IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

REAL-TIME CRASH PREDICTION USING ADAPTIVE BOOSTING IN MACHINE LEARNING

C.Kalaiarasi[1] Prasanth.K.C[2] Rakesh Kumar.T [3] Sakthivel.N[4]

Assistant Professor[1] ,Student[2,3,4]

Department of Computer Science and Engineering

PERI INSTITUTE OF TECHNOLOGY

Abstract

In this paper, we are going to look into the project which is going to be an advanced managing site which can be used to manage multiple educational institutions along with extra features like as 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, 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 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 the planning and management of traffic and would help us reduce a lot of road accidents in the future.

Introduction

Machine Learning is a system of computer algorithms that can learn from example through self-improvement without being explicitly coded by a programmer. Machine learning is a part of artificial intelligence which combines data with statistical tools to predict an output which can be used to make actionable insights.

The breakthrough comes with the idea that a machine can singularly learn from the data (i.e., example) to produce accurate results. Machine learning is closely related to data mining and Bayesian predictive modelling. The machine receives data as input and uses an algorithm to formulate answers.

A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience by personalizing recommendations.

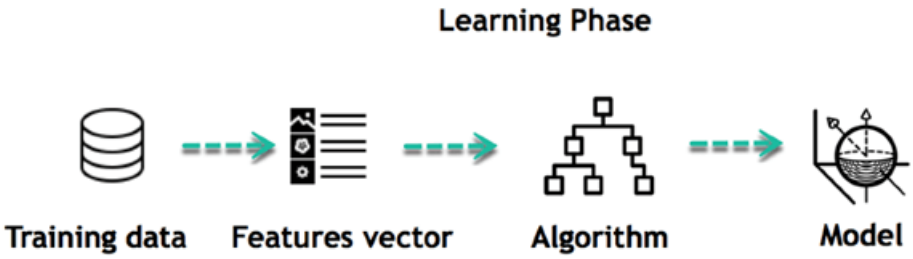

Fig 1: overview

Machine learning is also used for a variety of tasks like fraud detection, predictive maintenance, portfolio optimization, automated task and so on. Machine learning is the brain where all the learning takes place. The way the machine learns is similar to that of a human being. Humans learn from experience. The more we know, the more easily we can predict. By analogy, when we face an unknown situation, the likelihood of success is lower than the known situation. Machines are trained the same. To make an accurate prediction, the machine sees an example. When we give the machine a similar example, it can figure out the outcome. However, like a human, if it feeds a previously unseen example, the machine has difficulties predicting.

Inferring

When the model is built, it is possible to test how powerful it is on never-seen-before data. The new data are transformed into a features vector, go through the model and give a prediction. There are all the beautiful parts of machine learning. There is no need to update the rules or train again the model. You can use the model previously trained to make inferences on new data.

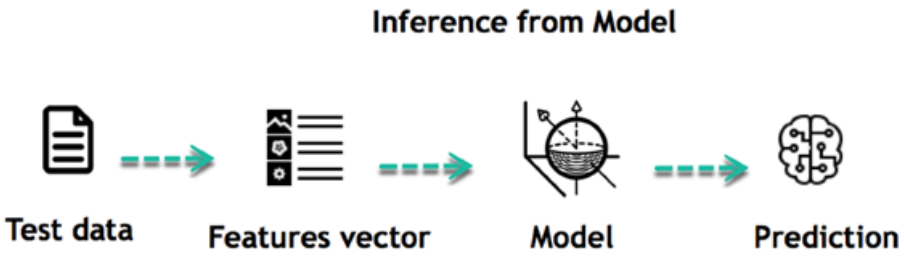

Fig 2: inferring process

Once the algorithm gets good at drawing the right conclusions, it applies that knowledge to new sets of data. ML is the best tool so far to analyze, understand and identify a pattern in the data. One of the main ideas behind machine learning is that the computer can be trained to automate tasks that would be exhaustive or impossible for a human being. The clear breach from the traditional analysis is that machine learning can take decisions with minimal human intervention Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a part of artificial

intelligence. ML algorithms build a model based on sample data, known as "training data", to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult to develop

Regression

When the output is a continuous value, the task is a regression. For instance, a financial analyst may need to forecast the value of a stock based on a range of features like equity, previous stock performances, and macroeconomics index. The system will be trained to estimate the price of the stocks with the lowest possible error.

AdaBoost

Classification or regression technique that uses a multitude of models to come up with a decision but weighs them based on their accuracy in predicting the outcome.

