AN ANALYSIS OF TRAFFIC SYSTEM BASED ON BIG DATA TECHNOLOGY

1st Veena Rathna Augesteelia
Research Scholar
School of computing sciences VISTAS
Chennai ,India veenajsph@gmail.com
2nd Dr.K.Rohini
Associate Professor
School of Computing sciences VISTAS
Chennai, India
rrohini16@gmail.com

Abstract—

The rapid advancement of huge technology of Big Data, its applications have grown increasingly diverse. The best gratitude for resolving traffic congestion in big cities is the appliance of Big data technologies in the intelligent system of transportation . The most common reasons of traffic congestion in large cities are examined in depth in this paper. Large scale data and associated traffic data supported by GPS are composedusing modern communication technologies and internet of things. To build a traffic prediction and vehicle prediction model the data analysis method is used. The forecasting model is used to estimate traffic stream in each and every direction of traffic intersections at a specific location and time. The likelihood of traffic flow and congestion at a specific location at a specific time, as well as pedestrian travel trajectory and travel habits predicted. Consider the effect of non-motorizedautomobiles and pedestrians on traffic congestion at the same time. To untangle traffic jam problems, use real-time traffic information monitoring and forecasting results. A helpful reference is provided for decision-making in metropolitan trafficjam and traffic flow solutions when combined with optimization and control strategy for traffic collaborative management.

Keywords— Big data technology, Traffic systems, Machine Learning, Intelligent Transportation system

DOI Number:10.14704/nq.2022.20.8.NQ44429 **NeuroQuantology 2022; 20(8): 3985-3992**

I. INTRODUCTION

The faster At present,the three-dimensional traffic monitoring platform is completed or built in most of the cities . The pedestrian GPS, Vehicle GPS camera and other monitoring tools are not usd to monitor road cross-section flow, monitor vehicle travel speed and intersection shunt. The real-time evaluation of the operational status of urban roads is possible using these data . Huge amount of real-time monitoring data forms massive traffic information.It provides effective data for traffic jam prediction, and it also requires traffic jam forecast to proficiently and comprehensively cover the whole urban transportation network. In many places, traffic congestion has gotten worse, and traffic accidents have become more common.

These issues created traffic management issues that needed to be addressed. The traditional method of traffic management appears to be inadequate in response to rise traffic request and the strain on transportation properties. To improve traffic management system the need of intelligent transportation system (ITS) has become increasingly important. To improve the efficiency of urban traffic network The traditional road traffic management methods are deeply rehabilitated by using contemporary information technology, supported by Big Data technology. Most cities have completed or are in the process of completing traffic monitoring platforms in three dimensional form. To track vehicle speed, road cross-section flow, and intersection shunt, thevehicle GPS, pedestrian GPS, cameras, and other monitoring equipment are frequently used. Real-time examination of the operating status of urban highways is possible using these data. Massive traffic information is formed by enormous and real time monitoring data, in which it provides active data support for traffic jam forecast and also demands traffic jam prediction efficiently and thoroughly to cover the entire urban transportation network. One of the main goals of intelligent transportation system is predicted traffic jamsaccurately. The self-similarity laws of urban traffic circumstances will be employed for prediction nevertheless, road segment and ecological elements generate road segment variation and external causes persuade road network dynamics and making traffic jam law extremely complex and undefined. The prediction model system not only adapt to the difficult road network conditions, but it must also completehigh efficiency and high precision predictioninforms in response to changes in the road network environment.

II. CAUSES OF URBAN TRAFFIC JAM

In Large Cities the growth of population , Vehicles are one of the most common reasons of traffic jam.The population growth leads to an increase in personal cars and vehicles. The following are the reasons for the population density in some locations is just too high, in public urban transportation, the investment of public transportation vehicles is also expanding the development of non motor vehicles and walkers has certain effect on traffic congestion. In Urban Basic Transportation there are many Infrastructure problem arises due to the land restrictions in several metropolitan cities and the width of many traffic highways is limited. The distance between some nearby traffic junctions is comparatively low due to the urban design issues. There are also traffic signals at the stop's intersection, in which it reduces the amount of buses that stop there. The frequency of switching traffic signal and the duration of green and red lights in a given direction is too long. Another issue that cannot be overlooked is that the entrances to institutions and hospitals are located near busy roadways. All of these scenarios have the potential to produce traffic congestion.

