# Demand prediction in a bike-sharing system using machine learning techniques

Research Project
Masters in Data Analytics (B)


Ashish Rawat
Student ID: X18185801

School of Computing
National College of Ireland


Supervisor: Paul Stynes

**National College of Ireland**
**Project Submission Sheet**
**School of Computing**

| | |
|---|---|
| **Student Name:** | Ashish Rawat |
| **Student ID:** | X18185801 |
| **Programme:** | Masters in Data Analytics (B) |
| **Year:** | 2020 |
| **Module:** | Research Project |
| **Supervisor:** | Paul Stynes |
| **Submission Due Date:** | 17/08/2020 |
| **Project Title:** | Demand prediction in a bike-sharing system using machine learning techniques |
| **Word Count:** | 4384 |
| **Page Count:** | 16 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

| | |
|---|---|
| **Signature:** | |
| **Date:** | 17th August 2020 |

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies). | ☐ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | ☐ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | ☐ |

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Demand prediction in a bike-sharing system using machine learning techniques

Ashish Rawat
X18185801

**Abstract**

A bike-sharing system provides people with a sustainable mode of transportation and has beneficial effects for both the environment and the user. And the city-wide accessibility and low cost has exponentially increased its popularity. Nonetheless, the increased usage has led to create issues like unavailability of bikes and docks at bike stations. Therefore, the study aims to predict the demands of a bike sharing system using machine learning models. And also analyses comprehensive effect of the time dependent and inter-station relationship on predicting the demands. The Metro bike-share data of 2019 consisting of more than 2 million records of bikes trip was used in the research. The original data files lacked the demand attributes, required for the research and demanded extensive data processing and transformation. Four machine learning models were employed for predicting the demands; ARIMA, LSTM, STGCN and TAGCN. And parameters such as RMSE, MSE and MAE were used for evaluating the models. The performance of the models was found to be vary with the time interval used in transforming the data. And the best performance was achieved by the STGCN with comparatively small RMSE values of 0.76 and 0.37 for bikes and docks demand respectively. The approach successfully addressed the shortcomings highlighted in Kim et al. (2019) in resolving the inclusion of new bike station in the system and performed well.

## 1 Introduction

The increased usage of private vehicles in metropolitan areas has resulted in a significant rise in fuel consumption's that have adverse effects on the climate. It has led people in today's society to accept problems like road traffic as the norm. Therefore, the government and organizations started adopting measures to facilitate sustainable development to address the issue. In that context, the Bike Share initiative was launched to tackle the public mobility problem. It provided the people with an alternative to using a sustainable mode of transport for a small distance at a minimal cost. And gave people the freedom to utilize the service by themselves. In a bike-share system, a user could lend a bike from any bike stations and return it to a bike station near the destination and since it involves the activity of pedalling the bike it has beneficial health effects. And the city-wide installation of bike stations improved the accessibility of areas by bikes.

However, the popularity of the bike-share system increased drastically which led to creating a gap between the supply and demands of bikes and docks at bike stations. And the most common issues faced by the users are the lack of bikes and docks available at bike stations. The growing concern led the bike operators to consider the matter seriously, and to develop strategies to dissipate the issue. The initial strategies implemented truck usage to move the bikes from one bike station to another to solve the issue of unavailability. The strategy was widely analysed in several studies to provide an appropriate implementation plan. While it was found to be successful for night-time maintenance operations, the underlying purpose of solving daytime issues was not accomplished. Since then, the objective to resolve the unavailability issue has been widely examined in many studies. But many of those studies produced qualitative results due to the lack of dependent features in the available data. Few empirical studies have manipulated the bike share data to derive the desired features and to present quantitative results by utilizing the historic records of bike trips and external factors that affect the system. In addition, very limited few have examined the impact of nearby bike stations on the availability of a particular bike station. The task of assessing both the inter-station effects and the time-varying data has been an ordeal to analyze using standard methods. But the recent introduction of new methodologies has made it possible to evaluate both the factors to comment on the availability issue.

Therefore, the fundamental aim of the research is to predict the demands in a bike-share system using machine learning methods. And the secondary objective is to analyse the collective inter-station and time-varying effect on predicting the demands for a bike station. To address the research objectives widely used models related to the study, autoregressive integrated moving average (ARIMA) model and long short-term memory (LSTM) model are included in the study. And to evaluate the second objective spatial-temporal graph convolution network (STGCN) model and topology adaptive graph convolution network (TAGCN) are utilised. The comprehensive evaluation of the models is performed on the root mean squared error (RMSE), mean squared error (MSE) and mean absolute error (MAE), to justify the research conclusion.

The research is structured in the following way: section 2 reviews the related work in the bike-sharing domain, section 3 describes the methodology followed in the study, section 4 provides details of implementation process, section 5 illustrates the evaluation metrics and critics the final results, and finally, section 6 concludes the work and comments on the future scope.

# 2 Related Work

Since the last decade, a lot of work has been presented on bike-sharing system but very few actually aims to quantitatively predict the demand at a bike station. Initial studies involved application of optimization algorithms which were proven to be ineffective for the situation (Raviv et al.; 2013). However, the application of machine learning models for bike-share network provided significant results which are briefly described in the sub sections.

