Project report

Prediction and comparison of daily demand for public bikes in Seoul with deep learning LSTM models

Hyunjung Byun, Yaejin Lee

Abstract: The purpose of this research is to predict the daily demand for public bikes in Seoul using deep learning methods. Dataset was created using Seoul's bike sharing system, with the addition of temperature, humidity, windspeed and micro dust data. To build various Long short-term memory (LSTM) models which are an artificial recurrent neural network, experiment was conducted using each of three LSTM models (Bidirectional LSTM, Vanilla LSTM, Stacked LSTM) to see which model would improve the accuracy the most. To evaluate the model performance, RMSE is used. This research was conducted by implementing TensorFlow and Keras libraries. After the result of the experiments are given, this paper concludes by giving further possible researches to conduct based on this research.

Keywords: Deep learning; RNN; Neural Network; LSTM; TensorFlow, Keras, Python; Bike sharing; Prediction; Vanilla LSTM; Stacked LSTM; Bidirectional LSTM

---

1. Introduction

The goal of this research is to predict the number of daily public bike within Seoul by forming a deep learning LSTM model. To do so, Seoul bike demand data was retrieved from the public data portal from 2018.01.01 to 2020.06.31 along with the humidity, temperature, wind speed and microdust data for the corresponding time period. The bike data used to form the model is the data of Seoul's public bike sharing system called "Ddareungi.[1]" It is also named "Seoul Bike" in English and was first introduced to Seoul in October 2015. Starting with 150 stations and 1500 bike, the number of stations has increased steadily to cover new districts other than Han River where it first started. In 2019, around 1540 stations and 20000 bikes are available and as of December 2020,

---

1. [1] https://en.wikipedia.org/wiki/Ddareungi

it has increased to more than 300000. Seoul still has a plan to increase the number of station and bike in future as the demand for Seoul bike is constantly growing. Thus, the research we provide will be useful and practical for Seoul in predicting the demand for bikes and will help Seoul in their decision makings and policies related to bike use.

2. Problem statement

Popularity of the public bike has been constantly increasing within last two years in Seoul, and consequently the need to accurately forecast the demand of these bikes has also grown. The demand of public bike more than two years ago was about 10,000, but its demand is doubled now, so further research for the future prediction is necessary. Consequently, the goal of this research is to predict the total number of daily public bike usage using deep learning LSTM methods which already has proved better performance and higher accuracy than compared other techniques such as machine learning. In most previous research on the topic of Seoul's public bike demand prediction, the models were formed using traditional machine learning methods and statistical techniques, and so this research specifically focuses on the deep learning methods for prediction. Moreover, this paper compares three LSTM models using a well known and popular LSTM models (Vanilla LSTM, Stacked LSTM and Bidirectional LSTM) to see whether there is a difference in accuracy between these different models. The results provided by this research will prove useful in assisting Seoul in making decisions or policies related to Seoul bikes by providing a more accurate prediction of its demand.

3. Related research

Recent research were conducted on prediction for Seoul's public bike daily usage by using techniques other than deep leaning method. "Development of Demand Forecasting Model for Seoul Shared Bicycle" conducted by Heejong Lim and Kwanghun Chung in 2019, propose and developed a forecasting model for demand for Seoul shared bicycle using statistical time series analysis and in particular the Holt-Winters method which was used to forecast electricity demand and sensitivity analysis on the parameters that effect on real demand forecasting.

In "Prediction for public bike demand within Seoul using cluster analysis" conducted by Changhwan Lee and Kungok Kim in 2019, researchers used machine learning techniques. The found out optimal clusters of public bike rental spots by applying time series K-means analysis several times and forecasted bike demand for each cluster using Random forest method, allowing efficient and accurate method of forecasting the demand for bikes compared to the traditional statistical methods.

In “Analysis of data and forecast of public bicycle demand according to weather factor and public bicycle rental rate” conducted by Sungmin Hong and HyunCheol Kim in 2017, ensemble leaning technique, gradient boosting regression and random forest regression was used to predict public bike demand in Silicon Valley area. They compared those machine learning models and found out the ensemble learning has the best performance. However, they mentioned at the end of the thesis that they need further implementation with building a deep learning model for better performance.

As reviewed above three recent papers, all of them used statistical testing and machine learning methods. Many other papers not mentioned specifically here also focused on using machine learning method. Thus in this research, deep learning LSTM model will be used to create the prediction models and models made with different LSTM models will be compared to see whether a difference in accuracy is present among them.

4. Idea for solution

To predict daily demand for bike within Seoul, “deep learning methods” are used. Specially, our data is time series dataset, so three various LSTM methods are used for forecasting and compared. First, Vanilla LSTM will be used. A Vanilla LSTM has a single hidden layer of LSTM units, and an output layer used to make a prediction. Then, Stacked LSTM techniques will be used for modelling. A Stacked LSTM has Multiple hidden LSTM layers which can be stacked one on top of another. Lastly, Bidirectional LSTM will be used. This LSTM utilizes a sequence processing model that consists of two LSTMs; one taking the input in a forward direction, and the other in a backwards direction. Bidirectional LSTMs effectively increase the amount of information available to the network, improving the context available to the algorithm. Moreover, Activation function “Relu” and optimizer “Adam” will be used with the dropout rate of 0.4. Detailed explanation on each of these LSTM models and how each of these models are constructed is given within the “design of neural networks and its implementation”.

4. Design of LSTM models

As mentioned previously, three different LSTM models will be used in this research such as bidirectional LSTM, Vanilla LSTM, and Stacked LSTM. A diagram is provided below to illustrate how each of these models work. Bidirectional LSTMs are an extension of traditional LSTMs that can improve model performance on sequence classification problems. In problems where all timesteps of the input sequence are available, Bidirectional LSTMs train two instead of one LSTMs on the input sequence. The first on the input sequence as-is and the second on a reversed copy of the input sequence. This can provide additional context to the network and result in faster and even fuller learning on the problem. A Vanilla LSTM is an LSTM model that has a single hidden layer of LSTM units, and an output layer used to make a prediction. A Stacked LSTM architecture can be defined as an LSTM model comprised of multiple LSTM layers. An LSTM layer above provides a sequence output rather than a single value output to the LSTM layer below. Specifically, one output per input time step, rather than one output time step for all input time steps.