III. EVALUATION OF TRAFFIC STREAM AND TRAFFIC CONGESTION

Evaluation of this traffic stream and congestion The flow of traffic is a serious aspect that donates to traffic congestion. Traffic flow is referred as the amount of number of passing cars per unit time in one of the direction of a road is referred to as traffic flow. The amount of vehicles travelling through a unit of time is during a specific direction at a traffic crossroads can be calculated and predicted using Big Data. The intelligent transportation industry generally uses three parameters to describe traffic flow: (1) The traffic flow is the first one, it is often known as traffic volume in which it reflects the number of cars travelling through a certain segment of road in a given amount of time. Vehicles per hour is the unit of measuremen (2) The traffic flow speed, is the second one,in which it is denoted by the flow and indicates the traffic flow speed in metres or kilometres (3) The traffic flow density, is third one in which it is measured in km/km and shows the degree of traffic flow density, is the number of vehicles incorporated within the length of the road unit.

IV.COLLISION WARNING SYSTEM

ALARM TYPE

- **FORWARD COLLISION WARNINGS (FCW)**
  This alarm system alerts drivers of an imminent rear end collision with a car ,truck or motorcycle.
- **URBAN FORWARD COLLISION WARNINGS (UFCW)**
  This system UFCW provides an alert before a possible low speed collision with the vehicle in front thus assisting the driver at a low speed in densely heavy traffic.
- **HEADWAY MONITORING WARNING (HMW)**
  The headway monitoring warning (HMW) helps drivers maintain a safe following distance from the vehicle ahead of them by providing visual and audible alerts if the distance becomes unsafe.
- **LANE DEPARTURE WARNINGS (LDW)**
  The LDW provides an alert when the vehicle un

intentionally departs from the driving lane without using the turn signals. If the turn signals are used when changing lanes, an alert is not generated. This is further classified into: (a) LDWL, for lane departures towards left lane and (b) LDWR, for the same towards right lane.

- **PEDESTRIANS AND CYCLIST DETECTION AND COLLISION WARNING (PCW)**
  The PCW notifies the driver of a pedestrian or cyclist in the danger zone and alerts drivers of an imminent collision with a pedestrian or cyclist.
- **OVERSPEEDING**
  Detects and classifies various visible speed limit signs and provides visual indication when the vehicle's speed exceeds the posted speed limit.

V SAMPLE DATA

	Code	latitude	longitude	wardName	alarmType \
0	8.645040e+14	12.984595	77.744087	Kadugodi	PCW
2	8.645040e+14	12.984595	77.744087	Kadugodi	PCW
4	8.645040e+14	12.987233	77.741119	Garudachar Playa	FCW
6	8.645040e+14	12.987233	77.741119	Garudachar Playa	FCW
8	8.645040e+14	12.987503	77.740051	Hudi	Overspeed
...	...	...	...	...	...
415224	8.645040e+14	12.976435	77.741516	Kadugodi	UFCW
415226	8.645040e+14	12.986425	77.745117	Kadugodi	UFCW
415228	8.639770e+14	12.969396	77.749886	Hagadur	Overspeed
415230	8.639770e+14	12.974123	77.746841	Hagadur	FCW
415232	8.639770e+14	12.975480	77.744125	Hagadur	PCW

	speed	recordedTime
0	32.0	2018-02-01T01:48:59.000Z
2	32.0	2018-02-01T01:48:59.000Z
4	41.0	2018-02-01T01:50:00.000Z
6	41.0	2018-02-01T01:50:00.000Z
8	37.0	2018-02-01T01:50:11.000Z
...	...	...
415224	0.0	2018-07-30T10:56:12.000Z
415226	12.0	2018-07-30T11:07:19.000Z
415228	17.0	2018-07-30T11:14:02.000Z
415230	36.0	2018-07-30T11:16:24.000Z
415232	28.0	2018-07-30T11:25:51.000Z

[207617 rows x 7 columns]

