BIG DATA ANALYTICS FOR A NOVEL ELECTRICAL LOAD FORECASTING TECHNIQUE

Usman Ali, Abdul Rauf, Umer Iqbal, Ijaz Ali Shoukat, Abu ul Hassan

Department of Computing, RIPHAH International Univeristy Faisalabad
r.usmaanali@gmail.com, umeriqbal15@gmail.com,
abdulrauf2000.pk@gmail.com, ijaz342@yahoo.com,
Hassan.superior1@gmail.com.
Pakistan

Abstract: In this paper focused on large size data of electrical load, which is called energy big data. With the help of this extremely large size of data, we performed short term electrical load forecasting. Conventional techniques are not very supportive to handle big data for forecasting. For accurate load forecasting in this paper using a framework that is called feature selection, extraction, and classification (FSEC). For feature selection used Random Forest (RF) and XGBoost (XGB). Feature extraction performed with Recursive Feature Eliminator (RFE) and for classification purpose, used Enhanced Support Vector Machine (ESVM). To show the improved performance of our Enhanced SVM, we made a comparison with some conventional techniques. Results prove that our proposed technique performed better than all.

Key words: Big data, Short-term, Electrical load forecasting, Support Vector Machine.

1. INTRODUCTION

Smart Grid is a modern way to manage efficiently the distribution, generation, and consumption of electricity [1]. It gives facility to consumers in the economical, secure and reliable manner. Management of Demand Side (DSM) is the part of Smart Grid. Consumers can manage and control their demand with the support of DSM techniques and can shift high voltage machines from more consumption hours to low consumption hours described in [2].

One of the most important tasks in the electrical grid is a prediction of the electrical load which is a need in the future that is called electrical load forecasting. There are commonly 3 types of electrical load forecasting [3], Medium Term, Long Term, and Short Term Load Forecasting (STLF). In this paper, we performed STLF.

STLF consists of one hour to one month prediction; MTLF consists of 1 month to 1 year and long term make a prediction from 1 year to several years.

The data of the real world is very large [4]. The data of SG is reviewed in detailed [5]. A huge size of data has a lot of information which helps the utilities to perform analysis that leads to making further improvement in planning of operation and management in the market. Conventional techniques are not very supportive to handle the big data, because the big data is very multifaceted and have a huge number of features. The big data analytics is very helpful to in mining of concealed patterns, trends of the market and additional important information.

In this paper, a framework is used to classify the big data, which is called Feature Selection, Extraction, and Classification (FSEC). To make this technique novel we used extremely large data which consist of 8 years, and perform feature selection with the combination of RF and XGB. RFE used to eliminate irrelevant features. Classification performed with combination of SVM and Random Search.

Next sections consist of related work, proposed model, result and conclusion.

2. RELATED WORK

There are many forecasting techniques which have discussed in literature. According to literature, there are 3 categories that used mostly in load forecasting: Classical, Artificial Intelligent and Data-driven. The classical method has NB, RF, and ARIME, etc. Artificial intelligent have ANN, PSO, and DNN, etc. methods. Data-driven methods are mostly considered the past data that predicted the desirable outcomes

In [7]-[9], DNN gave better results of forecasting, but SNN performed worst. In [10] authors used ReLU and RBM for forecasting. For tanning and processing, RBM used and through RelU performed forecasting. Feature selection performed with KPCA, and SVM is used for forecasting in [11]. Hybrid of CNN and LSTM used in [12], it helps to enhance the accuracy rate of forecasting. In [13], the authors used Gated-Recurrent-Units (GRU) technique for forecasting.

Through feature engineering, new features created using domain knowledge. The new features help machine learning algorithms for prediction. Feature engineering is part of the classifier. There are 2 important operations performed in feature engineering; extraction and selection. Many authors have taken help from the feature engineering technique to perform forecasting. In [14] - [16], many authors have been discussed about feature engineering existing techniques.

3. PROPOSED MODEL

The model that we have used is shown in fig. 1. Used model [6] basically has 3 steps.

