GUI Application of best traffic route prediction for accident analysis using machine learning algorithm

Praveen Kumar D[#1], Pravin Kumar A [#2], Nithyashri J [#3], Veeralakshmi P [#4]

[#1], Student, Department of Information Technology, Prince Shri Venketeshwara Padmavathy Engineering College, Ponmar, Chennai

[#2], Student, Department of Information Technology, Prince Shri Venketeshwara Padmavathy Engineering College, Ponmar, Chennai

[#3,4] Associate Professor, Associate Professor, Head of The Department, Department of Information Technology, Prince Shri Venketeshwara Padmavathy Engineering College, Ponmar, Chennai

Abstract: **Currently, road transport infrastructure failing to cope up with the exponential increase in vehicular population. To computing the fastest driving routes and accidents in the presence of varying traffic conditions is an essential problem in modern navigation systems. To prevent this problem is to investigate the transport department dataset with machine learning method for finding the best road selection without accident forecasting by prediction results of best accuracy calculations. The analysis of dataset by supervised machine learning technique (SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments and analyze the data validation, data cleaning/preparing and data visualization will be done on the entire given dataset. Additionally, to compare and discuss the performance of various machine learning algorithms from the given transport department dataset with evaluation of GUI based user interface air quality prediction by given attributes.**

INTRODUCTION

Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Process of training and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. Machine learning can be roughly separated in to three categories. There are supervised learning, unsupervised learning and reinforcement learning. Supervised learning program is both given the input data and the corresponding labeling to learn data has to be labeled by a human being beforehand. Unsupervised learning is no labels. It provided to the learning algorithm. This algorithm has to figure out the clustering of the input data. Finally, Reinforcement learning dynamically interacts with its environment and it receives positive or negative feedback to improve its performance.

Data scientists use many different kinds of machine learning algorithms to discover patterns in python that lead to actionable insights. At a high level, these different algorithms can be classified into two groups based on the way they "learn" about data to make predictions: supervised and unsupervised learning. Classification is the process of predicting the class of given data points. Classes are sometimes called as targets/ labels or categories. Classification predictive modeling is the task of approximating a mapping function from input variables(X) to discrete output variables(y). In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too. Some examples of classification problems are: speech recognition, handwriting recognition, bio metric identification, document classification etc.

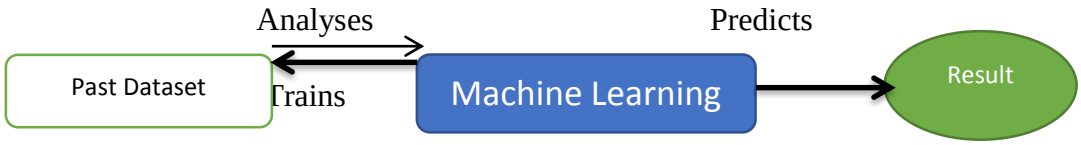

Fig: Process of Machine learning

Supervised Machine Learning is the majority of practical machine learning uses supervised learning. Supervised learning is where have input variables (X) and an output variable (y) and use an algorithm to learn the mapping function from the input to the output **is y = f(X).** The goal is to approximate the mapping function so well that when you have new input data (X) that you can predict the output variables (y) for that data. Techniques of Supervised Machine Learning algorithms include **logistic regression**, **multi-class classification**, **Decision Trees** and **support vector machines etc**. Supervised learning requires that the data used to train the algorithm is already labeled with correct answers. Supervised learning problems can be further grouped into **Classification** problems. This problem has as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for categorical for classification. A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes. A classification problem is when the output variable is a category, such as "red" or "blue".

RELATED WORKS

[1] Forecasting of road accident in Kerala: A case study (2018). Time series analysis is useful in discovering the trends in road accidents which enables the prediction of future patterns. It uses the SARIMA & Holt Winter (HW) model.

[2] Data mining methods for traffic accident severity prediction (2018). A decision system has been build using the model generated by the random forest technique that will help decision makers to enhance the decision-making process by predicting the severity of the accident. It uses decision tree, random forest, support vector machine, artificial neural network model.

[3] Graphical prediction of road accidents using data analysis (2018). To provide an efficient way to analyze and predict the accident zones and the severity. Data mining technique with big data

[4] Analysis of accident times for highway locations using K-Means clustering and decision rules extracted from decision trees (2018). Analyzing the traffic accident data play an important role in identifying the factors that affecting the repeated accidents and trying to reduce. It uses the data mining, data clustering technique, classification tree.

[5] Using hybrid data mining algorithm for analyzing road accidents data set (2019). The performance of both KNN and SVM algorithm using 'R' programming with large accident datasets and its results shows that hybrid model enhances the accuracy of road accident analysis. It uses the Machine learning, hierarchical learning, clustering, K-NN, support vector machine model.

