Report

Ocean Protocol Challenge

Storage Network Energy Consumption - Filecoin

Presenting Data Driven Insights Into
Energy Consumption of Filecoin Storage
Alongwith a Forecast of Storage Capacity
added per day in December.

Submitted by: Udit Gupta

14th December 2022

Github Submission Repository

LinkedIn Profile

Contents

1 Overview **1**

2 Part 1 **1**
2.1 Energy Performance 1
2.2 Storage Capacity Evolution 3
2.3 Share of Renewable Energy 4

3 Part 2 **5**
3.1 Correlation between Sealing Energy and FIL price 5
3.2 Correlation between Storage added per day and FIL price 6
3.3 Correlation between Energy Consumption and FIL price 6

4 Part 3 - Prediction of Storage Capacity for December **8**
4.1 Regression Splines[1] - (to predict) Daily Storage Added 8
4.2 Regression Splines[1] - (to predict) Cumulative Daily Storage Added . . 9
4.3 Prophet model[2] - (to predict) Daily Storage Added 10
4.4 Prophet model[2] - (to predict) Cumulative Daily Storage Added 11
4.5 Comparing Performance of 4 methods 11
4.6 Prediction for December 12

1. Overview

This report presents insights into the Filecoin Storage Network. The report is broadly divided into 3 sections, one for each question asked in the challenge.
Note: The analysis code for Part1 and Part2 is included in the Github Repo as ***eda_analysis.ipynb***.
The Storage capacity Forecasting code for Part 3 is named ***storage_prediction_algo.ipynb***

2. Part 1

2.1 Energy Performance

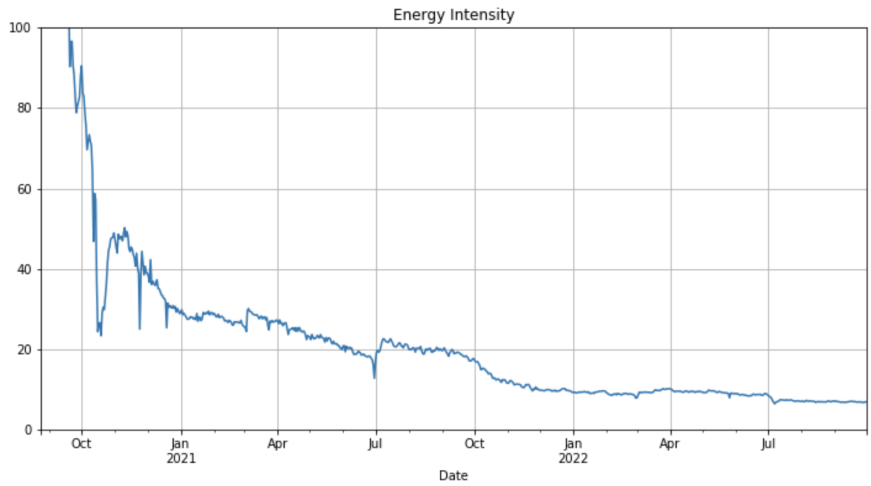

Figure 1: Energy Intensity

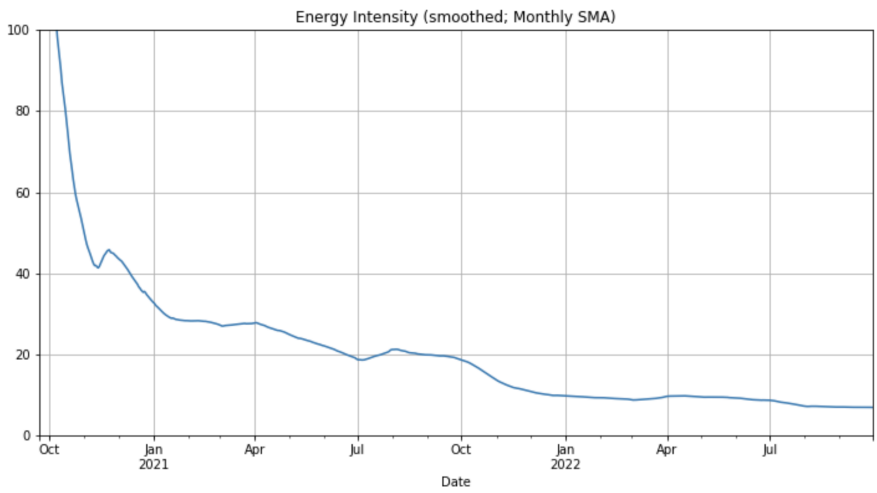

Figure 2: Energy Intensity (Smoothed)

Observation 1: Energy Intensity, that is the Energy consumed per unit data stored by the network is going down over time which means the Network is becoming more efficient in storing a unit of data.

