Determining and vigilance of the Road Accidents Hotspots using Machine Learning Algorithms

Mandarapu Hemanth[1*], Mavuluri Datha Sushma[1], Myla Krishna Rajitha[1], Mandava Giridhar sundar[1], and Swathi Mutyala[2]

[1]Department of Information Technology, Seshadri rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India
hemanthmadarapu13@gmail.com,mavulurisushma50@gmail.com,myla-krishnarajitha@gmail.com, giridharasundar@gmail.com
[2] Department of Information Technology, Seshadri rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India
swathigec7@gmail.com

Abstract. Worldwide, traffic accidents result in fatalities, injuries, and financial losses. Accurate models for predicting accident severity are essential for transportation systems. This study focuses on constructing injury severity classification models using key variables and various machine learning techniques. Supervised algorithms (Random Forests, Decision Trees, Logistic Regression, and K-Nearest Neighbors) are employed, with the SMOTE algorithm addressing data imbalance. Findings indicate that Logistic Regression and SVM models effectively determine injury severity. Additionally, leveraging user GPS data, the system proactively alerts users before reaching accident-prone areas, visually mapping these locations.
Keywords: Accident Hotspots, Machine Learning, GPS, Proactive alert System.

1 Introduction

Road Traffic Accidents on the Rise in Andhra Pradesh

Road traffic accidents (RTAs) are a growing threat to public health in Andhra Pradesh, with a significant increase reported in 2022. Recent data indicates a 20% rise in RTAs compared to 2021. This translates to a concerning number of fatalities and injuries: 7,039 deaths and 19,675 injuries in 2022 alone (up from 6,371 deaths and 16,188 injuries in 2021).

Overspeeding: A Persistent Culprit

The data highlights speeding as a major factor in RTAs. In 2021, over 16,600 accidents were attributed to exceeding the speed limit. This trend underscores the need for stricter enforcement and driver education programs to address risky behavior behind the wheel.

Harnessing Technology for Prevention

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_46

Advancements in surveillance technology offer a valuable opportunity. The vast amount of data now available can be used to develop targeted accident prevention strategies. By analyzing accident patterns and identifying high-risk areas, authorities can implement effective interventions to save lives.

.

Fig.1 Road Accident

A Call to Action

The rising number of RTAs in Andhra Pradesh demands immediate action. Combining stricter enforcement, educational campaigns, and leveraging data from surveillance technology can significantly reduce road accidents and fatalities. This multi-pronged approach is crucial to ensure safer roads for everyone.

1.1 Difference between Deep Learning and Machine Learning

Machine Learning (ML) requires high specificity from the user since the computer relies on manual input to interpret and search for features. Deep Learning (DL) stands out as it autonomously illustrates feature sets without manual input, ensuring high accuracy, speed, and reliability.

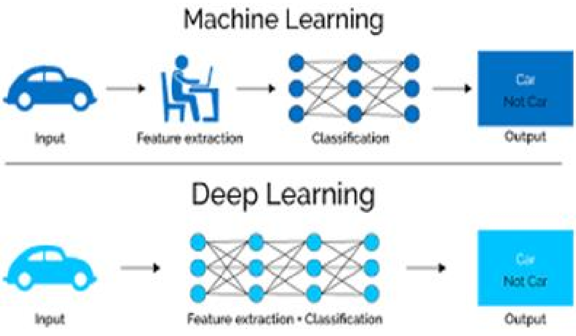

Fig.2 Machine Learning Vs Deep Learning

Enhancing Road Safety with Technology

Geographic Information Systems (GIS) play a crucial role in road safety. This technology allows for the collection, storage, analysis, and visualization of spatial data related to traffic patterns, road networks, and accident locations. By integrating various data sources, such as road layouts, traffic volume, weather conditions, and past accident records, GIS helps identify patterns and trends in accidents.