1. Preprocessing of data,
2. Feature selection and extraction (FSE),
3. Forecasting of the load.

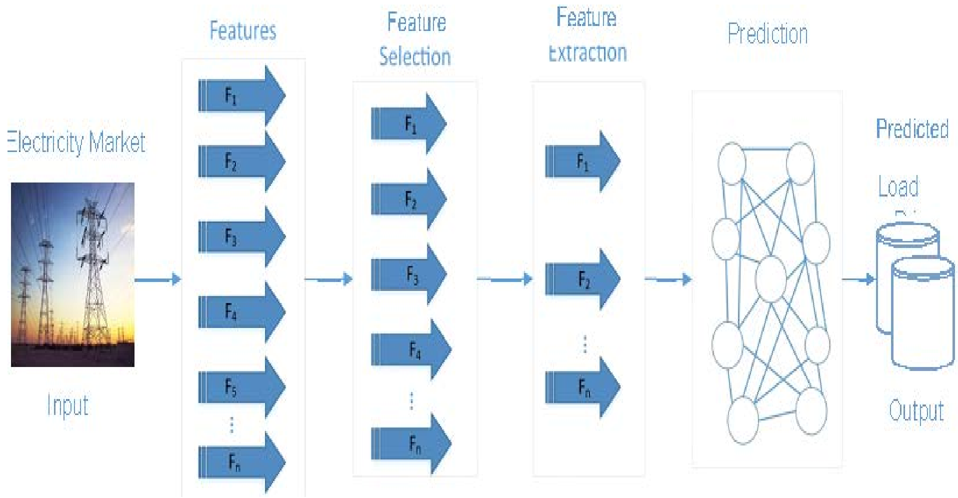

Fig. 1. Proposed Model of Electrical load forecasting

3.1. Pre-processing of data

In this model, we take data from ISO-NE that is 8 years of data, January 2011 to January 2018. We divided the data into testing and training feature. No. of 2191 features used for training and 731 features used for testing.

3.2. Feature Selection and Extraction (FSE)

We used the FSE model to select the most relevant values and eliminated non-important features. For this purpose, we used a hybrid of XGB and RF for feature selection. RFE used to remove non-important features. Feature extraction is a process to remove non-relevant features from selected features. This process generates new data on the basis of newly selected data to give accurate results than the original data [17]. We set the threshold values: XG_LowThreashold=0.8; RF_LowThreashold=0.7; RFE_threashold= True; to select the most relevant features and eliminated the non-important features.

3.3. Forecasting of load

Forecasting is a process to classify the data and make a prediction with the help of classification techniques. In this article, The Enhanced support vector machine (ESVM) used for classification. Manually tuning the parameters of SVM is a big problem. So, Random Search (RS) algorithm used to tune the parameters of SVM. In the ESVM kernel name, Radial basis function (RBF) used. C values, Gamma values, and other parameters tuned dynamically with RS and the number of iterations performed in ESVM is 15. Results have been evaluated with MAE, RMSE, MSE, and MAPE.

4. RESULTS AND DISCUSSIONS

Spyder (Python 3 packages) provided by Anaconda, is used for simulation. The system has 12GB RAM and Core i5 6th generation proc. is used for simulation.

4.1. Feature Selection and Extraction

The process of feature selection is performed with XGB and RF. The result discussed in the form of graphs. Combination of XGB and RF selected most important features relevant features. Fig 2 and 3 are showing the importance of feature selection using XGB and RF.