Defining Energy Performance:

$$Energy\ Performance = \frac{K(constant)}{Energy\ Intensity}$$

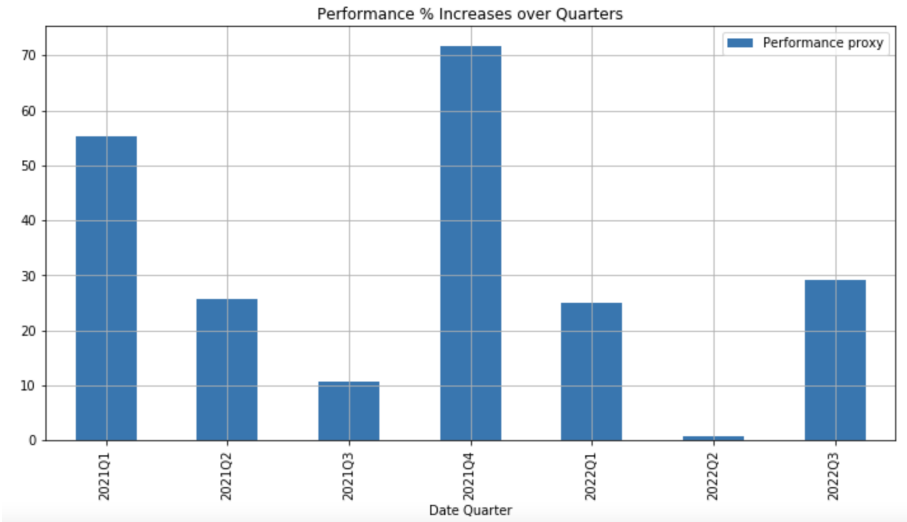

Figure 3: Energy Performance Percentage Increase over Quarters

Insights: The Performance of the Network (as measured by the inverse of Energy Intensity) is improving Quarter over Quarter which shows a positive trend and indicates that the network is becoming more efficient as time progresses.

2.2 Storage Capacity Evolution

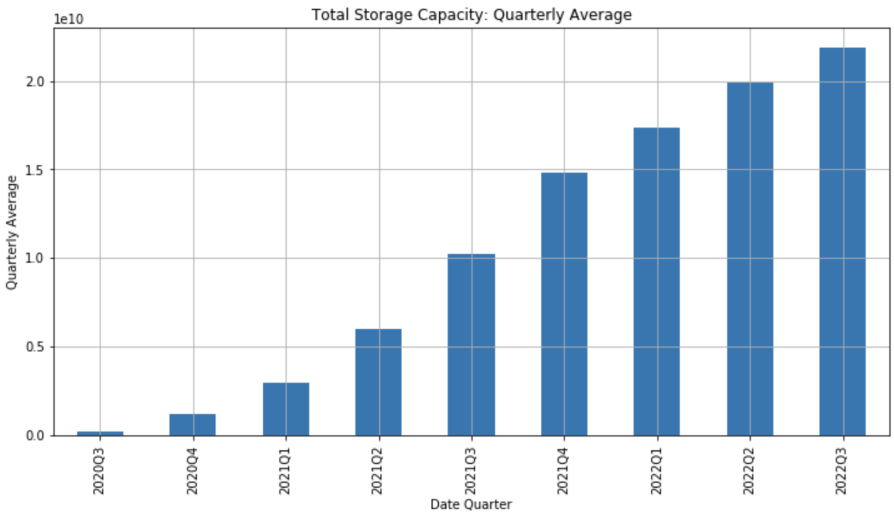

Figure 4: Storage Capacity Increase over Quarters

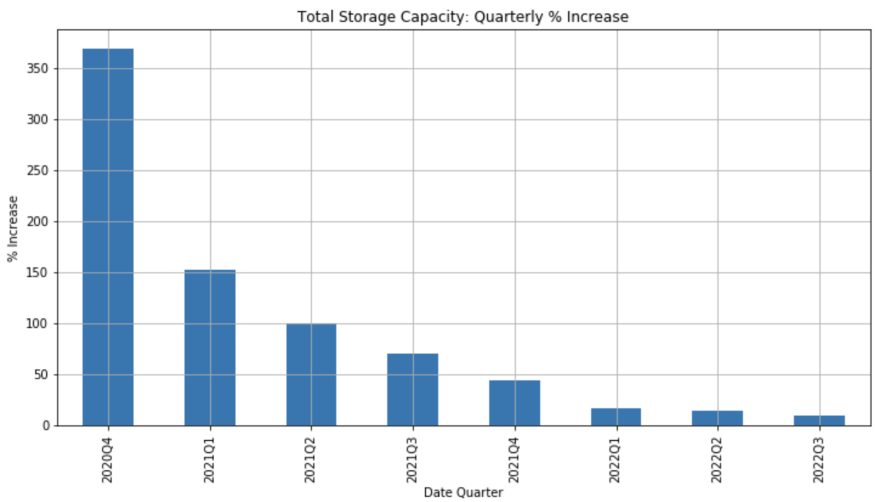

Figure 5: Storage Capacity Percentage Increase

Insights: The Total Storage Capacity of the network has been increasing as depicted in Figure 4.
However, the Percentage Increases over Quarters has been going down as is shown in Figure 5 which means the *Pace of New Storage Capacity expansion is slowing down.*