Real-Time Monitoring for Proactive Action

Live monitoring systems utilize sensors, cameras, and other data collection equipment installed along roadways. This real-time data on traffic conditions, vehicle speeds, and environmental factors allows authorities to proactively address potential threats and prevent accidents. By analyzing this data, they can quickly identify anomalies or risky situations.

Predicting Accidents to Save Lives

Predictive analytics can be a powerful tool for accident prevention. By analyzing historical accident data alongside relevant factors like traffic volume, weather, road design, and time of day, these techniques can identify areas with a high likelihood of future accidents (accident hotspots). Machine learning algorithms can be trained on past data to find patterns and correlations, leading to the development of accurate prediction models.

2 Literature Review

Studies on Accident Severity and Prediction

Impact of Accident Severity on Patterns: A study by Le et al. (2020) investigated how accident severity affects the location and timing of accident hotspots. They analyzed traffic accident data in Hanoi, Vietnam, considering factors like seasons, time of day, and accident severity. This research helps identify areas and times with a higher risk of severe accidents.

Predicting Injury Severity: Paprzycki et al. (2021) proposed a model to automatically categorize traffic accidents based on injury type and severity. This model could be valuable for developing targeted traffic safety policies based on accident patterns and driver behavior.

Reducing Accident Severity in Ethiopia: Beshah and Hill (2021) explored the link between road features and accident severity in Ethiopia. Their data mining approach aimed to identify road characteristics that contribute to serious accidents. This information can be used by traffic authorities to improve road safety measures.

3 Related Work

3.1 Existing System:

- Statistical Analysis
- Spatial Analysis
- Data Acquisition
- Data preprocessing

3.2 Limitations:

Challenges in Identifying Accident Hotspots

Finding accident hotspots involves several complex steps, each with its own challenges:

Statistical Analysis: Statistical methods used to identify hotspots may not always accurately reflect real-world causes. Additionally, complex models and small datasets can lead to unreliable results. Robust procedures, collaboration between different disciplines, and constant validation are crucial to improve the accuracy of hotspot identification.

Spatial Analysis: Techniques used to pinpoint hotspots on maps can be limited by factors like edge effects (accidents near borders appearing less frequent) and the scale chosen for analysis. Spatial resolution limitations can also affect the accuracy of hotspot detection.

Data Acquisition: The quality of data used for analysis is critical. Inconsistent reporting and missing information in accident records can significantly affect the reliability of hotspot identification.

Data Preprocessing: Cleaning and preparing data for analysis, such as handling missing values and outliers, can introduce bias if subjective judgments are made or complex algorithms are used.

3.3 Proposed System:

Machine Learning: The Core Steps

Machine learning (ML) algorithms rely on data to learn and make predictions. Here's a breakdown of the key steps involved:

Data Preparation: Providing high-quality data in the right format is crucial. This includes ensuring the data is relevant, scaled appropriately, and includes all necessary features.

Defining the Task: Clearly identifying the problem you want the ML model to solve is essential. This involves understanding what kind of output you expect from the data.

Model Selection: Choosing the right mathematical model is vital. This model should be able to learn the relationships between the input data and the desired outcome, and perform well on unseen data (data it wasn't trained on).

Training the Model: This stage involves feeding the prepared data into the chosen model. The model learns from this data and adjusts its internal parameters to make accurate predictions. Common training algorithms include Random Forest, KNN classifiers, and Decision Trees.

Evaluation: Once trained, the model's performance is assessed using separate test data. Metrics like accuracy, precision, recall, and F1-score are used to measure its effectiveness. Evaluation helps ensure the model generalizes well and produces reliable predictions in real-world scenarios.

Methodology

3.4 Logistic Regression:

Logistic Regression: From Probabilities to Classification

Logistic regression is a powerful tool for classification tasks. Unlike linear regression that predicts continuous values, logistic regression predicts the probability of an event happening (often labeled as "1"). It achieves this in two key steps:

Modeling with the Sigmoid Function: Similar to how linear regression assumes a linear relationship between data points, logistic regression uses a mathematical function called the sigmoid function. This function transforms the model's output into a probability between 0 and 1.