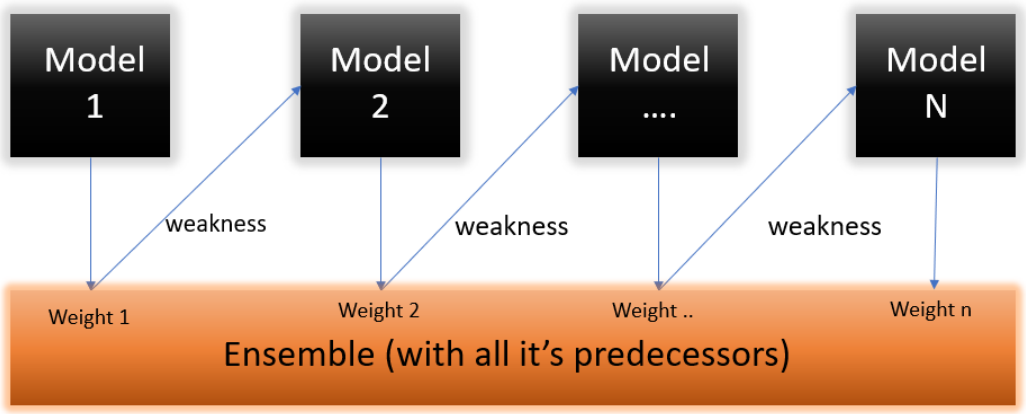

Fig 3: regression process

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. 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.

Literature Survey

Text mining analysis of railroad accident investigation reports

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.

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.

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.

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.

System design

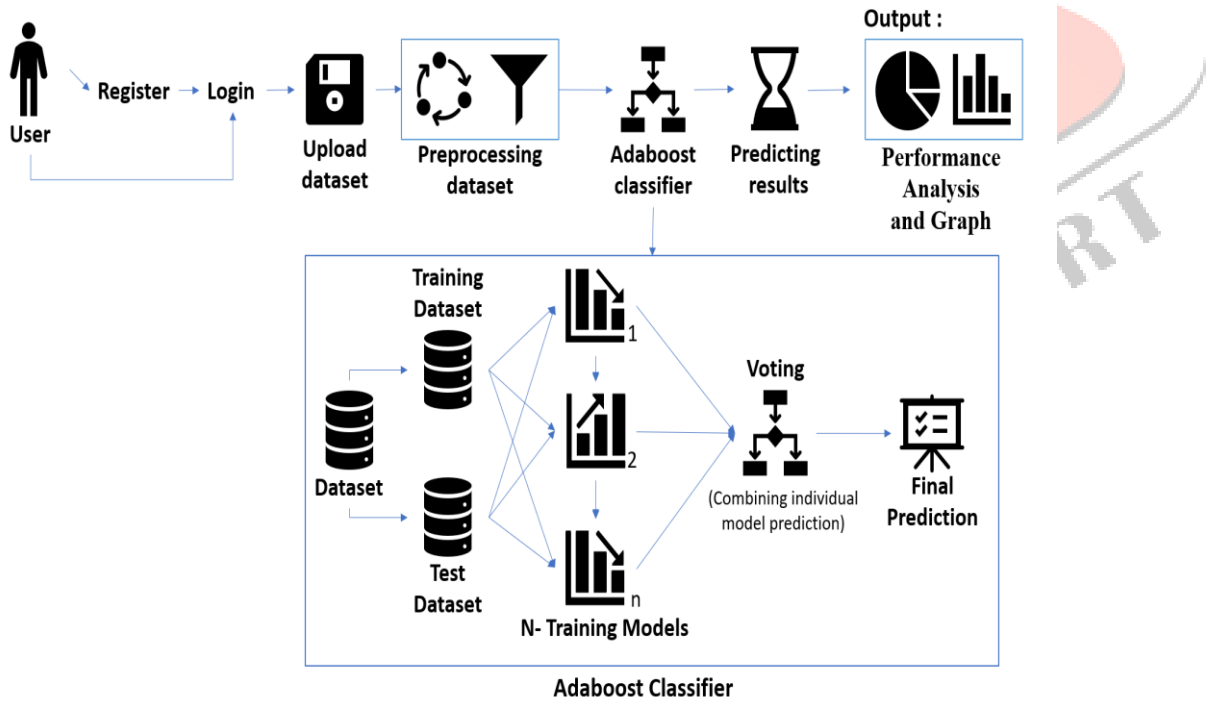

Fig 4: system design

Implementation

Modules

- Sign up/ Login
- Data hand
    - Data Collection
    - Dataset
    - Data Preparation
- Model Selection
- Analyze and Prediction
- Accuracy of the test set
- Saving the Trained Model

AdaBoost Algorithm

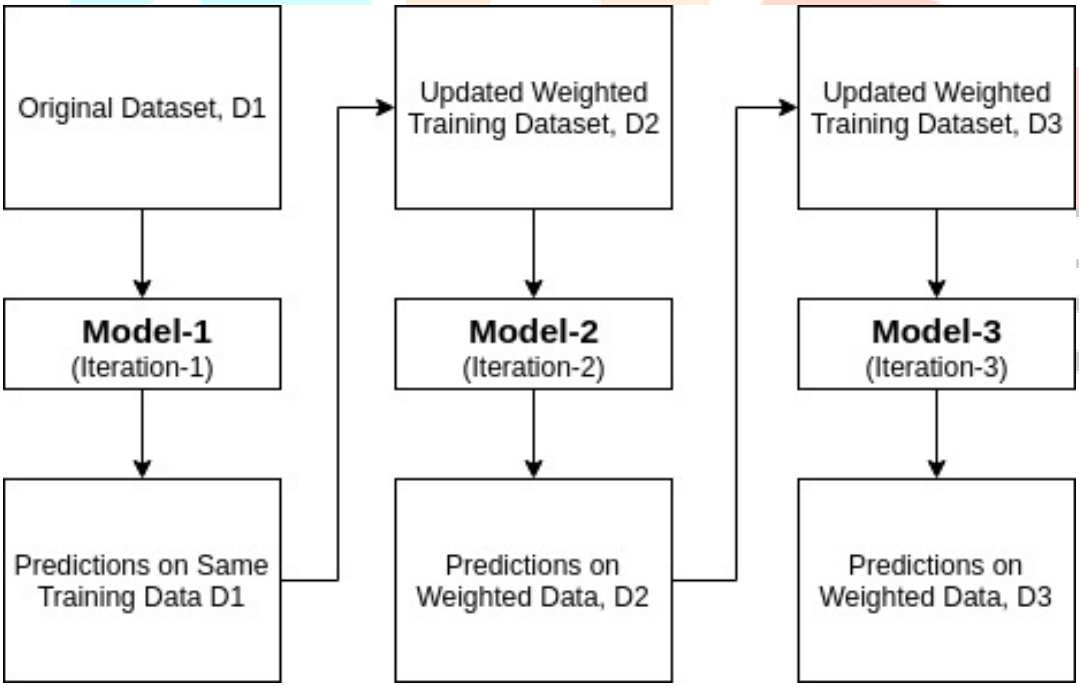

Fig 5:AdaBoost process

Performance Analysis

Meeting the Need for Speed: Data can be gathered and analyzed easily to generate instantly.

Addressing Data Quality: Data should be accurate and ensure data quality.

Displaying Meaningful Results: improving performance will produce more accurate results.

Timely Data: data should reach without wasting any time. Business Decisions- to support business decision and action taking in organization.High response-time: to obtain timely response Ensure security: to ensure security of data Avoid access **latencies:** to avoid any time delays.To Avoid Errors: To

generate error free data Data **Storage:** How data is stored across different nodes.Commit logs: setting the value for commit log before it grows.

Concurrency: how many threads can perform read and write operations: concurrent reads and concurrent writes. Caching: caching uses large amount of memory so we should tune them properly.

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. This model was implemented by making use of several data mining and machine learning algorithms applied over a dataset for India and has been successfully used to predict the risk probability of accidents over different areas with high accuracy.

References

1. Mpho Mokoatle, Dr Vukosi Marivate, Prof. Michael Esiefarienrhe "Predicting Road Traffic Accident Severity using Accident Report Data in South Africa", ACM Journals, 2020.
2. Mingchen Feng, Jiangbin Zheng, Jinchang Ren, Yanqin Liu, "Big Data Analytics and Mining for UK Traffic Accident Analysis, Visualization & Prediction", ICMLC: Conference on Machine Learning and Computing,2020.
3. Williams T, Betak J, Findley B. (2016). Text mining analysis of railroad accident investigation reports. At the 2016 Joint Rail Conference. American Society of Mechanical Engineers V001T06A009.
4. 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.
5. Sarkar S, Pateshwari V, Maiti J. (2017). A predictive model for incident occurrences in the steel plant in India. In ICCCNT 2017, IEEE, pp. 1-5. http://dx.doi.org/10.14299/ijser.2013.01.