The following subsections are structured as follows; 2.1 provides information on the data transformation techniques utilised in related works, 2.2 illustrates the details of widely used machine learning models for bike-share and finally 2.3 introduces the application of recent methods to also consider inter-station relationship in the bike-share network. At the end, a conclusion details the methods is provided.

## 2.1 Data Modification

The nature of the bike share data limits the option of methods, which can be utilised for analysis. Most of the bike share data consists of bike trip records and station location records, which usually does not include bikes and docks demand attributes. Hence, most studies usually focus on analysing the demographics of the data and how it affects the system (Faghih-Imani et al.; 2017).

However, Giot and Cherrier (2014) suggested to transform the original bike data of Washington DC to obtain one record per hour for all the days within time period before utilising it for training models. Similarly, Xu et al. (2018) also modified the bike data, but they developed different time interval datasets (10 min, 15 min, 20 min and 30 min). Although, they utilised the data of a station-less bike sharing system their data modification method can be utilised for any bike sharing data and generate features like demand of bikes and demand of docks. These features are vital for effectively training and evaluating the results of predictive models.

## 2.2 Machine and Deep Learning Methods

A bike-share system data majorly constitutes of time-dependent features. These features fluctuate randomly making it impossible to build a predictive model using static stochastic time series techniques (Kaltenbrunner et al.; 2010). Thus, Yoon et al. (2012) proposed the implementation of AutoRegressive Integrated Moving Average (ARIMA) models which effectively handles the dynamic nature of the data. They evaluated ARIMA models over two scenarios, short-term and long-term prediction and calculated the RMSE values of all the variations. The results concluded that for each experiment the ARIMA models generated the smallest RMSE values (0.92 and 3.50) and outperformed AutoRegressive Moving Average (ARMA) model and other algorithms. Siami-Namini et al. (2018) compared the performance of ARIMA and long short-term memory model (LSTM) over multiple time series data and assessed their performance using RMSE. Their results concluded that LSTM model which was trained on 70% of the data using MSE as loss function and ADAM as optimization function improved the prediction performance by 85% compared to ARIMA model.

Similarly, Xu et al. (2018) compared the performance of the LSTM with multiple models (ARIMA, ANN and XGBoost) however, they experimented it for different time interval. And according to their results the predictive accuracy of the LSTM model was 37.4% better than the ARIMA model and accordingly outperformed the other models. Pan et al. (2019) evaluated the performance of the LSTM model against DNN model and achieved low RMSE value of 3.67 and 2.7 for train and test set respectively. Zhang et al. (2018) utilised the LSTM Model for short term prediction of bike demands. Their primary assumption for the study was that, the LSTM model overcomes the drawback of retaining the memory of input values which is usually the case in RNN. The implemented LSTM model achieved least MAE and RMSE values (1.7 and 2.43 respectively), and also experimented the effect other public transport have on bike sharing system.

Likewise, many studies included external features which affected bike sharing network. These features in general comprised of weather information and in some cases, other modes of transportation (Singhvi et al.; 2015; Ashqar et al.; 2019; Pan et al.; 2019; VE and Cho; 2020). And through many works published till now, it has been verified that inclusion of these external features assist in getting more accurate prediction results. However, very few studies have analyzed the influence of neighbouring stations on the demands of a bike station. Yoon et al. (2012) stated that for short term predictions inter-station relationship plays a vital role and substantially affect the predictive capabilities of the model.

### 2.3 Graph Convolution Approach

Kipf and Welling (2016) introduced a novel approach (semi-supervised graph convolution) that used graph-structured data explicitly to identify nodes within the network. The interconnection of nodes in the graph enabled the data to maintain their spatial and temporal properties.Yu et al. (2018) developed a novel framework, spatial temporal GCN (STGCN) on the same principle. The model integrated the spatial convolution layers and temporal convolution layers by aligning them in sequence in the convolution block. Another novel framework, Topology Adaptive GCN (TAGCN) utilised the GCN approach to classify the nodes utilising the advantages of the traditional CNN and applying it to graph structured data (Du et al.; 2017).

Furthermore, Chai et al. (2018) presented a multi-graph convolutional neural network (MGCN) framework that convoluted features from multiple graphs (distance, interaction and correlation) to forecast results. The framework established was evaluated on RMSE parameter and compared against SARIMA, ARIMA, LSTM and GBRT models. The research concluded the application of MGCN outperformed other models which achieved least RMSE value.

In a related study, Kim et al. (2019) utilised the graph convolutional approach to predict demands of bikes and docks for Seoul bike share data. The model design stated the use of three different time window GCN modesl (hourly, daily and weekly) and concatenated the features obtained from each model. The process utilised the spatial as well as time varying features simultaneously. They verified the effectiveness of their novel framework by comparing it against heavily utilised models for bike share study like ARIMA, RNN and LSTM. The experimentation produced least MSE and APC values for the proposed framework which supported the hypothesis and encouraged the framework's application.

In conclusion, the machine learning techniques applied to derive the demand characteristics of a bike station has facilitated the process and provided it an computational edge than previous strategies. And, the recent introduction of GCN methodology has enabled to utilisation of spatial effects along with the temporal properties (Kim et al.; 2019; Chai et al.; 2018), encouraging its application for the verify the study objectives. The evaluation parameters in these studies (Zhang et al.; 2018; Kim et al.; 2019; Xu et al.; 2018) has also been utilised for this work to provide a comparison of the models performance.

