MACHINE LEARNING-BASED ENERGY DEMAND FORECASTING

[1] Ms. Isha Vyas, [2] Umesh Amate, [3] Bhaktvatsal Thakur, [4] Prajwal Nalawade, [5] Anuj Nirmal,
[1] *Asst. Prof.*, [2],[3],[4],[5] *Students*, Department of Electronics and Telecommunication,
Pimpri Chinchwad College of Engineering, Pune

ABSTRACT

In this study, the utilization of publicly accessible data to create forecasting models for electricity demand. Developed two machine learning-based models and evaluated their accuracy using various metrics. The data spanned from 2016 to 2018 and contained numerous measurements and observations regarding the building's electricity consumption and time-series weather variables, in varying time increments. These findings indicate that Support Vector Machine and random forest, among other machine learning algorithms, can achieve highly accurate electricity demand predictions when trained on such data.

Keywords— Electricity demand forecasting; Machine Learning Algorithms; Time Series Structured Data Analysis

INTRODUCTION

Energy consumption is a critical aspect of modern society, as it is directly linked to economic growth and quality of life. Accurate forecasting of energy demand is crucial for ensuring a reliable supply and meeting the needs of consumers. However, predicting energy consumption is a challenging task due to the complex and dynamic nature of energy systems. Traditional forecasting methods often rely on statistical techniques and historical data, but machine learning algorithms have gained attention in recent years for their ability to capture complex patterns and relationships in large datasets.

In this context, this study aims to develop and evaluate machine learning-based models for forecasting energy consumption, with a particular focus on electricity demand. Publicly available data is uesd from various sources to train and evaluate the performance of different machine learning algorithms, including SVM, random forest, and neural networks. Our objective is to compare the accuracy of these models using various evaluation metrics and identify the most effective approach for energy demand forecasting. By doing so, we hope to contribute to developing more reliable and accurate energy forecasting methods that can support policy-making and investment decisions in the energy sector.

RELATED WORK

Feature engineering is a crucial step in building machine learning models, as it involves selecting the relevant variables that can capture the key factors affecting energy consumption. Further exploration of feature selection methods could improve the accuracy and generalizability of energy demand forecasting models. Hybrid models that combine multiple machine learning algorithms, such as neural networks and decision trees, have shown promising results in energy consumption forecasting.

Investigating the effectiveness of our models in capturing the complex and dynamic nature of energy systems could be an interesting avenue for further research. The level of granularity and data availability can significantly impact the accuracy of energy demand forecasting models. Exploring the effect of different time intervals and data sources on model performance could improve our understanding of the factors that affect energy consumption. Interpreting and explaining the results of machine learning models can be challenging, especially in the context of energy systems. Developing models that provide interpretable and explainable results could increase their transparency and support decision-making processes in the energy sector.

