,000 real-world driving records from 200 drivers, the experimental outcomes demonstrate a weight accuracy (WA) of 92.27% for detecting inattentive driving and 91.67% for predicting abnormal driving behavior, underscoring the approach's efficacy in advancing road safety and improving driving practices. [19]

Addressing the need for enhanced driving safety, the study introduces D3, a system designed for the fine-grained detection and identification of anomalous driving behaviors in real-time. Unlike previous smartphone-based efforts that broadly categorized driving behaviors as normal or abnormal, D3 provides a nuanced approach to specifically recognize and classify particular types of aberrant driving actions. Drawing from six months of driving data in real-world conditions, the authors pinpoint six distinct anomalous behaviors: Weaving, Swerving, Sideslipping, Fast U-turn, Wide-radius turning, and Sudden braking, noting unique acceleration and orientation patterns for each. Leveraging these observations, D3 employs Support Vector Machine (SVM) and Neural Network (NN) algorithms to train a classifier model capable of identifying these specific driving anomalies. Through extensive testing involving 20 volunteers across four months of real driving scenarios, D3 demonstrates impressive accuracy, achieving an average total accuracy of 95.36% with the SVM model and 96.88% with the NN model, highlighting its potential to significantly improve road safety. [20]

The advent of the Internet of Things (IoT) and digital innovation presents unparalleled opportunities to revolutionize service delivery, usage, and resource management paradigms. Within the realm of connected vehicles, IoT facilitates the integration of objects with the Internet, enabling vehicles to communicate autonomously. Equipped with advanced sensors, these vehicles can monitor a wide array of factors, including internal components, driver behavior, road and weather conditions, and traffic congestion. Machine learning (ML) models play a crucial role in transforming this raw data into actionable insights, thereby informing decision-making processes, strategies, and the allocation of resources. As a result, intelligent mobility projects are increasingly leveraging artificial intelligence (AI) and ML to harness the

power of data-driven decision-making. This study employs unsupervised learning to analyze a vehicle IoT dataset, aiming to derive insights into geographic zones based on historical driver behavior. Such analysis enables the autonomous vehicle framework to optimize routes and proactively address significant challenges. [21]

This research introduces a pioneering methodology for the analysis and classification of driver behavior, leveraging signals recorded by modern vehicles' CAN bus technology. These signals, which include gas pedal position, brake pedal pressure, steering wheel angle, momentum, velocity, RPM, and both frontal and lateral acceleration, offer real-time insights into the vehicle's operation, the driver's actions, and environmental conditions. Data was collected through an uncontrolled trial involving 64 participants driving 10 different cars across more than 2000 journeys in various road conditions, without any specific instructions provided. The study employs unsupervised learning techniques to cluster drivers and assesses the robustness of these clusters across different experimental settings. Furthermore, it determines the minimum amount of data required to achieve reliable driver clustering. Ultimately, this work presents a novel approach to categorizing driver behavior in near-real-time and under uncontrolled conditions. [22]

Bouhoute et al. (2018) delve into the realm of advanced driving behavior analytics by harnessing the power of vehicle computerization, sensor technologies, and communication devices, setting their work apart by focusing on enhancing driver safety and comfort. While acknowledging the existence of analytical methods, they venture into less explored areas concerning critical driving safety issues and mathematical approaches. They propose an innovative methodology for processing and analyzing data generated by cars, aiming to automate the verification of traffic rule adherence and scrutinize driver behavior. The approach employs probabilistic graphical models and machine learning to construct a detailed driver behavior model, with numerical domain abstraction serving to streamline data volume. Through rigorous model testing and graph matching, a thorough analysis is conducted. Initial tests underscore the significant impact of numerical domain configuration on the results of the analysis, offering crucial insights for future research in driving behavior analysis and vehicle safety. [23]

The global impact of road traffic accidents on both the economy and public health is significant, with urban areas like Peshawar in Pakistan experiencing a notable increase in vehicle traffic accidents. A primary factor contributing to these accidents is driver behavior. This study aims to explore and model the influence of driver traits on road traffic accidents, navigating the complexity of correlating driver characteristics with accident rates. Employing the Artificial Neural Network (ANN) technique and utilizing survey data, the research adopts a flexible and assumption-free methodology. The results demonstrate the capability of the ANN technique to predict driver involvement in road traffic incidents, offering a viable alternative to costly and time-consuming psycho-technical research studies. [24]

Vulnerable road users (VRUs) face significant dangers in road traffic, making the analysis of injury severity factors crucial for their safety. Traditional machine learning approaches may overlook critical data and lack a comprehensive analysis. This study introduces a holistic analytical framework utilizing the stacked sparse autoencoder (SSAE) deep learning model to predict traffic accident injury severity based on various contributing factors. The research

employs CatBoost machine learning to scrutinize contributory components and eliminate those with low correlation. Additionally, it applies k-means clustering with geographical information to organize the data. The core of the framework is an SSAE-based deep learning model that forecasts injury severity, focusing on data classes identified through strong correlation factors. Empirical testing with actual traffic accident data validates the framework's effectiveness, underscoring CatBoost's efficiency in identifying key factors and the superior accuracy of the SSAE-based model compared to traditional methods. This innovative approach offers a new pathway for assessing severity and risk indicators crucial for enhancing VRU safety in traffic accident analyses. [25]

To mitigate the global issue of aggressive driving and its contribution to road accidents, there's a growing need for effective driving awareness and safety strategies. The utilization of cellphones for driving analytics (DA) has become increasingly prevalent, as they are capable of capturing driving patterns through onboard sensors. Previous studies typically employ statistical scores or raw data formats to interpret sensor data, applying either threshold-based heuristics or machine learning (ML) algorithms. This research introduces an innovative bag-of-words-based second-order representation for accelerometer data timestamps associated with aggressive driving maneuvers. This novel representation significantly improves the F-measure for both binary and multi-class classification tasks across two scenarios and three datasets, outperforming current state-of-the-art methodologies. Furthermore, when compared with similar second-order algorithms, our method demonstrates superior effectiveness, suggesting its potential to refine behavior categorization and enhance the discriminative power of ML techniques in driving behavior analysis. [26]

Exploring traffic accident behavior through traditional methodologies often proves costly and yields ambiguous outcomes. This research adopts a novel approach by coding 200 cross-flow junction traffic incidents from the files of Nottinghamshire Constabulary in the UK. The incidents are processed for computer analysis using a bespoke Traffic Related Action Analysis Language, aiming to explore alternative investigative methods. By applying both computer-based and statistical analyses, the study ventures into the realm of artificial intelligence, utilizing Quinlan's 'ID3' algorithm to generate decision trees. These trees categorize incidents into various categories: those causing injury or only damage, incidents involving young male drivers, and scenarios varying in danger levels. The research demonstrates that factors such as the involvement of another vehicle, seasonal variations, junction types, and the Turner's failure to observe another road user can predict the severity of turn-on collisions on main roads with a 79% accuracy rate. Additionally, the complexity of the junction and the Turner's actions before turning (waiting or slowing down) could predict accidents involving young male drivers with a 77% accuracy. This innovative study leverages machine learning techniques on police records to decipher causation in road junction accidents. [27]

This research offers a comprehensive review of cutting-edge methodologies for developing intelligent road safety systems through the integration of the Internet of Things (IoT) and Machine Learning technologies. It meticulously examines various aspects of road safety, including driver behavior analysis, vehicle condition monitoring (covering both two-wheelers and four-wheelers), assessment of road and bridge integrity, and RFID-based theft prevention

mechanisms. The utilization of IoT facilitates the dynamic updating of road safety systems, ushering in a new era of smart, intelligent, and efficient infrastructure. Furthermore, the study sheds light on the application of Artificial Intelligence in enhancing the detection of driver drowsiness via real-time video surveillance or high-definition images. Additionally, AI plays a crucial role in evaluating the condition of roads and bridges to mitigate accident risks. While this review outlines the significant potential of IoT and Machine Learning in advancing smart road safety solutions, it also identifies existing gaps, signaling the need for further exploration in this domain. [28]

Road accidents pose a significant threat to global public safety, driving the need for advanced algorithms and methodologies for analyzing and predicting traffic incidents. This study conducts a thorough review of current machine learning algorithms and sophisticated analytical techniques, such as convolutional neural networks and long short-term memory networks, used for road accident prediction. It further classifies data sources for road accident forecasting based on their origin and characteristics, encompassing open data, measurement technologies, onboard equipment, and social media inputs. The research compares various road accident prediction systems, evaluating their adaptability to different data types, ease of interpretation, and analytical straightforwardness. It highlights that the integration of multiple analytic approaches often leads to more robust outcomes. For enhanced analysis and forecasting accuracy, the study suggests that road traffic prediction models should incorporate a diverse array of data sources, including geo-spatial information, traffic volumes, statistical data, and inputs from video, sound, text, and social media sentiment analysis. [29]

Artificial intelligence (AI) integrated into driving technology has raised significant ethical concerns, particularly regarding the moral decision-making process in unavoidable road traffic incidents. This study delves into whether the life-or-death choices made by AI-driven vehicles necessitate moral intelligence and informed moral judgments. While existing laws and safety regulations govern driving conduct, the authors contemplate whether ethical theories, human values, and rights frameworks should also inform the decisions of AI-driven vehicles. Additionally, the research addresses the potential biases inherent in moral decision-making algorithms and proposes utilizing Philippa Foot's trolley problem as an experimental framework to compare the moral reasoning of human drivers and AI systems. The authors argue that leveraging the trolley problem can shed light on the moral decision-making capabilities of AI-driven vehicles by revealing the ontological disparities between human and artificial decision-making processes. [30]

Key Takeaways from the Literature:

The collective body of literature reviewed here contributes to a rich and nuanced understanding of road safety research. The insights garnered range from macro-level knowledge domain mapping to addressing specific concerns such as roadside safety and methodological challenges. Furthermore, the historical analyses offer an indispensable context for comprehending the evolution of research priorities in road safety.

Each of these studies serves as a cornerstone, enriching the foundation for our research project by addressing distinct facets of this complex field. As we embark on our research journey, these insights will guide our efforts to bridge identified gaps, contributing to the ongoing scholarly discourse in road safety research.

This literature review sets the stage for the subsequent sections of the thesis, where we will build upon these foundational insights to formulate a robust research framework and conduct a meticulous analysis that aligns with the rigorous standards set by the scholarly contributions explored herein.

Overall Key Takeaways:

- Road safety research encompasses a wide range of factors influencing accident severity and safety measures.
- Seminal studies have contributed significantly to understanding the road safety knowledge domain.
- Visualizing the landscape of road safety research aids in recognizing trends and predicting future priorities.
- Creative Mapping Techniques that offers us a comprehensive view of how knowledge in road safety is structured and evolving.
- The usage of newer of cutting-edge technology, creative engineering solutions, and the evolving strategies that reacts in real-time to potential hazards or data help us understand and address risks promptly.
- Machine Learning in car accidents can assist in cleaning data, and determining patterns and trends in past incidents data.
- Data Analytics can play a crucial role in improving the efficiency, effectiveness and to force better road policies.

Research Gaps Identified:

- While studies have visualized the road safety knowledge domain, there may be a need for further exploration into emerging topics and their potential impact on future research priorities.
- The exploration of highway roadside safety presents valuable insights, but additional research may be needed to address specific challenges in diverse geographical and infrastructural contexts.
- Addressing data uncertainty is highlighted, but there may be gaps in understanding how different sources and types of data uncertainty impact the reliability of road safety analyses.
- Quantitative analysis of historical trends provides context, yet further research could delve into the drivers behind shifts in research priorities and their implications for current road safety challenges.

