ELECTRICITY PRICE PREDICTION
Using Random Forest Regressor

V.M.Bhuvanesh (19bcd7088)[1], M.Naga Venkata SaiSreeja(19bce7428)[2], Nuthana Kolla (19bcd7061)[3] ,Guru Theja(19bcd7034)[4]

[1,2,3,4]Department of Computer Science and

Engineering Vellore Institute of

Technology,Amaravathi,Andhra Pradesh

[1]manikanta.19bcd7088@vitap.ac.in, [2]saisreeja.19bce7428@vitap.ac.in, [3]chowdary.19bcd7061@vitap.ac.in, [4]gurutheja.19bcd7034@vitap.ac.in

Abstract:

Our project is to predict whether a electricity bill using machine learning approach. This helps the electricity producers(govt) to predict according to the factors affecting the production.So that they can assign optimal price based on the location of a electricity plant. The data set consists of 18 columns and 38014 instances. Here we will predict based on these instances : "Day", "Month", "ForecastWindProduction","SystemLoadEA","SMPEA", "ORKTemperature", "ORKWindspeed","CO2Intensity","ActualWindProduction","SystemLoadEP2".

By using random forest regressor we will predict the electricity bill.

Code Link:
https://colab.research.google.com/drive/1e4-RBfmP0uejTAsEP9shyA0yXgccUBSL?usp=sharing

Introduction:

Forecasting electricity prices is an important task in an energy utility and needed not only for proprietary trading but also for the optimization of power plant production schedules and other technical issues. A promising approach in power price forecasting is based on a recalculation of the order book using forecasts on market fundamentals like demand or renewable infeed.

Literature Review:

S.NO	Author	Year of Publication	Paper Title	Methodology	Algorithm	Advantages	Disadvantages	Remarks
1.	B. Vincenzo	2013	N. Linear regression models to forecast electricity consumption in Italy. Energy Sources Part B.	Developed simple and multiple regression models to forecast electricity consumption using annual electricity consumption data of residential and nonresidential consumers for a period of 37 years in Italy.	Regression	Useful to analyse the relationships among the input variables.	Not fit for real-world applications.	The results showed that the selected predictor variables are strongly correlated to the electricity consumption.
2.	F. Rodrigues	2014	The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal.	Introduced ANN method to forecast hourly and daily electricity consumption based in a feed forward ANN and Levenberg marquardt algorithm by using 18 months long comprehensive data set of 93 households. The input parameters fed to the model were apartment area, number of occupants, electrical appliance consumption and Boolean inputs as hourly metering system.	ANN	Reduce the computation time and increase the prediction accuracy.	Overfitting occurs and difficult to generalise.	Good performance results have been found in the hourly prediction of electricity consumption analysis for first 3 days of the 3rd week of one randomly selected household.
3.	W. J. Lee	2015	A hybrid dynamic and fuzzy time series model for mid-term power load forecasting, Electrical Power and Energy Systems.	Developed a hybrid model based on dynamic and fuzzy time series for mid-term prediction applied on actual load data from the Seoul metropolitan area and then compared the predicted result with the Koyck and ARIMA models.	Hybrid	Improve performance of the model	High model complexity.	indicated that the hybrid model produces a less forecasting error and can be used for mid-term forecasting along with observed air temperature data.
4.	Tae-Young Kim	2018	Predicting the Household Power Consumption Using CNN-LSTM Hybrid Networks. Springer Nature Switzerland AG LNCS.	Focused on predicting household power consumption using a CNN-LSTM hybrid network that linearly connects CNN and LSTM. The dataset used in the proposed model comprised of one generation of power consumption having 9 attributes among which the active power variable was used for power demand forecasting.	CNN	High accuracy in prediction results.	More suitable to analyse image data, Needs lot of training data.	The predictive results visually confirmed that an excellent prediction performance was achieved even in situation where periodicity was not observed.
5.	K. Yan	2019	A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households.	A hybrid deep learning model was generated by combining an ensembled LSTM neural network with the SWT technique to improve the accuracy of forecasting.	LSTM	High accuracy	Need more time to train the model, Not suitable for short-term prediction.	The proposed SWTLSTM framework produces the most accurate forecasting results with time steps including 5, 10, 20 and 30 minutes.
6.	B. Vincenzo	2013	N. Linear regression models to forecast electricity consumption in Italy. Energy Sources Part B.	Developed simple and multiple regression models to forecast electricity consumption using annual electricity consumption data of residential and nonresidential consumers for a period of 37 years in Italy.	Regression	Useful to analyse the relationships among the input variables.	Not fit for real-world applications.	The results showed that the selected predictor variables are strongly correlated to the electricity consumption.