## 3 Methodology

This study uses different machine learning models to assess the demands in the network for bike stations. Initially, the lack of prediction factor in the original data is addressed, through the application of a data transformation strategy and the creation of demands for two-time intervals (30min and 60min). And after data transformation, the following steps are employed:

- Data pre-processing and transformation

- Developing and optimizing ARIMA model
- Developing and optimizing LSTM model
- Developing graph data
- Implementing STGCN and TAGCN models
- Evaluation and discussion

## 3.1 Data Collection

The data in this research was collected from the Metro bike share system in Los Angeles city which is available on the Metro Bike website[1]. The data sets constitutes of bike trip data with 290283 records of bike usage from $1^{st}$ January 2019 till $31^{st}$ December 2019, and station data with 286 bike stations detail. The features utilised from the data sets for the research are listed in Table 1.

Table 1: Features utilised in the research.

| | Columns | Description | Values |
|---|---|---|---|
| 1 | start_time | Trip start date and time | From 01/01/2019 to 31/12/2019 |
| 2 | end_time | Trip end date and time | From 01/01/2019 to 31/12/2019 |
| 3 | start_station | Id of the start station | 4 digit numbers like 3005, 3006 etc |
| 4 | end_station | Id of the end station | 4 digit numbers like 3005, 3006 etc |
| 5 | Station_Name | Name of the station | String values |

## 3.2 Data Pre-processing

In pre-processing, initially time interval data set was created, containing start date time and end date time with a gap of fixed interval (30min and 60min). The records present in trip start time (bike trip dataset) were iterated and added to a respective interval in the time interval dataset to produce bike demand feature. Similarly, process was repeated with trip end time to give dock demand feature. The unwanted features and missing values were dropped from the newly formed data set.

Moreover, the bike trip data was further processed to generate weights for edges connecting each station with every other station. An average percentage parameter was generated and used as edge weight. The dataset created was essential for developing a graph structured data, which is a necessity for the proposed graph convolution models. Structure of the dataset is shown in figure.

## 3.3 Machine learning models

In this research the following machine learning models have been adopted.

[1]https://bikeshare.metro.net/about/data/

### 3.3.1 ARIMA Model

Autoregressive integrated moving average (ARIMA) model utilises the periodic and seasonal trends in a linear time series data (Box and Pierce; 1970). It is characterized with three parameters; 'p' is the degree of autoregressive elements, 'd' is the degree of differentiation and 'q' is the degree of moving average elements. Depending on the nature of the data different data methods are applied to make the data stationary. In the research data is made stationary by differentiating it by first order.

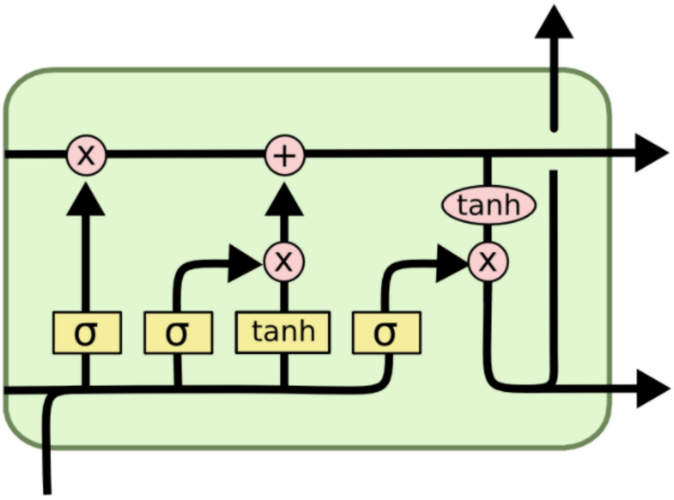


Figure 1: Structure of LSTM cell

### 3.3.2 LSTM Model

The long-term model (LSTM) is a form of recurrent neural network ( RNN) with two gates and a memory unit that allows the data to be stored for longer duration than standard RNN (Mikolov et al.; 2010). The two gate units are implemented by specifying the activation functions in the LSTM cell structure to regulate information flow within the network (figure 1) (Rathor; 2018). The LSTM model in this research consisted of 4 consecutive pair of LSTM layer and dropout layer. The dropout layer decreased the model's dependency on previous data, and the final output is obtained through the dense layer.

### 3.3.3 GCN Models

The graph convolution network (GCN) models utilises the node and edge representation of the graph structure data to incorporate both the spatial and temporal properties of the data (Kim et al.; 2019). The structure of the graph convolution model utilised in this research referred from Ohrn (2020) and is shown in figure 2.

In the 1D convolution layer, temporal data points from input tensors were, convoluted with a set kernel size. The tensor obtained was passed through an activation function that used one input channel as the controller and permitted data flow from the second channel.

