



ROAD ACCIDENT ANALYSIS USING DATA MINING

Mrs. Kavitha Bai A. S¹, Aiswarya Thankachan², E. Sai Krupa³, Farzana M⁴, Kavya S⁵

1. Assistant Professor, Department of CSE, Vemana Institute of Technology, Bangalore.

2, 3, 4, 5. UG student, Department of CSE, Vemana Institute of Technology, Bangalore.

1.kavithabai.pawar@gmail.com, 2.aishwaryath123@gmail.com, 3.elatesaikrupa@gmail.com,

4.farzfarzana08@gmail.com, 5.kavyachinnu193@gmail.com

Manuscript History

Number: IRJCS/RS/Vol.06/Issue06/JNCS10091

Received: 29, May 2019

Final Correction: 30, May 2019

Final Accepted: 02, June 2019

Published: June 2019

doi://10.26562/IRJCS.2019.JNCS10091

Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India

Copyright:©2019 This is an open access article distributed under the terms of the Creative Commons Attribution License, Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

ABSTRACT -The universal tribulation with ever rising trend is the increase in the number of road accidents. This most disastrous pileups are the main cause of increasing death rates and casualties worldwide. The manifoldness of macabre of these disasters leads to a need for effective analysis and prediction of road accidents with the aid of powerful analysis tools and techniques. In this project, we analyses all the possible aspects of accidents happening in India and predict the number of accidents in the forthcoming years using python by data mining techniques. Various methods like PCA, SVM, classification and clustering algorithms are used in the process. Here, with the help of predictive analysis and time forecasting methods, the number of road accidents that are likely to occur in the upcoming years for different states are predicted with greater accuracy. Android Application is developed that alerts the travelers about the nearing potholes and bumps present on the roads so that they can take precautionary measures while travelling. These analysis and prediction can help the RTA to plan accordingly and reduce the number of road accidents by adapting safety measures for the welfare of the society.

KEYWORDS – road accidents, PCA-Principal Component Analysis, SVM-Support Vector Machine, data mining, Android Application.

I. INTRODUCTION

Road accidents are the biggest threats the world is facing today. They are becoming quotidian over the globe. The pressure on roads has been on the increase and the number of vehicles is increasing by leaps and bounds. In order to prevent them and reduce their repercussions, they need to be systematically analyzed at various levels. Efficient analysis of road accidents is made possible by data mining techniques. Data mining is the process of discovering patterns in large data sets and establish relationships to solve problems through data analysis. It is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information from a data set and transform the information into a comprehensible structure for further use. Alongside raw analysis step, it also involves many processes like preprocessing, database and management model post-processing of discovered structures, visualization, and online updating. The first step in data mining is gathering relevant data critical for business. Company data is either transactional, non-operational or metadata. Organizations with a strong consumer focus deal with data mining techniques providing clear pictures of products sold, price, competition and customer demographics. Data mining has been proven as a well-founded technique to analyze road accidents and predict accurate results. Firstly, data can be extracted from different sources preferably from the sources that provide authenticated datasets so that the prediction fetches real time solutions. Then the data is loaded into python where the data preprocessing takes place. Data is cleansed to remove all the impurities, outliers and missing values present.

Finally, the final dataset is evaluated to find out patterns that contribute in predicting the number of road accidents that are likely to occur in the nearing years. With this analysis, a conclusion can be drawn about the importance of solving this tragedy in order to protect number of lives of the people.

Android application plays a vital role in providing user-friendly user interfaces. It helps a business to increase sales, enhance revenues and reach their targeted goals. Android OS has many handy and primordial features. It helps the developers to build excruciating applications. This platform also allows developing applications that are rich in features with great functionality and usability. The custom mobile application development gives an added advantage to be different from others. It gives the users a host of options so that the developers can experiment its creativity. Android gives a huge range of options for graphic design need and their fulfilments. The graphic design features offered by Android are way superior to any other mobile application development platform. Game applications development is considered as the most pleasing platform for facing intuitive and attractive graphic design. This Android app development platform provides high built-in support for constructing the best 2D and 3D graphics that are found attracted by users. In this project, an Android application is developed using Java programming language. This helps to alert the travelers about the nearing potholes and bumps on their way.