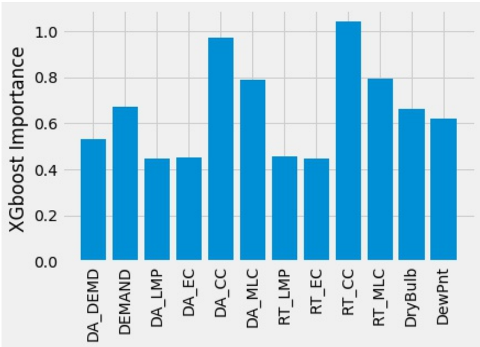

Fig. 2. Importance of feature selection using XGB

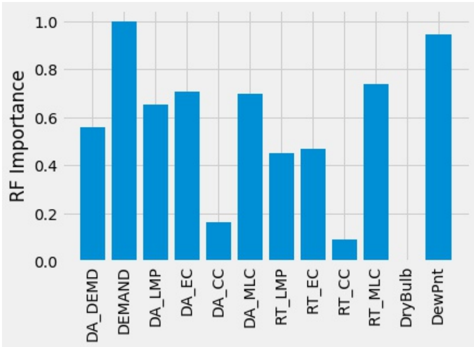

Fig. 3. Importance of feature selection using RF

RFE removed the non-important feature. After feature selection and extraction, a number of features that has been selected or rejected shown in table 1. 10 main features selected and 2 rejected from 12 total features.

Table 1. Number of features selected and rejected

Selected Features	1:- DA_DEMD, 2:- DEMAND, 3:- DA_EC, 4:- DA_CC 5:- DA_MLC 6:- RT_LMP 7:- RT_EC 8:- RT_CC 9:- RT_MLC 10:- DewPnt
Rejected Features	1:- DA_LMP 2:- Dry Bulb

4.2. Forecasted Results

Normalized data of 8 years displayed in Fig 4. It shows different variations between dissimilar days. After splitting the data in testing-data and training-data, the data is sent to the forecasting engine for future electrical load forecasting.

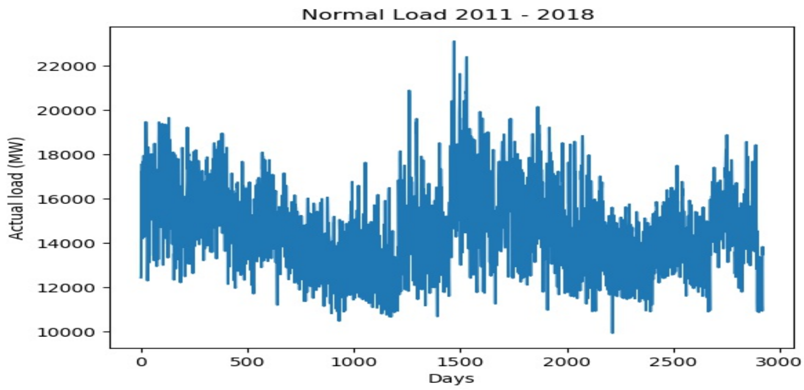

Fig. 4. Normalized data

Different months of data' variation is presented in Fig 5.

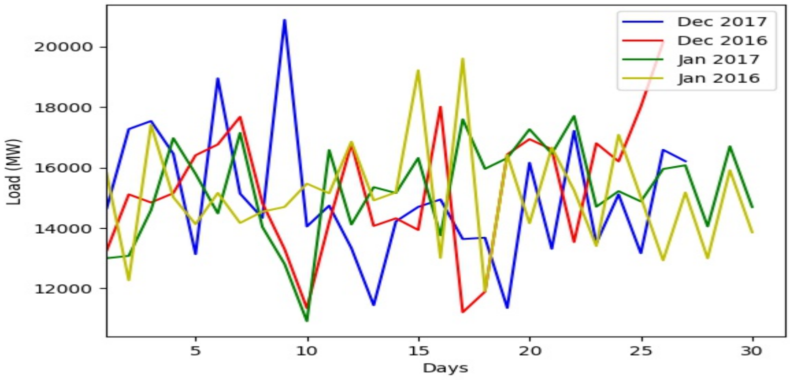

Fig. 5. Variation of the load

In Fig 5, there is very little change in similar months and have a big difference in the pattern of difference months. Fig 6 and Fig 7 is shown one week and one month forecasting load.