Preprocessing
Data size 207617 After Preprocessing 152276
55341 - Duplicate
Now Data Shape
(152276, 7)

VI TRAIN DATA INFORMATION

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 152276 entries, 0 to 207616
Data columns (total 7 columns):
deviceCode_deviceCode                152276 non-null int64
deviceCode_location_latitude         152276 non-null float64
deviceCode_location_longitude        152276 non-null float64
deviceCode_location_wardName         152276 non-null object
deviceCode_pyld_alarmType            152276 non-null object
deviceCode_pyld_speed                152276 non-null int64
deviceCode_time_recordedTime_$date   152276 non-null object
dtypes: float64(2), int64(2), object(3)
memory usage: 9.3+ MB
```

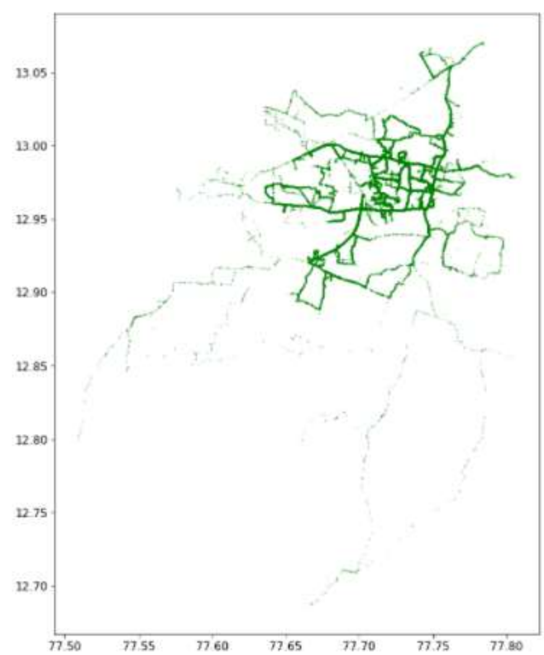
- Updated column names of train dataframe: Index(['deviceCode', 'latitude', 'longitude', 'wardName', 'alarmType', 'speed', 'recordedDateTime'],dtype='object')
- **Data Analysis**
- Range of latitude: 13.070075035095215 12.686662673950195Range of longitude: 77.80682373046875 77.5081787109375

Dataset statistics

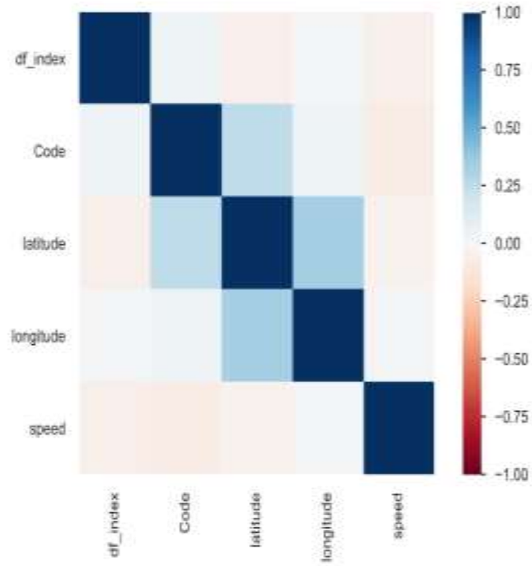
Number of variables	8
Number of observations	207617
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	12.7 MiB
Average record size in memory	64.0 B

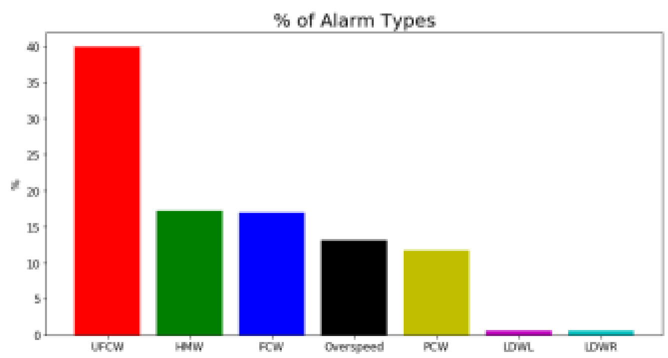
Variable types

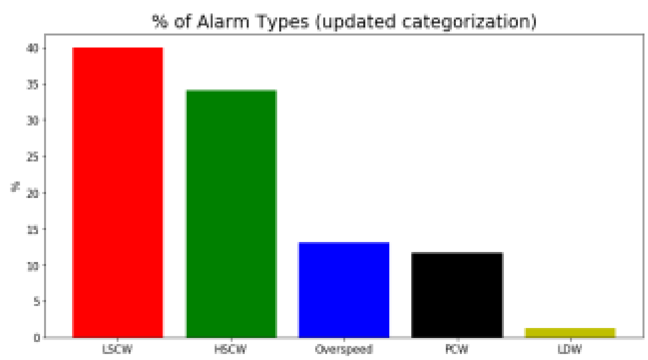
CAT	4
NUM	4



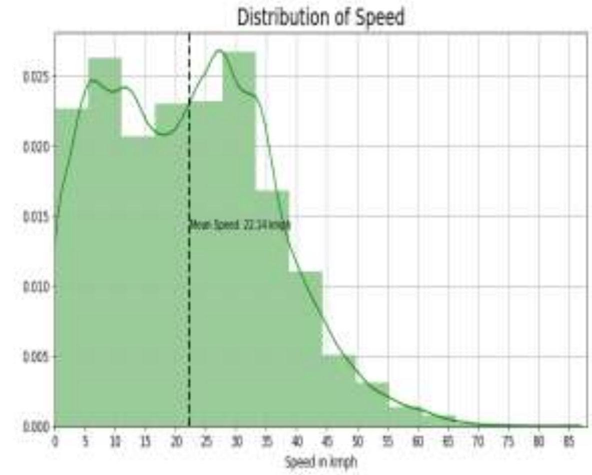
Correlations

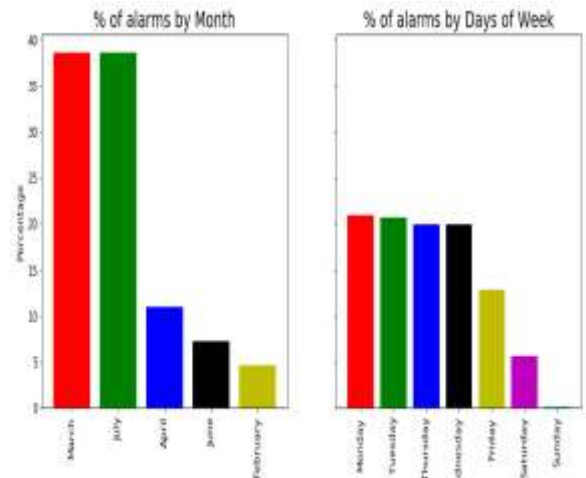


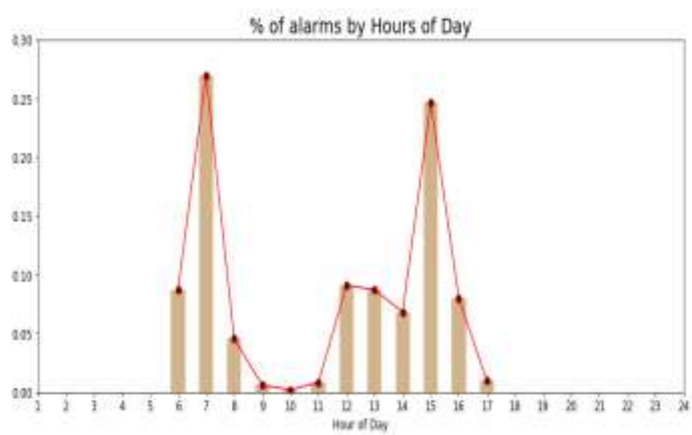


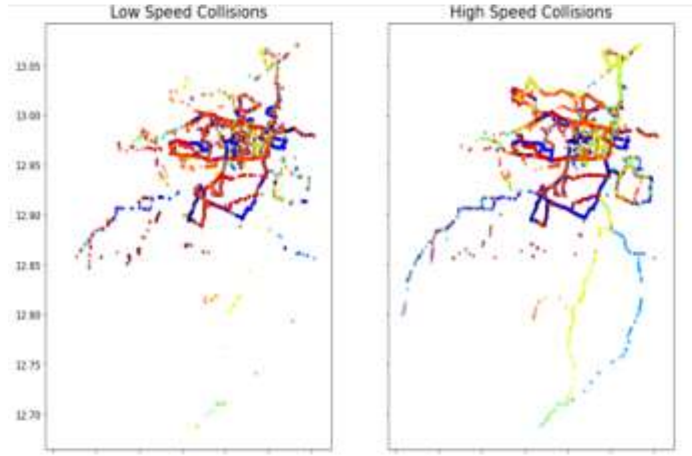


eISSN1303-5150