The main goal of road accident analysis is to predict the amount and number of road accidents happening in different states of India in the future coming years and give awareness to the Road Transportation Authority regarding the chaos. The predictions in different states can be viewed in Tableau dashboard linked to the application. The application will in turn notify the travelers about the bad road conditions like potholes, bumps by providing a voice message about the situation. With the help of this notification, the user can take necessary measures and appropriate steps for his/her safety. With these analysis and predictions, major decisions can be made concerning the safety of the people and the society. The impact on the number of accidents can be reduced by considering various factors of roads and vehicles by implementing welfare plans and policies and prevent the traffic accidents that questions the existence of people on roads.

II. METHODOLOGY

The project is partitioned in tiny groups called as modules for easy coding and understand. This paper mainly contains two modules which are analysis and prediction module and application module. These modules are briefly explained in the following sections.

A. Data Preprocessing

Datasets are collected from various sources for preprocessing. They are loaded into python to transform the raw data for further processing. Datasets undergo ETL(Extract, Transform, Load)process. As shown in fig 1the data is extracted and cleansed through processes such as filling in missing values(NA), smoothing the noisy data or resolving inconsistencies in the data. Data is transformed into a structured manner in tabular form for analysis. After the data transformation, master dataset is created by combining the structured datasets.This step is an added advantage for accurate prediction.

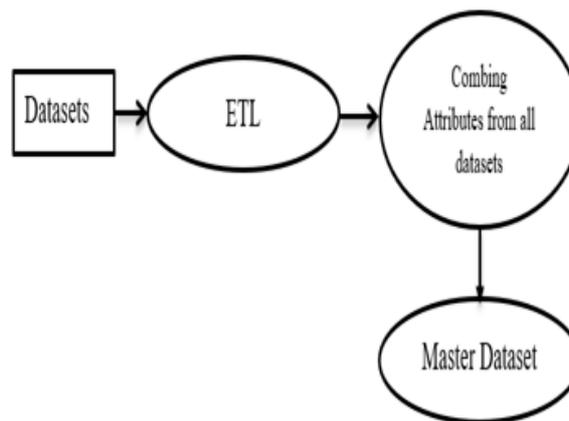


Fig 1 Data Preprocessing

B. Data Analysis and Correlation

As shown in Figure 2 the master dataset is generated which is used for further analysis. The static analysis of data is performed on the generated dataset. All the parameters are analyzed causing road accidents and plotting graphs for each parameter for different states. Attributes like climatic conditions, lane accidents, types of vehicles, presence of boulders are one of the reasons for the road accidents. After the analysis of data, correlation and dependencies is found between each dataset.

C. Predictive Analysis

The information is extracted from the dataset to predict future outcomes. Prediction of the road accidents can be done by the following two factors: Time Series Forecasting and Parameters in dataset. Time Series Forecasting involves the study of neural networks and machine learning. Neural network consists of LSTM and MLP algorithms. Mean Square Error (MSE) is calculated for obtaining accurate output. Prediction of future datasets and road accidents is done based on the historical dataset.

D. Parameter Based Prediction

Parameter based prediction is done by using algorithm like k means clustering, Principal component analysis, and SVM techniques. K means clustering is used for grouping similar data points into clusters. Principal component analysis is applied to the training model to reduce the dimension. SVM techniques are applied to find the hyperplane and to calculate the data points for predictions. The predictions and the analysis of the number of road accidents that are likely to occur in the upcoming years are predicted.

