Deep Learning for Short-term bike-sharing demand prediction: LSTM and CNN-LSTM

Tri Dung Huynh and
Md Asaduzzaman

Table Of Content

Literature review

- Spatiotemporal data (STD) contain both space and time information. Spatiotemporal data has become a broad domain in ecology and environmental management, public safety, transportation, earth science, epidemiology, and climatology to discover patterns and knowledge (Jiang et al., 2019)
- Castro et al. (2013) studied traffic dynamics using large-scale taxi pickup and drop-off data by tracking time and GPS data
- Caulfiled et al. (2017) use the logistic regression model to examine the usage trends of bike-sharing in a small city by patterns such as frequency of usage, temperature, distance traveled, and time

Literature review

- A deep learning model Spatio-temporal graph neural network (STGCN) was used to predict traffic congestion using the data collected from Bluetooth sensors of passing cars in the study by Cunnov et al. (2021)
- A combination of k-nearest neighbor and LSTM was used to predict the traffic flow and achieve promising results compared with autoregressive integrated moving average (ARIMA), support vector regression (SVR), wavelet neural network (WNN), deep belief networks combined with support vector regression (DBN-SVR), and LSTM models in Luo et al. (2019) works

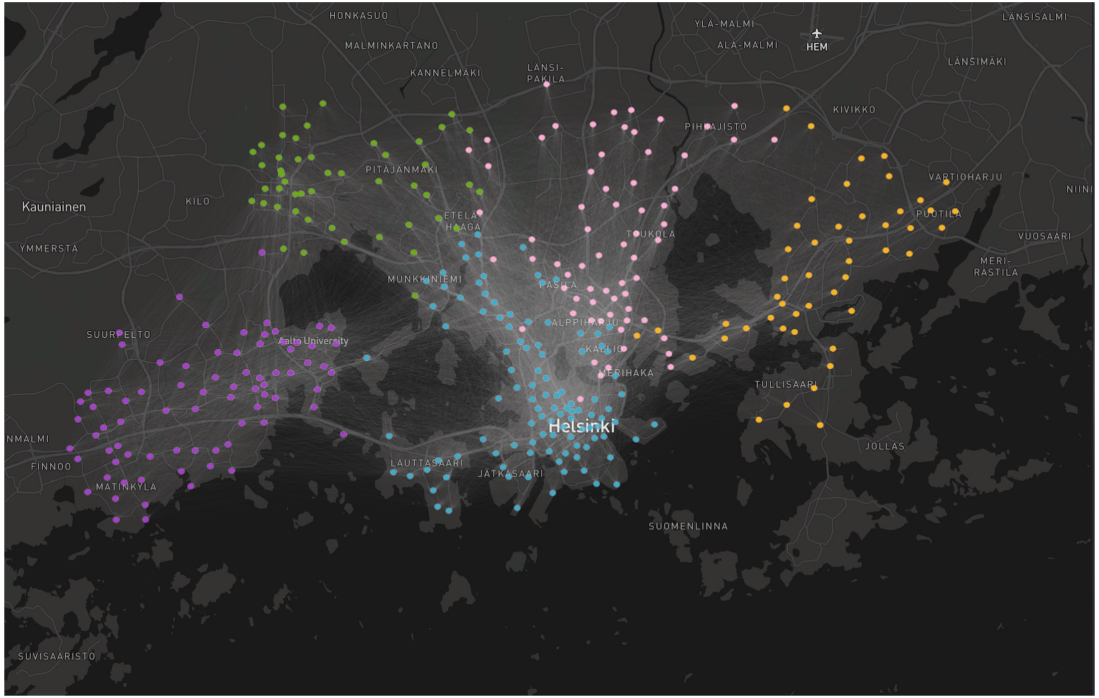
Literature review

- Ma et al. (2022) propose spatial-temporal graph attentional long short-term memory (STGA-LSTM) for short- term prediction of bike-sharing demand
- Mehdizadeh et al. (2022) apply the hybrid model of convolutional neural network (CNN) and long short-term memory (LSTM) to predict pickup demand for shared bikes in Montreal

Objective

- Predicting Short term demand of Helsinki Bike sharing service among 347 stations.
- Implementing forecasting model,including ARIMA, LSTM and CNN-LSTM for helsinki City bike data from 2016 to 2020.
- Compare there accuracy measured with different time lags.

