Can ML help to tackle Climate Change - A Case Study on Electricity Consumption

Ratul Chakraborty
ratulc720@gmail.com

1 Abstract

Climate Change has been a matter of concern for the Human race for a long time. Not only as it is directly linked with survivability and sustainability of the race as a whole, but also a matter of concern for our next generation. When it comes to causes of climate change human race themselves are the one to blame. As, mankind climbs the stairs of evaluation, it comes with inevitable side-effects, which are not really desirable to us. Rise in global average temperature, increase in GHG (Green House Gases) in the atmosphere, hole in the Ozone Layer are only a few of the many happening around us. Most of them are outcomes of cumulative 'hard-work' by humans in past few centuries. No outright solution for these is possible. But, we can take small steps towards better future, at the very least stop climate change on its track forward. Here we will show a simple of way of predicting consumption of Electricity, a part and parcel of our life also the biggest instigator of climate change, efficiently to reduce its impact on climate change through Machine Learning.

2 Introduction

Let's start with a simple story. You are a merchant. You sell a product which has great demand in the market. But, you are not the only one with the product. You have competitors. To outshine your competitors, you will sell the product to your customers at a minimum cost while maintaining your profit and meeting demands. Having less of the product is very detrimental to the business. So, you will procure more of it, even if most of it goes to waste.

The story is same for the Electricity Companies a.k.a Suppliers. Electricity market has been generating massive amount of profit over the years. As long as the World is heading towards digitization, demand of electricity is ever rising. Even if they come to know that their product is actually harmful for the planet in the long run, they are unlikely to change their ways. So, let us discover this problem with help of Machine Learning.

Electricity production comes with the issues of over-production, under-production, unequal distribution and etc. Equal distribution of energy is nothing but a fool's dream, as there are issues of connectivity and demand. But, enough distribution is always possible.

As for the other two problems, over-production leads to waste of energy and under-production leads to customer dissatisfaction. What Companies wish to avoid the most is Customer dissatisfaction. As long as they as it is possible to meet the demands, they are fine with overproduction even if it is unnecessary. We can ease the problems to some extent, through controlling production cleverly.

But, it gives rise to a question. Is there any way to predict energy consumption better, without solely relying on past consumption data? In other words, are there other components influencing people to consume more or less energy than the norm?

This is where Machine Learning (ML) comes into play. For years after years, Statistical Univariate Analysis has been the only way to counter this issue. But, ML may just bring change to already existing ways.

We can have assumptions about many such causes, that influence people's energy consumption. But, the one component that here we would like to promote is 'Climate'. To be more exact different climate components like Temperature and Rainfall in a region.

For example, Air Conditioners (AC) are highly used during Summer or Hot days. Whereas, it is not really used during Winter or Cold Days. Just like that, Climate is effecting our daily life without our knowledge. After all, humans are a comfort and luxury pursing species. So, Climate also gets effected indirectly through Electricity Consumption and our actions.

3 Related Work

The most related work for energy consumption was found from POSOCO. Two of their publications were especially remarkable.

- Southern Power Demand Scenario - A Forecast
  A comprehensive overview of demand and forecast in southern states - Andhra Pradesh, Karnataka, Kerala and Tamil Nadu w.r.t seasonal variation and daily load.

- Eastern Region Power Demand Scenario - A Forecast
  A comprehensive overview of demand and forecast in eastern states - BSEB, DVC, GRIDCO and West Bengal w.r.t load factor and load.

But, these two were on purely Uni-variate Analysis. And also the techniques used here are pretty simple, mostly regression and curve fitting.

There have also been works in the field of finding relation between energy climate change and electricity consumption. One of them have been mentioned in recent Climate Change And AI paper. It was published as Recommendations for Government Action from the recent Climate Change And AI meet.The UK National Grid Electricity System Operator (ESO) worked with Open Climate Fix to implement deep learning approaches to help optimize national electricity demand forecasts. After trying over five hundred variants of Temporal Fusion Transformer model, the error managed to be reduced to 1/3 of previous model.