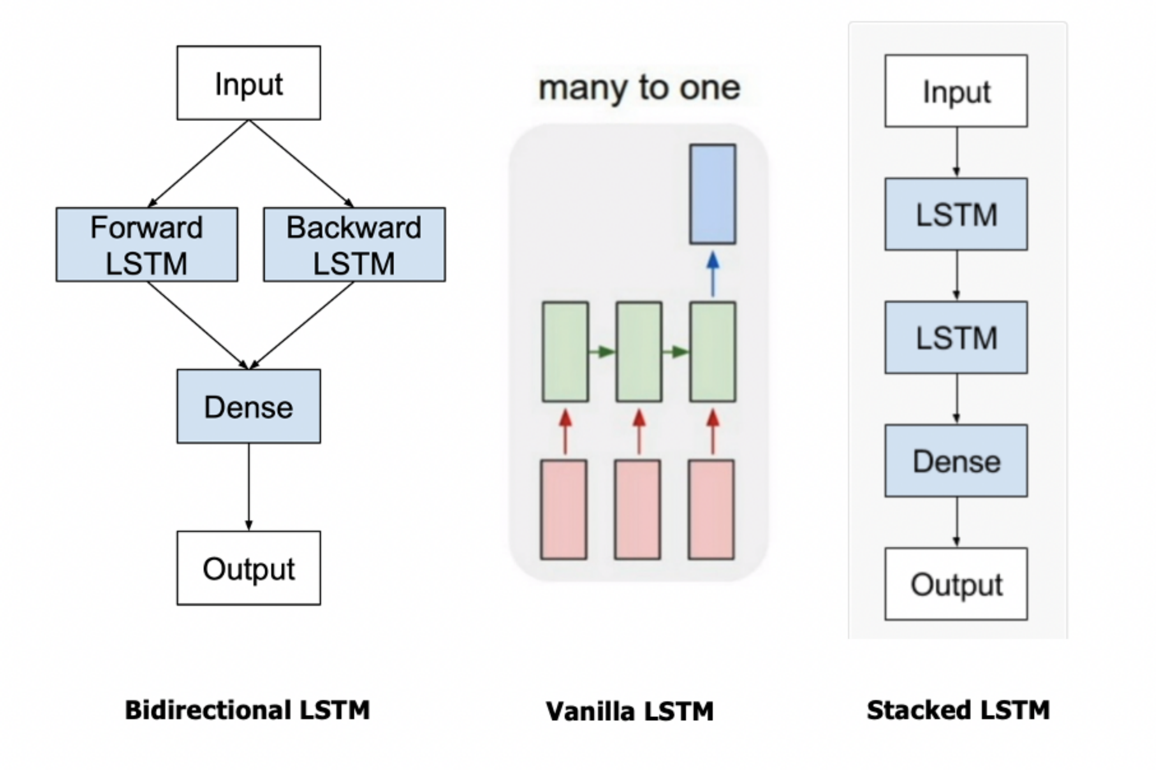


Figure 1. Above picture is three different LSTM techniques and visualized how they work each way. As followed the right to the left, bidirectional LSTM, Vanilla LSTM, and Stacked LSTM.

4. Dataset description

The dataset used for analysis has been retrieved from different sources. There are a total of 9 columns and 921 rows without any missing values in within the dataset which is given in the table below. The information of daily number of public bike rentals in Seoul from 1/1/2018 to 6/30/20 is from "공공데이터포털."Temperature(℃), Humidity(%),Wind speed(km/h), Micro dust(PM-2.5 (㎍/m3) are brought from "기상청 날씨누리." Micro dust(PM-2.5 (㎍/m3) is brought from "서울특별시 대기환경정보." The measurement of PM-2.5 can be described as particles smaller than 2.5 microns PM2.5 and ㎍/m3 means micrograms per cubic meter. In addition, other information about the day such as "day of the week", "weekend vs weekday", and "holiday vs working day" was used from calculating with excel. Day of week is described Sunday as 1, Monday is 2, ..., and Saturday as 7. If the day is weekend, it marks as 1. If not, it marks as 0. Holiday is also marked in the same way 1 or 0. So, the predictors are humidity, wind speed, micro dust, day of the week, weekend vs weekday, holiday vs working day. These data was used to examine how each variable has effects the daily number of public bike demand in Seoul.

Variables	Description	Data source
Date	1/1/2018 to 6/30/20	
Number of rentals per day	Daily number of public bike usage in Seoul	공공데이터포털
Temperature(°C)	Daily average temperature	기상청 날씨누리
Humidity(%)	Daily average humidity	기상청 날씨누리
Wind speed(km/h)	Daily average wind speed	기상청 날씨누리
Micro dust(PM-2.5 (㎍/m3)	Daily average micro dust	서울특별시 대기환경정보
Day of the week	1 to 7 / Sun to Sat	Excel
Weekend vs weekday	1 or 0 / Yes or No	Excel
Holiday vs working day	1 or 0 / Yes or No	Excel

Table 1. Above picture is our dataset which has nine variables. The number of rentals per day would be our dependent variable and the rest of them would be our independent variables. The rightest column indicated data sources.

5. Exploring dataset with EDA

In figure 2, time series graphs for temperature, humidity, windspeed, and micro dust from 1/1/2018 to 6/30/20 are given. There are cycles in temperature variable, since it is based on the four seasons of the year. In figure 3. the graphs show each numerical variable distribution such as daily number of public bike rentals, wind speed, and micro dust. You can find out that bike count, windspeed, and micro dust graphs are skewed to the right. In figure 4, graphs show the relationship between daily number of bike demand and temperature, micro dust, humidity, and wind speed. Daily number of bike demand increased as Temperature gets higher, but there is no visible correlation between humidity and Daily number of bike demand. In figure 5, graphs show public bike demand by month by year. Bike demand is higher in summer and fall season compared to winter season. Moreover, the demand is fast growing between 2018 and 2019 and still portrays increment of daily demand in 2020. In figure 6, graphs show differences of public bike demand followed by each day of week, weekend or weekday, and holiday or normal day. Average Bike demand on Tuesday has the highest point, and average Bike demand on weekdays is higher than weekend. Average Bike demand on holidays is lower than other days.

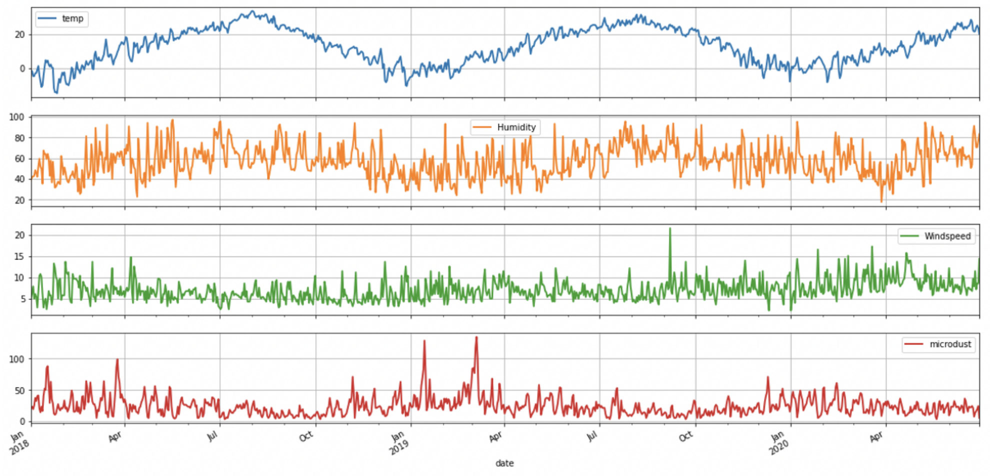


Figure 2. Above graphs from top to bottom are temperature, humidity, windspeed, and micro dust from 1/1/2018 to 6/30/20.