2.3 Share of Renewable Energy

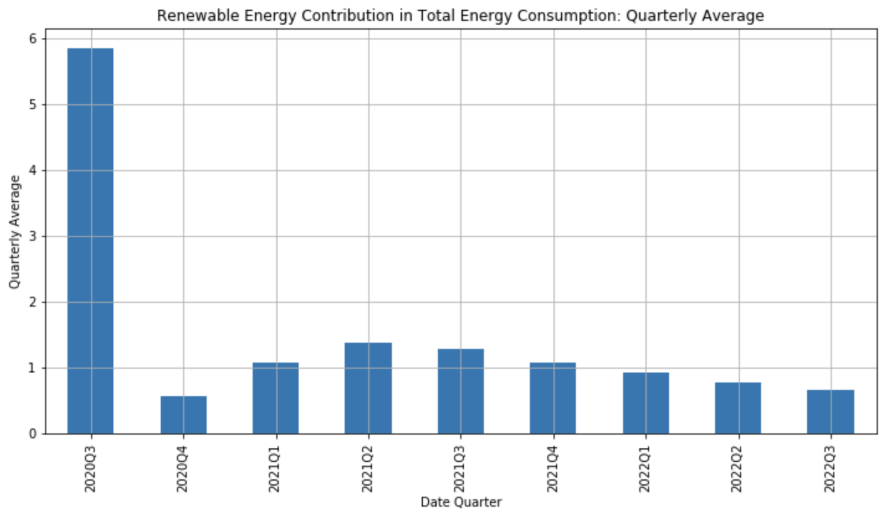

Figure 6: Fraction of Renewable Energy out of Total Energy consumption

Insights: Figure 6 shows the contribution of Renewable Energy in total energy consumption. Q32020 is somewhat of a outlier where the proportion of Renewable energy credits purchased was much higher.
The Renewable energy contribution was increasing until Q32021 which was a positive sign for the network as it had become Carbon Neutral. Q42021 onwards, the Renewable energy fraction has been in a downtrend and the according to the most recent Quarter, the fraction of Renewable Energy is only 60% of the total energy consumed. This indicates that the Carbon emissions are increasing over the past four quarters.

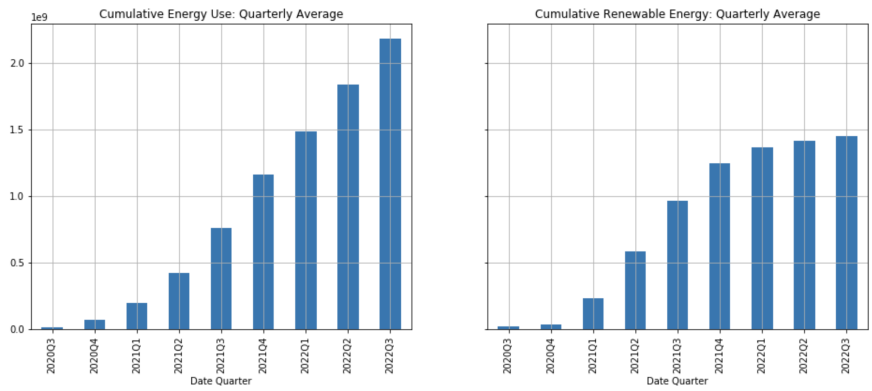

Figure 7: Cumulative Energy used vs Cumulative Renewable Energy used

Insights: It is evident from Figure 7 that though the Total Energy consumption has been consistently increasing over time (left chart), the Renewable Energy purchases (right chart) have reached a level of saturation over the past 4 Quarters.

3. Part 2

3.1 Correlation between Sealing Energy and FIL price

Observations: The Correlation between the FIL price[1] and the Energy used to Seal data comes out to be **0.58** considering all the available data. This shows that a decent level of correlation exists between the two entities. However, looking at the data within each Quarter reveals more interesting information.

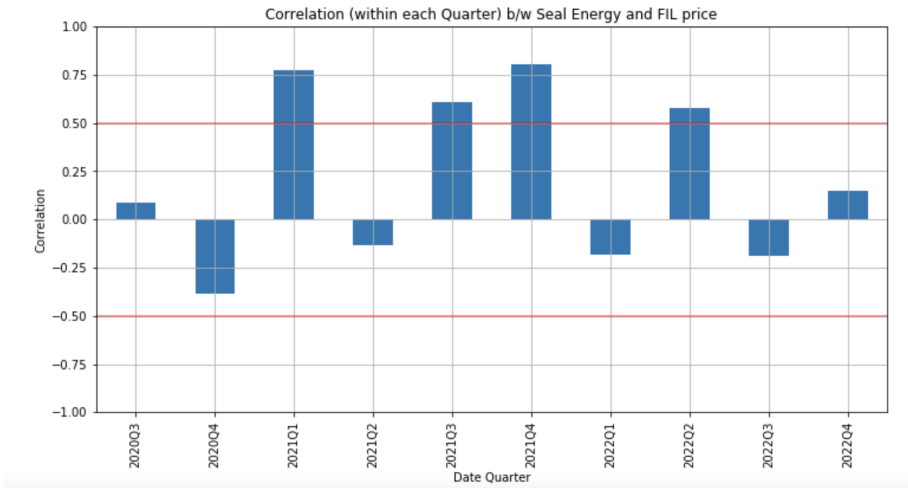

Figure 8: Correlation between Sealing Energy and FIL price (Within Quarters)

