ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN INTERACTIVE SYSTEM FOR ANALYZING MOVEMENT OF BUSES IN HAJJ

[1]EMAD FELEMBAN, [2]FAIZAN UR REHMAN, [3]ASAD ALI BIABANI, [4]ATIF NASEER, [5]OMAR HUSSAIN, [6]EHSAN ULLAH WARRIACH

[1] Computer Engineering Department, College of Computing and Information Systems, Umm Al-Qura University, Saudi Arabia
[2] Institute of Research and Consultation Studies, Umm Al-Qura University, Makkah, Saudi Arabia
[3] Science and Technology Unit, Umm Al-Qura University, Saudi Arabia
[4] Science and Technology Unit, Umm Al-Qura University, Saudi Arabia
[5] Science and Technology Unit, Umm Al-Qura University, Saudi Arabia
[6] Rochester Institute of Technology, United Arab Emirates
E-mail : [1]eafelemban@uqu.edu.sa, [2]fsrehman@uqu.edu.sa, [3]sabiabani@uqu.edu.sa, [4]anahmed@uqu.edu.sa, [5]omarkazem@gmail.com, [6]euwcad@rit.edu

ABSTRACT

Analyzing traffic behaviour performs a robust part to accomplish day-to-day activities of the users efficiently and effectively, particularly in densely populated urban areas. Activities drawing in huge numbers such as religious pilgrimage or sporting events, collisions among automotive traffic flows are likely resulting in interruptions and unsafe situations for pilgrims, often creating chaos and congestion. The issue is more ambitious as millions of pilgrim's travels over a fixed period throughout Hajj from one location to the other. Hajj is an Islamic pilgrimage which occurs annually during the month of Dhul al-Hijjah[1] which begins on the eighth day of the month and last for five days where millions of pilgrims from across the globe assemble in Makkah to perform the Spatio-temporal rituals. This article presents an interactive platform that utilizes large-scale GPS traces to know the motion of buses during Hajj. For a period of 60 days, GPS traces are gathered for over 20,000 plus buses for pilgrim's mobility to conduct their rituals. To save big spatial data, we developed an interactive platform that can help in analyses and visualization. The analyses have been done for various stakeholders which includes company authorities. They can visualize the movement of buses, identify driver's behaviour, speed violations, and area of the violations using maps-based visualization. Extracted expertise would be of benefit to stakeholders to improve the accessibility of pilgrims throughout the Hajj by creating a smart transport system like scheduling, evacuation, maximizing the number of buses or roads for each establishment.

Keywords: *GPS Data, Trajectory, Hajj, Categorization, Big Data, Clustering*

2. INTRODUCTION

Annually millions of Muslims from across the globe visit the holy Makkah city to perform Hajj, which is a mandatory pilgrimage for every individual who is healthy and financially affords it, once in their lifetime. Figure 1 shows the map of Mashaer[1] area including Mina, Muzdalifah and Arafat to perform the Spatio-temporal rituals. For different Hajj rites to be followed, pilgrims cross by different routes, of Makkah to Mina during Hajj days. On the aurora of 9th day of Dhul Hijjah pilgrims travel to Arafat and spend their day performing rituals till the time of sunset and proceeds to Muzdalifah and remain there till the next morning and then continue for further ritual's. These mass transportation of pilgrims restricts the region results in a severe degree of movement blockage at the same time and span.

During major activities like sports, religious groupings, a smooth and seamless traffic flow plays a vital role. Inappropriate traffic activity frequently contributes to prolonged delays and is a concern if not supervised and scheduled. Controlling an immense flow of crowd movement is a tough job for the officials to make appropriate conclusions. Besides, effective handling of pedestrians and traffic facilities needs constant observation of these components but also the distribution of regulation plus management measures, knowledge and instructions to drivers.

[1] 12th month of an Islamic lunar calendar

A case experiment for Hajj-pilgrim traffic management and regulation is presented in [1] manifesting a need to maximize the movement of pilgrim-traffic through the scheduled ritual places during Hajj shows its primary significance for the pilgrims' welfare. It is vital to analyze and evaluate mobility behaviour, particularly in Hajj periods, to control the transport network.

Pilgrims from around the world, based on their nationality, are split into establishments. The Mashaer area is divided into seven establishments where each establishment has its camps, routes and parking spaces to handle the movement of pilgrims as shown in figure 1. Connection to the internal roads in the Mashaer region is blocked to control the crowd and prevent a stampede, and the pilgrims are divided, in seven establishments. Thus, every establishment is designated to perform Spatio-temporal rituals in the Mashaer area, with a particular area and itineraries for their motions. Doing so will facilitate decision-makers to make effective decisions during an emergency.

Through this article, we provide an interactive platform for stakeholders to consider decisions about the pilgrims' improved movement throughout their stay in Makkah. We obtained the GPS traces from 20,000 plus buses for numerous pilgrim movements performing Spatio-temporal rituals. The data generated has been transferred to the Big Data framework to allow the system effectual and secure. Later the data has been analyzed considered the use cases for different stakeholders. Starting from driver's behaviour, movement of buses, violation of speeds, start and end points lots of analysis has been done. Interactive visualization for end-users to perceive the traffic activity at various hours of the day throughout the Hajj season is presented, by the system.

This article provides an expansion of our previous research [2], [3] by focusing on analysis and an interactive platform to visualize the bus travel. Section 2 describes the related work on GPS data, analysis and Big Data. Section 3 and Section 4 draws the data management and platform overview respectively. Section 5 focusses on the implementation and results. Section 6 presents data analysis; while Section 7 highlights the conclusion and future challenges.

2. LITERATURE REVIEW

Controlling transportation with such a massive crowd with minimal capacity is a vigorous activity, around the same period. Continuous traffic jams are an indication of traffic obstruction in a religious event such as Hajj in which interference with daily traffic flow is usual. Organizing and controlling crowd movement and enriching the travelling experience for the thousands of pilgrims is a challenge for the officials. Scientists are applying to technology such as GPS, Big Data Analysis to test new approaches and methods for optimizing transport system.

2.1 GPS Data and Analysis

Pilgrims journeying by bus during Hajj, get caught in a massive queue of traffic on their route to Arafat, followed by Muzdalifah and back to Mina. The increase in the number of buses and pilgrims in the last ten years has caused inadequate functioning of roads. Identification of travel mode is a vital step towards the detection of travel information with a global positioning system (GPS). Data consistency evaluation is part of the data quality test. General data output is distinguished by consistency satisfaction, completeness status, reliability, minimal [4] and is intended to address data loss, reduction, and dispute. As the Intelligent Transportation Systems (ITS) progresses, traffic data mining is becoming increasingly important [5], [6]. In recent years, the GPS proves vehicle data has shifted as a valuable data source for ITS due to extensive utilization of GPS technology. But because of the volatile estate, the stability of GPS traffic data is a challenge for data mining. Authors in [7] using appropriate data density and data ideality to explain the data quality of GPS traffic and propose a multidimensional cube approach to data quality.

The GPS systems are used to a great extent world-wide. It could provide a longterm consistent positi on and provide appropriate semipermanent robust precision for the location data [8]. Moreover, it is vulnerable to disruption outside, such as electronic jamming, and impossible to obtain a signal within a tunnel [8], [9]. For overcoming this flaw, GPS or Integrated Navigation System (INS) is used [10]. A novel back propagation neural network was proposed to result in the outages of INS/GPS by authors in [11]. There are many areas of problems reaching them for various circumstances and reasons such as its topology, atmosphere and nature formations. GPS technology allows for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

remote monitoring and data acquisition. In [12] a GPS data acquisition and analysis software are proposed over the IP platform.

GPS precisely quantifies the location and time of the tawaf (circumambulation of the Kaaba seven times) [13]. A spatial-temporal representation of the motility of the pilgrim from Mina to Jamarat is explored in [14], which can be used to define the constrictions confronting the pilgrim movement. There are several methods for implementing its road quality assessment that uses a sensor to collect data from smartphones. Authors in [15] use smartphone sensors to evaluate driving incidents and road phenomenon. They used smartphone accelerometer mounted on the dashboard to collect road and driving information. Such results are observed and analyzed by testing the parameters for different patterns. Authors proposed in [16] data study of road condition based on accelerometer and mobile GPS sensors. Approximation of the speed of travel on a road network obtained from public transit device GPS data using machine learning techniques is recommended [17]. Authors in [18] proposed an application for a method of collecting data using GPS for the freight route which defines the characteristics of the overall route along with the environmental impact of highways and the driver's behaviour. The prediction of congestion in traffic based on GPS trajectory data is proposed by the authors in [19]. A method to identify activities and trips from the GPS data, authors in [20] provided a research scheme to detect the information of the mode of travel and its purpose. A solution based on GPS data for traffic flow prediction is presented in [21] which comes with a scalable traffic simulator application. Authors in [22] develop an expert traffic congestion detection system and accidents from real-time GPS data is obtained from GPS trackers or the smartphone of the driver.

2.2 Big Data

In both research and technology, Big Data is a popular topic. It contains large complex sets of data derived from all forms of information. Big Data technologies, including data mining, machine learning, artificial intelligence, data fusion and social networking, are among the most common data processing techniques. Big Data is now a subject of intelligent transportation systems (ITS) analysis which is seen in several research projects around the world. Authors in [23] analyze the past and features of Big Data and ITS, examine issues relevant to the use of analytics in ITS.

To automate network operations and improve the efficiency of system data depository, Big Data technologies can be incorporated into the urban smart transport system. It can solve problems and determine the sources of a rapid reduction in error costs through the centralized collection of massive data. It provides a robust database for individual analyzes by data mining techniques to explore the connection between the massive volumes of data. It discusses the value and functionality of Big Data technologies and intelligent public transportation system through GPS technology in [24].

Due to advances in Big Data and IoT (Internet of Things), ITS can provide users with more opportunities to deliver smart services [25], [26]. The ITS has generated a large volume of Big Data on traffic in recent years that poses a problem of scalability if road traffic data, is being investigated. Authors use a MapReduce approach to overcome it in [27], [28], [29]. Precise visualization of Big Data traffic can allow users to find traffic information effectively, such as analyzing weather trends and relationships. In [30] the authors suggested a timeline analysis for bigdata traffic. To enhance the traffic forecast information platform in real-time, a dynamic predictive algorithm on Big Data analysis is proposed in [31].

For resolving transportation problems Big Data for socially transportation brings novel opportunities. To tackle realistic problems in transportation there is a need for substantial information on transit systems like visual analysis, crowdsourcing and data-based services. Authors in [32] provided a summary of data sources, analytical approaches, and application systems for social transportation.

In the areas of transport engineering and automotive sector, each automobile on the roadway may be viewed as a different data source that can provide massive data. Big Data is widely used as a paragliding concept when addressing data obstacles like quantity, range, velocity, veracity and value. Authors in [32] propose a distributed data model that solves the problems by depending for each of the data processing steps on a specific design option.

Hajj is a unique event where millions of pilgrims move from one place to another at the same time to perform specific rituals. Existing state-of-the-art discussed mining and analysis of the GPS data of buses and taxis particular to the city. In the case of Hajj, buses on both sides of the roads are moving in the same direction which makes it different from the existing approach and impose multiple challenges.

3. BIG DATA COMPONENTS

Figure 2: Big Data Framework

Data is the source of developing any system. Mostly the system is dealing with multiple sources of data as an input. The data comes from different channels using different formats and velocity. The Big Data framework steps are shown in figure 2. Data Collection is one of the primary building blocks. There are many sources of data like sensors, cameras, and IoT devices that can be used to collect the data. The collected data is stored in many formats at multiple locations. The data storage can be on the local servers or remotely on clouds. The key to the data management pyramid is data pre-processing. In pre-processing, the irrelevant data is removed, and missing data is filled correctly. Data transformation is necessary for data mining as it helps in data normalization, attribute selection, discretization, and hierarchy generation. Data reduction some time requires on the large scale of data as the analysis of a vast amount of data is sometimes harder. Data reduction increases the efficiency of storage and the analysis cost. The steps required for data reduction are Data cube dimensionality reduction, aggregation, numerosity reduction, and attribute subset selection. Data can be presented in various forms depends on the type of data we are using. The data can be shown in organized tables, charts, or graphs.

