Predicting NJIT's GDS capacity using Long Short-Term Memory (LSTM)

CS 301 Final Project

Nikhil Ramesh
Lazar Agoev
Sergio Torrico

Problem

Currently, there is no way to tell how many students are using NJIT's GDS and therefore it would be inconvenient/unsafe for students to leave without an estimate of how busy continuous dining is.

We decided to work in the implementation of an algorithm to predict number of people accessing the facility as well as number of meal swipes used during four specified time periods throughout the day.

Related Works

Upon research, we decided to use the article “5 Machine Learning Techniques for Sales Forecasting”. This article, discusses the following techniques which could all be applied in some way to solve our problem statement:

- Linear Regression
- Random Forest Regression
- XGBoost
- LSTMs (Long short-term memory)
- ARIMA (Autoregressive Integrated Moving Average)

Your Approach and Results

Upon discussing between Autoregressive Integrated Moving Average and Long Short-term Memory. We decided to go with LSTM.

Benefits of LSTM:

- Model can predict future values based on previous sequential data.
- LSTM can efficiently handle non stationary sequences.
- Even though LSTM takes longer to train, the results escalate the larger the set gets and the results are very accurate.

Technical Challenges

- Originally we thought about using existing datasets from Kaggle but it would not be as accurate as the real data.
- We thought about taking turns sitting outside GDS and manually counting each student that entered/exited GDS.
- We contacted the NJIT cafeteria through email to request access to data that shows the ID swipes via David Arluna, head of Gourmet Dining
- The dataset is not large enough to account for various outliers such as holidays, breaks, or special campus events influencing the amount of students using GDS for that specific period.

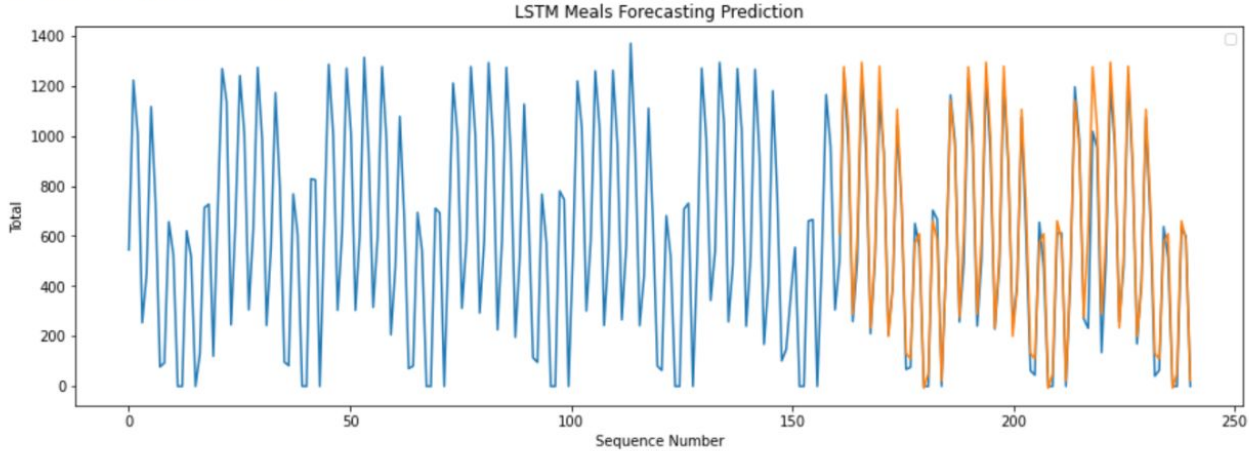
LSTM Implementation

- Import corresponding libraries
- Preparing the data:
  - Upload
  - Drop unnecessary columns
  - Split the data into subsets to train and to test
- Develop the helper functions to train, and plot the data using:
  - Keras
  - Matplotlib
- Develop function to create the model
- Perform corresponding predictions.
- Plot the results for comparison.

Results

Meals Forecasting Prediction
Blue-actual Data, Orange-prediction

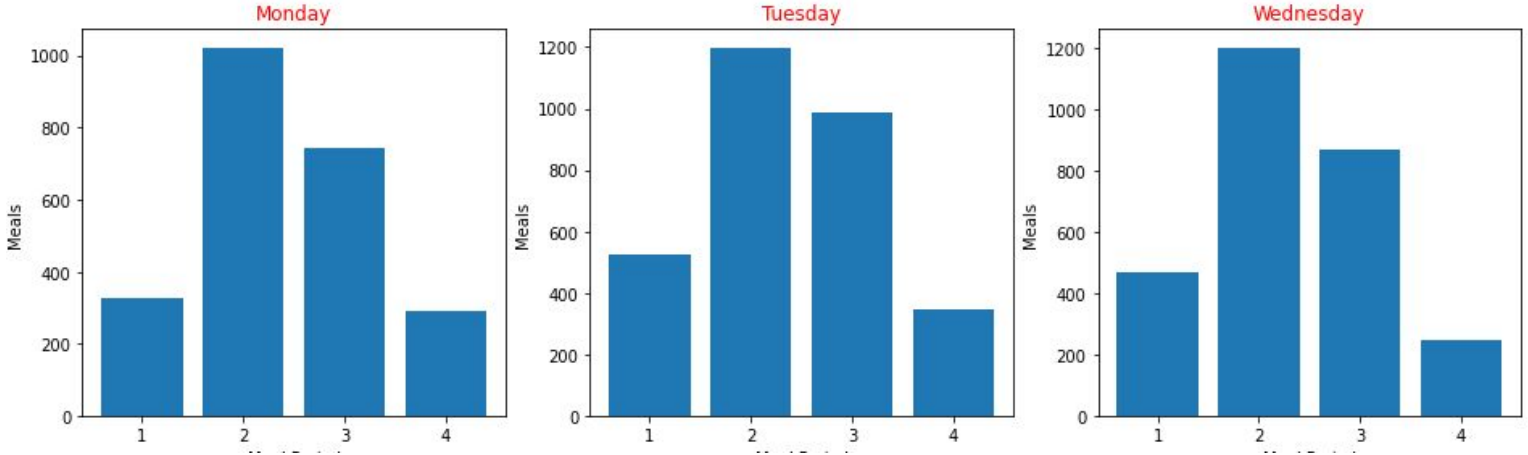
RMSE: 78.28115659055426
MAE: 56.19960941672325
R2 Score: 0.9616323526913827

