Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

A machine learning approach for predicting road accidents

Artur BUDZYŃSKI
abudzyns@uek.krakow.pl (Corresponding author)
https://orcid.org/0000-0002-5803-6749
Cracow University of Economics, Cracov, Poland

Mikołaj FEDEROWICZ
mf301741@student.polsl.pl
Silesian University of Technology, Gliwice, Poland

Adam JABŁOŃSKI
aj303403@student.polsl.pl
Silesian University of Technology, Gliwice, Poland

Wesam HASAN
wesahas021@student.polsl.pl
Silesian University of Technology, Gliwice, Poland

Jan GORSZANÓW
jangors584@student.polsl.pl
Silesian University of Technology, Gliwice, Poland

Taras NIKITISHYN
taras.nikitishyn@alogliwice.polsl.pl
Academic Secondary School of the Silesian University of Technology, Gliwice, Poland

Received: 19 July 2024 | Revised: 27 December 2024
Accepted: 27 December 2024 | Available online: 07 February 2025

Abstract

Road accidents pose significant challenges to public safety and necessitate proactive measures to mitigate them. This paper introduces a machine-learning approach for predicting road accident incidences, leveraging diverse datasets encompassing traffic patterns, weather conditions, and historical accident records. The proposed model integrates feature engineering techniques to capture the multifaceted nature of variables influencing accidents. Through the application of advanced machine learning algorithms, such as ensemble methods and neural networks, the model aims to discern complex patterns within the data, facilitating accurate predictions of accident likelihood. The study also explores the interpretability of the model outputs, providing insights into the key predictors and their interactions. Validation and performance assessment involve rigorous testing on diverse datasets to ensure the generalizability and robustness of the predictive model. The outcomes of this research hold promise for the development of proactive road safety strategies and the implementation of targeted interventions, ultimately contributing to reducing road accidents and their associated societal impacts.

Keywords: Forecasting, Machine Learning, Road Accidents, Safety

1. Introduction

a. Background

Road accidents represent a critical and pervasive public health concern with far-reaching societal implications. The complex interplay of factors contributing to road accidents, encompassing human behavior, vehicle dynamics, and environmental conditions, underscores the multidisciplinary nature of the issue. As a global phenomenon, road accidents result in substantial loss of life, injury, and economic burden, necessitating comprehensive research endeavors aimed at understanding the underlying mechanisms and formulating effective preventive strategies. The intricate network of variables involved in road safety spans driver psychology, traffic engineering, vehicle technology, and governmental policies, demanding a holistic approach to mitigating the frequency and severity of accidents.

b. Objectives

This scientific exploration seeks to unravel the intricacies of road accidents, examining the immediate causes and the systemic and contextual factors that influence their occurrence. Through empirical investigation and data-driven analyses, this research aims to

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

contribute to the development of evidence-based interventions and policies that promote safer road environments and ultimately reduce the societal toll of road accidents. Road accidents are undesirable. The problem affects the whole world. Accidents cause death and injury to many people. Road accidents are the result of a combination of factors. There are many such factors. The causes of accidents can be divided into direct and indirect. There are also human, technical, and external factors. The human factor causes mistakes. Indirect factors influence human errors. Fatigue is an indirect cause of human errors. The absence of fatigue can be a factor that can prevent an accident caused by an external or technical factor.

2. Literature review

a. Previous Studies on Road Accident Prediction

Smith (2021) conducted research on residents of South Wales. The study was conducted using questionnaires. The relationship between involvement in an accident and caffeine consumption was checked. The accident rate among those consuming caffeine was 1.9%, and among those not consuming caffeine was 3.3%. People who are younger, in poor health, stressed and take greater risks are more likely to be involved in road accidents.

Abeer Hassan et al. (2016) analyzed accidents in Sudan. The data came from questionnaires completed by the accident victim and an eyewitness. Police reports were also used. The accident had a geographical impact. Most occurred in the capital of the country – Khartoum – 46.7%. The influencing factor was the location of the event. Accidents at level crossings accounted for 46.8% of the occurrences. Driver errors caused the accidents. One of the mistakes is excessive speed. This is a conscious error that sets it apart from the rest, accounting for 14.6% of accidents.

Másilková (2017) conducted a textual analysis of documents to analyze the health and social consequences of road traffic accidents, focusing on the direct and indirect impacts on individuals and society.

Xing et al. (2023) established a road traffic accident prediction model using the BP neural network, selecting 14 main influencing factors as input variables and traffic accident deaths as the output variable, trained the model with data from 2000 to 2017, and verified its accuracy with a relative error within 1% for predicting 2018 traffic accident deaths in China.

Wu & Wang (2020) used principal component analysis to reduce the dimensionality of China's road traffic accident data and compared the prediction accuracy of Support Vector Regression (SVR) and BP neural networks, finding that SVR provided superior estimation accuracy for predicting road traffic accidents.