FLOWCHART

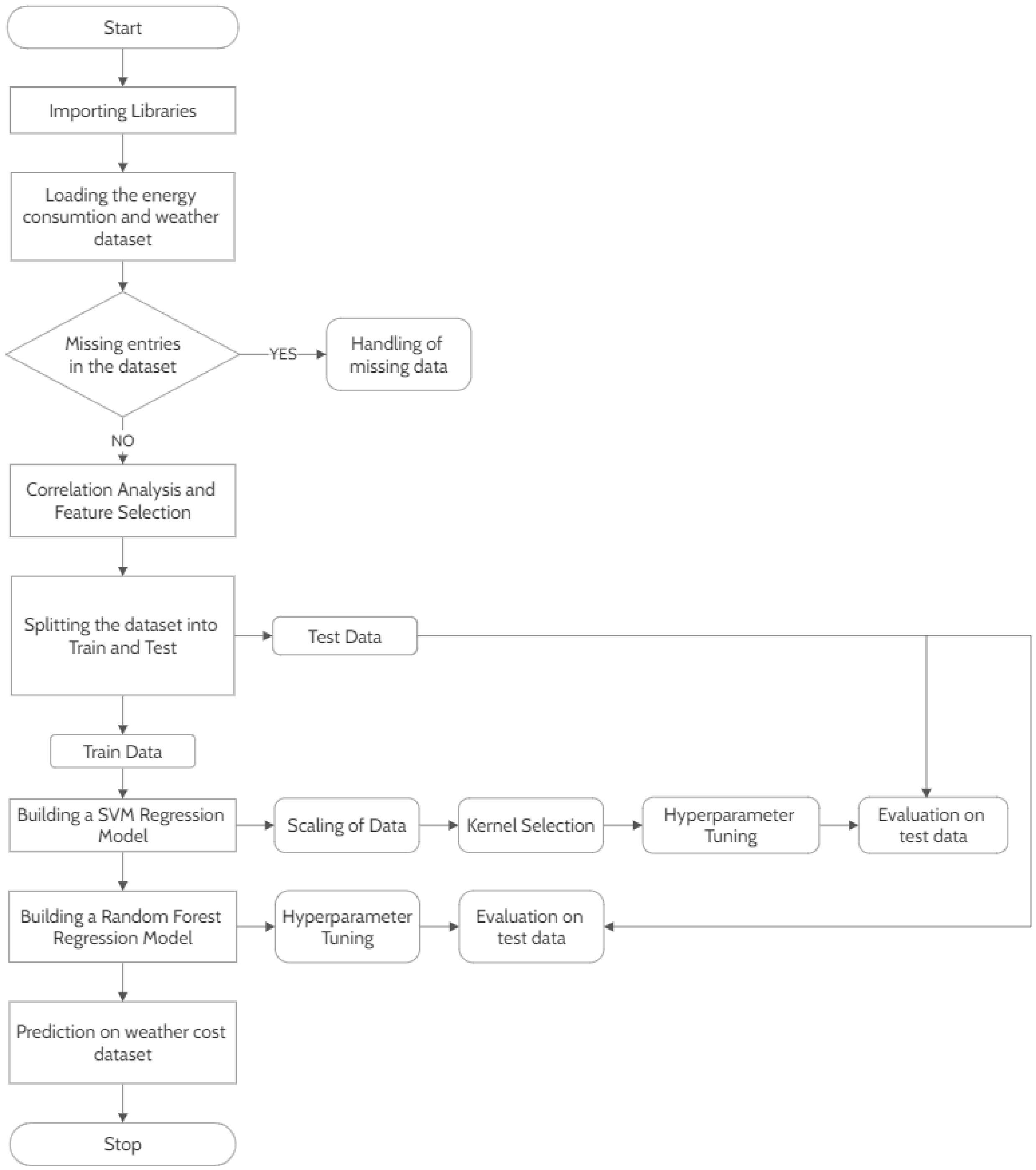

Figure 1 - Flowchart

METHODOLOGY

The prediction of energy demand and corresponding daily energy cost is estimated by performing predictive modeling on the collected dataset. The dataset is collected from a comprehensive collection of datasets on Kaggle. This dataset contains time-series structured data for 3 years of electrical consumption, generation, and building pricing from 2016 to 2018. Furthermore, it includes hourly weather data for the corresponding time instants.

Weather variables, viz. Hour of the day, Dew point temperature, Relative humidity, Air temperature, Hourly sum of precipitation, Specific humidity, Wind direction, Wind speed, Maximum wind gust speed, and Atmospheric pressure serve as feature or predictor variables, and building consumption as a target or output variable. Further, exploratory data analysis is performed to examine the presence of any null or incomplete entries. The data have been discovered to be complete, so, no feature imputations were needed. Besides, the dataset had a notably insignificant amount of outlier entries.

Python libraries Pandas, Numpy, Matplotlib, and Scikit Learn have been used for various data preprocessing, data manipulation, visualization, and regression tasks.

Most relevant features were identified with the computation of the Pearson correlation coefficient, which measures the linear correlation between two predictor variables. To bring all the features or variables of the dataset onto a similar scale, data normalization is performed using Standard Scalar, MinMax Scalar, and Robust Scalar techniques.

The dataset is then split to form train and test datasets. The train and test set share 80% and 20% proportions of the original dataset respectively. Subsequently, to train the model, the train and test sets were passed as arguments to instances of the Support Vector Machine and Random Forest regressor classes sequentially.

To produce more effective performance, an optimal set of hyperparameters were computed for SVM and RF models. The SVM hyperparameters C, Epsilon, Gamma, and the RF hyperparameter max_depth were tuned using an exhaustive grid search.

The predictive modeling has been performed on a new unseen time-series structured dataset containing weather parameters. This dataset is again pre-processed to handle the presence of any NaN values. Subsequently, the predictions for hourly energy consumption were made on a normalized version of this dataset using the prediction method of the Random Forest regressor. The predicted values were then converted to a Pandas data frame and the calculation of hourly cost has been done assuming standard charges of INR (₹) 4 per kilo-watt Hour of energy consumption. The hourly cost results were then resampled daily and visualizations were created to show the predicted hourly energy consumption, hourly cost, and daily cost over time.

