IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

A ROAD ACCIDENT PREDICTION MODEL USING DATA MINING TECHNIQUES

[1]P MOUNIKA, [2]T.KUSUMITHA

[1](Assistant Professor), MCA, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

[2]MCA, scholar, **DANTULURI NARAYANA RAJU COLLEGE(A) PG COURSES, BHIMAVARAM ANDHRA PRADESH**

ABSTRACT

Daily accident rates are rising at an alarming pace, mostly as a result of the skyrocketing availability of motor cars. These days, there are a lot of traffic mishaps and fatalities, thus the transportation department needs a way to predict how many accidents will happen in a certain time period so they can make informed judgments. To better understand this situation and develop strategies to mitigate accidents, it would be beneficial to examine their frequency of occurrence. It becomes apparent after a while that there is a certain pattern to the accidents that happen in a single region, even though most incidents are characterized by ambiguity. Making educated guesses about the frequency of accidents in a given region and creating models to do so may be aided by this regularity. Our research in this study has focused on the environmental elements that contribute to road conditions, the correlations between these variables, and the frequency and severity of road accidents. We used the Apriori algorithm and Support Vector Machines, two data mining tools, to build a model that can anticipate accidents. This research used publicly accessible records (from 2014 to 2017) on vehicle accidents in Bangalore. This study's findings may be valuable for a variety of stakeholders, such as public works agencies, contractors, and other automotive businesses, who can utilize the estimations to improve road and vehicle design.

1.INTRODUCTION

A major reason for worry is the increasing pace at which accidents are occurring in

ISSN 2347–3657

Volume 12, Issue 3, 2024

India. While it only has 1% of the world's vehicles, India is responsible for around 6% of all road accidents, according to current data [1]. A large number of accident instances are attributed to carelessness on two-wheelers, with overspeeding being another culprit. Also prevalent are accidents caused by drunk drivers or by other traffic infractions. Despite rules and traffic standards, many accidents still occur because drivers are careless with factors including vehicle speed, vehicle condition, and their personal safety, such as not wearing helmets. Although the proliferation of cars is said to be the primary culprit in road accidents, the state of the roads and other environmental elements also play a part.

The high body count from vehicular accidents in India is very concerning. With almost 137,000 persons injured in traffic accidents, the situation is rather bleak. This amount exceeds the yearly death toll from terrorist attacks by more than four times. Tragically, many innocent lives are lost in accidents involving commercial vehicles used for public transit, such as buses, or large cargo vehicles, such as trucks. Inclement weather, such as rain, fog, etc., might increase the likelihood of accidents. Consequently, it is helpful to have an accurate estimate of accidents as well as information about accident hotspots and contributing variables in order to take measures to decrease them. To do this, accident data must be carefully studied and models for accident prediction must be developed.

A well-designed road framework management system that takes road security into account should have an optimized accident prediction model that can assess the impact of current models on accident reduction and identify problems that may arise as a result of infrastructure failures. The most difficult part of making this kind of model is figuring out how to include all the factors that may have contributed to the disaster and how to give each factor the weight it deserves in the final product. A number of fields have already discovered data mining methods and models helpful for data interpretation, such as recommendation systems, credit risk management, fraud detection, healthcare informatics, and many more. These research have been further enhanced by methods that use artificial intelligence and machine learning. Our research for this study has focused on the ways in which environmental elements and

ISSN 2347–3657

Volume 12, Issue 3, 2024

underlying road conditions interact to increase the likelihood of traffic accidents. We may use data mining methods to extract essential facts from this massive amount of data, which would be worthless without the correct interpretation, as this research needs us to cover various elements causing accidents.

In this article, we'll go over how this kind of accident prediction model might help us spot potential dangers on the road. What follows is a discussion of previous efforts made in this area, specifically with regard to examining the many incidents that have occurred over the years. After that, the approach employed in this paper is summarized. In addition, the discussion covers the many aspects of implementation, such as the system architecture, software and languages utilized, simulation, user interface, and screenshots of the program that was produced. This research concludes with a discussion of its findings and an overview of its future directions in the final two parts. The study's findings have informed the development of a model that users may input to get an idea of how likely it is that traffic accidents will occur in a certain region.