Mehta et al. (2022) developed a road accident prediction system using the Xgboost machine learning model, incorporating variables like day, weather, and road type, and integrated it into a web application, achieving over 80% accuracy in forecasting accident severity.

b. Machine Learning in Road Safety

Kodieswari et al. (2024) introduced a critical examination of the challenges arising from the increased usage of vehicles, such as traffic congestion, travel delays, road accidents, and heightened energy consumption, over the past two decades. The proposed solution involves developing an intelligent transport system utilizing robust predictive models. The focus is on leveraging neural network and artificial intelligence (AI) models to design an intelligent transportation system. Artificial Neural Networks (ANNs) are employed to extract patterns crucial for the development of transportation infrastructure systems (TIS). Graph Neural Networks (GNN) play a pivotal role in estimating arrival times by analyzing traffic congestion, offering a proactive solution to overcome traffic jams. The use of Deep Convolution Neural Network (CNN) is highlighted for detecting driver behavior, with an emphasis on features such as acceleration, gravity, throttle, and revs per minute (RPM) to reduce the risks of accidents. The integration of these AI and neural network models is posited as a transformative approach, not only enhancing safety but also contributing to the development of a civic, cost-effective, and reliable transportation system.

Tayebikhorami & Sacchi (2022) compared the prediction performance of the support vector machine (SVM), decision tree (DT), and random forest (RF) machine learning algorithms with traditional safety performance functions (SPFs) for road safety predictive analytics, using data from urban intersections in Saskatchewan, and validated the ML algorithms' effectiveness in network screening for identifying collision-prone locations.

Jesi et al. (2023) developed an intelligent system using a combination of accelerometer sensors, vibration motion sensors, gyroscopic sensors, GPS technology, and machine learning techniques to detect road anomalies and provide driving support, integrating this system into vehicles to identify and validate road irregularities and store data in a cloud-based repository, ultimately enhancing road safety.

Subhan et al. (2023) developed an intelligent system to improve road safety by detecting road anomalies using a combination of accelerometer sensors, vibration motion sensors, gyroscopic sensors, GPS technology, and machine learning techniques, and integrated this system into vehicles to identify and validate road irregularities, ultimately enhancing road safety.

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

3. Methodology

The research process for predicting road accidents was structured into a series of sequential steps, as illustrated in Figure 1. The workflow begins with data collection, utilizing comprehensive records from the SEWIK system of the Road Traffic Office for 2015–2022. This is followed by data preprocessing, where inconsistencies and missing values were addressed to prepare the dataset for analysis. Exploratory data analysis (EDA) was conducted to uncover patterns and relationships within the data, which informed the subsequent feature engineering phase. During feature engineering, new variables were created to enhance the predictive capabilities of the model. The process then advanced to building the model, where a Decision Tree Classifier was developed and fine-tuned. This model was rigorously evaluated for accuracy and interpretability, and its performance metrics were analyzed. Finally, the insights and findings were synthesized into actionable conclusions and applications, providing a foundation for targeted interventions to improve road safety.

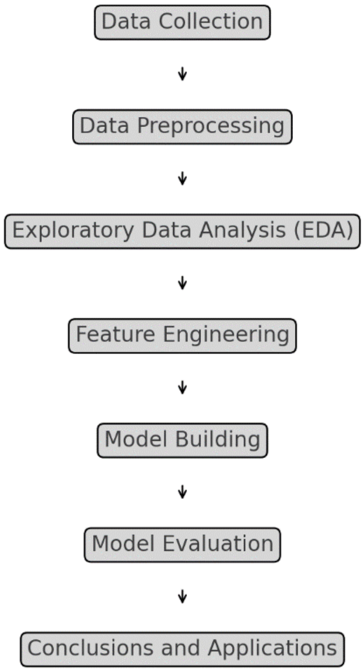

Figure 1. Workflow of the research process for road accident prediction. The authors' own work.

Statistical methods were used in the research work. The data used for the study concern accidents that occurred in Poland in 2015-2022. The data was received from the Police Headquarters. The data concerns over three million road accidents and collisions. According to the Central Statistical Office's definition of an "injured person," an injured person is someone who did not die immediately as a result of an accident or has not died within 30 days but suffered injuries that usually require medical treatment. A seriously injured person is an injured person who, as a result of an accident, stays in hospital for more than 24 hours. A slightly injured is a person injured, excluding death and serious injuries. People who have minor injuries, such as minor cuts or bruises, are not usually recorded as injured. Based on this information, there is a fine line between classifying a person as seriously injured or dead. For example, a person who dies on the 31st day after an accident is statistically seriously injured. There is a fine line between classifying a person as seriously injured or slightly injured. After an accident, accident participants experience many emotions related to stress. A person who feels well immediately after the event and does not require medical attention may feel worse when the emotions stop working. For the purposes of the project, it was decided to calculate the additional variable "victims" as the sum of those killed, seriously injured and slightly injured.

The project was written using the (*Python*) programming language. Python has evolved into a robust and widely adopted tool in the scientific, academic, and industrial communities. Its syntax emphasizes code readability, employing indentation and whitespace to define program structure, making it particularly accessible to beginners. Python's extensive standard library provides an array of modules and packages, facilitating diverse applications in areas such as data science, machine learning, artificial intelligence, and web development. The language's dynamic typing and automatic memory management contribute to its flexibility and ease of use, while its open–source nature fosters a collaborative and supportive global community. Python's adaptability, coupled with its comprehensive ecosystem, positions it as a preferred language for scientific research and development.