RESULTS

Performance of SVM Kernels -

SVM Kernel	R2 – Score	Mean Squared Error
Linear Kernel	0.6440430907072687	14.222181951190796
Polynomial Kernel	0.6354473894568301	14.5656213507089
Radial Basis Function	0.8676607662388549	5.287585695616701

Table 2- Performance of SVM Kernels

Among the three SVM kernel functions, Radial Basis Function yields the higher R2 – Score and a significantly lower mean squared error. It explains 86.77% of the variance of the dependent features.

Hyperparameter Tuning Results -

Hyperparameters for a regression algorithm are a set of configurable parameters with certain default values. This set needs to be adjusted for the optimal performance of the model. SVM hyperparameters C, Epsilon, Gamma, and max_depth for Random Forest are tuned exhaustively using a grid search approach. The set of hyperparameters that resulted in the best performance on the test set are -

Best SVM Hyperparameters - `'C': 14, 'epsilon': 0.03, 'gamma': 3`

```
R2 – Score - 0.9861953911819542
Mean Squared Error - 0.5485543150502656
```

Best RF Hyperparameters: `'max_depth': 27`

```
R2 – Score - 0.8988574151607829
Mean Squared Error – 4.019107102575941
```

Comparison Results for the Performance of SVM and RF Models –

Algorithm	R2 – Score		Mean Squared Error	
	Train Set	Test Set	Train Set	Test Set
Support Vector Machine Algorithm	0.9921768284888686	0.9861953911819542	0.31257313950656923	0.5485543150502656
Random Forest Algorithm	0.9862465310884323	0.9064192603728158	0.5495168999781948	3.7186217447162955

Table 3 - Comparison of performance for SVM and RF Models

From Table 3, it can be interpreted that the R2 score for the SVM model is 0.9861953911819542 for the test set. This indicates that the SVM model can explain 98.6% of the variance in the dependent variable based on the independent variables. The MSE for the SVM model is 0.5485543150502656, which is relatively low. A low MSE indicates that it can accurately predict building energy consumption based on the predictor features.

The R2 score for the Random Forest (RF) model is 0.9064192603728158, which is lower than that of the SVM model. The MSE for the RF model is 3.7186217447162955, which is relatively high. So, the Random Forest regressor may not be able to understand the data patterns and is relatively less accurate while predicting energy consumption.

Fig. 2 consists of a line plot representing the fitting of predicted values against the original data for the SVM regression model. Yellow and blue colored plots represent the predicted data and the original data. This plot clearly explains the overfitting of the SVM model, as predicted data appears to be almost exactly fitting with the original data.