Experimental Framework:

Based on the ideas drawn from the mentioned papers we had improved our model by using random forest regressor which has good training period among the other proposed algorithms and predict the electricity bill.

Proposed Methodology:

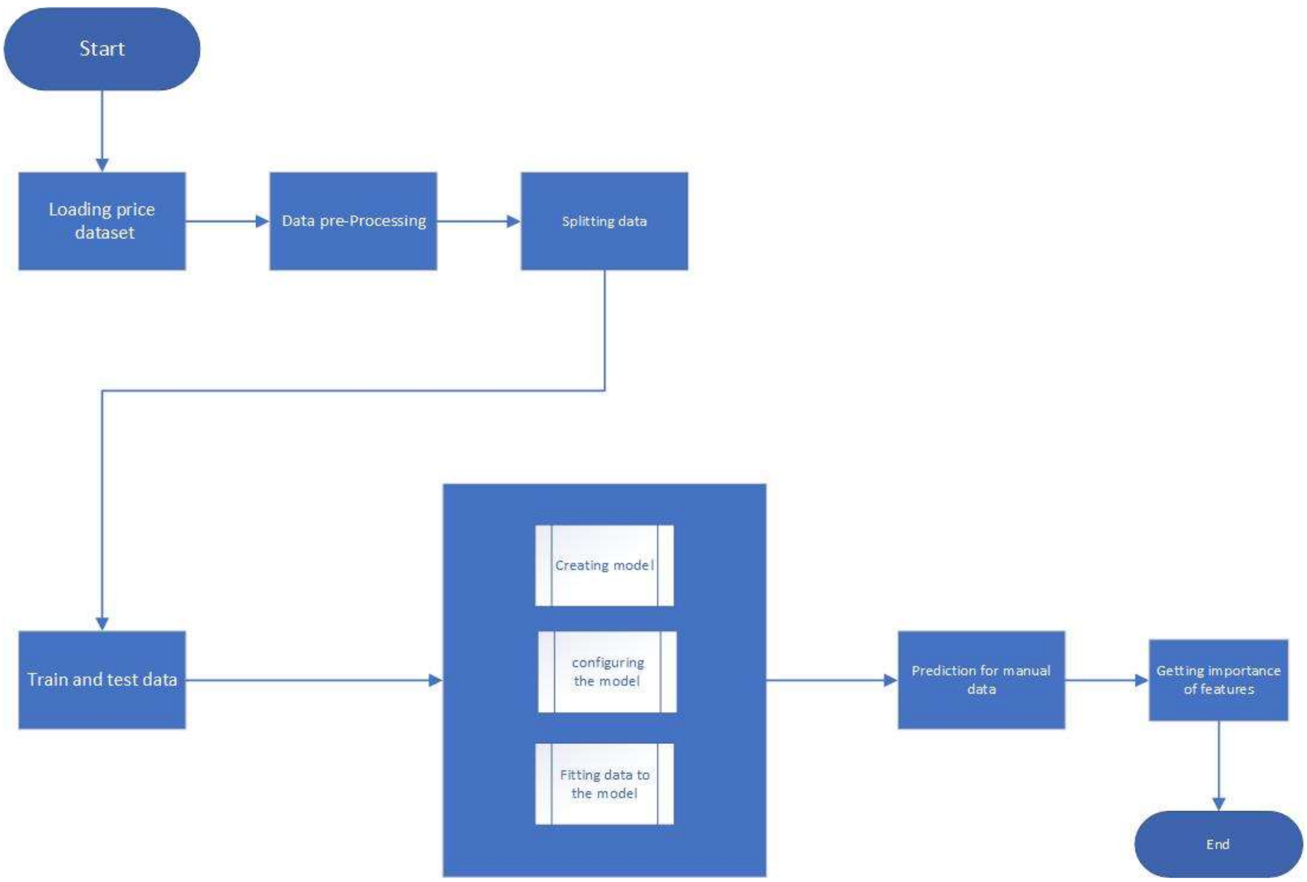
Algorithm:

We worked on electricity bill dataset obtained from kaggle; The data set consists of 18 columns and 38014 instances. Here we will predict based on these instances : "Day", "Month", "ForecastWindProduction", "SystemLoadEA","SMPEA", "ORKTemperature", "ORKWindspeed", "CO2Intensity", "ActualWindProduction", "SystemLoadEP2".