But, here we don't use any overly complicated model to get to the simple goal of predicting electricity consumption.

4 Methodology

4.1 Data Collection and Preprocessing

Here, we are concerned about West Bengal, India. So, we will be dealing with Electricity and Climate data of West Bengal only. Data used in this project mainly comes from two sources, as the data itself is assembled from two different types of data.

- Energy Consumption Data - from POSOCO.
  The weekly reports of each year were used to collect energy consumption data in each state for the past 7 days from date of report being isssued.

- Climate Date - from IMD.
  The official IMD (Indian Meteorological Department) website contains grided rainfall and Temperature data for more than 50 years (nearly).

- State Coordinate - from Maps of India.
  West Bengal's Coordinate were collected to be later used to extract climate data for the state from grided data, as here we are solely focusing on West Bengal.

Data Preprocessing was also done in mainly two parts -

- Energy Consumption Data from weekly reports were extracted. Weekly reports were all pdfs, giving brief overview of all sorts of national and international energy transfer records along with energy consumption in different states, union territories and government owned projects.

- Useful tables were extracted as dataframes from the weekly report pdfs, using python libraries and some hard coding. These dataframes were assembled according to the need of the project.

- Climate Data from IMD was grided data. In other words, it contained rainfall and temperature from different coordinates (longitude-latitude). For rainfall it was at an interval of 0.25 degrees. For temperature it was at an interval of 1 degree.

- The West Bengal state coordinates came handy here. It was used on the coordinates from IMD data through KNN (K-nearest neighbourhood) with k=1. Thus climate data for different coordinates of West Bengal was obtained. We took the average of all these coordinates. And, used as climate data for West Bengal for the sake of simplicity.

- After extracting energy consumption of West Bengal, it was paired up with climate data for West Bengal.

4.2 Regressions

Rest of the work was of simple forecasting. The data was divide into train (0.85) and test (0.15) data. Day wise Energy Consumption data was used as dependent feature (target of prediction), day wise Climate data was used as independent features (features to be used for prediction).

Train	Test
0.85	0.15
01/01/2018 to 04/07/2020	19/07/2020 to 31/12/2020
854 Data Points	151 Data Points

Different types of regressions were used here. From, the most basic Ordinary Least Square to the much more stable Random Forest Regression. Random Forest Regression was an obvious choice in this case. As, it is known for its stability.

Random Forest Regression is a supervised ensemble learning method. It mainly is a bunch of random Stumps (Small Decision Trees), combined together to bypass the One of the major issues of a Single Decision Tree, which is High Variance (tendency to overfit training data). Whereas, for Random Forest a bunch of stamps are trained through Row-sampling and Feature-sampling, not on the whole dataset to bypass Overfitting. In case of, Random Forest Classifier decision is made through Majority Voting among the random Stumps. But, for Random Forest Regressor, we take the Mean of random stumps outcomes (in Sci-Kit Learn). So, we get a single value as outcome.

By averaging the decisions of Random Stumps in a Forest, it proves to be better than a single Decision Tree, hence improves its accuracy and reduces over-fitting.

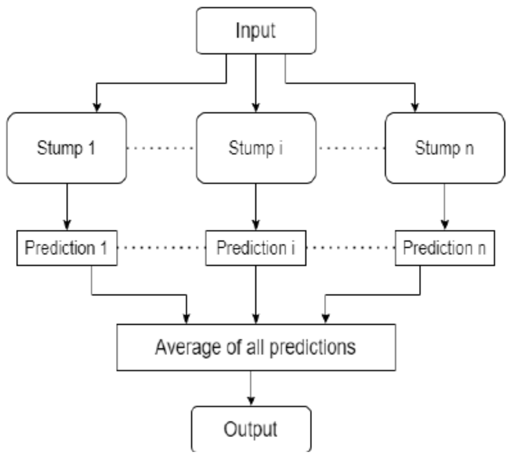

Figure 1: Random Forest Regressor

4.3 LSTM Neural Network

Train	Validation	Test
0.6375	0.2125	0.15
640 Data Points	214 Data Points	151 Data Points

We have also used, Long Short Term Memory network (LSTM) here. LSTM works really well on time dependent data. It was originally developed as an extension of Recurrent Neural Network (RNN). RNN has a very short term memory and tends to forget what it has been trained 5 - 10 discreet instances ago.