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

The code was written using the Jupyter Notebook by Kluyver et al. (2016). It is an interactive online environment for writing and sharing code. The Jupyter Notebook, an interactive computing environment widely embraced in scientific research, education, and data analysis, provides a flexible platform for creating and sharing documents that integrate live code, equations, visualizations, and narrative text. Originally derived from the IPython project, Jupyter supports multiple programming languages, including Python, Julia, and R, fostering interdisciplinary collaboration and accommodating diverse computational workflows. The notebook interface allows users to break down complex problems into manageable, executable cells, enabling an interactive and iterative approach to code development. With its seamless integration of multimedia content and the ability to generate dynamic visualizations, the Jupyter Notebook enhances the reproducibility and transparency of scientific analyses. Moreover, the open-source nature of Jupyter facilitates the creation of interactive documents that can be effortlessly shared, encouraging collaboration and accelerating the dissemination of scientific findings. The Jupyter Notebook has become an integral tool in the scientific community, providing a versatile and accessible environment for conducting, documenting, and communicating research. The Jupyter Notebook was utilized by Khan & Hussain, (2024) for traffic accident prediction, where machine learning algorithms like the random forest, linear regression, and decision tree were developed to analyze road geometry, traffic data, and other factors, achieving an 84.4% accuracy in predicting accident hotspots.

The Pandas library by McKinney (2010) was used to process the data. Creating new features was done by summing column values in Data Frame Pandas. The Pandas library, a powerful and widely employed data manipulation and analysis tool for the Python programming language, plays a pivotal role in the scientific and data science communities. Pandas introduces two key data structures, namely series and data frame, which efficiently handle labelled and structured data. Leveraging NumPy arrays, Pandas excels in its ability to manipulate, clean, and analyze datasets, offering functionalities for indexing, merging, and reshaping data. Its seamless integration with other libraries, such as NumPy and Matplotlib, contributes to its popularity in the data science ecosystem. The library's comprehensive set of tools for data manipulation, exploration, and preprocessing establishes Pandas as an essential component in the data analysis workflow, facilitating tasks ranging from data ingestion to transformation and statistical analysis. With its intuitive syntax and extensive documentation, Pandas continues to be an indispensable resource for researchers, analysts, and data scientists engaged in diverse scientific domains.

The Matplotlib by Hunter (2007) and Seaborn by Waskom (2021) libraries were used to visualize the data. Matplotlib was used to create bar and line charts. Seaborn was used to create a heat map. Matplotlib and Seaborn, both prominent Python libraries, are indispensable tools for creating high–quality, publication–ready visualizations for scientific research and data analysis. Matplotlib, initially developed by John D. Hunter, offers a comprehensive set of plotting functionalities for static, interactive, and animated visualizations. With support for a wide array of plot types and customization options, Matplotlib provides researchers and analysts with the flexibility to represent complex data in a visually informative manner. Seaborn, built on top of Matplotlib, complements its capabilities by simplifying the creation of aesthetically pleasing statistical graphics. Developed by Michael Waskom, Seaborn specializes in producing attractive and informative visualizations for exploratory data analysis, leveraging concise syntax and default themes. Together, Matplotlib and Seaborn form a powerful duo, allowing scientists and data practitioners to generate insightful and compelling visualizations that enhance the interpretability of data, aiding in the communication and dissemination of scientific findings.

4. Analysis

Table 1 shows basic information about the data set, which includes 3,499,245 entries. The dataset encompasses a comprehensive set of features related to road safety incidents, each providing valuable information for understanding and analyzing the occurrences of road accidents. The key features include the KSIP event ID and event ID for unique identification, the date of the event for temporal analysis, and location–related details such as county, community, voivodeship, town, street, house number, road number, and KM HM (chainage). Details about the spatial characteristics are further augmented by information on roundabouts, junctions, intersections with streets or roads, and distances to intersections, specified by GPS coordinates (GPS X and GPS Y). The dataset captures crucial details about the accident or collision, including the area, road geometry, and intersection type. Factors influencing incidents, such as lighting, weather conditions, and road surface conditions, are meticulously documented. Other relevant features cover event type, permissible speed, road type, presence of traffic lights, road markings, and additional factors contributing to the incident. This rich and diverse set of features forms the foundation for a comprehensive analysis of road safety incidents, enabling the application of advanced analytical techniques, including machine learning, to predict and prevent future occurrences.

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

Table 1. Basic information about the data set.

