Wind Power Forecasting: Jan-March 2020

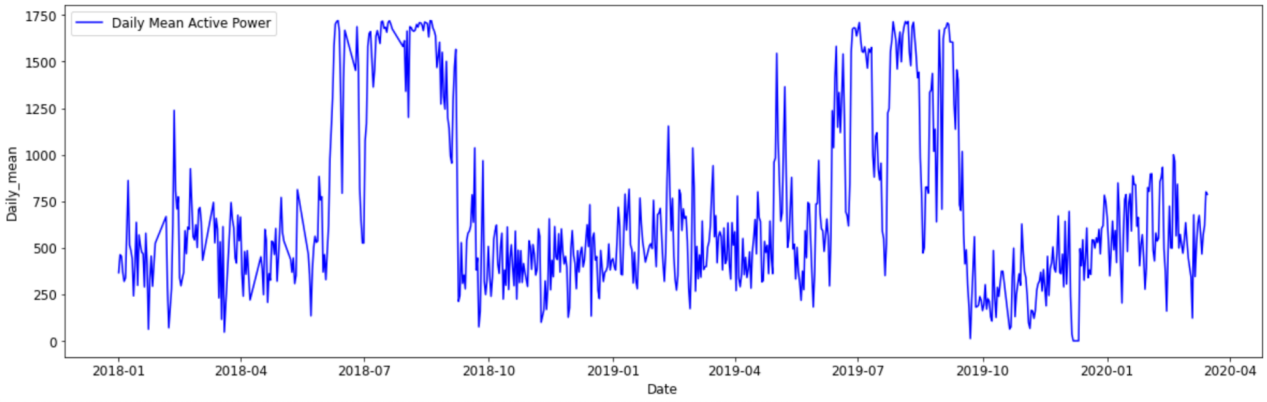
Final Project
MSDS 604
Group 4

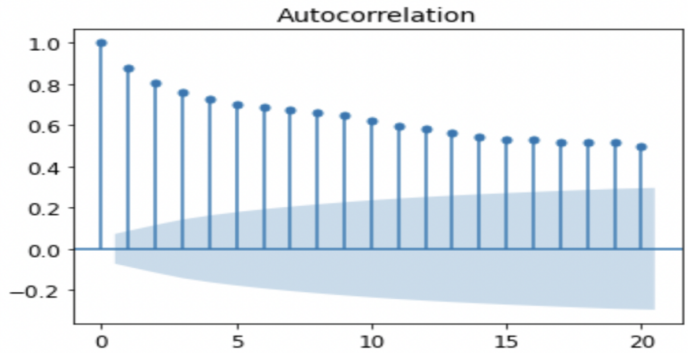
Sanchita Jain
Daniel Tinoco
Camilo Chaves

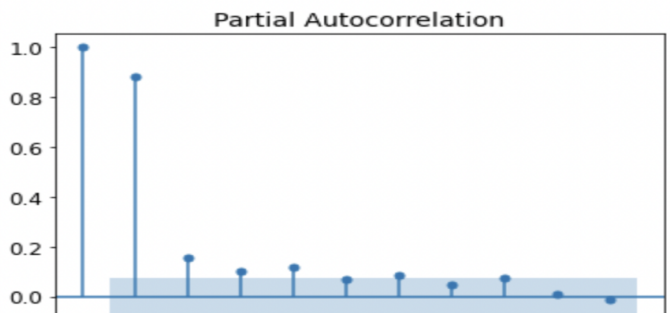
Data and Methods Description:

The goal of this project is to construct a time series analysis and model to give a fifteen day forecast of the mean daily wind power generated by a particular windmill. Furthermore, we will compare our predictions with the actual mean daily power readings recorded from March 16, 2020 to March 30, 2020. Over 77,000 temperature, turbine, rotor and other measurements were taken from the windmill from 2018 to 2020.

