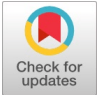# Smart Traffic Analysis using Machine Learning

Aditya Krishna K.V.S, Abhishek K, Allam Swaraj, Shantala Devi Patil, Gopala Krishna Shyam

Abstract: Congestion is costly as well as annoying. India is the second largest road network in the world. Out of the total stretch of 5.4 million km of road network, almost 97,991 km is covered by national highways.The major cause leading to traffic congestion is the high number of vehicle which was caused by the population and the development of economy[1].Typical urban residents spend more than ten hours a week driving of which (one to three hours) occurs in congested situation. In smart city roads would be equipped with the sensors for analyzing the trafficflow and also there are few traffic analysis / prediction methods use neural network and other prediction models which are not so efficient and suitable for many real world application [1]. So, here in this paper solution for traffic analysis using random forest algorithm is being proposed which would select only part of data for analyze like two third of entire data and predict the traffic congestion of specific path and notifying well in advance the vehicles intending to move to move on that specific path. Thus accurate traffic flow information help road users for fast and safe transporting.

Index Terms: Machine Learning, Traffic analysis, Styling, Random Forest

I. INTRODUCTION

A **smart city** is an urban area that uses different types of electronic data collection sensors to supply information which is used to manage assets and resources efficiently [2]. This includes data collected from citizens, devices, and assets that is processed and analyzed to monitor and manage traffic and transportation systems, power plants, water supply networks, waste management, law enforcement, information systems, schools, libraries, hospitals, and other community services [2]. The traffic analysis in the smart cities is very often fully automated. So the traffic information has to be provided accurately on the internet for the user purposes. So, smart traffic analysis will play a very major role in the major cities of India and also other countries.

There is no definitive explanation of a smart city because of the breadth of the Technologies that can be incorporated into a city in order for it to be considered a smart city. Mark Deakin defines it as a city that utilizes ICT to meet the demands of its citizens, and that community involvement in the processes is a necessity for a smart city [2]. From the definition given by HusamAlWaer and Mark Deakin in their research publication "From Intelligent to Smart Cities, "the factors that contribute to a city being classified as smart are [2]:

☐ The application of a wide variety of digital and electronic technologies to the city and its communities

☐ The application of ICT top lift life and the working environments in the region

☐ The embedding of such ICT withingovernment systems

☐ The territorialization of practices that bring the people and ICT together in order to foster innovation and enhance the knowledge that they offer.

Smart traffic analysis will help the users in finding the shortest path to the destination without any loss of time. The system that is being developed will help in time of travel of a person and also would help in the safety of that person by avoiding the accident prone regions.

II. LITERATURE SURVEY

In the below section we will discuss about the literature survey conducted on this topic based on the previous researches performed in this topic:

1. **Paper Name**: Traffic accident analysis using Machine learning Paradigms [9]
   Problem Statement: Modeling the severity of injury that occurred during traffic accidents [9]
   Mechanisms: Artificial Neural Networks using hybrid learning, decision tress, DTANN [9]
   Advantages: The behavioral and roadway accident patterns can be useful to develop traffic safety control policies [9]
   Disadvantages: Data set used doesn't provide actual speed during the accident occurred [9]
   Gaps: 3% the use of ml algorithms for traffic accident analysis
   Future work: To find out the speed of vehicle during the accident [9]

2. **Paper Name:** Vehicle crashes and machine learning [7]
   Problem Statement: Predicting crash fatalities with machine learning [7]
   Mechanisms: Regression problem, Random forest algorithm(a tree based algorithm) [7]
   Advantages: Helps to understand the causes and the factors that affect car crash severity[7]
   Disadvantages: Random forest algorithm provides accuracy only up to 87% [7]
   Gaps: 25% use of random algorithm
   Future work: Find some other algorithm to increase the model performance [7]

3. **Paper Name:** Using machine learning to predict car accident risk [10]

   Problem Statement: Predict accident risk per road segment per hour within the specified area using ML [10]

   Mechanisms: Supervised Machine learning, Gradient boosting [10]

   Advantages: Predicts accident risk per road segment every hour [10]

   Disadvantages: It does not use real time information, accidents are random occurrences [10]

   Gaps: 2% use of real time data

   Future work: Use of Real time traffic information to improve the model significantly [10]

4. **Paper Name:** Machine learning based traffic congestion prediction in a IoT based smart city [1]

