FORECASTING OF ENERGY CONSUMPTION IN AN INDUSTRIAL FIRM USING STATISTICAL AND MACHINE LEARNING MODELS

H.W.A. Suranga, N. Karunathunga, S.S.N. Perera and S.A.K.P. De Silva

Research & Development Centre for Mathematical Modeling,
Department of Mathematics, Faculty of Science, University of Colombo,
Sri Lanka

amilasl85@gmail.com

ABSTRACT: *Manufacturing industry is a key contributor to the economy. These manufacturing firms are high energy consumers and currently facing challenges in their energy demand due to scarcity, cost, subsequent environmental impact on heavy consumption, social and consumer pressure for carbon footprint of consumer goods and industry specific global compliances and regulations. Then an accurate short-term forecast of energy consumption is a must to maintain optimal supply and usage while minimizing negative concerns against higher consumption. This will enable a smooth operation with a minimum risk on energy related inventories. Electricity, biomass, furnace oil and diesel are the main energy sources in industry. In current context, studies are conducting on electricity consumption in industries, buildings, and residential boundaries. However, predictions on total energy consumption on manufacturing firms are not frequently studied. Both conventional statistical models and deep learning models are widely used for this task. In this exercise, a manufacturing entity has been selected to predict its forthcoming month's energy consumption using historical energy consumption and manufacturing figures. Existing literature related to energy consumption prediction suggested, four models to predict energy consumptions. Fb prophet, vector auto regression, Long Sort Term Memory (LSTM) and Auto Regressive Integrated Moving Average (ARIMA) models were taken to make predictions for energy consumption in terms of electricity and biomass. Performance of each models were compared using root mean squared error, mean absolute error and mean percentage absolute error. LSTM outperformed over all other models for both types of energies and the same model was selected to predict monthly energy consumption of selected entity.*

Keywords: Energy forecasting, Electricity, Biomass, Machine learning, Deep learning, Vector auto regression, LSTM and ARIMA

1. INTRODUCTION

Overconsumption of resources is a main environmental issue in current era. According to European commission circular economy action plan, by 2050, the globe will be consuming three times resources than required. Further they expect a doubled consumption of fossil fuels, minerals and metals, and biomass during next forty years. All these consumptions are accountable of global warming, climate changes and waste accumulation and subsequent pollution over the time. Industrial sector is the largest energy consumer and now focusing on advanced energy management practices to optimize their consumptions. UN sustainable goals, cleaner production concepts, imposed regulations and taxes are widely used to regulate the excess energy usage and wastage during industrial practices.

Addition to the above controls, firms are implementing good energy management systems to lean their consumptions while fulfilling their demands, there by optimize the energy consumption accurately in order to meet those sustainable implementations and cost benefits minimizing wastes.

Consumption monitoring is one of good practice to identify the consumption patterns of a firm. Such continuous monitoring system generate a time series data on energy consumption and this information could be effectively used to identify the behavior of energy consumption. Such information is important in statistical and mathematical model development and thereby predict future consumption figures with given other inputs.

There are some published articles on building energy consumption prediction and country wise electricity prediction based on historical consumption patterns. However the studies conducted on industrial energy consumption is very limited and most of them are focusing only on electricity consumption.

In this paper, I focused on historical energy consumption of an energy-oriented manufacturing firm which uses both electricity and biomass as the energy sources for their operations and make 30-day prediction on both energies separately. In addition to each consumption data, I have used daily production figures of their two production sectors and daily rainfall data. Based on the literature survey conducted on the topic, I have used two statistical prediction models and two deep learning prediction models. The best models are selected from the figures of mean absolute error, root mean squared error and mean percentage absolute error.

2. METHODOLOGY/ METHODS/ DEVELOPMENT OF MODELS

Data collection and preparing

Daily electricity consumption in kWh and daily biomass consumption in kg are recorded by the engineering division of the selected entity. Daily energy consumption for energy was taken separately from those records. In addition to that, daily production figure pertaining to two manufacturing sections were extracted.

In order to avoid unpredictable issues due to cleanliness of raw data, dataset was cleaned by removing outliers and filling not available values by last 7-day averages. Datapoints which lies on between (Q1-1.5*IQR) and (Q3+1.5*IQR) was taken as the range for accepted data. Due to industry norm on higher firewood consumption on rainy days, daily rainfall data was collected from department of meteorology with respect to nearest collection point. Then the data was normalized to make the all the inputs carry equal weights. Normalization reduces the range of all data between 0 and 1.