To perform our analysis, we grouped records that were taken on the same given day and calculated the mean power for the day. After, we splitted our dataset into all the records with dates prior to March 1, 2020 and the fifteen remaining records. Then, we did some preprocessing and exploration (e.g checking there were no null or negative values) to make sure there were no obviously faulty recordings. Finally, we plotted a time series plot for the daily mean power generated by the windmill, along with its respective ACF and PACF plots:







To check if our data is stationary, we performed an ADF test on our training data:

```
Results of Augmented Dickey-Fuller Test:
Test Statistic   -3.237316
p-value           0.017924
dtype: float64
```

Since the p-value is less than alpha so the TS is stationary

We will now use ETS, ARIMA, SARIMA, Prophet ensemble forecasting methods to model our data. We will pick the best model amongst each forecasting strategy, and then we will compare their RMSEs to choose our final model and find how it performs against the mean daily wind power recordings on the test data.

1. For each model, the column(s) you have used from your data?

   ANS - The date and active power.

2. Did you handle missing values, if yes, how?

   ANS - There were some dates missing, but we ultimately excluded them due to having little effect on the forecast.

3. Did you aggregate the data, if yes, how (since the data was based on 10-minute intervals, you could choose to aggregate to be daily average, etc.)

   ANS - We chose to aggregate by daily average, since the goal itself is to forecast the daily average for 15 days.

4. The validation method you have used to select the model (cross-validation? How many folds? And etc.)

ANS - We chose cross-validation with an increasing train size and rolling window of 15. For ARIMA, we used 10 folds, and 5 for SARIMA and ETS.

5. If you have included features in your model, how did you provide the feature values for the forecast?

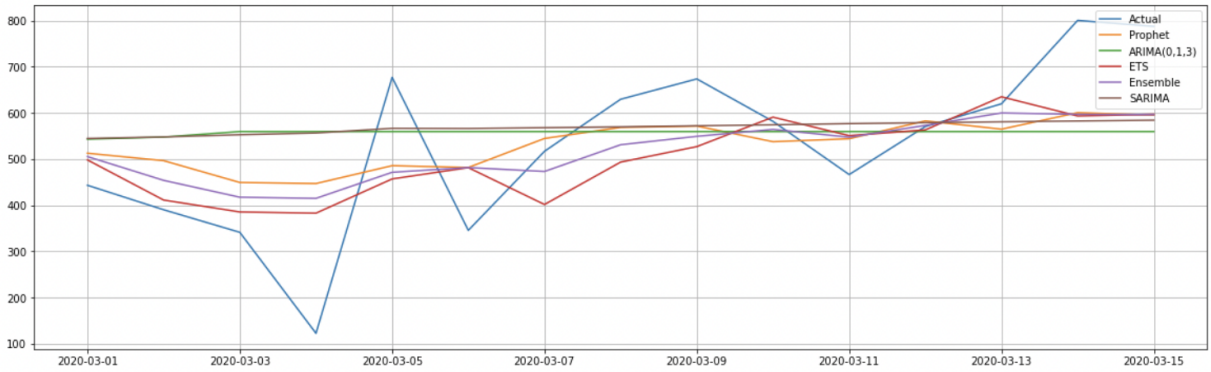
   ANS - None others included.

RMSE COMPARISON TABLE

Model	*Parameters*	*Test RMSE*
ARIMA	*(0,1,3)*	*178.08630729758502*
SARIMA	*(6,0,0)(0,0,0,12)*	*172.87062968603107*
Prophet	*changepoint_range= 0.1 changepoint_prior_scale= 0.02 seasonality='daily', period=345, fourier_order=55*	*139.35284463516473.*
ETS	*trend='add', seasonal='add', seasonal_period=38, damped_trend=True*	*136.6388678425477*
Ensemble	*ETS + Prophet*	*134.77254833721696*