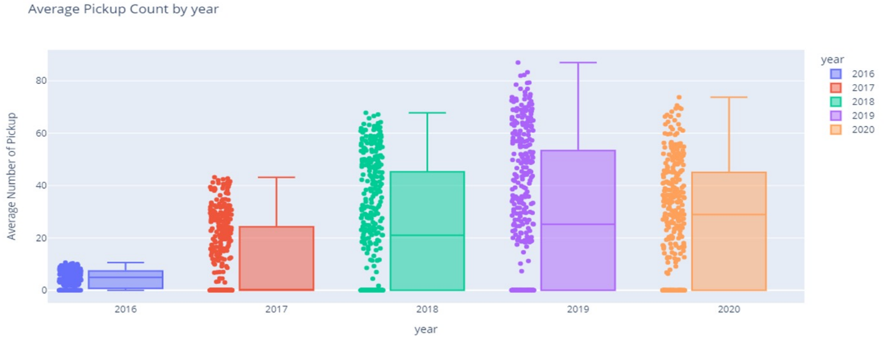
Helsinki City Bikes

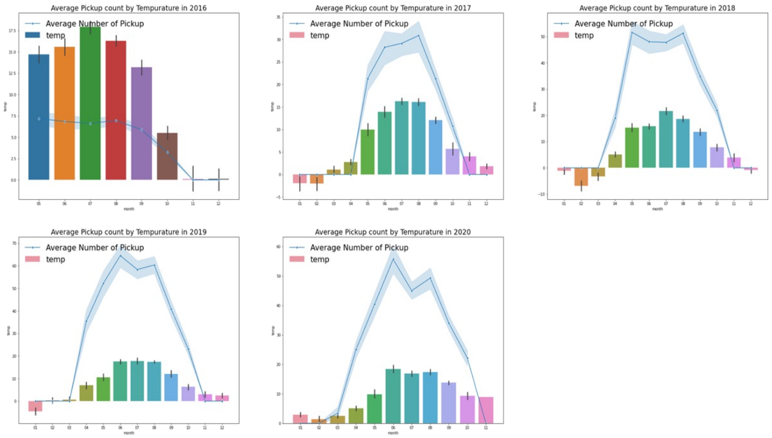


- Since May 2016
- 3,510 bikes in season 2020
- 241 bike stations in Helsinki and 110 stations in Espoo
- Dataset contains 12.138.008 entries with 14 attributes from 02.5.2016 to 01.11.2020
- Historical hourly weather data provided by Visual Crossing Weather.

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

Descriptive Analysis





- In 2016, the average pickup demand per year was max. 10.54, which is gradually increasing in 2017 is 43.17, in 2018 is 67.80, in 2019 is 86.99 and finally in 2020 is 73.74, slightly decreasing due to the pandemic Covid attacked by worldwide.
- The trip duration variability is most likely attributed to weather conditions (Temp, windspeed), as bike users tend to cycle more in warm conditions.

OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

Descriptive Analysis

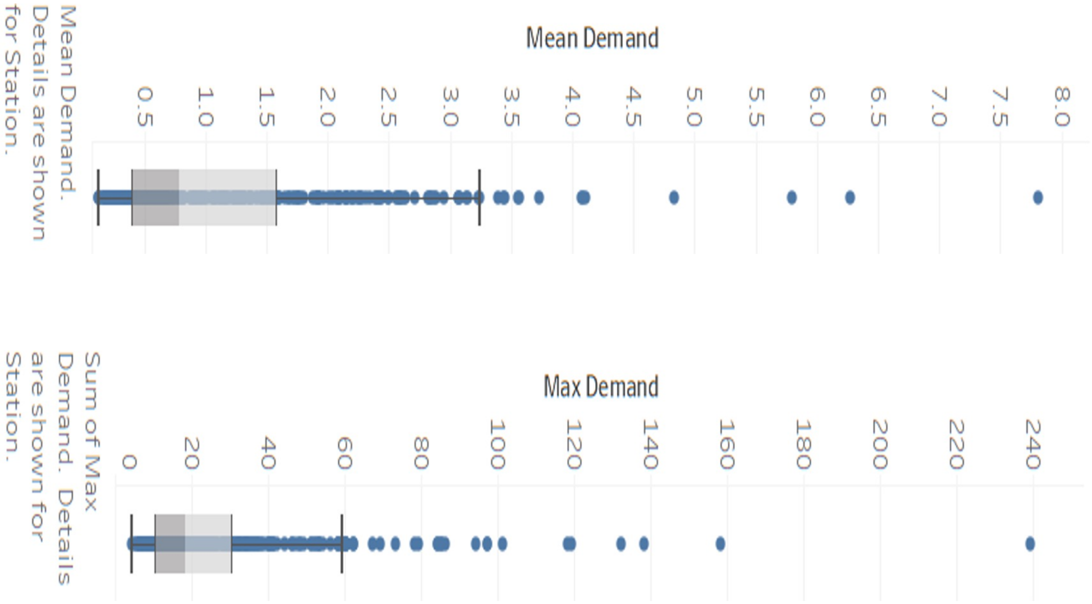
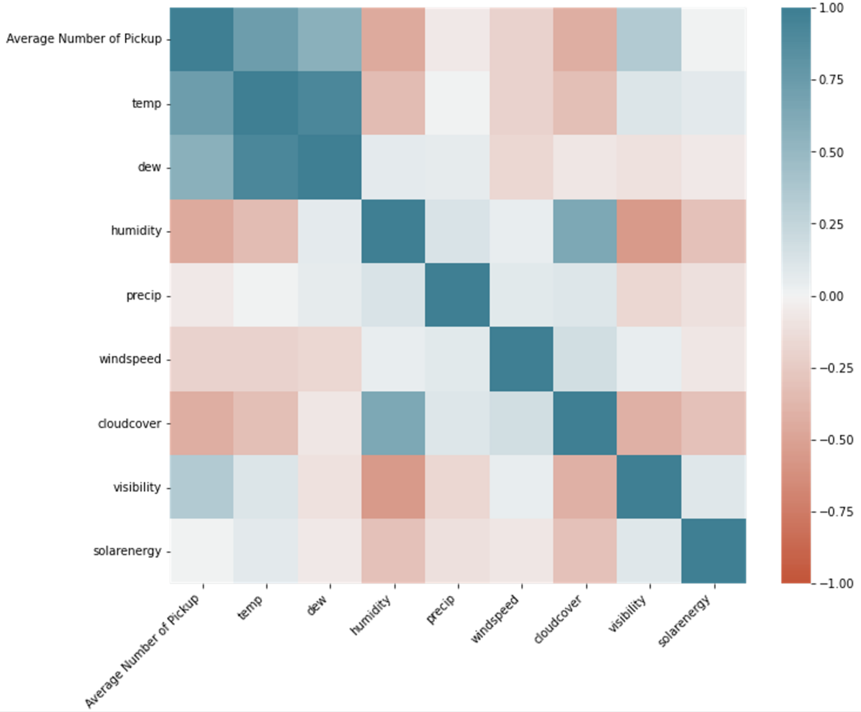


Figure. Distribution of Mean and Max Demand of 347 stations.

- In mean demand, the upper whisker, upper hinge, median, lower hinge, and lower whisker are respectively 3.234, 1.567, 0.772, 0.389, and 0.110. The interquartile range (IQR) is here 1.178. Station Itämerentori shows the max outlier range with pointing 7.794.
- The maximum and minimum demand noted in this analysis at station Kaivopuisto is 239, and Itäkeskus Metrovarikko is 4, respectively. The upper whisker, upper hinge, median, lower hinge, and lower whisker are respectively 59,30,18,10, and 4. The interquartile range (IQR) is 20. Station Kaivopuisto showed the highest outlier in this analysis.

Descriptive Analysis



- Average number of pickup is ideally has strong positive correlation with weather features.
- Strong positive correlations among features, 0 indicate no correlations and 1.0 indicates positive correlation and vice versa.