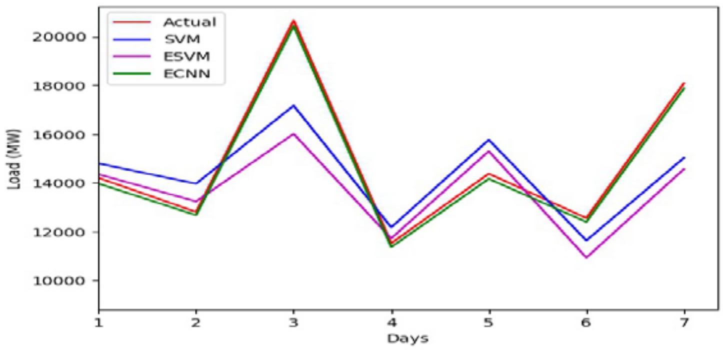

Fig. 6. 1 Week Forecasting

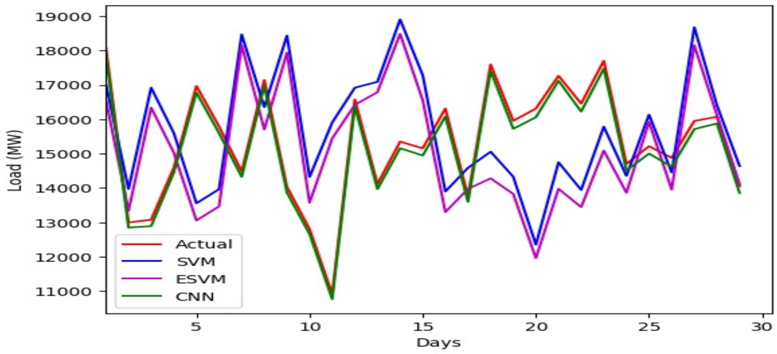

Fig 7. 1 month Forecasting

According to four evaluated metrics the results shown in fig 8.

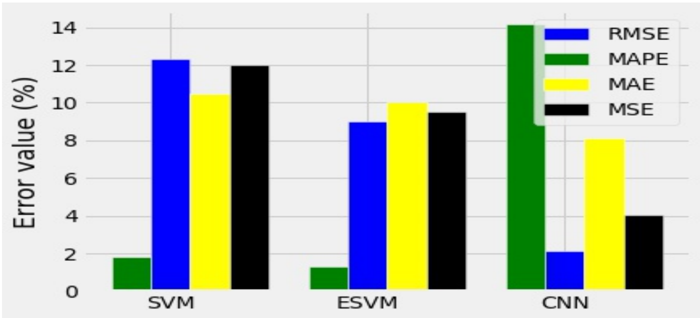

Fig 8. Error rates of forecasting techniques

Fig 8 is shown, our ESVM performed better results than traditional SVM and traditional CNN. Further, in table 2 and table 3 shows the comparison of ESVM with SVM and CNN according to error rates and accuracy rates.

Table 2. Error rates

Techniques	MSE	MAE	RMSE	MAPE
SVM	12	10.50	12.30	1.79
CNN	4.03	8.11	2.10	14.15
ESVM	9.50	10.0	9	1.2

Table 3. Accuracy rates

Techniques	MSE	MAE	RMSE	MAPE
SVM	88	89.5	87.7	98.21
CNN	95.97	91.89	97.9	85.85
ESVM	90.5	90	91	98.8

ESVM got the minimum accuracy rate 1.2 and highest accuracy rate 98.8 through MAPE.

5. CONCLUSION

In this paper, we discussed about electrical load forecasting and performed one week & one month electrical load forecasting. For forecasting purpose, we used feature selection and extraction framework to select the most relevant features, and classification to predict electricity load that required in the future. XGB, RF performed well to select relevant features, and RFE eliminated redundant values very accurately. It is a difficult task to tune the parameters manually. So, we used the RS algorithm for tuning the parameters of SVM. Our ESVM performed well and gave the lowest error rates.

REFERENCES

[1] S. Mujeeb, N. Javaid, M. Ilahi, Z. Wadud, F. Ishmanov, and M. K. Afzal. Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities. *Sustain.*, vol. 11, no. 4, 2019, pp. 1–29.

[2] P. Samadi, V. W. S. Wong, and R. Schober. Load Scheduling and Power Trading in Systems with High Penetration of Renewable Energy Resources. *IEEE Trans. Smart Grid*, vol. 7, no. 4, 2016, pp. 1802–1812.