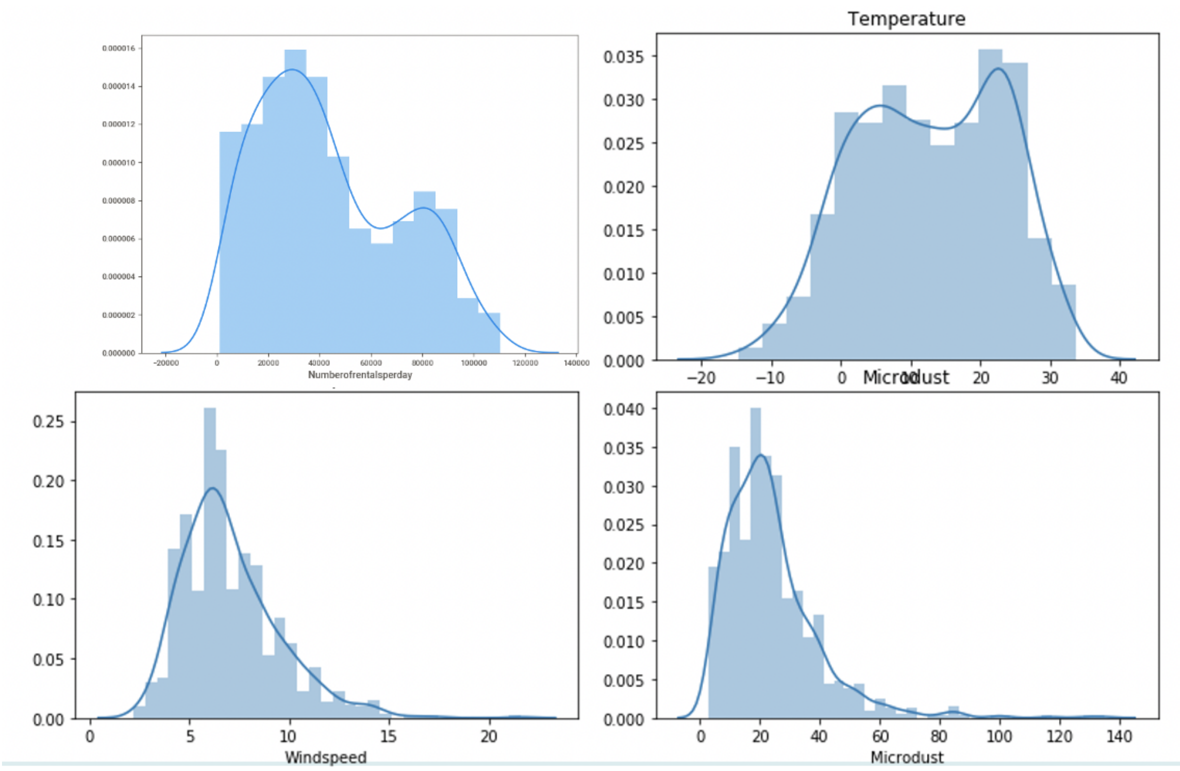


Figure 3. Above graphs show each numerical variable distribution. The top- left graph shows the distribution of daily number of public bike rentals in Seoul. The top- right graph shows the distribution of temperature. The bottom- left graph shows wind speed and the bottom- right graph, micro dust.

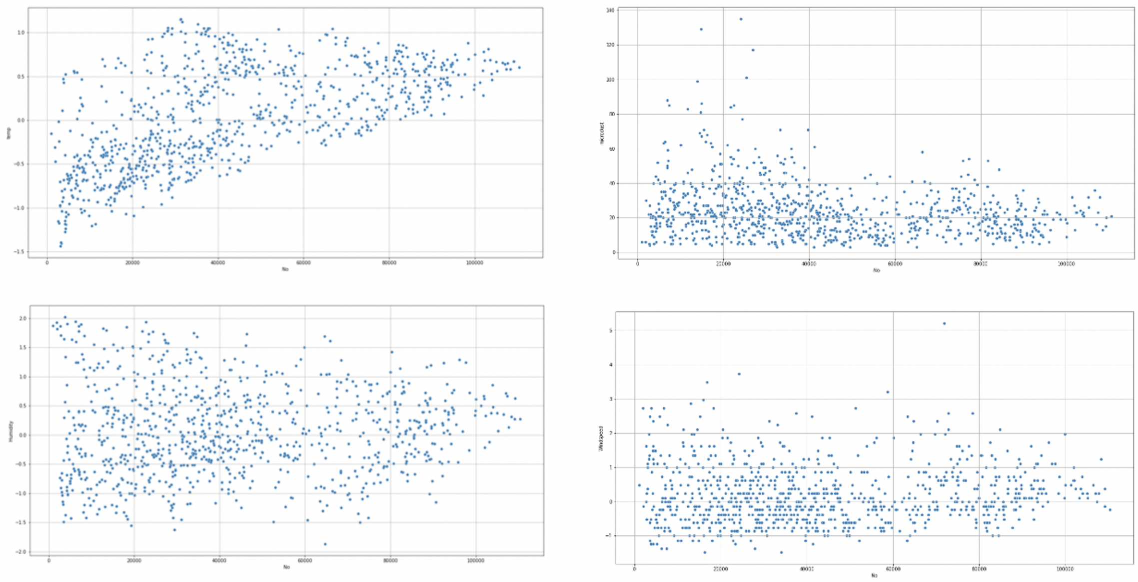


Figure 4. Above graphs show the relationship between daily number of bike demand and temperature, micro dust, humidity, and wind speed as followed as the top-left, top-right, bottom-left, and bottom-right.

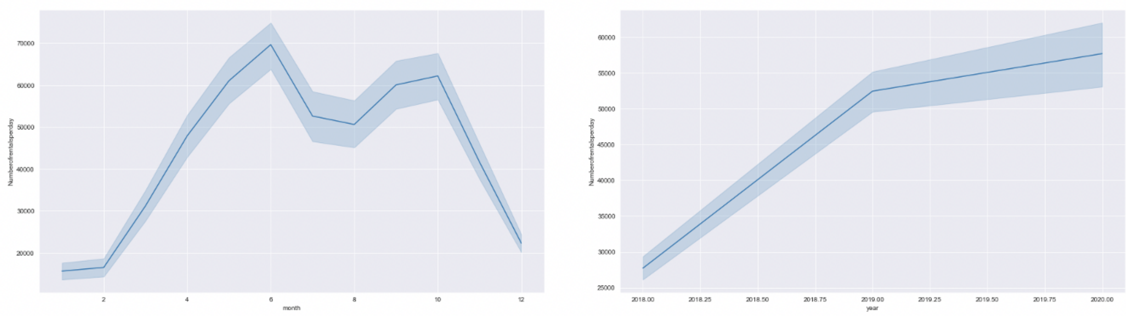


Figure 5. Above graphs show public bike demand by month on the left, and public bike demand by year on the right.

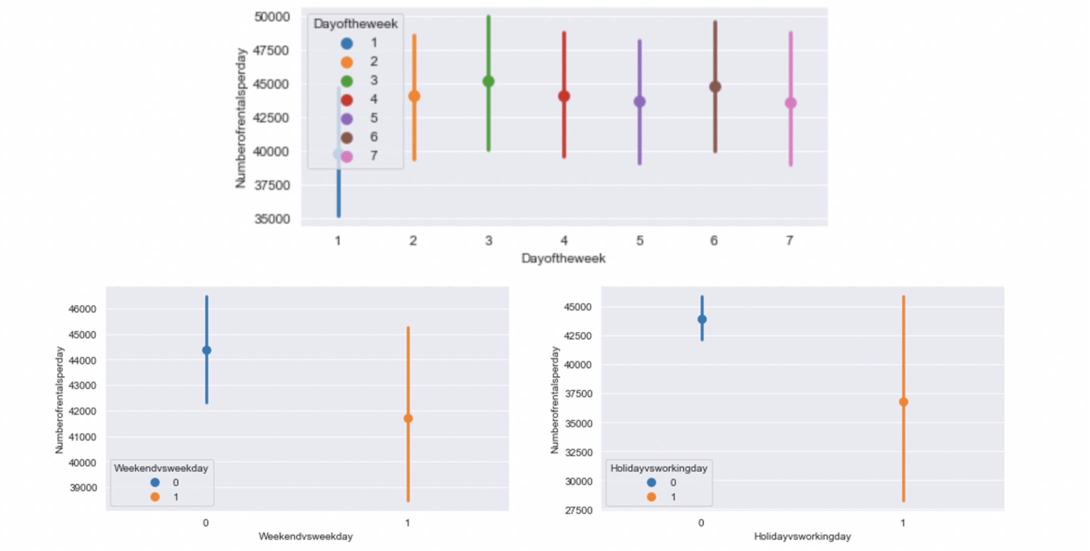


Figure 6. Above graphs show differences of public bike demand followed by each day of week, weekend or weekday, and holiday or normal day.