.

2.LITERATURE SURVEY

The rising number of accidents in India has stimulated a great deal of academic interest in the causes and consequences of traffic mishaps. In the past, many types of data mining approaches were used to construct road accident prediction models. This is due to the fact that data mining techniques do not rely on specific assumptions between dependent and independent variables, as do conventional statistical methods. In creating these models, researchers have concentrated on several sets of characteristics. Studying accidents at crossing locations has mostly been the focus of Srivastava and Ghazizadeh et al. [3]. The first one used a more effective Multi-layered perceptron (MLP) technique to classify accidents according to severity, while the second one used a feed forward MLP that used back propagation learning to examine how factors like time of day, traffic conditions, and more affect accidents. Research by Chen et al. [4] indicates that highways are the most often reported locations for accidents.

Research by Williams et al. [5] shows that a driver's age and level of experience are significant factors in accident rates. The performance of various classification

ISSN 2347–3657

Volume 12, Issue 3, 2024

algorithms, including linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, Random Forest, and gradient boosting algorithm, was compared in a paper by Suganya, E. and S. Vijayarani [6]. The algorithms were evaluated based on accuracy, error rate, and execution time. According to their findings, KNN outperforms the competition. A comparative research on the types of roadways that are prominent in accidents has been conducted by Sarkar et al. [7]. While looking at additional factors linked to accidents, they discovered that highways had a higher accident rate than regular roads [4]. A neural network model for accident prediction was constructed using actual data by Stewart et al. [8]. They discovered that this model outperformed the ones being utilized in the models developed for Indian roads in terms of speed.

The variety of injuries sustained in car accidents has been investigated by Zheng et al. [9], who have also examined the potential role of the drivers' emotions as a contributing factor. In their comprehensive review, Arun Prasath N and Muthusamy Punithavalli [10] cover the many approaches, methods, and strategies employed in road accident detection throughout the years, as well as their pros and cons.

Current approaches utilized in the creation of accident prediction models on an international basis are described in the study by George Yannis et al. [11]. In order to determine which model may be most effective for accident prediction, they used questionnaires to gather detailed information on several models.

In order to identify and anticipate global atmospheric degradation, Anand, J. V. [12] devised a technique to ascertain the impact of several factors. This approach was created using R-studio, the ARIMA framework, and fuzzy C means clustering. The effect of different variables on traffic accidents may also be investigated using a similar method. To determine the relative importance of each element in road accidents, it is necessary to examine their root causes. In order to categorize road accident data according to the types of road users, Tiwari et al. [13] used self-organizing maps, K-mode clustering algorithms, Support Vector Machines, Naïve Bayes, and Decision trees.

Accident hotspots may be better understood by analyzing the historical data. N. Singh et al. [15] used this to provide the groundwork

ISSN 2347–3657

Volume 12, Issue 3, 2024

for a model to identify potential accident hotspots. Additionally, Kaur, G. et al. [14] have developed a model for predicting accidents on state highways and regular district roads by studying data on traffic collisions and road accidents using R and several visualization approaches.

The most important thing to remember from all those studies that were done in the past is that if we can tell people how likely it is that an accident will happen, it will help new or inexperienced travelers be more careful on the road. The government will learn what causes accidents, what factors (like weather and transportation) are most impactful in accident-prone areas, and they will be able to help draw connections between the many factors that contribute to accidents, whether they are directly or indirectly involved. If you know which areas are prone to accidents caused by drunk driving, other distractions (such as talking on a phone), aggressive or careless driving, a lack of respect for traffic laws, or driver fatigue, you can share that information with the regional transport office. The RTO may use this data to implement stringent measures, including verifying drivers' licenses, doing breathalyzer tests, or even stationing more traffic cops in certain locations. Our goal is to also provide a hand when it comes to traffic management.