[3] M. Zahid et al., .Electricity Price and Load Forecasting using Enhanced Convolutional Neural Network and Enhanced Support Vector Regression in Smart Grids. *Electronics*, vol. 8, no. 2, 2019, p. 122.

[4] H. Hahn, S. Meyer-Nieberg, and S. Pickl. Electric load forecasting methods: Tools for decision making," *Eur. J. Oper. Res.*, vol. 199, no. 3, 2009, pp. 902–907.

[5] Jiang, Hui, Kun Wang, Yihui Wang, Min Gao, and Yan Zhang. Energy big data: A survey. *IEEE Access,* vol.4, 2016, pp. 3844-3861.

[6] N. Ayub, N. Javaid, S. Mujeeb, M. Zahid, W. Z. Khan, and M. U. Khattak. Electricity Load Forecasting in the Smart Grids using Support Vector Machine. *in Proceedings Int. Conf. on Adv. Information and Networking Applications, Springer*,2019, pp. 1-13.

[7] J. P. Liu and C. L. Li. The short-term power load forecasting based on sperm whale algorithm and wavelet least square support vector machine with DWT-IR for feature selection. *Sustain.*, vol. 9, no. 7, 2017.

[8] A. Ghasemi, H. Shayeghi, M. Moradzadeh, and M. Nooshyar. A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand-side management. *Appl. Energy,* vol. 177, 2016, pp. 40–59.

[9] K. Wang, C. Xu, Y. Zhang, S. Guo, and A. Y. Zomaya. Robust Big Data Analytics for Electricity Price Forecasting in the Smart Grid. *IEEE Trans. Big Data*, vol. 5, no. 1, 2017, pp. 34–45.

[10] S. Ryu, J. Noh, and H. Kim. Deep neural network based demand side short term load forecasting. *2016 IEEE Int. Conf. Smart Grid Commun. SmartGridComm,* vol. 10, 2016, pp. 308–313.

[11] C. Fan, F. Xiao, and Y. Zhao. A short-term building cooling load prediction method using deep learning algorithms, *Appl. Energy*, vol. 195, 2017, pp. 222–233.

[12] J. H. Zhao, Z. Y. Dong, and X. Li. Electricity price forecasting with effective feature preprocessing. *in Proceedings of IEEE Power Engineering Society General Meeting,* 2006, pp. 8.

[13] M. Ramin, and W. Jianhui. A hierarchical framework for smart grid anomaly detection using large-scale smart meter data. *IEEE Transactions on Smart Grid*, vol. 9, no. 6, 2017, pp. 5820-5830.

[14] Z. W. Qiu. Mutivariable mutual information based feature selection for electricity price forecasting, *in Proceedings of International Conference on Machine Learning and Cybernetics,* 2012, pp. 168-173.

[15] O. Abedinia, N. Amjady, and H. Zareipour, A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems. *IEEE Trans. Power Syst.*, vol. 32, no. 1,2017, pp. 62–74.

[16] H. Qian and Z. Qiu. Feature selection using C4.5 algorithm for electricity price prediction. *in Proceedings of International Conference on Machine Learning and Cybernetics,* 2014, pp. 175-180.

[17] B. Bilalli, A. Abelló, T. Aluja-Banet, and R. Wrembel, Intelligent assistance for data pre-processing. *Comput. Stand. Interfaces*, vol. 57, 2018, pp. 101–109.

Information about the authors:

Usman Ali – is a student of MSCS Degree program and currently working in the data-sciences field.

Dr. Abdul Rauf – has a Ph.D. degree in computer science and currently working as a professor at RIPHAH International University Faisalabad, Pakistan.

Dr. Umer Iqbal – has a Ph.D. degree in computer science and currently working as a professor at RIPHAH International University Faisalabad, Pakistan.

Dr. Ijaz Ali Shoukat – has a Ph.D. degree in computer science and currently working as a professor at RIPHAH International University Faisalabad, Pakistan.

Abu ul Hassan – is a student of MSCS Degree program and also working as a lecturer in Superior University, Faisalabad.

Manuscript received on 12 July 2019