Distribution of Speed
Speed in kmph


% of alarms by Month
% of alarms by Days of Week
Percentage
March
July
April
June
February
Monday
Tuesday
Thursday
Wednesday
Friday
Saturday
Sunday


% of alarms by Hours of Day
Hour of Day



Distribution of Speed

- **Model Accuracy**

Logistic Model

- Train Accuracy of Logistic model: 39.72
- Test Accuracy of Logistic model: 39.62

Linear Model

- Train Accuracy of Linear model: 39.72
- Test Accuracy of Linear model: 39.62

K-Nearest Neighbour Model

- Train Accuracy of knn model: 88.37 {'n_neighbors': 2}
- Test Accuracy of KNN model: 66.27

Naïve Bayes Model

- Train Accuracy of Naive Bayes model: 39.72
- Test Accuracy of Naive Bayes model: 39.62

Decision Tree Model

- Train Accuracy of Decision Tree model: 99.25
- Test Accuracy of Decision Tree model: 77.1

<u>Algorithms Accuracy</u>

	Model	Score_train	Score_test
4	Decision Tree Classifier	77.10	77.10
2	k-Nearest Neighbors	66.27	66.27
0	Logistic Regression	39.72	39.62
1	Linear	39.72	39.62
3	Naive Bayes	39.62	39.62

V.CONCLUSION

It can solve the problems of traffic congestion prediction analysis and processing, traffic stream forecast, scientific planning of transportation organization using massive data technology. It needs the traffic information of the entire city and its supply practical data and solutions for traffic guidance and the concrete planning. The web technology of Things is used to collect data, retrieve real time data, and historical data. Using Big Data technology, the material is cleaned and preprocessed, and the right algorithm is then selected to build a traffic forecast model. With the increasing adoption of driverless technology in large cities, intelligent transportation systems powered by Big Data can deliver precise and highly reliable traffic data for the web of vehicle.

