Railway Infrastructure and Traveller usage Prediction and Rendering Solutions

Krishna Mohan Ankala, Jyothirmai Kanigolla

Abstract: ***This project introduces the primary establishments of Big Data connected to Smart Cities. An IOT based mechanism is proposed to be connected to various areas. In this project, we are trying to predict and provide the solution to improvise the railway / bus infrastructure and their services. Indian local & state railways or buses are a mode of transport service where thousands of people process every minute. Thus our proposed system involves data collection of the users based on id, username, gender, age, the timing of travel, station source and destination to monitor the user travel behavior. Thus the collected data can be used for analytics and prediction. Predicting the consumer's count and behavior who uses the railway services are solved through the R Programming. The data analytics are performed using R studio. For this work, In R programming, we use K-means algorithm for clustering and use Naive Bayes algorithm for machine learning and solution defining. Finally, the predictive output is sent for public access using shinyapps.io. These results are useful to the travelling systems for giving better services to passengers.***

Index Terms: Clustering, Classification, IOT, Smart Card.

I. INTRODUCTION

A Smart City evolves when the urban framework is developed through information as well as communication technologies (ICT). [2] In urban environments, there is a huge amount of different data sources. A lot of sensors are fixed around urban communities, the majority of them introduced in indoor spaces. This circumstance has brought new examination systems and instruments that give understanding enabling us to have a powerful and community-oriented approach to work the machines. Furthermore, we have numerous mobile data sources like smart-cards, wearable and onboard sensors, in the case of vehicles. All these sensors plays vital role in generating information ,by using the information we can detect mobility patterns. In any case, most existing administration frameworks of urban communities are not ready to use completely and effectively this tremendous measure of information and, therefore, there are substantial volumes of
information, which aren't misused. Toward this path, numerous AI methods in Computer Science have been acquainted with the handling of colossal measure of information to extract valuable data from Raw data, this is known as Big Data.[1]
This project work is intended to understand the interest of big data on smart cities.[5]
The challenges that smart cities still have to face are:

- Individual intelligence and local reasoning.
- Learning and adaptation.
- Privacy of user and security handling mechanisms.
- Integration of sensors and abstraction mechanisms.
- Dynamic human-centric services.[4]

II. EXISTING SYSTEM

The Railways carries more than 24 million passengers over a route of 67,000 km, passing through more than 8,100 railways stations and employing over 1.4 million people. In Existing system, the user prediction is not at all taken into account for improvisation. Railway facilities have not improved mere substantially over the past few decades which leads to congestion and infrastructure damage. Also, there is no initiative to improve the quality of service. In the existing system, traveler data are heterogeneous and blindly improvisation happens which leads to loss of money. No existing system is there which can provide improvisation based on analyzing the traveler demand.
In the existing system, large volumes of variety of data for their use in smart city applications. In this project work, we focus on presenting implementations carried out in railways to achieve energy and time efficiency. The best example is in the public tram service in the City of Chennai, great amount of data generated by the service's transit cards. In this case, big data techniques are applied to extract travelling patterns in public transport of Chennai.[1]

III. RELATED WORK

Step 1 - Input and organize your data in Excel
Organize your data in an Excel worksheet, such that the first row (Row 1) contains the column names and each subsequent row contains all the necessary information for each data point in the experiment [i.e. classification levels and measurement(s)].
Step 2 - Save your worksheet as a comma separated values (.csv) file type .This will be your master file that you can always return to in order to modify things, add new data, etc. Next, to create a version of your data to input into R, click "Save As…" and save it with the appropriate name.
Step 3 - Import your data into R Studio
When we import data into R studio, Smart Object is created and inner objects will be created from columns we given in excel.

Revised Manuscript Received on September 05, 2019.
Dr. Krishna Mohan Ankala, CSE, JNTU Kakinada, Kakinada, India.
Jyothirmai Kanigolla, CSE, JNTU Kakinada, Eluru, India.

```
Smart          25 obs. of 9 variables
 Id : num 1001 1002 1003 1004 1005 ...
 NAME : chr "Rasika.SL" "Tamil Selvi" "Ann Maria .
 AGE : num 21 22 22 21 22 22 21 22 22 21 ...
 SOURCE : chr "Beach" "Beach" "Beach" "Beach" ...
 TIME_S : num 7.45 7 7 6.3 7 7 7.45 7 7 7.45 ...
 AM/PM..6 : chr "AM" "PN" "PM" "AM" ...
 TIME_D : num 9.45 8.3 8.3 8.3 8.3 8.3 9.45 8.3 8.
 AM/PM..8 : chr "AM" "PN" "PM" "AM" ...
 DESTINATION: chr "Chengalpattu" "Kanchipuram" "K.
```

Fig:1 Sample Input data

We use Server.r, Ui.r, Smart.r scripts for user interface ,data processing and smart.r for clustering and classification

Beach	25 obs. of 9 variables
Chengalpattu	15 obs. of 9 variables
Kanchipuram	10 obs. of 9 variables
kmeans_Smart	List of 9
kmeans_Smart2	List of 9
kmeans_Smart3	List of 9
model	List of 4
Smart	25 obs. of 9 variables
Smart_new	25 obs. of 9 variables
Smart_samp	25 obs. of 9 variables
smart1	25 obs. of 1 variable
Smart2	25 obs. of 1 variable
Smart3	25 obs. of 1 variable
test	10 obs. of 9 variables
train	15 obs. of 9 variables
train_class	15 obs. of 8 variables
Values	
confusionMat	'table' int [1:2(1d)] 15 10
pred	Factor w/ 2 levels "Chengalpattu",..: 1
Smart_table	'table' int [1:2(1d)] 15 10
Smart_table1	'table' int [1(1d)] 25
smarttab	'table' int [1:5, 1:2] 6 0 6 5 7 0 1 0
split	logi [1:9] FALSE FALSE FALSE FALSE TRUE

Fig:2 Objects created after running Smart.r

In "Ui.r" module, we use shiny,ggplot2,e1071,caret,catools libraries of R. We have created an user interface in this module where user can input his choice for clustering and classification . For clustering ,user has choice of selecting 4 variables, they are age,id, src_time, dest_time in input panel .User has given chance to give cluster count. Based on user input , K-means clustering is performed in smart.r which is handed over "server.r" and server sends output to user interface in graphical format.

We have classification input labels as source and destination and number of observations to view is given. Based on user input Navie Bayes classification is performed on data and output is sent to server.r and server sends output to user interface of having subsets of data based on given input number of observations to view and suggestions like extra of infrastructure , tickets will be printed in user interface based on traveler data set.

K-Means Clustering We have to clean the data using preconditions. We use K-mean clustering in this work ,to cluster different kind of users based on 4 defined variables ID, Age, Source time and destination time. By this we can predict the number of users going from one particular source to different intermediate stations followed by destination at that particular time. This helps in giving accurate count to travels.We take help of built in R packages to perform Unsupervised learning of K-Means clustering.[4]

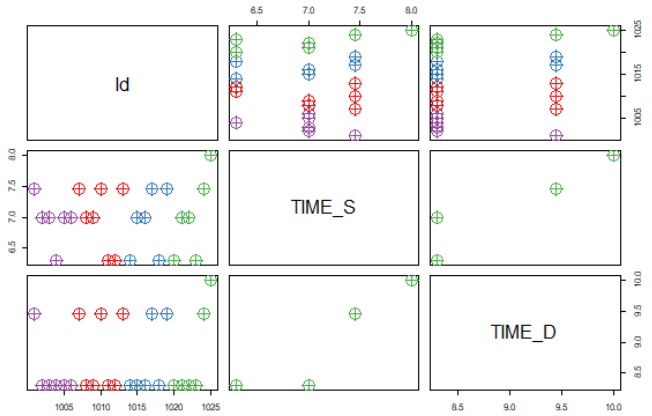

Fig:3 K-Means Clustering of traveler data set of cluster count-4

Similarly various group of parameters we can cluster using this (id, src_time ,dest_time),(age ,id, src_time) etc.

A. *Naive Bayes algorithm*

We use Navie Bayes classification in order to estimate the probability of demand at one particular station at a given point of time based on that we can suggest to travel system the same. We will provide the estimated results based on demand at that particular time. Naive Bayes is a characterization calculation for twofold and multi-class order issues. There is a very strong assumption that is most unlikely in real time data, i.e. that the attributes do not interact. Nevertheless, the method performs shockingly better on information where this suspicion does not hold. A list of probabilities is stored to file for a learnt model .We define class and conditional probabilities for each input value given in training data set.[4]

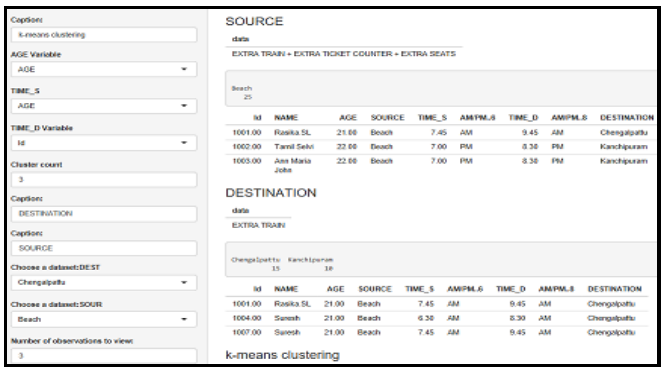

Fig:4 Classification output for traveller data set with number of observations-3

B. Prediction output

With the shiny app, you can upload your text (.csv) files, then we do initial preprocessing of data followed by clustering analysis and view predictions using your own uploaded data.

Benefits of Using Shiny:

1) Ability to use all R functionality and packages without any constraints.

2) Open-source R and JavaScript help to create highly customizable applications.

3) Can expose model parameters using controls like sliders, text fields, and drop-down lists.

4) For primary use-cases, only having R knowledge is enough to get by.

5) Easy to handle events using reactive components.
6) Quickly deploy and share Shiny applications.

```
Confusion Matrix and Statistics

              Reference
Prediction     Chengalpattu Kanchipuram
  Chengalpattu            8           0
  Kanchipuram             0           3

               Accuracy : 1
                 95% CI : (0.7151, 1)
    No Information Rate : 0.7273
    P-Value [Acc > NIR] : 0.03011

                  Kappa : 1
 Mcnemar's Test P-Value : NA

            Sensitivity : 1.0000
            Specificity : 1.0000
         Pos Pred Value : 1.0000
         Neg Pred Value : 1.0000
             Prevalence : 0.7273
         Detection Rate : 0.7273
   Detection Prevalence : 0.7273
      Balanced Accuracy : 1.0000

       'Positive' Class : Chengalpattu
```

Fig:5 Result Analysis of confusion matrix and statistics

IV. ARCHITECTURE DIAGRAM

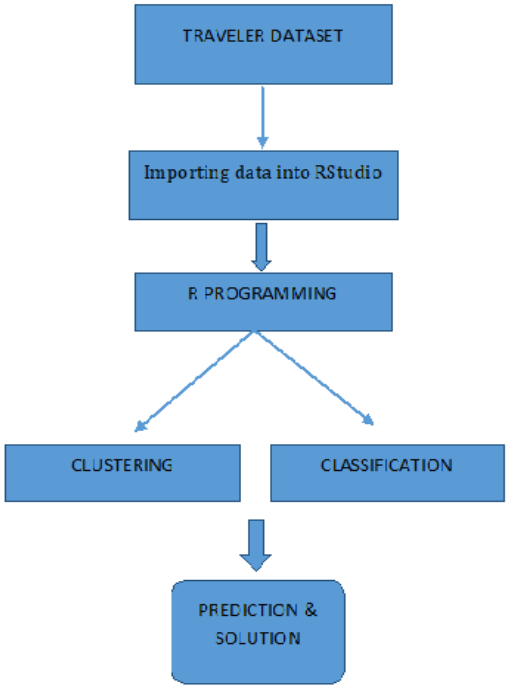

Fig: 6 Architecture Diagram of proposed method

V. CONCLUSION

This paper focuses on applying big data techniques over railway/bus traveler data deployed in smart cities. Thus, this project collects the real time data and process using big data analytics tools /libraries supported by R Studio. The predicted result would provide an efficient solution for railway/buses department to improvise the service and infrastructure. Initially, we have demonstrated that pattern recognition carried out using data related to the public tram service of the city of Chennai, tests were performed for the travel systems in order to find how passengers utilizing transportation system. On the other hand, in the case of the public tram service, data coming from crowd sensing vendors will be merged to improve the findings of the urban mobility patterns. Our experimental results using R Studio, which provides high-efficiency results as shown above in fig-3.

In future perspective, we can design a smart transit card where it tracks user moves by integrating with sensor, it could have unique id , it would be easy to know user behavior in terms of travelling ,proposed transit cards to detect the consumer behavior and count. The real sensor data are transmitted into systems using COM
port and these real time data are obtained using Net beans and saved in SQL database, as that data is more accurate so we can give accurate results by using data analysis tool.

REFERENCES

1. N. Komninos, “Intelligent cities: variable geometries of spatial intelligence,” Intelligent Buildings International, vol. 3, no. 3, pp. 172–188, 2011.
2. L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010.
3. L. Da Xu, W. He, and S. Li, “Internet of things in industries: a survey,” Industrial Informatics, IEEE Transactions on, vol. 10, no. 4, pp. 2233–2243, 2014.
4. R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, “Big data analytics: Computational intelligence techniques and application areas,” Int. J. Inf. Manage, pp. 10–15, 2016.
5. M. Victoria Moreno, Fernando Terroso-Saenz, Aurora
6. Gonzalez-Vidal, Mercedes Valdes-Vela et al.
7. "Applicability of Big Data Techniques to Smart Cities
8. Deployments", IEEE Transactions on Industrial Informatics, 2017

AUTHOR PROFILE

Dr. Krishna Mohan Ankala, He received his PhD degree in CSE and he has 14 years of teaching experience., research interests include Data Mining and Big Data, web technologies, software Engineering. He received best teacher award based on outgoing student evaluation .He published many Research papers.

Jyothirmai Kanigolla, She is pursing final year of M.Tech degree in JNTUK in the stream of CSE. She graduated in B.Tech degree from Rama Chandra College of Engineering, Eluru .She did her project work on Gene Ontology, in data analytics in B.TECH in 2017. Her research areas are Data Analytics, Data Mining.