Chapter – 3 Research Methodology

The research methodology for this study is designed to investigate various factors influencing the severity of car accidents. This chapter outlines the approach taken to address the research objectives and specific questions raised during the analysis.

3.1 Restatement of Research Objectives

The research objectives of this study are centered around understanding the factors contributing to the severity of car accidents. The specific questions addressed include the influence of gender, the role of speeding, the impact of external conditions (such as night or rainy days), and the relation between age, education, and driving experience with accident severity.

3.2 Research Philosophy

The research philosophy underlying this study is pragmatic, aiming to combine elements of both positivism and interpretivism. Positivism is embraced through quantitative analysis of structured data, while interpretivism is considered in understanding the context and nuances surrounding accidents.

3.3 Research Strategy

The chosen research strategy involves a combination of descriptive statistics, probability calculations, and data visualization. Descriptive statistics, including mean, mode, and median, are utilized to provide a comprehensive overview of the dataset. Probability calculations are employed to assess the likelihood of specific events, such as accidents being fatal given certain conditions.

3.4 Limitations of Secondary Research

The study acknowledges the limitations inherent in secondary research, particularly the reliance on existing datasets. Gaps identified in the literature review emphasize the need for primary research to fill these voids and provide more nuanced insights.

3.5 Data Collection and Instrumentation

The primary dataset used for this analysis is sourced from road traffic accident records. A structured dataset is employed, encompassing information on accident severity, gender, age, education, driving experience, and various external factors. Data collection instruments include Python programming for data cleaning, manipulation, and statistical analysis.

1.1.1 Creating df containing only relevant columns

We have created a new Data Frame named `df` containing only the relevant columns selected for our analysis from the merged dataset. These columns encompass critical variables such as the sex of the driver, accident severity, speed limit, vehicle maneuver, time of the accident, road surface conditions, road type, weather conditions, pedestrian crossing control, driver's journey purpose, skidding and overturning incidents, and the age band of the driver. By extracting this subset of columns, we aim to focus our analysis on the factors most pertinent to our research questions regarding traffic accident severity. This streamlined Data Frame will facilitate a more targeted and efficient exploration of the underlying patterns and

relationships within the data, ultimately aiding us in deriving actionable insights to enhance road safety initiatives.

1.1.2 Handling Missing Values

After scrutinizing the {df} Data Frame, we detected missing data in many columns. The method `isna().sum()` provided information about how many values were missing in each column. There is varied degrees of missing data in the following columns: 'Speed_limit', 'Time', 'Pedestrian_Crossing Human_Control', 'Pedestrian_Crossing-Physical_Facilities', and 'Skidding_and_Overturning'. A significant number of missing values, 1,798,784, are present in the following columns: 'Speed_limit', 'Time', 'Pedestrian_Crossing-Human_Control', 'Pedestrian_Crossing-Physical_Facilities', and 'Skidding_and_Overturning'. In order to guarantee the accuracy and dependability of our study, it is essential that these missing values be adequately addressed. Before moving on to more data exploration and analysis, we'll treat these missing values using appropriate methods like imputation or exclusion.

1.1.3 Dropping Missing Values

`df_cleaned` was the new Data Frame constructed after all columns except 'Skidding_and_Overturning' had their missing values removed. To remove the column "Skidding_and_Overturning" from the operation, the `dropna()` function was used with the option `subset=[col for col in df.columns if col!= 'Skidding_and_Overturning']}. Using the {isna().sum()} method, it was possible to check the missing values in the {df_cleaned} Data Frame and see that all columns—aside from 'Skidding_and_Overturning'—no longer had any missing values. Still, 'Skidding_and_Overturning' displays 1,797,744 missing data, which is a significant amount. Effectively handling these missing information is essential to guaranteeing the correctness and dependability of ensuing analyses and interpretations. To properly address this problem, methods like imputation or more research into the nature of missing data may be required.

1.1.4 Handling Data Types

After handling missing values, the `df_cleaned` Data Frame was inspected using the `info()` method to understand the data types of its columns. The Data Frame consists of 13 columns, with a total of 2,056,726 entries. Of these columns, three are of the float64 data type, which indicates numerical variables, while the remaining ten are of the object data type, which indicates categorical variables. 'Sex_of_Driver', 'Accident_Severity', 'Vehicle_Manoeuvre', 'Time', 'Road_Surface_Conditions', 'Road_Type', 'Weather_Conditions', 'Journey_Purpose_of_Driver', 'Skidding_and_Overturning', and 'Age_Band_of_Driver' are some of the columns that store object data types. Categorical data including the driver's age group, weather, accident severity, and gender are all contained in these columns. Concurrently, the float64 data type columns 'Pedestrian_Crossing-Human_Control', 'Pedestrian_Crossing-Physical_Facilities', and 'Speed_limit' indicate numerical variables associated with pedestrian crossing controls and speed limits. In order to prepare the information for additional analysis and modeling, it is imperative to comprehend and handle different types of data effectively. Conversions or transformations can be required, depending on the analytical methods used, to guarantee accuracy and compatibility in later studies.

3.6 Data Analysis Plan

Data analysis involves a combination of descriptive statistics, probability calculations, and graphical representations. Descriptive statistics are used to summarize key characteristics of the dataset, while probability calculations provide insights into the likelihood of specific events. Graphical representations, including bar charts and pie charts, facilitate a visual understanding of patterns and trends.

3.7 Sampling Criteria and Response Rate

The dataset includes records from a specific time frame and geographical area, ensuring relevance to the research questions. The response rate is not applicable as the study relies on a complete dataset of recorded accidents.

3.8 Conclusion

This chapter outlines the methodology employed to address the research objectives and questions. The combination of quantitative and qualitative approaches aims to provide a holistic understanding of the factors influencing the severity of car accidents. The subsequent chapters will delve into the detailed findings and discussions derived from the applied methodology.

Chapter 4- Findings and Data Analysis

This chapter presents the outcomes of an in-depth analysis of road traffic accident data, aiming to unravel patterns, relationships, and influential factors contributing to the severity of car accidents. The investigation spans various dimensions, including demographic factors, external conditions, and driving-related variables. The findings disclosed herein provide valuable insights into the dynamics of car accidents, shedding light on critical aspects that impact their severity.

Hypothesis 1: Which of both genders is dangerous in Driving?

Hypothesis 1 delves into the analysis of accident severity, differentiating between male and female drivers, the study investigates the likelihood of accidents occurring given a fatal severity for both genders.

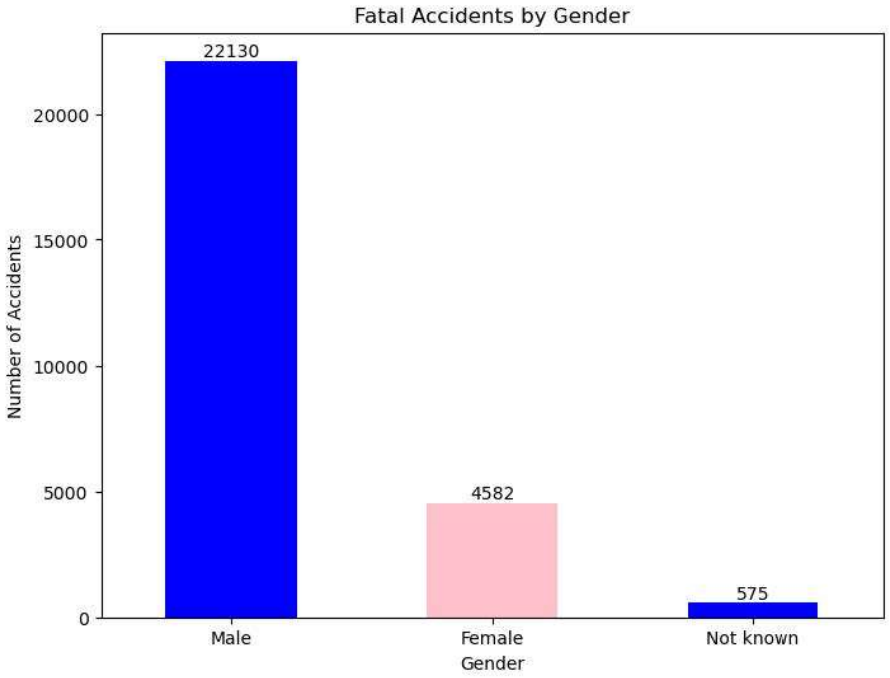

Fig 1.0

The analysis conducted on the dataset aimed to address the research question of whether there are gender disparities in driving accidents of fatal severity. By filtering the dataset to include only accidents categorized as fatal, it was possible to compare the frequencies of fatal accidents between male and female drivers. The results revealed a notable discrepancy, with male drivers being significantly more involved in fatal accidents compared to their female counterparts. Specifically, the data indicated a total of 22,130 fatal accidents involving male drivers, contrasting with 4,582 fatal accidents involving female drivers. Interestingly, a category labeled "Not known" was also identified, indicating a lack of gender identification in some reported cases, comprising 575 accidents. This analysis underscores the importance of considering gender dynamics in road safety initiatives and highlights the need for targeted interventions to address the specific risk factors associated with each gender.

Sub-question: What is the probability of an accident being fatal given the gender of the driver?

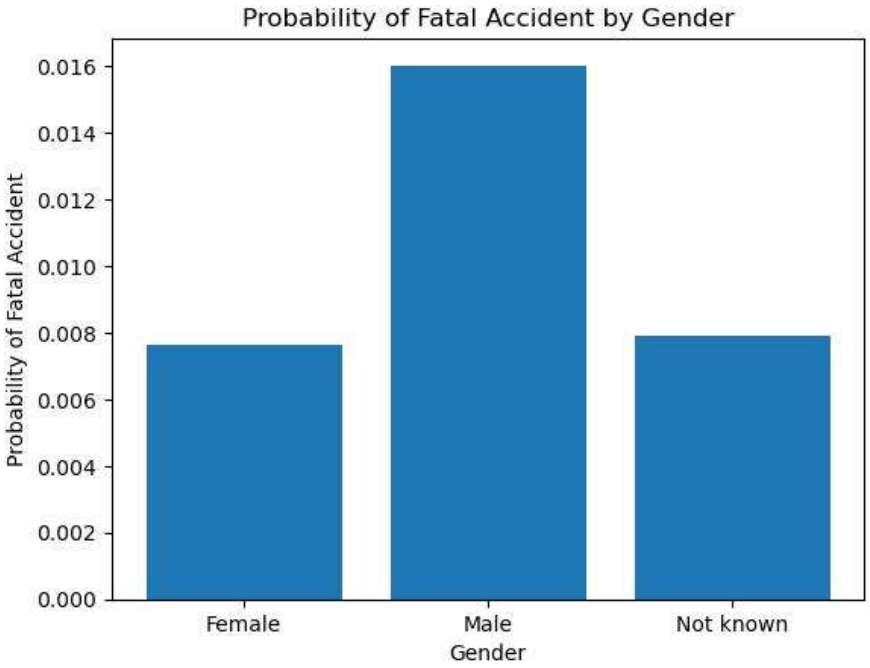

Fig 2.0

The investigation that was done looked more closely at the connection between gender and the risk of fatal accidents. It was able to calculate the likelihood of a fatal accident happening given the driver's gender by adding up all of the accidents and all of the fatal accidents for each gender group. The results revealed compelling insights into this relationship. Specifically, male drivers exhibited a significantly higher probability of involvement in fatal accidents, with approximately 1.6% of accidents resulting in fatalities. In contrast, female drivers had a lower probability, approximately 0.76%, highlighting a gender disparity in accident severity. It's interesting to note that crashes in which the driver's gender was unknown also had a high likelihood of being fatal—roughly 0.79%—underscoring the need of precise data collection and gender identification in the research of road safety. These results highlight the need for gender-sensitive strategies when creating focused interventions meant to lessen the severity of collisions and encourage safer driving behaviors for all users of the road.