VI.REFERENCES

1. The templat Camilo Gutierrez – Osorio, Cesar Pedraza, “Modern data sources and techniquesfor
2. Raffaele Mauro, Andrea Pompigna, “State of the art and computational aspects of time-dependent waiting models for non-signalized intersections”, Journal of Traffic and Transportation Engineering September 2019.
3. ZhenhuaWang,YangsenYu,DangchenJu,“Analysis and Prediction of Urban Traffic Congestion Based on Big Data”, International Journal on Data Science and Technology, October 2018.
4. Hua-pu Lu, Zhi-yuan Sun and Wen-cong Qu, “Big Data-Driven Based Real-Time Traffic Flow State Identification and Prediction”, Discrete Dynamics in Nature and Society, Volume2015.
5. Sachin Kumar, Durga Toshniwal, “A data mining framework to analyze road accident data”, Journal of Big Data, a Springer Open Journal,2015.
6. Sung, Hong Ki and Chong, Kyu Soo, “Development of Road Traffic Analysis Platform Using Big Data”, International Conference on Advances in Big Data Analytics,ABDA’17.
7. Chen, “Analysis and Forecast of Traffic AccidentBig Data”, ITM Web of Conferences 12,2017.
8. Yachao Jia and Xin Wang, “Intelligent Traffic
9. Decision Analysis System Based on Big Data Mining”, IOP Conference Series: Materials Science and Engineering2018.
10. M. Koroliuk, C. Connaughton, “Analysis of big data set of urban traffic data”, Analysis of big data set of urban traffic data,2015.