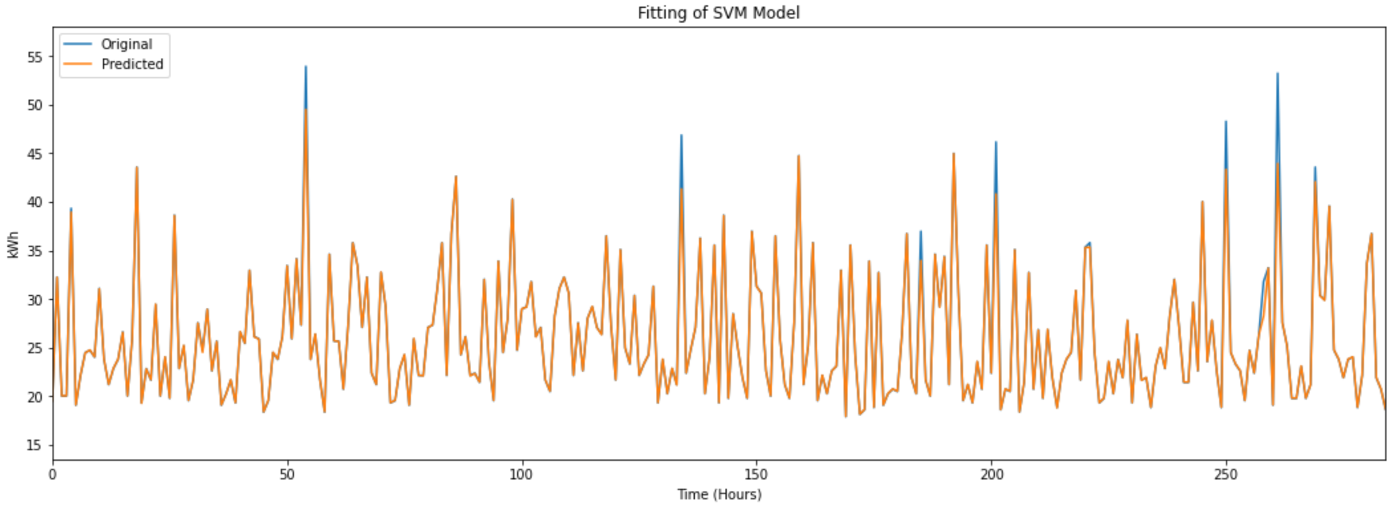

Figure 2 - Fitting of SVM Regressor

Fig. 2 consists of a line plot representing the fitting of predicted values against the original data for the Random Forest regression model. Magenta and blue colored plots represent the predicted data and the original data. The RF model fits inferior to the SVM model but also avoids the overfitting issue.