The STGCN model used the GCNConv layers (Kipf and Welling; 2016) to determine the characteristics over which the temporal dimensions were assessed. The results were

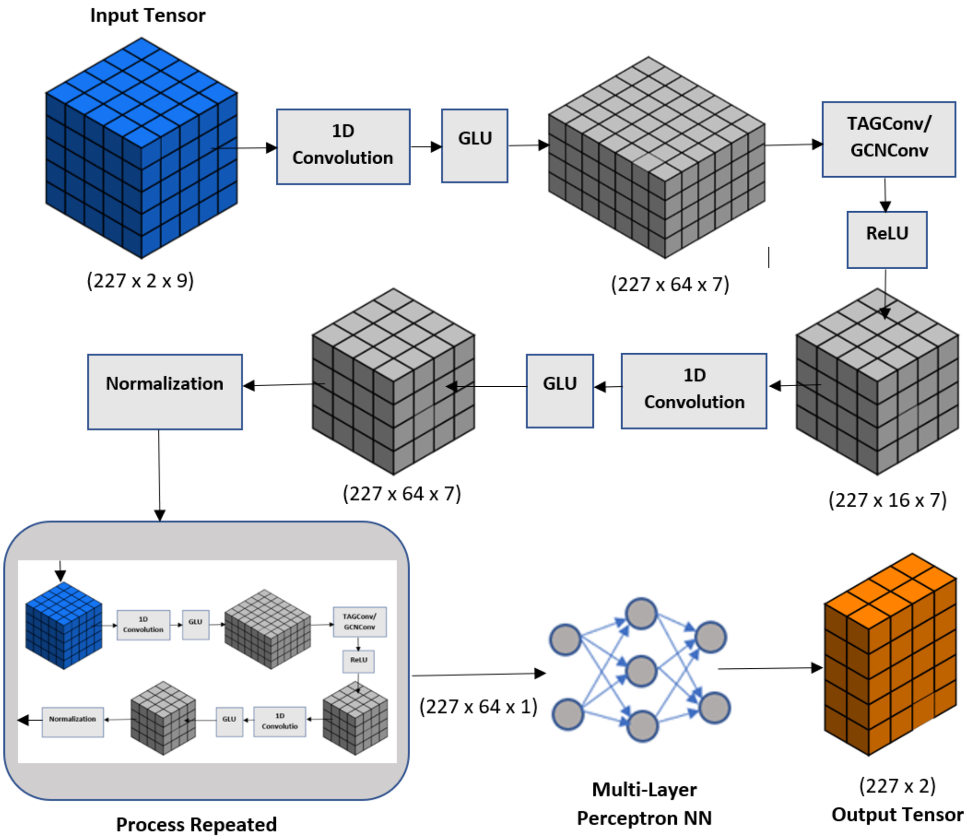


Figure 2: Architecture for GCN models

then passed through an activation function to transform the tensors shape. In a similar way, TAGCN model (Du et al.; 2017) utilised the TAGConv layers for determining the weights and characteristics of the data.

Normalization layers were, configured to mitigate the problems created by multi-connected systems and scale the dimension of tensors. The tensor thus obtained was repeated in the same cycle, and the final prediction was achieved via the multilayer perceptrons neural network (Eunbee Jang; 2019). The resulting tensor included the forecast of demands for both the bikes and docks for each station in the network.

# 4 Implementation

The research was implemented on Google colaboratory framework using Python 3.6 and some models GPU was was also utilised. The following sections constitutes the implementation: 4.1 provides data preparation details for each model, 4.2 presents the implementation setup for ARIMA model, 4.2 provides details to replicate the LSTM execution and 4.3 provides brief details on the parameters used in GCN models.

## 4.1 Data Preparation

The data obtained after pre-processing consists of 4979760 records and includes bike demand and dock demand features. However, the data contained records for all the stations, thus individual station datasets were generated from it for application in ARIMA and LSTM model. The variation in bikes and docks demand at station 3005 is represented in figure 3.

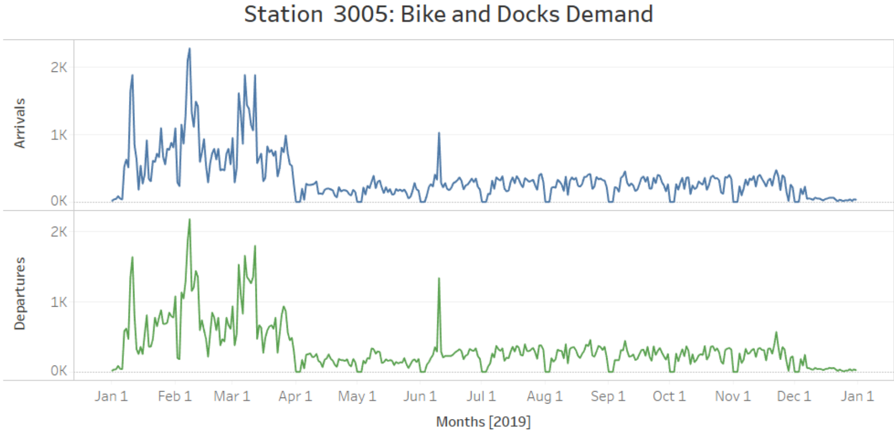


Figure 3: Bikes and docks demand of station 3005

However, to apply the graph convolutional methods the two dimensional data required conversion into multi-dimensional graph structure. The transformation process was administered using Pytorch Geometric package[2]. It utilises both the transformed and graphs weight datasets (from pre-processing). The nodes represent stations in the network and edges transcribe the relation between start station and end station. These edges are assigned the weights correspondingly from the graph weights dataset, which illustrate the frequency of use among those stations.And according to the sample size value, the network graph was disintegrated to generate into equivalent number of sub graphs. These graphs were stored into a directory during runtime for later utilisation. It should be noted that since, nodes are assigned for every value present in the station dataset, the model to be developed will not susceptible to new stations which was an issue faced by Kim et al. (2019).