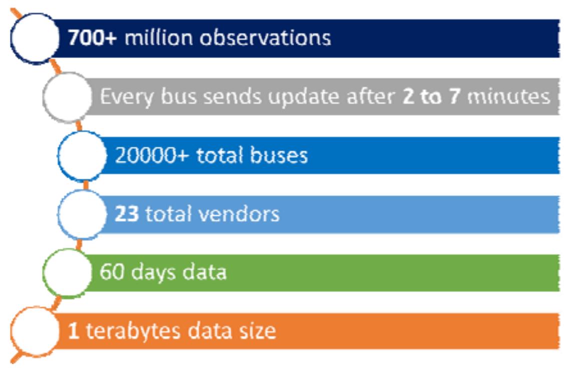


Figure 3: Summary of Dataset

3.1 Data Collection

Data is collected during the season of Hajj by automatic vehicle location (AVL). Ministry of Hajj and Umrah collected data of 20,380 buses using almost 10 AVL service providers. We received the data from the Data source (Naqaba), who is the official transport authority of the Ministry of Hajj and Umrah. The data collected from Naqaba is approximately 737,145,410 location history records. These locations contain the data of 20,000 buses, the companies that are operating these busses, offices of companies, garage, establishments, zones, and movement stages, etc. arranged in almost 98 relational tables. Table 1shows the location history that has record_id, bus_id, company_id, angle, latitude, longitude, GSM signal, speed, and record_time. Figure 3 shows the detail of the dataset that was used in this paper. The data available in MYSQL is not cleaned correctly. A lot of data is redundant in most of the tables that required cleaning and pre-processing.

Figure 4 shows the relationship of busses, establishments, offices and zones. This relational data is crucial to calculate the number of busses according to each establishment and their routes. The data available in MYSQL is not cleaned correctly. A lot of data is redundant in most of the tables.

Journal of Theoretical and Applied Information Technology
15th November 2020. Vol.98. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

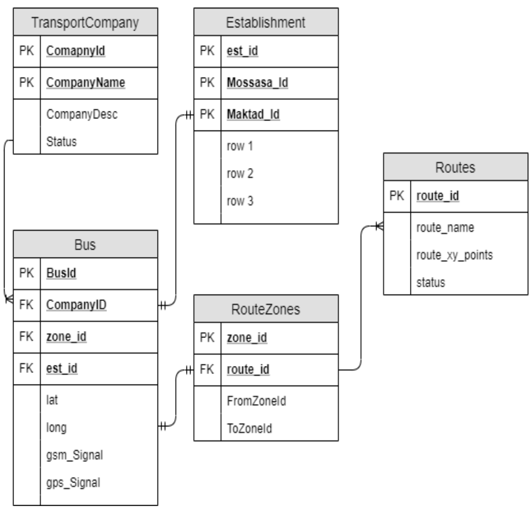


Figure 4: ER Diagram of the partial Schema

3.2 Data Storage

The collected data is initially stored at multiple locations. The AVL data of almost 20,000 buses are stored separately into the MySQL database. The data is then stored in the MS SQL server dump format. To retrieve and access this dump data took a long time as a basic query require approximately 12-15 minutes or higher to recover the data.

3.3 Data Pre-processing

In any data analytics application, pre-processing is the most critical and work-intensive step. The pre-processing stage is used to format the data in a standard way as most of the raw data is not always in a required format. The buses data that we acquired for experimentation has redundancy and errors. Also, the available data has a lot of incorrect and blank data. So, pre-processing is applied to the collected data. The data manipulation algorithms are applied to the data that cuts the required variables into ranges. Some data cleansing techniques were also applied the correct the missing and incorrect information.

3.4 Data Transformation

As compared to data pre-processing, data transformation is also an important step before the actual analysis. The data transformation transforms the current data into a new set of attributes. These newly created attributes are essential for analyzing the data. We acquire almost 20,000 busses records that have some missing attributes that are required for analysis. Like, we apply the cleansing and extract some new features like bus travelling distance (in Kilometers), the distance between the buses, and the speed of each bus between a specific time. The travelling distance, speed, and time are essential to calculate the total distance travelled by each bus in a specific area under each establishment and total time a bus is parked at some particular parking area. After applying the data transformation on the available data, we can now easily identify the bus route, parking area, bus station point, and the routes assigned to each establishment and offices.

4. PLATFORM OVERVIEW

The Naqaba data is saved in MS SQL server dump format and a basic query require approximately 12-15 minutes or higher to recover the data. To increase the efficiency of the platform, we transferred the data from MS SQL server to the Big Data platform through the Cassandra cluster. Figure 5 shows the high-level view of a Big Data platform that we developed to analyze the data of buses.

Data lake layer encompasses Master data service and MS SQL service. The master data is used for every relation to visualising the data on certain circumstances like establishment, offices or bus number. The MS SQL consists of the original data that we received from Naqaba.

Big Data layer consists of a Cassandra cluster and Big Data Aggregation service. The location history data is transferred to the Cassandra cluster for cleaning and removing noise using ETL engine. The reason for using Cassandra cluster is to improve the competency and scalability using a distributed, wide column store, NoSQL database management system. The Big Data aggregation service is an amalgamation of Hadoop and Presto. Presto is highly competent for distributed SQL analytical queries on data in the Hadoop distributed file system (HDFS) and is utilized by a big giant like Facebook. The benefits of using Hadoop for batch-based analytics and Cassandra is for time-based queries.

REDIS cache is an open-source (BSD-licensed), in-memory data structure store used as cache. It lets storing keys and values pairs with a large number of data into its cache effortlessly. The system's performance can be improved by Redis cache.

Naqaba REST API enlists the APIs to handle the request of front-end. The front-end calls the API and it fetches the data from Master data

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

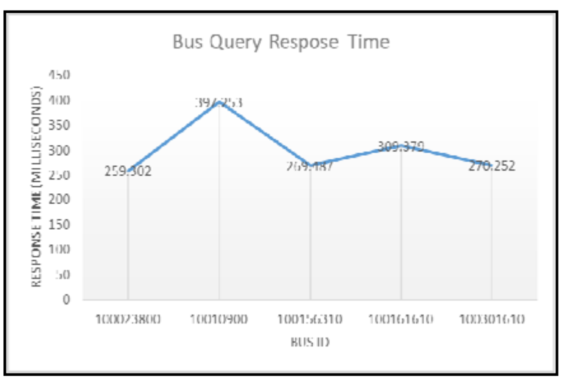
services, Big Data aggregation service or REDIS cache and returns the results that are visualised in front-end.

The Naqaba API server provides the front-end visualization of data. This interface allows user to select establishment (Mosasa), offices (Maktab), number, company, bus number and route.

5. IMPLEMENTATION AND RESULTS

The Big Data-based interactive platform has been developed to visualize the movement of 20,000 buses and various analyses such as identify driver's behaviour, speed violations, and area of the violations using maps-based visualization. The front-end implementation is done in leaflet JavaScript on the top of google map's layer. The back-end and ETL's are coded in Python. The REST APIs are implemented in PHP to handle the queries of the application server. The Cassandra cluster is running on two nodes with 4GB RAM and 100GB storage. The nodes are running on the workstation with 64 GB RAM, 2 TB SSD storage and Intel Xeon CPU E5-1620 v2 @ 3.70 GHz processor. The platform is scalable i.e. can store the data of another hajj for comparison or prediction using machine language algorithm, on increasing the nodes in Cassandra cluster.