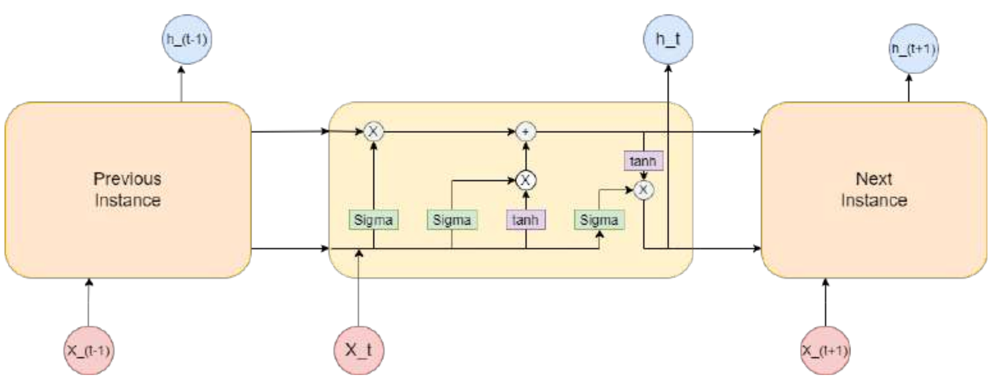

Figure 2: An example of LSTM structure

LSTM bypasses this issue through receiving two inputs at each instance, namely Hidden State (Short Term Memory, used to be a part of RNN) and Self State (Long Term Memory, addition to new LSTM). In the architecture of LSTM, three types of gates are added, namely Input Gate (decides what to take as input from previous instance), Output Gate (decides what to send to next instance) and Forget Gate (decides what information to forget and modifies the Self State). Through these LSTM carries over partial information from many instances ago, also modifies itself to fit the current instance.

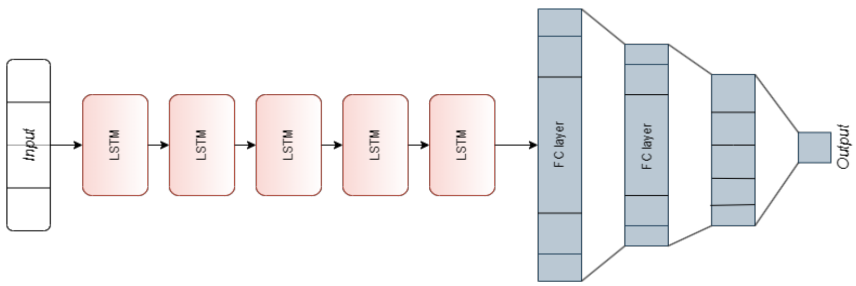

Figure 3: LSTM architecture used for Prediction

The Neural Network used here has 5 layers of sequential LSTM, followed by 4 fully connected layers, with activation functions of ReLu (Rectified Linear unit). As, the data itself was time dependent, LSTM was inevitable.

5 Experimental Results

5.1 Ordinary Least Square

Ordinary Least Square have been used as one of the models fitted here. It has fit the data with a R2-Score of 0.7521. In other words, the fitted model is emulating around 75% of the real world data.

In Figure 4, we can clearly see that predictions from OLS are not really following the original data faithfully and at the end (right most) the prediction is overestimating the data quite a bit, while at some places the data is being underestimated.