6. Design of neural network and its implementation

Before explaining further on the different LSTM models used for modelling the prediction of bike demand, it is necessary to briefly look into the basic concept of LSTM. LSTM(also known as Long Short Term Memory) is a special type of RNN model in order to solve the problem of vanishing or exploding gradients. This problem of vanishing gradient or exploding gradient happens because the information that RNN can access and utilize as the time goes by becomes limited and consequently, the back-propagated error blows up or diminishes. Thus, by using the LSTM model, it is possible to address issues that we were previously unable to address such as carrying information from time steps far back. The most important concepts for LSTM model include forget gate, input gate, output gate and the cell state. These cells help in retaining the relevant information used in future predictions and acts as a transportation highway that "delivers" relevant information down the cell sate. It can thus be especially useful within the field of Natural Language Processing since capturing only the necessary information is critical in inferring information from long passages or with long sentences. With the previous RNN model, the information from the earlier sentences get lost as the paragraph becomes long. But with the "forget gate", information from the previous state and the information from the current input passes into the so called "activation function"(with sigmoid and relu being two of the most popularly used) and the output is given as a value between 0 and 1. The closer the value is to 0, it is to be forgotten, the close the value is to 1, it is to be remembered. The "input gate" then decides which value to add in the current step while the output gate determines what the next hidden state should be.

There are many different types of LSTMs such as Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, CNN-LSTM, TCN(Temporal Convolutional Network) LSTM, Multivariate Attention LSTM-FCN(MALSTM-FCN). The most popular and commonly used LSTMs are Vanilla, Stacked and Bidirectional LSTM which were also used within the experiment to make deep learning bike models. Consequently, prior to looking at the models made, a detailed explanation of each 3 types of LSTMs is given.

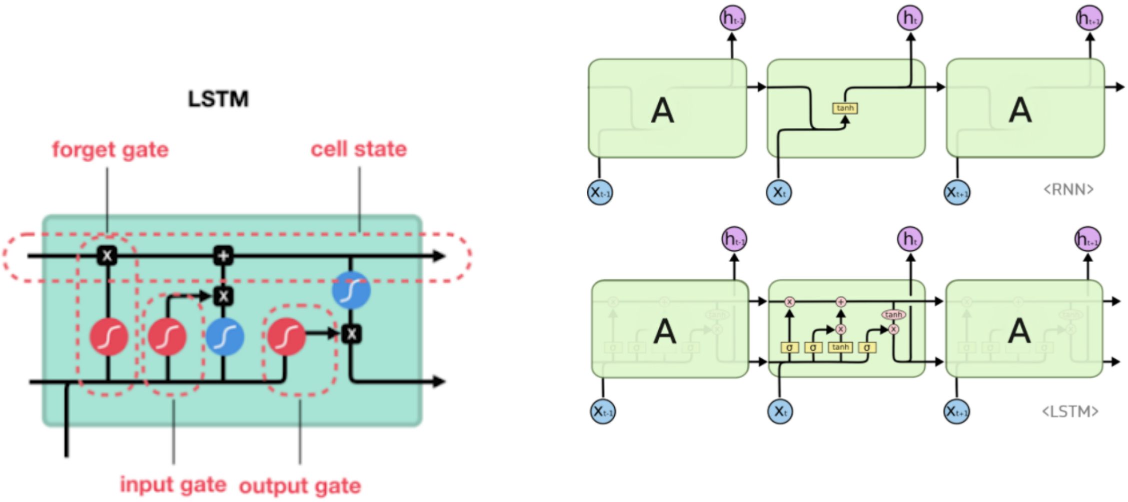
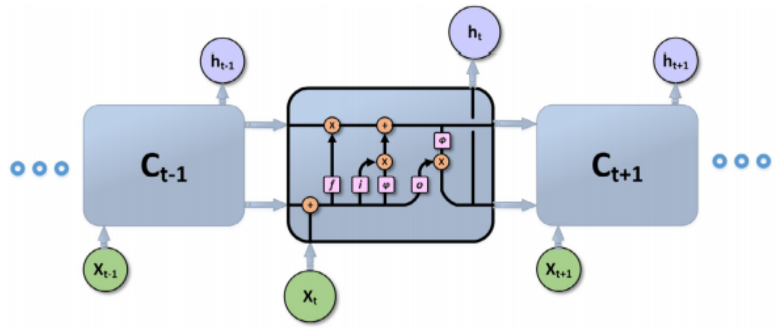


Figure 7. The left figure highlights the important terminologies: "forget gate", "cell state", "input gate" and "output gate"> The right figure shows the difference between the RNN and LSTM.

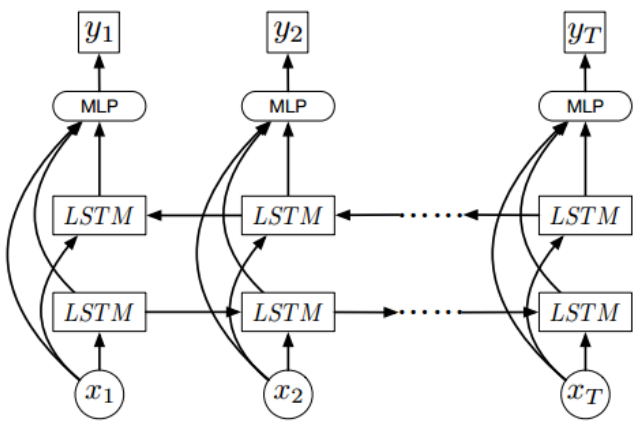
1) Vanilla LSTM

Vanilla LSTM is the simplest form of an LSTM model which has a single layer of LSTM units and an output layer used to make a prediction. It is widely used for the univariate time series forecasting but can still be used for multivariate time series models as well. Even though Vanilla LSTM is composed of a single layer of LSTM units, the authors that first introduced the LSTM model claimed that LSTM performs reasonably well on various datasets and consequently, Vanilla LSTM has been used by many scholars proving and showing a significant improvement in the performance compared to the original RNN model (Yuting, Mei, Shaopeng, Yingqi, 2017) Moreover, since it is considered the "basic" model, it is also used as a baseline when comparing it with other models (Xue, Huynh, Reynolds, 2017). For example, in the research *"High Precision Dimensional Measurement with Convolutional Neural Network and Bi-directional LSTM"* (Wang and Chen, 2019), the authors used CNN and Bi-directional models and compared them with the Vanilla LSTM model to show which model was doing better. The term "Vanilla" is given since it refers to the original Vanilla ice cream as not having anything topping and was also introduced with Vanilla RNN as well. A diagram shown below is a good visualization of the Vanilla LSTM in that it clarifies the three gates and how they are associated with the activation functions and the cell states. It is not so different from the diagram shown when LSTM was first introduced and hence it is also the easiest to understand since it has a relatively simple structure compared to other LSTM models.

