Advances in Nonlinear Variational Inequalities
ISSN: 1092-910X
Vol 28 No. 2s (2025)

Using Mining Techniques to Develop a Road Accident Prediction Model

Dr. Varaprasada Rao P[1], KVSL Harika[2] ,Kakumanu Venkata Vamshi Krishna[3]

[1]Professor in CSE, Gokaraju Rangaraju Institute of Engineering and Technology,prasadp.griet@gmail.com,

[2]Assistant professor, Gokaraju Rangaraju Institute of Engineering and Technology,harika1742@grietcollege.com

[3]Student, Gokaraju Rangaraju Institute of Engineering and Technology , kakumanuvamshi1999@gmail.com

Article History:

Received: *13-09-2024*

Revised: *28-10-2024*

Accepted: *07-11-2024*

Abstract:

A daily tally of accidents is climbing at a frightening pace, paralleling the exponential growth in the volume of automobiles on the street. Given the current high rate of traffic-related events and fatalities, it is crucial for the transportation department to have the capability to envisage the frequency of traffic accidents within a certain timeframe in order to make data-driven choices. In this case, it would be wise to study accident rates in order to utilize that information to develop strategies for lowering them. Although most accidents include some degree of uncertainty, there does seem to be some pattern to the incidents that happen in the same place over time. By utilizing this pattern, ML designs can be constructed for predicting unfortunate incidents and arrive at informed estimations regarding the frequency of mishaps occurring in a specific area. The present study explored the relationships among roadway circumstances, environmental factors, and the number of crashes. A predictive model for accident prediction has been developed by employing the AdaBoost machine learning algorithm. . Various online datasets covering traffic accidents in Bangalore from 2014 to 2017 were used for this research. Public works agencies, contractors, and other sectors of the automotive industry are just a few of the many potential beneficiaries of this study's findings, which may inform more accurate road and vehicle design.

Keywords: Accidents, Regularity, Occurrence, Transportation department, Vehicles based, Mining techniques, Alarming rate, Number of accidents, Road accidents, Traffic accidents.

.

1. Introduction

Term "machine learning" refers to an umbrella term for a set of algorithms implemented in computers that, without human intervention, may "learn" from past mistakes and better themselves. ML, a component of artificial intelligence (AI), utilizes empirical approaches to analyze facts and provide estimates concerning potential results from which humans may draw useful conclusions. A major step forward is the concept that a computer may independently learn from examples (data) to generate reliable outcomes. There is a tight relationship between data mining, machine learning, and Bayesian predictive modeling. Data is fed into the machine using an algorithm, which then generates output. Giving a suggestion is one common machine learning assignment. Every movie or TV show that Netflix suggests to its subscribers is based on their viewing habits and preferences. When it comes to personalized recommendations, tech companies are using unsupervised data mining to make things better for users. ML has many other applications as well, including identification of fraudulent activity, maximizing the efficiency of portfolios, automation of chores, predictive upkeep, and many more.

1.1 ML vs. Conventional Computing

There is an important distinction between conventional

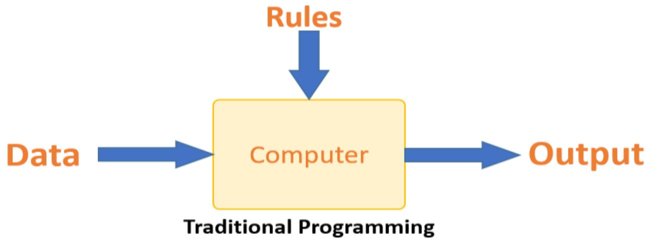


Fig. 1.1. Traditional Programming

computing and ML. Typically, while developing applications, the developer would contact a specialist in the relevant field before coding every single specification. The system will carry out a result of the conceptual assertion that each criterion is founded on. Additional rules must be drafted as the system's complexity increases. Keeping it up might become an impossible task in no time.

This problem is meant to be solved via ML. After figuring out the correlation between the input and output data, the computer constructs a rule. Every time fresh data comes in, the programmers won't have to design new rules. Eventually, the effectiveness of the techniques is enhanced as they adjust to fresh information and lessons learned.

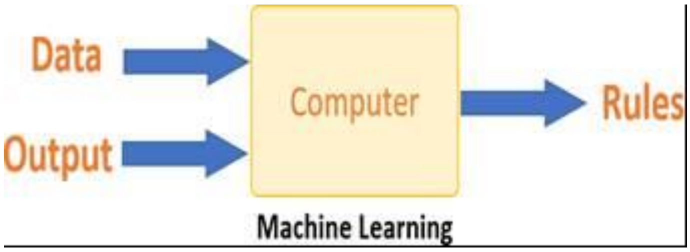


Fig. 1.2. Machine Learning

The central processing unit (CPU) in ML is the learning process. The machine's learning process is somewhat comparable to that of a person. People gain knowledge via doing. Our ability to foretell outcomes improves as our knowledge grows. The probability of being successful declines in an unforeseen circumstance compared to a recognized one, according to analysis. All machines learn the same way. The system needs to observe an example in order to provide a precise forecast. The machine can determine the result when we provide it with a comparable example. Machines aren't any better at prediction than humans when presented with novel examples. Understanding and prediction are at the heart of ML. The computer learns, first and foremost, by seeing patterns. The data allowed for this finding to be made. Data scientists play an essential role in making informed decisions about which pieces of information to feed into machines. Feature vectors are collections of characteristics used to solve problems. A vector of parameters is like a data filter that helps you solve the issue in question.

1.2. How does Machine Learning Work?

A set of sophisticated algorithms allows the computer to abstract this finding into a representation while simultaneously simplifying the underlying facts. To characterize the evidence and consolidate it into a hypothesis, the learning process is used.

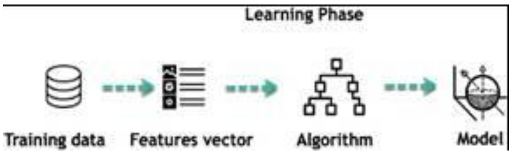


Fig. 1.3.Learning Phase

As an illustration, the algorithm is attempting to deduce how someone's earnings relate to their propensity to dine at a fine dining establishment. As it turns out, the model reveals a favorable correlation involving pay and dining at upscale restaurants.

1.3. Inferring

Upon completion of the strategy, its efficacy may be evaluated using novel evidence. A characteristics vector is created from the updated data, which is subsequently inputted into the ML algorithm in order to provide a forecast. The most appealing aspect of ML is this whole thing. Retraining the model or revising the rules is unnecessary. Inferring from fresh data is possible with the help of the developed algorithm.

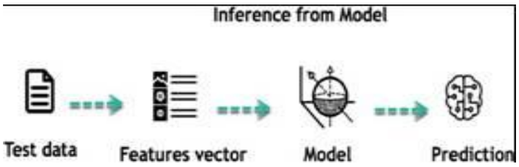


Fig. 1.4. Inference from Model

Here is a rundown of the typical lifecycle of a ML training:To produce a prediction using the model, follow these steps: (1) formulate a query; (2) gather evidence; (3) interpret the findings; (4) train the algorithm; (5) evaluate the approach; (6) gather evaluations; (7) revise the technique; and (7) repeat steps 4–7 till outcomes have satisfactory results.

2 LITERATURE REVIEW

Studies evidence on train accidents analyzed through text processing

According to the writers Williams T ET AL. [6], the National Transportation Safety Board in the United States and the Transportation Safety Board of Canada both issue reports regarding significant railroad incidents. These reports are compiled by the respective organizations. The content from these accident reports was evaluated using text mining methods such as probabilistic topic modeling and k-means clustering in order to find the recurring themes that were present in significant train incidents. As a result of these assessments, it has been shown that the accidents that occurred on railroads may be effectively classified into several categories. An further conclusion that can be drawn from the findings is that the sorts of accidents that occur often include track flaws, wheel defects, grade crossing accidents, and switching accidents. The result that accidents involving bridges are found to be more prevalent in the Canadian reports has been identified as a significant distinction between the data from the United States and those from Canada.

Researchers Chen et al. [7] have used a wide range of methodologies in order to explore factors that are connected to one another and develop efficient prediction models in order to reduce the amount of damage that is caused by traffic accidents. Researchers employed two nonparametric and one statistical data mining technique—logistic regression (LR), classification and regression tree (CART), and random forest (RF)—in their study to compare the predictive power of each method, identify the key essential variables—identified by LR and CART or RF, respectively—that are highly correlated with the severity of traffic accidents, and identify the variables that have a significant positive influence on prediction performance. Specificity, sensitivity, and accuracy are the three main characteristics of prediction performance assessment that are performed in order to identify which integrated model is the most successful. In their investigation, the researchers determined that the accuracy and specificity of RF are superior to those of LR and CART when the 15 original variables are input, and when just a few unique significant factors identified by LR or critical variables discovered by CART are input into RF. This was the conclusion reached by the researchers. Thus, the researchers came to the conclusion that radio frequency (RF) is better to both of the technologies.

The authors Suganya, E. et al. [8] state that the classification is a model finding procedure that is used for the purpose of segmenting the data into several classes based on a given set of criteria. Through the use of several classification algorithms, this study examines the data set pertaining to road accidents in India. These algorithms comprises of linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, RF etc. To assess efficiency model metrics involving accuracy, error rate, and time required for processing are used. Using the data mining program R, this analysis was carried out. There are alternative algorithms, however KNN has a superior performance than the others.

The predictive model for incident occurrences at steel plants in India was developed by Sarkar S et al. [9], as stated by the authors. It is generally agreed that the steel industry is one of the economic sectors that has a greater number of incidents. During the course of their workday, employees in this sector are forced to contend with a broad range of potential dangers. The database that is kept in industry is thus different in terms of the sorts of data that indicate the nature of accidents, the causes of accidents, the date and time stamp, and other similar information. Based on free-text data or narratives that have been stored in the database from prior events report, the purpose of this research is to provide a predictive solution for the occurrence of accidents in the steel sector. The results of a text extraction approach have been passed to three different classification algorithms, which include SVM, RF, and Maximum Entropy. In order to determine which model is the most suitable for predicting the incidence of injuries in workplaces, these classifiers are now being evaluated on an experimental basis. The record set used consists of 9488 records of workplace incidents that happened at an integrated steel production facility in India from 2010 to 2013. In addition, the causes of injuries may be predicted with the assistance of the same three classifiers that will be discussed before. When compared to other classifiers, the model that were constructed using methodologies Maximum Entropy and Random Forest have been proven to be the most effective in pertaining to AUC-ROC value in binary and multi- class machine learning models, correspondingly.

Mohamed K. Nour, one of the creators of the Road Traffic Accidents Injury Data Analytics [10], makes the following statement: Road safety experts who have been working on road accident data have seen success in the study of road traffic accidents via the use of data analytical tools.However, there has been very little progress achieved in the prediction of road injuries. The purpose of this research is to demonstrate how sophisticated data analytics techniques may be used to forecast injury severity levels and assess their effectiveness. The research makes use of methodologies that are based on predictive modeling in order to identify risk and important elements that contribute to the severity of accidents. This analysis makes use of data that is readily available to the public from the United Kingdom Department of Transport and spans the years 2005 through 2019. In this study, a technique is presented that is sufficiently broad to be applicable to a variety of data sets originating from various nations. Based on the findings, it was determined that tree-based approaches, such as XGBoost, achieved better results than regression-based techniques, such as ANN. In addition to the article, it finds intriguing linkages and acknowledges challenges that are associated with the quality of those interactions.The writers Annie Racheal Rajkumar **ET AL.** [11] state that the prediction of the severity of road accidents is as follows: Through the use of the Machine Learning Algorithm, it has been shown that injuries resulting from traffic accidents are among the most prominent causes of mortality, aside from health-related disorders. According to the World Health Organization, during the year 2016, injuries sustained in traffic accidents were responsible for an estimated 1.35 million fatalities around the globe. In other words, one person is murdered every quarter of a minute. Because of this, it is necessary to do research on automobile collisions and the elements that contribute to them, as well as to devise a strategy for lowering the likelihood that such collisions would take place. The investigation of the severity of road accidents was carried out by putting an accident dataset through a number of different machine learning classification algorithms. The goal of this process was to determine which model performed the best in effectively categorizing the accidents into different severity classifications, such as minor, severe, and deadly. It was shown that the maximum accuracy score was achieved with the use of logistic regression for the purpose of performing multi- label classification. An other observation that was made was that the severity of the collision was influenced by a variety of variables, including the volume of automobiles that have taken part, the lighting conditions, and the characteristics of the road.The seriousness of road unfortunate incidents is a worldwide worry, mainly in countries that are still in the process of developing. By gaining an understanding of the primary factors and factors that contribute to traffic accidents, it may be feasible to reduce the severity of these incidents. The findings and the most significant target-specific variables for the severity of road accidents were discovered by the researchers Yassin et al. [12]. Both K-means and random forest (RF) techniques were used in the research project in order to determine the most significant characteristics that have an impact on the occurrence of road accidents. Random forest and K- means are two methods that we use to group datasets that are connected to one another. This allows us to classify the components of road accidents according to the severity variable. When compared to other conventional models, such as Logistic Regression, k Nearest Neighbor, and Support Vector Machine, the K- Means + random forest combination demonstrates superior performance. The author of the research demonstrated that the proposed strategy has a one hundred and eighty-six percent accuracy rate. In addition, the author came to the conclusion that factors such as driving experience and day, lighting circumstances, driver age, and the year the vehicle was in operation were important contributors to serious injury, moderate damage, and severe damage according to this sequence.

The approach used by the authors is superior to the performance of traditional machine learning algorithms and matches the standards of their inquiry. Incorporating k-means clustering into the training set, in addition to classification algorithms and random forest, in order to give severity ratings to the causes of road accidents, yielded findings that were persuasive. According to the findings of the experiment, the addition of a new cluster to the already existing training set leads in a considerable improvement in classification accuracy. The Random Forest algorithm was able to reach an accuracy of 99.86%. This study also provides an explanation of the target- specific influence of the variable. In a nutshell, the purpose of the study was to illustrate the advantages of combining Clustering and Classification in order to improve the accuracy of the model and to identify the major factors that led to this improvement.

In spite of the fact that engineers and other professionals in the automobile industry may do all of the research and development in the world, accidents on the road will still occur. To further our knowledge of the elements that lead to hazardous traffic events, the creation of a prediction system that is capable of automatically identifying the severity of injuries sustained in various kinds of traffic accidents would be of great assistance. Understanding these patterns of behavior and the roadways that they travel might provide valuable information that could be used to enhance policies around road safety. For effective policymaking, it is vital to conduct stringent scientific examinations on the factors that cause accidents and the severity of injuries. A number of algorithms inside the system make use of machine learning in order to establish an assessment of the severity of injuries. The authors P. Chirag et al. [13] provided a classifier for the early forecast of road unfortunate incidents by employing ML strategies such as support vector machines, decision trees, and Random forests. Using the dataset of traffic incidents that was provided by Kaggle, the researchers were able to achieve a success rate of 98% by using the strategy that was given. The researchers sought for indicators that would indicate the significance of road accidents. Some of these criteria were the kind of vehicle that was involved, the code of the device, its location (including latitude and longitude), speed, date, and time.

3 SYSTEM ANALYSIS AND DESIGN.

3.1. FEASIBILITY STUDY

An evaluation of the project's feasibility and the presentation of a research strategy occur at this stage. including a broad concept for the project along with some cost estimates. During the process of system analysis, it is necessary to conduct a feasibility assessment for the proposed system. This is to guarantee that the suggested solution does not impose any excessive strain or hardship on the firm. In order to conduct a feasibility study, it is crucial to have a comprehensive grasp of the key needs for the system. The feasibility study involves three crucial considerations: economic feasibility, technical feasibility, and social feasibility.

3.1.1. Economic feasibility: This research is conducted to assess the financial implications that the system will have on the company. The company's capacity to provide funds for the research and development of the system is restricted. Justification for the expenses is required. Therefore, the system was constructed under the allocated budget due to the use of mostly publicly accessible technology. Only the items that were tailored to specific requirements needed to be bought.

3.1.2. Technical feasibility: This research is conducted to assess the technical feasibility, namely the technical requirements of the system. The development of any system should not place a significant burden on the existing technological resources. This will result in a significant increase in the demand for the available technical resources. This will result in the customer facing significant expectations..

3.1.3. Social feasibility: This element of the research aims

to assess the degree of user acceptability of the system. This encompasses the procedure of instructing the user to use the system proficiently. The user should not see the system as a source of intimidation, but rather acknowledge it as an essential need. The degree of user acceptance is contingent upon the strategies utilized to educate and familiarize the user with the system. In order for him to provide valuable feedback as the ultimate decision-maker, his level of confidence has to be enhanced, enabling him to provide constructive criticism.

The current document does not include any sources. The user is referring to the system's series.

3.2. EXISTING SYSTEM

A method for determining the influence of many variables on the detection and prediction of atmospheric deterioration on a worldwide scale has been developed by Anand, J. V. [14]. The ARIMA framework, R-studio, and Fuzzy C-means clustering were used to construct the approach. Analyzing the impact of various variables on car accidents may also be done using a similar approach. Because each factor contributes to traffic accidents in its own unique way, understanding what causes them is essential. In order to classify road accident data based on the kind of road traffic, Tiwari et al. [15] employed decision tree techniques, Naïve Bayes, Support Vector Machines, K-mode clustering approaches, and self- organizing maps.

–Problems with the existing system: The present system only considered one particular city.

- Performance accuracy dropped.
- Complexity rose.

3.3 PROPOSED SYSTEM

3.3.1 We propose the following system: An app that estimates the chance of traffic accidents using current data is developed and presented in this research. The road accident data must undergo data pre-processing before it can be used as a dataset. Normalizing and removing null and unnecessary values are part of data preparation. The next step is feature selection, which involves picking out relevant characteristics from the source dataset to include in the finished product.

To prepare the dataset that will be used as input for the model, the raw data from road accidents is first preprocessed. The model is trained using the training data that is supplied and then used to predict the possible accident risk for a user-specified location. A graphical representation is then presented to the user based on the statistics that have been gathered.

This research details the steps used to create and test a model for predicting road accidents that takes into account a wide range of possible causes. In particular, the studies only look at a small set of factors, such as road conditions, weather, and accident causes.

3.3.2. Advantages of the proposed system: A visual representation of the features that have historically led to accidents at a certain place is shown in the user interface of the model-based software. The algorithm then uses this data to provide a high or low accident probability prediction for the given area. Thanks to the all- encompassing model, we now understand the myriad of factors that might lead to deadly accident scenarios.Predicting accident severity level is a crucial problem in road risk research and modeling. The suggested method offers a prediction model for determining the severity degree of accidents, hence enabling the prevention of future accidents.

3.4. SYSTEM DESIGN

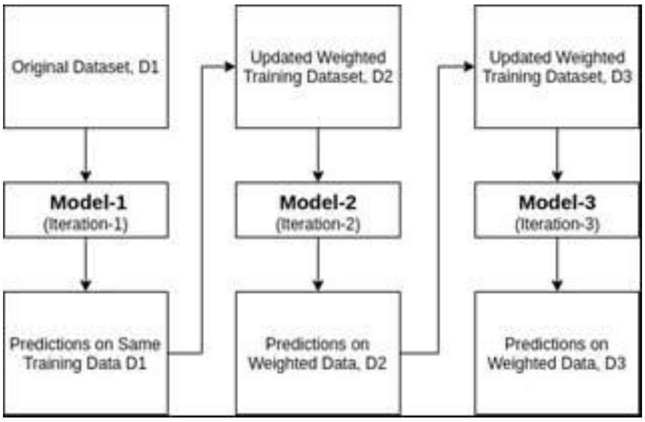


Fig. 3.1. System Architecture

3.4.1. *data flow diagram:* To visually depict the transfer content from one location to another in an action, often a data-containing system, one may use a data-flow diagram. The DFD also specifies the inputs and outputs of the approach for each unit.

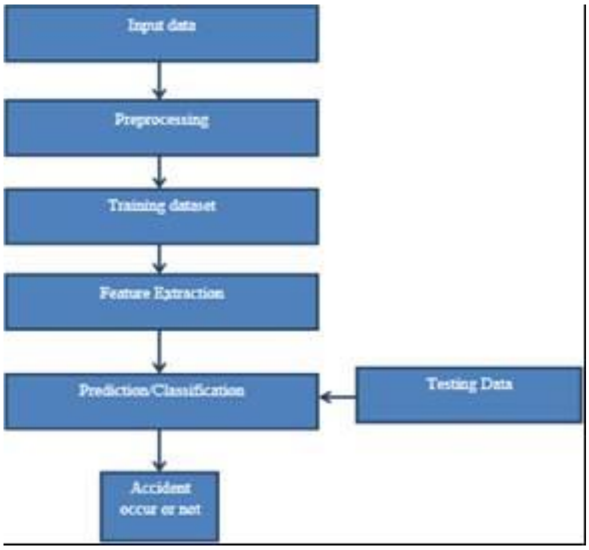


Fig.3.2. Data Flow Diagram

Among the modeling tools that are considered to be among the most important is the data flow diagram (DFD). In order to simulate the components of the system, it is used. The dynamic flow diagram (DFD) illustrates how the information undergoes a series of transformations and how it is transferred across the system. Information flow and the changes that are done to data as it goes from input to output are shown via the use of this graphical approach.

The DFD is sometimes referred to as the bubble chart. When representing a system at any level of abstraction, a

DFD may be used as a representation tool. It is possible to segment DFD into levels that reflect increasing degrees of information flow and functional depth via the use of partitioning.

3.4.2. *Diagrams created using the UML:* Unified Modeling Language is what the acronym UML stands for. The Unified Modeling Language (UML) is a general- purpose modeling language that has been standardized and is used in the area of object-oriented software engineering. This standard was developed by the Object Management Group, which is also responsible for its management.

A single language for the creation of models of object- oriented computer software is what the Unified Modeling Language (UML) aims to eventually become. A Meta- model and a notation are the two primary components that make up the Unified Modeling Language (UML) in its present configuration. In the future, the Unified Modeling Language (UML) could also include a method or process of some kind, or it might be related with it.

The Unified Modeling Language is a standard language that may be used for business modeling and other non-software systems, as well as for describing, visualizing, constructing, and documenting the artifacts of software application systems.

The Unified Modeling Language (UML) provides a compilation of the most effective engineering approaches that have been shown to be effective in the modeling of big and complex systems.

When it comes to the process of producing software and developing object-oriented software, the

Unified Modeling Language (UML) is an extremely significant component. Graphical notations are the primary means by which the Unified Modeling Language (UML) expresses the design of software projects.

4 RESEARCH METHODOLOGY.

4.1. MODULES:

- Obtaining Data,
- Creating a Dataset,
- Cleaning Up Data,
- Choosing a Model,
- Running the Model,
- Evaluating Results,
- Making Predictions, and Finally,
- Saving the Trained Model

4.2. MODULES DESCRIPTION:

The gathering of data is the first and most important stage in developing a machine learning

4.2.1 Model: The success of the model is dependent on this crucial stage; the more and better data we collect, the better the model will be. Many methods exist for gathering this information, such as manual interventions, site scraping, and so on. Using Data Mining Techniques, A Model for Predicting Road Accidents

4.2.2 Dataset: The collection contains information from 576 unique individuals. The following is a description of each of the fifteen columns that make up the dataset.

- States/UTs : States and union territories of India
- Junction: Types of junctions road
- Vehicle Age: In year
- Human Age and Sex: human age and male / female
- person without safety precautions
- area: Types of area in India
- type of place : Urban or Rural
- load of vehicle: Types of load of vehicle
- traffic rules violation: Types traffic rules violations
- weather: weather condition
- vehicle type and sex: Types of vehicle and male / Female

- type of road
- license: License Valid Permanent/Without License/Learner's License
- time
- accident occurrence: yes or no

4.2.3 *Preparation of the Data:* and Maintain control of the data and get it ready for training. That which may need cleaning (such as removing duplicates, correcting mistakes, dealing with missing values, normalization, and data type conversions, etc.) should be cleaned. By randomizing the data, we are able to eliminate the influence of the specific sequence in which we gathered and/or otherwise prepared our data. Data visualization may be used to assist in the identification of meaningful correlations between variables or class imbalances (beware of bias!), or other exploratory analysis can be carried out. Sets for training and assessment should be separated.

4.3 Model Selection: Because we were able to achieve an accuracy of 94.7% on the test set by using the decision tree AdaBoost-Classifier method, we decided to put this technique into practice. AdaBoostClassifier: Ada-boost, also known as Adaptive Boosting, represents the sort of ensemble machine learning algorithm that was designed

by the researchers Yoav Freund and Robert Schapire in an academic paper published in 1996. Combining a number of different classifiers results in an improvement in the accuracy of the classifiers. One example of an iterative ensemble algorithm is AdaBoost. By merging a number of classifiers that perform badly, the Ada-Boost classifier is able to construct a powerful classifier. This allows for the creation of a strong classifier that is reliable and accurate. The central concept involved in Adaboost is to train the model in each iteration and adjust the outcome values of the classifiers in such a way that it guarantees correct predictions of observations that are not typical. It is possible to utilize any machine learning algorithm as a basic classifier provided that the algorithm accepts weights assigned to the training set. Two requirements need to be satisfied by Adaboost: It is recommended that the classifier be taught by interactive training on a variety of weighted training instances.The AdaBoost classifier is educated in an iterative style by selecting the input data set based on one's ability to accurately anticipate the results of the most recent training.In order to ensure that the observations that were incorrectly categorized have a high likelihood of categorization in the subsequent iteration, it gives these observations a larger weight than they would otherwise have.Additionally, it distributes the weight to the trained classifier in each iteration using the accuracy of the classifier as the basis for the assignment of weight. A higher weight will be given to the classifier that is more accurate.This algorithm will continue to loop until the whole training data set fits without any errors or until the maximum number of estimators meets the requirements that have been established.

To classify, perform a "vote" across all of the learning algorithms you built.

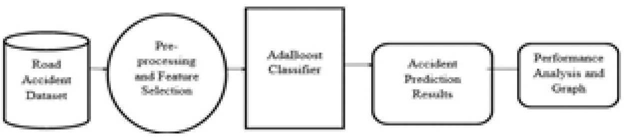


Fig. 4.1. Strategy of AdaBoost Algorithm

4.4. *Analyze and Prediction:* In the actual dataset, we chose only 14 features :

0 States/UTs : States and union territories of India

1 junction : Types of junctions' road

2 vehicle age : In year

3 human age and sex : human age and male / female

4 person without safety precautions

5 area : Types of area in India

6 type of place : Urban or Rural

7 load of vehicle : Types of load of vehicle

8 traffic rules violation : Types traffic rules violations

9. weather : weather condition

10 .vehicle type and sex : Types of vehicle and male / Female type of road

11. License: License Valid Permanent/Without License/Learner's License

12.Accident occurrence : yes or no

4.5 **RESULT**:

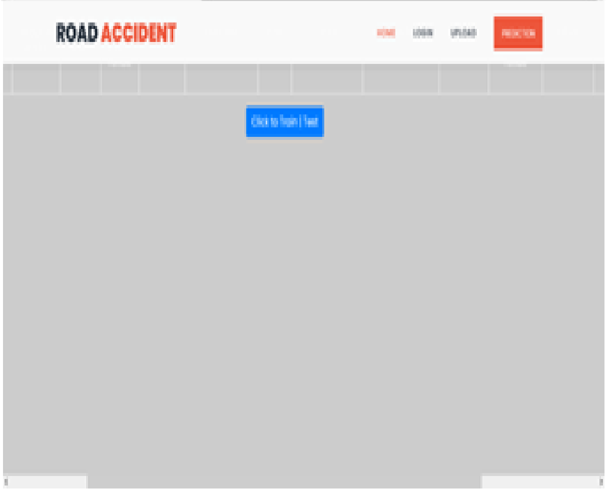
The model exhibited the accuracy of 94.80% on test set. Storing the Trained Model: When you are sufficiently satisfied in the performance of your trained and tested model and want to use it in a production setting, the first step is to store it as a .h5 or .pkl file using a library such as pickle.

Ensure that you have the pickle module installed in your environment. Next, we will import the module and save the model as a .pkl file.

SCREEN SHOTS


ROAD ACCIDENT


ROAD ACCIDENT
A ROAD ACCIDENT PREDICTION
Login


ROAD ACCIDENT


A ROAD ACCIDENT PREDICTION
Road Accident Prediction
Prediction is:


ROAD ACCIDENT


ROAD ACCIDENT

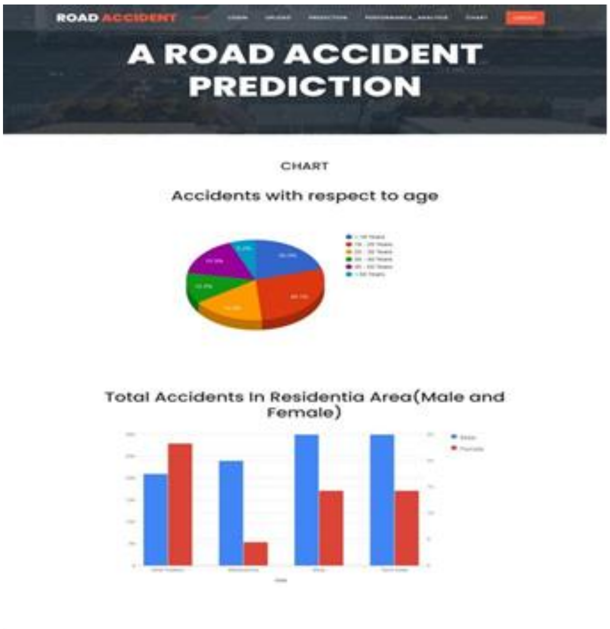

A ROAD ACCIDENT PREDICTION
Road Accident Prediction
Prediction is:


ROAD ACCIDENT
preview


A ROAD ACCIDENT PREDICTION
Road Accident Prediction
Prediction is:Yes, There is a Chance Of Road Accident! Be Careful.


A ROAD ACCIDENT PREDICTION
Road Accident Prediction
Prediction is:


ROAD ACCIDENT
A ROAD ACCIDENT PREDICTION
PERFORMANCE ANALYSIS
Precision and recall
Recall Precision
No Road Accident(0) 0.98 0.99
Road Accident(1) 0.95 0.97
Confusion Matrix


ROAD ACCIDENT
A ROAD ACCIDENT PREDICTION
CHART
Accidents with respect to age
Total Accidents In Residentia Area(Male and Female)

5 CONCLUSIONS

Numerous lives may be irrevocably altered by one accident. To slow this rising tide, we must all do our part. To a certain degree, this may be accomplished by implementing safe driving practices. Since no two incidents have the same underlying cause, both the road development authority and the car industry need to take precautions while planning road structures and developing new models of vehicles to reduce the likelihood of fatalities. Predicting the likelihood of an accident using historical data and observations is something we can do, and it can help these authorities and sectors. With the use of variables including category of automobile, age of motorist, year of automobile, climate, route layout, and so on, this research was able to successfully develop an app that may aid in the efficient prediction of road accidents. This model has been applied to a dataset for India using various data mining and machine learning methods. It has been used to accurately forecast the risk likelihood of accidents across different places.Future optimizations of the model may include other limitations that were not considered in this work. The government may make good use of these improved models to enact legislation that would make roads safer and to decrease the number of accidents that occur on them. Creating a smartphone app to assist drivers in route selection is another aspect of this project's scope. In addition to providing instructions, the mapping service may also notify the motorist of the danger probability 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. Using the data collected from this model, we may improve road safety signage and place them along highways.

REFERENCES

[1] S. Das and N. Roy, ―Applications of Artificial Intelligence in Machine Learning : Review and Prospect,‖ vol. 115, no. 9, pp. 31–41, 2015.

[2] N. J. Nilsson, ―INTRODUCTION TO MACHINE LEARNING AN EARLY DRAFT OF A PROPOSED TEXTBOOK Department of Computer Science,‖ 2005, [Online]. Available: https://github.com/ec2ainun/books-ML-and- DL/blob/master/introduction-to-machine- learning BY Nils J. Nilsson.pdf

[3] Duda, ―Duda2001 - Pattern classification.pdf.‖ [Online]. Available: https://github.com/rohinarora/EECE5644- Machine_Learning/blob/master/Richard O. Duda%2C Peter E. Hart%2C David G. Stork -Pattern classification (2001%2C Wiley).pdf

[4] L. Vanneschi and S. Silva, *Bayesian Learning*. 2023. doi: 10.1007/978-3-031-17922-8_9.

[5] A. M. Turing, ―Computing Machinery and Intelligence,‖ *Brain Physiol. Psychol.*, pp. 212– 240, 2023.

[6] B. F. Trefor Williams, John Betak, ―Text mining analysis of railroad accident investigation reports,‖ 2016. doi: https://doi.org/10.1115/JRC2016-5757.

[7] M. M. Chen and M. C. Chen, ―Modeling road accident severity with comparisons of logistic regression, decision tree and random forest,‖ *Inf.*, vol. 11, no. 5, 2020, doi: 10.3390/INFO11050270.

[8] E. Suganya; S. Vijayarani, ―Analysis of road accidents in India using data mining classification algorithms,‖ *Int. Conf. Inven. Comput. Informatics,* 2017, doi: 10.1109/ICICI.2017.8365315.

[9] S. Sarkar, V. Pateshwari, and J. Maiti,―Predictive model for incident occurrences in steel plant in India,‖ *8th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2017,* no. December, 2017, doi: 10.1109/ICCCNT.2017.8204077.

[10] M. K. Nour, A. Naseer, B. Alkazemi, and M. A.Jamil, ―Road Traffic Accidents Injury Data Analytics,‖

vol. 11, no. 12, pp. 762–770, 2020.

[11] A. M. P. Annie Racheal Rajkumar, Srihari Prabhakar, ―Prediction of Road Accident Severity Using Machine Learning Algorithm,‖ *Int. J. Adv. Sci. Technol.*, vol. 29, no. 06, 2020, [Online]. Available: http://sersc.org/journals/index.php/IJAST/ar ticle/view/11302

[12] S. S. Yassin and Pooja, ―Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach,‖ *SN Appl. Sci.*, vol. 2, no. 9, pp. 1–13, 2020, doi: 10.1007/s42452-020-3125-1.

[13] P. Chirag and M. Supreetha, ―Road Accident Prediction and Classification using Machine Learning,‖ *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Sect. Int. Conf.*, no. 2, pp. 48–59, 2022, doi: 10.1109/MysuruCon55714.2022.9972671.

[14] A. J.V., ―A Methodology of Atmospheric Deterioration Forecasting and Evaluation through Data Mining and Business Intelligence,‖ *J. Ubiquitous Comput. Commun. Technol.*, vol. 2, no. 2, pp. 79–87, 2020, doi: 10.36548/jucct.2020.2.003.

[15] P. Tiwari, S. Kumar, and D. Kalitin, ―Road-user specific analysis of traffic accident using data mining techniques,‖ *Commun. Comput. Inf. Sci.*, vol. 776, no. October, pp. 398–410, 2017, doi: 10.1007/978-981-10-6430-2_31.