## 4.2 ARIMA Model

ARIMA model was implemented using the statsmodel package. And for identify the stationarity of the data, rolling mean statistics and Dickey-Fuller test was performed. To pass the stationary test the data should have a p-value less than 0.5, otherwise the data is differentiated and again tested for stationarity. The degree of differentiation may vary from 0 to N depending on the data. The first order differentiated data of station 3005 is represented in figure 4.

After the data was made stationary 'p', 'd' and 'q' parameters for the ARIMA model were identified using a function which iterated these parameters over a range of values (0 to 5) to identify the best combination, for which AIC (Akaike information criterion) value was the smallest. To train the model 80% of the bike data was processed and predictions were made on the remaining 20% of the data.

[2]https://pytorch-geometric.readthedocs.io/en/latest/

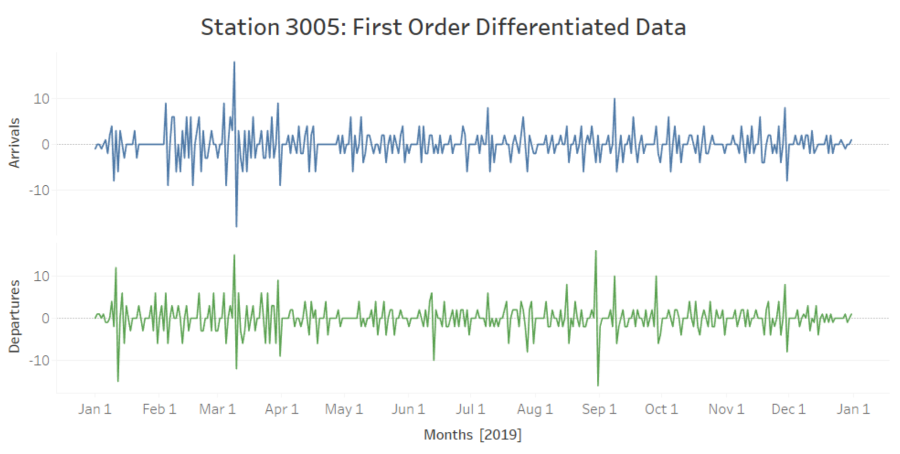


Figure 4: First order differentiated data of station 3005

## 4.3 LSTM Model

To implement the LSTM model the keras library were utilised in this research. At the initial step, features were scaled in the range of 0 to 1 using MinMaxScaler and a sequential neural network was developed. Adam was employed as the optimization function and mean squared error for the loss function was employed for training the model. The model was trained on 80% of the data and 20% from the train data was used for validation. Furthermore, the model's performance were experimented on different batch sizes and epochs and finally, following hyperparameters were established; 30 epochs, 32 batch size and earlystopping function with patience set to 5 epochs.

## 4.4 GCN Models

The two GCN models, STGCN and TAGCN were implemented using Pytorch package with CUDA 1.5.0 with GPU enabled backend. Both models have a similar structure except for the embedded spatial layer which uses GCNConv for STGCN model (Kipf and Welling; 2016) and TAGConv for TAGCN model (Du et al.; 2017). And the framework utilised mainly two activation functions, gated linear unit (GLU) for 1D convolution layer and rectified linear unit (ReLu) for spatial convolution layer. The stochastic gradient descent with a learning rate of 0.01 and a momentum value of 0.9, was used for optimization. And the models were finally executed for 5 epochs, 64 batch size due to memory limitations. For training the model 80% of time bound was identified and hard coded while the remaining was used for training.

# 5 Evaluation

The performance of the model was evaluated using the following parameters: MSE, MAE, and RMSE value. These parameters are commonly used to evaluate regression models and have been used, in many related works. The models were executed to obtain demands at bike stations for the next 30 min or 60 min time interval accordingly.

The following experiments were performed with the objective to validate the hypothesis. Experiment 1 presents the results of ARIMA model and comment on its complexity, experiment 2 verifies the use of LSTM to effectively predict time series data and experiment 3 presents the results obtained by STGCN model and experiment 4 illustrates TAGCN performance. And finally, a conclusion is provided which the discusses each techniques and highlights their advantages and disadvantages.

## 5.1 Experiment 1: ARIMA Model

For this experiment, the ARIMA model was trained for the data of station 3005 and was executed individually to generate both the predictions (demands for bike and docks). The table 2 describes the performance of the model for both 30 min and 60 min intervals. And from there it can be observed that for bike trip data with 30 min interval, the model produced a comparatively lower MSE than data with 60 min interval. It implies that ARIMA model performed better for the prior case and lower RMSE, for data with 30 min interval supports the model response in predicting the test data and indicates a better fit the short interval.