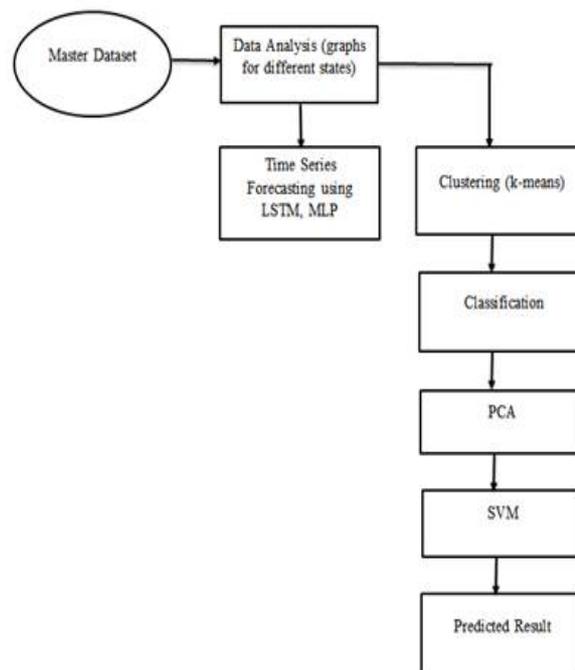


Fig 2 Data Preprocessing

Android Application Development Modules:

A. GPS Location

As shown in Figure 3 the location of the place is acquired by using GPS. The latitude and longitude of the particular location is displayed in the application. The obtained location with the longitude and latitude is sent to the database server.

B. Uploading Images to Server

The images of the potholes, bumps present on the roads are captured. The captured images are then uploaded to the server with the help of GPS. The images are stored on the database server.

C. Alerting Travelers

The travelers receive the voice notification about the nearing potholes and bumps about 100 meters prior in distance. They are alerted with a message about the road conditions so that they can take measures while travelling. The analysis of accidents and the graphs for the predicted analysis of road accidents can be viewed in the android application.

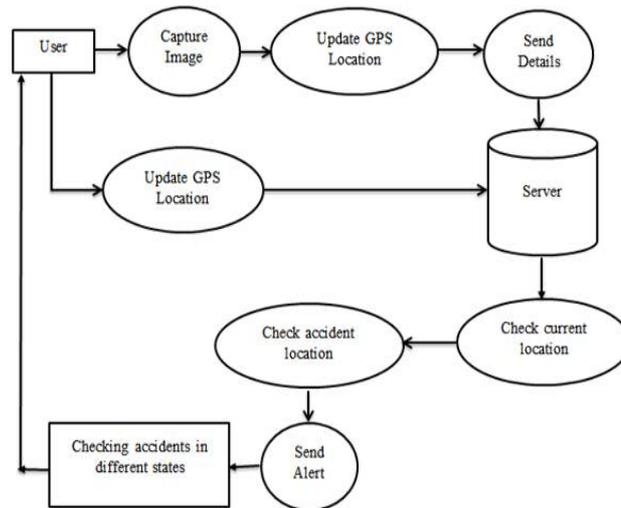


Fig 3 Android Application

III. SYSTEM ARCHITECTURE

In this section we describe how the connections take place between two modules this mainly composed of

- i) Data Mining module
- ii) Android application development module

The Figure 4 represents the system architecture of Data Mining module, where the datasets are collected from various sources for preprocessing. They are loaded into python to transform the raw data for further processing. Datasets undergo ETL(Extract, Transform, Load)process. Data is extracted and cleansed through processes such as filling in missing values(NA), smoothing the noisy data or resolving inconsistencies in the data. Data is transformed into a structured manner in tabular form for analysis. After the data transformation, master dataset is created by combining the structured datasets. The information is extracted from the master dataset to predict future outcomes. Prediction of the road accidents can be done by the following two factors: Time Series Forecasting and Parameters in dataset. Time Series Forecasting involves the study of neural networks and machine learning. Neural network consists of LSTM and MLP algorithms. Parameter based prediction is done by using algorithms like K-mean clustering, Principal component analysis, and SVM techniques. Prediction of future datasets and road accidents is done based on the historical dataset.