ARIMA

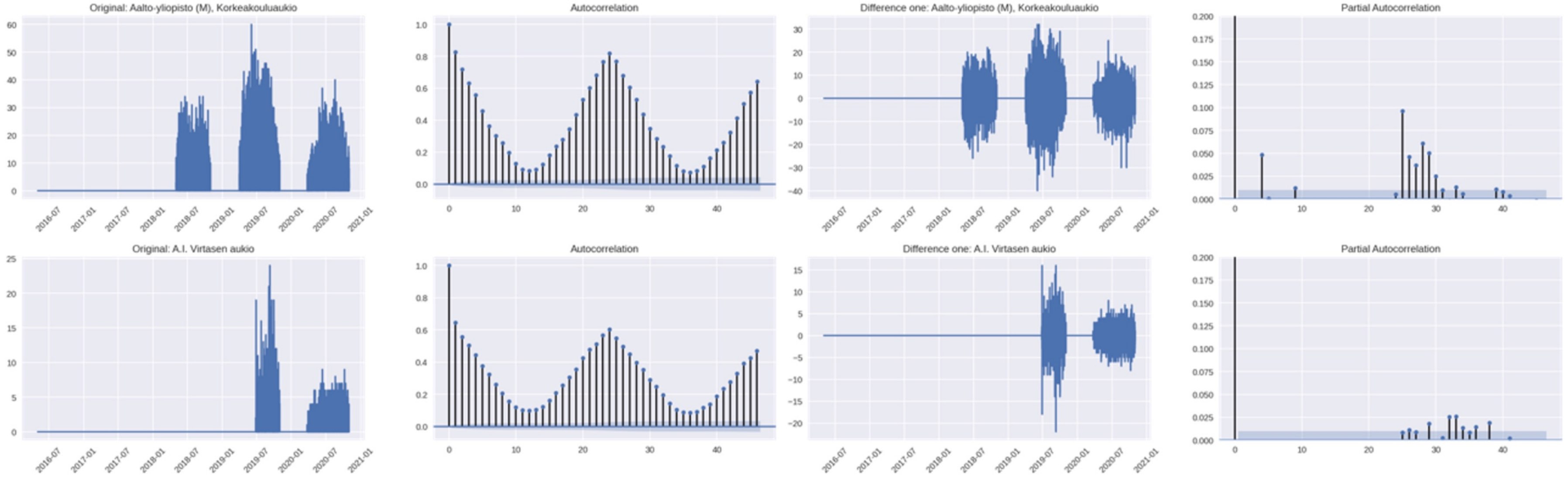
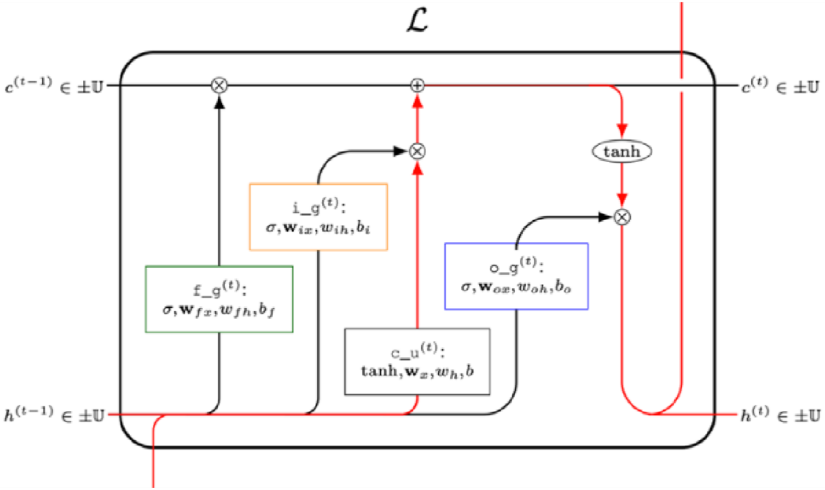


Figure. ACF & PACF of station different stations

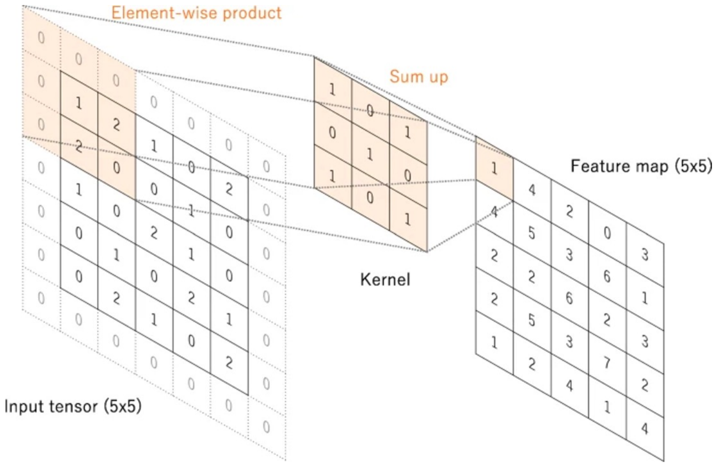
OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

Methodology: Long Short-Term Memory Recurrent Neural Networks (LSTM)



- LSTM is a particular recurrent neural network capable of remembering the values from earlier stages and forgetting unnecessary trends in each network cell (Ojo et al., 2019).
- LSTM can learn over 1000 discrete- time steps by allowing constant error carousels through special cells in layers (Hochreiter et al., 1997)
- The fundamental model consists of three layers: an input layer, a hidden layer, and an output layer

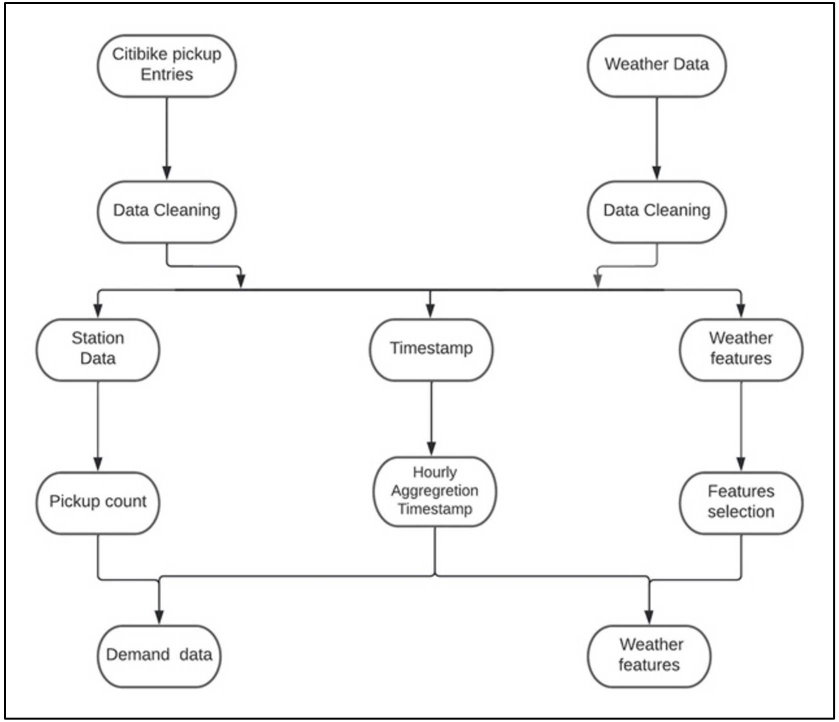
Methodology: Convolutional neural networks (CNN)



- CNN can discover meaningful features automatically without relationship instruction
- CNNs are modeled after the structures of human and animal brains, similar to a conventional neural network.
- CNN is far more data hungry because of its millions of learnable parameters to estimate, and, thus, is more computationally expensive

(Yamashita et al. (2017)

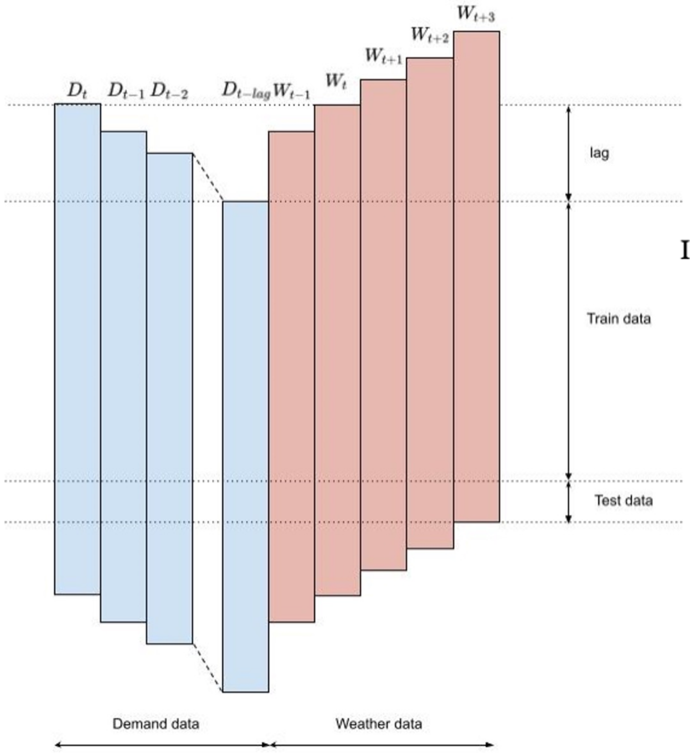
Data Preprocessing



$$\text{Demand data} = \begin{bmatrix} D_0^0 & D_0^1 & \cdots & D_0^c \\ D_1^0 & D_1^1 & \cdots & D_1^c \\ D_2^0 & D_2^1 & \cdots & D_2^c \\ \cdots & \cdots & \cdots & \cdots \\ D_{T-1}^0 & D_{T-1}^1 & \cdots & D_{T-1}^c \\ D_T^0 & D_T^1 & \cdots & D_T^c \end{bmatrix}$$

$$\text{Weather features} = \begin{bmatrix} W_0^0 & W_0^1 & \cdots & W_0^f \\ W_1^0 & W_1^1 & \cdots & W_1^f \\ W_2^0 & W_2^1 & \cdots & W_2^f \\ \cdots & \cdots & \cdots & \cdots \\ W_{T-1}^0 & W_{T-1}^1 & \cdots & W_{T-1}^f \\ W_T^0 & W_T^1 & \cdots & W_T^f \end{bmatrix}$$

Data Structure Design



$$\text{Input} = \begin{bmatrix} D^c_t & D^c_{t-1} & \dots & D^c_{t-lag} & W^1_{t-1} & \dots & W^f_{t-1} & W^1_t & \dots & W^f_t & \dots & W^1_{t+3} & \dots & W^f_{t+3} \\ D^c_{t+1} & D^c_t & \dots & D^c_{t-lag+1} & W^1_t & \dots & W^f_t & W^1_{t+1} & \dots & W^f_{t+1} & \dots & W^1_{t+4} & \dots & W^f_{t+4} \\ D^c_{t+2} & D^c_{t+1} & \dots & D^c_{t-lag+2} & W^1_{t+1} & \dots & W^f_{t+1} & W^1_{t+2} & \dots & W^f_{t+2} & \dots & W^1_{t+5} & \dots & W^f_{t+5} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ D^c_{T-1} & D^c_{T-2} & \dots & D^c_{T-lag-1} & W^1_{T-2} & \dots & W^f_{T-2} & W^1_{T-1} & \dots & W^f_{T-1} & \dots & W^1_{T+2} & \dots & W^f_{T+2} \\ D^c_T & D^c_{T-1} & \dots & D^c_{T-lag} & W^1_{T-1} & \dots & W^f_{T-1} & W^1_T & \dots & W^f_T & \dots & W^1_{T+3} & \dots & W^f_{T+3} \end{bmatrix}$$

$$\text{Output: } y^c = [D^c_{t+1} + D^c_{t+2},\quad D^c_{t+2} + D^c_{t+3},\quad \dots,\quad D^c_T + D_{T+1},\quad D^c_{T+1} + D^c_{T+2}]$$

The layered structure of LSTM and CNN-LSTM models

CNN

LSTM

Layer	Output shape	Param #
LSTM	(None, 1, 30)	48960
Dropout	(None, 1, 30)	0
LSTM	(None, 1, 30)	7320
Dropout	(None, 1, 30)	0
LSTM	(None, 30)	7320
Dropout	(None, 30)	0
Dense	(None, 1)	31

Total params: 63,631

Layer	Output shape	Param #
Conv1D	(None, 1, 30)	56580
MaxPooling1D	(None, 1, 30)	0
LSTM	(None, 1, 30)	7320
Dropout	(None, 1, 30)	0
LSTM	(None, 1, 30)	7320
Dropout	(None, 1, 30)	0
LSTM	(None, 30)	7320
Dropout	(None, 30)	0
Dense	(None, 1)	31

Total params: 78,571

Time Lags with LSTM and CNN-LSTM in Prediction Accuracy

Models	Average R-Squared	Average MAE	Average RMSE
LSTM 6	0.260	1.652	2.623
LSTM 12	0.281	1.638	2.561
LSTM 24	0.369	1.448	2.344
LSTM 48	0.381	1.429	2.328
LSTM 96	0.391	1.421	2.315
LSTM 120	0.395	1.412	2.303
LSTM 168	0.421	1.352	2.237
LSTM 336	0.425	1.348	2.230
LSTM 504	0.427	1.360	2.243
LSTM 672	0.416	1.379	2.281
CNN-LSTM 6	0.281	1.617	2.568
CNN-LSTM 12	0.272	1.632	2.559
CNN-LSTM 24	0.372	1.448	2.342
CNN-LSTM 48	0.388	1.419	2.306
CNN-LSTM 96	0.392	1.399	2.283
CNN-LSTM 120	0.401	1.390	2.273
CNN-LSTM 168	0.429	1.323	2.186
CNN-LSTM 336	0.429	1.322	2.182
CNN-LSTM 504	0.416	1.335	2.204
CNN-LSTM 672	0.411	1.339	2.220

