Augusto Rodrigues da Costa Bedin, Lionel Rodin Rajaona, and Madhavee Kaushik

Group 20

Augusto Rodrigues da Costa Bedin, Lionel Rodin Rajaona, and Madhavee Kaushik

[[1]](#footnote-1)

Progress Report

*Abstract*—Machine learning has been used in various datasets to detect patterns, predict the future with said patterns and aid decision-making with those algorithms. This paper discusses the role of an Additive Model to provide an effective model. Here an energy consumption dataset from PJM Evolution is used on an attempt to forecast energy consumption in the following years based on extensive record data of hourly energy consumption through the period of 1998 to 2018. A study about how the old data have a pattern and its working has been made and the similar methods are used to make the same pattern of power consumption to predict result for future using machine learning. This specific case cannot be resolved by algorithms such as linear or logistic regressions, as we're dealing with time series type of data. In this case, models such as auto-regressive or neural networks are viable, valid options.

# INTRODUCTION

T

HIS current project covers the analysis of the contents of a dataset which includes the hourly consumption of power in some areas of USA from Jan 2002 to Jan 2018. As with the increasing data storing procedure, it is important to use this information in order to make the system ready to follow a regular pattern every time. As for something unexpected, it can never be analyzed or recorded in a system to be used later.

The data is taken from the PJM Interconnection LLC which is a regional transmission organization in the United States of America over the last 10 years. The data recorded is on hourly basis and is calculated in Megawatts. This PJM Interconnection LLC covers a large area of USA which includes Eastern Interconnection grid, an electric transmission system.

It consists of one of the most unstructured data which is seen as only some entries of power on daily basis. The goal of Machine learning is to make this data useful in order to get the desired result for predictions just by using the older datasets.

First, the static database has been used. After its successful running, trial for dynamic database has been put out. In the beginning, small quantity of datasets was used and then will take more datasets as the coding improves.

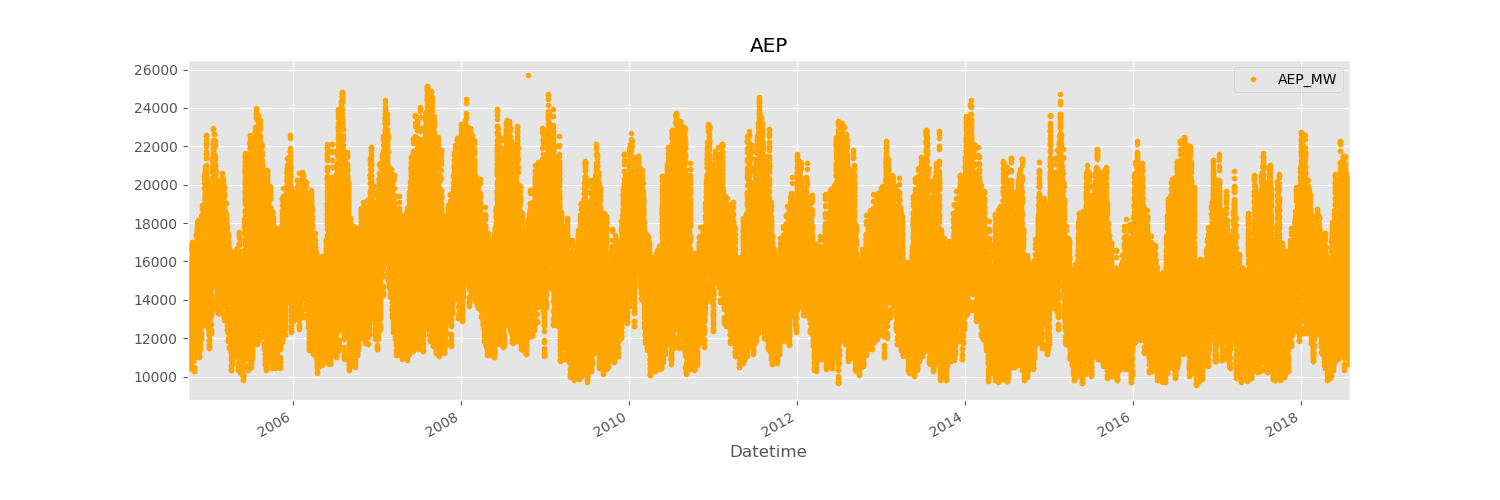
As the result obtained was deterministic, it was easier to cross check the result and find bugs. This improved project has been made by implementing various logical methods.