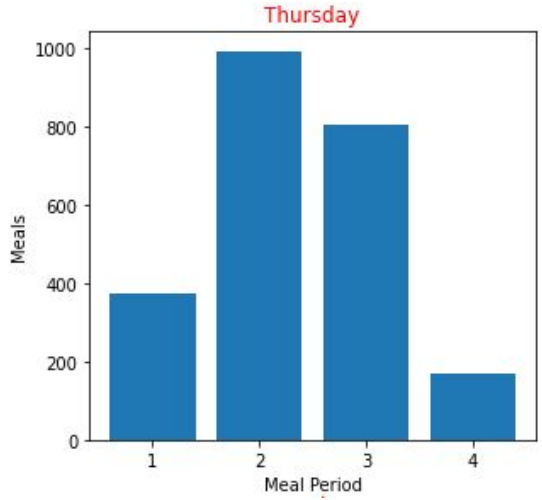


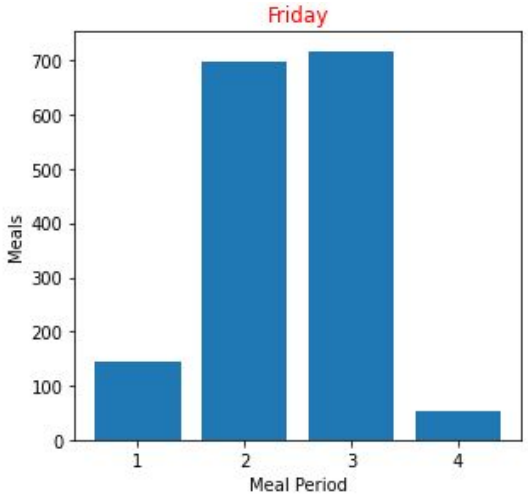
Week predictions

Each index represents the following periods

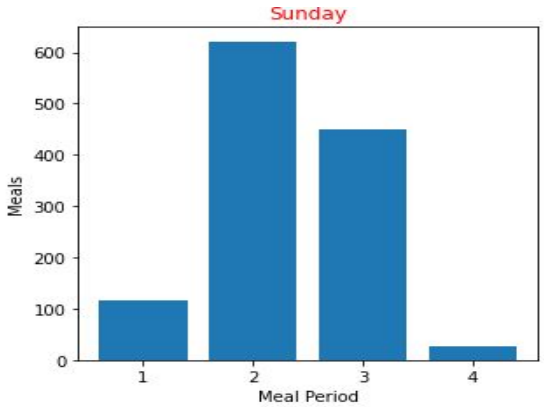
7am-10:30am, 10:30am-3:30pm, 3:30pm-8:00pm, 8:00pm-10:00pm




Thursday
Meals
Meal Period
0
200
400
600
800
1000
1
2
3
4


Friday
Meals
Meal Period
0
100
200
300
400
500
600
700
1
2
3
4


Saturday
Meals
Meal Period
0
100
200
300
400
500
600
1
2
3
4


Sunday
Meals
Meal Period
0
100
200
300
400
500
600
1
2
3
4

Broader Impact

- Many of the students who dorm at NJIT use GDS as their main food source
  - Wait time and capacity are huge concerns for on campus students
- Covid-19 and social distancing is more relevant than ever due to new variants and potential new waves
- Currently there is no algorithm/monitoring service for NJIT GDS
- Algorithm could be presented to NJIT to be integrated to the NJIT Go App as a free service for students

Limitation and future improvements

- Access a larger GDS dataset of at least 3 years.
- Receiving more precise data in order to create more precise predictions.
- Compare our predicted data with the corresponding weeks in the future.

Resources

https://towardsdatascience.com/5-machine-learning-techniques-for-sales-forecasting-598e4984b109

Pandas - data manipulation and analysis

Numpy - multidimensional array manipulation/storage

Keras - machine learning implementation

Sklearn - statistical analysis (msqe and R2)

Matplotlib - graphical/plot of data

Real Meal Period Data courtesy of David Arluna, head of Gourmet Dining