Solar energy prediction using machine learning

Shrinivas Gollalappa Kadaganchi
700750193
dept.Computer Science
University of Central Missouri
sxk01930@ucmo.edu

Vamsi inampudi
700747651
dept.Computer Science
University of Central Missouri
Vxi76510@ucmo.edu

Chaitanya Phani Kumar Akula
700740502
dept.Computer Science
University of Central Missouri
cxa05020@gmail.com

Abstract—The world is largely dependent on fossil fuels but these resources are finite. Leaving the era of fossil fuels behind is gaining enormous pace. Solar energy is the cheapest way to produce electricity without using fossil fuels in light of the rising greenhouse effect. Solar panel electricity can be harvested directly from the sun even on overcast days without the need for a costly setup beyond the installation of solar panels. Even the pollution of the air and water is decreased by this procedure. Solar photovoltaic (PV) and concentrated solar power (CSP) are the two methods used to collect solar energy. Firstly, electrical devices are used to transform solar energy into electricity. The energy from sun rays is captured using mirrors in the later method. In this project, we are proposing ensemble regression models to predict both generated energy and exported energy. The main objectives of the project are: To select the 2 individual datasets for experimental analysis. Analyzing the data with Exploratory data analysis.Building multiple regression models to predict solar energy generation and exported energy using weather and plant parameters.Evaluating the models and conducting a comparative analysis of the models. For the experimental analysis dataset is collected from PV-Output.org and the features are: Generated energy, Peak Generation, Exported energy, minimum temperature, maximum temperature, and generate/exp ratio.

[1] *Index Terms*—concentrated solar power (CSP), Solar photovoltaic (PV), minimum temperature, maximum temperature, and generated energy.

I. INTRODUCTION

Solar energy forecast features exhibit a complex relationship between the weather parameters and solar energy parameters which leads to bias in predictions. Multiple evaluation metrics are incorporated to assess the model performance Mean Absolute Error (MAE), Rooted Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The complexity of the parameters is reduced by choosing the ensemble models. Ensembling techniques allow one to choose individual models by reducing the bias in predictions. Aggregating the results of multiple data model performance.

Academic study has increasingly centered on examining the dynamics of solar energy prediction. In this section, we review the related works on solar-generated energy and exported energy using machine learning.

Support Vector Machines Based on Multi-source Data Fusion: This paper proposes the SVM-DF(Support Vector Machine) model and ANN(Artificial Neural Network) models to predict the solar power generated and solar power emission.

SVM-DF is an extension of general SVM. To perform experiments weather parameters and power-related data are used. To improve the performance another set of weather data NWP is used. In the experimental analysis, SVM-DF outperformed the ANN regressor.

Solar Power Generation Prediction from Weather Forecasts using Machine Learning: This paper proposes a strategy for building smart power grids. Optimization of the number of grids in the solar energy system is gaining popularity. To assist this information of weather data and geographical location is needed. Weather data is collected from NSRDB(National Solar Radiation Database). Multiple linear regression models like Ridge and Lasso are implemented to forecast solar energy.

Solar Energy Prediction using Meteorological Variables: This study presents a time series solar energy forecasting model for 7 days in advance. To predict solar energy two types of parameters are considered one is plant related parameters and temperature parameters. In the first category inverter, cables are considered and in the second category statistical parameters of the ambient temperature are considered. Experimental results showed accurate results except on overcast days.

Because of its environmental and economic benefits, solar energy is increasingly being integrated into smart grids and numerous utilities. However, the uncertainty of accessible solar energy poses issues in terms of power-generating reliability and, as a result, consistency in production-consumption balance. In an ensemble, support vector regressors are employed as base predictors, and Multi-layer Perceptrons, Decision Trees, and K-Nearest Neighbour Regressors are utilized as meta-learners to combine.