Insights: Figure 8 shows the Correlation within each Quarter, and plots this correlation for all Quarters. It can be seen that, **only 4 out of 10 quarters depict strong positive correlation.**

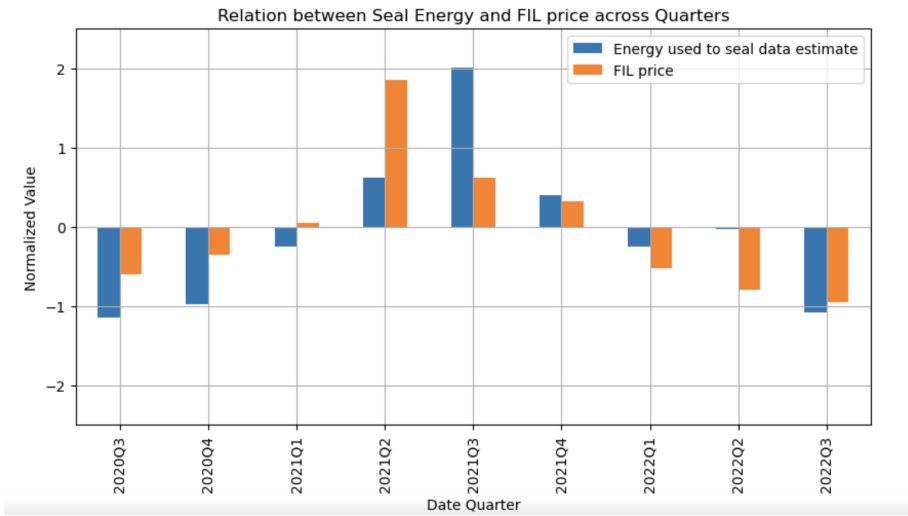

Figure 9: Average Normalised value of Sealing Energy FIL price

[1] Adjusted Close Price used for computing Correlation throughout the report

Insights: Figure 9 compares the values of FIL price and Sealing Energy. The comparison is done using the Average of Normalised values so that they are on the same scale. It can be seen that **9 out of 10 times both quantities have same signs** (i.e. positive and negative signs). The Correlation between Average values across Quarters is **0.66**.

Findings: Within a Quarter, the FIL price and Sealing Energy do not show strong correlation indicating that the values are too volatile or noisy at the daily level. However, when we Zoom out and look at the average values across Quarters, then we can definitely see some trend which indicates that the **Correlation exists but only when compared on an average basis and over a long enough time horizon**.

3.2 Correlation between Storage added per day and FIL price

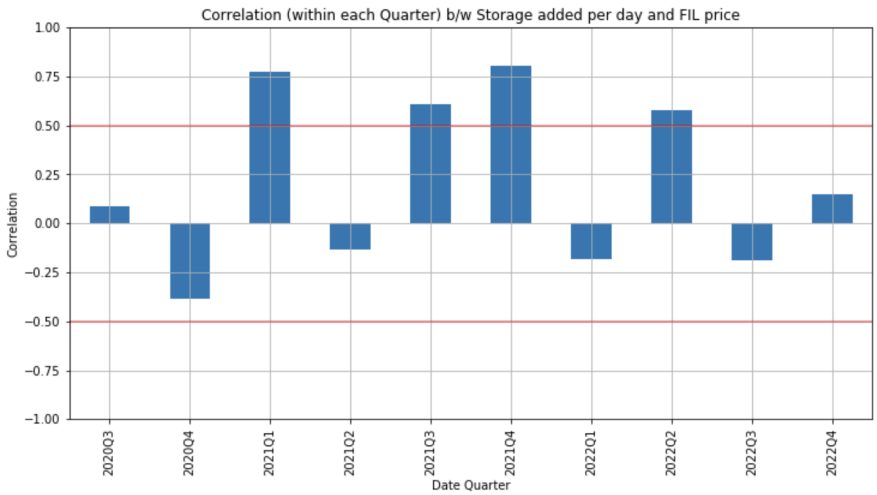

Figure 10: Correlation between Storage added per day FIL price (Within Quarters)

Insights: This is exactly same as the details provided in 3.1. This is because by definition, the Sealing Energy is estimated by multiplying a constant with Storage capacity over a given time. Since multiplication by a constant doesn't affect correlation, the insights for 3.2 will be similar to the insights presented in 3.1

3.3 Correlation between Energy Consumption and FIL price

Observations: The Overall Correlation between Energy Consumption and FIL price is **0.08** which indicates that these two entities are uncorrelated.

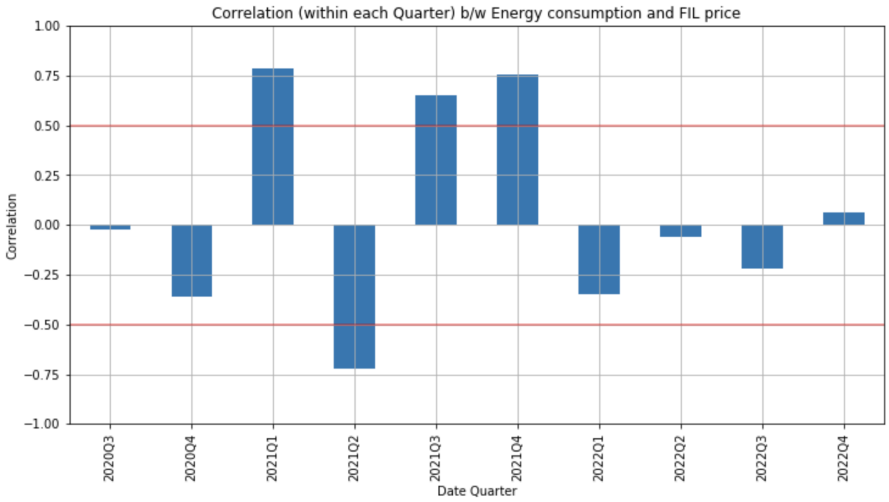

Figure 11: Correlation between Energy Consumption and FIL price (Within Quarters)

Insights: In Figure 11 there are 3 strong positive correlations and 1 strong negative correlation, i.e the direction of correlation switches in different Quarters. Overall, there are 5 positively correlated Quarters and 5 negative correlated Quarters which means that **predicting correlation within a Quarter is a coin toss** and there is no statistical evidence that higher energy consumption leads to higher FIL price.

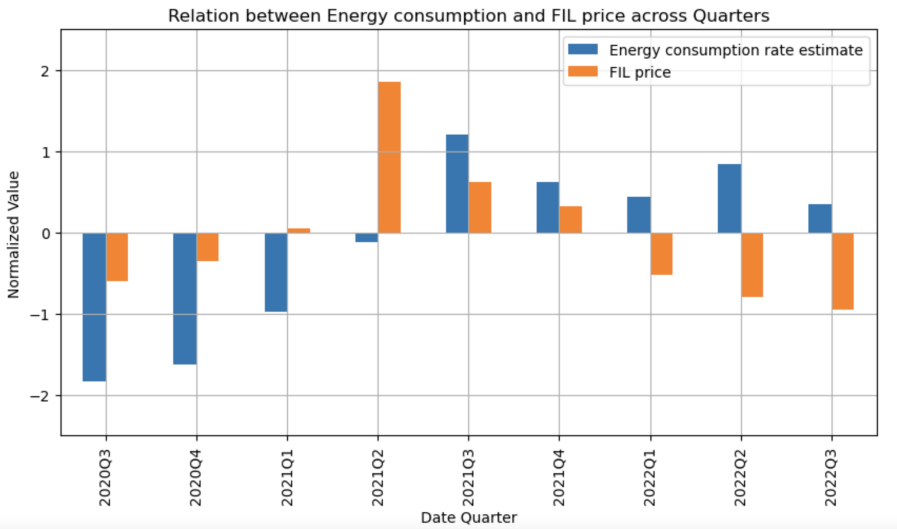

Figure 12: Average Normalised value of Energy Consumption and FIL price

Insights: In the first 2 Quarters i.e 2020Q3 and 2020Q4 and in the middle quarters 2021Q3 and 2021Q4, both the entities had same relationship i.e. both were either positive or both negative. However, in the last 3 Quarters i.e 2022Q1, 2022Q2 and 2022Q3, the direction has been opposite. Here, though the Energy consumption has been high but still the FIL price has been relatively low. This again indicates no clear correlation between the two entities.

4. Part 3 - Prediction of Storage Capacity for December

Predicting the Storage capacity added per day for December is non trivial because the data provided is till 16th October 2022. So the lead time is from 16th October to 1st December which is 46 days. And the prediction horizon is from 1st December to 31st December, therefore the prediction required is 46 days to 77 days ahead compared to the data available.

To accomplish this task, experimentation is required to see which approach works the best. To evaluate which technique works best, the data from last 30 days i.e. 17th September to 16th October 2022 is used as a test set, and the remaining data before 17th September 2022 is used for training the algorithm.