Table 2: ARIMA experiment results

| Bikes Demand | | | | | |
|---|---|---|---|---|---|
| Interval | (p, d, q) | AIC | MSE | MAE | RMSE |
| 30min | (3,1,4) | 132771.98 | 1.12 | 0.78 | 1.06 |
| 60min | (4,1,4) | 93086.59 | 3.26 | 1.38 | 1.81 |
| Docks Demand | | | | | |
| Interval | (p, d, q) | AIC | MSE | MAE | RMSE |
| 30min | (4,1,4) | 135473.82 | 1.16 | 0.82 | 1.08 |
| 60min | (4,1,4) | 93794.94 | 3.37 | 1.46 | 1.83 |

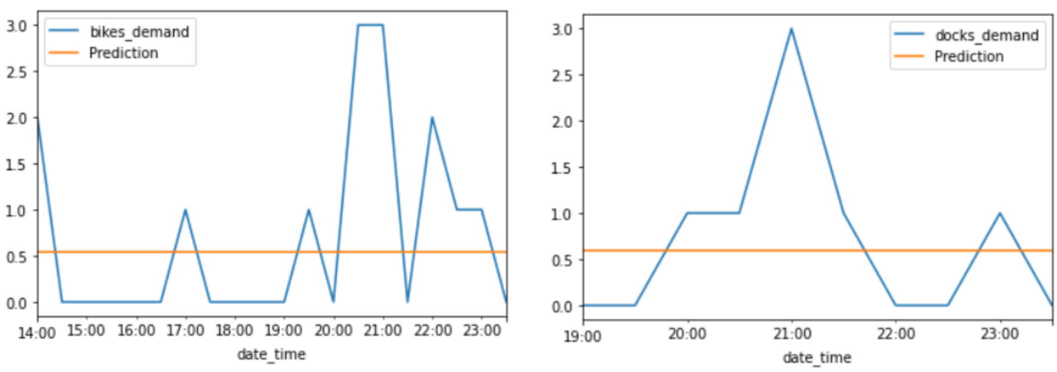


Figure 5: ARIMA model bikes and docks demand prediction for 30 min interval

However, from figure 5 it can be observed that model was not able to adapt to the variations in the data and produces a constant prediction. Therefore, exclusively on the basis of evaluation metrics the effectiveness of the model cannot be determined. Hence the LSTM model which has been referred, in similar studies to analyse bike trip data has been experimented in the next section.

## 5.2 Experiment 2: LSTM Model

Multiple experiments were performed on the LSTM model to fine-tune it to optimum configuration. The final model build was with parameters: 30 epochs with EarlyStopping function with patience of 5 epochs, 32 batch size and 0.2 dropout rate. The results in table 3 shows a similar trend like ARIMA model, where data with 30 min interval generates lower prediction error. However, the values obtained for the data with 60 min interval are significantly lower for the LSTM models which implies that the model is also effectively studying the variation of demands in the 60 min intervals and producing satisfactory results.

Table 3: LSTM experiment results

| Bikes Demand | | | |
|---|---|---|---|
| Interval | MSE | MAE | RMSE |
| 30min | 0.83 | 0.56 | 0.91 |
| 60min | 2.68 | 1.10 | 1.63 |
| Docks Demand | | | |
| Interval | MSE | MAE | RMSE |
| 30min | 0.88 | 0.60 | 0.94 |
| 60min | 2.42 | 1.09 | 1.55 |

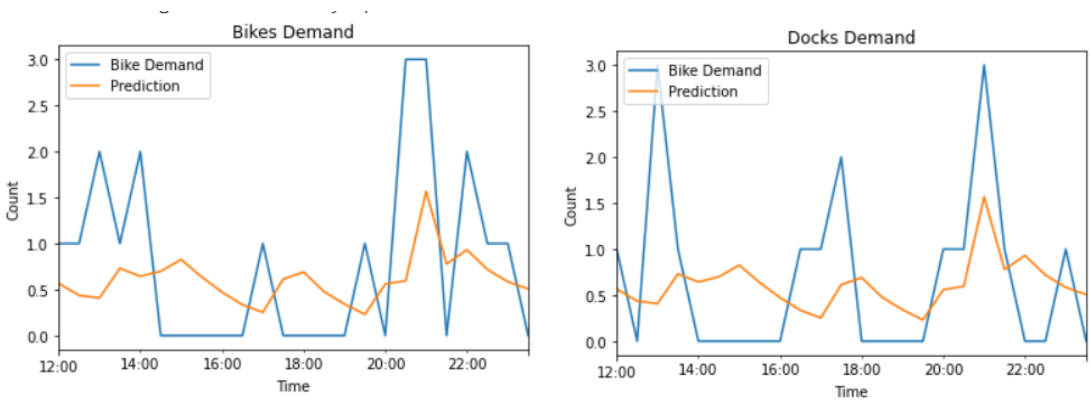


Figure 6: LSTM model bikes and docks demand prediction for 30 min interval

The results obtained in figure 6 indicated the effectiveness of model to adapt to the variation of demands in the time interval. However, the LSTM model utilised only the time dependent characteristics of the bike trip data to make predictions. Thus, to study the influence of adjacent bike stations of the demands STGCN model and TAGCN model is studied in the following sections.

## 5.3 Experiment 3: STGCN Model

The STGCN model was trained for only 5 epochs with a batch size of 64 due to the system RAM limitation. And during experimentation it was observed as the epoch size was increased from 0 to 5 the model performance significantly improved and lower MSE, MAE and RMSE were obtained. The results obtained on the test data is illustrated in Table 4. It is quite clear from the table 4 that model performed poorly in forecasting bike demands for 60 min interval and produced a high MSE of 8.26. Nonetheless, the results obtained for the data with 30 min interval are quite good.