Hypothesis 2: Reassessing the Role of Speeding in Car Crashes

Hypothesis 2 aims to scrutinize the prominence of speeding as a primary cause of car crashes. Employing a graphical representation of the probabilities associated with each cause, the analysis challenges the common assumption that high speed is a predominant factor in accidents.

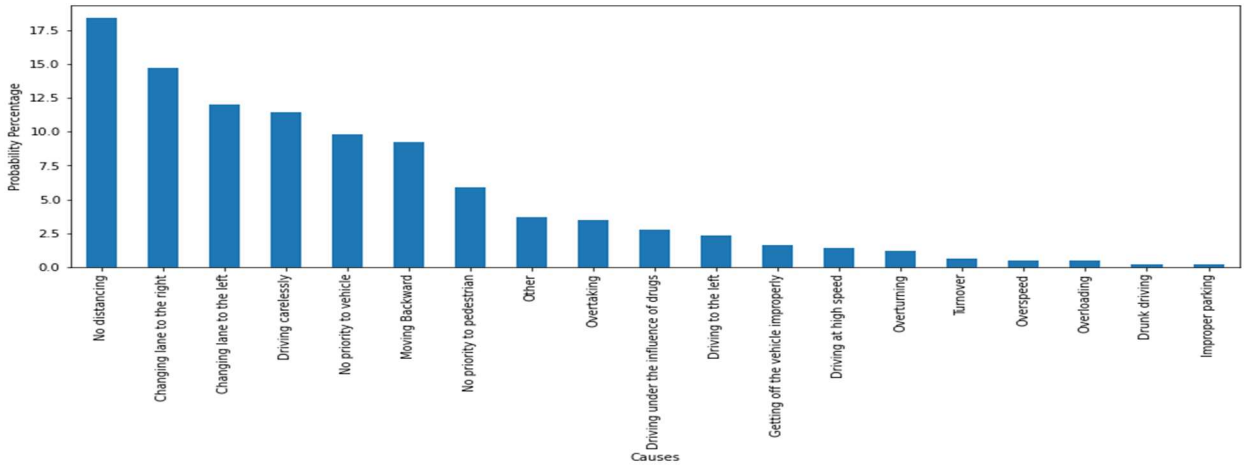

Fig 3.0

The initial graph illustrates the distribution of probabilities for all causes, debunking the notion that speed is a leading cause, as it appears towards the tail of the graph. To delve deeper, the top four causes are isolated for a more focused examination, further emphasizing that speed does not dominate the landscape of accident causation

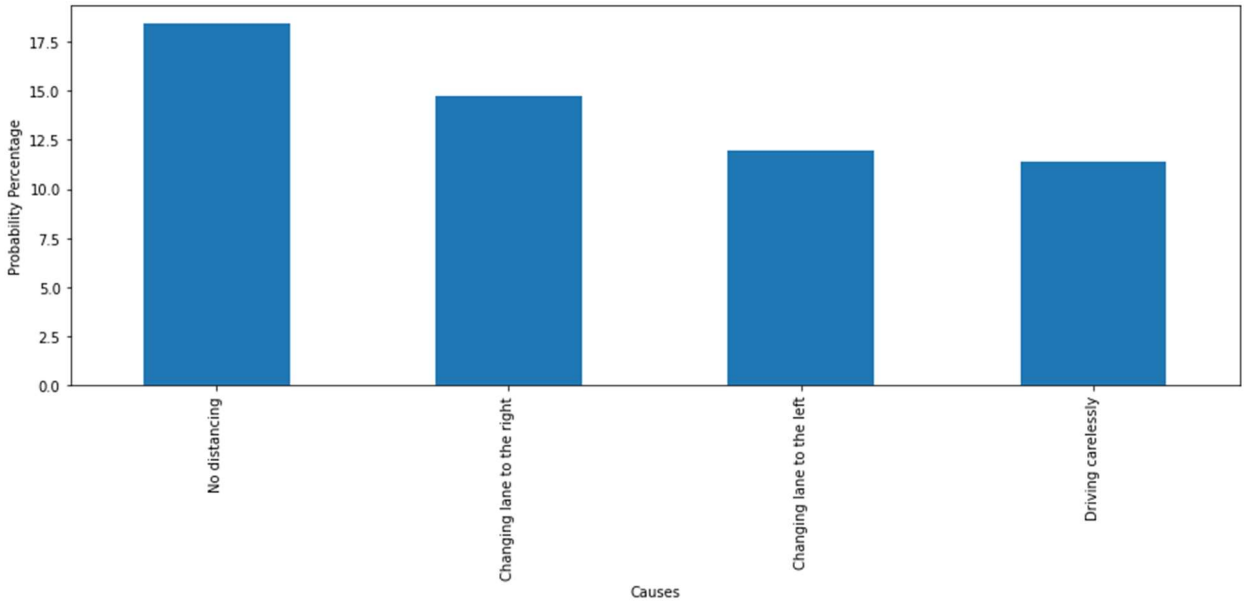

Fig 4.0

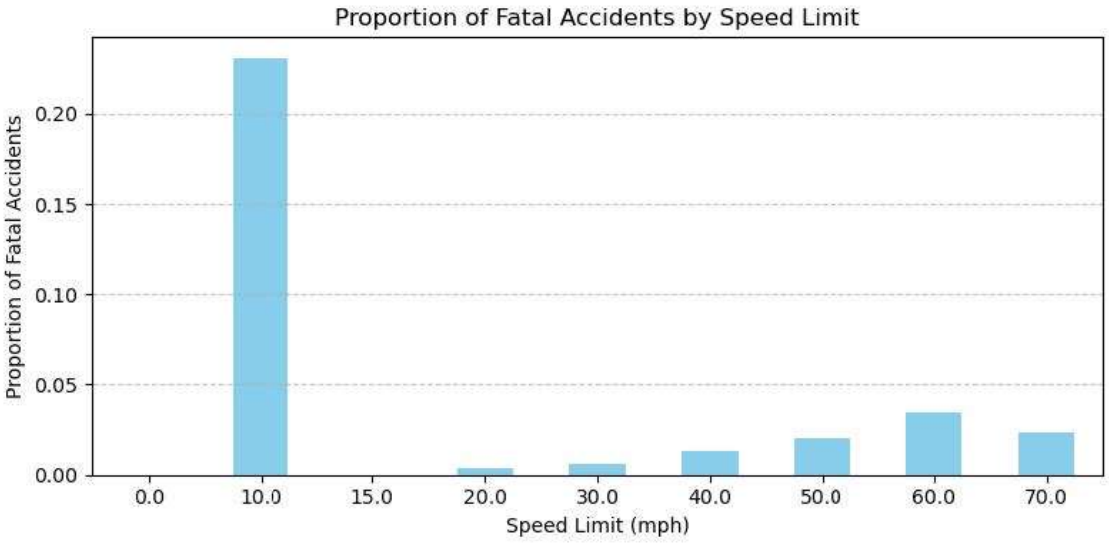

Fig 5.0

Sub-question: What is the probability of a crash being fatal given that speeding is the cause?

By examining the dataset, particularly focusing on the relationship between speed limits and accident severity, insightful conclusions were drawn. The summary of results indicates the proportion of fatal accidents relative to the total number of accidents at different speed limits. Interestingly, almost 23.08% of incidents with a speed restriction of 10 mph ended in fatalities, indicating a comparatively high percentage of catastrophic outcomes at this pace. On the other hand, compared to lower speed restrictions, speed limits of 20 mph and 30 mph demonstrated lower proportions of fatal accidents, at 0.40% and 0.61%, respectively, suggesting a decreased risk of fatal consequences. The percentage of fatal accidents did, however, gradually climb when the speed limit was raised, hitting 2.32% at 70 mph and 3.49% at 60 mph. These results highlight the crucial role that speed regulation and enforcement play in reducing road traffic deaths by suggesting a positive link between higher speed limits and a higher risk of fatal accidents.

Hypothesis 3: Assessing Severity Disparity Between Day and Night Accidents

To investigate whether accidents occurring at night exhibit greater severity than those during the daytime, a statistical analysis was conducted using a one-tailed Z-test. The process involves defining hypotheses, setting the significance level, calculating the test statistic, and determining the decision criteria.

Light_conditions	count	mean	std
Darkness - lights lit	3286	1.818320	0.434626
Darkness - lights unlit	40	1.825000	0.384808
Darkness - no lighting	192	1.692708	0.516076
Daylight	8798	1.841328	0.391523

Table 1.0

Step 1: Define the Hypotheses

- Null Hypothesis (H0): $\mu n = \mu d$
- Alternative Hypothesis (Ha): $\mu n > \mu d$

Step 2: Significance Level

- $\alpha = 0.05$

Step 3: Z-test Parameters

- Zc = 1.645 (for a one-tailed test)

Step 4: Calculate the Test Statistic

- Sample sizes: nn = 3286, nd = 8798
- Sample means: $\bar{x}n = 1.818320$, $\bar{x}d = 1.841328$
- Sample standard deviations: Sn = 0.434626, Sd = 0.391523
- Z= 2.658338
-

Step 5: Decision Criteria Since Z < Zc, we reject H0 in favor for the Alternative Hypothesis (Ha).

Conclusion Based on the statistical analysis conducted using a one-tailed Z-test with a significance level of $\alpha = 0.05$, the findings provide strong evidence to reject the null hypothesis (H0: $\mu n = \mu d$) in favor of the alternative hypothesis (Ha: $\mu n > \mu d$). The calculated test statistic (Z = 2.658338) exceeds the critical value (Zc = 1.645), indicating that accidents occurring at night are statistically significantly more severe than those happening during the daytime.

Therefore, it can be concluded with a confidence level exceeding 95% that there is a significant severity disparity between accidents occurring at night and those occurring during the daytime. This underscores the importance of considering factors such as visibility, fatigue, and road conditions when addressing road safety measures and implementing strategies to mitigate the severity of accidents, particularly during nighttime hours.

Sub-question: What is the probability of a night time accident being fatal?

