Real Time Taxi Prediction

Sourav Dey | Sweta Kumari | Arushi Das | Prof. Krishnamoorthy A | SCOPE

Introduction

One of the most crucial questions for all taxi drivers is where to find a passenger. People are forced to wait for extended periods of time in risky conditions, lowering the taxi service's overall satisfaction rating, drivers are unsure of where to seek for the next fare after dropping off a passenger, and taxi drivers are hesitant to travel to a rather remote place for fear of not finding any customers and wasting time and fuel. If the demand for cabs can be foreseen, such problems can be avoided.

Motivation

Many multibillion-dollar businesses have risen from the provision of such services to consumers via Internet apps, portals, or local services. An imbalance in the distribution of taxi drivers in the city is a major issue, particularly in well-established metropolitan regions where need for taxis is strong but supply is few.

Scope of the Project

The objective of our project is to predict the number of pickups in a particular location at an equal time gap. Few locations require large number of taxis to suffice the demand in that particular area, while some may need less taxis in that area, so to have a better understanding of this problem we have made certain algorithms over a dataset so that we can predict where the requirements are needed and fulfilled accordingly.

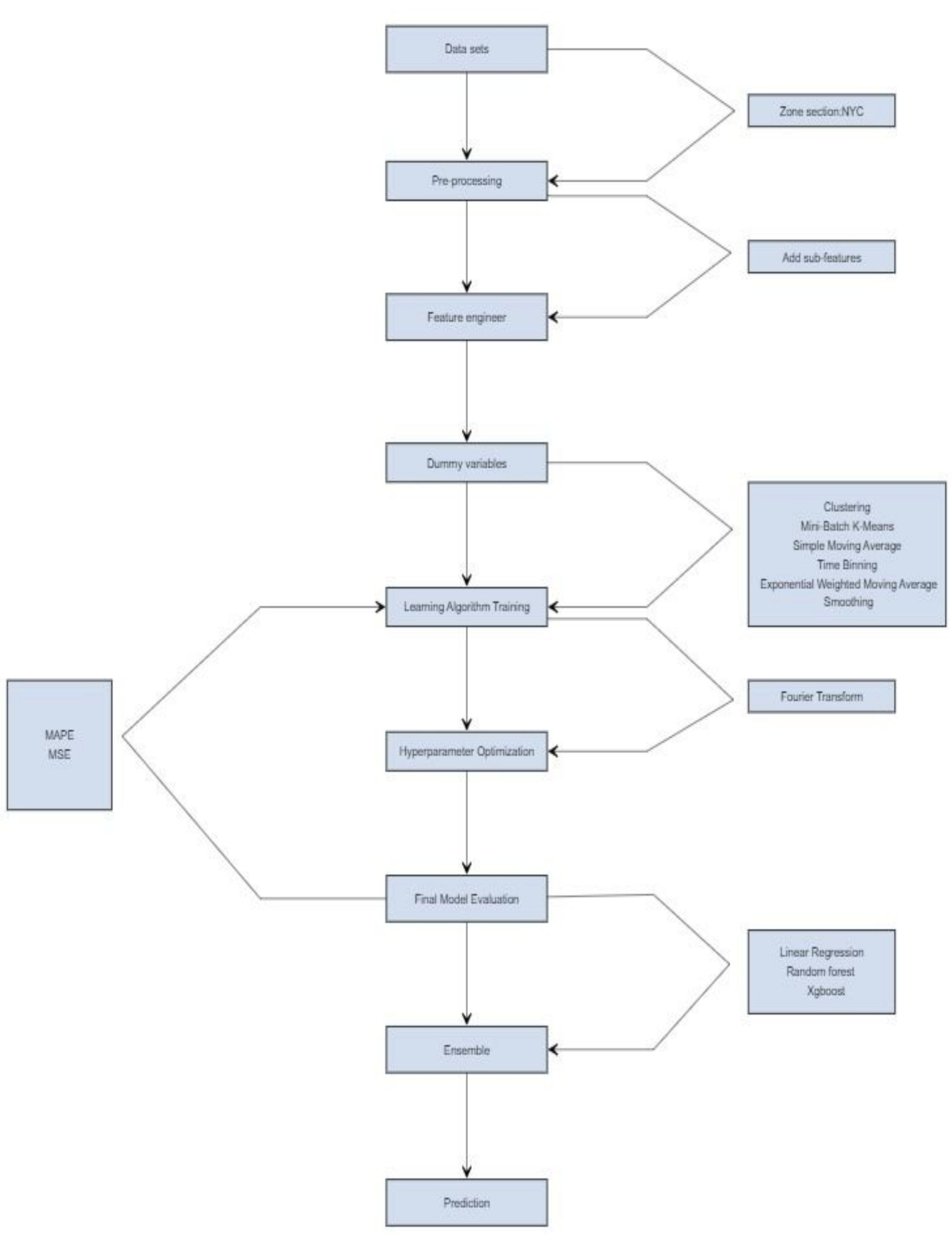
Methodology

Workflow Diagram

The framework for the entire process is depicted in Figure 1. To begin, we divided the dataset into two parts: training and testing. Two effective classification algorithms based on the training set are Clustering, Mini-Batch K-Means, Simple Moving Average, Time Binning, Exponential Weighted Moving Average, Smoothing, Regression Models, Fourier Transform, Random Forest, XGBoost for Regression was implemented to develop the detection models. Finally, we categorised the new and unknown test data using the generated Machine Learning models and chose the model with the highest precision.

Preprocessing

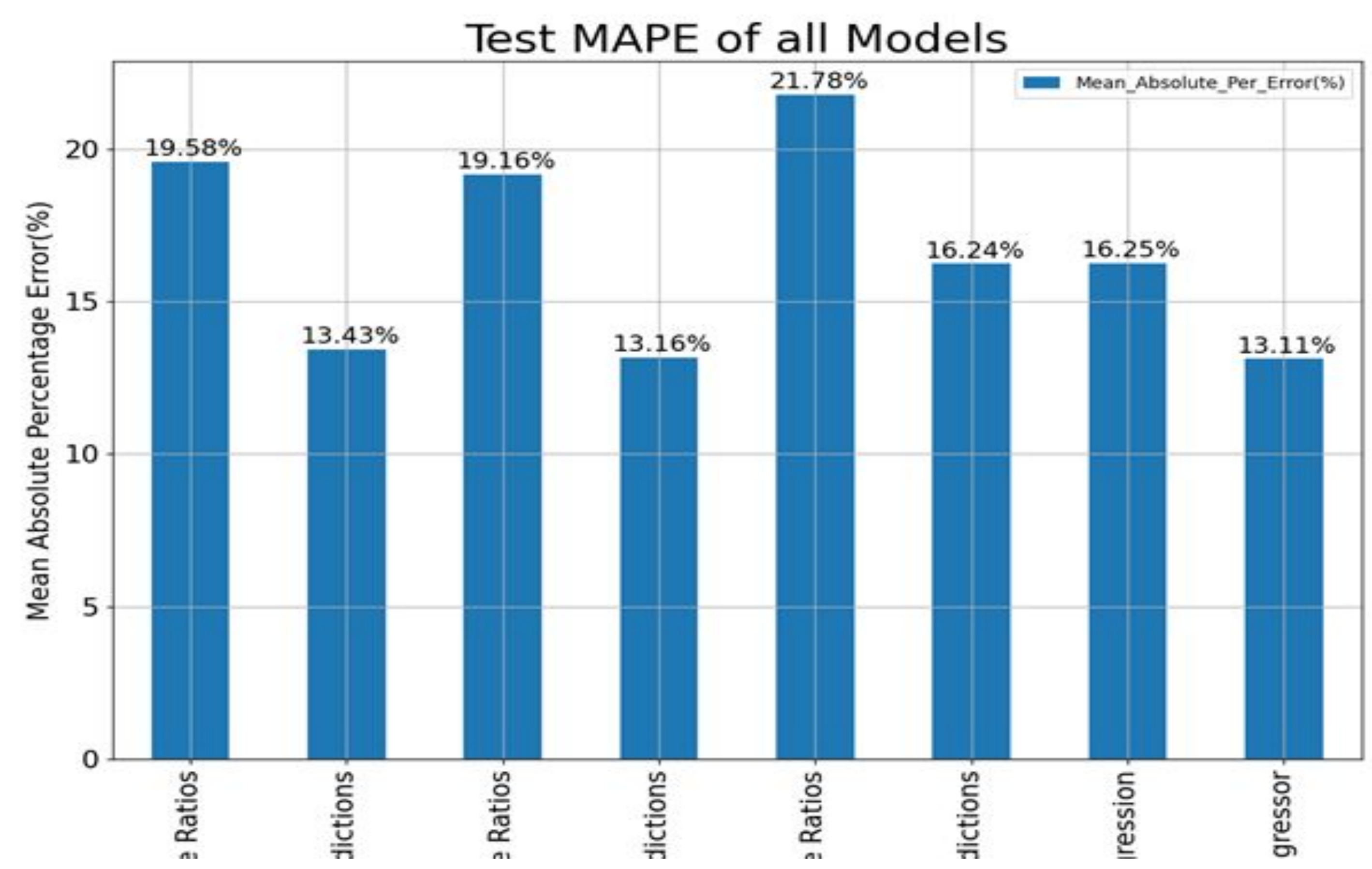
- *Data preprocessing*: The machine learning stage in which information is processed, transformed or encoded into formats which can be efficiently scanned by a machine[5]. The pre-processing was done as follows:
- *Feature Encoding*: The process of continual data modifications so that it may be utilized as an algorithm input while keeping its original significance.
- *Train / Validation / Test Split:* It is advisable, before deciding on the algorithm, to split the datasets into two or three pieces. Machine Learning methods, like any other algorithm, must first be trained on correctly dispersed data, then reviewed and verified before being applied to real-world data.



Results

The data set contains dependent or target variables along with independent variables. We build models using independent variables and predict dependent or target variables. If the dependent variable is numeric, regression models are used to predict it. MSE is used to evaluate the models.

We also calculated the Mean Absolute per Error (MAPE), the sum of the individual absolute errors divided by the demand (each period separately). It is the average of the percentage errors, despite being a poor-accuracy indicator. As you can see in the formula, MAPE divides each error individually by the demand, so it is skewed: high errors during low-demand periods will significantly impact MAPE. Due to this, optimizing MAPE will result in a strange forecast that will most likely undershoot the demand.



Model	Mean_Absolute_Per_Error(%)
Simple Moving Average Ratios	19.123931
Simple Moving Average Predictions	13.220558
Weighted Moving Average Ratios	18.739608
Weighted Moving Average Predictions	12.957807
Exponential Weighted Moving Average Ratios	21.245534
Exponential Weighted Moving Average Predictions	16.007452
Linear Regression	18.381391
XGBoost Regressor	13.232612

Conclusion

By implementing our model, we will be able to get insights into the data to help make effective taxi demand forecasting decisions. This study uses Clustering, XGBoast, Random Forest, Linear Regression, Simple Moving Average, Linear Regression, Exponential Weighted Moving Average, and Weighted Moving Average to evaluate taxes and regression models and compare them to one other.

The weighted motion average technique is the smallest loss with a MAPE of 12,957807. This method has a greater accuracy than other regression-based algorithms. Future research can include a comparison of other more modern, yet less sophisticated, algorithms.

References

[1] Zhao, K., Khryashchev, D., Freire, J., Silva, C., & Vo, H. (2016, December). Predicting taxi demand at high spatial resolution: Approaching the limit of predictability. In 2016 ieee international conference on Big data (big data) (pp. 833-842). IEEE.

[2] Seow, K. T., Dang, N. H., & Lee, D. H. (2009). A collaborative multiagent taxi-dispatch system. IEEE Transactions on Automation Science and Engineering, 7(3), 607-616.

[3] Davis, N., Raina, G., & Jagannathan, K. (2016, November). A multi-level clustering approach for forecasting taxi travel demand. In 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC) (pp. 223-228). IEEE.

[4] Lee, D. H., Wang, H., Cheu, R. L., & Teo, S. H. (2004). Taxi dispatch system based on current demands and real-time traffic conditions. Transportation Research Record, 1882(1), 193-200.

[5] Gers, F. A., Eck, D., & Schmidhuber, J. (2002). Applying LSTM to time series predictable through time-window approaches. In Neural Nets WIRN Vietri-01 (pp. 193-200). Springer, London.