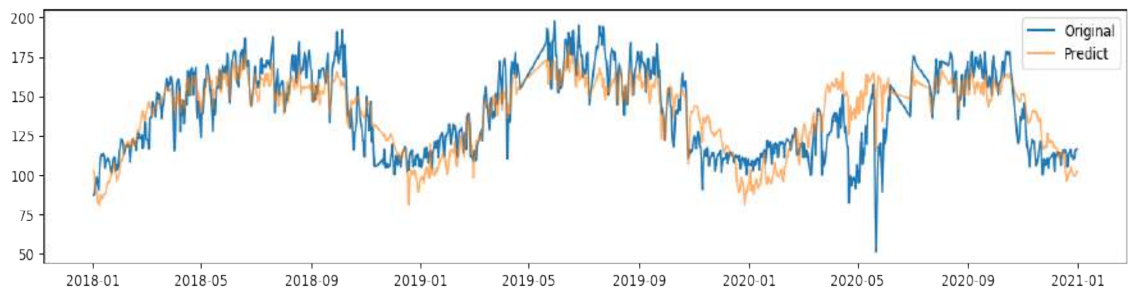

Figure 4: Comparing Original Data with Fit from OLS

5.2 Least Absolute Shrinkage and Selection Operator

Least Absolute Shrinkage and Selection Operator (LASSO) have been used as one of the models fitted here. It has fit the data with a R2-Score of 0.7711. In other words, the fitted model is emulating around 77% of the real world data.

In Figure 5, prediction from LASSO is estimating the original data quite a bit better than OLS predictions. It is still overestimating and underestimating the data nearly at the same places as OLS, but the over estimations and underestimations are a little less compared to OLS.

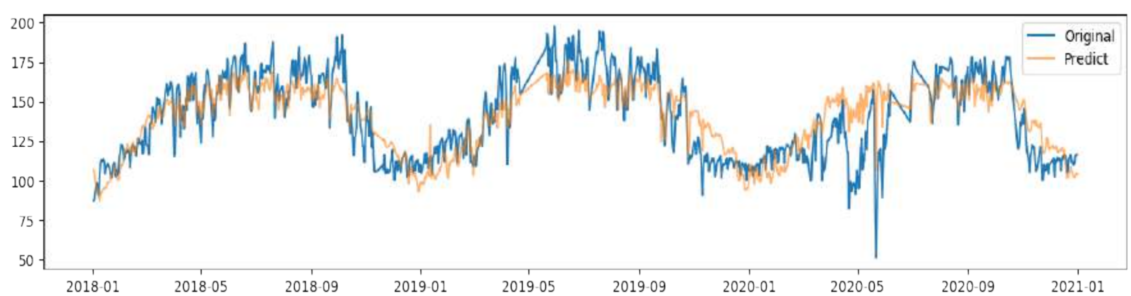

Figure 5: Comparing Original Data with Fit from LASSO

5.3 RIDGE

RIDGE Regression have been used as one of the models fitted here. It has fit the data with a R2-Score of 0.7699. In other words, the fitted model is emulating around 76% of the real world data.

Figure 6 is nearly similar to LASSO predictions.

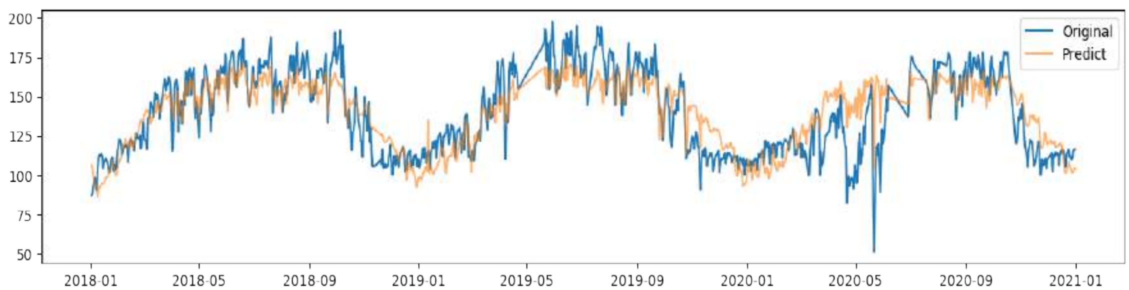

Figure 6: Comparing Original Data with Fit from RIDGE

5.4 Random Forest Regressor

Random Forest Regression have been used as one of the models fitted here. It has fit the data with a R2-Score of 0.8827. In other words, the fitted model is emulating around 88% of the real world data.