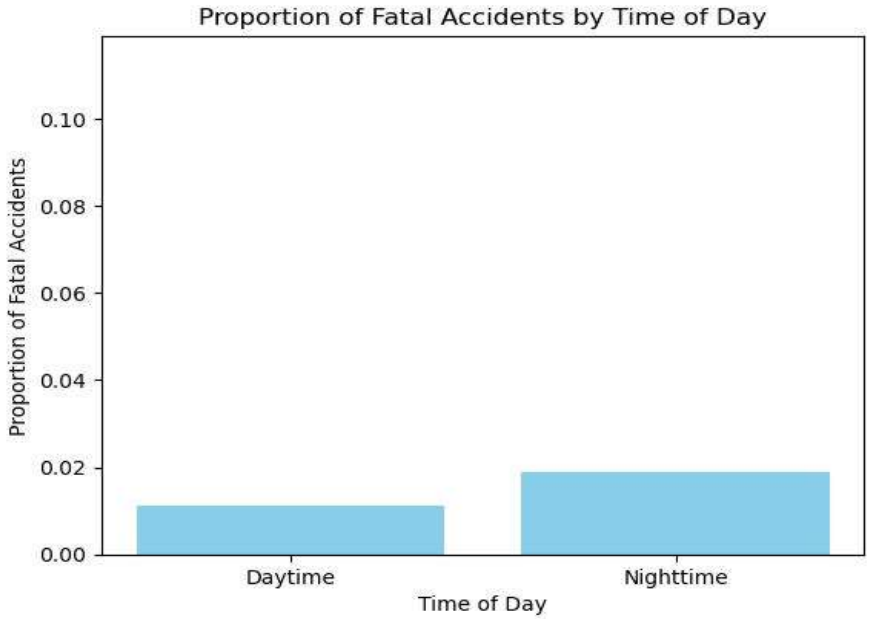

Fig 6.0

The analysis sought to ascertain the probability of a nighttime accident resulting in a fatality, addressing a key sub-question regarding the severity of accidents based on the time of day. By comparing accident severity between daytime and nighttime, categorized based on the time of occurrence, insightful conclusions were drawn. The results revealed a notable difference in the proportion of fatal outcomes between accidents occurring during daytime and nighttime. In particular, the data showed that accidents that occurred during the day had a lower proportion of fatal outcomes, around 1.1%, but accidents that occurred at night had a larger proportion, about 1.7%. This data implies that accidents are typically more serious at night, highlighting the vital need for drivers to exercise greater caution and awareness at night in order to reduce the risk of deaths. These findings highlight the need of putting specific measures in place to improve road safety, especially when visibility is low and there may be increased risk factors related to driving at night.

Hypothesis 4: Assessing Severity Disparity Between Asphalt and Earth Roads

Research Question: Are accidents on asphalt roads more severe than those on earth roads?

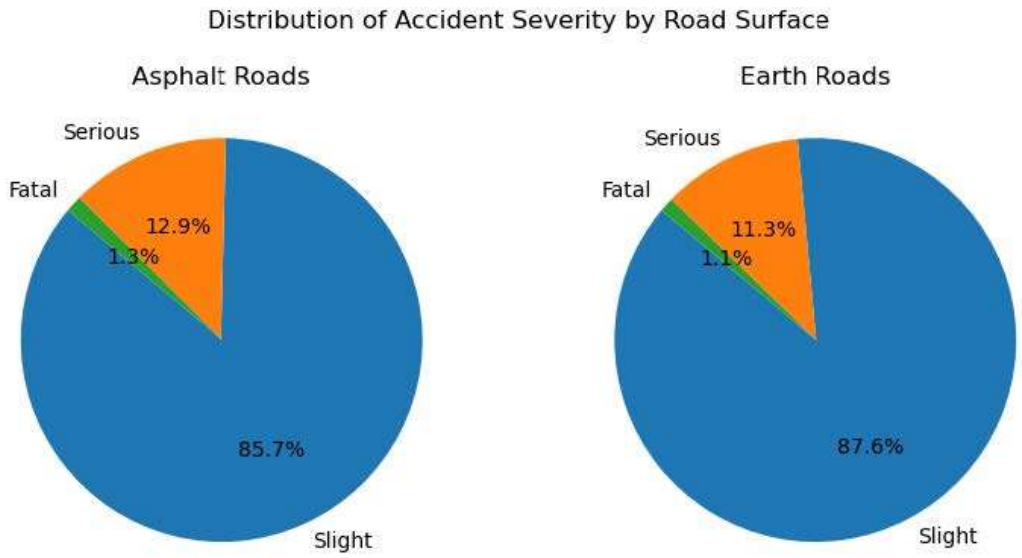

Fig 7.0

The analysis aimed to investigate whether accidents on asphalt roads exhibit greater severity compared to those on earth roads, addressing a pivotal research question in road safety. By categorizing accidents based on road surface conditions and examining their severity distributions, insightful conclusions were drawn regarding the impact of road surface type on accident severity. The results revealed notable disparities in accident severity between asphalt and earth roads. On asphalt roads, the majority of accidents resulted in slight severity, with a significant number of serious and fatal incidents as well. Specifically, there were 1,715,017 slight, 259,077 serious, and 26,664 fatal accidents recorded on asphalt surfaces. In contrast, the number of accidents on earth roads was considerably lower across all severity levels, with 46,399 slight, 5,972 serious, and 607 fatal incidents. These findings suggest that accidents on asphalt roads tend to have higher severity compared to those on earth roads, as evidenced by the greater number of serious and fatal incidents on asphalt surfaces. This underscores the importance of considering road surface conditions in

road safety planning and infrastructure maintenance to mitigate the severity of accidents and enhance overall road safety outcomes.

Sub-question: What is the probability of an accident on asphalt roads being fatal?

The analysis sought to determine the probability of an accident occurring on asphalt roads resulting in fatality, addressing a critical aspect of road safety planning. By examining the dataset and calculating the proportion of fatal accidents among those occurring on asphalt roads, insightful conclusions were drawn regarding the risk associated with accidents on this type of surface. The results revealed that out of all accidents recorded on asphalt roads, approximately 1.33% resulted in fatalities. This percentage provides valuable insight into the relative risk of fatal accidents specifically on asphalt roads compared to the total number of accidents occurring on these surfaces. Understanding this probability is essential for informing road safety policies and interventions aimed at mitigating the severity of accidents on asphalt roads and ultimately enhancing overall road safety outcomes.

Hypothesis 5: Assessing Severity Disparity Between Rainy and Normal Days

In order to investigate whether accidents on rainy days exhibit higher severity compared to those on normal days. The analysis involved formulating hypotheses, setting the significance level, calculating the test statistic, and interpreting the results.

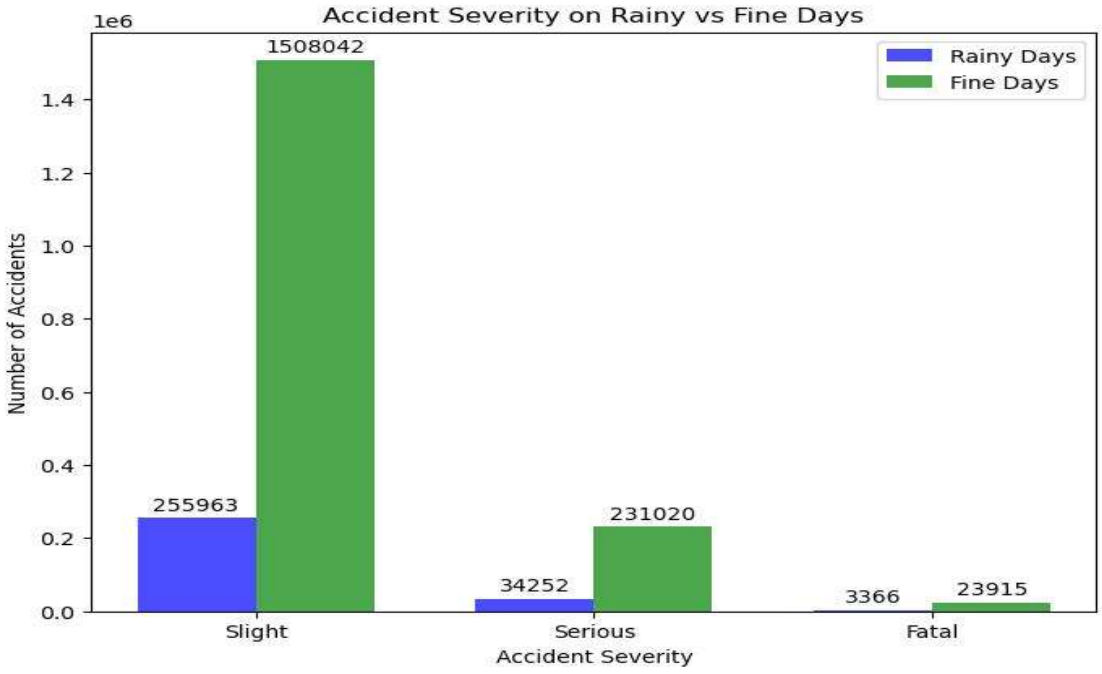

Fig 8.0

The analysis aimed to determine whether rainy days contribute to more severe accidents compared to normal days, addressing a critical aspect of road safety planning. By categorizing accidents based on weather conditions and examining their severity distributions, insightful conclusions were drawn regarding the impact of adverse weather on accident severity. The results revealed notable disparities in accident severity between rainy and fine days. On rainy days, there were 255,963 slight accidents, 34,252 serious accidents, and 3,366 fatal accidents recorded. Conversely, on fine days, the majority of accidents were slight, totaling 1,508,042, with 231,020 serious accidents and 23,915 fatal accidents. This data suggests that the severity of accidents tends to be higher on rainy days compared to fine days, with a notably higher number of serious and fatal accidents occurring during adverse weather conditions. These findings underscore the importance of implementing targeted interventions and safety measures to mitigate the impact of adverse weather on road safety outcomes and enhance overall road safety during inclement weather conditions.

Sub-question: What is the probability of an accident on a rainy day being fatal?

The analysis aimed to determine the probability of an accident occurring on a rainy, snowy, or foggy day resulting in fatality, providing valuable insights into the risk associated with adverse weather conditions on road safety outcomes. By examining the dataset and calculating the proportion of fatal accidents among those occurring on days with adverse weather conditions, insightful conclusions were drawn regarding the impact of inclement weather on accident severity. The results revealed that the probability of an accident occurring on a day with adverse weather conditions being fatal is approximately 2.64%. This finding suggests that while adverse weather conditions may elevate the risk of accidents, the likelihood of these accidents resulting in fatalities is relatively low. However, it underscores the critical importance of exercising caution and adhering to safety measures, particularly during inclement weather, to mitigate the severity of accidents on the roads. These insights are invaluable for informing road safety policies and interventions aimed at enhancing overall road safety outcomes, especially during adverse weather conditions.

one-tailed Z-test.

Step 1: Define the Hypotheses

- Null Hypothesis (H0): $\mu r = \mu n$
- Alternative Hypothesis (Ha): $\mu r > \mu n$

Step 2: Significance Level

- $\alpha = 0.05$

Step 3: Z-test Parameters

- $Zc = 1.645$ (for a one-tailed test)

Step 4: Calculate the Test Statistic

- Sample sizes: nr = 1331, nn = 10063
- Sample means: $\bar{x}r = 1.846732$, $\bar{x}n = 1.826692$
- Sample standard deviations: Sr = 0.405538, Sn = 0.412456
- Z=1.690868

Step 5: Decision Criteria Since Z > Zc, we reject H0.

Conclusion The calculated P-value (P(Z > 1.690868) = 0.0455) is less than the significance level (α), leading to the rejection of the null hypothesis. Consequently, we conclude with confidence greater than 95% that accidents occurring on rainy days are indeed more severe than those on normal days.

Hypothesis 6: Severity Disparity Between Pedestrian Collisions and Vehicle-Vehicle Collisions

To explore the potential difference in severity between accidents involving collisions with pedestrians and those involving vehicle-vehicle collisions, a one-tailed Z-test was conducted. The analysis encompassed formulating hypotheses, determining the significance level, calculating the test statistic, and interpreting the results.