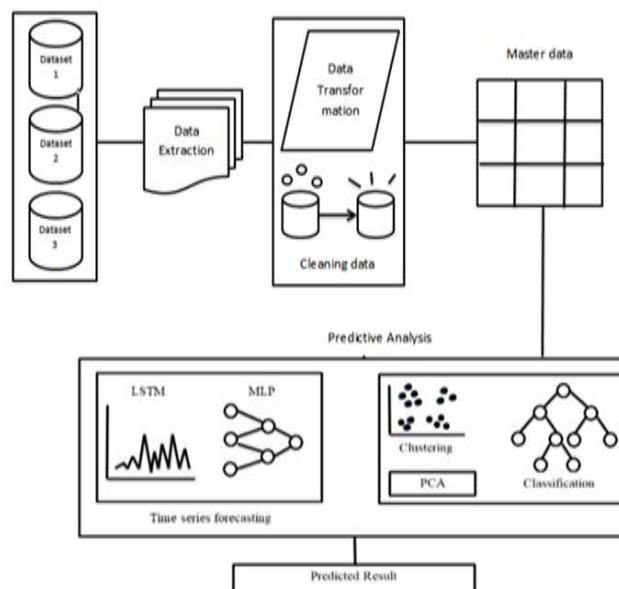


Fig 4 System Architecture of Data Mining

The Figure 5 represents the system architecture of Android Application Development module. It contains a user model where the location of the place, uploading images, and analysis is done. The location of the place is acquired by using GPS. The latitude and longitude of the particular location is displayed in the application. The obtained location with the longitude and latitude is sent to the database server. The images of the potholes ,bumps present on the roads are captured and images are then uploaded to the server with the help of GPS. The images are stored on the database server. The travelers receive the voice notification about the nearing potholes and bumps about 100 meters prior in distance. They are alerted with a message about the road conditions so that they can take measures while travelling. The analysis of accidents and the graphs for the predicted analysis of road accidents can be viewed in the android application. The results of data mining module and the upcoming prediction can be viewed on the tableau dashboard. The working of the application is shown below.

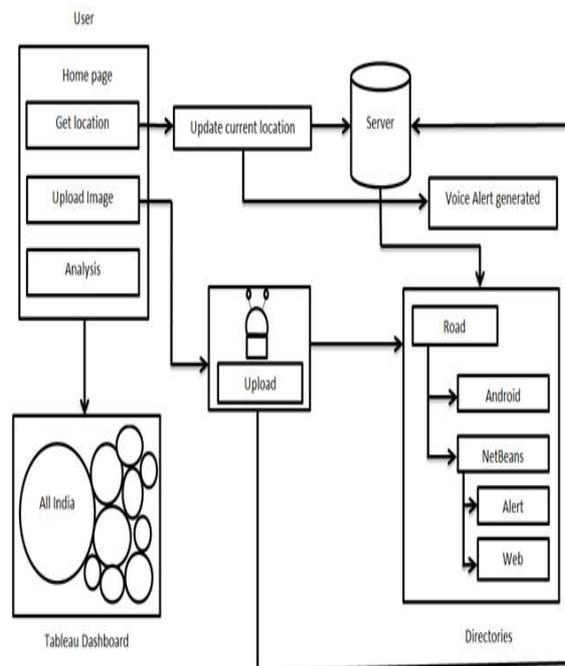


Fig 5 System Architecture of Android

IV. IMPLEMENTATION

Implementation is one of the most important stages of system development life cycle. The stage of implementation includes converting the design phase into a real system using various programming languages and scripting languages. The idea of implementation and the idea of the system will be in the design phase. There are various algorithms and techniques used for predictive analysis like time series forecasting method which uses LSTM(Long Short-Term Memory) and MLP(Multilayer Perceptron),K means clustering, PCA(Principal Component Analysis).

The pseudo code for LSTM is

1. Decide the data parameters.
2. Model those parameters.
3. Load and save the data.
4. Performed stop word elimination and clean data (remove special characters).
5. Save parameters to file.
6. Performed cross validation of data.
7. data for classifier.
8. Used LSTM classifier.
9. Repeat the validation of data until the end of training data.

The pseudo code for MLP is

1. Choose an initial weight vector w
2. Initialize weights w to the input x
3. Linear combination
4. while error did not converge do
5. for all ($\sim x, \sim d$) D d

6. apply $\sim x$ to network and network output
7. calculate $\delta e(\sim x)$
8. end For
9. calculate $\delta E(D)$
10. for all weights summing over all patterns
11. perform one update step of minimization approach
12. end while

The pseudo code for K-Means Clustering is

1. k (number of clusters)
2. D (a set of lift ratios)
3. Output: A set of k clusters
4. Method: Arbitrarily choose k objects from from the D as the initial cluster centers.
5. Repeat:
 - Reassign each object to the cluster to which the object is the most similar, based on the mean value of the cluster.
 - Update the cluster means i.e., calculate the mean value of the objects for each cluster.
6. Until: no change

The pseudo code for Principal Component Analysis(PCA):

1. Subtract the mean from the datasets in all n-directions.
2. Calculate the covariance matrix of this mean subtracted variance.
3. Calculate the eigenvalues and eigenvectors of covariance matrix.
4. Forming a feature vector by selecting the eigenvalues. These are the principal components of datasets.