Fig 1

Hourly Power Consumption Data From PJM

# Algorithm used

In the beginning, the facebook prophet, which is a open source library based on decomposable models, is used to help making the predictions more accurate with increased accuracy and a lot of customizable tools. It helps better in forecasting by a thorough understanding about the model’s working. It uses three main components which are trend, seasonality and holidays. All these changes are combined together along with a error for unusual change which cannot be inserted into the model. Using this algorithm, hourly time series forecasting is done with the present hourly power consumption data.

As the power consumption have unique characteristics, the output would be a little bit interesting. To see some changes in the trend following, features such as hour of day, time of year, week’s day etc are added.

For making a test and a training set to work on, the dataset have been divided in two. Training set is the one before Jan,01 2015 and test set is the one after Jan 01, 2015. This splitting can be seen on the plotted graph of the dataset.

After that the training set is plotted along with its prediction. This forecasting can actually be checked using the actual test set made from the data sets after the date Jan 01, 2015.

In the end, error metrics are made to compute the data values from the test set.

# Results

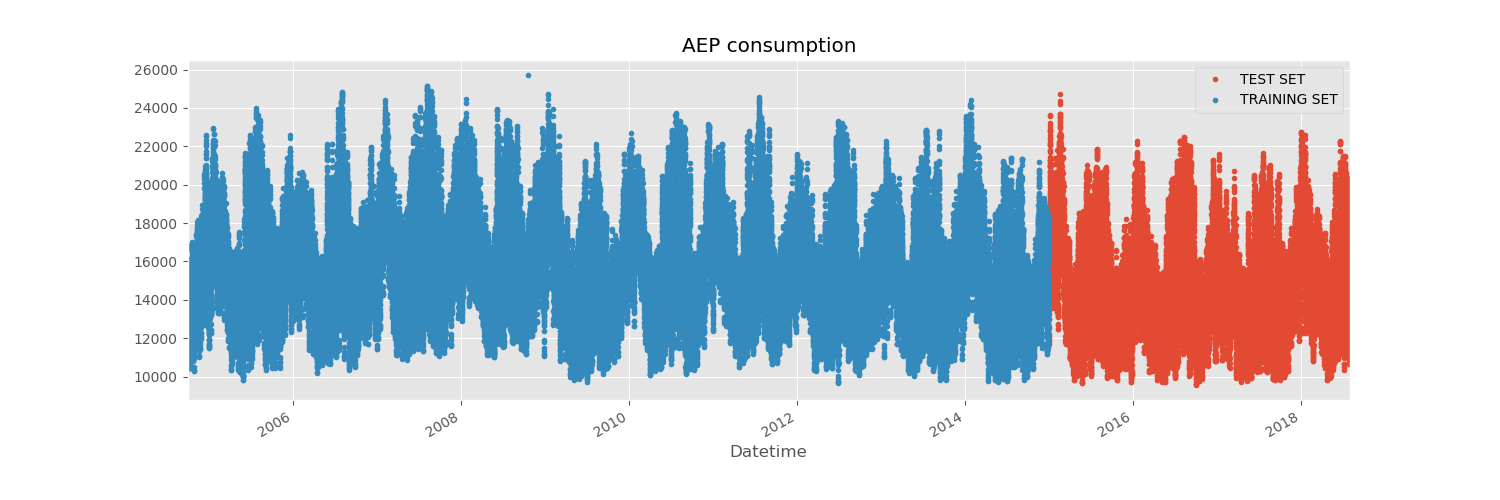


Fig 2

Splitting Training set and Test set

Blue = Training, Red = Test

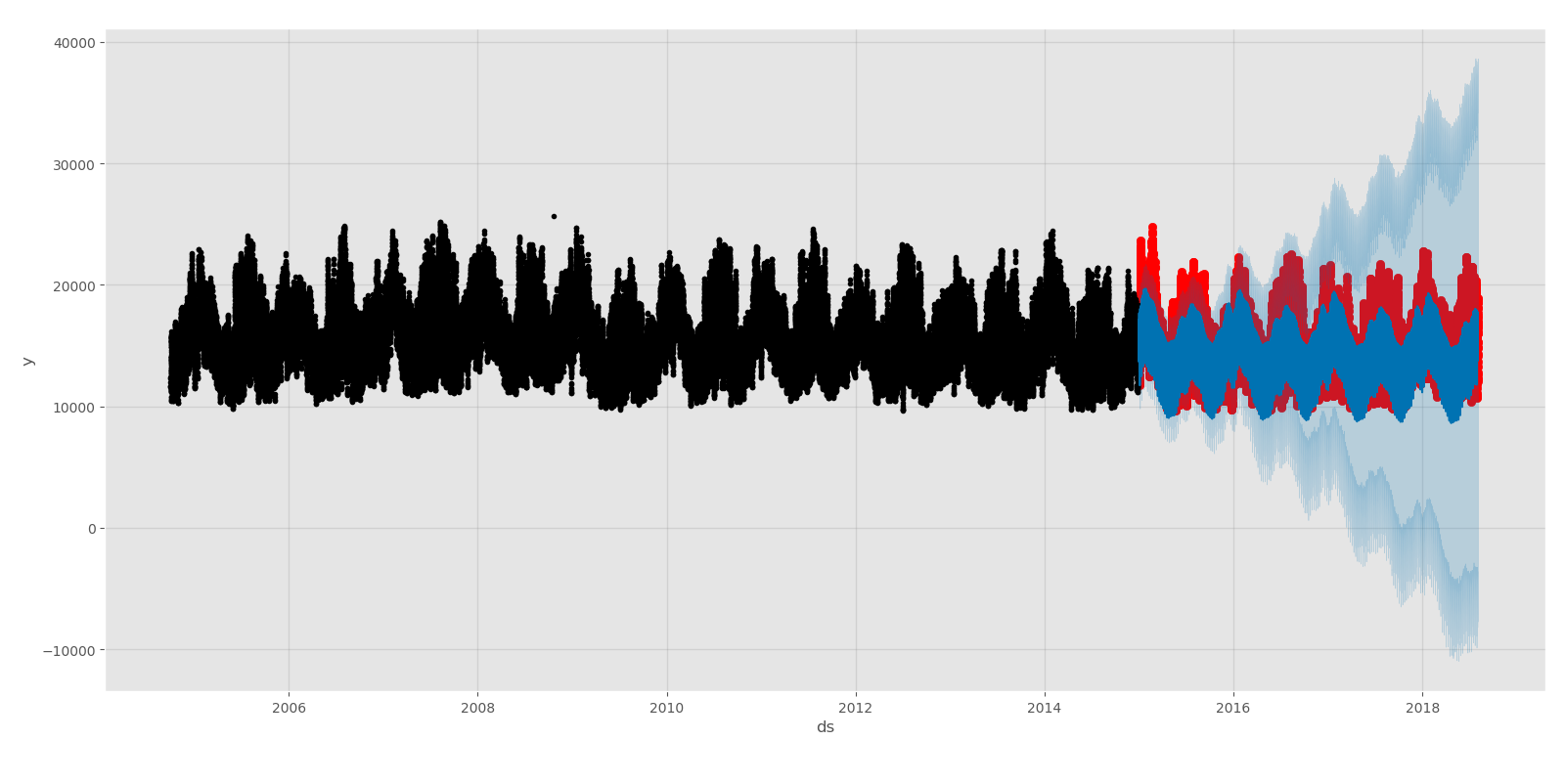


Fig 3

AEP prediction model for Test Set

Blue = Predicted, Red = Actual

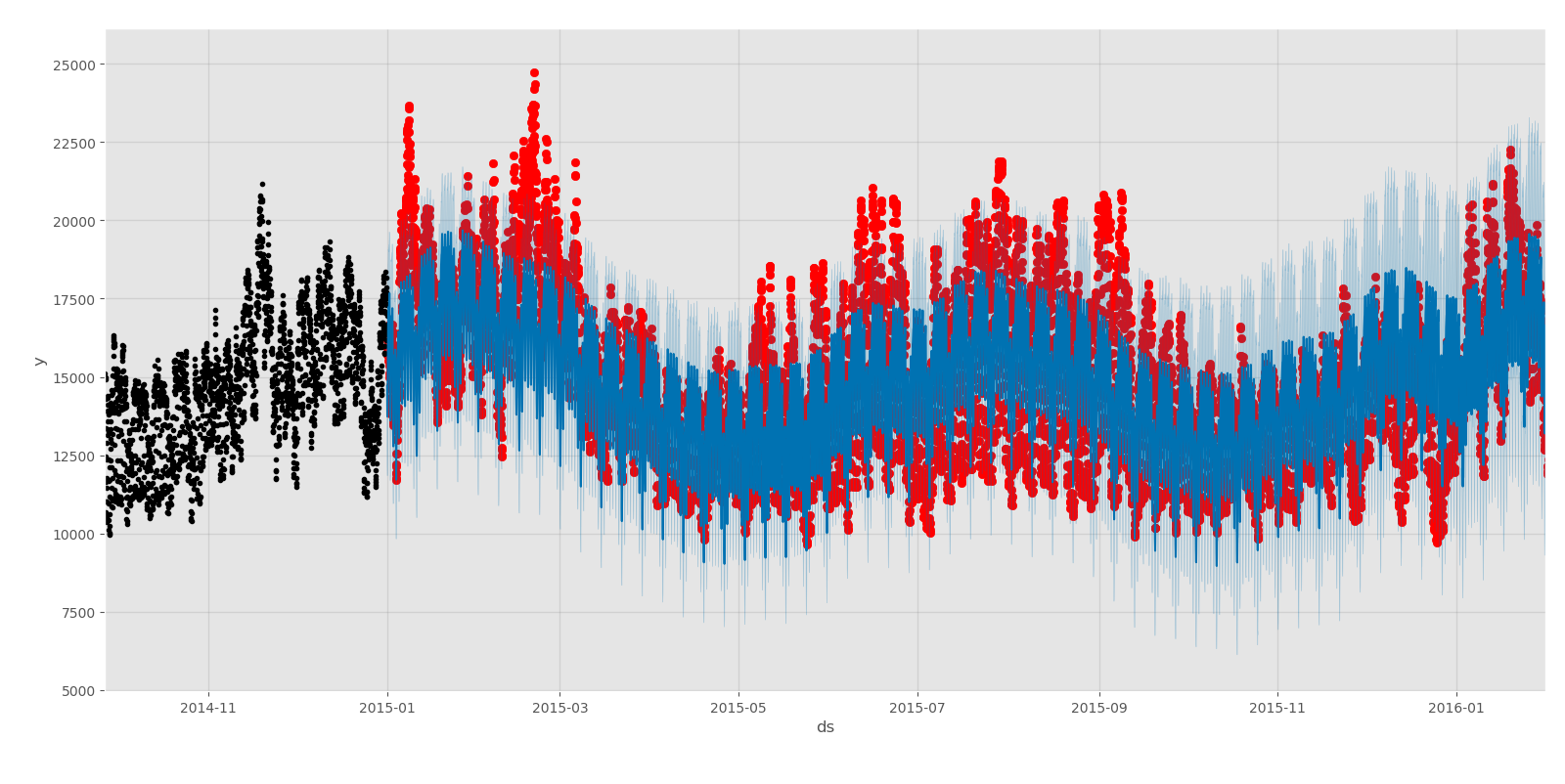


Fig 4

AEP prediction model on a monthly basis

Blue = Predicted, Red = Actual

# https://lh5.googleusercontent.com/3333LO1LLlkxT7eYGhy6GeX01fjnXMJMZnghwU8NZThQm2Qa30ujnem9Blevxy7OLSjeyGKNlGFn1l1T2BfbjoAtL2SgyAc08Xy4Y5-uOdQNOplyQknDgJPP3_AW8oXzEuzA0knR

Fig 5

AEP prediction model for on a weekly basis

Blue = Predicted, Red =Actual

These are metrics concerning the validation set.

The following are the measured errors:

Mean Squared error is 6787383.499449232 MW²

Root-mean squared error is 2605 MW

Mean Absolute error is 2082.0541852746464 MW

Mean Absolute percentage error is 14.18%

Accuracy is 85.82%

Given the following results we can argue on a good model to start on. Indeed, accuracy is not perfect, but we are certain that there is no overfitting issue, which is not what we are looking for. Thanks to this algorithm, we can continue by forecasting values from 2018 to 2019.

# Future Work

For future work, holidays and their predictions can be added to the model. The changes can be seen by plotting the holidays along with their predictions in the model and comparing the training model with the test model. Using all these features, the model will be effective to give predictions about the data after Jan 2018.

# Contributions

## All the work was divided equally between all member. Meetings were held in order to distribute the labor and after completion of one member’s task, the datasets and operations were passed to the other member. Members were responsible for providing their views, concerns and solutions on each step of the way. Results are discussed, pondered upon and implemented when everyone agrees and see fit.

# Conclusion

The model which was expected has been made to some extent and has given satisfactorily results. By adding more features, the prediction model became more understandable and consistent to work upon.

References

[1] Dataset

<https://www.kaggle.com/robikscube/hourly-energy-consumption?fbclid=IwAR1CKf-8alwBUfwajdxHjysGF6ogd9sAttYZYlKdu0hpp1e92sikDWdQnjA>

1. [↑](#footnote-ref-1)