Energy Demand Predictor

Tushar Mohandas, Sai Srujan Eppalapally, Dinesh Balani

Department of Computer Science, Stony Brook University
100 Nicolls Rd, Stony Brook, NY 11794, USA

{tmohandas,saeppalapall,dbalani}@cs.stonybrook.edu

I. Introduction

Energy Demand Predictor is a model to predict the energy demand for a region on a weekly basis. This model also compares electricity pattern usage at state level and clusters the states showing similar pattern together. We have also developed a prototype to make this comparisision at county level and cluster all the counties showing similar consumption pattern together. We used technologies like Spark, TensorFlow and scikit-learn to develop this model. This report has been divided into following sections. Firstly, it describes how this model is aligned with Sustainable Development Goal 7 - Affordable and Clean Energy referred as SDG 7 from here on. In the next section, we have described the source, authenticity, volume, various features and a hierarchy of the data used for analysis. Section 4, gives an in detail explanation of the methods and technologies used to train an autoregressive model to make energy demand predictions and clustering techniques applied to perform analysis. We have stated the results that were obtained while testing the trained model and the analysis made through clustering in section 5, and discussed some of the inferences we have learned through these results and how it can be useful in achieving SDG 7 in section 6. Finally section 7, presents our conclusion.

II. Sustainable Development Goal & Background

Sustainable Development Goal 7 aims to ensure affordable, reliable and sustainable and modern energy for all. Energy plays the most essential role in achieving almost all of the other Sustainable Development goals. SDG 7 has been shortlisted/divided in 3 subgoals.[1]

- Ensure universal access to affordable, reliable and modern energy services[1]
- Increase substantially the share of renewable energy in the global energy mix[1]
- Double the global rate of improvement in energy efficiency[1]

These goals defined by the United Nations, have specific indicators attached to them to track the progress of the work being done, and calculate the percentage of the target achieved.

A. Related Work and progress

Numerous amount of projects and policies laid down by UN, governments and other private organizations has led to improvement in access to affordable and clean energy.

With the growth of smart sensors and the big data they produce, and new companies that know how to crunch that information, energy users from huge factories down to individual households can track and reduce waste of electricity in a way that simply wasn't possible just a few years ago[2]

Statistical analysis of energy consumption data for different types of buildings across the USA has been done using big data technologies[8].

These improvements fall short of what is needed to be achieved. There is a requirement for high level of financing, bolder policy and mutual cooperation amongst the countries to leverage new technologies like Big Data analysis on a much wider scale.[1]

B. Our Model

Energy is central to nearly every major challenge and opportunity the world faces today.[4] Electricity is one of the important forms in which energy is consumed. We aim to improve access to electricity and promote efficient energy usage by performing analysis on electricity consumption data and gaining insights about the usage patterns. Specifically, our model performs time series analysis on the consumption data and predicts the energy consumption requirement for the coming weeks. Understanding the future requirement is very critical, since it will help in better planning and provisioning of resources leading to formation of strong policies for production and distribution of electricity. This cost efficient approach will help in directing finances towards renewable sources of energy. Performing techniques like clustering will help us identify different regions with similar consumption patterns in a country. A region can be a state or a county or a city. The information gained from these analyses go hand in hand. While clustering helps us promote efficient energy usage and help save energy resources, demand forecasting helps in directing those resources to regions with a supply shortfall. Next section describes the specific database on which we performed the analysis to demonstrate a working prototype.

III. Data:

The dataset used for prediction contains commercial and residential load profiles for all TMY3 locations in the United States. The data was obtained from OpenEI and authored by Office of Energy Efficiency & Renewable Energy [3]

TABLE I

Data Description

S.No.	Observation	Values
1	Data Size	19GB
2	Data Features	*State, County, Time Period, Unit Type, Net Electricity Consumption, Net Natural Gas consumption*
3	Number of files	936
4	Number of observations	8760
5	Total Observations	8199360

The data obtained was organized in folders for each TMY3 locations. Inside each folder there are 16 files for 16 types of Commercial Building and a separate folder contained data for 3 types of Residential Buildings. Each file contains hourly load profile observations. The observation were in unit of Kilowatt/hour (kWh).

For cluster analysis we obtained the data regarding per capita energy consumption across all the states in U.S from U.S Energy Information Administration [5] and multiplied the per capita consumption with the state population from CDC[7] to obtain the total energy consumption for the state. We have also collected the county wise energy consumption for California State from California Energy Commission[6] and performed county level cluster analysis.

IV. Methods

A. Pre Processing

The dataset had data about 936 locations where each location had 19 types of buildings which includes 16 commercial and 3 residential types with hourly electricity and natural gas consumption data for a year. We used Spark Map Reduce to perform data pre-processing and aggregation. We aggregated the building-wise energy consumption data for each location and further aggregated the hourly data to obtain weekly data for each of the 936 locations.