V. RESULTS

Sports are one of the most important activities or events that will be taking place at colleges. The traditional system of management system was used to manage sports at colleges. As, there were large number of sports and large amount of data to be manage, it was tedious to manage all this data manually. In this, sports event management platform we try to bring out a simple process where all these data about sports and students can be managed automatically by the database.

Some of the snapshots for the working of the admin module are shown in the following figures

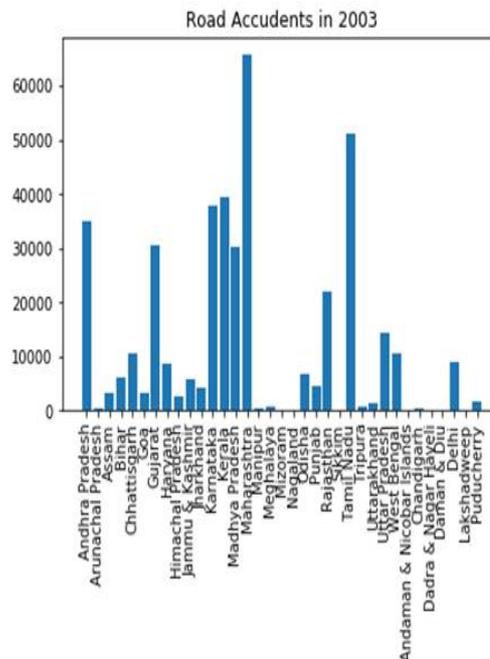


Fig 6 Data Mining

Figure 6 represents the accidents occurred in every state in 2003.

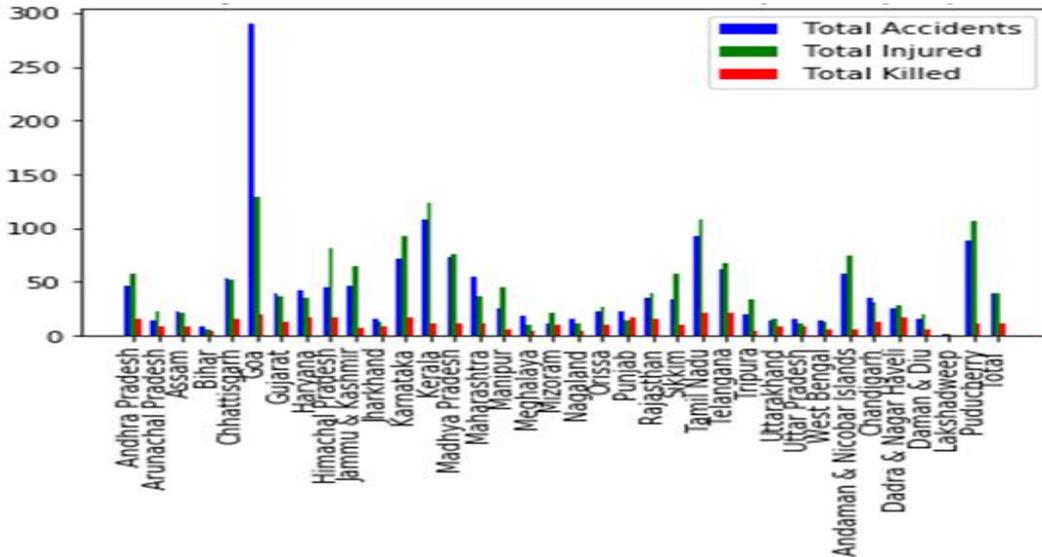


Fig7 Accident based on lane

Figure 7 displays total number of accidents, injured and killed for each state.

