IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

Real Time Crash Prediction using Machine Learning Algorithm

P. Manikanda Prabu[1], S. Venkatesh[2], M. Sayed Ibrahim[3], S. Sathiyan[4], B. S. Viswanathan[5]
Assistant Professor, Department of Science and Computer Engineering[1]
Students, Department of Science and Computer Engineering[2,3,4,5]
Anjalai Ammal Mahalingam Engineering College, Kovilvenni, Tiruvarur, Tamil Nadu, India

Abstract: *Road traffic accidents (RTAs) have a significant impact on individuals, their families and the nation. Without knowledgeable action, road traffic injuries are said to be the seventh leading cause of death. With the exponentially increasing number of vehicles, road safety is a matter of huge concern. Road accidents kill 1.2 million people every year. It causes loss of lives and economical damage, due to which is a serious concern which needs to be solved We have used machine learning algorithms to predict the severity of an accident occurring at a particular location and time. Factors like speed limit, age, weather, vehicle type, light conditions and day of the week have been used as parameters for training the model. We have created a web app for user input and output display and a notification is sent to the police to take preventive measures. The model will run the with the input data and predicts the severity of an accident occurring at the respective location of the user. This model will play an important role in planning and management of traffic and would help us reduce a lot of road accidents in the future.*

Keywords: Accident prediction, Data mining, Adaptive Booster algorithm, Data Analysis

I. INTRODUCTION

Road traffic injury is a leading cause of death globally for children and was the leading cause of death for young people aged 15–29 in 2010. Without urgent action, it is forecast that road traffic injury will be the 7th leading cause of death for all by 2030.Death and serious injury from road crashes is preventable if crash energies are managed so that they do not exceed human tolerances for serious and fatal injury and this is accomplished with effective, results-focused and resourced road safety management. The Safe System goals and strategies focus on providing a road traffic system free from death and serious injury. The Safe System guides the planning, design, management, operation and use of the road traffic system so as to provide safety in spite of human fallibility.

Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its algorithm, rather than having human programmers specify every needed step. The discipline of machine learning employs various approaches to teach computers to accomplish tasks where no fully satisfactory algorithm is available. In cases where vast numbers of potential answers exist, one approach is to label some of the correct answers as valid. This can then be used as training data for the computer to improve the algorithm(s) it uses to determine correct answers. For example, to train a system for the task of digital character recognition, the MNIST data set of handwritten digits has often been used.

II. LITERATURE REVIEW

Text mining analysis of railroad accident investigation reports Williams T, Betak J, Findley B.The National Transportation Safety Board in the United States and the Transportation Safety Board of Canada publish reports about major railroad accidents. The text from these accident reports was analyzed using the text mining techniques of probabilistic topic modelling and k-means clustering to identify the recurring themes in major railroad accidents. The output from these analyses indicates that the railroad accidents can be successfully grouped into different topics. The

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

output also suggests that recurring accident types are track defects, wheel defects, grade crossing accidents, and switching accidents.

Analysis of road accidents in India using data mining classification algorithms**.** Suganya, E. and S. Vijayarani Classification is a model finding process which is used for segmenting the data into different classes based on some constraints. This work analyzes the road accidents in India data set using classification algorithms namely linear regression, logistic regression, decision tree, SVM, Naïve Bayes, KNN, Random Forest and gradient boosting algorithm. Performance measures used are accuracy, error rate and execution time. This analysis is done in R data mining tool. The performance of KNN is better than other algorithms. Predictive model for incident occurrences in steel plants in India. Sarkar S, Pateshwari V, Maiti J.Steel industry is considered to be an economic sector with higher number of accidents. Workers in this industry are exposed to a wide variety of hazards during working hours. Thus, the database maintained in the industry varies in terms of the types of data indicating the nature of accidents, causes of accidents, date and time-stamp etc. The objective of this study is to give a predictive solution to accident occurrences in a steel industry based on free-text data or narratives logged in the database from previous incident reports.