Table 4: STGCN experiment results

| Bikes Demand | | | |
|---|---|---|---|
| Interval | MSE | MAE | RMSE |
| 30min | 0.74 | 0.58 | 0.76 |
| 60min | 8.26 | 1.79 | 1.34 |
| Docks Demand | | | |
| Interval | MSE | MAE | RMSE |
| 30min | 0.094 | 0.13 | 0.37 |
| 60min | 0.70 | 0.54 | 0.73 |

## 5.4 Experiment 4: TAGCN Model

The training and testing parameters for the TAGCN model were identical to the STGCN model. From table 5 it can be observed that TAGCN is not effectively able to analyze and forecast the demands for bikes as indicated by the higher MSE value. But for the 30 min interval variation the demands are effectively assessed. And the difference in MAE and RMSE for both intervals is almost similar which indicates identical variation in error is present irrespective of the data samples.

Table 5: TAGCN experiment results

| Bikes Demand | | | |
|---|---|---|---|
| Interval | MSE | MAE | RMSE |
| 30min | 1.01 | 0.54 | 0.73 |
| 60min | 3.63 | 1.26 | 1.12 |
| Docks Demand | | | |
| Interval | MSE | MAE | RMSE |
| 30min | 0.48 | 0.36 | 0.60 |
| 60min | 1.78 | 0.73 | 0.86 |

## 5.5 Discussion

The study evaluated the performance of the models for short and long interval data (30 min and 60 min respectively). The comparison of the model for short interval indicated that performance of STGCN model outer performed the other models with MSE of 0.74

and 0.094 for both bikes and docks demand. The results obtained by the STGCN model were comparatively better then the results obtained by the multi-graph convolution model framework in Kim et al. (2019) for a bike share network. Thus, implying the effectiveness of incorporating neighbouring stations as a factor for prediction. LSTM model was also quite effective in forecasting demands and provided a balanced result for both bikes and docks. It also verified the works that proposed the LSTM model's superiority for precisely determining forecasts than the ARIMA model. (Siami-Namini et al.; 2018). The better performance of STGCN than TAGCN model inferred that the GCNConv layer (Kipf and Welling; 2016) was more effective in assigning weights to edges for the metro bike-share data.

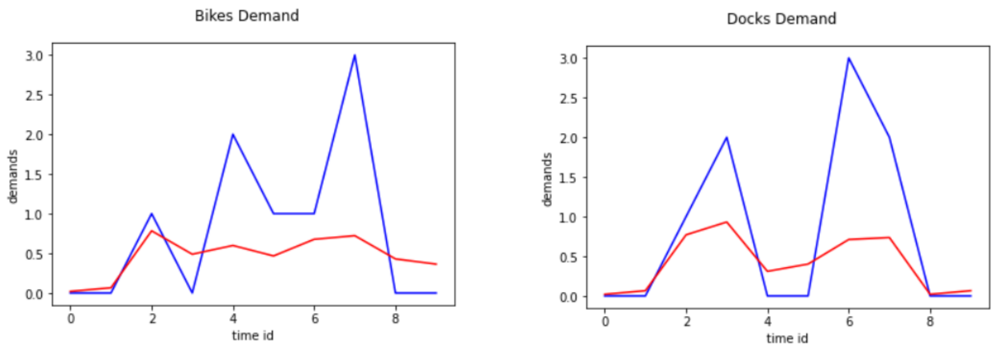


Figure 7: STGCN model bikes and docks demand prediction for 60 min interval

But for the long interval a similar trend of increased error rates was observed for all the models. However for this scenario, the LSTM was found to be more effective than the rest. The increased MSE with the increase of interval size can be associated to the high variation of demands present in the adjacent intervals. And it can be observed from figure 7 that while the STGCN model is capable of stimulating short fluctuations in demand, consecutive large fluctuations are still out of scope. A similar pattern was observed for TAGCN in figure 8 and even though TAGCN produced low MSE, its predictive capability was found to be poor. It can also be concluded that increasing the interval size increases the variance between the intervals, making it difficult to predict demands.

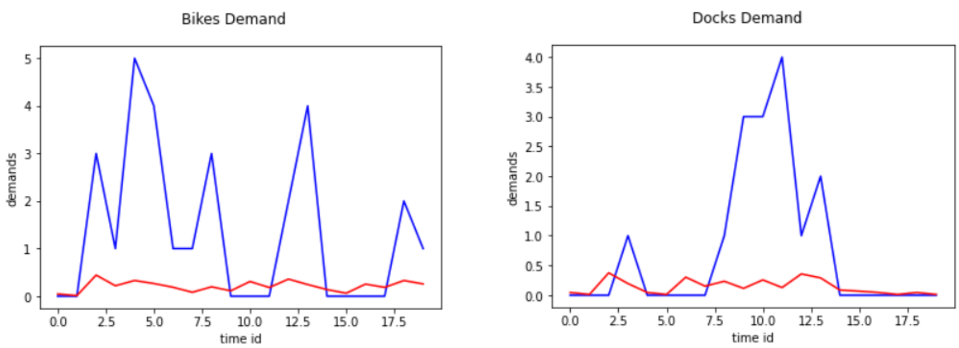


Figure 8: TAGCN model bikes and docks demand prediction for 60 min interval

# 6 Conclusion and Future Work

This study proposed the use machine learning techniques to identify the demands in a bike-sharing system. The study utilised the conventional ARIMA and LSTM model along with STGCN and TAGCN model predict demands. Moreover, the application of STGCN and TAGCN also calculated the adjacent bike stations effects on the another bike stations demand characterstics.