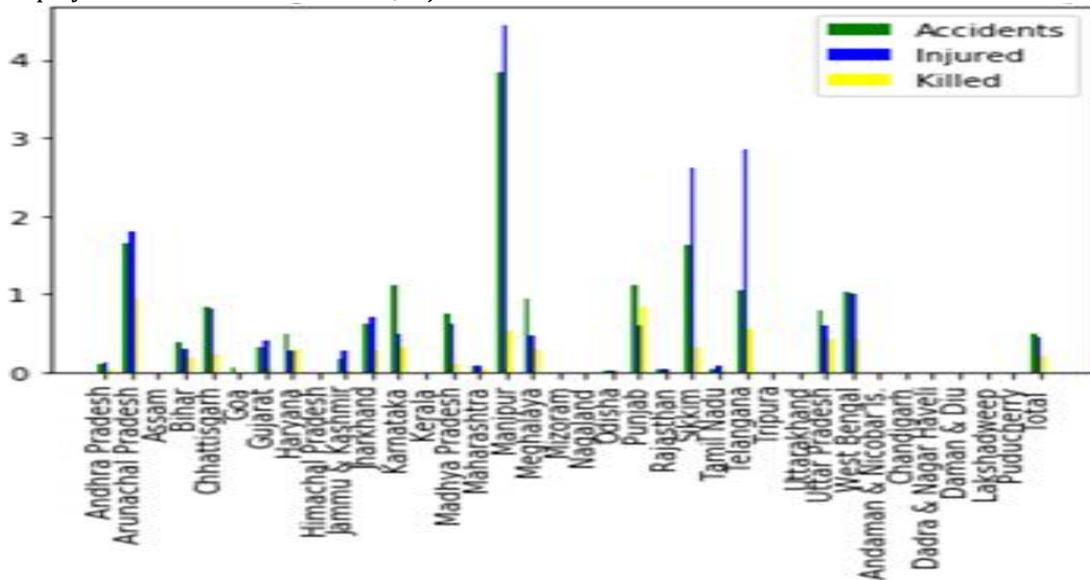


Fig8 Accident based on reason

Figure8 displays number of accidents occurred based on weather conditions

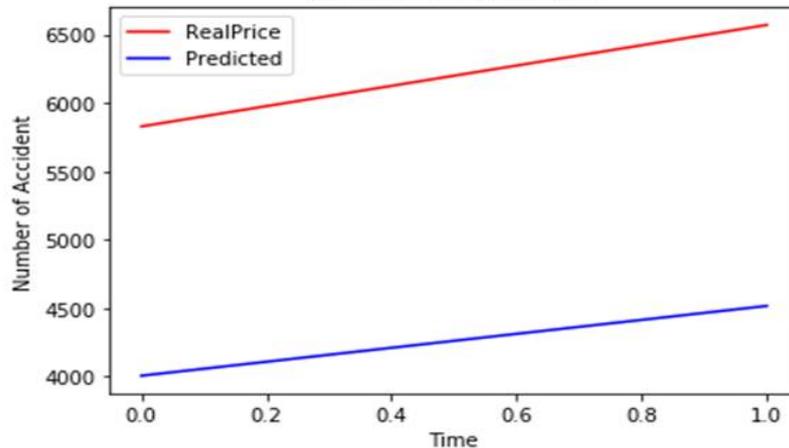


Fig 9 Prediction of accident rate

Figure9 displays the predicted accident rate for upcoming years.

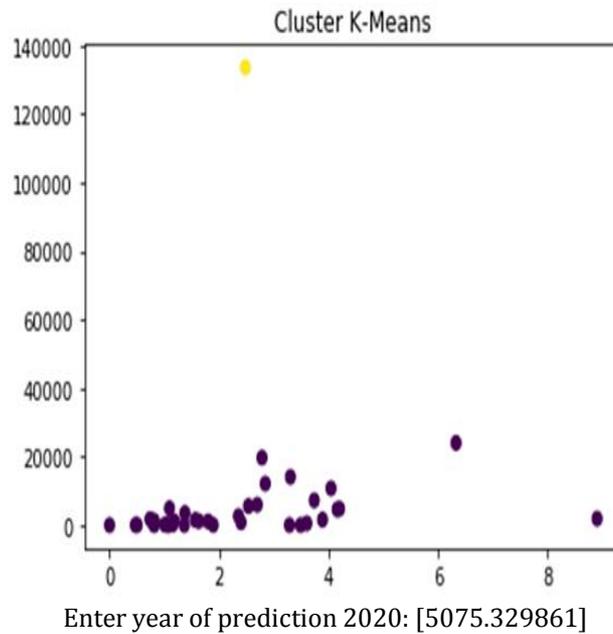


Fig 10 Predicted result

Figure10 displays predicted result for upcoming years



Fig 11 Cover page of application

Figure 11 shows the cover page of android application.