Road Traffic Accidents Injury Data Analytics ,Mohamed K Nour. Road safety researchers working on road accident data have witnessed success in road traffic accident analysis through the application of data analytic techniques, though, little progress was made in the prediction of road injury. This paper applies advanced data analytics methods to predict injury severity levels and evaluates their performance. The study uses predictive modelling techniques to identify risk and key factors that contribute to accident severity. The study uses publicly available data from the UK department of transport that covers the period from 2005 to 2019. The paper presents an approach which is general enough that can be applied to different data sets from other countries. The results identified that tree-based techniques such as XGBoost outperform regression-based ones, such as ANN. In addition, the paper identifies interesting relationships and acknowledged issues related to the quality of data.

Prediction of Road Accident Severity Using Machine Learning Algorithm, Annie Racheal Rajkumar, Srihari Prabhakar, A Meena Priyadharsini. Injuries due to road accidents are one of the most prevalent causes of death apart from health-related issues. The World Health Organization states that road traffic injuries caused an estimated 1.35 million deaths worldwide in the year 2016. That is, a person is killed every 25 seconds. This calls for the need to analyse road accidents and the factors affecting them and come up with a method to reduce the probability of their occurrence. The analysis of road accident severity was done by running an accident dataset through several machine learning classification algorithms to see which model performed the best in classifying the accidents into severity classes such as slight, severe and fatal. It was observed that logistic regression to perform multilabel classification gave the highest accuracy score. It was also observed that factors such as the number of vehicles, lighting conditions and road features played a role in determining the severity of the accident.

III. METHODOLOGY

There are five methodologies Namely they are Data Collection, Preprocessing, Model Selection, Feature Extraction, Prediction.

Data Collection

This is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get, the better our model will perform.There are several techniques to collect the data, like web scraping, manual interventions, etc.A Road Accident Prediction Model Using Data Mining Techniques

Preprocessing

Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data. Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis. Split into training and evaluation sets features are extracted using Grey Level Co-occurrence Matrix (GLCM) as it is robust method with high performance.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

This texture feature extraction method is very competitive as using smaller number of grey levels shrinks the size of GLCM which reduces the computational cost of the algorithm and also maintains the high classification rates. These features are used to differentiate between normal and abnormal brain.

Model Selection

Model selection is the process of selecting one final machine learning model from among a collection of candidate machine learning models for a training dataset.

Feature Extraction:

Working with a large number of features may affect the performance because training time increases exponentially with the number of features. Even, it has also the risk of overfitting with the increasing number of features. So, for getting a more accurate prediction, feature selection is a critical factor here.

Prediction:

In the actual dataset, we chose only 14 features:

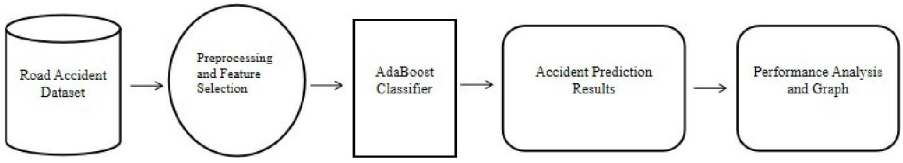
1. States/UTs : States and union territories of India
2. JUNCTION : Types of junctions road
3. VEHICLE AGE : In year
4. HUMAN AGE AND SEX : human age and male / female
5. PERSON WITHOUT SAFETY PRECAUTIONS
6. AREA : Types of area in india
7. TYPE OF PLACE : Urban or Rural
8. LOAD OF VEHICLE : Types of load of vehicle
9. TRAFFIC RULES VIOLATION : Types traffic rules violations
10. WEATHER : weather condition
11. VEHICLE TYPE AND SEX : Types of vehicle and male / Female
12. TYPE OF ROAD LICENSE : License Valid Permanen/Without Licence/Learner's Licence
13. TIME ACCIDENT OCCURRENCE : yes or no