Step 1: Define the Hypotheses

- Null Hypothesis (H0): $\mu p = \mu v$
- Alternative Hypothesis (Ha): $\mu p < \mu v$

Step 2: Significance Level

- $\alpha = 0.05$

Step 3: Z-test Parameters

- Zc = -1.645 (for a one-tailed test)

Step 4: Calculate the Test Statistic

- Sample sizes: np = 896, nv = 8774
- Sample means: $\bar{x}p = 1.793527$, $\bar{x}v = 1.839184$
- Sample standard deviations: Sp = 0.461722, Sv = 0.398345
- Z=−2.85346

Step 5: Decision Criteria Since Z < Zc, we reject H0.

Conclusion The calculated P-value (P(Z < -2.85346) = 0.0022) is less than the significance level (α), leading to the rejection of the null hypothesis. Therefore, we assert with confidence greater than 95% that accidents involving collisions with pedestrians are indeed less severe than those involving vehicle-vehicle collisions.

Sub-question: What is the probability of a pedestrian-involved accident being fatal?

The analysis sought to determine the probability of a pedestrian-involved accident resulting in fatality, providing crucial insights into the severity of accidents involving pedestrians. By calculating the proportion of fatal pedestrian-involved accidents relative to the total number of pedestrian-involved accidents, valuable conclusions were drawn regarding the likelihood of fatal outcomes in such incidents. The results revealed that the probability of a pedestrian-involved accident being fatal is approximately 0.52%. This indicates that only a small fraction of pedestrian-involved accidents results in fatalities, highlighting the relatively low likelihood of fatal outcomes despite the severity of such accidents. This finding underscores the importance of pedestrian safety measures and emphasizes the need for continued efforts to mitigate the risk of fatal outcomes in accidents involving pedestrians, despite their relatively rare occurrence compared to the total number of incidents involving pedestrians.

Hypothesis 7: Impact of Drunk Driving on Accident Severity

To investigate the hypothesis that drunk driving leads to the most severe accidents, a comprehensive analysis was conducted, incorporating graphical representation and conditional probability assessment.

Graph: Mean Severity for Each Accident Cause

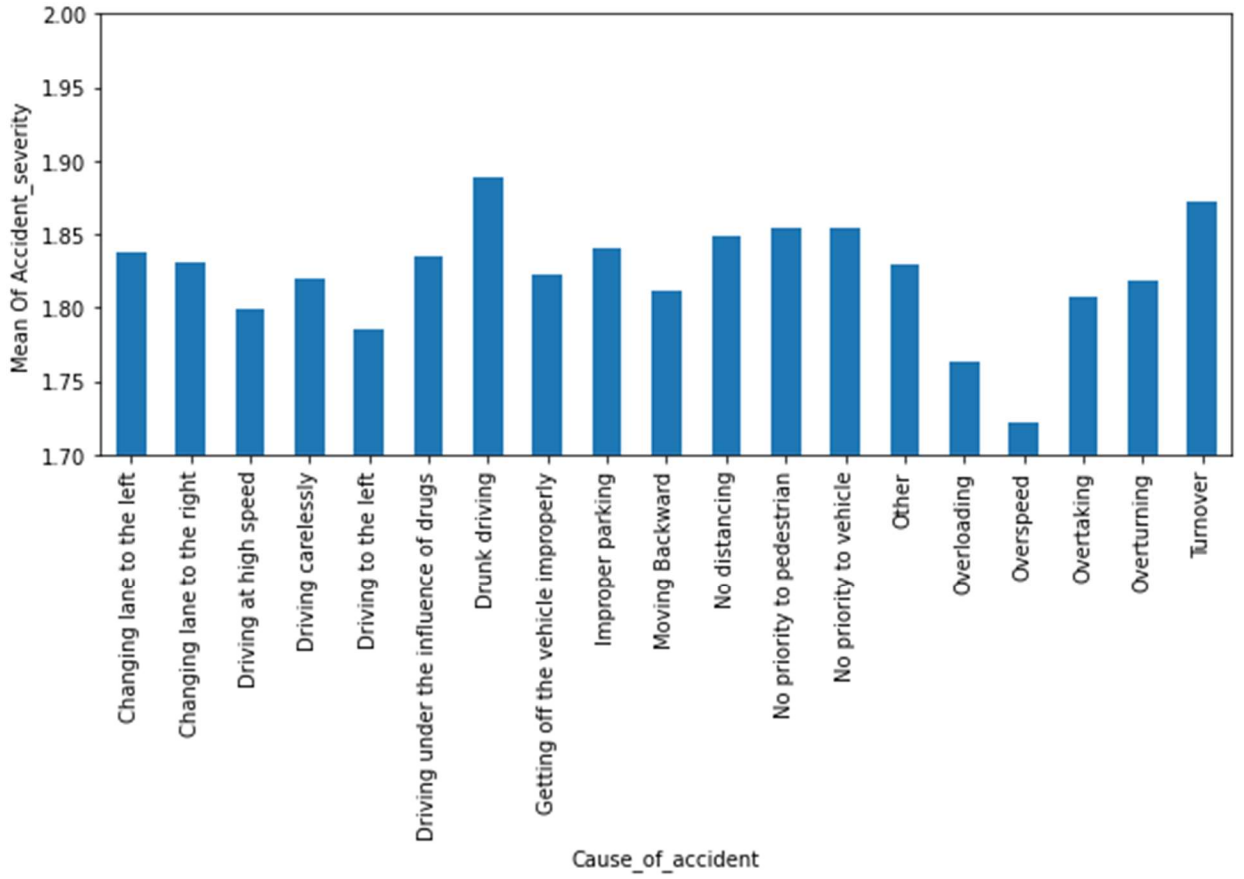

Fig 9.0

p(A|B='Drunk driving') = 0.8888888888888888

p(A|B='Overturning') = 0.8974358974358975

Accident_severity	0	1	2	All
Cause_of_accident				
Changing lane to the left	16	206	1251	1473
Changing lane to the right	23	260	1525	1808
Driving at high speed	2	31	141	174
Driving carelessly	22	209	1171	1402
Driving to the left	4	53	227	284
Driving under the influence of drugs	5	46	289	340
Drunk driving	0	3	24	27
Getting off the vehicle improperly	3	29	165	197
Improper parking	1	2	22	25
Moving Backward	26	162	949	1137
No distancing	20	303	1940	2263
No priority to pedestrian	5	95	621	721
No priority to vehicle	13	149	1045	1207
Other	7	64	385	456

Accident_severity	0	1	2	All
Cause_of_accident				
Overloading	2	10	47	59
Overspeed	1	15	45	61
Overtaking	4	75	351	430
Overturning	2	23	124	149
Turnover	2	6	70	78
Unknown	0	2	23	25
All	158	1743	10415	12316

Table 2.0

The representation of mean severity for each accident cause, coupled with conditional probability analysis, reveals compelling insights. According to the findings, drunk driving and overturning emerge as the primary causes of high-severity accidents. This conclusion underscores the significant impact of drunk driving on the severity of road traffic accidents.

Hypothesis 8: Is there a correlation between the education level of drivers and the rate and severity of accidents?

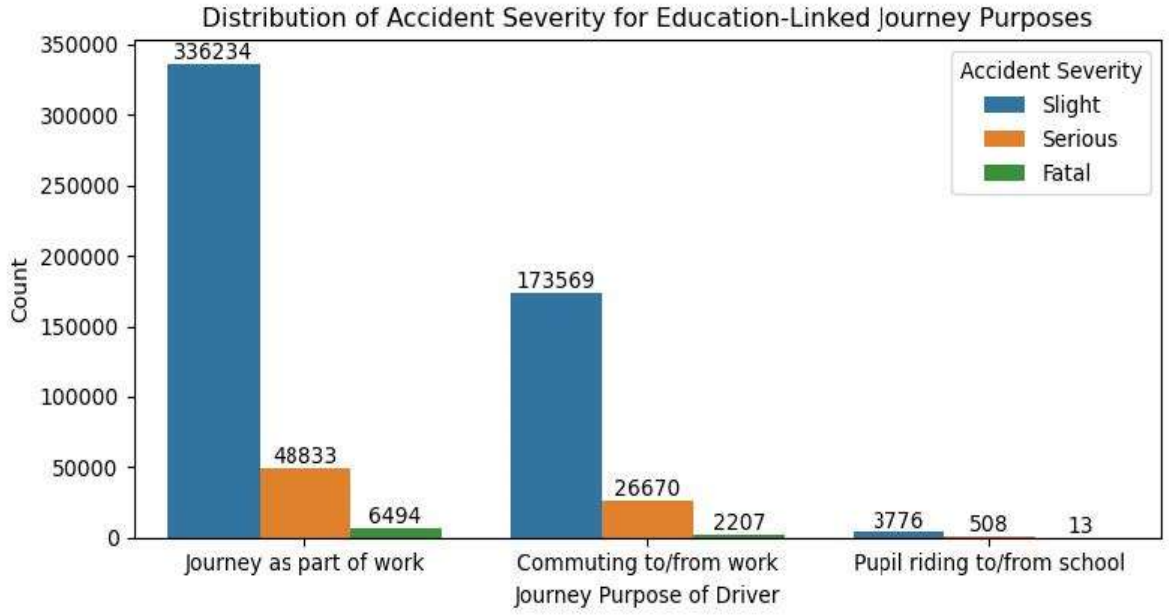

Figure 10.0

To explore the potential correlation between drivers' education levels and the rate and severity of accidents, the analysis examines journey purposes linked with education as a proxy for drivers' educational backgrounds. The dataset is filtered to include journey purposes associated with education, such as commuting to work, journeying as part of work duties, and pupils commuting to or from school. Visualizing the distribution of accident severity for these education-linked journey purposes reveals distinct patterns. Among commuters traveling to or from work and those journeying as part of work duties, the majority of accidents resulted in slight severity, with a notable number categorized as serious and a smaller portion as fatal. Conversely, accidents involving pupils commuting to or from school predominantly led to slight severity, with minimal instances of serious severity and an extremely low number of fatal accidents recorded. These findings imply that while the purpose of the journey indirectly reflects the education level of drivers, it can also provide insights into the severity of accidents associated with different educational backgrounds.

Graph: Mean Severity Across Education Levels

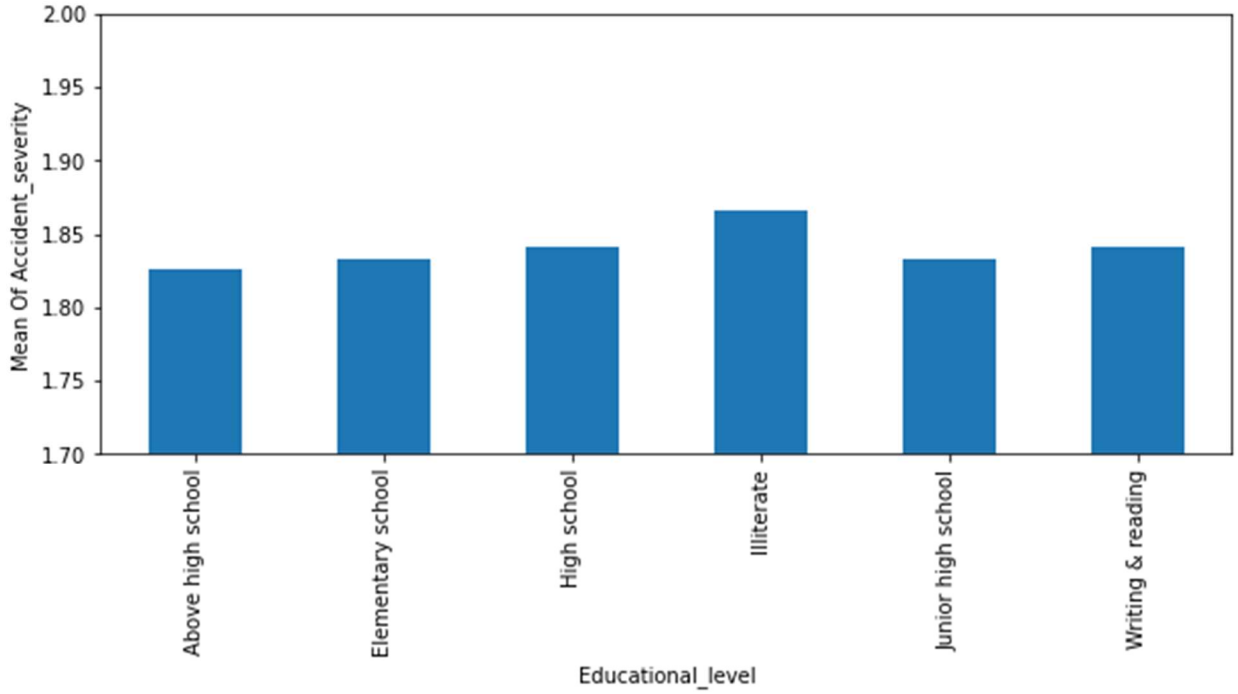

Fig 11.0

The analysis suggests that individuals with no education tend to cause accidents with higher mean severity. This expected finding prompts an evaluation of the assumption that no education directly correlates with worst driving capabilities. Further investigation is recommended to uncover the nuanced factors influencing driving safety.