Experiments are conducted with 4 *Time Series Forecasting* techniques to find out which works best in this scenario. These techniques are:

1. Regression Splines[1] - (to predict) Daily Storage Added
2. Regression Splines - (to predict) Cumulative Daily Storage added
3. Prophet Model[2] - (to predict) Daily Storage Added
4. Prophet Model - (to predict) Cumulative Daily Storage added

4.1 Regression Splines[1] - (to predict) Daily Storage Added

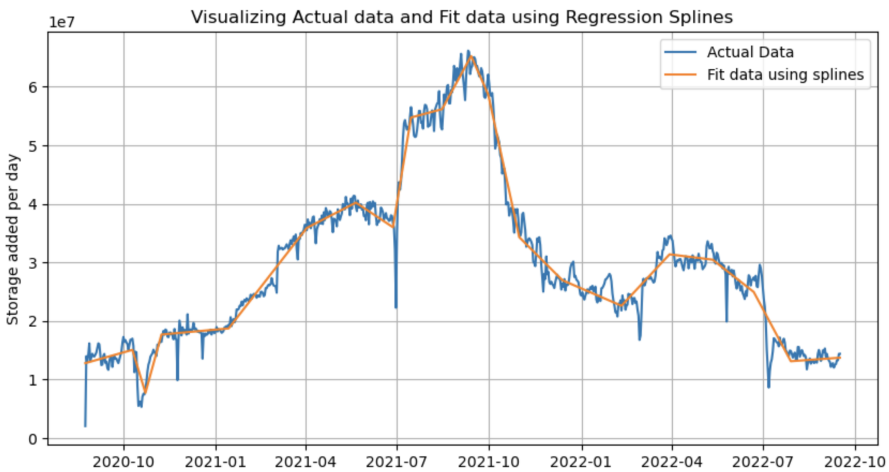

Figure 13: Using the Algorithm to predict on Training data

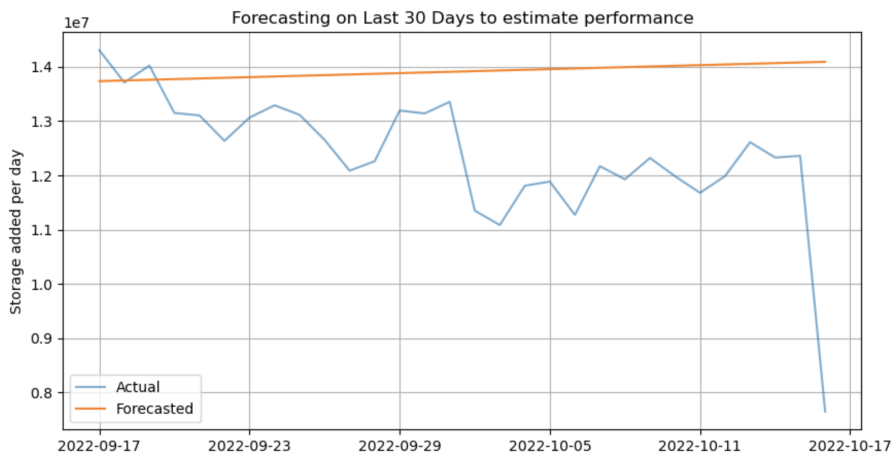

Figure 14: Using the Algorithm to predict on Test data

The Regression spline algorithm is fit on the *Daily Storage data added* as shown in Figure 13. We can see that the data fit by the algorithm is close to the actual data. To know the efficacy of this technique, we have to make a Forecast on unseen data and compare it with actual data. The Forecast is shown in Figure 14.
Note: The py-earth package by scikit-learn [3] was used for this application.