IV. ALGORITHM

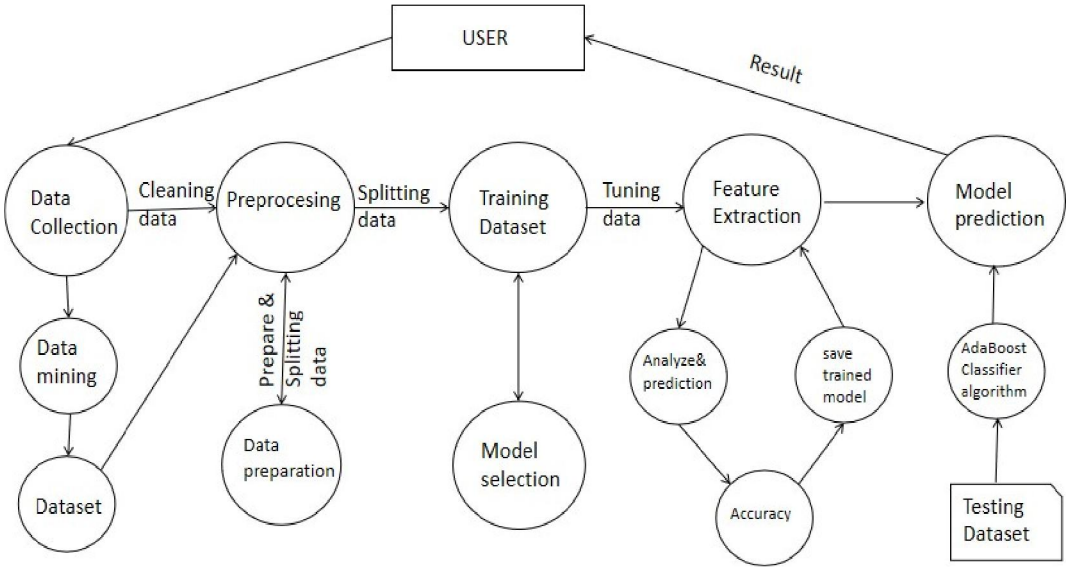
Ada-boost or Adaptive Boosting is one of ensemble boosting classifier proposed by Yoav Freund and Robert Schapire in 1996. It combines multiple classifiers to increase the accuracy of classifiers. AdaBoost is an iterative ensemble method. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that you will get high accuracy strong classifier. The basic concept behind Adaboost is to set the weights of classifiers and training the data sample in each iteration such that it ensures the accurate predictions of unusual observations. Any machine learning algorithm can be used as base classifier if it accepts weights on the training set. Adaboost should meet two conditions: The classifier should be trained interactively on various weighed training examples. In each iteration, it tries to provide an excellent fit for these examples by minimizing training error.It works in the following steps: Initially, Adaboost selects a training subset randomly. It iteratively trains the AdaBoost machine learning model by selecting the training set based on the accurate prediction of the last training. It assigns the higher weight to wrong classified observations so that in the next iteration these observations will get the high probability for classification. Also, It assigns the weight to the trained classifier in each iteration according to the accuracy of the classifier. The more accurate classifier will get high weight. This process iterate until the complete training data fits without any error or until reached to the specified maximum number of estimators.To classify, perform a "vote" across all of the learning algorithms you built.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