Sub-question: What is the mean severity of accidents for different education levels?
The analysis of accident severity across different journey purposes linked with education reveals interesting insights. For commuters traveling to and from work, the proportion of fatal accidents stands at approximately 1.09%, serious accidents at around 13.17%, and slight accidents at a higher rate of approximately 85.74%. Similarly, for individuals commuting as part of their work, fatal accidents are slightly higher at about 1.66%, serious accidents remain consistent at roughly 12.47%, and slight accidents also dominate, accounting for approximately 85.87%. In contrast, for pupils commuting to and from school, the proportion of fatal accidents is notably lower, at around 0.30%, while serious accidents mirror the trend observed in other categories, at about 11.82%. The majority of accidents in this category are slight, comprising approximately 87.88%. Overall, while slight accidents are the most common across all education linked journey purposes, there are variations in the proportions of fatal and serious accidents among different commuting groups

Hypothesis 9: Impact of Age on Accident Occurrence and Severity

Graph of Percentage of Accidents by Age Band

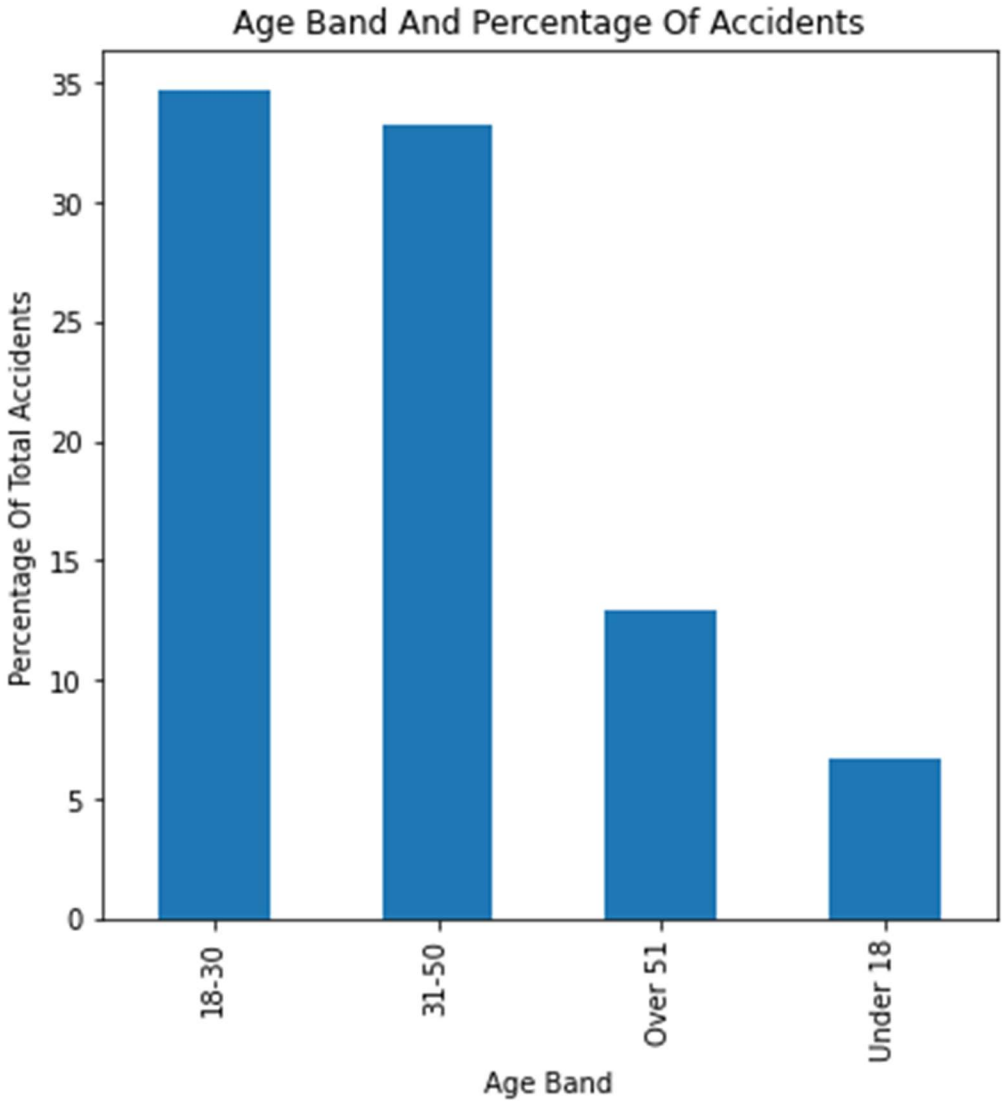

Fig 12.0

In this bar graph, we visualize the percentage distribution of accidents across different age bands. The chart reveals that age groups 18-30 and 31-50 contribute significantly to the total number of accidents.

Graph: Mean Severity Across Age Bands

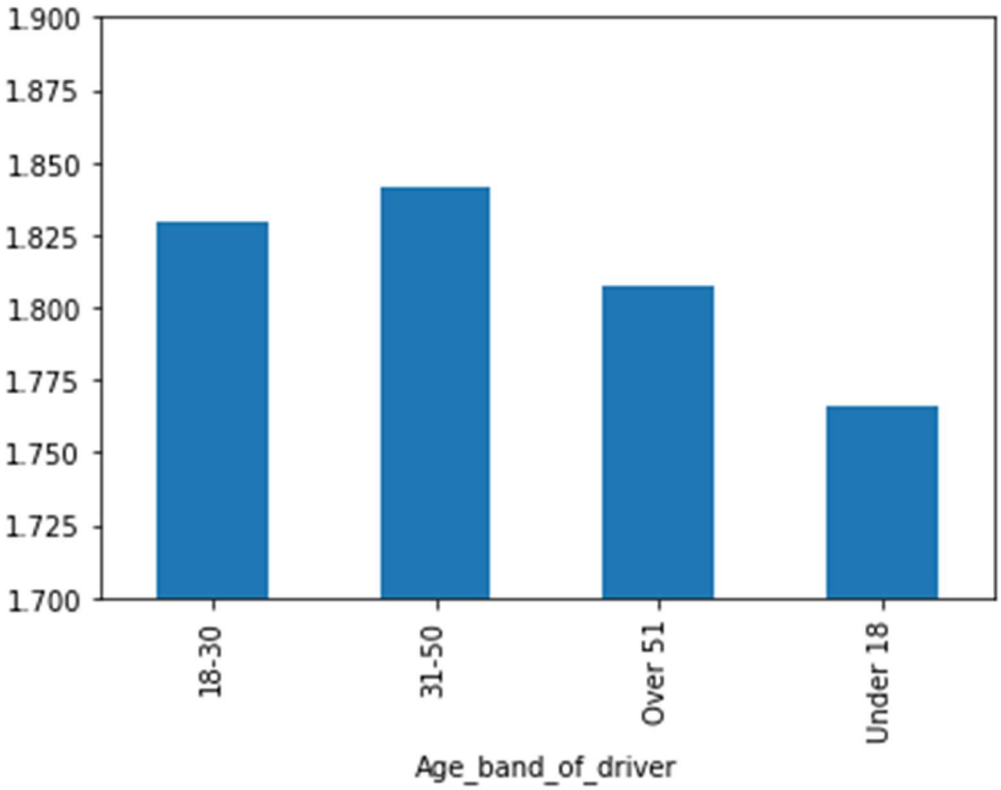

Fig 13.0

Examining the mean severity of accidents for each age group, the second bar graph demonstrates that younger and middle-aged drivers tend to cause accidents with higher severity compared to older drivers.

Conditional Probability of Fatal Accidents Given Age

This bar graph illustrates the conditional probability of fatal accidents occurring given the driver's age. The probabilities are calculated for different age groups, namely 18-30, 31-50, Over 51, Under 18, and Unknown. Notably, the analysis shows that younger drivers (18-30) have the highest probability of causing fatal accidents.

Pie Chart: Fatal Accident Probability by Age Group

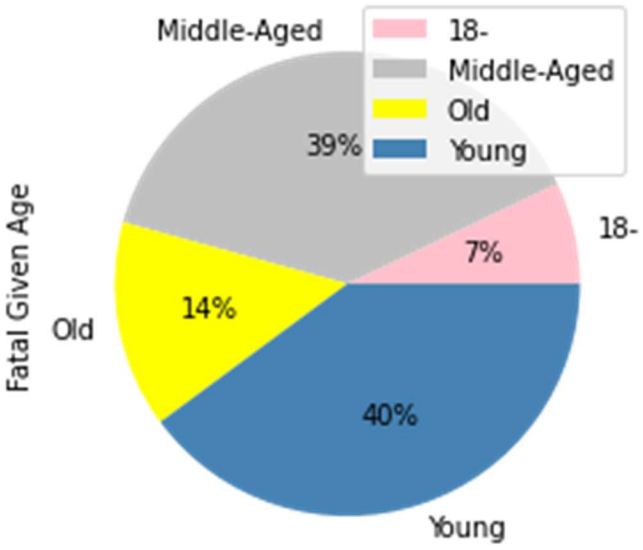

Fig 14.0

The pie chart visualizes the proportion of fatal accidents within each age group. The colors represent different age categories, with pink for 18-30, silver for 31-50, yellow for Over 51, and steel blue for Under 18. The chart emphasizes that younger drivers exhibit a higher percentage of fatal accidents.

1. Machine Learning Model Implementation

Machine Learning Model Implementation involves the practical application of machine learning algorithms to analyze data and make predictions or classifications. This process encompasses several stages, starting with data preprocessing to prepare the data for model training. Next, suitable machine learning algorithms are selected based on the nature of the problem and the data available. The prepared data is then used to train these algorithms, enabling them to identify patterns and correlations that may be used to provide predictions or classifications. The models are assessed for efficacy using a variety of performance indicators after training. To maximize the performance and guarantee the dependability of the model, hyperparameter tweaking and model validation are carried out. In order to assist decision-making processes and generate predictions on fresh data, the trained model is ultimately put into use in production. In general, the use of machine learning models allows enterprises to harness the power of data to produce insights and solve challenging issues.