2) Stacked LSTM

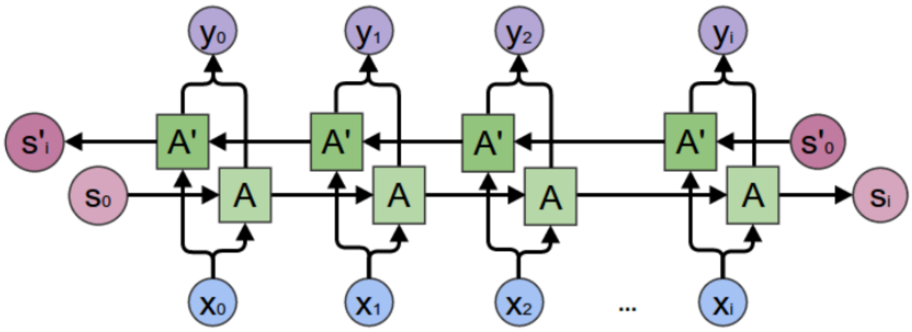


Stacked LSTM model refers to the LSTM model that is comprised of multiple LSTM layers. By having multiple layers, it is possible to have one output per input time step and is also widely known as the deep LSTM. This model was first applied to the speech recognition problem and has been widely known since due to its success in solving the speech recognition within Natural Language Processing. (Graves, et al, 2013), Having multiple layers mean that it is able to add levels of abstraction on input observations over time allowing the effect of "chunking" observations over time or "representing the problem at different time scales". The stacked LSTM offers a stable technique for challenging sequence prediction problems with a greater model complexity and so allows a more complex input pattern. Alongside with Vanilla LSTM model, Stacked LSTM is also popularly used and many researches presents the application of Stacked LSTM model performing better than the usual machine learning methods. For example, Du, Zhang, Nguyen, Han(2017) proposed in their paper *"Stacked LSTM Deep Learning Model for Traffic Prediction in Vehicle-to-Vehicle Communication"* that stacked LSTM model worked better than the logistic regression in terms of accuracy and precision.



3) Bi-directional LSTM

Bidirectional LSTM are extension to the traditional LSTMs in the problem since where all timesteps of the input sequence are available, bidirectional LSTM trains a forward sequence of LSTMs and another backward of sequence of LSTMs. This architecture is achieved by duplicating the first recurrent layer in the network. The idea of Bidirectional LSTM is understood well in terms of applying it in the NLP context: For example, when we try and solve the problem of speech recognition, we try and use the whole context of what is being used to interpret rather than going in a single "linear" direction for interpretation. Graves, Schmidhuber, and Phoneme claimed in their paper "*Classification with Bidirectional LSTM and Other Neural Network Architectures*" (2005) that although bidirectional LSTMs may not make sense for all the sequence prediction problems, but can offer benefits in terms of better results within the "appropriate domains". There is still an on-going research in clarifying to which "appropriate domains" the bidirectional LSTMs work out, but so far, researchers claim that Bidirectional LSTMs work the best with Natural Language Processing.



The three different LSTM models introduced in the previous paragraphs were implemented in this research to model the demand prediction of Seoul bikes and to see whether there are any differences in the accuracy of the prediction among these three models. For comparison of theses different models, this research modelled three LSTMs using the same optimizer, dropout rate, activation function, epochs, batch size. For the loss function mean squared error was used and the root mean squared error was used to evaluate and compare the models. Moreover, 10% of the last time series were used for test data since for time series data are not picked at random and allotted 30~40% but uses the last 10~20% of data that is not mixed.

7. Experiments

The steps for the experiment are as follows:

1) Load the dataset
2) Undergo preprocessing and feature engineering
   - Scale the data using RobustScaler
   - Transform the data into a sequence dataset
3) Use Vanilla, Stacked and Bidirectional LSTM model
   - Activation function: "relu"
   - Dropout rate: 0.4
   - Optimizer: Adam

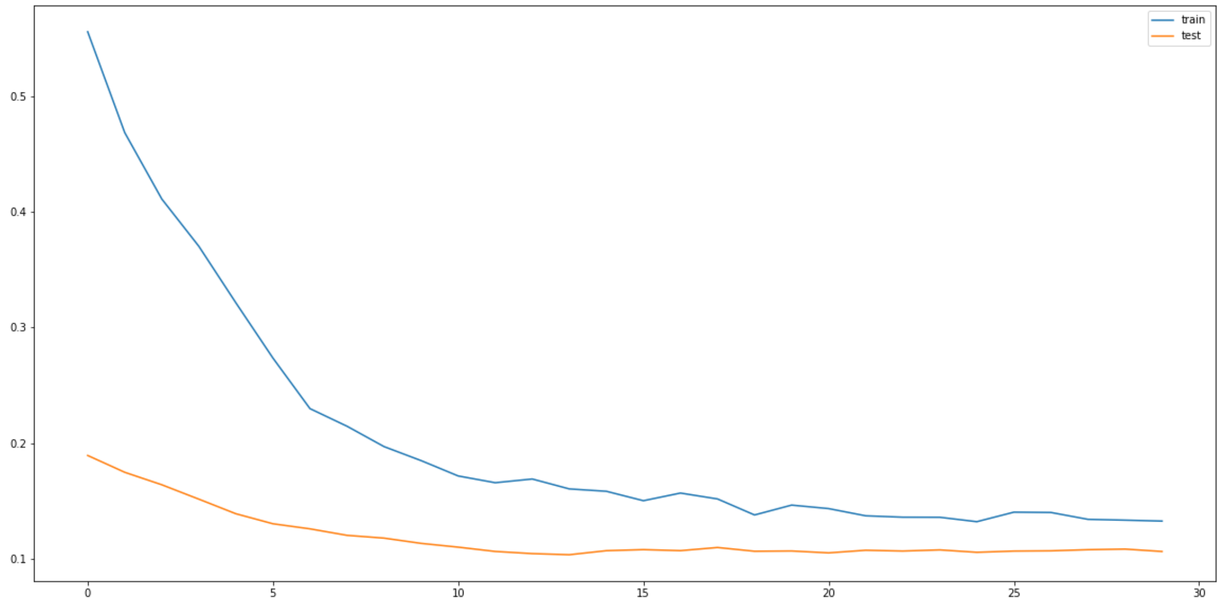
- Epochs: 30
- Batch Size: 50
- Loss: Mean squared error

Note that after several initial experiments including the weekday vs weekend, holiday vs working day and the day of the week variable, the RMSE for including these datasets were considerably high. Therefore, the model selected and used temperature, humidity, wind speed and microdust as explanatory variables when setting up the demand forecasting model.

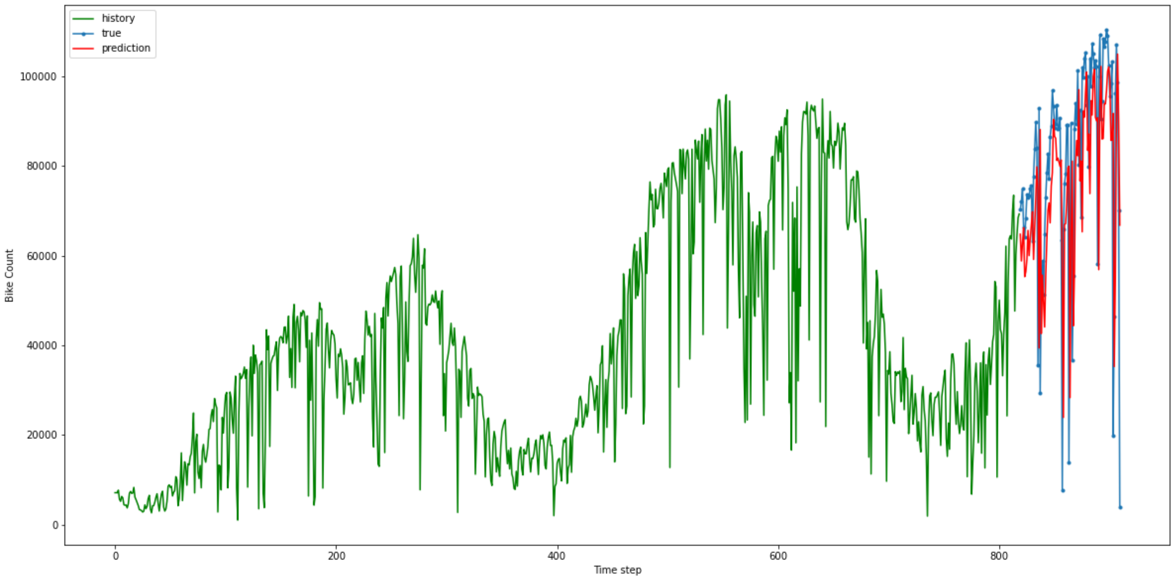
4) Evaluate each model using the calculated RMSE for each model
- Scale back the data and use the RMSE for evaluation criteria