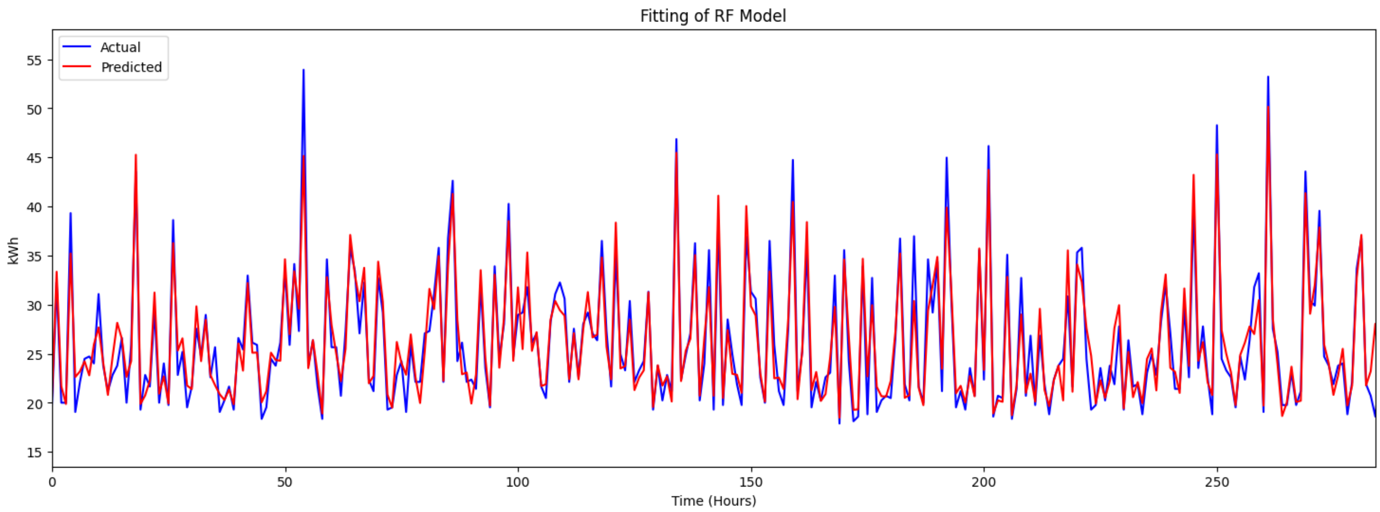

Figure 3 - Fitting of RF Regressor

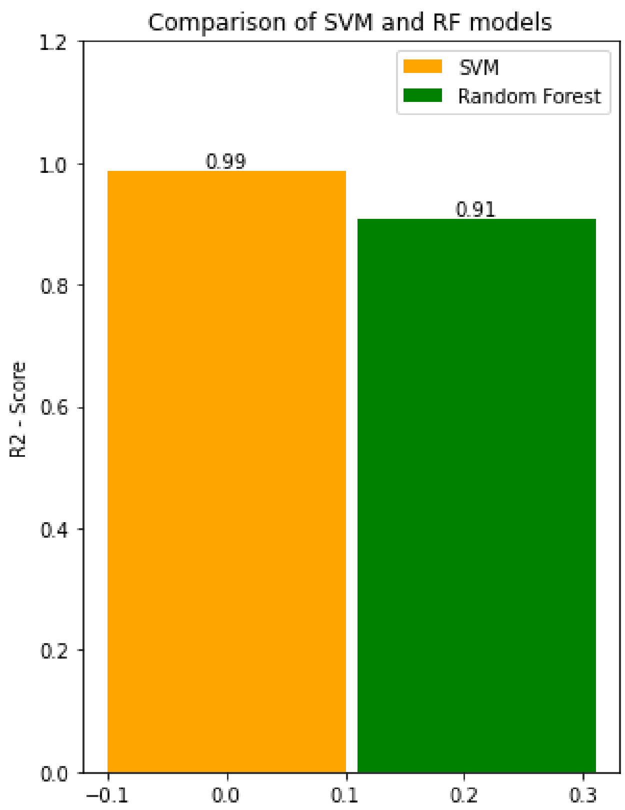

Figure 4 - Bar plot for R2 - Scores of SVM and RF

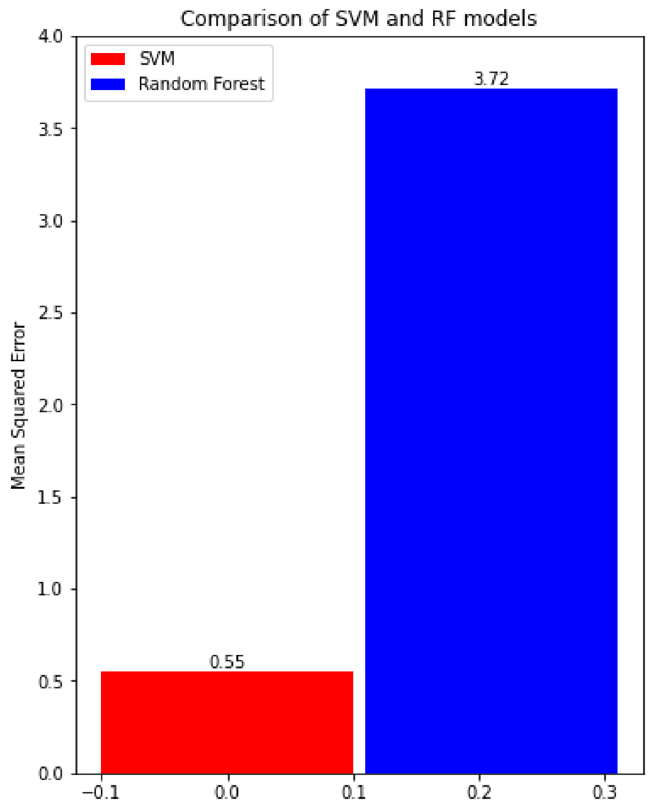

Figure 5 - Bar plot for MSEs of SVM and RF

Predicted Hourly Energy Consumption -

Time	kWh	Time	kWh	Time	kWh
2019-01-01 00:00:00	45.124553	**2019-01-17 12:00:00**	45.593154	**2019-01-31 15:00:00**	21.692187
2019-01-01 01:00:00	45.108070	**2019-01-17 13:00:00**	45.383579	**2019-01-31 16:00:00**	23.802066
2019-01-01 02:00:00	45.174004	**2019-01-17 14:00:00**	44.545279	**2019-01-31 17:00:00**	23.528912
2019-01-01 03:00:00	45.277614	**2019-01-17 15:00:00**	21.993599	**2019-01-31 18:00:00**	22.864865
2019-01-01 04:00:00	45.185777	**2019-01-17 16:00:00**	22.617614	**2019-01-31 19:00:00**	22.664709
2019-01-01 05:00:00	45.117489	**2019-01-17 17:00:00**	22.982604	**2019-01-31 20:00:00**	22.636452
2019-01-01 06:00:00	45.176358	**2019-01-17 18:00:00**	22.810705	**2019-01-31 21:00:00**	22.612904
2019-01-01 07:00:00	45.185777	**2019-01-17 19:00:00**	22.189045	**2019-01-31 22:00:00**	22.565809
2019-01-01 08:00:00	45.329419	**2019-01-17 20:00:00**	22.711805	**2019-01-31 23:00:00**	22.311493

34

Figure 6 - Predicted Hourly Energy Consumption

Fig. 6 shows some values of predicted hourly energy consumption in kilo-Watt Hours against corresponding time instants.