An ensemble learning strategy is proposed to improve the forecasting of solar radiation strength on horizontals. To forecast solar radiation, two types of machine learning models are used: recurrent neural networks and support vector regressors. A multi-layer perceptron model serving as an ensemble learning technique is used to combine the forecasts of ensemble models using a stacking strategy. The combiner performs automatic weighted averaging of the forecasters' results. The proposed method aids in enhancing the accuracy of one-day-ahead solar energy predictions. The performance of the combining technique is tested over a year using meteorological data. The trials demonstrated that the learning-based combinatory model outperformed single models and

---

[1] https://github.com/shrinivas1881/Shrinivas-Kadaganchi.git
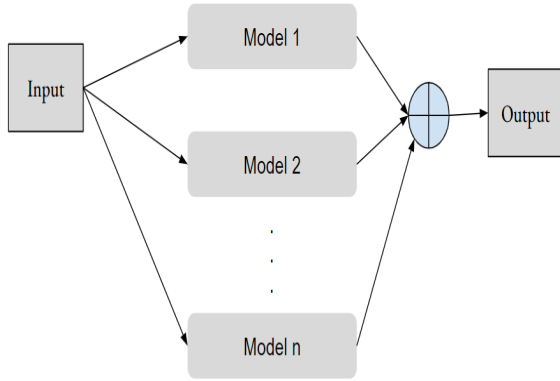
alternative combining strategies.

Figure 1. Architecture of ensemble models

Ensemble modeling is utilized in this project to introduce diversity in forecasts. To anticipate a result, the ensemble modeling process employs numerous distinct base models known as weak learners. This methodology is based on the wisdom of the crowd's method. The goal of using ensemble models is to make model performance more resilient and to reduce error. The technique aims to make a forecast. It has real-time data applications because of its robustness. There are 3 different types of ensemble methods. Bagging methods: Bagging is a basic method in ensemble learning. Bagging is often known as bootstrap aggregation. The main objective of the bagging ensemble approach is to reduce variance in a noisy dataset. The work of bagging includes selecting the random sample multiple times and training the model. In the end, the predictions are aggregated to produce the final predictions. Boosting methods: The main difference between bagging and bootstrapping is: bagging works on the principle of sampling with replacement and boosting method works based on the principle of resampling with replacement. Stacking models: The stacking method contains weak learners and the predictions from all the weak learners are passed to the meta-learner for final predictions.

II. MOTIVATION

Solar photovoltaic is gaining popularity for its modularity and low cost. By the end of 2020, there were 710 GW of solar PV installations worldwide. Many nations are entering the markets for selecting the most affordable electric equipment. We need to be aware of the device power forecast to accomplish this. Machine learning regression models are used to carry out this task to produce improved utilization. The motivation for this idea derives from the growing use of renewable energy plants, such as solar power plants. With the increased use of power plant setups, there is a great need to create cost-effective setups. To design a cost-effective set up knowing the output or generated and exported energy is important. Existing systems failed to capture the data

complexity. Therefore building an effective machine learning model to understand and predict real-time data is important.

III. OBJECTIVES

The main objectives of this project are:
- Constructing machine learning models to predict the generated and exported energy using weather parameters and derived features of energy at a particular Photovoltaic power plant.
- Observing the variability in the data and taking measures to reduce the variability.
- Applying various ensemble models like bagging regression models, boosting regression models, and stacking regression models to reduce the variability of the data and to improve the efficiency of the predictions.
- Comparing the performance of the bagging, boosting, and stacking ensemble models.
- Analyzing the best-fit line of predictions of the models using a scatter plot, feature importance plot, and permutation importance test.

IV. RELATED WORK

The distributed energy in a virtual power plant is computed in this study. There are five components involved in predicting distributed energy. One is wind power and photovoltaic generation energy computation, while the other parameters include power saving potential, pollution emission, and cost reduction. The model's scalability and interpretability are observed by gathering all of these parameters. The final step of the analysis of the results reveals that environmental elements influence the computation of VPP [20].

Virtual Power Plants (VPP) are becoming increasingly common as the use of distributed energy sources grows. Currently, the system is centralized, and it gives renewable energy projection information to the buyer who purchases the information. Consumers are mostly interested in learning about power trade trends. To regulate the information pass, a decentralized system algorithm is presented in this work. Experimental results demonstrate that a decentralized system is more efficient than a centralized system [2].

With the development of Artificial Intelligence, it is now possible to forecast complicated data correlations. One of the projects is the forecast of thermal power generation. In this paper, the generated thermal power is projected using data from the Jamshoro power plant. The data includes the data's historical parameters. The prediction procedure is divided into three stages: gathering raw data from thermal power plants using a dependable resource, preprocessing the data, and model construction. After obtaining data from a credible source, redundant characteristics are deleted, and numerous prediction models are developed before selecting the best model based on performance [14].