   Problem Statement: Traffic congestion prediction using machine learning [1]

   Mechanisms: Decision tree, Random forest algorithm, SVM, MLP, Logistic Regression [1]

   Advantages: It can be effectively used by any devices which has less computation capability & with less resources [1]

   Disadvantages: It requires huge past and present real time traffic information [1]

   Gaps: 15% use of random forest algorithm

   Future work: The traffic congestion can be predicted using various hybrid techniques [1]

5. **Paper Name:** Smart Traffic Management system [8]

   Problem Statement: Predicting traffic using RFID technology [8]

   Mechanisms: RFID controller, RFID tag [8]

   Advantages: RFID technology easy and cost efficient to install when compare to other techniques [8]

   Disadvantages: The frequency noticed may not be accurate sometimes [8]

   Gaps: 2% the concept of predicting traffic

   Future work: It can be enhanced by using other powerful communication network other than GSM[8]

III. MOTIVATION

India is a country of huge population. The Road traffic in all cities of India is of greater concern. There is always a long wait for the people on the roades of the cities. India is among the top countries with large traffic index in the worldand, it is also 4th among the traffic index rankings of 2019 [3]. With high time index and also the C02 (Carbon di oxide) percent among all the cities [3]. So our main motivation was to provide a small remedy for this problem. Our project will describe more about the some traffic effected in Bangalore which is the fourth highest in the country for traffic

IV. METHODOLOGY

The dataset for this project is created for some random places on our own by comparing with the other datasets. The data in general has the details about the places in the form of nodes, time and distance for the traffic data. In our traffic analysis using machine learning we would be using the random forest algorithm to analyze the delay

between the nodes. The delay will help us to choose the route of less traffic.The random forest algorithm will help in classifying whether the traffic is more or less. The datset is created by comparing some available datasets.The reason for creating own dataset was to use data that can be analyzed on the roads of India. We have created a data for Yealahanka in Bangalore for easy understanding.

The reason for using the random forest is that the random forest will actually construct a number of decision trees while training and outputting the class that is the mode of all classes or can also be said as mean prdiction of the decision trees. The random forest will actually take two-third of the original data in the dataset and will try to construct a multitude of decision trees on this dataset.

V. SYSTEM DESIGN

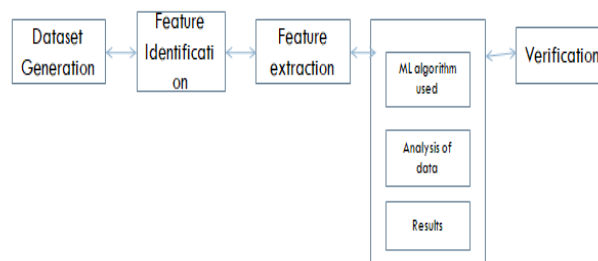The TAM Algorithm that is identifed for this project will have the following set of modules as follows:

Fig 1: System Design Of TAM Algorithm

1) **Dataset Generation**:
   The dataset for this project is generated based on available datasets for traffic analysis. The dataset is created for a particular location in bangalore called Yelahanka For easy understanding. The dataset will be in the form of a .csv file.

2) **Feature Identification**:
   The neccessary features for the project are to be identified like time, distance, delay, Vehicle Number etc. The features which are associated with the project are identified for the dataset by using which the analysis could be easily performed.

3) **Feature Extraction:**
   Feature extraction will in general make use of the dimensionality reduction procedure to reduce and consider only those neccessary attributes neccessary for the project like time ,distances, Nodes between which the traffic in general is identified.

4) **Machine Learning Algorithm used for Analysis**:
   The Machine learning algorithm that is used for the traffic analysis we have used for our TAM algorithm is the Algorithm.The Algorithm will help in classifying whether the traffic is more or less in a particular area based on the dataset loaded to the algorithm.

5) **Verification:** The Verification step will check whether the analysis done on the dataset is proper or not. This means that the analysis step is giving the proper result or not.

VI. RESULTS

The main aim of this project is to analyze the traffic between two particular nodes with the help of the data that is available in the dataset. The results that are observed in this project are given as follows:

- The Random Forest Algoithm will predict the Mean square error(MSE) which means the measure of prediction accuracy of the Random Forest model.
- It will also calculate Mean Absolute error(MAE) which means the difference between two continuous variables may be X and Y.
- It will also calculate the Root mean squared error(RMSE) which means the frequently used measure of difference in the values predicted by th machine learning model.