Figure 01 : Graphical illustration of sample raw dataset

Model selection and literature search recommendations
Detailed literature search was conducted to identify the research conducted on energy predictions. The published articles and research papers suggested models in a wide range. Then the most appropriate and most related research to industrial energy and electricity consumption prediction was identified.

Fadhilah et al. (2009) conducted analysis and prediction for Malaysian national electricity company on their load forecast using ARIMA, Naïve method, Seasonal holt-winters method, regression with ARMA errors methods. Their outcome was AR (2) model as the best model for electricity consumption forecasting.

Miao (2015) discussed an ARIMA model to forecast total energy consumption in China using historical consumption figures. Miao has successfully developed ARIMA (1,1,1) model for the total energy consumption (in mega ton coal equivalent) and the relative error is less than 2% for five consecutive years.

A study conducted by Ribeiro et al. (2020) on Energy forecasting of a manufacturing firm to predict energy consumption using different machine learning and statistical models. In this exercise, ARIMA, internal calculation method, recurrent neural network, long short-term memory, support vector regression and random forest models has been evaluated. The list contains both statistical and machine learning models. Model performance and parameter setting has been measured using Diebold-Mariano test, mean absolute percentage error (MAPE), root mean squared error (RMSE) and mean absolute error (MAE). Based on outcome machine learning methods has performed well compared to the statistical methods. However, ARIMA model also performed well, and predictions were accurate into significant level.

Vector auto regression is another successful model used to forecast electricity consumption by Guefano et al. (2021). Further vector autoregression is suitable predictive tool when multiple time series influence on each other.
An Indian research on monthly electricity demand forecast, conducted by Chaturvedi et al.(2022), has revealed that Fb Prophet has better performance over LSTM recurrent neural network and SARIMA models.

ISBN: 978-955-627-013-6

Composition and behavior of energy consumption of above research are interconnected to the qualities of dataset selected. Considering the similarities and conclusions extracted on literature search, following four models were selected to predict energy consumption using the dataset.

1. ARIMA
2. Vector auto regression
3. FB Prophet
4. LSTM recurrent neural network

Even though the studies on biomass type energies are infrequent, characteristics of consumptions were similar to that of electricity consumption. Then both energy series were tested using selected models.

Quantitative Approach

Three evaluation calculations were used in evaluation the best for energy prediction over selected models. MAE used to quantify the performance of each model due to its ease of use and dependability of data scale. Simply MAE is the average of absolute difference on two Y series of same X series. Here it measures the absolute deviation of each individual pertaining to same X and Finally presented as the average of each. Pi and Ri are the predicted value and real value while n equals to the number of observations.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Pi - Ri| \tag{1}$$

Additionally, RMSE and MAPE was calculated to selected model interpretation to provide a better clarity on model performance. MAPE directly interpret the accuracy of the model and can be used to explain the accuracy of the prediction as a percentage.

RMSE is the square root of mean squared error. In other words, quadratic mean of differences between predicted value and real values of the model outcome. Unlike other evaluation metrics RMSE is more sensitive to outliers as it takes the squared value of the difference of predicted and real values. Latter two values were calculated to interpret the accuracy of the selected model.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Pi-Ri}{Ri}\right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Pi - Ri)^2} \tag{3}$$

ARIMA model

ARIMA model is a statistical model widely used in demand prediction and stock price prediction. Here, the time series required to be stationary in order to make the prediction. Stationarity of the series to be obtained by differentiation the series and once the series is stationary, predictions could be generated.

“pm. autoarima” function was used to determine the best fit for ARIMA (p, d, q) for electricity and biomass. (p-order of auto regressive model, d-degree of differentiation and q- order of moving average model) the best model was evaluated considering minimum AIC (Akaike Information Criterion). Once the best fitting model is identified, prediction was made. Mean absolute value of each prediction was taken to compare the ARIMA models with other models.

Vector Auto regression model

Vector autoregression is a statistical method used to capture the relationship of multiple timeseries. After loading the data series, their relationship was tested using Granger's causality test. This is hypothesis test that conclude whether one time series useful in predicting other series. Null hypotheses assume that previous lagged value of series-1 does not explain the series-2. When the p value obtained is lesser than the considered significance level null hypothesis can be rejected. Then cointegration test was conducted to check whether the selected data series are statistically significant. After the data series were split into train data and test data. Augmented dicky fuller test was conducted to check the stationarity of the time series and relevant differentiation has to be conduct make the series stationary. Once the stationarity obtained the lag of the vector auto regression was determined considering the minimum AIC value. Then the model was trained using train data set validated with test data. Once the model was verified, predictions were drawn. MAE was calculated and kept for comparison purpose.