In Figure 7, the predictions from random forest regressor is following the original data very closely, to the point they are overlapping quite a bit. Overestimations and underestimations are also very less both in prediction and numbers.

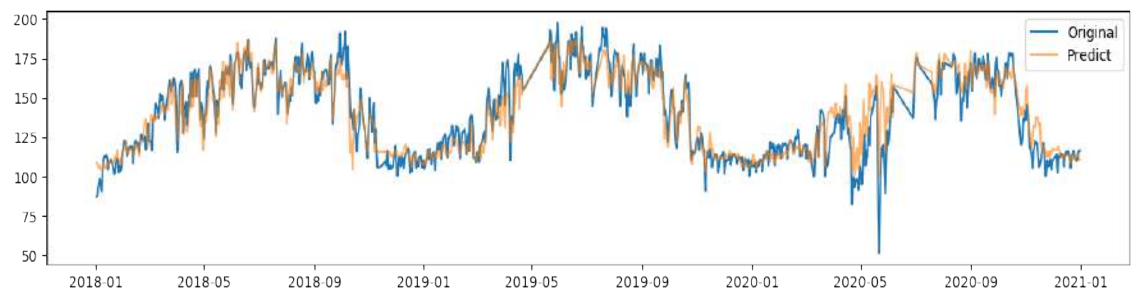

Figure 7: Comparing Original Data with Fit from Random Forest Regressor

5.5 Comparison

Here, we have put the original data against all the fits. From this comparison here, it is clear to see which model has fitted best. As, said from the previous arguments we can clearly see that Random Forest Regression has done the best work, as clearly follows the original input data very closely. Whereas, for other regression quite a bit of deviation is evident.

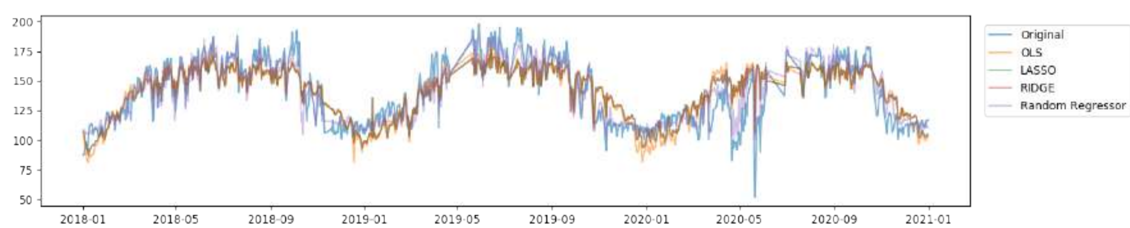

Figure 8: Comparing Original Data with different fits

5.6 LSTM

Long Short Term Memory have been used as the main model fitted here. It was an obvious choice as the data itself is Time Dependent. It has fit the data with a R2-Score of 0.9147. In other words, the fitted model is emulating around 91% of the real world data.

In Figure 9, even if the prediction is not really overlapping with the original data for most of the time. But, it is following the original data nearly parallel for most of the time.