The formulas depicting the Eq 1. mean square error , Eq 2. mean absolute error and also theEq3.. root mean square error are actually depicted below and also the output for our project depicting all these quantities given above is displayed below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(P_i - O_i\right)^2}{n}}$$ ----Eq (1)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$ ----Eq (2)

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$ ---Eq (3)

@easycalculation.com

Fig 5: Results of the Random Forest Analysis

VII. DISCUSSIONS

In the previous research papers using machine learning algorithms they try to predict the traffic congestion, predicting the crash victims [8] . Here were are also using the same machine learning algorithms like Random forest algorithm, etc to predict the traffic in the particular areas which user want to know. In previous papers they were predicting traffic and crash fatalities, here we are predicting traffic and accidents in area, which the helps user to predict the traffic oriented areas and accident zones [8]. Which helps the user to find out easiest and fast path to the destination with the help of the prediction of traffic and accident areas. In previous papers they were using RFID technology to predict the vehicle traffic which is bit costly here we are using the datasets to get the information about the traffic information and accident areas information [8]. Using the data sets we predict the traffic in the particular areas and accident prone regions. We are using random forest algorithms because it gives the accurate analysis of the traffic data and accident data.

We are using more accurate algorithms to predict the information from the data sets of traffic analysis and accident prone regions that is random forest algorithm because it cannot take a small part of the entire data which may be $2/3^{rd}$ of the entire data. With the help of this data the analysis could be more accurate. We are using new algorithm called TAM algorithm, this algorithm will help in classifying whether the traffic is more or less in a particular area based on the data set loaded. Where as in previous papers they take data through RFID technology but we are using the data sets to get information about vehicle traffic and accident areas information [8].

VIII. CONCLUSION AND FUTURE WORK

The Traffic analysis model based on the random forest algorithm is presented . The main aim behind doing this project is to analyze the traffic and provide a model which could actually reduce the traffic to some extent in a particular area. And in this regard the training model is created which could analyze the Traffic. The main idea behind this project was not only to analyze the traffic but also to recommend a new route with less traffic with a good accracy to the user who has requested. So, the training model in this regard will help in analyzing the traffic by constructing decision trees and predicting the result more accurately. The main reason for chosing the random forest algorithm was because it can be used for easy classification of the data into less or more traffic effected and as the algorithm would construct decision trees based on the dataset provided. I would like to conclude by saying the need for using the Machine Learning for this project. It is because the analysis is being performed on random data the Machine learning Algorithms will help in training the model for new data which basically means for learning. In the future work on this project:

1. We would like to extend the area of traffic analysis from one particular place to a number of places
2. We would like to add the reinforcement learning to this model for self learning in the future without any extra data needed.
3. We would like to create a mobile application which could work as a user interface for providing user data in the form of a voice

We would Like to collect dynamic data with the help of some sensors from many parts of the city.

REFERENCES

1. Suguna Devi.,Neetha T.,2017.,Machine Learning based traffic congestion prediction in a IoT based Smart City,Department of computer science and engineering ,Brilliant School Educational Institutions,Hayathnagar,Hyderabad,Telangana,India,2017.
2. "Smart City", Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 18 Apr. 2019. Web. 19 Apr. 2019.
3. "Traffic Index of 2019", numbeo.com, Apr 2019.
4. "Machine Learning an Introduction to mean square error and regression lines",Moshe Binieli,Freecodecamp.com, Mean Square Error Image
5. "Mean Absolute Error",easycalculation.com.
6. "Root mean square error",easycalculation.com.
7. "Vehicle Crashes and Machine Learning",Abidishakur,towardsdatascience.com, Jan 01 2019
8. NinadLanke.,SheetalKoul., Smart Traffic Management System,Sahakarnagar,Pune,Department of Information Technology,SKNCOE,Pune, August 2013
9. Miao chong.,Ajithabraham.,and Marcin Paprzycki, Traffic Accident Analysis Using Machine Learning Paradigms,ComputerScienceDepartment,Oklahoma State University,USA, School of Computer science and engineering,Chung-AngUniversity,Korea, Computer Science ,SWPS,Warszawa,Poland,Decemeber 20 2004
10. "Using Machine Learning to predict Car accident Risk", Daniel Wilson, May 3 2018