The 20,000 buses data after migrating to Big Data
the platform took around 300 milliseconds with two nodes of Cassandra as compared to a couple



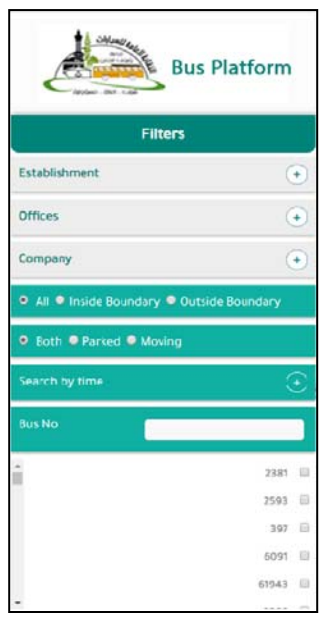
of

Figure 6: Buses Query Time on our Platform

minutes in MS SQL server. Figure 6 shows the query response time on bus data after migrating to our platform. The figure shows the data of 5 buses with their ids and time in milliseconds to fetch their locations.

Figure 7 displays the front end of the interactive platform that was developed to visualize the GPS traces data of buses in Hajj 2018 (1439H). The end-user can click on any bus number to see its movement. The filter panel of the interactive platform shown in figure 8 is provided to facilitate users to filter buses data, based on their desired requirements like buses assigned to each establishment, offices and company details, bus location (inside and outside the boundary) and its movement (parked or moving), search by time and bus number.

A visualization of the GPS data points for the buses assigned to the South-East Asia Establishment in Hajj 2018 is shown in figure 9. The other prominent features of the platform that visualize data of buses are shown in figure 10 that includes(a) no. of buses assigned to each office, (b) data records for the buses owned by the user's specified company, (c) extraction and visualization of the data of all buses within the user's specified boundary and its last location that are in the Mashaer (Mina, Muzdalifah and Arafat boundaries)and outside. The data points in the red



indicate the parked or still vehicles while in green are moving.

Figure 8: Filter Panel of the Interactive Platform

The platform provides an interface to visualise the movement of an individual bus by clicking on the bus number as shown in figure 11. It also allows users to simulate the historical data of individual bus by selecting the start and end time to play the simulation as shown in figure 12.

Journal of Theoretical and Applied Information Technology
15th November 2020. Vol.98. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6. DATA ANALYSIS

In this section, we analyze the bus speed data to understand the driver's behaviour, violations, and area of the violations using maps-based visualization. The speed profile for any bus driver for the entire period of Hajj can be shown by the timeline as in figure 13. This figure demonstrates the speed patterns for Bus No. 1 and its maximum and minimum reached speeds in Hajj 2018, that are 102 and 81 km/h respectively.

The violation of buses to the road speed limit is very likely to happen. Thus, identifying bus speed violation is accomplished by observing the continuous number of observations that exceeded speed limit threshold of 80 km/h with starting & ending timestamp, then by calculating the duration of violation from starting & ending timestamp in minutes & seconds as shown in table 2. Now we classify each violation into different categories concerning their duration, as in table 3, for detailed analysis.

Using the violation classifier and speed data, we identified the worst and best bus drives in Hajj 2018. The drivers of buses No. 152 and 132 are the worst in terms of violations as shown in figure 14, whereas the drives of the buses No. 4038 and 4004 are the best as shown in figure 15.

Figure 18: Bus violations heatmap

In terms of companies performance towards violations, as seen in figure 16, that the companies with ID 4, 15, and 21 are the worst performed companies in Hajj 2018, and companies with ID 19, 10, and 13 are the best-performed companies with only a few violations as shown in figure 17.

We visualize the locations and intensity of bus violations using the Heat and Geo maps as shown in figure 18 and figure 19. These interactive maps can display the intensity and number of violations at any specified area on the map. For example, in Fig. 10 we can see that the bus has violated 145 times in the area around the exit to Jeddah on the Makkah-Madinah highway.

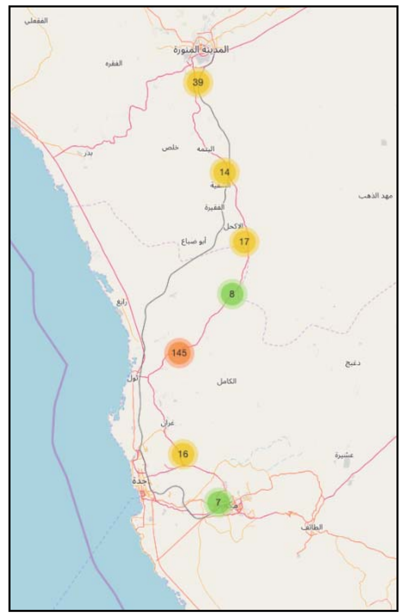


Figure 19: Bus violations frequency

7. CONCLUSION

The annual Hajj pilgrimage at Makkah, Saudi Arabia is a challenging event where millions of Muslims assemble at the said date and time to conduct their rituals. Every year a new initiative is being made to fight the stampede that can arise because of strong crowd movement. We present an interactive platform to grasp the bus movement using GPS traces on a large scale. As a prototype, in the Mashaer area during various hours of hajj days, we analyzed the data and defined the routes, parking spaces allocated to each establishment and travel time of routes for different establishments. We identified the best and the worst drivers and companies in Hajj 2018 using the violation classifier and performance of the companies towards violations. Big spatial data can be stored and visualize using our platform which can be of great benefit to the stakeholders and decision-makers in developing an intelligent transportation system. Future work involves forecasting the use of machine learning algorithms, further analytics like the pick-up and the drop off time, travel time

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

for each shuttle service trip for an individual establishment etc.

ACKNOWLEDGEMENT:

This work is funded by grant number NSTIP-10-INF1235-10 from the Long-Term National Plan for Science, Technology and Innovation (LT-NPSTI), the King Abdul-Aziz City for Science and Technology (KACST), Kingdom of Saudi Arabia. We thank the Science and Technology Unit at Umm Al-Qura University, Makkah 21955, Saudi Arabia for their continued logistics support. We would also like to thank Eng. Osama Fatheldin from The General car Syndicate (Naqaba) for providing us with the buses data.