3. EXISTING SYSTEM

According to research by Williams et al. [5], driver experience and age are also important factors in accident rates. In their paper, Suganya, E. and S. Vijayarani [6] examined road accidents in India and contrasted various classification algorithms' performance metrics, including accuracy, error rate, execution time, and decision trees. The algorithms included SVM, Naïve Bayes, KNN, Random Forest, and gradient boosting. Compared to the alternatives, they discovered that KNN performed better.

Disadvantages

1) The system doesn’t have facility to train and test on large number of numbers.

2) The system doesn’t measure an accurate road accident due to poor classification models.

3.1 Proposed System

The suggested solution incorporates an app that can forecast the likelihood of accidents occurring using existing data on road accidents. In order to create a dataset, this data on traffic accidents is pre-processed.

ISSN 2347–3657

Volume 12, Issue 3, 2024

Once the data has been normalized and any null or trash values removed, the next phase in data preparation is feature selection, which involves taking just the most important characteristics from the source dataset and adding them to the final dataset. Afterwards, several data mining methods are applied to the dataset. This dataset undergoes clustering. The clusters are further processed using other algorithms such as Apriori and Support Vector Machines (SVM). Due to the non-disclosured nature of the study's data distribution, we must employ support vector machines (SVMs) to forecast the likelihood of accidents and a priori methods (Apriori) for rule mining, i.e., to produce a set of frequently occurring items using the provided confidence and support values.

The many kinds of roads and weather conditions necessitated the establishment of rules that take into account the myriad of circumstances that might contribute to accidents. For sets of items that occur often, the greater the confidence and support values, the more likely it is that a certain combination of qualities would cause an accident. Take the training dataset as an example; rule mining suggests that, even in perfect conditions, the likelihood of a collision at a crossroads due to overspeeding is significant and might be deadly. Using support vector machine classification, each accident occurrence has been categorized as either high or low risk. The accident dataset is subjected to a number of data mining and exploratory visualization approaches in order to get the interpreted findings.

The Benefits

1) The government may make effective use of these improved models to lower accident rates and establish road safety rules.

2) The model as a whole has been useful in illuminating the permutations of causes that have resulted in catastrophic accidents.

4. OUTPUT SCREENS

User login:

User Profile:

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

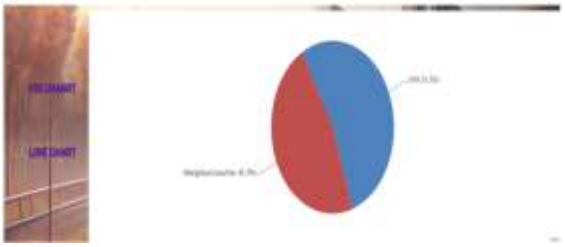
Predict Road Accident:

Admin Login:

Accuracy:

View in Bar Chart:

Accuracy Results In Pie Chart:



View All Road Accident Prediction Status:

Find Road Accident Prediction Type Ratio:

View all remote users:

ISSN 2347–3657

Volume 12, Issue 3, 2024

5. CONCLUSION

Numerous lives may be irrevocably altered by one accident. To slow this rising tide, we must all do our part. Adopting safe driving practices may help make this a reality. Since no two incidents have the same underlying cause, both the road construction authorities and the car industry need to take safety precautions while planning road structures and developing new models of vehicles with reduced mortality rates. We can help these authorities and enterprises by predicting the likelihood of accidents using historical data and observations. With the use of variables including vehicle type, driver age, vehicle age, weather, and road structure, this research was able to successfully develop an app that may aid in the efficient forecast of road accidents. With the use of a dataset for Bangalore and a number of data mining and machine learning methods, this model was able to accurately forecast the risk likelihood of accidents across various places.