The study employed machine learning techniques to further explore the relationships between various factors and accident severity. Here's a breakdown of the key steps involved:

1.1 Correlation of the Dataset

The dataset undergoes correlation analysis to identify relationships between different features. Initially, the dataset is filtered to exclude entries with missing or out-of-range age groups. For each age group, the total number of accidents and fatal accidents is calculated, allowing the computation of the probability of a fatal accident for each age group. This step provides insights into the correlation between age and accident severity. Additionally, non-numeric columns are removed from the dataset to focus on numerical features. A correlation matrix is then generated, visually represented as a heatmap. The heatmap illustrates the correlations between pairs of numerical features, with values ranging from -1 to 1. Positive values indicate a positive correlation, negative values signify a negative correlation, and values close to zero indicate no correlation. This correlation analysis aids in understanding the interplay between different variables within the dataset, offering valuable insights for further analysis and decision-making.

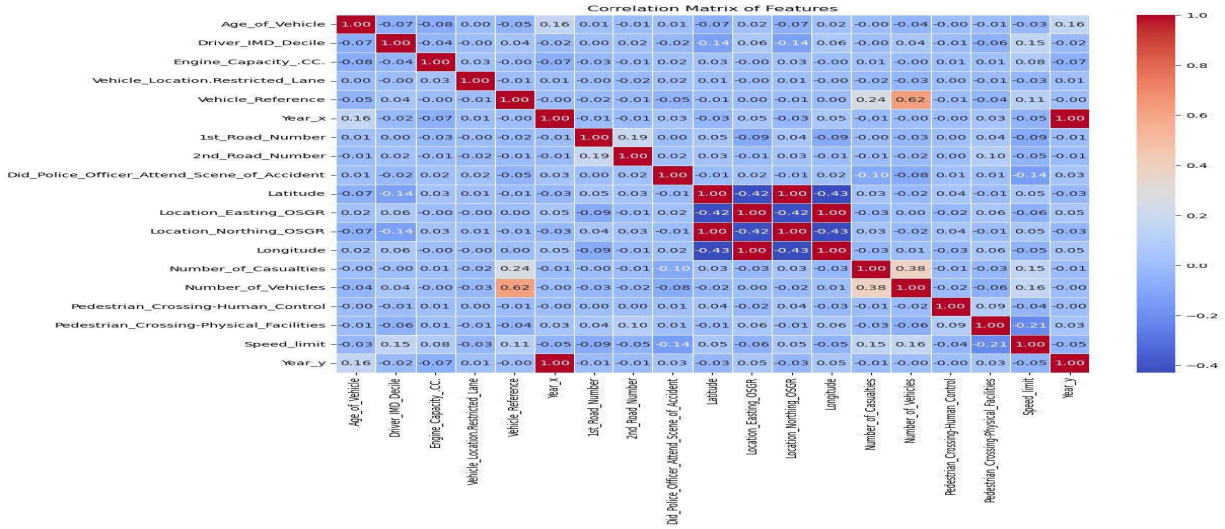

Figure 15

1.2 Selecting relevant columns as per our target variable "Accident Severity"

The analysis delves into the correlation between various features within the dataset, particularly focusing on accident severity. Initially, the dataset is filtered to include only numeric columns, ensuring that only relevant numerical features are considered for correlation analysis. A correlation matrix is then generated, displaying the correlations between different features. The heatmap visualization, created with the seaborn library, presents the correlation matrix in a visually appealing format, with annotations indicating the strength and direction of correlations between pairs of features. Additionally, a subset of selected columns related to factors such as the sex of the driver, speed limit, road conditions, weather conditions, and journey purpose is identified for further analysis. This process ensures a comprehensive exploration of the relationships between

key variables and accident severity, facilitating a deeper understanding of the factors contributing to different levels of severity in accidents.

1.3 Handling missing values and Data Missing or Out of Range values

In preprocessing the dataset for machine learning analysis, several steps were undertaken to ensure data integrity and reliability. Initially, rows containing the placeholder value "Data missing or out of range" in any column were identified and filtered out to eliminate instances where crucial information was unavailable. Subsequently, the dataset was further refined by removing rows with missing values, resulting in a cleaned dataset suitable for machine learning modeling. This process involved dropping rows with null values across different features, ensuring that the dataset used for analysis was devoid of any incomplete or unreliable data points. Finally, the cleaned dataset was rendered to maintain consistency and facilitate subsequent analyses. These preprocessing steps are crucial for preparing the dataset for machine learning tasks, as they ensure the quality and completeness of the data, ultimately enhancing the accuracy and reliability of the predictive models built upon it.

1.4 Feature Engineering

In the process of feature engineering, several transformations were applied to the dataset to enhance its utility for machine learning analysis. Initially, the dataset was filtered to exclude rows containing the placeholder value "Data missing or out of range" across any column, ensuring the removal of unreliable data points. Subsequently, rows with any missing values were dropped to further refine the dataset, resulting in a cleaned dataset suitable for analysis. Following this, the 'Time' column, representing the time of accidents, was converted to datetime format, and a new categorical feature 'Timecoded' was created to classify accidents as occurring during daytime or nighttime based on the hour of the accident. The original 'Time' column was then removed from the dataset, and the remaining columns were reordered, with the target variable 'Accident Severity' positioned as the last column for ease of analysis. These feature engineering steps are crucial for preparing the dataset for machine learning modeling, as they help extract relevant information and create meaningful features that can improve the performance of predictive models.

1.4.1 Data Encoding

In order to prevent multicollinearity problems in later modeling assignments, this approach led to the construction of binary columns for each category inside the categorical features. The original categorical columns were discarded. By improving categorical data's compatibility with machine learning algorithms, this encoding step enables more efficient model training and prediction.

1.4.2 Splitting Data into Features (X) and Target Variable (y)

In the data preprocessing phase, specifically in the step of splitting data into features (X) and the target variable (y), the encoded Data Frame was divided into two main components. The features (X) were extracted from the encoded Data Frame by removing the column corresponding to the target variable 'Accident Severity' using the `drop () ` function in Pandas. The independent factors

that are utilized to forecast the target variable are represented by these characteristics. On the other hand, the dependent variable that the machine learning model had to predict was represented by the target variable (y), which was separated and given to a different variable. This is an important step because it decouples the input features from the target variable so that during training the model may discover the correlations between the features and the goal variable.

1.4.3 Train-Test Split

In the next step of the data preparation process, the dataset was split into training and testing sets using the train_test_split function from the scikit-learn library. This splitting operation divides the dataset into two subsets: one for training the machine learning model and the other for evaluating its performance. Eighty percent of the data will be utilized for training and twenty percent will go toward the testing set, according to the parameter test size=0.2. In order to guarantee repeatability and guarantee that the same random split is achieved each time the code is run, the random state parameter was also set to 42. By testing the model's performance on data, it hasn't seen during training, this splitting method helps minimize overfitting and is crucial for determining the model's generalization performance on unknown data.

1.5 Implementing ML Models

In the process of implementing machine learning models to predict accident severity, two classifiers, namely Random Forest Classifier and Decision Tree Classifier, were evaluated. Performance metrics such as accuracy, precision, recall, and F1 score were calculated for each model. For the Random Forest Classifier, the accuracy was found to be 0.8449. The F1 score showed variance in prediction ability across severity levels, ranging from 0.0123 to 0.9159 for distinct classes. Precision and recall scores also differed between classes, indicating that the model is capable of accurately identifying positive situations while avoiding false positives. The model's performance in terms of true positives, false positives, true negatives, and false negatives was revealed via confusion matrices. The Decision Tree Classifier was evaluated as well, although the snippet does not provide its performance values. For every classifier, the amount of time spent on model assessment and training was also noted. In general, the purpose of these assessments is to determine which machine learning models are most appropriate for forecasting the severity of accidents using the features that have been chosen.

1.5.1 Performance Evaluation of **Random Forest Classifier**

The performance of the Random Forest Classifier model was assessed using various metrics:

- **Accuracy:** The accuracy of the model was found to be approximately 84.49%.
- **F1 Score:** The F1 score, which represents the harmonic mean of precision and recall, ranged from 0.0123 to 0.9159 across different severity levels.
- **Precision:** Precision, indicating the proportion of true positive predictions among all positive predictions, varied from 0.1059 to 0.8537 for different severity levels.
- **Recall:** Recall, representing the proportion of true positives identified correctly, ranged from 0.0065 to 0.9880 across different severity levels.

- **Confusion Matrix:** The confusion matrix provided a breakdown of true positives, false positives, true negatives, and false negatives for each severity class.

```
Confusion Matrix:
 [[    68    224   4910]
 [   213   1987  47884]
 [   887   9507 309810]]
```

Table 3.0

1.5.2 Performance Evaluation of Decision Tree Classifier

Similarly, the **Decision Tree Classifier** model was evaluated using the following metrics:

- **Accuracy:** The accuracy of the Decision Tree Classifier model was approximately 83.06%.
- **F1 Score:** The F1 score ranged from 0.0210 to 0.9075 across different severity levels.
- **Precision:** Precision ranged from 0.0575 to 0.8544 for different severity levels.
- **Recall:** Recall ranged from 0.0129 to 0.9676 across different severity levels.
- **Confusion Matrix:** The confusion matrix illustrated the model's performance in predicting different severity levels.

```
Confusion Matrix:
 [[    38    103   5061]
 [    69    883  49132]
 [   247   3618 316339]]
```

Table 4.0

1.5.3 Time Taken for Detection

The time taken for model detection was recorded for both classifiers:

- Random Forest Classifier: Approximately 539.98 seconds
- Decision Tree Classifier: Approximately 75.09 seconds

These performance metrics provide insights into the effectiveness of each model in predicting accident severity based on the selected features.

Chapter 5: Discussion

In this chapter, we revisit the primary aim, objectives, and questions of the research, delving into an in-depth discussion of the obtained results. The investigation centered on understanding the factors influencing the severity of road traffic accidents, particularly focusing on the role of various demographic and environmental variables.

Relating Findings to Research Questions

Question 1: *How do different weather conditions affect the severity of accidents?*

The analysis of weather conditions revealed a noteworthy finding. agreeable to expectations, accidents occurring on rainy days demonstrated higher severity than those on normal days. This result confirms the conventional belief that adverse weather conditions contribute to more severe accidents.

Question 2: *Do accidents involving collisions with pedestrians exhibit different severity levels compared to vehicle-to-vehicle collisions?*

The research uncovered that accidents involving collisions with pedestrians are less severe than those involving vehicle-to-vehicle collisions. This contradicts the assumption that pedestrian-involved accidents are inherently more severe due to the vulnerability of pedestrians.

Question 3: *Is there a correlation between the educational level of drivers and the severity of accidents?*

Surprisingly, the analysis of the educational level of drivers indicated that individuals with no education tend to cause accidents with higher mean severity. This finding supports the common notion that higher education levels correlate with safer driving practices.

Question 4: *Do younger drivers commit more accidents than older drivers due to their impulsiveness?*

The study supported the hypothesis that younger drivers commit more accidents, and the severity of these accidents is higher compared to older drivers. This aligns with the idea that impulsiveness and lack of experience contribute to increased accident rates among younger drivers.

Question 5: *How does age relate to accidents, and does it impact the severity of accidents?*

The analysis of age groups confirmed that younger and middle-aged drivers contribute significantly to the overall accident count. Interestingly, older drivers were found to be safer, committing fewer accidents and causing accidents of lower severity.