The demands were obtained for two-time intervals (30 min and 60 min), and the results for those intervals concluded that the demand frequency varies significantly in long intervals, finding it challenging for machine learning models to work effectively. Nevertheless, if the data is individually analyzed for each bike station then from past records, other demand characteristics can be defined to forecast demands like the LSTM model. On the other hand, the frequency of demands is widely distributed in short interval which facilitates the model's ability to predict accurate demands for the succeeding interval. Therefore, the STGCN model that, considers inter-station dependence and time-varying characteristics was comparatively more efficient in obtaining demands close to the actual value.

To further increase the performance of the model's external parameters can be included which affects the bike usage. Moreover, imitating the multi-graph convolution methodology (Chai et al.; 2018) for different time interval of data can also effective. And besides, the study scope allows it to be applied to other modes of transportation.

# References

Ashqar, H. I., Elhenawy, M. and Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions, *Case studies on transport policy* **7**(2): 261–268.

Box, G. E. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *Journal of the American statistical Association* **65**(332): 1509–1526.

Chai, D., Wang, L. and Yang, Q. (2018). Bike flow prediction with multi-graph convolutional networks, *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 397–400.

Du, J., Zhang, S., Wu, G., Moura, J. M. F. and Kar, S. (2017). Topology adaptive graph convolutional networks.

Eunbee Jang, A. (2019). Pytorch: Introduction to neural network—feedforward neural network model.
**URL:** *https://medium.com/biaslyai/pytorch-introduction-to-neural-network-feedforward-neural-network-model-e7231cff47cb*

Faghih-Imani, A., Hampshire, R., Marla, L. and Eluru, N. (2017). An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville, *Transportation Research Part A: Policy and Practice* **97**: 177–191.

Giot, R. and Cherrier, R. (2014). Predicting bikeshare system usage up to one day ahead, *2014 ieee symposium on computational intelligence in vehicles and transportation systems (civts)*, IEEE, pp. 22–29.

Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J. and Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system, *Pervasive and Mobile Computing* **6**(4): 455–466.

Kim, T. S., Lee, W. K. and Sohn, S. Y. (2019). Graph convolutional network approach applied to predict hourly bike-sharing demands considering spatial, temporal, and global effects, *PloS one* **14**(9): e0220782.

Kipf, T. N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks.

Mikolov, T., Karafiát, M., Burget, L., Cernocky, J. and Khudanpur, S. (2010). Recurrent neural network based language model, in 'eleventh annual conference of the international speech communication association'.

Ohrn, A. (2020). London bike ride forecasting with graph convolutional networks.
**URL:** *https://towardsdatascience.com/london-bike-ride-forecasting-with-graph-convolutional-networks-aee044e48131*

Pan, Y., Zheng, R. C., Zhang, J. and Yao, X. (2019). Predicting bike sharing demand using recurrent neural networks, *Procedia computer science* **147**: 562–566.

Rathor, S. (2018). Simple rnn vs gru vs lstm :- difference lies in more flexible control.
**URL:** *https://medium.com/@saurabh.rathor092/simple-rnn-vs-gru-vs-lstm-difference-lies-in-more-flexible-control-5f33e07b1e57*

Raviv, T., Tzur, M. and Forma, I. A. (2013). Static repositioning in a bike-sharing system: models and solution approaches, *EURO Journal on Transportation and Logistics* **2**(3): 187–229.

Siami-Namini, S., Tavakoli, N. and Namin, A. S. (2018). A comparison of arima and lstm in forecasting time series, *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, pp. 1394–1401.

Singhvi, D., Singhvi, S., Frazier, P. I., Henderson, S. G., O'Mahony, E., Shmoys, D. B. and Woodard, D. B. (2015). Predicting bike usage for new york city's bike sharing system, *Workshops at the twenty-ninth AAAI conference on artificial intelligence*.

VE, S. and Cho, Y. (2020). A rule-based model for seoul bike sharing demand prediction using weather data, *European Journal of Remote Sensing* pp. 1–18.

Xu, C., Ji, J., Liu, P. and Peng, L. (2018). Forecasting the travel demand of the station-free sharing bike using a deep learning approach, *Technical report*.

Yoon, J. W., Pinelli, F. and Calabrese, F. (2012). Cityride: a predictive bike sharing journey advisor, *2012 IEEE 13th International Conference on Mobile Data Management*, IEEE, pp. 306–311.

Yu, B., Yin, H. and Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* .
**URL:** *http://dx.doi.org/10.24963/ijcai.2018/505*

Zhang, C., Zhang, L., Liu, Y. and Yang, X. (2018). Short-term prediction of bike-sharing usage considering public transport: A lstm approach, *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1564–1571.