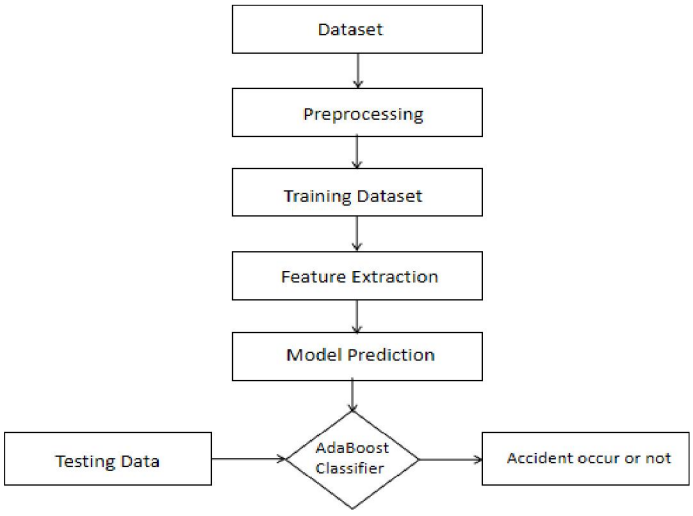
V. SYSTEM ARCHIETECTURE



VI. DATAFLOW DIAGRAM



VII. FLOWCHART



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

VIII. FINAL PROTOTYPE

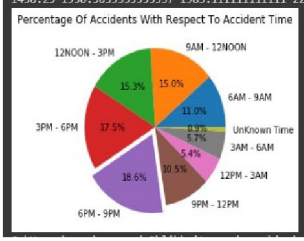
```
#Plotting a Pie Chart

sizes = [mean_EarlyMorning, mean_Morning, mean_Noon, mean_MidDay, mean_Evening, mean_Night, mean_MidNight, mean_VeryEarlyMorning,mean_UnKnownTime]
labels = ['6AM - 9AM','9AM - 12NOON','12NOON - 3PM','3PM - 6PM','6PM - 9PM','9PM - 12PM','12PM - 3AM','3AM - 6AM','UnKnown Time']
explode = (0, 0, 0, 0, 0.1, 0, 0, 0, 0)
plot.pie(sizes, labels = labels, explode = explode, autopct = '%.1f%%')
plot.title("Percentage Of Accidents With Respect To Accident Time")
plot.show()

#Plotting the Total Number Of Accidents With Respect To Accident Time In Increasing Order

data_AccidentTime['Total Accidents'] = data[['Time 06-0900hrs - (Day)',"09-1200hrs - (Day)",'12-1500hrs - (Day)','15-1800hrs - (Day)','18-2100hrs - (Night)','21-2400hrs - (Night)',
sorted_AccidentTimeTotalAccidents = data_AccidentTime.sort_values('Total Accidents')
sorted_AccidentTimeTotalAccidents['Total Accidents'].plot.bar(figsize=(15,5))
plot.ylabel('Number of Accidents')
plot.title("Total Accidents With Respect To Accident Time")
```

```
1430.25 1950.3055555555557 1983.111111111111 2267.1944444444443 2416.277777777778 1365.611111111111 705.75 738.0833333333334 116.86111111111111
```



IX. RESULT

Our Dataset contains several types of dataset values based on the accident possibilities and collected from different online resources. Kaggel contains real cases of refined accidents datasets . In this work, acuuracy of predicting the accident is high and it is performed by using convolution neural network. Simulation is performed by using python language. The accuracy is calculated and compared with the all-other state of arts methods. The training accuracy, validation accuracy and validation loss are calculated to find the maximum accuracy of proposed accident predction.

```
#Use Label Encoder to convert the categorial data into Numerical data for Prediction
labelencoder = LabelEncoder()
columnsToEncode = list(accidentData.select_dtypes(include=['category','object']))
for feature in columnsToEncode:
    accidentData[feature] = labelencoder.fit_transform(accidentDataDummy[feature])

#Split the dataset into Training and Testing Data
target = pd.Series(labelencoder.fit_transform(target))
accidentData_train, accidentData_test, target_train, target_test = train_test_split(accidentData, target, test_size = 0.3)

# Create adaboost classifer object
adaboostClassifier = AdaBoostClassifier(n_estimators = 50, learning_rate = 1)

# Train Adaboost Classifer
trainedModel = adaboostClassifier.fit(accidentData_train, target_train)

#Predict the response for test dataset
target_pred = trainedModel.predict(accidentData_test)

#Printing the Accuracy
print("Accuracy:", metrics.accuracy_score(target_test, target_test))
```

```
C:\Users\jpinf\Anaconda3\lib\site-packages\ipykernel_launcher.py:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  """
Accuracy: 0.9479768786127167
```