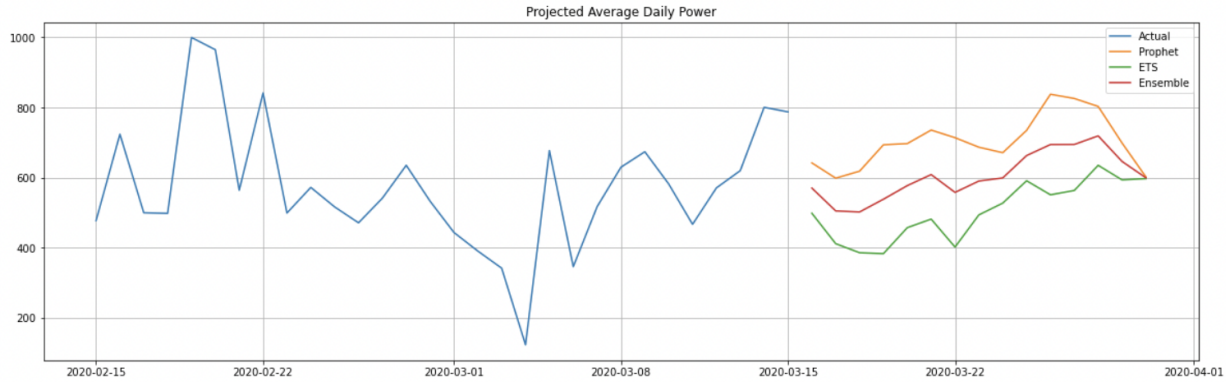
MODEL COMPARISON GRAPH (TEST DATA)

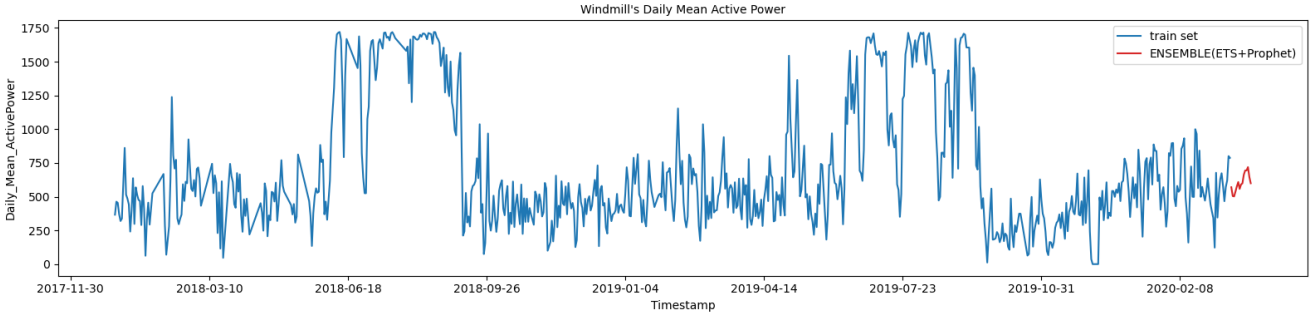
```
Test RMSE for ARIMA(0, 1, 3): 178.08630729758502.
Test RMSE for SARIMA((6,0,0),(0,0,0,12)): 172.87062968603107.
Test RMSE for Prophet model: 139.35284463516473.
Test RMSE for ETS model: 136.6388678425477.
Test RMSE for Ensemble model: 134.77254833721696.
```



FINAL MODEL SELECTION

We have chosen the ensemble model as our final model.




Windmill's Daily Mean Active Power
train set
ENSEMBLE(ETS+Prophet)
Daily_Mean_ActivePower
1750
1500
1250
1000
750
500
250
0
2017-11-30
2018-03-10
2018-06-18
2018-09-26
2019-01-04
2019-04-14
2019-07-23
2019-10-31
2020-02-08
Timestamp

	Date	Forecastings
0	2020-03-16	569.827437
1	2020-03-17	504.576027
2	2020-03-18	501.636445
3	2020-03-19	538.001008
4	2020-03-20	576.802150
5	2020-03-21	608.346056
6	2020-03-22	557.509647
7	2020-03-23	589.881718
8	2020-03-24	598.786681
9	2020-03-25	662.601849
10	2020-03-26	694.097338
11	2020-03-27	694.288503
12	2020-03-28	718.826249
13	2020-03-29	645.862741
14	2020-03-30	599.205819