Several restrictions that were not considered in this research may be added to the model in future optimizations. The government may make effective use of these improved models to enact regulations aimed at reducing traffic accidents. Making a smartphone app that drivers may use to plan their routes is another potential outcome of this project. In addition to providing instructions, the mapping service may also notify the motorist of the danger likelihood along the selected route. Once this is in place, service provider businesses like Ola, Uber, and others may use it down the road. Better monitoring of accident-prone locations and the provision of emergency services are two more areas where this will be helpful. The model's findings on potential dangers may also inform the development of more effective road safety signage for highways.

6.REFERENCES

1. . F.M.O.I. Forensic Medicine Organization of Iran; Statistical Data, Accidents, online avail-able on: http://www.lmo.ir/?siteid=1&pageid=1347
2. A.T. Kashani et al., “A Data Mining Approach to Identify Key Factors of Traffic Injury Severity”,

ISSN 2347–3657

Volume 12, Issue 3, 2024

PROMETTraffic& Transportation, 23(1), pp. 11-17, 2011.

3. L.Y. Chang, H.W. Wang, "Analysis of traffic injury severity: An application of non-parametric classification tree techniques", Accident Analysis and Prevention, 38(5), pp. 1019-1027, 2006.
4. S. Yau-Ren et al. "The Application of Data Mining Technology to Build a Forecasting Model for Classification of Road Traffic Accidents", Mathematical Problems in Engineering, Volume 2015 (2015), pp. 1-8., 2015. F. Babi and K. Zuskáová • Descriptive and Predictive Mining on Road Accidents Data– 92
5. R. Nayak et al., "Road Crash Proneness Prediction using Data Mining". Ailamaki, Anastasia & AmerYahia , Sihem (Eds.) Proceedings of the 14th International Conference on Extending Database Technology, Association for Computing Machinery (ACM), Upp-sala, Sweden, pp. 521-526, 2011.
6. V. Shankar, J. Milton, F. Mannering, "Modeling accident frequencies as zero-altered probability processes: An empirical inquiry", Accident Analysis & Prevention, 29(6), pp. 829-837, 1997.
7. A Araar et al., "Mining Road traffic accident data to improve safety in Dubai", Journal of Theoretical and Applied Information Technology, 47(3), pp. 911-927, 2013.
8. S. Vigneswaran et al., "Efficient Analysis of Traffic Accident Using Mining Techniques", International Journal of Software and Hardware Research in Engineering, Vol. 2, No. 3, 2014, pp. 110- 118, 2014.
9. L. Martin et al. "Using data mining techniques to road safety improvement in Spanish roads", XI Congreso de Ingeniería del Transporte (CIT 2014), Procedia - Social and Behavioral Sciences 160 (2014), pp. 607– 614, 2014.
10. P. Flach et al., "On the road to knowledge: Mining 21 years of UK traffic accident reports", Data Mining and Decision Support: Aspects of Integration and Collaboration, Springer, pp. 143-155, 2003.
11. H. Zhang et al., "In-Memory Big Data Management and Processing: A

ISSN 2347–3657

Volume 12, Issue 3, 2024

Survey", IEEE Transactions on Knowledge and Data Engineering, Vol. 27, No. 7, pp. 1920–1948, 2015.

12. J. Hipp, U. Güntzer, G. Nakhaeizadeh, "Algorithms for Association Rule Mining &Mdash; a General Survey and Comparison", SIGKDD Explor Newsl 2, pp. 58–64, 2000.
13. A.T. Kashani et al., "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", PROMETTraffic& Transportation, 23(1), pp. 11-17, 2011.
14. P.J. Ossenbruggen, J. Pendharkar et al., "Roadway safety in rural and small urbanized areas", Accidents Analysis & Prevention, 33(4), pp. 485-498, 2001.
15. R. Agrawal, T. Imieliski, A. Swami, "Mining Association Rules Between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, pp. 207–216, 1993.
16. R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules in Large Data-bases", Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 487-499, 1994.
17. L. Breiman, "Random Forests", Machine Learning, Vol. 45, pp. 5-32, 2001.