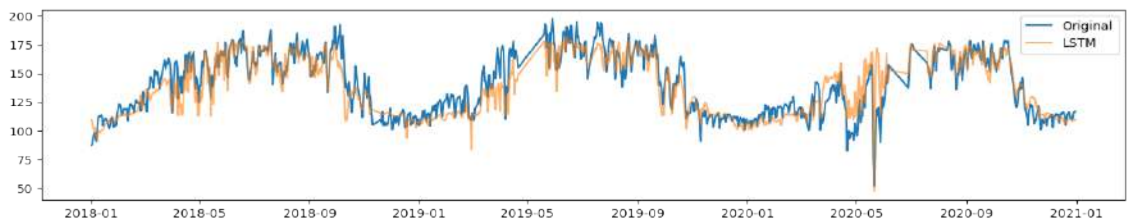

Figure 9: Comparing Original Data with Fit from LSTM

6 Conclusion

In this work, we don't use any fancy method to reach our target. The main target is to increase awareness about our climate. As, our climate changes so do we. Our dependency on modern facilities is the main reason behind it, but we also can't really abandon them. Most of the time electricity is made through burning natural resources (Fossil Fuels). In the production part of electricity through fossil fuels a huge amount of GHG (Green House Gas), and other toxic elements are created as byproducts, which in turn effects Climate in a negative way. A better climate may lead to a better future.

In the last Conference of Parties meeting held on October 2021 at Glasgow, UK, definite steps towards climate control were discussed. More than 2030 new emission sources have been identified. Over 90% of them are being covered by net zero commitments. It has been estimated that restoring ecosystem and managing land sustainably can reduce GHG emission by 7 giga tonnes. But, most importantly 190 countries have agreed to phase down by coal power, by reducing pipelines of new coal plants by 76%.

Some countries have already become aware of the effects of climate change and how much electricity effects it. Some (like UK) have been using ML as a way to gain better control over Electricity production while meeting population demand and wasting as little as possible. They may serve as an inspiration for others to adapt ML as one of effective tools against Climate Change.

Here, we have described a simple process of collecting and assembling data to make your own Climate related dataset. Also, that even the simplest of the models may prove to be pretty decent depending on the nature of data.

One of the limitations of this work would be to not have enough amount of data. As, we are predicting daily Electricity Consumption its practical use is quite limited. It would have been much more practical if, we had enough data to predict Electricity Consumption for finer intervals, like one instance for an hour or even finer at one instance for every quarter of an hour. Also, being unable to deploy the model, which was the original intention. But, had to abort as new climate data is yet to be updated in the IMD website. So, we do not have enough Climate data for up-to date predictions.

7 References

References

[1] Maps of india. https://www.mapsofindia.com/lat_long/.

[2] UN Climate Change Conference UK 2021. Cop26 negotiations explained. https://ukcop26.org/wp-content/uploads/2021/11/COP26-Negotiations-Explained.pdf.

[3] UN Climate Change Conference UK 2021. Cop26 the glasgow climate pact. https://ukcop26.org/wp-content/uploads/2021/11/COP26-Presidency-Outcomes-The-Climate-Pact.pdf.

[4] David Rolnick Et al. Tackling climate change with machine learning. https://arxiv.org/pdf/1906.05433.pdf.

[5] Colah. Understanding lstm networks. https://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[6] India Meteorological Department. Free download yearly gridded data binary file. https://www.imdpune.gov.in/Clim_Pred_LRF_New/Grided_Data_Download.html.

[7] Government of India. Power system operation corporation limited : National load despatch centre. https://posoco.in/.

[8] Global Partnership on AI. Climate change and ai - recommendations for government actions. https://gpai.ai/projects/responsible-ai/environment/climate-change-and-ai.pdf.

[9] The Click Reader. Random forest regression explained with implementation in python. https://medium.com/@theclickreader/random-forest-regression-explained-with-implementation-in-python-3dad88caf165

[10] L.K.Kanungo U.K.Verma S.P.Barnwal Shamreena.V R.G.Yadav, S.K.Soonee. Southern region power deamnd scenario - a forecast. https://posoco.in/papers/Southern%20Region%20Power%20Demand%20Scenario%20-%20A%20Forecast.pdf.

[11] S.S.Barpanda K.K.Ram S.K.Sinha, S.K.Soonee. Eaastern region power demand scenario - a forecast. https://posoco.in/papers/Eastern%20Region%20Power%20Demand%20Scenario%20-%20A%20Forecast_GMPS_2000.pdf.