FB Prophet.

FB Prophet is developed by Facebook Inc, an open-sourced database used for time series data analyze purpose. Prophet has the capability identify the seasonal trends, holiday effects and works with daily periodic data and compatible with date series having missing data as well. The same data set was used and split in to two for training and validation. Once the validation is conducted, trained model was used to generate predictions.MAE was calculated to comparison purpose.

LSTM Recurrent neural network

LSTM, Long Short-Term Memory is a type of recurrent neural network which is capable in identifying long term patterns and dependencies of data series for prediction purposes and the application of LSTM expands in machine- translation, speech recognition and handwriting recognition etc.

LSTMs have memory cells referred as memory units. Data can be filled into these cells using a gate system which consists forgot gate, input gate and output gate. Data input and previous lag output is multiplied by a weight matrix and the result is converted to binary output and the initial date would get forgotten. Then the information is adjusted using mathematical functions (Sigmoid and tanh). Finally, useful information is extracted, and the individual data memories will be deleted. The process is similar reading a book where we do not remember each and every word, but the entire story is absorbed finally. Using that information, the model identifies the characteristics of the data series and the trained model can be used to draw predictions.

As usual same data series is separate into train and test data. Train data set is loaded and designed as 7 data lags. As example first 7 inputs gives the 8th datum as output. Similarly, next 7 inputs provide 9th datum as the output. Model was optimized using the Adam optimizer. Adam optimizer is an algorithm that can be used to update weights according to the iterations during training process. Trained model is then used to make predictions. Mean absolute value of the predicted data and real data was taken for comparison purpose.

The best performing model was selected by considering the mean absolute error obtain energy d for each model. Then the model with least MAE value was selected to predict energy consumption of the firm. The last 30 figures of the data set were separated for comparison purpose and predictions were carried out for the same last

30 figures. In addition to the MAE value, MAPE and RMSE values were calculated to explain the accuracy level of the selected model

3. RESULTS AND DISCUSSION

Below table presents the mean absolute error value obtained for both electricity and biomass predictions during initial model selection conducted for ARIMA, Vector auto regression, FB prophet and LSTM models.

Table 01: MAE value comparison for evaluated models

	MAE for Electricity consumption prediction (kWh)	MAE for Biomass consumption prediction (kg)
ARIMA	16025.4	7891.7
Vector auto regression	4177.6	57622.6
FB Prophet	4916.8	27763.9
LSTM	2545.5	5602.1

In the ARIMA, model (p, d, and q) parameters have to be defined to identify the best model. ARIMA (1,1,0) was the best fitting model for the electricity consumption figures and ARIMA (1,2,1) was the best fitting for biomass consumption figures.

Based on the MAE values obtained, ARIMA (1,2,1) and LSTM outperformed over VAR and FB Prophet for biomass consumption prediction. VAR has given the worst prediction for biomass consumption prediction while LSTM is the best model. However, vector auto regression and LSTM are the best candidates for Electricity prediction. ARIMA is the worst prediction model with higher MAE value while LSTM is the most suitable model for Electricity energy prediction purpose. Then LSTM is the most convenient model for both biomass and electricity prediction as per the used dataset.

Generally, time series models are difficult to use in predictions using statistical methods. However according to above plots and MAE data, ARIMA and Vector auto regression model's performance is significant compared to others. Those regression methods require additional information other than the targeted time series in order to make predictions. Therefore, the accuracy and their correlation play a critical role on the accuracy of the prediction. The basic issues like, degree of compatibility of developed model and parameters, noise of collected data, negative behavior of identified influencing factors, unidentified factors that have greater influence on time series are the main concerns that affect on the accuracy of the model.

However even with those challenges ARIMA (1,2,1) for biomass and VAR for electricity are definitely good alternative candidates. The importance of such statistical prediction model is that the ability to feed known inputs to predict the future energy demand.

According to literature search in most circumstances machine learning models outperform statistical models in predictions. With this result obtained for the data series we can identify both statistical and machine learning models perform equally better, and there could be fluctuations or errors based on the characteristics of the data series. Otherwise, VAR and ARIMA are good candidates according to the MAE values they obtained for model comparison.

Figure 05: Monthly energy prediction for electricity and Biomass using LSTM method

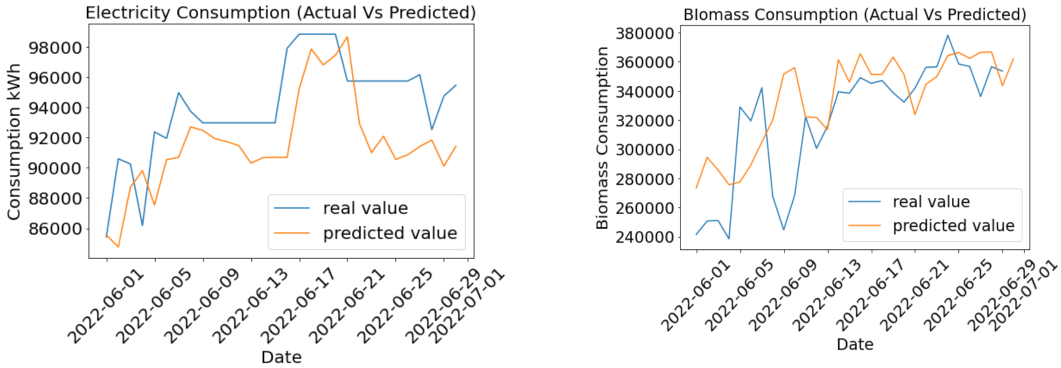

Table 02: RMSE, MAE and MAPE values obtained for Electricity and Biomass prediction using LSTM model

	Electricity Prediction	Biomass Prediction
RMSE	2545.5	6485.7
MAE	2048.2	5602.1
MAPE	0.022	0.025

As we can see in the outcome all RMSE, MAE and MAPE values confirm that the closer prediction obtained using LSTM method. Predicted curve shows the similar fluctuations with that of real values. Finally, machine learning LSTM model outperformed all other selected models. The LSTM model do not require supporting information other than same time series to make predictions, but the time taken for the training and the data requirement to train the model is comparatively higher than statistical models. The selected models are performed well in short term duration predictions.

4. CONCLUSION

Energy is critical input in manufacturing sector. With the current economic stress and the limitations and constraints on supply chain, manufacturing firms get squeezed in managing their daily energy demand. Volatility of prices of alternative energy sources like fuel oil and diesel minimize the space on backup energy solutions. So, with those considerations and environmental concerns including sustainability initiative, green manufacturing concepts organizations shall adopt a good energy management system to optimize their energy usage. In this exercise I examine the opportunity to predict monthly energy demand of a manufacturing firm. This manufacturing facility uses electricity and biomass as their main energy sources.

Using the historical energy consumption and manufacturing data, I developed four models for monthly energy consumption prediction. Among the tested model's LSTM machine learning model outperformed all other three models in predicting 30-day energy consumption for both electricity and biomass.

Then I would like to conclude that LSTM neural network machine learning method is the best model to predict the energy demand of next 30-days, out of selected four models. The model needs to be keep updating with latest data for the latest and most accurate outcome. Further performance of ARIMA (1,2,1) is a good candidate for biomass energy prediction with daily production figures as inputs. With that concern and the level of accuracy, further research could be conducted to identify other relevant input information for ARIMA model improvement. As an example, individual production line production figures, machine downtimes, changeover times, moisture content of biomass and the enthalpic values of each biomass consignment might be critical in accuracy of such ARIMA model. Similarly, VAR model can further improve by installing additional electricity measuring tools to identify exact electricity consumption taken for the production process and the remaining consumption for building and office consumption separately. Information related to product changeovers, product specifications, raw material consumptions might be good inputs for improving the accuracy of VAR model in electricity consumption prediction. Internet of things (IOTs) and dedicated reading devices could be coupled to the prediction model to feed real-time data in order to keep the model at the most update state and draw the real time predictions always.

REFERENCES

Razak, F. A., Shitan, M., Hashim, A. H., & Abidin, I. Z. (2009). Load forecasting using time series models. Jurnal Kejuruteraan, 21, 53-62.

Miao, J. (2015, August). The energy consumption forecasting in China based on ARIMA model. In International Conference on Materials Engineering and Information Technology Applications (MEITA 2015) (pp. 192-196). Atlantis Press.

Ribeiro, A. M. N., do Carmo, P. R. X., Rodrigues, I. R., Sadok, D., Lynn, T., & Endo, P. T. (2020). Short-term firm-level energy-consumption forecasting for energy-intensive manufacturing: a comparison of machine learning and deep learning models. Algorithms, 13(11), 274.

Guefano, S., Tamba, J. G., Azong, T. E. W., & Monkam, L. (2021). Methodology for forecasting electricity consumption by Grey and Vector autoregressive models. MethodsX, 8, 101296.

Chaturvedi, S., Rajasekar, E., Natarajan, S., & McCullen, N. (2022). A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India. Energy Policy, 168, 113097.