REFERENCES:

[1] O. Tayan and A. Al-binali, "Evaluation of a Proposed Intelligent Transportation Framework using Computer Network Concepts : A Case Study for Hajj-Pilgrim Traffic Monitoring and Control," vol. 10, no. 2, pp. 325–331, 2013.

[2] E. Felemban, F. U. Rehman, A. A. Biabani, A. Naseer, and U. Alabdulwahab, "Towards Building an Interactive Platform for Analyzing Movement of Buses in Hajj," in *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, Dec. 2019, pp. 3775–3778, doi: 10.1109/BigData47090.2019.9005521.

[3] E. Felemban, F. U. Rehman, H. Wadood, and A. Naseer, "Towards Building Evacuation Planning Platform using Multimodal Transportation for a Large Crowd," in *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, Dec. 2019, pp. 4063–4066, doi: 10.1109/BigData47090.2019.9006226.

[4] D. Aebi and L. Perrochon, *Towards improving data quality*. 1993.

[5] S. M. Turner and L. P. Albert, "ITS Data Quality Control and the Calculation of Mobility Performance Measures," 2000.

[6] L. Tang *et al.*, "Multidimensional Traffic Data Analysis : A Congestion Cube Approach," pp. 1–26, 2010.

[7] C. Zhang, Z. Zheng, F. Zhang, and J. Ren, "Multidimensional traffic GPS data quality analysis using data cube model," in *Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering, TMEE 2011*, 2011, pp. 307–310, doi: 10.1109/TMEE.2011.6199204.

[8] J. Park, D. Lee, and C. Park, "Implementation of Vehicle Navigation System using GNSS, INS, Odometer and Barometer," *J. Positioning, Navig. Timing*, vol. 4, no. 3, pp. 141–150, Sep. 2015, doi: 10.11003/jpnt.2015.4.3.141.

[9] J. A. Farrell and M. Barth, *The Global Positioning System & Inertial Navigation*, no. 4. 1999.

[10] A. Budiyono, "Principles of GNSS, Inertial, and Multi-sensor Integrated Navigation Systems," *Ind. Robot An Int. J.*, vol. 39, no. 3, 2012, doi: 10.1108/ir.2012.04939caa.011.

[11] G. Wang, X. Xu, Y. Yao, and J. Tong, "A Novel BPNN-Based Method to Overcome the GPS Outages for INS/GPS System," *IEEE Access*, vol. 7, pp. 82134–82143, 2019, doi: 10.1109/ACCESS.2019.2922212.

[12] H. R. Chamorro, A. M. Cruz, F. Niño, R. Ferro., E. Gomez, and E. L. Cantor, "GPS data analysis and acquisition software over IP platform," in *2011 IEEE 12th Annual Wireless and Microwave Technology Conference, WAMICON 2011*, 2011, doi: 10.1109/WAMICON.2011.5872906.

[13] N. Koshak and A. A. M. Fouda, "Analyzing Pedestrian Movement in Tawaf Using GPS and GIS," *Proc. Remote Sens. Arab.*, no. July, 2008,

[14] N. Koshak, "A GIS-based spatial-temporal visualization of pedestrian groups movement to and from Jamart area," in *Proceedings of Computers in Urban Planning and Urban Management (CUPUM '05) Conference June 29-July 1, 2005, London, UK*, 2005, pp. 1–12, [Online]. Available: http://www.cadgis.com/n/publications/jamarat_movement_simulation.pdf.

[15] N. Kalra, G. Chugh, and D. Bansal, "Analyzing Driving and Road Events via Smartphone," *Int. J. Comput. Appl.*, vol. 98, no. 12, pp. 5–9, 2014, doi: 10.5120/17233-7561.

[16] B. Y. Amirgaliyev, K. K. Kuatov, and Z. Y. Baibatyr, "Road condition analysis using 3-axis accelerometer and GPS sensors," in *Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings*, Jul. 2017, doi: 10.1109/ICAICT.2016.7991662.

[17] T. Kyaw, N. N. Oo, and W. Zaw,

ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

"Estimating travel speed of Yangon road network using GPS data and machine learning techniques," in *ECTI-CON 2018 - 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, Jan. 2019, pp. 102–105, doi: 10.1109/ECTICon.2018.8619908.

[18] P. Pluvinet, J. Gonzalez-Feliu, and C. Ambrosini, "GPS Data Analysis for Understanding Urban Goods Movement," *Procedia - Soc. Behav. Sci.*, vol. 39, pp. 450–462, Jan. 2012, doi: 10.1016/j.sbspro.2012.03.121.

[19] S. Sun, J. Chen, and J. Sun, "Traffic congestion prediction based on GPS trajectory data," *Int. J. Distrib. Sens. Networks*, vol. 15, no. 5, p. 155014771984744, May 2019, doi: 10.1177/1550147719847440.

[20] Z. Fang, J. Y. Lv, J. J. Tang, W. Xiao, and G. Fei, "Identifying activities and trips with GPS data," *IET Intell. Transp. Syst.*, vol. 12, no. 8, pp. 884–890, Oct. 2018, doi: 10.1049/iet-its.2017.0405.

[21] E. Necula, "Dynamic Traffic Flow Prediction Based on GPS Data," in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, Dec. 2014, vol. 2014-Decem, pp. 922–929, doi: 10.1109/ICTAI.2014.140.

[22] E. D'Andrea and F. Marcelloni, "Detection of traffic congestion and incidents from GPS trace analysis," *Expert Syst. Appl.*, vol. 73, pp. 43–56, May 2017, doi: 10.1016/j.eswa.2016.12.018.

[23] D. Pevec, H. Vdovic, I. Gace, M. Sabolic, J. Babic, and V. Podobnik, "Distributed data platform for automotive industry: A robust solution for tackling big challenges of big data in transportation science," in *ConTEL 2019 - 15th International Conference on Telecommunications, Proceedings*, Jul. 2019, doi: 10.1109/ConTEL.2019.8848542.

[24] Y. Liu, "Big Data Technology and Its Analysis of Application in Urban Intelligent Transportation System," in *Proceedings - 3rd International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2018*, Apr. 2018, vol. 2018-Janua, pp. 17–19, doi: 10.1109/ICITBS.2018.00012.

[25] P. Y. Kusmawan, B. Hong, S. Jeon, J. Lee, and J. Kwon, "Computing traffic congestion degree using SNS-based graph structure," in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, 2014, vol. 2014, pp. 397–404, doi: 10.1109/AICCSA.2014.7073226.