X. CONCLUSION

An accident can change the lives of many people. It is up to each of us to bring down this increasing number. This can be made possible by adopting safe driving measures to an extent. Since all instances of accidents cannot be attributed to the same cause, proper precautionary measures will also need to be exercised by the road development authorities in designing the structure of roads as well as by the automobile industries in creating better fatality reducing vehicle models. One thing within our capability is to predict the possibility of an accident based on previous data and observations that can aid such authorities and industries. This project was successful in creating such an application that can help in efficient prediction of road accidents based on factors such as types of vehicles, age of the driver, age of the vehicle, weather condition and road structure, so on

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, June 2022
Impact Factor: 6.252

REFERENCES

[1] https://www.statista.com/topics/5982/road-accidents-in-india/
[2] Srivastava AN, Zane-Ulman B. (2005). Discovering recurring anomalies in text reports regarding complex space systems. In Aerospace Conference, IEEE. IEEE 3853-3862.
[3] Ghazizadeh M, McDonald AD, Lee JD. (2014). Text mining to decipher free-response consumer complaints: Insights from the nhtsa vehicle owner's complaint database. Human Factors 56(6): 1189-1203. http://dx.doi.org/10.1504/IJFCM.2017.089439.
[4] Chen ZY, Chen CC. (2015). Identifying the stances of topic persons using a model-based expectationmaximization method. J. Inf. Sci. Eng 31(2): 573-595. http://dx.doi.org/10.1504/IJASM.2015.068609.
[5] Williams T, Betak J, Findley B. (2016). Text mining analysis of railroad accident investigation reports. In 2016 Joint Rail Conference. American Society of Mechanical Engineers V001T06A009-V001T06A009. http://dx.doi.org/10.14299/ijser.2013.01.
[6] Suganya, E. and S. Vijayarani. "Analysis of road accidents in India using data mining classification algorithms." 2017 International Conference on Inventive Computing and Informatics (ICICI) (2017): 1122-1126.
[7] Sarkar S, Pateshwari V, Maiti J. (2017). Predictive model for incident occurrences in steel plant in India. In ICCCNT 2017, IEEE, pp. 1-5. http://dx.doi.org/10.14299/ijser.2013.01.
[8] Stewart M, Liu W, Cardell-Oliver R, Griffin M. (2017). An interactive web-based toolset for knowledge discovery from short text log data. In International Conference on Advanced Data Mining and Applications. Springer, pp. 853-858. http://dx.doi.org/10.1007/978-3- 319-69179-4_61.
[9] Zheng CT, Liu C, Wong HS. (2018). Corpus based topic diffusion for short text clustering. Neurocomputing 275: 2444-2458. http://dx.doi.org/10.1504/IJIT.2018.090859.
[10] ArunPrasath, N and Muthusamy Punithavalli. "A review on road accident detection using data mining techniques." International Journal of Advanced Research in Computer Science 9 (2018): 881-885.
[11] George Yannis, Anastasios Dragomanovits, Alexandra Laiou, Thomas Richter, Stephan Ruhl, Francesca La Torre, Lorenzo Domenichini, Daniel Graham, Niovi Karathodorou, Haojie Li (2016). "Use of accident prediction models in road safety management – an international inquiry". Transportation Research Procedia 14, pp. 4257 – 4266.
[12] Anand, J. V. "A Methodology of Atmospheric Deterioration Forecasting and Evaluation through Data Mining and Business Intelligence." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 2, no. 02 (2020): 79-87.
[13] Prayag Tiwari, Sachin Kumar, Denis Kalitin (2017). "Road-User Specific Analysis of Traffic Accident Using Data Mining Techniques". International Conference on Computational Intelligence, Communications, and Business Analytics. 10.1007/978- 981-10-6430-2_31.
[14] Kaur, G. and Er. Harpreet Kaur. "Prediction of the cause of accident and accident prone location on roads using data mining techniques." 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (2017): 1-7.
[15] Irina Makarova, Ksenia Shubenkova, Eduard Mukhametdinov, and Anton Pashkevich, "Modeling as a Method to Improve Road Safety During Mass Events", Transportation Research Procedia 20 (2017) 43.