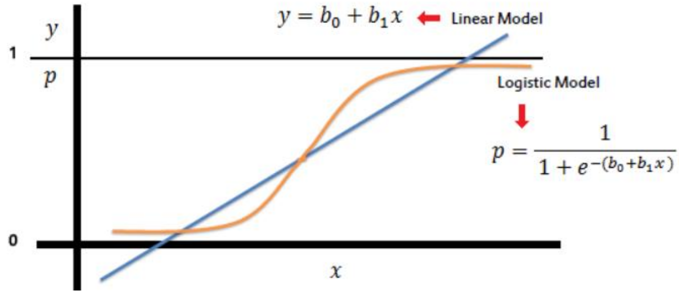

Fig.3 Sigmoid Function

3.5 Decision Tree (DT):

The decision tree stands out as one of the most effective and commonly employed techniques for prediction and categorization. It adopts a tree structure resembling a flowchart, where each internal node represents an attribute test, each branch corresponds to a test result, and each leaf node, also known as a terminal node, holds a class label. This intuitive structure makes decision trees highly interpretable and suitable for a wide range of classification and regression tasks.

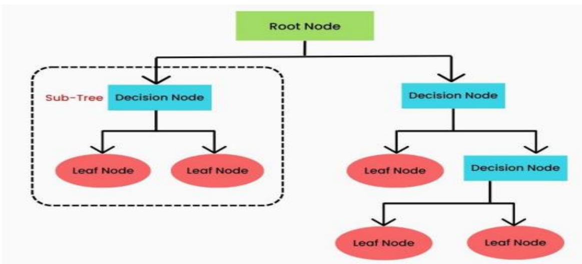

Fig.4 Decision Tree

3.6 Random Forest

Random Forest: Strength in Numbers

Random Forest is a popular machine learning algorithm for tackling both classification (predicting categories) and regression (predicting continuous values) problems. It works by creating a multitude of decision trees, each trained on a different subset of the data. These trees then "vote" to make a final prediction, improving accuracy compared to a single decision tree.

A key advantage of Random Forest is its ability to handle various data types. It can work effectively with datasets containing both numerical values (common in regression) and categorical data (like text labels), making it a versatile tool for real-world applications. This flexibility makes Random Forest robust and widely applicable across different scenarios.

4 Results and Discussions

4.1 Output:

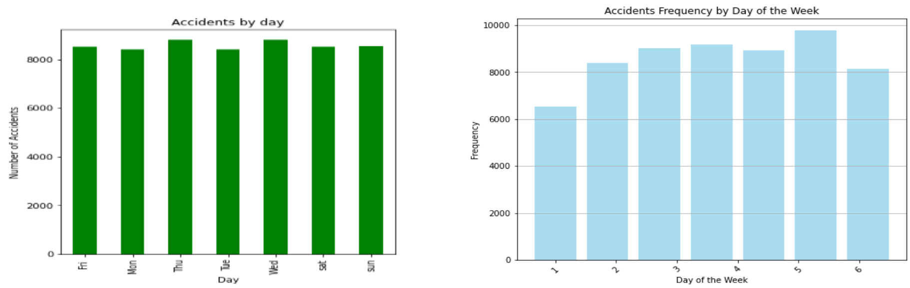

Fig.5 Accidents by day and week

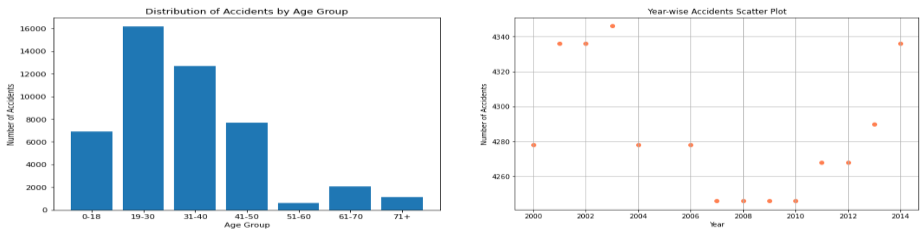

Fig.6 Distribution of Accidents by Age

```
Telemetry data sent:  {'ts': 1708733641048, 'values': {'latitude': 14.72402585, 'longitude':
78.61039332}}
Telemetry data sent:  {'ts': 1708733642057, 'values': {'latitude': 14.76235346, 'longitude':
78.53404158}}
Telemetry data sent:  {'ts': 1708733643060, 'values': {'latitude': 14.74560635, 'longitude':
78.47087739}}
Telemetry data sent:  {'ts': 1708733644075, 'values': {'latitude': 14.66712796, 'longitude':
78.55799399}}
Telemetry data sent:  {'ts': 1708733645084, 'values': {'latitude': 14.70344275, 'longitude':
78.57643143}}
Telemetry data sent:  {'ts': 1708733646091, 'values': {'latitude': 14.73735895, 'longitude':
78.53105133}}
Telemetry data sent:  {'ts': 1708733647091, 'values': {'latitude': 14.77949108, 'longitude':
78.51672417}}
Telemetry data sent:  {'ts': 1708733648107, 'values': {'latitude': 14.74135682, 'longitude':
78.57055271}}
```