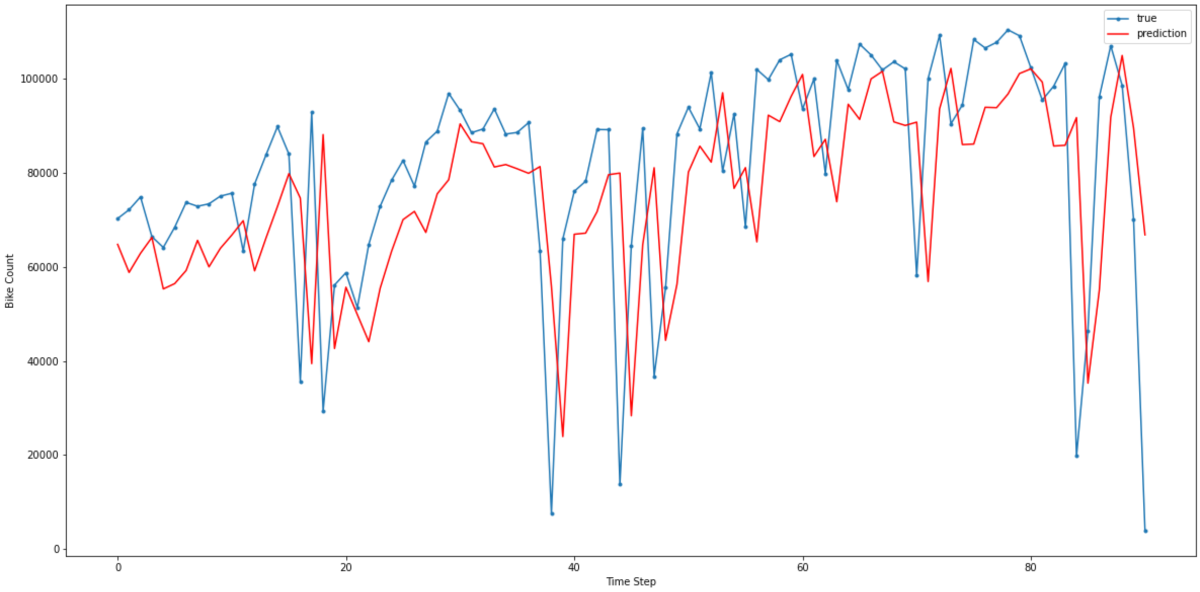
The results Vanilla LSTM model:

The loss(mean squared error) for train and test data decreases rapidly but does not overfit the data even after 30 epochs.



The true data and the predicted data for the last 10% is visualized below

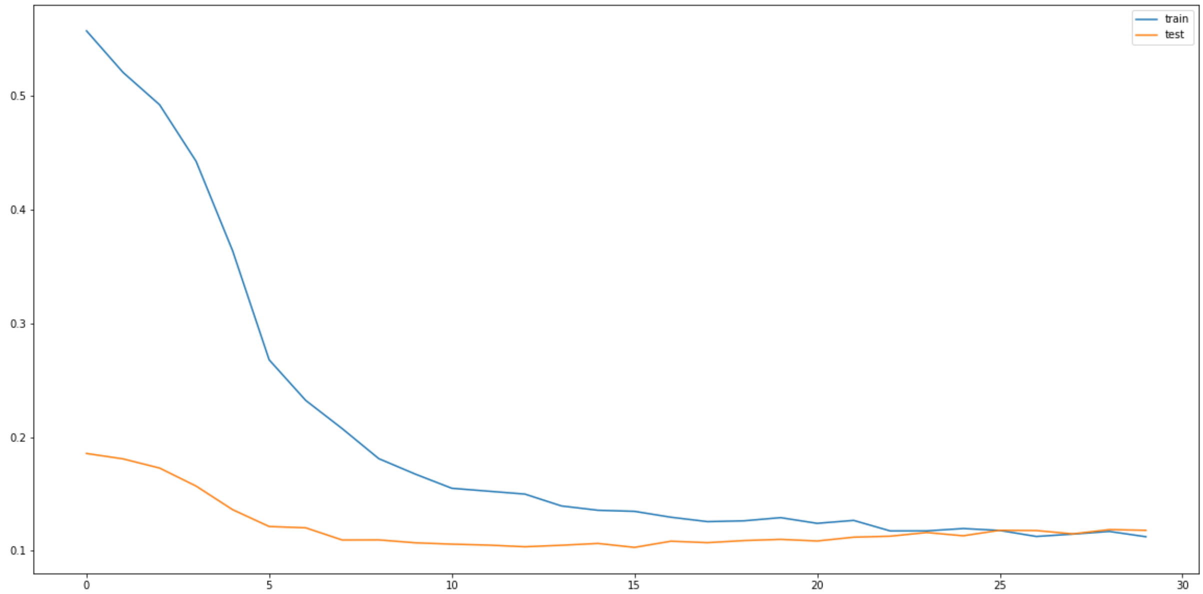


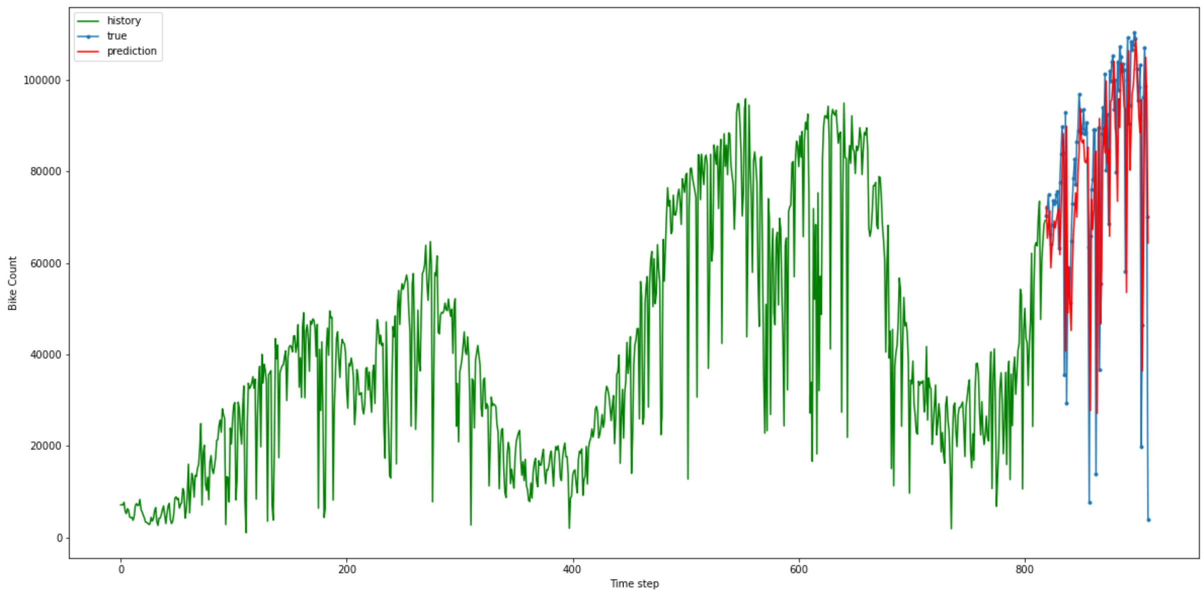


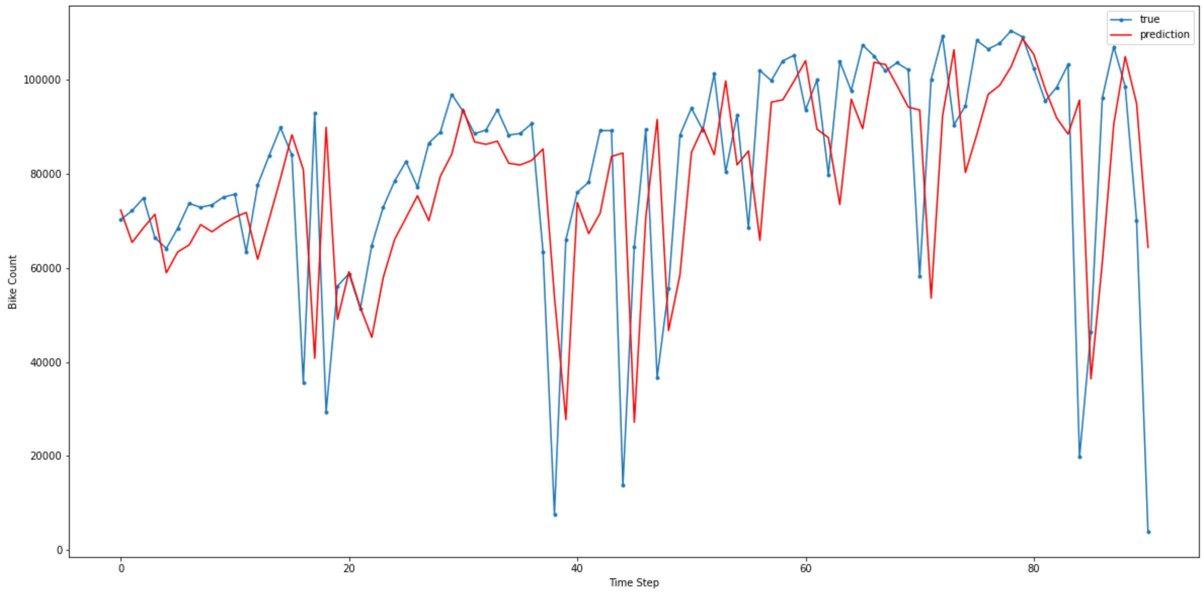
It is possible to see that the predicted red lines are able to pick up the trend of the true values and the overall RMSE of the Vanilla LSTM model is 151.9

The results for Stacked LSTM model:

Unlike the Vanilla LSTM, we can see an overfitting of the data after around 22 epochs


train
test
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30

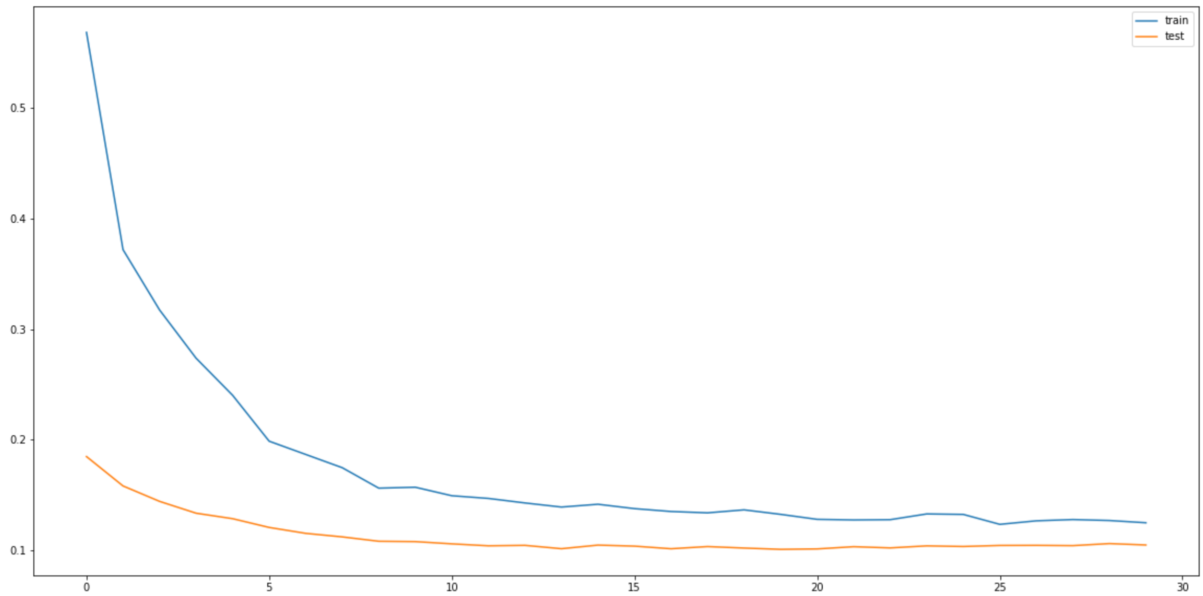

history
true
prediction
Bike Count
100000
80000
60000
40000
20000
0
0
200
400
600
800
Time step



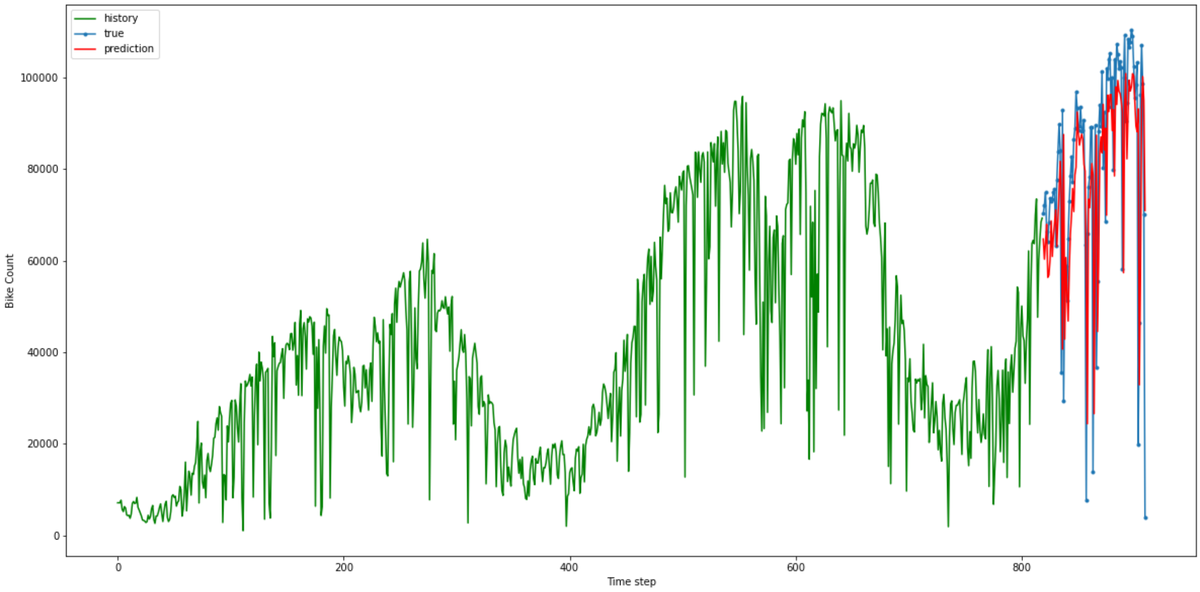
Just by looking at the visualized graph for prediction of the data, the prediction seems closer to the true data but calculated RMSE for the stacked LSTM model is 150.9 which is only a slight drop from the Vanilla LSTM. In addition, compared to other models, stacked LSTM model shows the tendency to overestimate the data at some points more than that of other models.