PROBLEM DEFINATION

- **Exploration data analysis of variable identification**
  - Loading the given dataset
  - Import required libraries packages
  - Analyze the general properties

- Find duplicate and missing values
- Checking unique and count values

➢ **Uni-variate data analysis**

- Rename, add data and drop the data
- To specify data type

➢ **Exploration data analysis of bi-variate and multi-variate**

- Plot diagram of pair plot, heatmap, bar chart and Histogram

➢ **Method of Outlier detection with feature engineering**

- Pre-processing the given dataset
- Splitting the test and training dataset
- Comparing the Decision tree and Logistic regression model and random forest

➢ **Comparing algorithm to predict the result**

- Based on the best accuracy

SYSTEM ARCHITECTURE

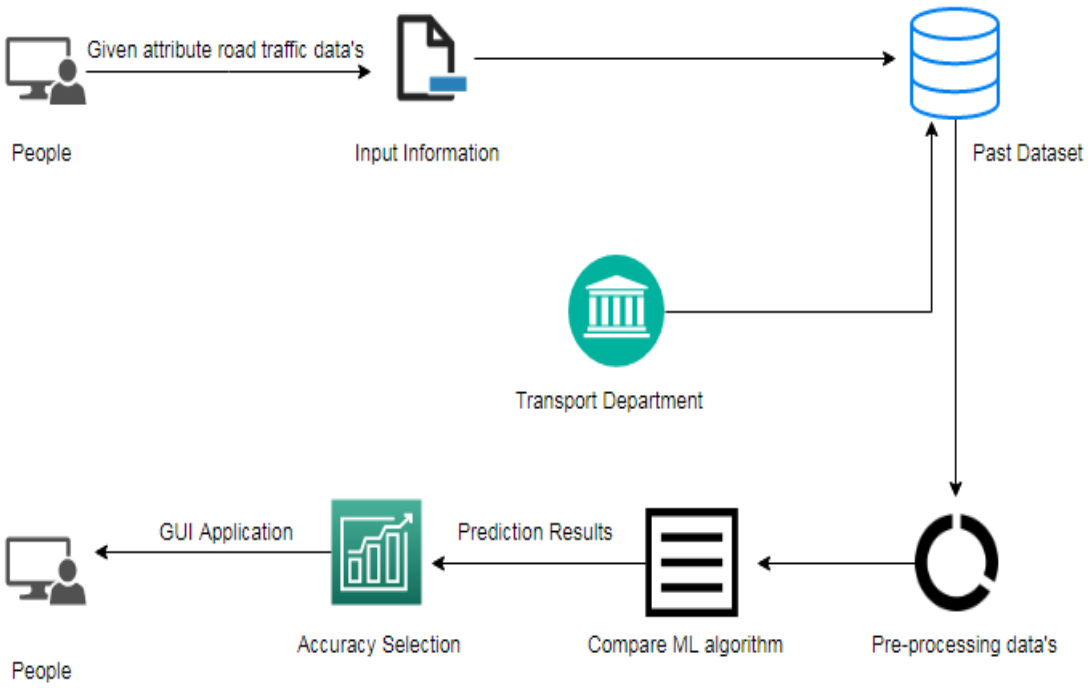

Fig. SYSTEM ARCHTECTURE

PROBLEM DESCRIPTION

METHODOLOGY: To predict the application for best traffic route for accident analysis, some steps are required to be followed. The steps can be defined in the following ways:

• DATA WRANGLING • DATA COLLECTION • BUILDING THE CLASSIFICATION MODEL • PREPROCESSING

• CONSTRUCTION OF PREDICTVE MODEL • GUI BASED PREDICTION FOR ACCIDENT ANALYSIS

DATA WRANGLING

In this section of the report will load in the data, check for cleanliness, and then trim and clean given dataset for analysis. Make sure that the document steps carefully and justify for cleaning decisions.

DATA COLLECTION

The data set collected for predicting given data is split into Training set and Test set. Generally, 7:3 ratios are applied to split the Training set and Test set. The Data Model which was created using Random Forest, logistic, Decision tree algorithms, K-Nearest Neighbor (KNN) and Support vector classifier (SVC) are applied on the Training set and based on the test result accuracy, Test set prediction is done.

BUILDING THE CLASSIFICATION MODEL

The predicting the air quality problem, decision tree algorithm prediction model is effective because of the following reasons: It provides better results in classification problem.

- It is strong in preprocessing outliers, irrelevant variables, and a mix of continuous, categorical and discrete variables.
- It produces out of bag estimate error which has proven to be unbiased in many tests and it is relatively easy to tune with.

PREPROCESSING MODEL

The data which was collected might contain missing values that may lead to inconsistency. To gain better results data need to be preprocessed so as to improve the efficiency of the algorithm. The outliers have to be removed and also variable conversion need to be done.

GUI BASED PREDICTION FOR ACCIDENT ANALYSIS

SVM:

- Its representation of different classes in a hyper-plane in multidimensional space. The hyper-plane will be generated in an iterative manner. So, that the error can be minimized and to divide the given datasets into classes to find a maximum marginal hyper-plane (MMH).
- It will generate hyper-planes iteratively that segregates the classes in best way.
- It will choose the hyper-plane that separates the classes correctly.

KNN:

- Load the given dataset and initialize ‘k’ value by chosen number of neighbors.
- Calculate the distance between trained customer behavior’s and new customer behavior.
- Sort the collection of distances in ascending order of customer behaviors by the distances and pick the first ‘k’ entries from sorted collection.
- So, get the labels of the selected ‘k’ entries and return the mode of the ‘k’ labels.

RFA:

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

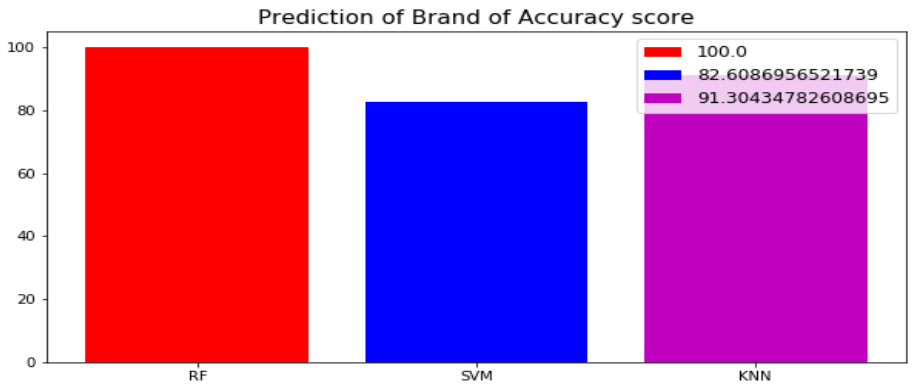

Fig. Comparison of machine learning accuracy results

RESULTS

- We can predict the best traffic route for accident analysis using the past data by machine learning algorithms.
- These reports are to the investigation of applicability of AI application for traffic accidental forecasting in operational conditions.
- It highlights some observations on future research issues, challenges, and needs.
- This will reduce huge number of accidents

CONCLUSION

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation.

FUTURE WORK

- Transport Department wants to automate the detecting the best route by not accident from eligibility process (real time) based on the account detail.
- To automate this process by show the prediction result in web application or desktop application.
- To optimize the work to implement in Artificial Intelligence environment.

REFERENCE PAPER

[1] D. Schrank, B. Eisele, T. Lomax, and J. Bak, "2015 Urban mobility scorecard," Texas A&M Transp. Inst., College Station, TX, USA, 2015.

[2] T.-Q. Tang, Z.-Y. Yi, and Q.-F. Lin, "Effects of signal light on the fuel consumption and emissions under car-following model," Phys. A, Stat. Mech. Appl., vol. 469, pp. 200–205, Mar. 2017

[3] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," Proc. IEEE, vol. 91, no. 12, pp. 2043–2067, Dec. 2003.

[4] B. Asadi and A. Vahidi, "Predictive cruise control: Utilizing upcoming traffic signal information for improving fuel economy and reducing trip time," IEEE Trans. Control Syst. Technol., vol. 19, no. 3, pp. 707–714, May 2011.

[5] M. A. S. Kamal, J.-I. Imura, T. Hayakawa, A. Ohata, and K. Aihara, "A vehicle-intersection coordination scheme for smooth flows of traffic without using traffic lights," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 3, pp. 1136–1147, Jun. 2015.

[6] P. Koonce et al., "Traffic signal timing manual," FHWA, Washington, DC, USA, Tech. Rep. FHWA-HOP-08-024, 2008.

[7] P. R. Lowrie, "The Sydney coordinated adaptive traffic system— Principles, methodology, algorithms," in Proc. Int. Conf. Road Traffic Signalling, 1982, pp. 67–70.

[8] H. Jiang, J. Hu, S. An, M. Wang, and B. Park, "Eco approaching at an isolated signalized intersection under partially connected and automated vehicles environment," Transp. Res. C, Emerg. Techn., vol. 79, pp. 290–307, Jun. 2017.

[9] BusinessTech Writer. South africs, shocking road death number at highest level in 10 years. Technical report, BusinessTech, 2017.

[10] S Vasavi. Information and Communication Technology, chapter Extracting Hidden Patterns Within Road Accident Data using Machine Learning Techniques, pages 13–23. Springer, 2018

[11] S Kumar and D Toshniwal. A data mining approach to characterize road accident location. Journal of Mod. Transport, 2016.

[12] National Road Safety Council. Study on the use of alcohol and its effects on road users in namibia: Transport management issues. Technical report, National Road Safety Council of Namibia, 2016