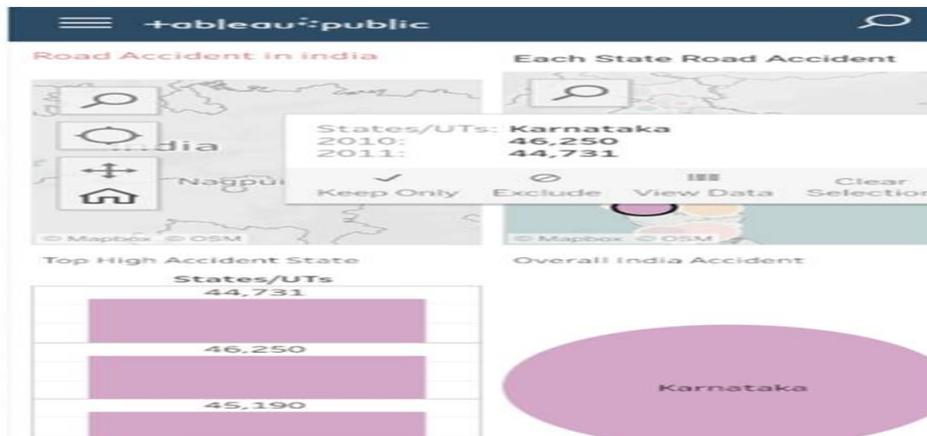


Fig 12 Tableau Dashboard

Figure 12 represents number of accidents and number of people injured in every state.

VI. CONCLUSION

Data mining has been proven as a reliable technique in analyzing critical decision-making factors. The analysis of road accidents using data mining has resulted in creating accurate models to predict the number of accidents that are likely to happen in the future coming years. This prediction alerts the authorities to take measures for maintenance of roads and ensures the safety of the people. Also, with the android application, the predictions can be demonstrated by detecting potholes, bumps which can help the society to take precautionary measures while travelling and avoid dangerous and collision prone areas.

REFERENCES

1. Ms. E. Suganya, Dr. S. Vijayarani, "Analysis of Road Accidents in India using Data Mining Classification Algorithms", International Conference on Inventive Computing and Informatics, 2017.
2. S. Shanthi, R. Geetha Ramani, "Classification of Vehicle Collision Patterns in Road Accidents using Data Mining Algorithms", International Journal of Computer Applications Vol-35, December 2011.
3. Dheeraj Khera, Williamjeet Singh, "Prediction and Analysis of Injury Severity in Traffic System using Data Mining Techniques", International Journal of Computer Applications, 2015.
4. S. Shanthi, R. Geetha Ramani, "Feature Relevance Analysis and Classification of Road Traffic Accident Data through Data Mining Techniques", Proceedings of the World Congress on Engineering and Computer Science 2012.
5. Olutayo V.A, Eludire A.A, "Traffic Accident Analysis Using Decision Trees and Neural Networks", I.J. Information Technology and Computer Science, 2014.
6. Baye Atnafu, Gagandeep Kaur, "Survey on Analysis and Prediction of Road Traffic Accident Severity Levels using Data Mining Techniques in Maharashtra, India", International Journal of Current Engineering and Technology, Vol-7, 2017.
7. Prajakta S. Kasbe, Apeksha V. Sakhare, "A Review on Road Accident Data Analysis Using Data Mining Techniques", International Conference on Innovations in Information Embedded and Communication Systems, 2017.
8. Tibebe Beshah, Shawndra Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road- related Factors on Accident Severity in Ethiopia", International Journal of Current Engineering and Technology, 2015.