The motivation for Dynamic Linear Prediction Model Based on Energy Storage System originates with the cause of randomness and volatility of Wind energy. This nature can be predicted using dynamic linear prediction models based on an

energy storage system (ESS). It starts with collecting the data and predicting the energy JITL(Just in Time Learning) algorithm is deployed. Wind Power Energy errors are considered to be serious; these errors can be reduced using various data mining techniques [17].

In this paper, the power consumption prediction model of the Taïba plant is formulated which has 157 MW of installed capacity. The production value of all intermittent power plants is determined by the environmental factors of the area in which it is located. Instability in the energy grid can be caused by bad weather. It is vital to employ methods for forecasting its output. With the forecasting models how much energy to produce to meet demand is easier. Raw data is divided into 80-20 ratios to train and test the system. The experimental results are quantified and summarised using appropriate parameters [4].

Renewable energy integration is now prevalent in all industries, and with it comes the difficulty of estimating renewable energy output. Large-scale renewable energy integration is expected to have a significant impact on power system scheduling and control. The key challenge here is to, ensure optimal scheduling and control of a high-penetration renewable energy infrastructure. On the one hand, assessing the complementary characteristics of sources of renewable energy assists in smoothing the irregularities of renewable energy output. As a result, this study develops wind-photovoltaic integrated power prediction based on complementary characteristics and, wind-photovoltaic storage [18].

The current distribution grid renewable energy penetration rate continues to increase and the distribution grid is in dire need of peak and frequency control resources to strengthen its regulation capacity. As the cost of ESS(Energy Storage Systems) becomes cheap they are integrated everywhere. One of the major challenges is weak centralized architecture. In this project, a CTDE algorithm is proposed to reduce the limitations of centralized architecture [3].

Stationary sectors extensively use Battery storage systems for the Optimisation of renewable generation plants' self-consumption, lowering peak loadsUninterruptible power supply (UPS), and a variety of other reasons. To identify time windows this study presents a way for analyzing energy storage algorithms by analyzing power and capacity profiles. The applications of the study can be used to create a predictive control algorithm and these deployment alternatives are the most cost-effective. The future work is to create an algorithm, which will be published in a forthcoming work [1].

The key factors to ensure the safe and stable functioning of the power grid are a high rate of renewable energy and high-precision new energy power projection. Traditional methods lack accurate predictions. To address the shortcomings of traditional methods of power prediction accuracy approaches, a unique accuracy evaluation approach is proposed which precisely represents power forecast data [5].

Failure of AC side converter station commutation has a direct influence on a single, prone-to-failure DC transmission network. This work provides a reactive control power strategy to handle the issue of communication failure threat. Simulations are used in the experimental investigation to demonstrate the practicality of the control strategy [12].

In nature there are several tasks which are almost impossible to predict with less effort. In that category rainfall prediction is one of the tasks. With the advent of machine learning these tasks became simpler. But a single machine learning model can yield good results for this kind of complex natural data. With the evolution of ensemble techniques this task became easy in that stacking models produced far better results [11].

This research investigates the usefulness of utilising the Genetic programming algorithm to take the outputs of each model as input and learn how to effectively combine the input predictions to make predictions. Stacking ensemble refers to the process of combining predictions from many regression models using another regression model. This research investigates the usefulness of utilising the Genetic programming algorithm to take the outputs of each model as input and learn how to effectively combine the input predictions to make predictions. Stacking ensemble refers to the process of combining predictions from many regression models using another regression [16].

A high accuracy of the outcome with minimal resource and time expenses for the operation of the chosen algorithm is required for an efficient prediction task solution. When high accuracy of the acquired output is a top priority, ensemble learning is a good choice. This research describes a prediction method based on a new stacking-based GRNN ensemble model [10].

Malicious PDF files being highly disguised and difficult to detect.To efficiently identify malicious PDF files that are heavily veiled, a machine learning ensemble model that is stacking ensemble is deployed. The process involves combination performance of several base learners, and experiments to identify the best combination of base learners and meta learner. Following studies, NB, RF, and DT are chosen as the best Stacking model base learners, with Logistic Regression serving as the meta learner. The best Stacking model gets 98.70 percent of the time [15].

Diabetes, being one of the chronic diseases endangering human health, relies heavily on early detection and treatment for disease control and progression. A single machine learning model's generalisation capacity is limited, therefore it frequently has limits in diagnostic applications.To overcome the limitations of the existing systems, this research sheds a light stacked ensemble earning-based approach to detect early diabetes prediction model. Gradient Boosting Decision Tree, Adaboost, and Random Forest were utilised as estimators in this model, while logistic regression was used as meta learners [19].

In this study for predicting water level time series in the short term a C-Stacking ensemble model is used. In the first step adaptive noise method is used to decompose the water level into frequency and residual components. In the second stage SVR(Support Vector Regressor) and LSTM(Long Short Term Memory) is used as the primary learners [13].

Access to the internet and mobile platforms are increasing substantially. Simultaneously, web application security has become a serious concern on the internet. Inadequate validation of user input information, software designed without rigorous attention to safety standards, susceptibility of reusable software libraries, programme weakness, and other factors contribute to the threat. To address the issue a stacking ensemble-based classifier model for detecting attacks is proposed. Base-learners, such as k-Means, Random Forest, and Decision Tree, to accurately detect XSS attacks. Logistic Regression is used as a meta-learner to more accurately forecast the attack [7] [6].

An overview of the film and provide input to film filmmakers about their films based on public mood. Movie reviews are incredibly important to people since they assist viewers as well.However, due to the vast amount of data, manually analysing sentiment is tough. In this study we investigate a machine learning strategy for classifying movie review sentiment in this study. As a classification strategy, the Ensemble Stacking model was employed for this sentiment analysis example. As the base-learners in the classification step, three algorithms (Nave Bayes, K-Nearest Neighbours, and Logistic Regression) are applied [8].

To pick the most favourable variables for the model as auxiliary variables and maximum mutual information coefficient (CV-MIC) a stacking ensemble learning architecture is proposed. As base learners, the model employs LSTM-GPR, SDAE-SVR, and XGBoost, and DNN as the meta-learner for ensemble learning. The model is applied to the prediction of the rotor deformation of the boiler air preheater in thermal power plants to validate its effectiveness. According to the experimental data, the soft sensor model based on ensemble learning has a greater prediction accuracy [9].

V. Dataset

Data in this project is collected from the source PV-Output.org. The original data is maintained by the solar energy department Queensland, Australia. The dataset contains 21 files which is the consolidation of different profiles who installed the Photovoltaic plants at their business or home. This is made available through GitHub https://github.com/gomesramos/PV-Output-Datasets. The features in the dataset are: Generated energy, Peak Generation, Exported energy, minimum temperature, maximum temperature and generate/exp ratio. In this project, we have conducted the experiment on 6 danny.csv files. This file contains 1280 records with 15 columns. The dataset contains annual and daily statistics.

VI. Proposed Framework

The project implementation is divided into 4 steps:

A. Data collection and preparation:

In the machine project life cycle data collection and preparation is the first step. After collecting the raw data data is cleaned by removing the null values or missing values. In

| S.No.	Column name	Description
1.	Date	Time stamp
2.	Mes	Month
3.	Ano	Year
4.	Gen	Generated solar energy in kWh
5.	SubGen	SubGen energy
6.	mmGen	anonymous parameter
7.	Exp	Exported solar energy kWh
8.	subExp	Sub exported solar energy kWh
9.	mmExp	anonymous parameter
10.	PP	Peak Power
11.	Cond	Weather condition
12.	Temp min	Minimum temperature in a day
13.	Temp max	Maximum temperature in a day
14.	Temp med	Median temperature in a day
15.	exp/gen	Derived feature exported vs generated energy ratio

Table I
DATASET FEATURE DESCRIPTION
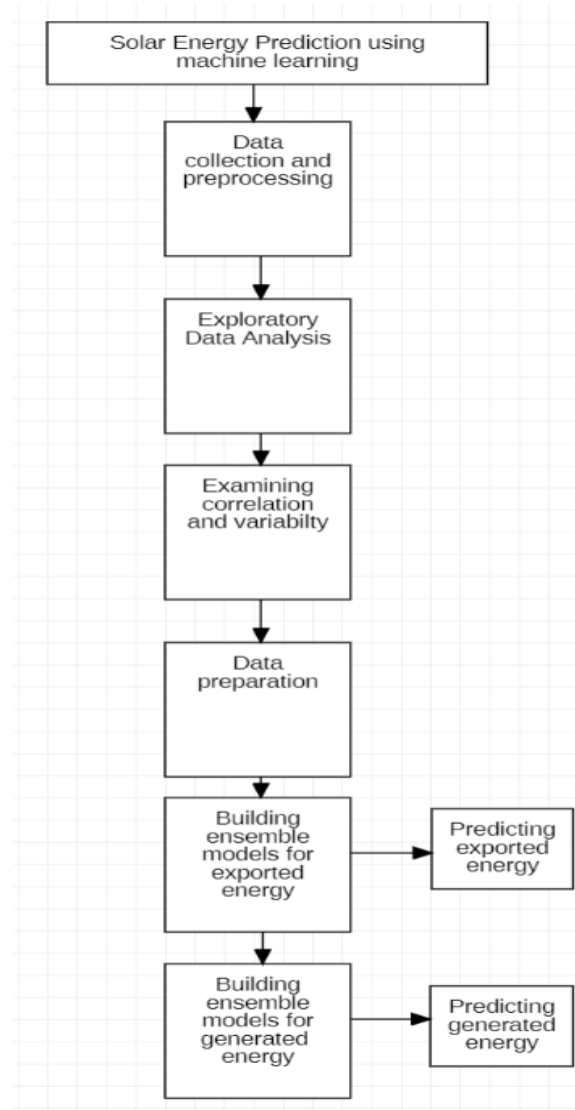
Figure 2. Workflow diagram

the data preparation step, categorical variables are encoded using a Label encoder. Data is split into training and testing.

B. Exploratory Data Analysis:

In this step, the distribution of the features is analyzed. A few observations from the EDA are:

- The majority of the samples are Fine weather condition-based samples. The least number of samples are showers data.
- There are no outliers in exported data and outliers are observed in generated energy samples.
- Outliers are present in Peak Power and the variance is very low
- Distribution of generated energy shows that there is a dip in the month of June and peaks are observed in the months of March, August, October, and December. Similarly exported energy also

Figure 4. Time series distribution of generated energy

Figure 3. Countplot of weather condition

C. Machine learning models

In this step machine models and ensemble regression models are constructed. To predict the results bagging regressor, gradient boosting regressor, and stacked regressor are trained. Model performance is evaluated using the test predictions and actual values. For each model r2 score, RMSE(Root Mean Squared Error), MSE(Mean Squared Error), MAE(Mean Absolute Error), and MAPE(Mean Absolute Percentage Error) is evaluated. To explore the data nature correlation strength is calculated for features. To perform this task heatmap is constructed using seaborn with correlated features. After the analysis, it is observed that there are features that are highly correlated. The following pairs have strongly correlated features
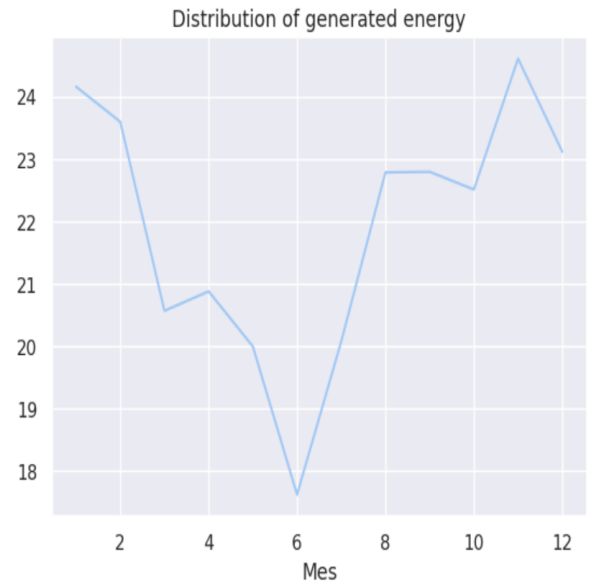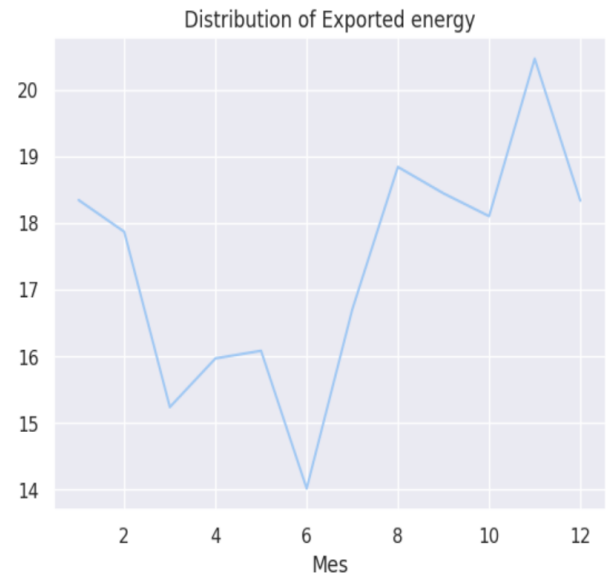
Figure 5. Time series distribution of exported energy

