IDENTIFYING ACCIDENT PRONE ZONE USING BIG DATA ANALYSIS

1Prof Shantanubhushana B.M,B.E,M.Tech.,(Ph.D), 2Meghana Rasalkar, 3Shweta kakodra, 4Kirti Kemmankeri, 5Tanushree Yapaladinni.

1Professor , 2Student, 3Student, 4Student, 5Student

1Computer Science And Engineering,
1AGM Rural College of Engineering and Technology, Hubli, Karnataka, India

Abstract : Identifying accident prone zone using big data analysis of the large-scale data of transportation and accidents has many potentials and it can give very useful insights from the hidden relationship of data. Accidents datasets are used to find main causes of traffic accidents which mainly causes traffic causality and congestion. This paper uses classification algorithm to analyze and predict the possibilities of traffic accidents in python coding environment. The k-means clustering algorithm predicts the possibilities of outcomes being a sign of any accident. Checking for the accurate results from algorithms rules and policies are made which is submitted to decision makers for further road safety measures.

Index Terms —Big data, Python.

I. INTRODUCTION
Traffic safety refers to the risk of a person being killed or seriously injured while using the traffic network as a pedestrian, cyclist or user of public or private transport. The importance of traffic safety is obvious, According to the world health organization, traffic accounts for many deaths and injuries. Almost every day, people face traffic accident in one way or the other leading congestion in one or more. Traffic management is a key branch within logistics. It concerns the planning, control and purchasing of transport services needed to physically move vehicles (for example aircraft, road vehicles, rolling stock and water craft). Big data ecosystem has the ability to store, manipulate, analyze and mine large accident datasets.

II. METHODS AND MATERIALS
A. Related Work
Bigdata analytics on Traffic Management Using Big Data Analytic Tool [1] by Saranya Krishna, Sharanya K S using Bigdata Techniques. This study revealed some common misconceptions about road incidents. There results revealed that driver’s attributes such as age and sex could be predicted correctly up to 70% by providing other attributes for an accident. Bigdata analytics on CCTV Images for Collecting Traffic Information [2] by Hyeongsoon Im and Bonghee Hong using Bigdata Techniques. They propose a new method for automatically calculating traffic volume and vehicle speed by pattern analysis using pixel data extracted from CCTV Video image. CCTV can detect Vehicles in wider areas because it has spinning function. Further, it can detect situations beyond a particular area, enabling possible traffic monitoring and collecting information.Predicting Traffic of Online Advertising in Real-time Bidding Systems from Perspective of Demand-Side Platforms [3] by Hsu-Chao Lai and Wen-Yuen Shih using Bigdata Techniques. They propose a method predicting traffic of requests from perspective of DSPs. They chose prediction model with closed-form solution. Online advertising works through real-time bidding systems. RTB will hold auctions between platforms and advertisers, QoS-aware Traffic-efficient Web Service Selection over Bigdata space [4] by T.H.Akila S and Banage T.G.S Kumara using Bigdata Techniques. They proposed successful web service selection method based on Bigdata space while addressing internal traffic concerns during the selection process of the web services.

Analysis of Movement and stay pattern using LTE signal [5] by sungil Ryu and Hyejoo Kim using Bigdata Techniques. They proposed change of traffic flow by day and time was analyzed using LTE signal in Seoul and nearby suburbs.


B. Proposed System
Our main aim of this project is to analyze the previous accident as well as vehicular causality records. This analysis done from the collected datasets will be given to the decision makers so that they can formulate rules and policies in order to enhance road safety, decrease the number of accident and avoid the traffic congestion. Here we have consider previous accident datasets with attributes. The attributes for accident dataset is source and destination and numbers of accidents happen previously. Large number of accidents from a cluster it is indicated by red spots. When User is nearer to the red spot it...
generates voice message like follow the traffic rules for eg. Speed limit, one way or two way road, work under progress, and school ahead etc. We have used k-means clustering algorithm to test the accuracy of the result. The System architecture of the application is given below, collected data is stored in the database on which with the help of clustering algorithms data is clustered and grouped indication a single marker on the map. Later voice message is sent to the user if he is not using the application. The collection of data or data set is also done with the help of review in this application.

SYSTEM ARCHITECTURE

![System Architecture](image)

**K-means clustering algorithm:** k-means is one of the simplest unsupervised learning algorithm that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given dataset through a certain number of clusters fixed Apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as possible far away from each other.

The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from previous step. Finally, this algorithm aims at minimizing an objective function know as squared error function given by

\[ J(V) = \sum_{i=1}^{c} \sum_{j=1}^{c_j} \left( \left\| x_i - v_j \right\| \right)^2 \]

Where, ‘\( \left\| x_i - v_j \right\| \)’ is the Euclidean distance between \( x_i \) and \( v_j \).
‘\( c_j \)’ is the number of data points in \( j^{th} \) cluster.
‘\( c \)’is the number of cluster centers.

K-means cluster algorithm is fast, robust, easier to understand and relatively efficient. Gives best result when data set are distinct or well separated from each other.

III. RESULTS AND DISCUSSION

A. Accidents data Analysis

Traffic accident in India are a major source of deaths, injuries and property damage every year. The National Crime Records Bureau (NCRB) 2016 report states there were 496,762 roads, railway crossing-related traffic accidents in 2015. Of these, road accidents accounted for 464,674 accidents which caused 148,707 traffic related deaths in India. The three highest total number of fatalities were reported in Uttar Pradesh, Maharashtra and Tamil Nadu, and together they accounted for about 33% of total Indian traffic facilities in 2015. India reported a traffic accident rate of about 0.8 per 1000 vehicles in 2015.

B. Graph

India reported a traffic accident rate of about 0.8 per 1000 vehicles in 2015. Identified the major causes of traffic collisions as driving over speed limit, driving under the influence, and not using helmets and seat belts. Failure to maintain lane or yield to oncoming traffic when turning are prime causes of accidents on four lane, non-access controlled National Highway.
Total number of persons killed injured due to road accidents, from 2001 to 2010
According to the 2013 global survey of traffic accidents by UN World Health Organization, India suffered a road fatality rate of 16.6 per 100,000 people in 2013.

IV. CONCLUSIONS
The major human factors that contribute to the potency of road accident causation include drunken drivers, indecisiveness, fatigue, distraction, and confusion. In addition, in most of the cases the drivers are found to be inexperienced, risk takers, impulsive, aggressive, casual and unaware of road signals. Big data analytics tools helps us to enhance the road safety and decrease traffic crashes. In our study we use PYTHON environment to evaluate the datasets. K-means classification used as a classifier. Form our experiments k-means gave the optimum results, with a lowest computation time. This study revealed some misconceptions about road incidents. Our analysis showed that the human behaviors has strong impact on the traffic flow and safety decisions. Our results revealed that decrease in road accidents.

V. FUTURE WORK
• Our future directions is to use the results obtained so far in developing a traffic real-time mining system to prevent roadway accidents and congestions.
• There is a need to design an intelligent traffic cloud by making use of cloud computing to solve the problems related to real-time.
• Usage of PLCs and SCADA system in intelligent transportation system for smooth traffic flow can be considered for further work on traffic issue.
• Designing a promising traffic management system to provide smooth traffic flow in non-recursive congestion situation can be an interesting issue for future research.

VI. ACKNOWLEDGEMENT
We have brought this paper work as a mobile application with the Complete help of team work and guide support. AGM Rural College of Engineering and Technology, Varur, Hubli, Karnataka.

REFERENCE