B. Forecasting Energy Demand

For predicting the energy usage, we ran autoregressive model of order 3 (AR3) on 936 TMY3 locations across USA. We normalized the weekly electric consumption data for each location and ran regression analysis using 60% of the data i.e. 31 weeks, aggregated for each of the location as our training data and made a single autoregressive model based on the below equation

$$X_t = \beta_0 + \beta_1 Xt_{-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \text{error}$$

We used the Gradient Descent Optimizer provided by TensorFlow to minimize the error and find the optimal beta values and performed testing on the remaining 40% of the data from the 936 locations based on the derived beta values and found the Mean Squared Error, Mean Absolute Error and Pearson Correlation for each of the 936 locations. Finally, we took the average of all the values per metric and found the average Mean Squared Error, Mean Absolute Error and Pearson Correlation for the model. Using a similar process, we also developed an AR5 model.

C. Finding Regions with Similar Consumption Patterns

In order to achieve our goal of promoting efficient use of energy, we decided to find regions with similar consumption patterns. For this purpose, we used K-Means Clustering. The features selected were per capita energy consumption and total energy consumption of the regions over a period of a year. After normalizing the data, we performed the analyses at two different regional aggregation levels viz. state-wise and county-wise. At the state aggregation level, we obtained 4 clusters and at the county aggregation level, for the state of California, we obtained 3 clusters.

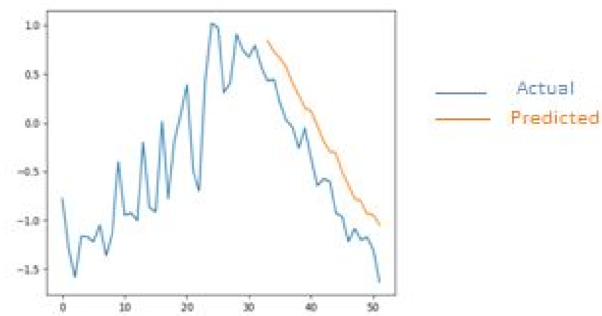

Fig 1: AR5 - Cambria County, Pennsylvania

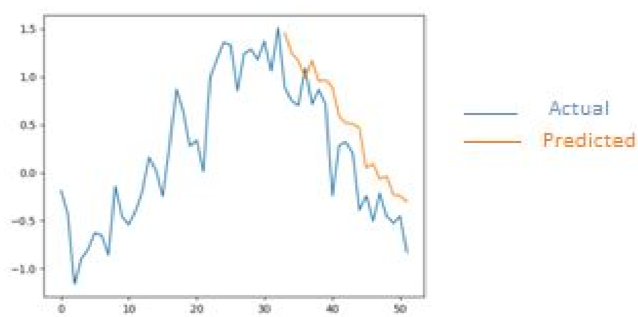

Fig 2: AR5 - Suffolk County, New York

V. Results/Evaluations

1. AR 5 model with 60% training data and 40% test data

Mean Squared Error	0.24236329817224034
Mean Absolute Error	0.40126880589224811
Pearson Correlation	0.84690542346941533

BetaValues [0.15082027, 0.37889251, 0.18437363, 0.14899378, 0.05128516, 0.21138208]

2. AR 3 model with 60% training data and 40% test data

Mean Squared Error	0.17957280035758252
Mean Absolute Error	0.33124294895069917
Pearson Correlation	0.84099051131524427

BetaValues [0.1120691, 0.43689749, 0.27935618, 0.24804688]