[26] T. I. Damaiyanti, A. Imawan, and J. Kwon, "Extracting trends of traffic congestion using a NoSQL database," in *Proceedings - 4th IEEE International Conference on Big Data and Cloud Computing, BDCloud 2014 with the 7th IEEE International Conference on Social Computing and Networking, SocialCom 2014 and the 4th International Conference on Sustainable Computing and C*, 2014, pp. 209–213, doi: 10.1109/BDCloud.2014.97.

[27] X. Y. Chen, H. K. Pao, and Y. J. Lee, "Efficient traffic speed forecasting based on massive heterogeneous historical data," in *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, 2014, pp. 10–17, doi: 10.1109/BigData.2014.7004425.

[28] X. Zheng and S. Wang, "Study on the Method of Road Transport Management Information Data Mining based on Pruning Eclat Algorithm and MapReduce," *Procedia - Soc. Behav. Sci.*, vol. 138, pp. 757–766, Jul. 2014, doi: 10.1016/j.sbspro.2014.07.254.

[29] J. Yu, F. Jiang, and T. Zhu, "RTIC-C: A big data system for massive traffic information mining," in *Proceedings - 2013 International Conference on Cloud Computing and Big Data, CLOUDCOM-ASIA 2013*, 2013, pp. 395–402, doi: 10.1109/CLOUDCOM-ASIA.2013.91.

[30] A. Imawan and J. Kwon, "A timeline visualization system for road traffic big data," in *Proceedings - 2015 IEEE International Conference on Big Data, IEEE Big Data 2015*, Dec. 2015, pp. 2928–2929, doi: 10.1109/BigData.2015.7364125.

[31] Z. X. Yang and M. H. Zhu, "A Dynamic Prediction Model of Real-Time Link Travel Time Based on Traffic Big Data," in *Proceedings - 2019 International Conference on Intelligent Transportation,Big Data and Smart City, ICITBS 2019*, Mar. 2019, pp. 330–333, doi: 10.1109/ICITBS.2019.00087.

[32] X. Zheng *et al.*, "Big Data for Social Transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 620–630, Mar. 2016, doi: 10.1109/TITS.2015.2480157.

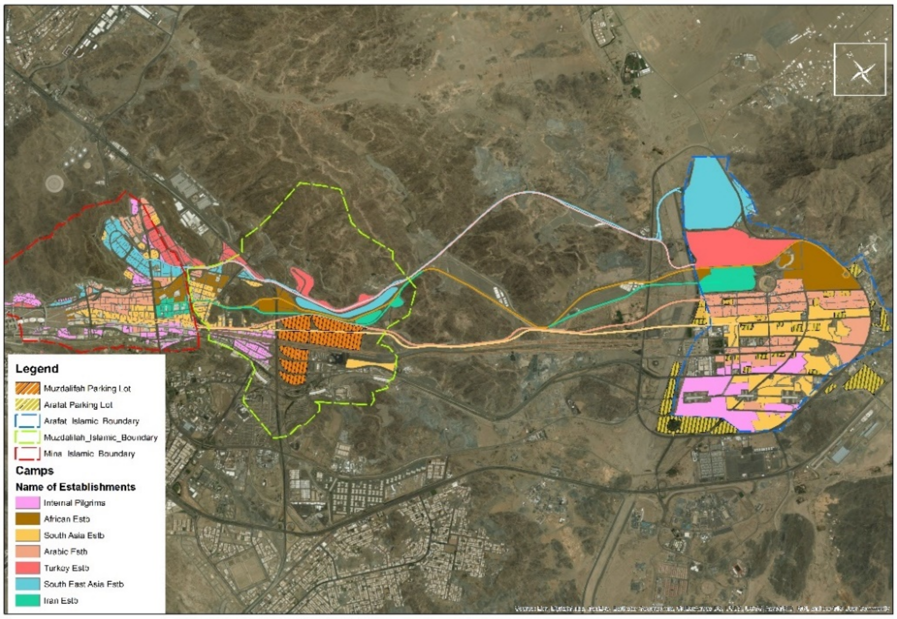


Figure 1: Map-based distribution of establishments in Mashaer area (Mina, Muzdalifah and Arafat), their assigned routes and parking lots

Table 1: Sample Location data of buses

bus_id	company_id	angle	lat	long	GM_signal	speed	record_time
108041310	1	144	24.49599	46.78958	1	39	2018-03-13 17:36:17
10708310	1	282	24.50138	46.79533	1	0	2018-03-23 11:03:55
12516310	1	61	24.50132	46.79531	1	0	2018-09-25 21:05:27

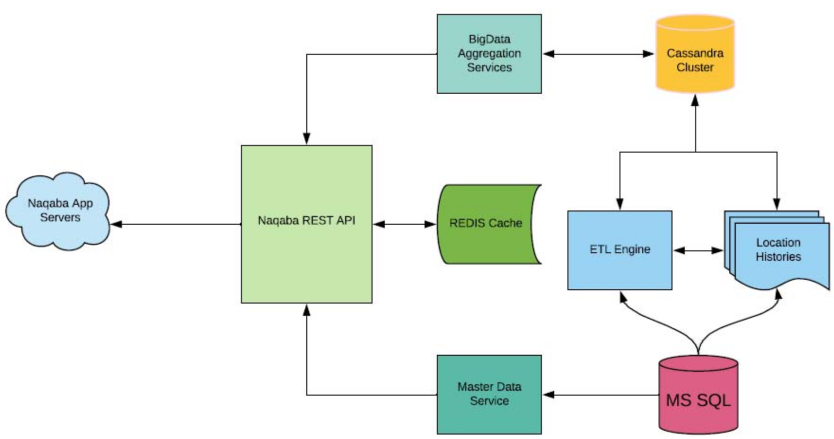


Figure 5: Overview of the platform

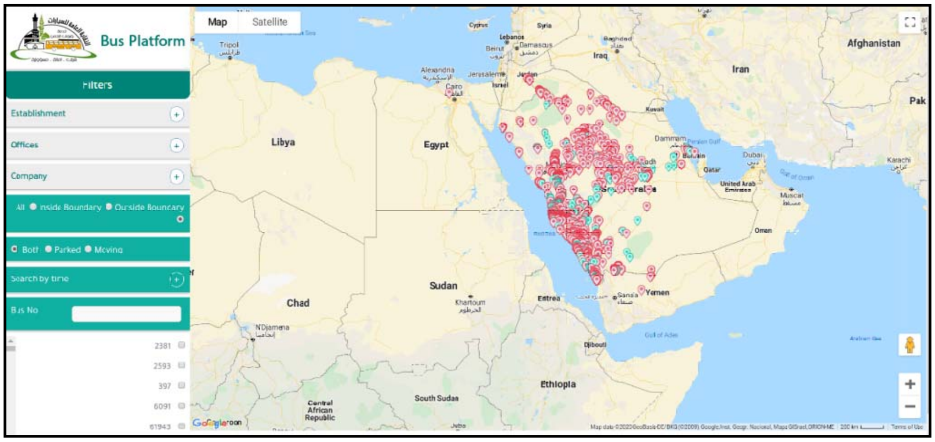


Figure 7: Interactive platform to visualize the buses data

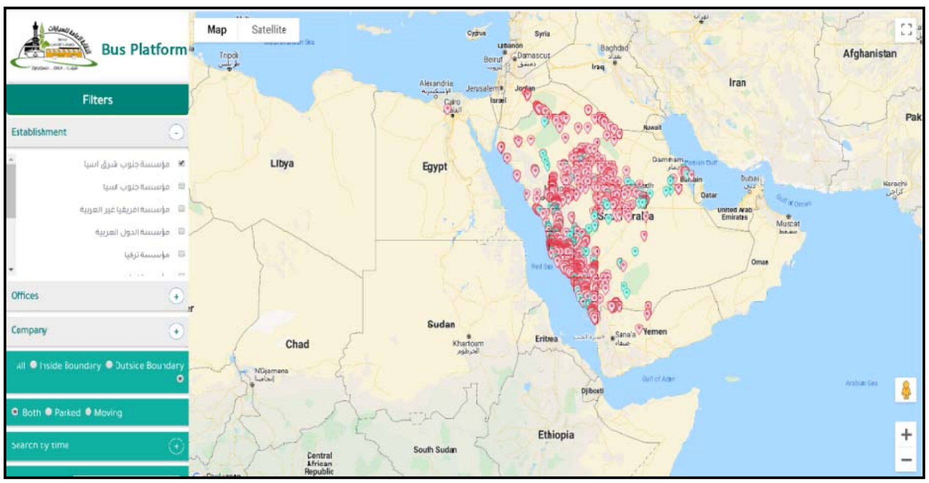


Figure 9: Zones and parking lots in Mashaer area based on the establishment

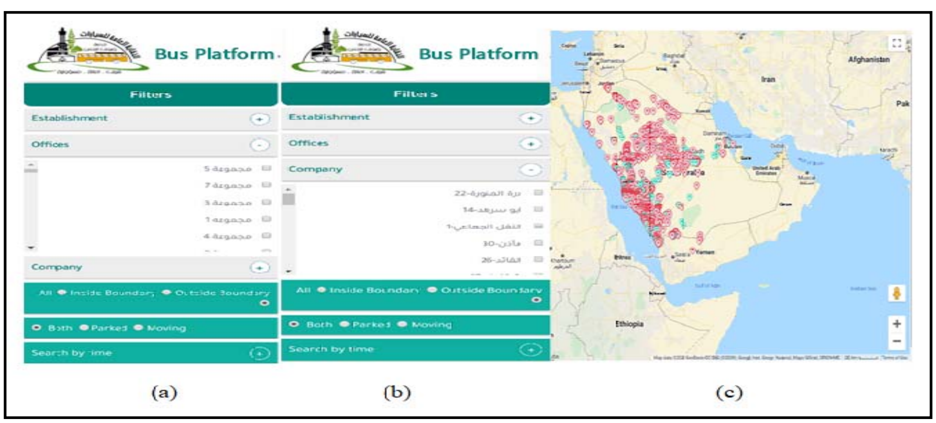


Figure 10:Features of the platform

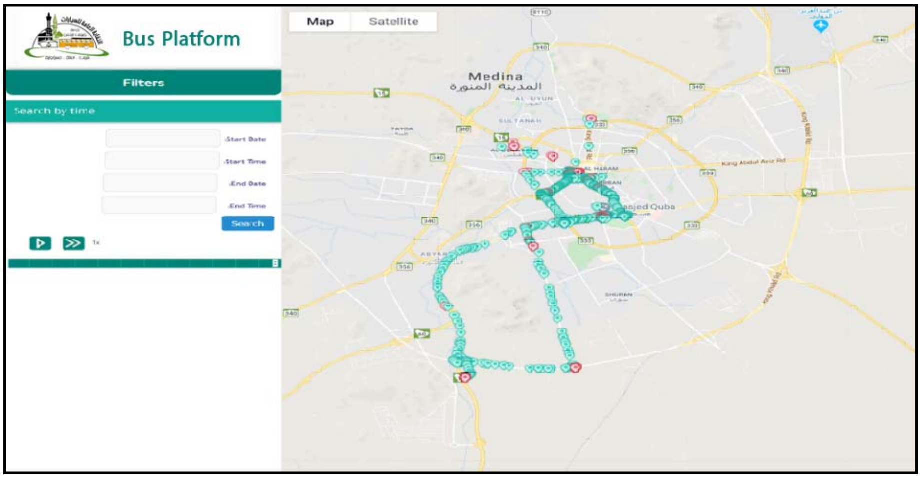


Figure 11: Movement of the individual Bus

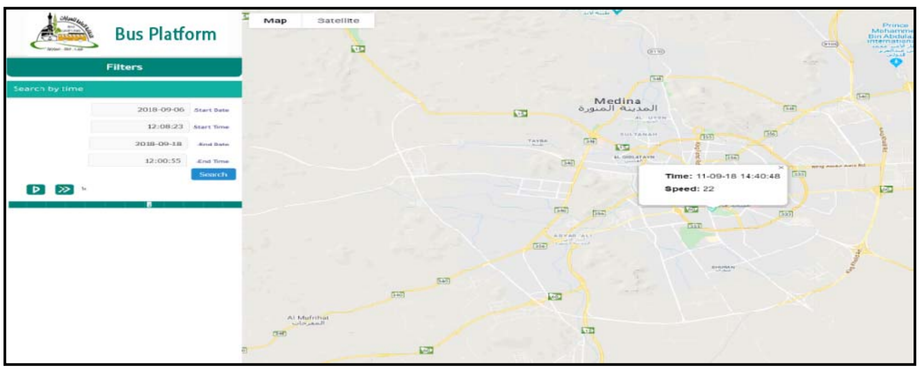


*Figure 12:*Simulation of an individual bus on historical data

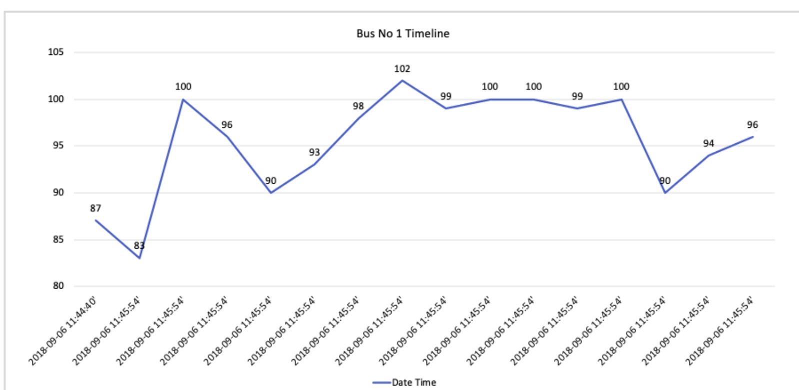


Figure 13: Driver's Behaviour (Speed Timeline)