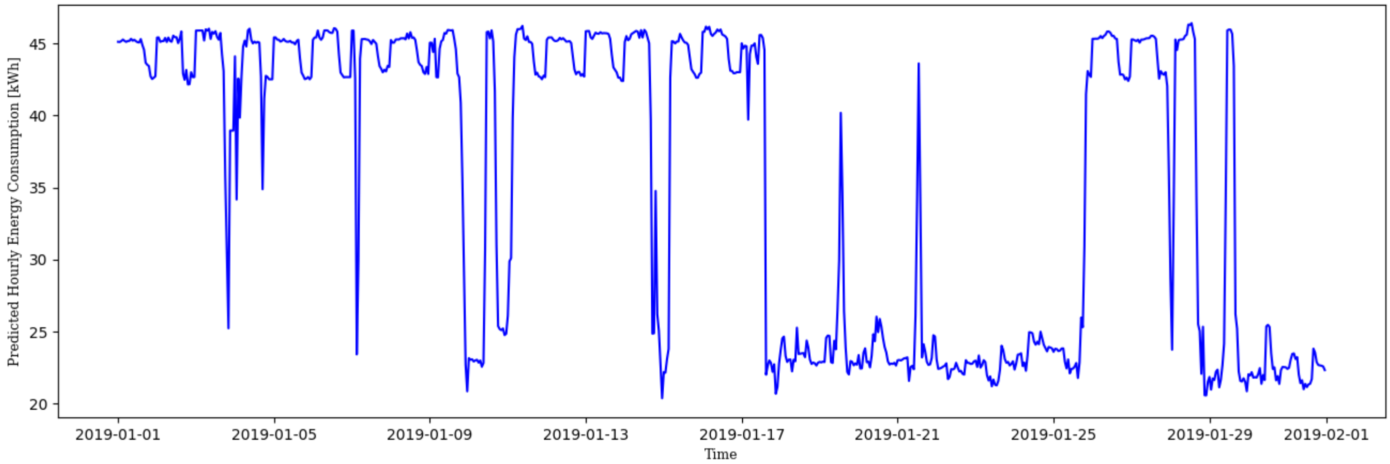

Figure 7 - Line plot of energy consumption

Fig. 7 is a line plot showing the trend of the predicted hourly energy consumption over time. The x-axis indicates the date and the vertical axis shows the predicted energy consumption in kilowatt-hours (kWh).

Predicted Hourly Cost -

Time	Cost (₹)
2019-01-01 00:00:00	180.498213
2019-01-01 01:00:00	180.432280
2019-01-01 02:00:00	180.696014
2019-01-01 03:00:00	181.110455
2019-01-01 04:00:00	180.743110
2019-01-01 05:00:00	180.469956
2019-01-01 06:00:00	180.705434
2019-01-01 07:00:00	180.743110
2019-01-01 08:00:00	181.317675

Time	Cost (₹)
2019-01-17 12:00:00	182.372614
2019-01-17 13:00:00	181.534314
2019-01-17 14:00:00	178.181115
2019-01-17 15:00:00	87.974394
2019-01-17 16:00:00	90.470455
2019-01-17 17:00:00	91.930416
2019-01-17 18:00:00	91.242822
2019-01-17 19:00:00	88.756179
2019-01-17 20:00:00	90.847220

Time	Cost (₹)
2019-01-31 15:00:00	86.768749
2019-01-31 16:00:00	95.208263
2019-01-31 17:00:00	94.115647
2019-01-31 18:00:00	91.459461
2019-01-31 19:00:00	90.658838
2019-01-31 20:00:00	90.545808
2019-01-31 21:00:00	90.451617
2019-01-31 22:00:00	90.263235
2019-01-31 23:00:00	89.245973

Figure 8 - Predicted Hourly Energy Cost

Fig. 8 shows the hourly charges (in Indian Rupees) for corresponding predicted hourly energy consumption values.