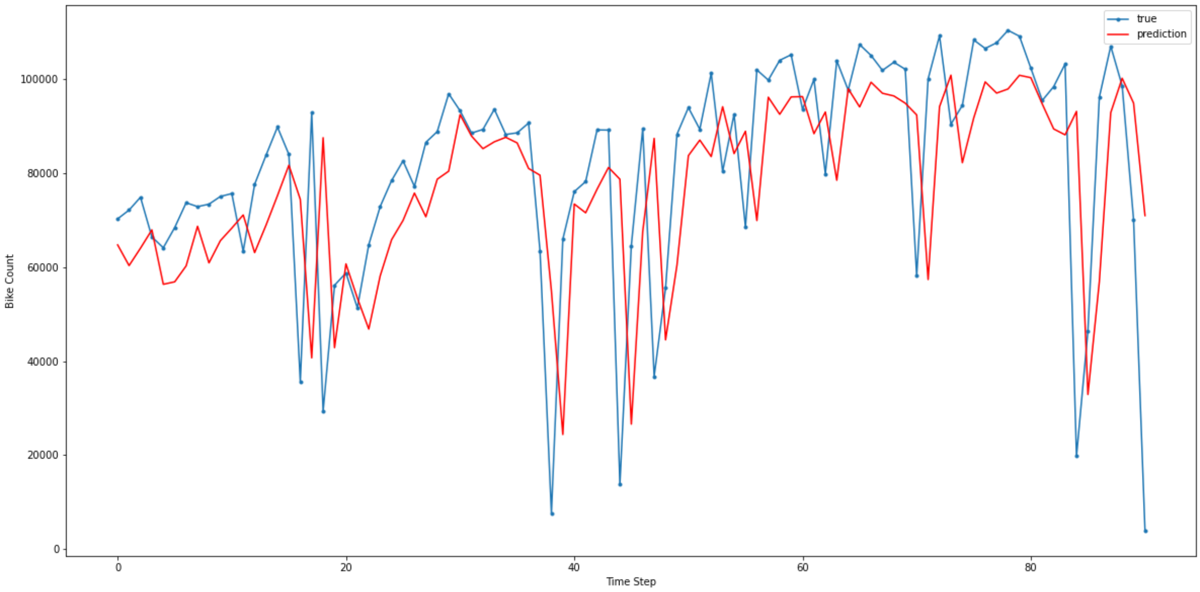
The results for the Bidirectional LSTM:

The mean squared error for the Bidirectional LSTM shows a similar trend after each epoch but shows no sign of data overfitting



The predicted visualizations are as follows:





Just from looking at the visualized graphs, the Bidirectional LSTM seemed to form predictions that were even more similar to the true values of the data, but the RMSE for Bidirectional LSTM is 149.7 which is only slightly lower compared to Vanilla LSTM and Stacked LSTM model.

8. Conclusion

To summarize, RMSE for each model is summarized in the table below. Although RMSE for the Bidirectional LSTM was the lowest with 149.7, the RMSE for each Vanilla LSTM and Stacked LSTM is 151.9 and 150.9 and so does not show a drastic difference between each of the models. The goal of this paper was to produce a deep learning LSTM model to predict the demand of Seoul bikes and also to compare whether there were differences in RMSE between different types of LSTM models. It is possible to conclude that Bidirectional LSTM model has the lowest RMSE and so is the model to predict the demand but the Vanilla LSTM and Stacked LSTM can also be used in the sense that the RMSE results do not show a drastic decrease.

Model	RMSE
Vanilla LSTM	151.9
Stacked LSTM	150.9
Bidirectional LSTM	149.7

The result of this experiment is in line with what the scholars said about "Vanilla LSTM performing reasonably well on various datasets".

9. Future work

The limitations of this research can be summarized as follows:

1) Limit in retrieving data of hourly demand for bikes
It was possible to retrieve data showing the daily demand for bikes but were not able to find an hourly demand for bikes in Seoul over a span of more than a year. If hourly demand for bikes were able to be retrieved over a span of more than 2 years, the models can be fine tuned even more to produce models with lower RMSE and consequently higher accuracy.

2) Use of only three main LSTM models for comparison
In this research, only the three main LSTM models(Vanilla LSTM, Stacked LSTM and Bidirectional LSTM models were used and compared. However, there are other types of LSTM models such as CNN-LSTM(Convolutional Neural Network LSTM), TCN-LSTM(Temporal Convolutional Network LSTM), MA LSTM(Multivariate Attention LSTM). Future works could compare the results of these models and verify whether there is a substantial decrease in the RMSE and if so, which model is most suitable in calculating the RMSEs.

References

Heejong Lim, Kwanghyun Jung (2019), *Development of Demand Forecasting Model for Seoul Shared Bicycle*, The Journal of the Korea Contents Association (한국콘텐츠학회논문지), Volume 19 Issue 1, p. 132-140, 1598-4877(pISSN), 2508-6723(eISSN)

Chanhwan Lee, Kyoungoh Kim (2019)*, Prediction of Bicycle Sharing System Demand in Seoul through Clustering Analysis*, 학위논문(석사)-- 서울과학기술대학교 대학원 : 데이터사이언스학과 2020. 2, I804:11034-200000291549

Sungmin Hong, Hyuncheol Kim (2017), *Analysis of data and forecast of public bicycle demand according to weather factor and public bicycle rental rate*

A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke and J. Schmidhuber,(2009), "*A Novel Connectionist System for Unconstrained Handwriting Recognition*," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 5, pp. 855-868, May 2009, doi: 10.1109/TPAMI.2008.137.

Yuting Wu, Mei Yuan(2017), *Remaining Useful Life Estimation of Engineered Systems using vanilla LSTM Neural Networks,* DOI: 10.1016/j.neucom.2017.05.063

H. Xue, D. Q. Huynh and M. Reynolds, (2018) *"SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction*," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, 2018, pp. 1186-1194, doi: 10.1109/WACV.2018.00135.

Wang, Y., Chen, Q., Ding, M., Li, J., (2019) *High Precision Dimensional Measurement with Convolutional Neural Network and Bi-Directional Long Short-Term Memory (LSTM)*, Sensors 2019, 19, 5302.

X. Du, H. Zhang, H. V. Nguyen and Z. Han, (2017), *"Stacked LSTM Deep Learning Model for Traffic Prediction in Vehicle-to-Vehicle Communication," 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, Toronto, ON, 2017, pp. 1-5, doi: 10.1109/VTCFall.2017.8288312.

Razvan Pascanu, Caglar Gulcehre, Kyunghyun Cho, Yoshua Bengio, (2014), *How to Construct Deep Recurrent Neural Networks,* arXiv:1312.6026

Graves A, Schmidhuber J. Framewise phoneme,(2005), *"Classification with bidirectional LSTM and other neural network architectures. "* Jun-Jul;18(5-6):602-10. doi: 10.1016/j.neunet.2005.06.042. PMID: 16112549