Fig.7 Data Being sent to ThingsBoard

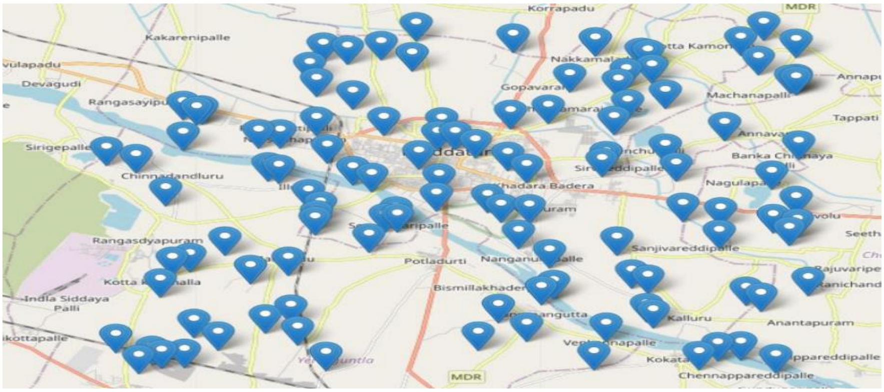

Fig.8 Hotspot Locations on Map

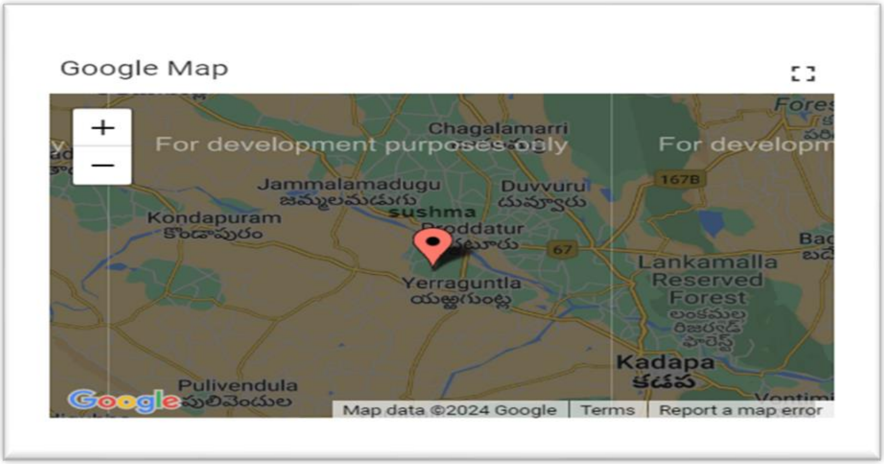

Fig.9 Hotspot Locations on ThingsBoard Widget

5 Conclusion

This study compared three machine learning algorithms (Decision Tree, Logistic Regression, and Random Forest) for their effectiveness in predicting freeway accidents. The results suggest that Logistic Regression outperformed the other models, achieving an accuracy rate of 86.92% in predicting accidents. This information can be valuable for transportation planners and highway engineers in designing safer roads. However, further research is needed to explore the impact of additional factors on accident occurrence. While Random Forest shows promise for predicting accidents, the study acknowledges limitations. The lack of specific data on factors like driver behaviour, traffic conditions, and accident severity restricts the model's ability to predict these aspects.

6 References

1. Le, H., Vu, H., Tran, M., & Nguyen, T., 2020. This study investigates how GIS and statistical methods can be used to identify accident hotspots in Hanoi, Vietnam, considering factors like time of day and season.
2. Chong, S., Abraham, A., & Paprzycki, M., 2004. This paper explores the use of machine learning techniques for analyzing traffic accident data.
3. Dejene, M., Tadele, M., & Tesfaye, T., 2012. This research focuses on data quality, ensemble methods, and trend analysis in traffic accident data mining to improve road safety in Ethiopia.
4. Krishnaveni, S., & Hemalatha, M., 2011. This study examines traffic accidents from a data mining perspective, utilizing relevant techniques for analysis.
5. Beshah, T., & Hill, S., 2010. This research investigates the role of road-related factors on accident severity in Ethiopia, using data mining techniques to improve road safety.
6. Nazneen, S., Vachal, K., & Benson, A., 2018. This paper explores the factors contributing to severe traffic accidents on the Fort Peck Indian Reservation.
7. WHO | Road traffic injuries, 2017.
8. Michigan State Police, Michigan Traffic Crash Decade-At-A-Glance, 2018.
9. Traffic Accident Analysis Using Machine Learning Paradigms, May 2005 by Miao M Chong, Ajith Abraham, Marcin Paprzycki.
10. Identification of Accident Hotspots Using Clustering Algorithms in Machine Learning
11. K. Ravikanth, K. Chandra Shekar, K. Shashi Kethana, V. Sai Praveena & CH. Sharon Rachel
12. A Machine Learning Approach for Classifying Road Accident Hotspots by Brunna de Sousa Pereira Amorim, Anderson Almeida Firmino, Claudio de Souza Baptista, Geraldo Braz Júnior, Anselmo Cardoso de Paiva and Francisco Edeverton de Almeida junior.
13. Machine Learning Approaches to Traffic Accident Analysis and Hotspot Prediction by Daniel Santos, José Saias, Paulo Quaresma and Vítor Beires Nogueira.
14. Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis and Prevention, 41(3), 359–364. https://doi.org/10.1016/j.aap.2008.12.014

15. Sagberg, F. (1999). Road accidents caused by drivers falling asleep. Accident Analysis and Prevention, 31(6), 639–649. https://doi.org/10.1016/s0001-4575(99)00023-8
16. Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Thomsen, L. K. (1995).] Accident Analysis and Prevention, 27(1), 1–20. https://doi.org/10.1016/0001-4575(94)e0023-e
17. Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-Temporal Clustering of road Accidents: GIS based analysis and assessment. Procedia: Social & Behavioral Sciences, 21, 317–325. https://doi.org/10.1016/j.sbspro.2011.07.020
18. Edwards, J. B. (1998). The relationship between road accident severity and recorded weather. Journal of Safety Research, 29(4), 249–262. https://doi.org/10.1016/s0022-4375(98)00051-6
19. Mohan, D. (2009). ROAD ACCIDENTS IN INDIA. IATSS Research, 33(1), 75–79. https://doi.org/10.1016/s0386-1112(14)60239-9
20. Jones, A. P., & Jørgensen, S. H. (2003). Accident Analysis and Prevention, 35(1), 59–69. https://doi.org/10.1016/s0001-4575(01)00086-0