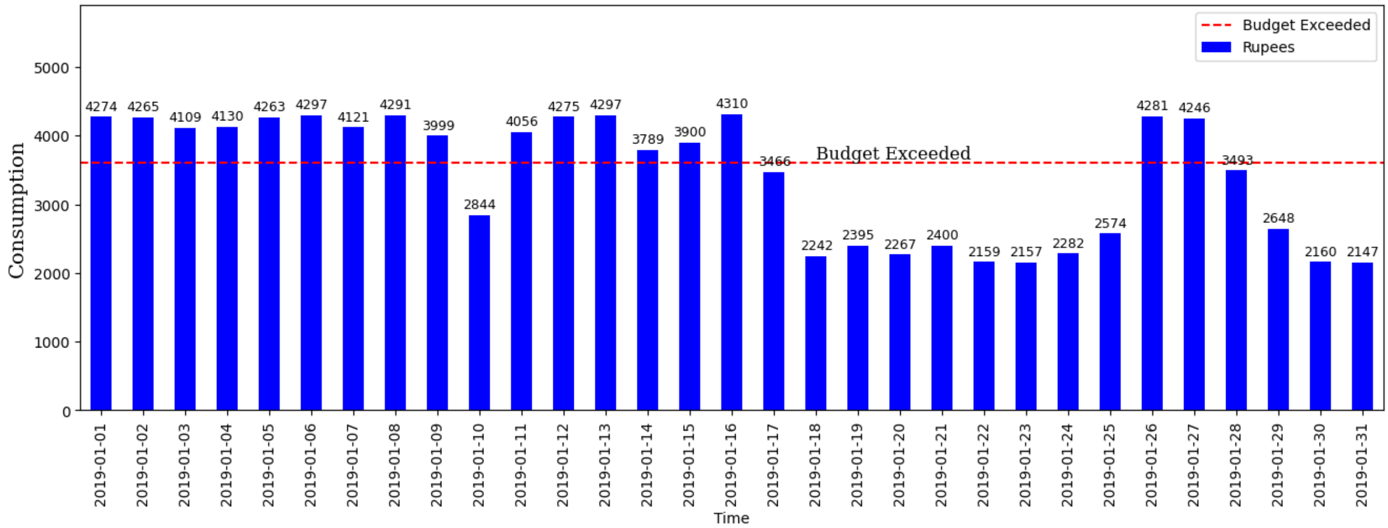

Figure 9 - Daily energy consumption costs over a period of time

Fig. 9 is a bar plot with each bar representing daily energy consumption cost. The plot has a red dashed line at the budget limit of 3600, which is the threshold that shouldn't be exceeded. The text labels on each bar indicate the exact value of the energy consumption cost for that particular day. The Y-axis shows the consumption cost values, while the X-axis shows the dates.

CONCLUSION

Two machine learning algorithms, Support Vector Machine (SVM) and Random Forest (RF) found to be better to fit on the given dataset. Standard normalization and Radial Basis kernel provided the best fitting and prediction results. Standard scaling and RBF explain 86.50% and 86.76% of the variance in the dependent variable based on the independent variables respectively. The R2 score for the Random Forest (RF) model is 0.9064192603728158, which is lower than that of the SVM model. The MSE for the RF model is 3.7186217447162955, which is relatively higher than SVM. This implies, SVM fits better to the data compared to the RF regressor but is proven to be prone to overfitting. Energy consumption varies across regions and countries, and developing models that can be transferred to different settings could increase their applicability and scalability. Investigating the transferability of machine learning models across different regions and energy systems could be a useful direction for future research.

REFERENCES

[1] 2017-UBMK-ElectricPrediction-v1

[2] M. A. Ortega-Vazquez, J. M. Flores-Vazquez, and J. P. Catala-German, "Electricity Demand Forecasting: A Review of the State-of-the-Art with a Look into the Future," Energies, vol. 13, no. 3, p. 676, Feb. 2020, doi: 10.3390/en13030676.

[3] T. L. Goh, H. M. Lee, and W. Y. Ong, " Electricity demand forecasting using machine learning algorithms: A review," Renewable and Sustainable Energy Reviews, vol. 113, p. 109297, Nov. 2019, doi: 10.1016/j.rser.2019.109297.

[4] C. C. M. Prajapati and R. K. Kamalapurkar, "Short-term Load Forecasting using Machine Learning Techniques," 2018 International Conference on Computing, Communication, Control, and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8542172.

[5] S. N. Alam, N. M. Saad, and R. Saidur, " Energy Demand Forecasting: A Review of the State-of-the-Art with a Look into the Future Directions," Renewable and Sustainable Energy Reviews, vol. 91, pp. 742-755, Oct. 2018, doi: 10.1016/j.rser.2018.04.081.

[6] J. C. Guzman, R. A. Fuentes-Fernandez, and J. L. Salmeron-Lissén, "Short-term forecasting of electricity demand with machine learning techniques," Energy, vol. 142, pp. 937-953, Nov. 2017, doi: 10.1016/j.energy.2017.08.097.

[7] Hourly energy demand generation and weather: https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather