**DCU School of Computing Assignment Submission**

**Student Name(s): Vikas Chhillar**

**Student Number(s): 16212887**

**Programme: MCM**

**Project Title: Creating the web interface to predict location and timings**

**Project Due Date: 24/08/17**

**Declaration**

**I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying is a grave and serious offence in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion, or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references.**

**I have not copied or paraphrased an extract of any length from any source without identifying the source and using quotation marks as appropriate. Any images, audio recordings, video or other materials have likewise been originated and produced by me or are fully acknowledged and identified.**

**This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study. I have read and understood the referencing guidelines found at http://www.library.dcu.ie/citing&refguide08.pdf and/or recommended in the assignment guidelines.**

**I understand that I may be required to discuss with the module lecturer/s the contents of this submission.**

**I/me/my incorporates we/us/our in the case of group work, which is signed by all of us.**

**Signed:Vikas Chhillar**

**Creating the web interface to predict location:**

**Abstract:** Passenger information system about bus arrival time at bus stop is an integral part of any transportation application. Bus reliability and regularity heavily affect the rider’s attitude for choosing public transportation and it is also one of the important indicator for transit performance evaluation. There are different approaches that try to make transportation system more desirable to commuter so that people start shifting from personal mode to public mode for the transportation. One of the important factor that the passenger expects is estimated arrival time prediction at the bus stops. There have been many studies done which looks into the problem of bus arrival time prediction. However, studies that uses the historical data to estimate the predicted time are limited. The present studies attempts to give the prediction of the arrival time based on the historical data of the buses under heterogeneous traffic conditions that exist in India. The results obtained from the forecasting are compared with the historical data.

**Introduction:** One of the biggest problem faced by urban transport around the world due to increasing vehicular population on the road is congestion. It is likely to become severe and may result in increased travel time and population. The best possible way to tackle this situation could be to encourage people to use the public transit. There are different ways to make the public transport more attractive and one of the services could be to provide the user with the predicted time for the arrival of the bus at the bus stop. The present study will provide bus arrival time to the users in advance, which will make the bus transportation system more reliable. There have been many studies done to give the predictions but most of them use the GPS of the buses to tell the location.There is no GPS system in buses of Delhi transport corporation and the time table submitted by the transportation system is often unreliable. This makes the bus system unreliable for most of the people in the city. Creating the web interface and telling people about the location and timings of the buses, has the potential to improve the transportation system of the city. With over 220 million users of smartphones in India, India has become the second largest user of smartphones which will make it easier for most of the people to use the information on the web interface(**Correspondent, S. (2016). *With 220mn users, India is now world’s second-biggest smartphone market*. [online] The Hindu. Available at: http://www.thehindu.com/news/cities/mumbai/business/with-220mn-users-india-is-now-worlds-secondbiggest-smartphone-market/article8186543.ece [Accessed 5 Dec. 2016].**

). Anyone with internet access will be able to know the information about buses regardless of where they are located and which machine they are using. This could convince more people to use the transportation system and improve the pollution level of the city and overall transportation system.The objective of this study is to improve the travel reliability and provide economic benefits.

2.**Literature review:** Various techniques are used to predict the timing of the buses such as statistical techniques, machine learning techniques and historic and real time approaches. Historic methods predicts the timings by averaging the time period over several previous time periods. In case of real time methods, travel time can be calculated for the next time period by using the present time period value.

Statistical techniques like regression and time series methods are used to calculate the timings of the buses. These techniques are based on the assumption that the current and future time patterns depends only on historical data. Regression techniques solves the value by using set of independent variable such as traffic conditions, road condition to solve the dependent variable. The accuracy depends upon identifying the suitable independent variable. (<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7581300>)

Generally speaking, transport timings methodologies can be isolated into

six classes:Kalman Filtering, K-Nearest-Neighbor, Artificial Neural Network, Support Vector Machine, time series and hybrid models, which are described below.

A. Support vector machine

Yu et al. [[22] **Y. Bin, Y. Zhongzhen, and Y. Baozhen, “Bus arrival time prediction using support vector machines,” J. Intell. Transp. Syst., vol. 10, no. 4, pp. 151–158, 2006**.] proposed this method to predict the timings of buses by considering the segment-level travel time and weather condition. Later, they improved the prediction by considering the multiple routes, which resulted in higher accuracy.

KNN model is also used by integrating cluster analysis and principal component analysis to predict timings using GPS data.

C. Kalman Filtering Model

**(J. Yang, “Travel time prediction using the GPS test vehicle and Kalman filtering techniques,” in Proc. Amer. Control Conf., Jun. 2005, pp. 2128–2133.)** presented this method to predict short-term arrival time of bus using real time GPS data in Seattle, Washington.

**A. Shalaby and A. Farhan, “Prediction model of bus arrival and departure times using AVL and APC data,” J. Public Transp., vol. 7, no. 1, p. 3, 2004**. applied this approach for predicting time of the bus based on link travel time and passenger demands from AVL and APC systems.

D. Artificial Neural Network Model

**R. Jeong and R. Rilett, “Bus arrival time prediction using artificial neural network model,” in Proc. 7th Int. IEEE Conf. Intell. Transp. Syst., Oct. 2004, pp. 988–993.** Proposed an Artificial Neural Network Modelfor stop-level bus dwelling time prediction with Automatic vehicle location data. Presented both link based and stop based ANN models for arrival time prediction, and used travel speed, volumes, as well as dwell time as inputs. The algorithm was tested in a microscopic simulation model with the best prediction performance.

E. Time series model

**M. D’Angelo, H. Al-Deek, and M. Wang, “Travel-time prediction for freeway corridors,” Transp. Res. Rec., J. Transp. Res. Board, vol. 1676, pp. 184–191, 1999** Predicted bus travel time pattern to find out the time lag for time series, and then implemented the exponential smoothing technique to find out the arrival time of buses. The result shows that the predicted arrival times are close to the actual values.

Most prediction model take help of GPS data to give the predicted bus arrival time. The GPS data is the recent data and it helps in giving a good prediction but some cities still do not have the GPS in the buses which is a big problem for any growing city.

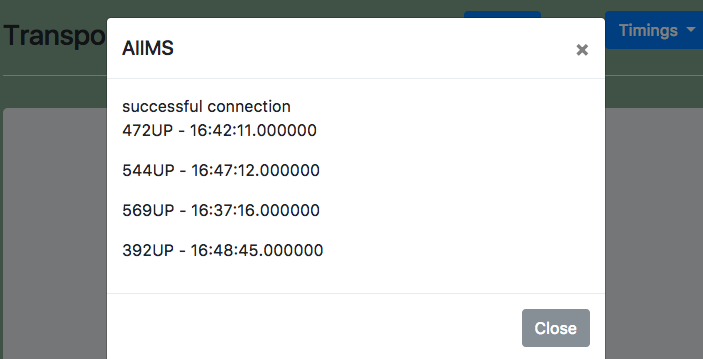
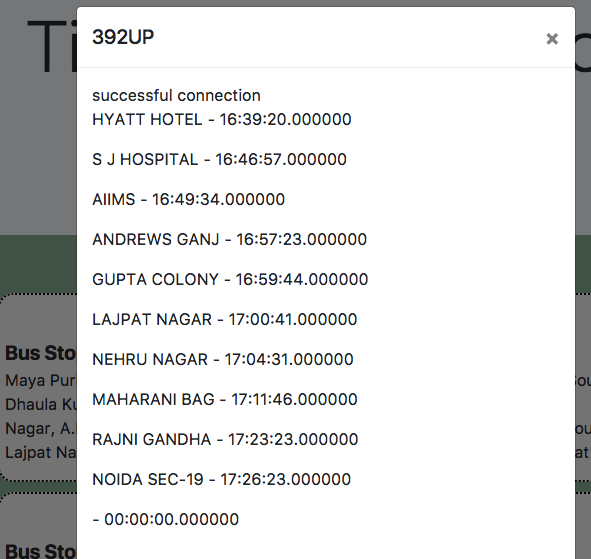
F. Hybrid Model: Liu et al mixed a state-space network model and KF model to predict the travel time. Van Lint et al. presented a state-space neural network for predicting the travel time with missing data. Park and Lee worked on loop detector data and Dedicated Short-Range Communication data to develop a bayesian model for estimation of travel time. Kuchipudi and Chien predicted travel time with link based and path based model from freeway toll data.

Summary: KNN and SVM are sensitive to the setting of parameters, which may also result in low prediction accuracy. Kalman filtering model works well for one-time-step-period-ahead bus travel prediction, but fails for multi-step-period-ahead bus travel prediction. The result of ANN has great randomness since the initial weight matrix is random. Time series models perform well in linear relationship, while arrival time may exhibit irregular and nonlinear feature during peak hours and thus thus time series models fail in capturing such feature. All of the above methods give the point value prediction instead of predictive confidence interval. Most prediction models rely on GPS data to predict the stop-level bus arrival time for single vehicle. These models do not consider the interaction between two consecutive buses at same stops.

**Data description**: This study uses the data of Delhi Transport Corporation(DTC) that operates on many routes in Delhi and neighbouring states. A total of 773 routes are covered across Delhi by DTC. The dataset was collected by Right to information act and it has live duty wise monitoring report. Key information includes Route, Date, passenger ticket timings, boarding and alighting information about passenger and it is collected from electronic ticket machine. The dataset is historical and it is not real time dataset.

**Data processing**: The dataset does not contain the timings at which bus reached the bus stop but it is calculated from the dataset using the information about boarding report of passenger. The transaction time associated with the first passenger is considered as the bus arrival time at the bus stop. The time attribute is converted into decimal to analyse it and find the results.

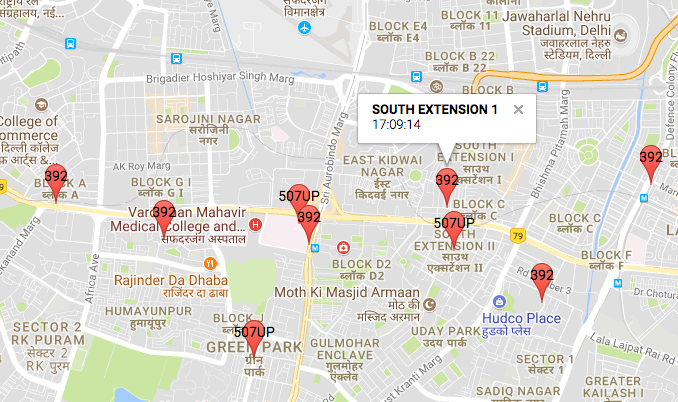
3. **System concept:** The prediction of bus timing is done from the historical data of the time series by using moving averages and regression for forecasting. This result is imported to the MySql database which also contains latitude and longitude of bus stops. MySql is connected to the web interface and maps which gives the result to the user about the predicted time of buses at different stops and also show the results on the maps. The results below are for one bus stop and one route which are displayed on the web interface.



Data collection and analysis

3.**Methodology:** In the present study, a bus prediction time travel method was proposed which will help the passengers to know the predicted time.The database is imported in MySql database by using Mamp. The web interface is created and the connection is established between MySql and the web interface.

3.1 Plotting on the map

The latitude and longitude of the bus stops are found by using Google maps. These latitude and longitude are described in the table of MySql. The data is then output as XML using php. The XML file is loaded and the markers are customized with different routes. It gives information about the predicted time to the user for different bus stops. 

4. **Implementation**: The historical time of the buses at the bus stop is normalized into decimals. The time series of 4 days is selected to see the trend. Moving averages(M.A) and centered moving averages(C.M.A) is calculated for the 4 bus stops on four days of the week to give forecasting of next days.

S(t),I(t)=y(t)/(C.M.A)

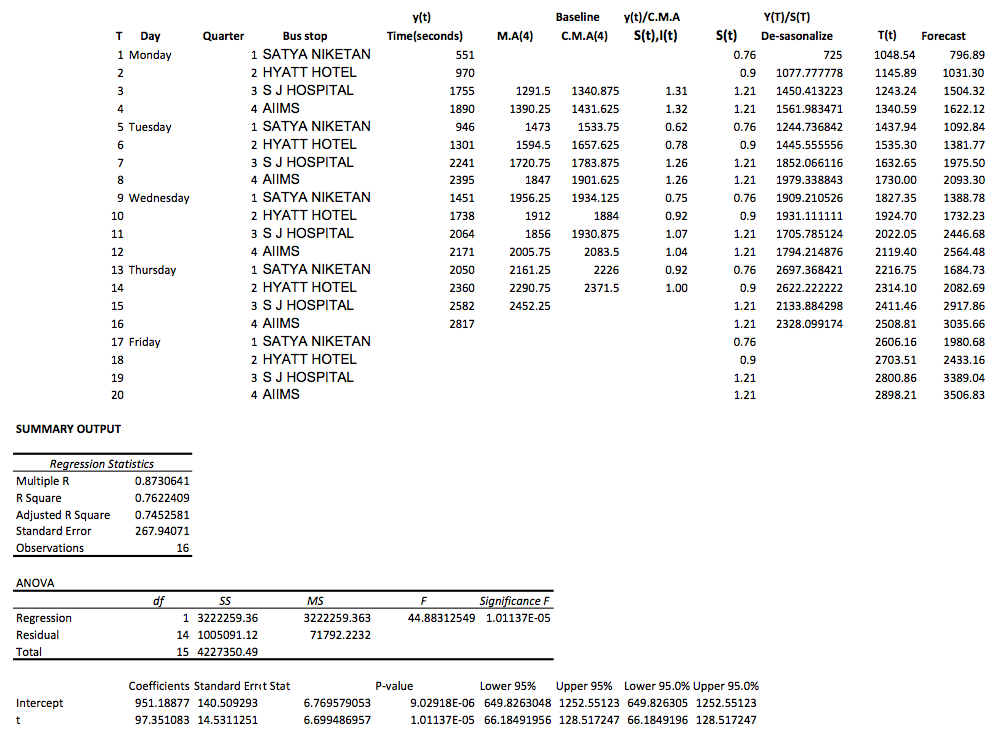
Where S(t) is the seasonal component and I(t) is the irregular component. M.A and C.M.A helps in smoothing the time series. Smoothing helps in taking out the seasonal component and irregular component. The model learns with the difference between original data and smoothed data. De-seasonality is also calculated using Y(t)/S(t). It helps in getting rid of seasonality and irregularity. S(t),I(t) tells how much percent the seasonality and irregularity component combined is below or above the smoothed curve.

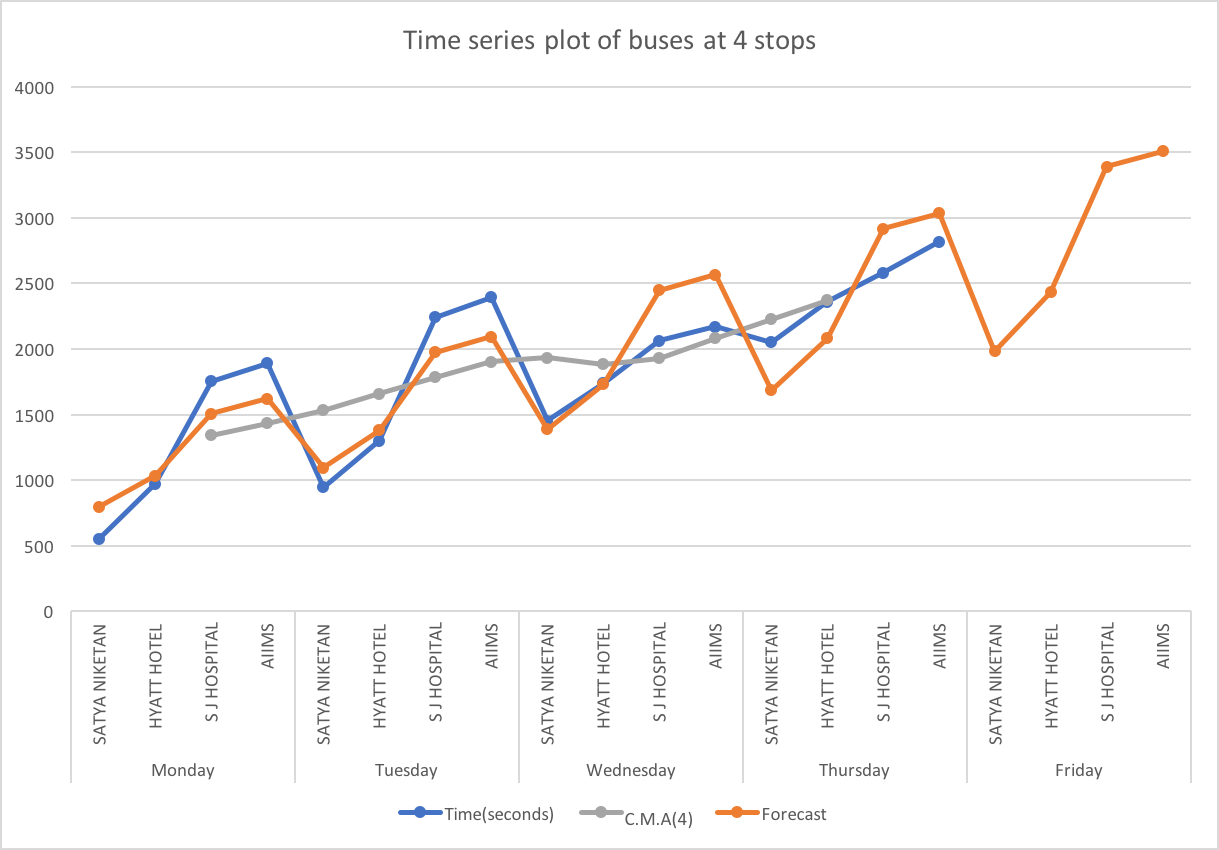
Y(t)=S(t)\*I(t)\*T(t)

Where Y(t) is the time series value at time t and T(t) is the trend component.

For calculating the trend component T(t), regression is used where y-axis is De-seasonalize and x-axis is t. T(t)=y intercept+slope\*t.

Forecast=S(t)\*T(t) gives us the prediction about the data using the historical value. In the figure below the dataset gives prediction about Friday by learning from the historical data. The values obtained in forecasting can be imported to the database to give the results on the web interface and Google maps. The implementation can be done separately for weekend and weekdays as the timings vary a great.





5. Evaluation:

6.conclusion: The main aim of bus arrival prediction application is to encourage people to use public transport and thus helps to reduce the traffic congestions on roads. However, for this to happen in reality, the users should be given estimated time of the buses. A reliable estimation of bus arrival time can reduce unnecessary passenger waiting time and prevent bus bunching. However, in reality, it is hard to predict the arrival time due to random passenger arrival time and unstable traffic condition. The present study proposed a prediction method which uses moving average and linear regression to give the forecasting under heterogeneous traffic conditions that exist in India. The web interface is deployed on the cloud so that everyone can access it.

7. Future work: There is no GPS in buses of Delhi transport corporation due to which the prediction about timings is difficult. If there is GPS in the buses, the accuracy of bus prediction will be more accurate as there are so many factors that affects the traffic and the arrival time of buses. GPS gives the real time data that can be processed to give better results to the users.