4.2 Regression Splines[1] - (to predict) Cumulative Daily Storage Added

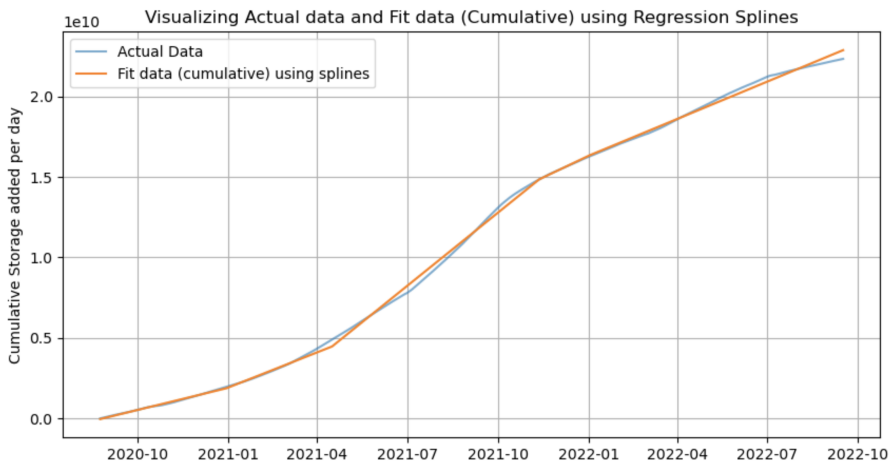

Figure 15: Using the Algorithm to predict on Training data

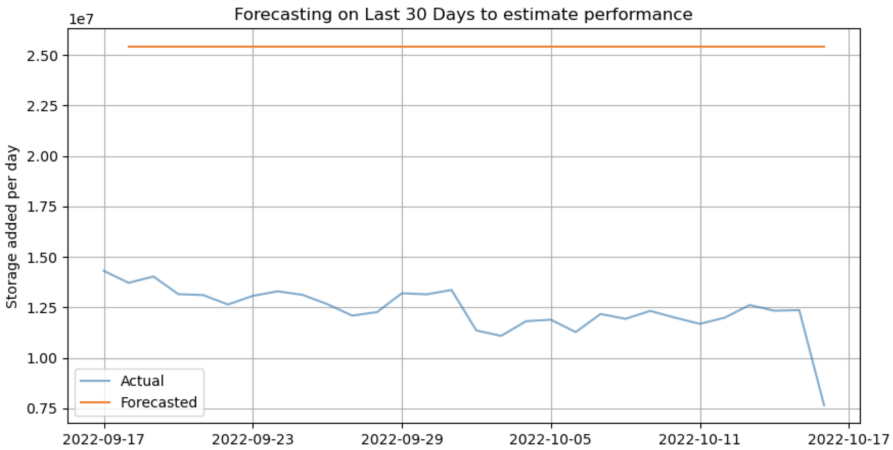

Figure 16: Using the Algorithm to predict on Test data

In this section, the *Cumulative Storage added per day* is used for learning the model. In Figure 16 we can see that the algorithm overpredicts the storage added per day i.e. this algorithm does not perform well.

4.3 Prophet model[2] - (to predict) Daily Storage Added

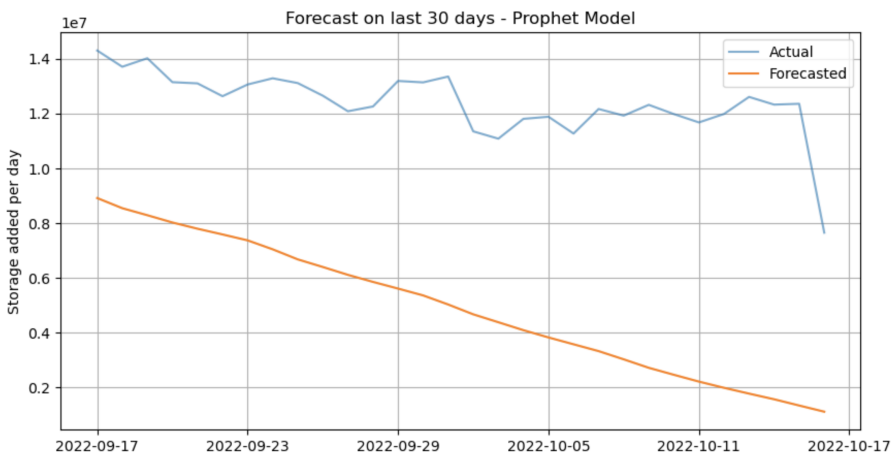

Figure 17: Using the Algorithm to predict on Test data

The Prophet model[2] is used to make the forecast for Storage added daily. Figure 17 shows that in this case, the Prophet model underpredicts the Storage added daily. This estimation by the model is also not good.
Note: The kats[4] library is used for learning the Prophet model.

4.4 Prophet model[2] - (to predict) Cumulative Daily Storage Added

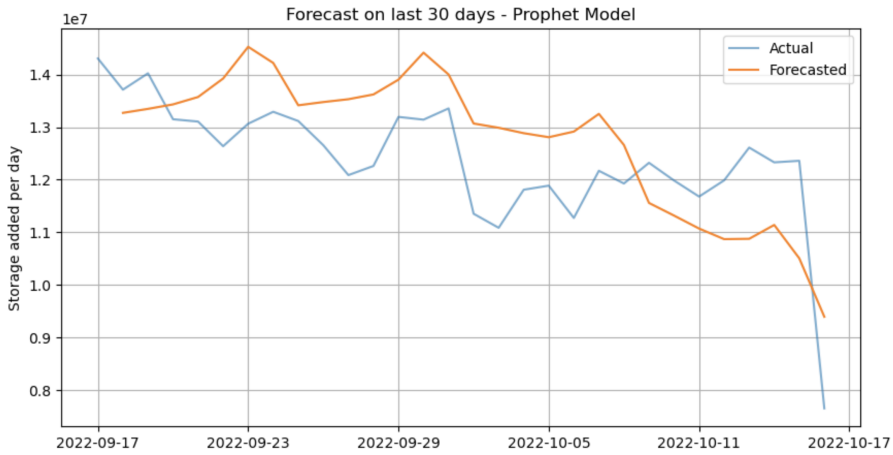

Figure 18: Using the Algorithm to predict on Test data

Here, we fit the Prophet model on *Cumulative Storage* data. Figure 18 shows that the **Forecasted values are very close to the Actual values** which means that the performance of this model is very good.