Table. Prediction model testing results for 200 stations

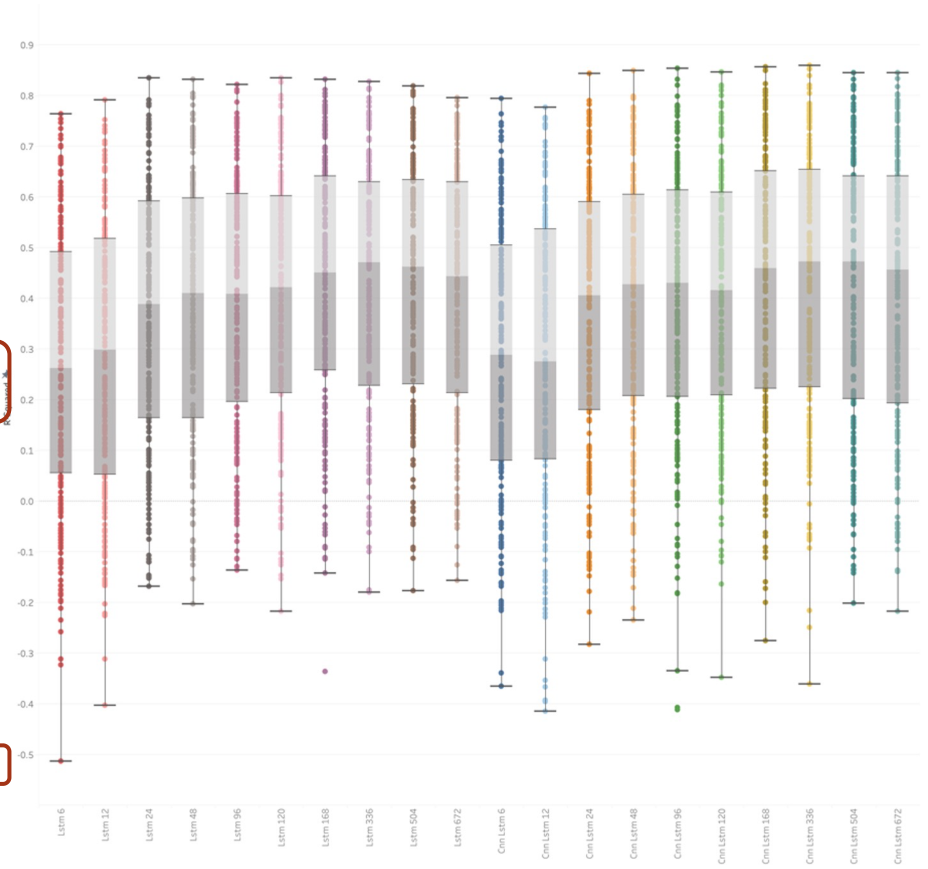


Figure. Boxplot distribution of R-Squared results for 200 stations

Average Pickups Demand in Prediction Accuracy

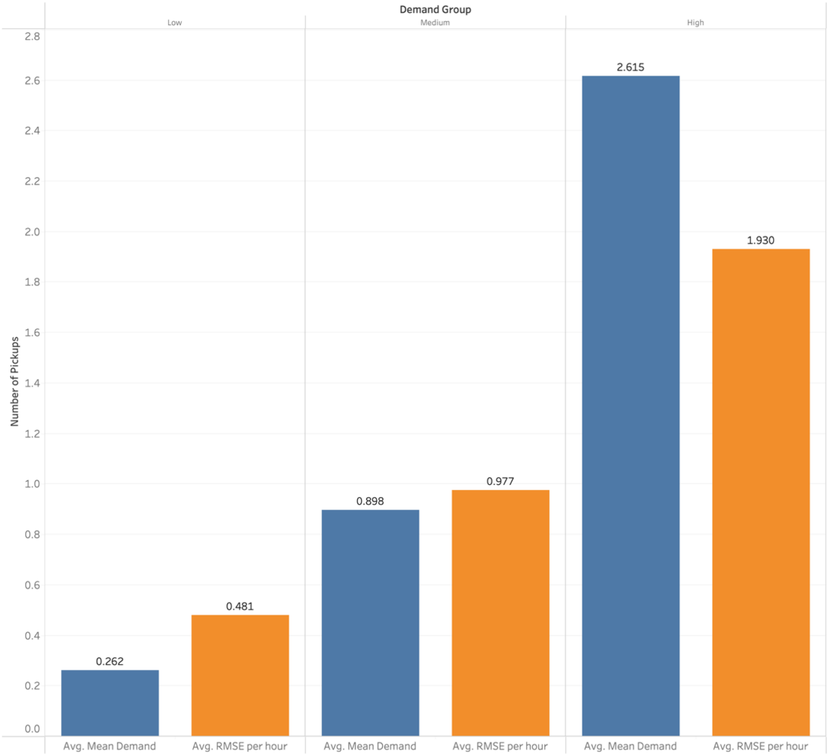


Figure. Comparison between Average of mean demand with RMSE per hour of CNN-LSTM 336 by demand groups

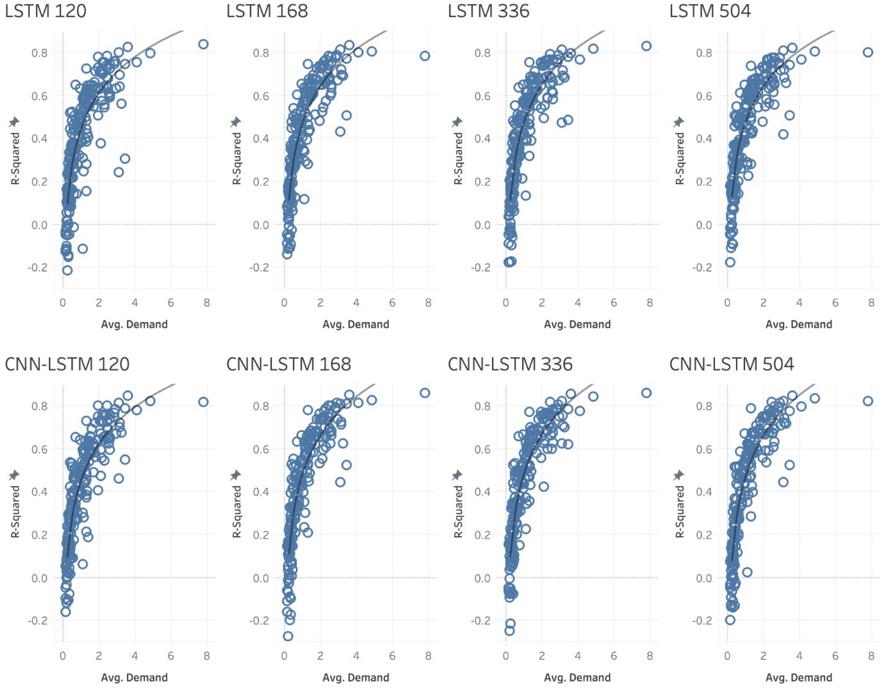


Figure. Logarithmic relation of R-squared and Mean demand at each station

Weather Forecasting in Prediction Accuracy

Model	R-Squared	MAE	RMSE
CNN-LSTM 336_2	0.417	1.334	2.205
CNN-LSTM 336_3	0.429	1.322	2.182
CNN-LSTM 336_4	0.435	1.320	2.182
CNN-LSTM 336_5	0.432	1.321	2.187
LSTM 336_2	0.434	1.347	2.218
LSTM 336_3	0.425	1.348	2.230
LSTM 336_4	0.428	1.350	2.229
LSTM 336_5	0.420	1.355	2.239

Table. Prediction model testing results with different weather forecasting interval

- The period interval considering for pickup demand is important, and 336 hours is the optimal lags for prediction models
- The user decides to use the bike by the weather forecast for the outgoing and return routes.
- Users look at weather forecasts for their trips for the next 4 hours

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

FAKULTÄT FÜR WIRTSCHAFTSWISSENSCHAFT

Questions and Discussion