Step-1: Importing the packages pandas , numpy, Pyplot, seaborn, RandomForestRegressor and initializing the data.
Step-2: The dataset is first cleansed and processed using pre-processing techniques like Data Integration, Data transformation, Data reduction, and Data cleaning using pandas tool.
Step-3: Encoding.we have to change that values into number format instead of names.
Step-4: Splitting the data to test and train. Based on the split criterion, the cleansed data is split into 80% training and 20% test.
Step-5: Creating a model and added estimators.
Step-6: Fitting the trained data in the model.
Step-7: Prediction of the manual data
Step-8: Finding importance of each feature in dataset while building model and plotting the model using pyplot.

Implementation Details:

Flow chart:

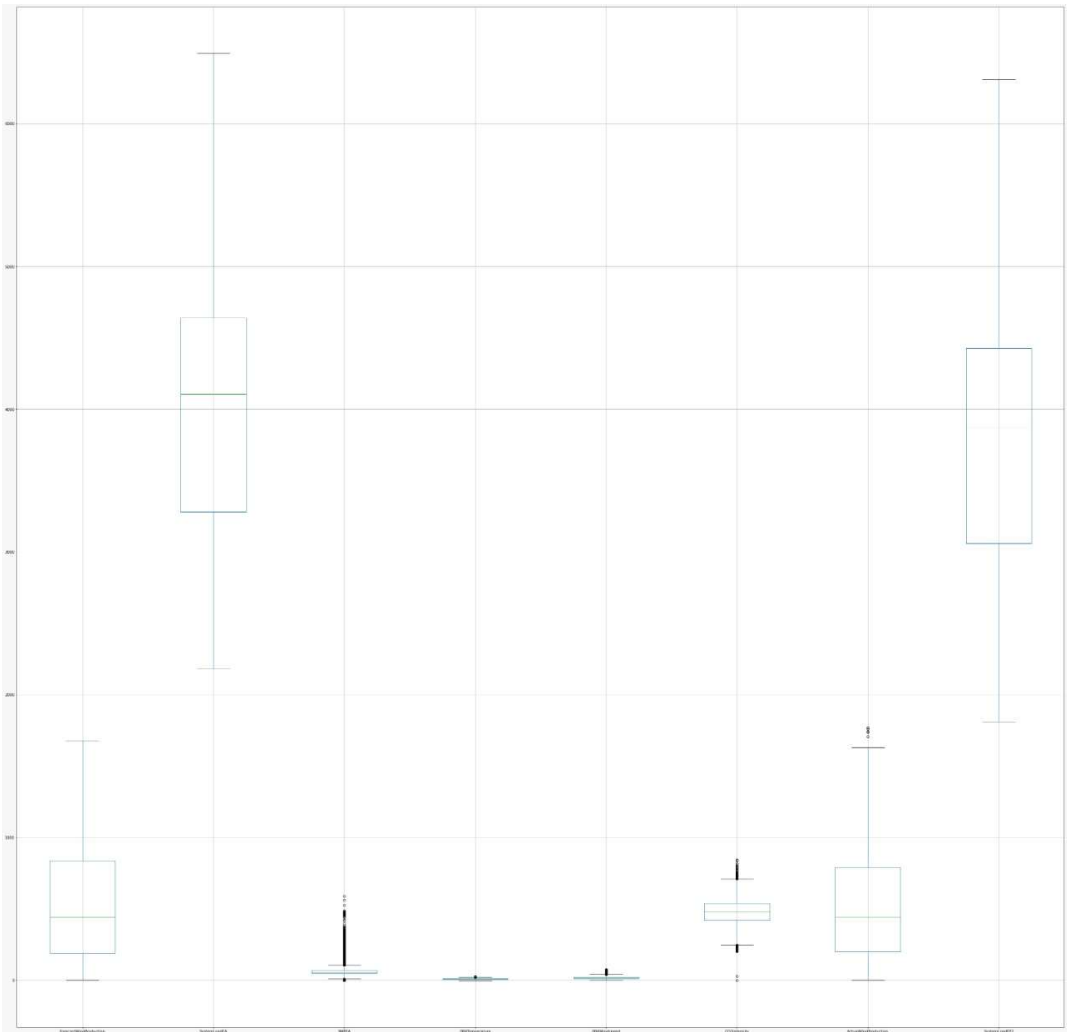


Dataset description:

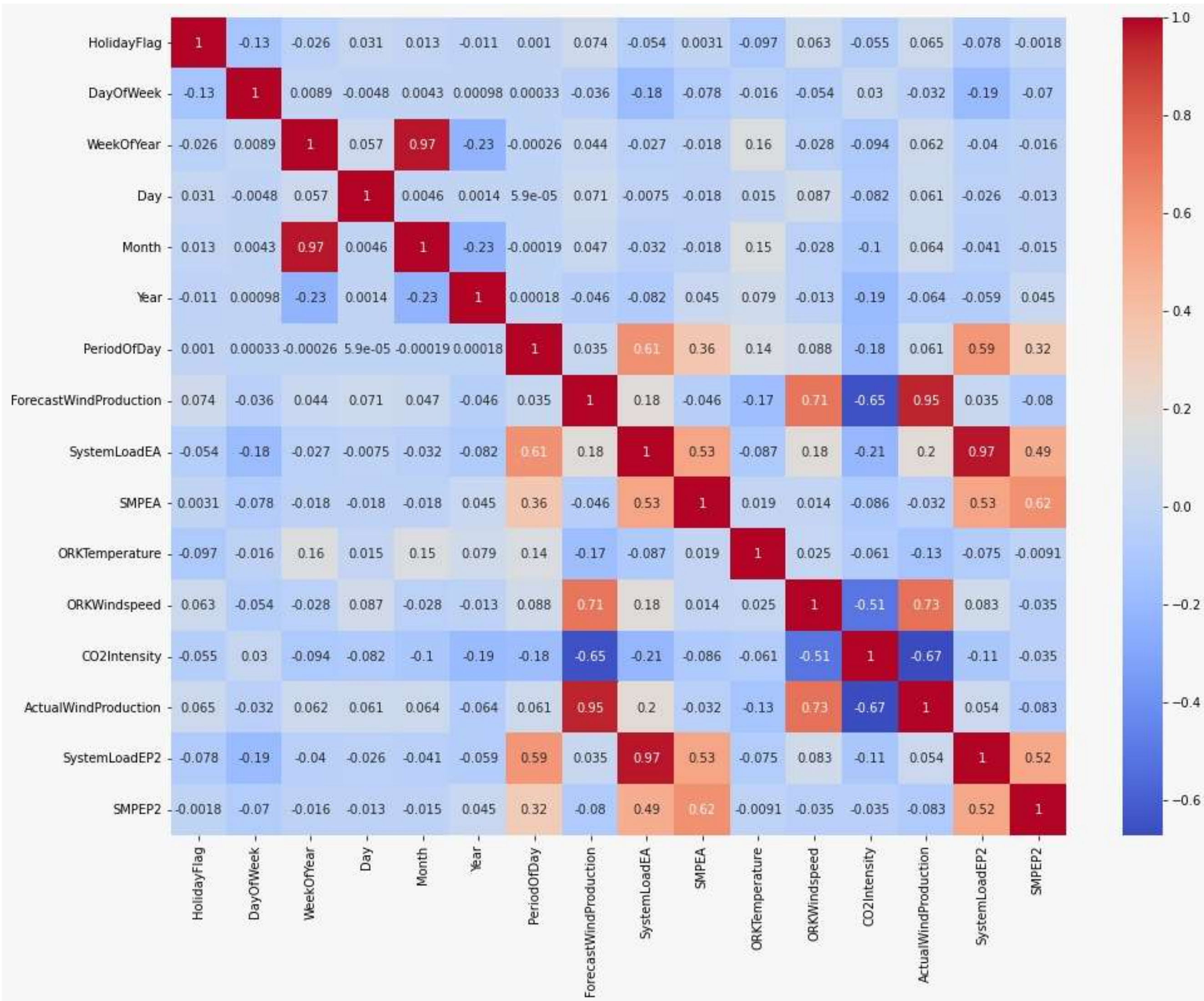
The data set consists of 18 columns and 38014 instances. Here we will predict based on these instances : "Day", "Month", "ForecastWindProduction", "SystemLoadEA","SMPEA", "ORKTemperature", "ORKWindspeed", "CO2Intensity", "ActualWindProduction", "SystemLoadEP2".