Table 2: Violations details of bus no. 1

Start Timestamp	End Timestamp	No of Consecutive Observations	Violation Duration in Seconds	In Minutes
2018-08-23 04:47:44	2018-08-23 04:55:44	2	480.0	8
2018-08-23 06:05:46	2018-08-23 06:13:45	2	479.0	7.98
2018-08-23 06:31:55	2018-08-23 06:31:55	1	0.0	0
2018-08-24 05:51:27	2018-08-24 05:51:27	1	0.0	0
2018-08-25 06:29:29	2016-09-25 06:37:29	2	480.0	8
2018-08-25 11:43:35	2016-09-25 12:19:36	17	2161.0	36.01
2018-08-25 12:29:43	2016-09-25 12:35:37	3	354.0	5.90

Table 3: Violations classification

Severity	Explanation
Once	Some violations do not have continuous number violating observations. These violations occur only in single observation so we cannot get starting timestamp & ending timestamp, so we called it **Once**.
Normal	If the violation's duration is less than 10 minutes, then we called **Normal** violation
High	If the violation's duration is between 10 to 20 minutes, then we say it is **High**
Severe	If the violation's duration is more than 20 minutes, then we say its **Severe**

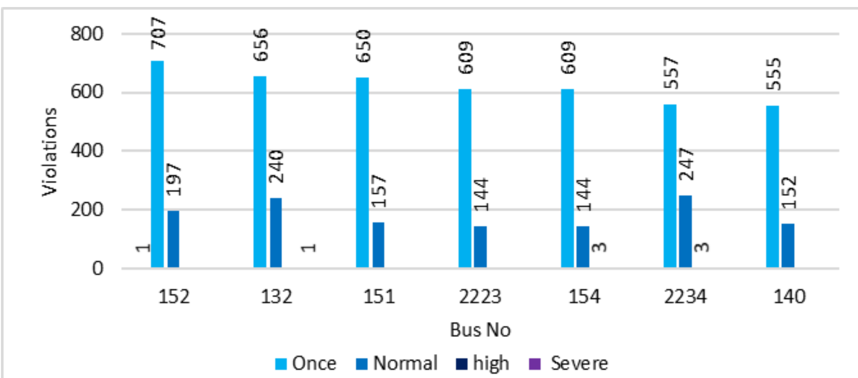


Figure 14: Identification of worst drivers

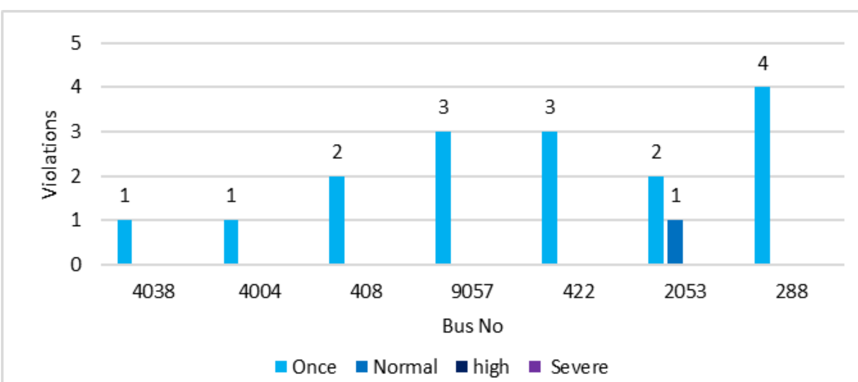


Figure 15: Identification of best drivers

Journal of Theoretical and Applied Information Technology
15th November 2020. Vol.98. No 21

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

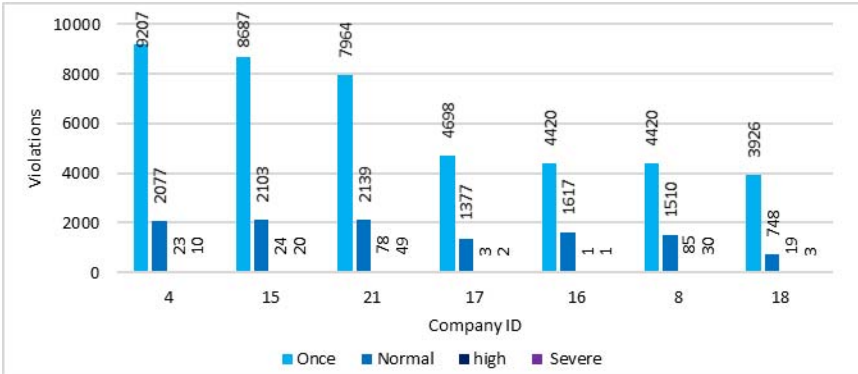


Figure 16: Worst performing companies

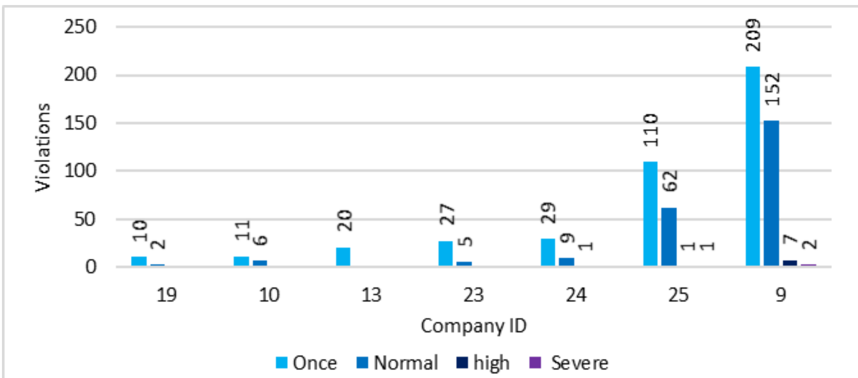


Figure 17: Best performing companies