Feature	Dtype
KSIP event ID	object
event ID	int64
date of the event	object
County	object
Community	object
Voivodeship	object
Town	object
Street	object
house number	object
road number	object
KM HM (chainage)	object
roundabout/junction	object
intersection with a street	object
intersection with a road	object
distance to the intersection	float64
GPS X	object
GPS Y	object
accident/collision	object
Area	object
road geometry	object
Intersection	object
characteristics of the scene of the incident	object
Lighting	object
weather conditions	object
type of event	object
permissible speed	float64
type of road	object
road surface	object
surface condition	object
traffic lights	object
road markings	object
other reasons	object

The presented data in Figure 2 shows the counts of road safety incidents categorized by voivodeship (province), providing insights into the regional distribution of such occurrences. The “Mazowieckie” voivodeship exhibits the highest count, with 537,011 incidents, underscoring its prominence in the dataset. Following closely, “Śląskie” ranks second with 433,880 incidents, suggesting a notable concentration of road accidents in this region. The subsequent voivodeships with substantial incident counts include “Dolnośląskie” (323,373), “Wielkopolskie” (288,164), and “Małopolskie” (264,476). Conversely, regions such as “Opolskie” (81,649) and “Lubuskie” (94,527) exhibit relatively lower incident counts. The chart provides a foundational overview of the geographic distribution of road accidents, setting the stage for further in–depth analyses and targeted interventions to address road safety challenges in specific voivodeships.

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

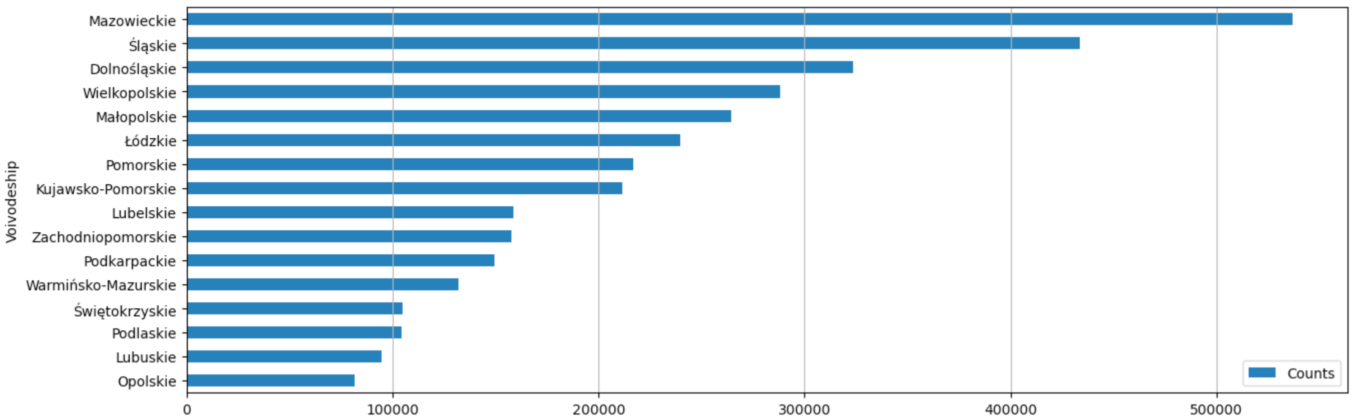

Figure 2. The number of road incidents by voivodeship. The authors' own work. Data obtained from the SEWIK system of the Road Traffic Office of the Police Headquarters

The provided data in Figure 3 delineates the distribution of road safety incidents based on the type of incident, distinguishing between "Accident" and "Collision." The count for "Accident" is recorded at 229,363 instances, while "Collision" demonstrates a substantially higher count of 3,269,882 occurrences. This classification sheds light on the prevalence of different incident types within the dataset, with collisions representing the predominant category. Understanding the composition of incident types is pivotal for tailoring preventive measures and safety interventions, as it provides a nuanced perspective on the nature of road safety challenges. Further analysis of contributing factors specific to each incident type can offer valuable insights for developing targeted strategies to reduce both accidents and collisions on the road.

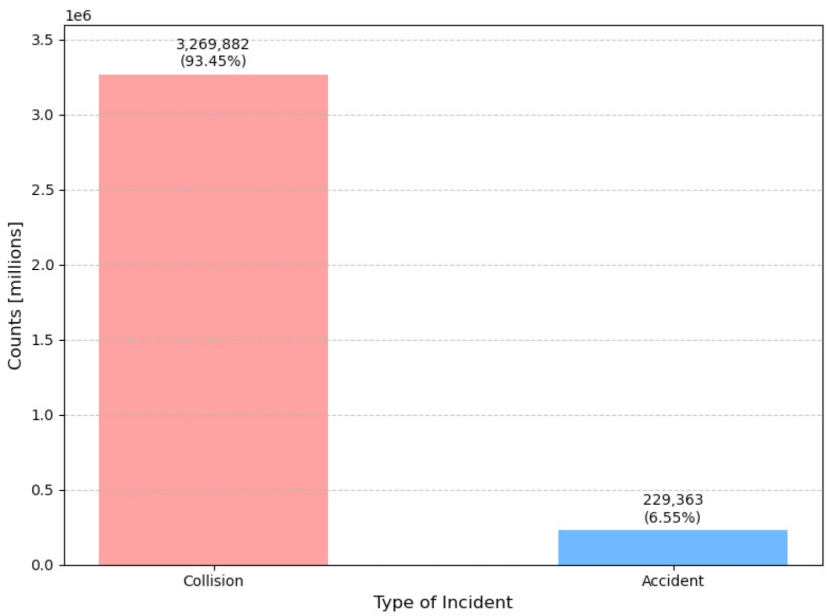

Figure 3. The number of collisions and accidents. The authors' own work. Data obtained from the SEWIK system of the Road Traffic Office of the Police Headquarters

The provided data in Figure 4 delineates the distribution of road safety incidents based on the area's classification into "Developed" and "Undeveloped" categories. The count for incidents occurring in "Developed" areas is recorded at 2,702,805 instances, while "Undeveloped" areas demonstrate a count of 796,341 occurrences. This categorization provides a key distinction between incidents transpiring in urban or developed environments and those transpiring in rural or undeveloped

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

settings. Analyzing the distribution of incidents across these areas can inform targeted safety measures and infrastructure improvements specific to the unique challenges posed by developed or undeveloped regions. Understanding the dynamics of road safety incidents in different areas is crucial for formulating effective policies and interventions tailored to the specific characteristics of the environment.

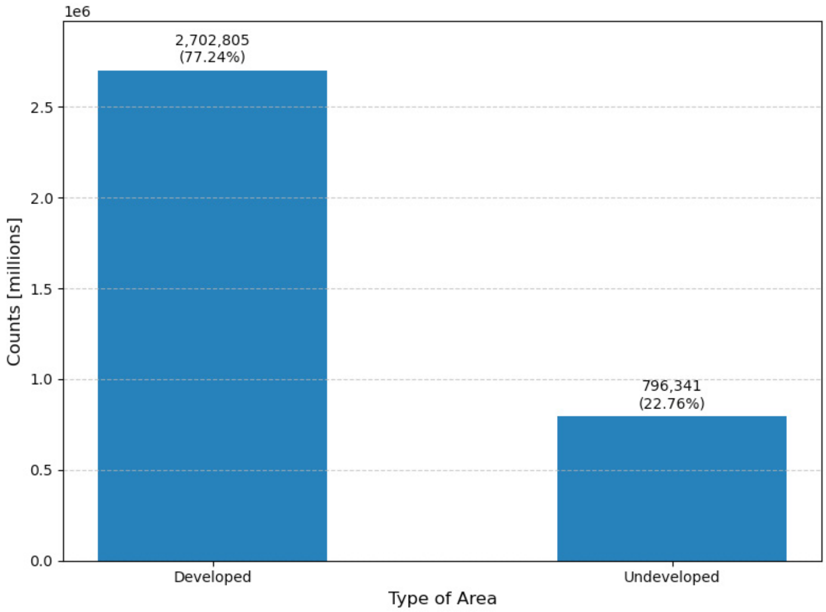

Figure 4. The number of incidents by area. The authors' own work. Data obtained from the SEWIK system of the Road Traffic Office of the Police Headquarters

The provided data in Figure 5 presents the distribution of road safety incidents based on the road geometry, categorizing incidents into different types such as "Straight," "Curve," "Hill," and combinations thereof. The most prevalent incidents occur on "Straight" roads, with a count of 2,176,323 instances. Following closely, "Curve" roads exhibit a substantial count of 272,413 incidents. Noteworthy variations include incidents on "Straight, Downhill" (37,322) and "Curve, Downhill" (14,652), indicating the significance of road slope in certain incident scenarios. The dataset further distinguishes incidents occurring at specific topographic locations, such as "Top of the hill" and "Curve, Top of the hill," suggesting the influence of elevation changes on road safety. This breakdown of incidents by road geometry provides valuable insights into the spatial characteristics contributing to accidents, offering a foundation for targeted safety interventions and infrastructure improvements tailored to different road configurations. Attention is drawn to the least numerous category, i.e., "Straight, Curve", which is mutually exclusive. This is presumed to be a recording error.

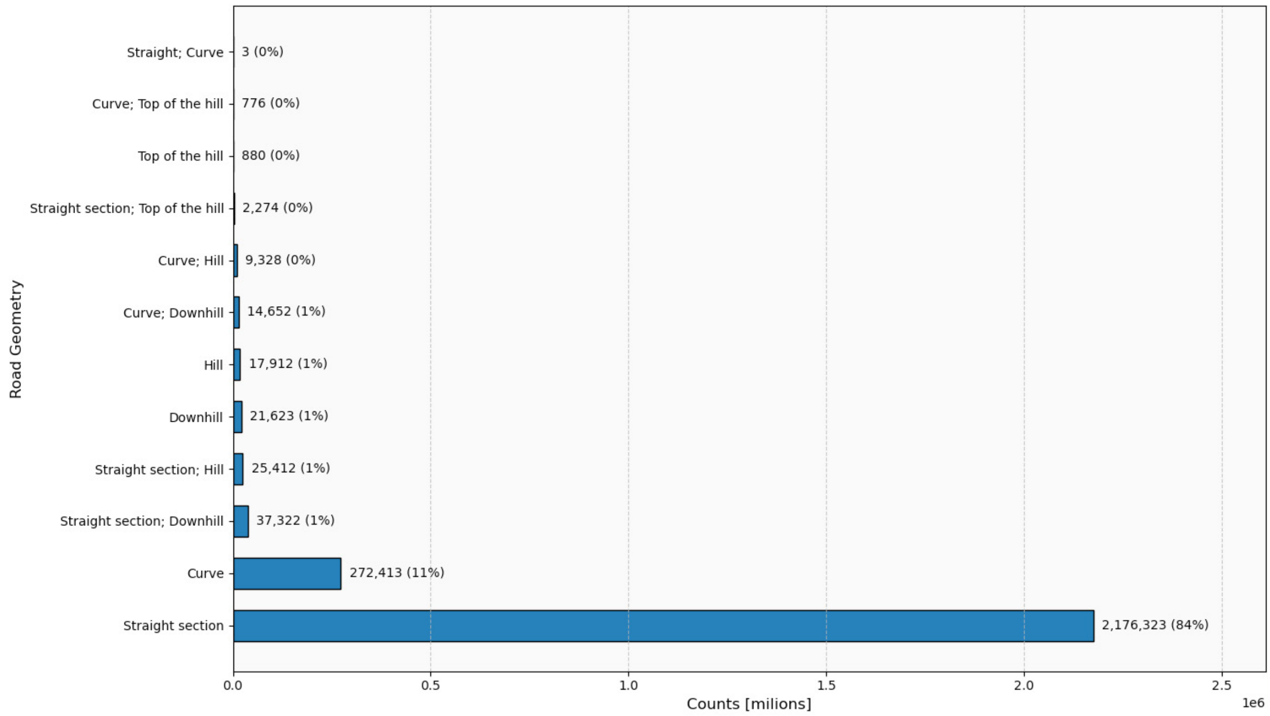

Figure 5. The number of incidents by road geometry. The authors' own work. Data obtained from the SEWIK system of the Road Traffic Office of the Police Headquarters

The presented data in Figure 6 delineates road safety incidents based on the type of intersection, categorizing them as "with priority," "Roundabout," and "equivalent." The highest count is associated with incidents occurring at intersections "with priority," totaling 732,221 instances. Notably, incidents at "Roundabouts" exhibit a count of 114,488, while those classified as "equivalent" intersections have a count of 22,895. Understanding the distribution of incidents across different intersection types is crucial for developing targeted safety measures, as each type presents distinct challenges and requires specific interventions. This information can inform traffic management strategies, infrastructure improvements, and educational initiatives aimed at enhancing safety at intersections and reducing the occurrence of road accidents.

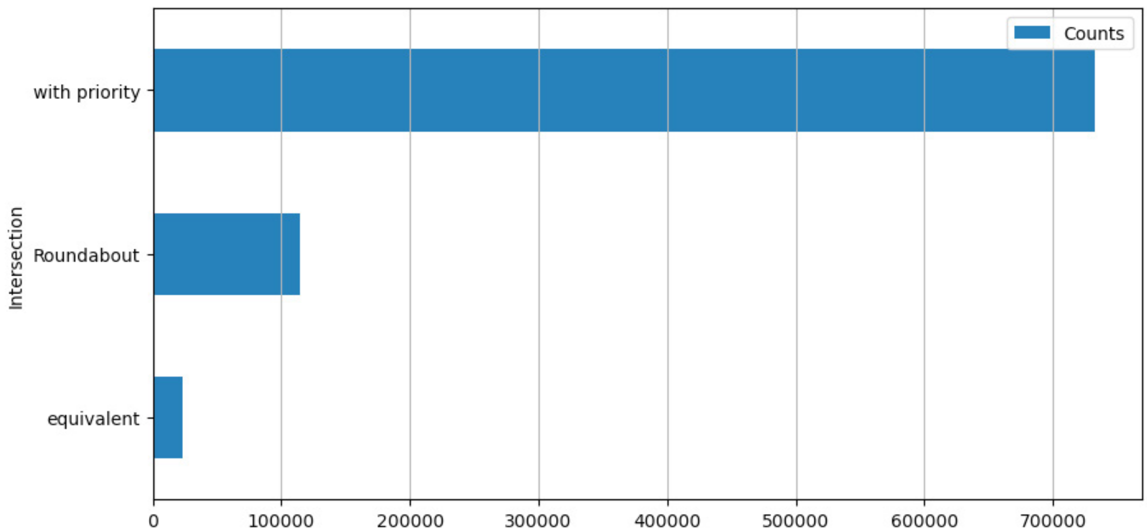

Figure 6. The number of incidents by intersection. The authors' own work. Data obtained from the SEWIK system of the Road Traffic Office of the Police Headquarters

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

5. Modelling

A classification machine learning model has been trained. Its purpose is to predict whether a collision or accident will occur under given conditions. It was decided to choose the Decision Tree Classifier model. The precision of the model is 0.94.

The provided weights in Figure 7 for each feature in the Decision Tree Classifier model offer insights into the relative importance of different factors in predicting whether a collision or accident will occur under given conditions. The higher the weight, the more influential the corresponding feature is in the model's decision–making process.

In this context:

- Type of Incident (Weight: 0.9059): The type of incident emerges as the most critical feature in the model, indicating that the nature of the event significantly influences the prediction. This aligns with the common understanding that different incident types may have distinct characteristics affecting their likelihood.
- Characteristics of the Place of the Event (Weight: 0.0316): This feature holds a moderate weight, suggesting that the specific characteristics of the event's location contribute meaningfully to the model's predictions.
- Road Number (Weight: 0.0138): The road number is also considered in the model, albeit with less weight compared to the type of incident and characteristics of the place.
- Intersection (Weight: 0.0132): The presence and type of intersections play a role in the model's predictions, indicating their relevance to road safety incidents.
- Street (Weight: 0.0122): Similar to the road number, the specific street where the event occurs is considered but with a slightly lower weight.
- Lighting (Weight: 0.0079): The lighting conditions at the time of the event are considered, although with a relatively lower impact compared to other features.
- Condition of the Surface (Weight: 0.0056): The surface condition is a factor, but it has a lower weight in the model's decision–making process.
- County (Weight: 0.0038): The county where the event occurs is considered but has a relatively minor impact on the model.
- Year (Weight: 0.0026): The temporal aspect, represented by the year of the event, is a factor, but its contribution is comparatively lower.
- Voivodeship (Weight: 0.0019): The voivodeship, representing the administrative region, has the lowest weight among the features considered.

Understanding feature importance aids in interpreting the model's behavior and can guide future interventions or data collection efforts to further enhance the model's predictive capabilities.

Weight	Feature
0.9059	The type of incident
0.0316	Characteristics of the place of the event
0.0138	Road Number
0.0132	Intersection
0.0122	Street
0.0079	Lighting
0.0056	The condition of the surface
0.0038	County
0.0026	Year
0.0019	Voivodeship
	… 18 more …

Figure 7. Top model features. The authors' own work. Data obtained from the SEWIK system of the Road Traffic Office of the Police Headquarters

6. Results

The Decision Tree Classifier model, trained to predict the occurrence of road accidents or collisions under given conditions, has yielded promising outcomes. The precision of the model, a key performance metric, stands at an impressive 0.94, signifying the model's accuracy in correctly identifying instances of accidents or collisions. This high precision indicates a robust ability to minimize false positives, a crucial aspect in applications where misclassifying non-accident instances as accidents has significant consequences.

Feature importance analysis further provides valuable insights into the factors influencing the model's predictions. The most influential feature is the "Type of Incident" with a weight of 0.9059, underscoring its paramount role in the model's decision-making process. The "Characteristics of the Place of the Event" (weight: 0.0316) and "Road Number" (weight: 0.0138) follow closely,

Safety & Defense 10(2) (2024)
DOI: https://doi.org/10.37105/sd.230

contributing meaningfully to the predictive capabilities of the model. Other features such as “Intersection,” “Street,” “Lighting,” and “Condition of the Surface” also play roles in shaping predictions, although with varying degrees of importance.

It is important to note, however, that the scope of this research was limited to data available in police statistics, which primarily include measurable and well-documented factors such as road geometry, lighting, and weather conditions. This statistical approach inherently excludes significant human-related factors, such as driver fatigue, stress, or inattention, which are not captured in the available datasets. While these variables can profoundly impact road safety, their absence in the analysis underscores a limitation of the current study. For example, the study investigates the impact of factors such as driving duration, body mass index (BMI), road types, and gender on driving fatigue, using EEG signals to analyze cognitive impairments and their correlation with fatigue.

7. Discussion

The rapid increase in global vehicle usage has resulted in significant challenges, including traffic congestion, delays, accidents, and elevated energy consumption. Addressing these issues requires innovative solutions, and the application of artificial intelligence (AI) models offers a promising approach. In the context of road traffic safety, predictive models, such as the Decision Tree Classifier used in this study, provide valuable insights into accident risks based on measurable factors like road geometry, lighting, and weather conditions.

This study highlights the potential of data-driven approaches to inform targeted interventions. However, it is essential to acknowledge limitations. The reliance on police statistics excluded critical human factors, such as driver fatigue or distraction, which are known contributors to accidents. Moreover, the imbalanced dataset limited the model’s ability to predict rare but critical events, such as fatalities. Future research should explore integrating real-time data sources, such as IoT sensors or wearable devices, to capture these nuanced factors.

While other studies, such as Tai et al. (2018), have developed systems to streamline accident processing and evidence collection, our findings emphasize the need for predictive tools to proactively mitigate risks. The use of models like Artificial Neural Networks (ANNs) could further enhance accuracy and capture complex patterns, offering opportunities for integrating real-time data into intelligent transportation systems. However, these approaches require addressing challenges such as data quality, model interpretability, and ethical considerations, particularly in driver behavior analysis.

Future research directions include refining existing models, unifying data collection frameworks, and exploring the societal impacts of advanced transportation technologies. By addressing these challenges, intelligent systems can evolve into more effective tools for enhancing road safety and reducing societal costs associated with traffic accidents.

8. Conclusions

The journey through the exploration of intelligent transportation systems powered by smart statistical AI models culminates in profound insights and implications for the future of urban mobility. The escalating global population has propelled a surge in vehicular usage, necessitating innovative solutions to mitigate challenges like traffic congestion, delays, accidents, and increased energy consumption. The adoption of artificial intelligence (AI) and neural network models emerges as a transformative approach, offering not only immediate solutions but also shaping the trajectory of intelligent transportation systems (ITS) in the years to come.

The primary contributions of this study lie in the successful application of AI models to address critical facets of transportation challenges.

The amalgamation of these smart statistical AI models paints a vision of a civic transportation system.

While the models showcased substantial efficacy, acknowledging certain limitations is crucial for future endeavors. Challenges include ensuring data quality, enhancing model interpretability, and addressing ethical considerations, particularly in driver behavior detection. Future research directions may involve refining existing models, incorporating real–time data sources, and exploring the broader societal impacts of ITS.

In conclusion, this study underscores the transformative potential of smart statistical AI models in reshaping the landscape of transportation. The deployment of these models is not merely a solution to existing problems but a step towards fostering a dynamic, responsive, and sustainable transportation future. As technology continues to evolve, the integration of AI in transportation systems holds the promise of making urban mobility safer, more efficient, and conducive to the evolving needs of society. The journey towards intelligent transportation is ongoing, and this study contributes a foundational understanding and framework for future advancements in the field.

The study analyses the health and social consequences of road traffic accidents, focusing on their impact on individuals, families, and society, using textual analysis of documents as the research method.

Acknowledgments
The project was implemented as part of the X competition for financing project-oriented education – PBL, as part of the Excellence Initiative – Research University of the Silesian University of Technology program.
The authors would like to thank the Road Traffic Office of the Police Headquarters for providing data from the SEWIK system.

Declaration of interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References
1. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55
2. Ho, J. X., Shamsuddin, S., Kamat, S. R., Ibrahim, M. S., & Setiawan, R. (2025). Investigation on Factors Affecting Cognitive Skills in Detection of Driving Fatigue. Journal of Advanced Research in Applied Sciences and Engineering Technology, 51(2), 270–280. https://doi.org/10.37934/araset.51.2.270280
3. Jesi, V. E., L, J. M., Saranya, S., C, P., A, M. T., & Saranya, K. (2023). Advancements in Sensor Technology and Computer Vision for Enhanced Road Safety and Anomaly Detection using Machine Learning Approach. *2023 Third International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, 457–462. https://doi.org/10.1109/ICUIS60567.2023.00081
4. Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J. B., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Team, J. D. (2016). Jupyter Notebooks—A publishing format for reproducible computational workflows. *International Conference on Electronic Publishing*. https://api.semanticscholar.org/CorpusID:36928206
5. Khan, A. A, Hussain, J. (2024). Utilizing GIS and Machine Learning for Traffic Accident Prediction in Urban Environment. *Civil Engineering Journal, 10*(6), 1922–1936. https://doi.org/10.28991/CEJ-2024-010-06-013
6. Kodieswari, A., Sabarmathi, K. R., Remya, K., Gayathiri, N. R., Nithyapriya, S., & Malathi, T. (2024). Statistical AI Model in an Intelligent Transportation System. In *Artificial Intelligence for Future Intelligent Transportation: Smarter and Greener Infrastructure Design*.
7. Másilková, M. (2017). Health and social consequences of road traffic accidents. *Kontakt*, *19*(1), e43–e47. https://doi.org/10.1016/j.kontakt.2017.01.007
8. Mehta, K., Jain, S., Agarwal, A., & Bomnale, A. (2022). Road Accident Prediction Using Xgboost. *2022 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, 50–56. https://doi.org/10.1109/ICETCI55171.2022.9921367
9. Smith, A. P. (2021). Caffeine and road traffic accidents. *World Journal of Pharmaceutical and Medical Research, 7*(12), 1–4. https://orca.cardiff.ac.uk/id/eprint/145222/1/article_1635574300.pdf
10. Subhan, F., Ali, Y., & Zhao, S. (2023). Unraveling preference heterogeneity in willingness-to-pay for enhanced road safety: A hybrid approach of machine learning and quantile regression. *Accident Analysis & Prevention*, *190*, 107176. https://doi.org/10.1016/j.aap.2023.107176
11. Tai, W.-K., Wang, H.-C., Chiang, C.-Y., Chien, C.-Y., Lai, K., & Huang, T.-C. (2018). RTAIS: Road Traffic Accident Information System. *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, 1393–1397. https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00230
12. Tayebikhorami, S., & Sacchi, E. (2022). Validation of Machine Learning Algorithms as Predictive Tool in the Road Safety Management Process: Case of Network Screening. *Journal of Transportation Engineering, Part A: Systems*, *148*(9), 04022068. https://doi.org/10.1061/JTEPBS.0000719
13. Waskom, M. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021
14. Wu, D., & Wang, S. (2020). Comparison of road traffic accident prediction effects based on SVR and BP neural network. *2020 IEEE International Conference on Information Technology,Big Data and Artificial Intelligence (ICIBA)*, 1150–1154. https://doi.org/10.1109/ICIBA50161.2020.9277150
15. Xing, Y., Song, W., Liu, W., & Gao, S. (2023). Road Traffic Accident Prediction Based on BP Neural Network. In W. Wang, J. Wu, X. Jiang, R. Li, & H. Zhang (Eds.), *Green Transportation and Low Carbon Mobility Safety* (Vol. 944, pp. 651–660). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-5615-7_46