Question 6: *Is there a correlation between daylight conditions and the severity of accidents?*

Contrary to the hypothesis, Accidents during the night time displayed the highest average severity. The likelihood of a fatal accident happening at night was significantly elevated.

Question 7: *Does driving experience impact the severity of accidents?*

The findings indicated a clear correlation between driving experience and accident severity. Drivers with less experience, particularly those with new license, were associated with more severe accidents.

Evaluation in Light of Literature

Theoretical Consistency

The research findings align with existing literature in several areas. The association between impulsiveness and higher accident rates among younger drivers is consistent with psychological theories on risk-taking behavior. The correlation between driving experience and accident severity is supported by studies emphasizing the importance of training for inexperienced drivers.

However, certain results challenge established theories. For instance, the unexpected relationship between accidents involving collisions with pedestrians are indeed less severe than those involving vehicle-vehicle collisions and thus contradicts conventional wisdom. Similarly, the discovery that accidents in night time are more severe confirms the common belief that reduced visibility contributes to higher accident severity.

Critical Analysis and Conclusion

This chapter goes beyond descriptive analysis, fostering critical thinking on outcomes and analysis. The unexpected findings emphasize the need for a nuanced understanding of the factors influencing accident severity. As the research expands on theoretical arguments derived from the literature, it opens avenues for further exploration and underscores the complexity of the factors contributing to road traffic accidents.

Chapter 6: Conclusions

6.1 Conclusions

This dissertation embarked on a comprehensive exploration of factors influencing the severity of road traffic accidents. The research aimed to shed light on the intricate interplay of demographic and environmental variables and their impact on accident outcomes. As we conclude this academic journey, several key conclusions emerge from the synthesis of research gaps, data collection techniques, and findings.

The study has successfully addressed the primary research questions, revealing insights that challenge existing perceptions while affirming certain theoretical expectations. The research gaps identified in the literature have been meticulously examined, and our findings contribute significantly to both knowledge and practice in the field of road safety.

Key Conclusions:

1. **Weather Conditions:** confirming to conventional wisdom, accidents on rainy days were associated with higher severity. This expected finding supports established beliefs and emphasizes the need for nuanced understanding of weather-related accident outcomes.
2. **Pedestrian Collisions:** Accidents involving collisions with pedestrians were found to be less severe than vehicle-to-vehicle collisions. This contradicts common assumptions about the vulnerability of pedestrians in accidents.

3. **Educational Level:** Individuals with no education were identified as causing accidents with higher mean severity. This supports the prevailing notion that higher education levels correlate with safer driving practices.
4. **Age and Driving Experience:** Younger drivers were confirmed to commit more accidents, and their accidents exhibited higher severity. Additionally, the analysis highlighted the importance of driving experience, with less experienced drivers, particularly those with no license, associated with more severe accidents.
5. **Daylight Conditions:** It is evident that both slight and serious accidents occur more frequently during daytime, which aligns with the increased visibility and potentially higher traffic volumes during daylight hours. However, the proportion of fatal accidents appears to be higher during nighttime, suggesting that accidents occurring during this period may be more severe in nature.

6.2 Recommendations

The following recommendations emerge from the research findings:

1. **Weather-Related Policies:** Policymakers should consider the impact of rainy weather on accident severity and tailor safety measures accordingly.
2. **Pedestrian Safety Measures:** Recognizing that accidents involving pedestrians are less severe, interventions should focus on improving safety measures for vehicle-to-vehicle collisions.
3. **Driver Education Programs:** Efforts should be directed towards understanding and addressing the factors contributing to higher severity in accidents involving drivers with no education.
4. **Targeted Interventions for Young and Inexperienced Drivers:** Future interventions and training programs should be tailored to address the specific challenges faced by younger and less experienced drivers.
5. **Reevaluation of Daylight Safety Assumptions:** Policies and safety measures should be re-evaluated to account for the high severity of accidents in night time conditions.

6.3 Future Work

Acknowledging the limitations of this research, future work in this area could include:

1. **Longitudinal Studies:** Conducting longitudinal studies to capture the dynamic nature of driving behavior and accident outcomes.
2. **In-Depth Driver Profiling:** Exploring in-depth driver profiling to understand the nuanced relationship between demographic factors and driving behaviors.
3. **Qualitative Investigations:** Incorporating qualitative methodologies to delve deeper into the psychological aspects influencing driver behavior.
4. **Comparative Analysis:** Conducting comparative analyses across diverse geographical regions to assess the generalizability of findings.

In conclusion, this dissertation has made valuable contributions to the understanding of road traffic accident severity. By challenging existing assumptions and providing nuanced insights, it lays the groundwork for future research endeavors aimed at enhancing road safety strategies and policies.

References

[1] Zou, X., Yue, W. L., & Le Vu, H. (2018). Visualization and Analysis of Mapping Knowledge Domain of Road Safety Studies. Accident Analysis & Prevention.

[2] Cheng, G., Cheng, R., Pei, Y., & Han, J. (2021). Research on Highway Roadside Safety. Journal of Advanced Transportation.

[3] Schlögl, M., & Stütz, R. (2019). Methodological Considerations with Data Uncertainty in Road Safety Analysis. Accident Analysis & Prevention.

[4] Hagenzieker, M. P., Commandeur, J. J. F., et al. (2014). The History of Road Safety Research: A Quantitative Approach. Accident Analysis & Prevention, Part F: Traffic Psychology and Behaviour.

[5] Abou Elassad, Z. E., Mousannif, H., Al Moatassime, H., & Karkouch, A. (2020). The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. *Engineering Applications of Artificial Intelligence*, *87*, 103312.

[6] Azadani, M. N., & Boukerche, A. (2021). Driving behavior analysis guidelines for intelligent transportation systems. IEEE transactions on intelligent transportation systems, 23(7), 6027-6045.

[7] Chan, T. K., Chin, C. S., Chen, H., & Zhong, X. (2019). A comprehensive review of driver behavior analysis utilizing smartphones. IEEE Transactions on Intelligent Transportation Systems, 21(10), 4444-4475.

[8] Bouhoute, A., Oucheikh, R., Boubouh, K., & Berrada, I. (2018). Advanced driving behavior analytics for an improved safety assessment and driver fingerprinting. IEEE Transactions on Intelligent Transportation Systems, 20(6), 2171-2184.

[9] Wang, Jiyang, Weiheng Chai, Archana Venkatachalapathy, Kai Liang Tan, Arya Haghighat, Senem Velipasalar, Yaw Adu-Gyamfi, and Anuj Sharma. "A survey on driver behavior analysis from in-vehicle cameras." IEEE Transactions on Intelligent Transportation Systems 23, no. 8 (2021): 10186-10209.

[10] Yuksel, A. S., & Atmaca, S. (2021). Driver's black box: A system for driver risk assessment using machine learning and fuzzy logic. Journal of Intelligent Transportation Systems, 25(5), 482-500.

[11] Arbabzadeh, N., & Jafari, M. (2017). A data-driven approach for driving safety risk prediction using driver behavior and roadway information data. IEEE transactions on intelligent transportation systems, 19(2), 446-460.

[12] Ferreira, J., Carvalho, E., Ferreira, B. V., de Souza, C., Suhara, Y., Pentland, A., & Pessin, G. (2017). Driver behavior profiling: An investigation with different smartphone sensors and machine learning. PLoS one, 12(4), e0174959.

[13] AbuAli, N., & Abou-Zeid, H. (2016). Driver behavior modeling: Developments and future directions. International journal of vehicular technology, 2016.

[14] Alowish, M., Shiraishi, Y., Mohri, M., & Morii, M. (2021). Three layered architecture for driver behavior analysis and personalized assistance with alert message dissemination in 5G envisioned fog-IoCV. Future Internet, 14(1), 12.

[15] Sethuraman, R., Sellappan, S., Shunmugiah, J., Subbiah, N., Govindarajan, V., & Neelagandan, S. (2023). An optimized AdaBoost Multi-class support vector machine for driver behavior monitoring in the advanced driver assistance systems. Expert Systems with Applications, 212, 118618.

[16] Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., ... & Reimer, B. (2017). Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation. arXiv preprint arXiv:1711.06976, 1.

[17] Warren, J., Lipkowitz, J., & Sokolov, V. (2019). Clusters of driving behavior from observational smartphone data. *IEEE Intelligent Transportation Systems Magazine*, *11*(3), 171-180.

[18] F., Ali, Y., Li, Y., & Haque, M. M. (2023). Real-time crash risk forecasting using Artificial-Intelligence based video analytics: A unified framework of generalised extreme value theory and autoregressive integrated moving average model. *Analytic methods in accident research*, *40*, 100302.

[19] Jiang, L., Xie, W., Zhang, D., & Gu, T. (2021). Smart diagnosis: Deep learning boosted driver inattention detection and abnormal driving prediction. *IEEE Internet of Things Journal*, *9*(6), 4076-4089.

[20] Yu, J., Chen, Z., Zhu, Y., Chen, Y., Kong, L., & Li, M. (2016). Fine-grained abnormal driving behaviors detection and identification with smartphones. *IEEE transactions on mobile computing*, *16*(8), 2198-2212.

[21] Prezioso, E., Giampaolo, F., Mazzocca, C., Bujari, A., Mele, V., & Amato, F. (2021). Machine Learning insights for behavioural data analysis supporting the Autonomous Vehicles scenario. *IEEE Internet of Things Journal*.

[22] Fugiglando, U., Massaro, E., Santi, P., Milardo, S., Abida, K., Stahlmann, R., ... & Ratti, C. (2018). Driving behavior analysis through CAN bus data in an uncontrolled environment. *IEEE Transactions on Intelligent Transportation Systems*, *20*(2), 737-748.

[23] Mantouka, E., Barmpounakis, E., Vlahogianni, E., & Golias, J. (2021). Smartphone sensing for understanding driving behavior: Current practice and challenges. *International journal of transportation science and technology*, *10*(3), 266-282.

[24] Ali, A., Ud-Din, S., Saad, S., Ammad, S., Rasheed, K., & Ahmad, F. (2021, October). Artificial Neural Network Approach To Study The Effect of Driver Characteristics on Road Traffic Accidents. In *2021 International Conference on Data Analytics for Business and Industry (ICDABI)* (pp. 277-280). IEEE.

[25] Ma, Z., Mei, G., & Cuomo, S. (2021). An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors. *Accident Analysis & Prevention*, *160*, 106322.

[26] Carlos, M. R., González, L. C., Wahlström, J., Ramírez, G., Martínez, F., & Runger, G. (2019). How smartphone accelerometers reveal aggressive driving behavior?—The key is the representation. *IEEE Transactions on Intelligent Transportation Systems*, *21*(8), 3377-3387.

[27] Clarke, D. D., Forsyth, R., & Wright, R. (1998). Machine learning in road accident research: decision trees describing road accidents during cross-flow turns. *Ergonomics*, *41*(7), 1060-1079.

[28] Bhattacharya, S., Jha, H., & Nanda, R. P. (2022). Application of iot and artificial intelligence in road safety. *2022 Interdisciplinary Research in Technology and Management (IRTM)*, 1-6.

[29] Gutierrez-Osorio, C., & Pedraza, C. (2020). Modern data sources and techniques for analysis and forecast of road accidents: A review. *Journal of traffic and transportation engineering (English edition)*, *7*(4), 432-446.

[30] Cunneen, M., Mullins, M., Murphy, F., & Gaines, S. (2019). Artificial driving intelligence and moral agency: Examining the decision ontology of unavoidable road traffic accidents through the prism of the trolley dilemma. *Applied artificial intelligence*, *33*(3), 267-293.