1. DateTime: Date and time of the record
2. Holiday: contains the name of the holiday if the day is a national holiday
3. HolidayFlag: contains 1 if it’s a bank holiday otherwise 0
4. DayOfWeek: contains values between 0-6 where 0 is Monday
5. WeekOfYear: week of the year
6. Day: Day of the date
7. Month: Month of the date
8. Year: Year of the date
9. PeriodOfDay: half-hour period of the day
10. ForcastWindProduction: forecasted wind production
11. SystemLoadEA forecasted national load
12. SMPEA: forecasted price
13. ORKTemperature: actual temperature measured
14. ORKWindspeed: actual windspeed measured
15. CO2Intensity: actual C02 intensity for the electricity produced
16. ActualWindProduction: actual wind energy production
17. SystemLoadEP2: actual national system load
18. SMPEP2: the actual price of the electricity consumed (labels or values to be predicted)

Box Plot:



Heat Map:

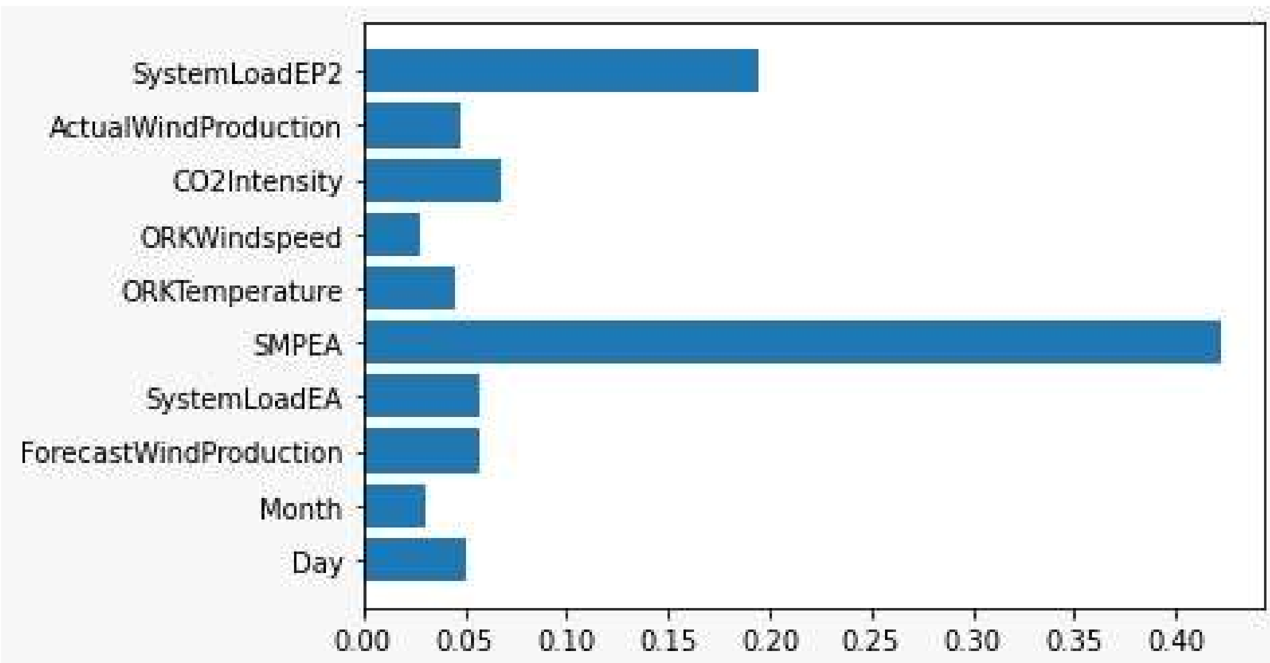


System Specification:

Software: Google Colaboratory using python.

Results:

Importance of Features:



Conclusion:

Here by using the random forest regressor which we had created using train data and predicted the price for manual data. We also found the importance of each feature while building the model.

References:

- Rodrigues F., Carderia C., Calado J. M. F. The daily and hourly energy consumption and load forecasting using artificial neural network method: a case study using a set of 93 households in Portugal. Energy Procedia. 2014; 62 220-229.
- Din G. M. Ud., Marnerides A. K. Short Term Power Load Forecasting Using Deep Neural Networks, International Conference on Computing. Networking and Communications (ICNC): Green Computing, Networking and Communications, 2017; 978-1-5090- 4588-4/17
- Kim T.Yand Cho S.B. Predicting the Household Power Consumption Using CNN-LSTM Hybrid Networks. Springer Nature Switzerland AG LNCS. 2018; 11314, 481–490. https://doi.org/10.1007/978- 3-030-03493-1_50
- Yan K., Li W., Ji Z., Qi M. and Du Y. A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households. IEEE Access, 2019; vol. 7, pp. 157633-157642, doi:10.1109/ACCESS.2019.2949065.
- Vincenzo B., Oronzio M., Sergio N. Linear regression models to forecast electricity consumption in Italy. Energy Sources Part B. 2013; 8:86–93
- Lee W. J., Hong J. A hybrid dynamic and fuzzy time series model for mid-term power load forecasting, Electrical Power and Energy Systems, 2015; vol-64 pp. 1057-1062