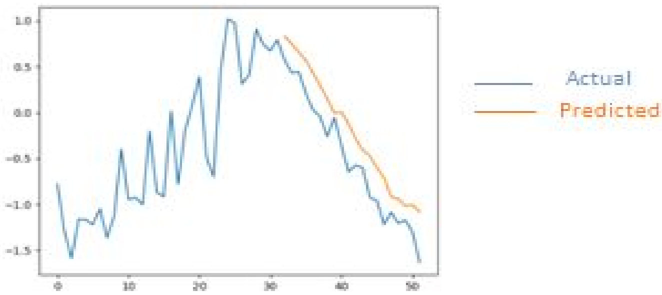

Fig 3: AR3 - Cambria County, Pennsylvania

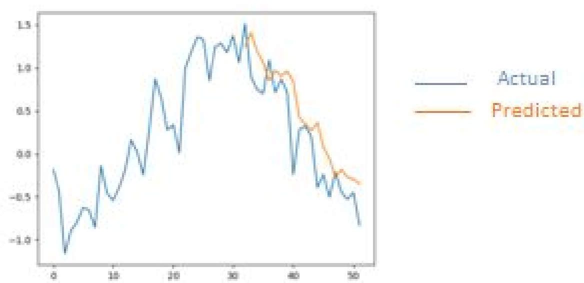

Fig 4: AR3 - Suffolk County, New York

3. Results of Clustering:

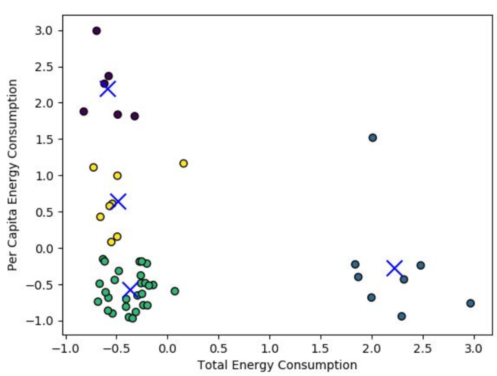

Fig 5: State-Level 4-Mean Clustering

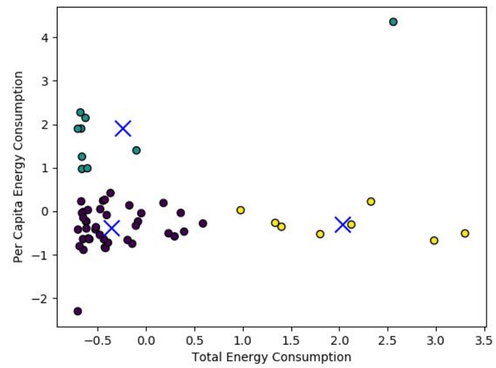

Fig 6: County-Level 3-Mean Clustering (California State)

VI. Discussion

As observed in results though the correlation is similar, the gap between the actual and predicted curve in AR5 model is larger than AR3 model for the test data, which is resulting in an higher error rate in AR5 than in AR3. As the error metrics obtained in the AR3 model were better than AR5, we propose AR3 model for energy demand prediction over AR5.

The knowledge derived from the Energy Demand Prediction can be used to help improve the access to electricity. In the long term, knowing the energy requirements of a region would help the governments and other stakeholders to do better capacity planning. Knowing the demand in advance would also give them enough time to invest in non-renewable energy resources and harness them in place of fossil-fuels and other conventional sources of energy. In the short-term, this knowledge can be used to drive inter-region (National/ International) energy trade. For example, an energy surplus nation could make arrangements to transmit the surplus electricity to a neighboring nation with a shortfall of electricity. This will not only help improve access of electricity but also help improve trade and potentially help work towards other SDGs.

The information about similar regions can be used to drive an incentive-based program to encourage the consumers to use energy more efficiently. One of the ways in which an incentive program could be devised is as follows. The energy distribution organization using the data about the regions with similar consumption patterns from the model we developed, can derive an "optimum" value for the consumers of that region. Consumers consuming energy below that decided value can be incentivized. There are various instruments of incentivization and we would not be going into those details. The idea here is that such incentive schemes would promote the efficient use of electricity and could drive users to be more careful about their energy use.

VII. Conclusion

Towards the goal of "Electricity for all" we developed an AR3 time series model called Energy Demand Predictor, based on the hourly electricity consumption data for TMY3 locations across USA made available by OpenEI. The insights derived from such a model can be used to drive inter-region trade of electricity and also to better plan the allocation of resources for generation of electricity and towards generation of electricity using cleaner, renewable sources. We also used K-Means Clustering algorithm to identify regions with similar electricity consumption patterns. This information can be used to help promote more efficient use of electricity by incentivizing the consumers to use lesser energy. By implementing these models we can move closer to achieving the goals of SDG 7.

Contributions

Data Collection:	Dinesh, Tushar, Sai Srujan
Proposal Presentation:	Sai Srujan, Dinesh, Tushar
Data Preprocessing:	Sai Srujan, Tushar
AutoRegression:	Sai Srujan, Dinesh
Clustering:	Tushar, Dinesh
Report:	Tushar, Sai Srujan, Dinesh
Presentation:	Dinesh, Sai Srujan, Tushar.

References

[1] https://sustainabledevelopment.un.org/sdg7.
[2] http://science.time.com/2013/03/26/smart-power-why-more-bytes-will-mean-fewer-and-cleaner-electrons/
[3] https://openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-states
[4] http://www.un.org/sustainabledevelopment/energy/
[5] https://www.eia.gov/state/rankings/
[6] http://ecdms.energy.ca.gov/elecbycounty.aspx
[7] https://wonder.cdc.gov/population-projections.html
[8] https://github.com/alexanderkell/Building-Energy-Analysis