4.5 Comparing Performance of 4 methods

R^2 score[5] is used to evaluate the performance of each method on last 30 days of data i.e. 17th September to 16th October 2022. This score can vary from $-\infty$ $to + 1$. When the predicted and actual values are exactly the same, then a score of +1 is obtained. On the other hand, when the predicted values starts deviating from actual values, then the score starts to decrease and can go down in an unbounded way towards $-\infty$.

Spline-Non Cumulative	Spline-Cumulative	Prophet-Non Cumulative	Prophet Cumulative
-1.81	-131.8	-41.95	-0.04

Table 1: R^2 Score Comparison for 4 methods

Insights: From Table 1 it is evident that the 4th model i.e Prophet Model which is fit on Cumulative Storage data, performs the best out of all other models. Therefore this method is chosen to make forecasts for the month of December.

4.6 Prediction for December

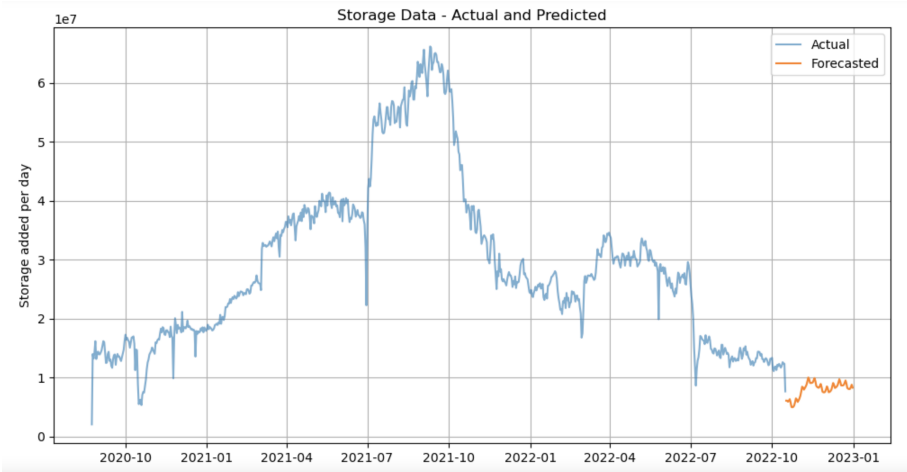

Figure 19: Storage Capacity - Actual and Predicted

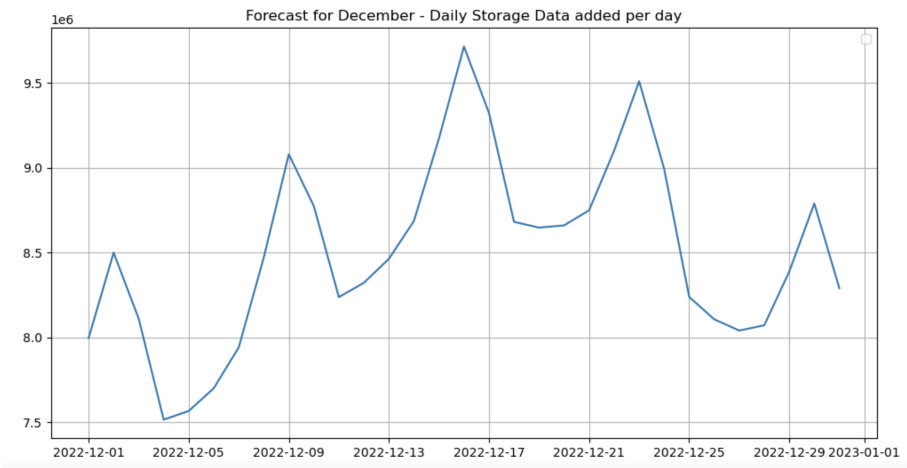

Figure 20: Forecast for December

In Figure 19, the orange curve depicts the Prediction for *Storage capacity added daily*. Figure 20 shows just the Forecasted values for December.

The actual forecasted values are also saved as a csv file in the github repo with the name ***forecast_december.csv***

Bibliography

[1] Bio Statistics collboration of Australia. Piecewise regression and splines. URL https://bookdown.org/tpinto_home/Beyond-Linearity/piecewise-regression-and-splines.html.

[2] Facebook. Forecasting at scale - prophet, 2017. URL https://peerj.com/preprints/3190/.

[3] Scikit Python Package. Py earth documentation. URL https://contrib.scikit-learn.org/py-earth/content.html#multivariate-adaptive-regression-splines.

[4] Facebook Research. Kats - time series analysis framework. URL https://github.com/facebookresearch/Kats.

[5] Wikipedia. Coefficient of determination - r squared. URL https://en.wikipedia.org/wiki/Coefficient_of_determination.