- Gen - mmGen : 0.83
- Gen - Exp: 0.88
- mmGen - mmExp : 0.85
- Exp - mm Exp: 0.85
- Temp min - Temp med: 0.95
- Temp max - Temp med : 0.

Predicting the solar energy(exported and generated) is a challenging task due to the high variance the in data. Ensemble models like bagging, boosting, and stacking models help to reduce the variance in the data. This can be achieved through the aggregation of predictions and by incorporating diverse
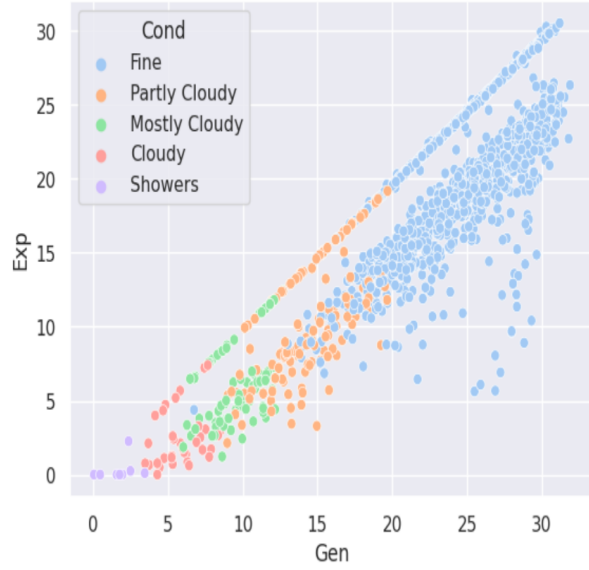
Figure 6. Linearity check for exported and generated energy

is calculated using Euclidean distance. RMSE: Root Mean Squared Error is calculated as the aggregation or squared root of MSE. This is calculated using squared distances of the actual and predicted values. The values of RMSE are always positive values and 0 values represent the idea case which is never achieved. Lower values of RMSE represent a good regression model but in practical terms, the good RMSE value depends on the data. Mean Absolute error is similar to the metric of RMSE but the interpretation of MAE is simpler compared to the RMSE. MAE is known as the mean aggregated value of the distance between the actual and predicted values. MAPE(Mean Absolute Error Percentage) is a relative error metric to calculate the deviation of error percentage deviation. It is used in various regression models due to its simple interpretation.
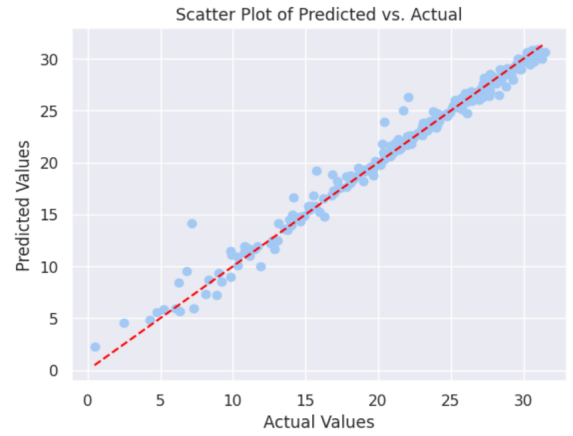

Figure 7. Gnerated energy best-fit line

models. Achieved r2 is 0.98 on both exported and generated energy prediction tasks.

The main features of the project are inducing diversity to improve model generalization on new data. In the bagging method multiple individual models or similar models are composed and trained on a randomly sampled subset of data and in the end the predictions are aggregated to produce accurate predictions. In boosting methods weak learners are trained sequentially improving the predictions of each step and in the end, the predictions are produced based on weightage. In the stacking method, multiple weak learners or base models are trained and the collected predictions are passed to the meta-model for final predictions.

Python programming language is selected to implement the project. Pandas, Seaborn, matplotlib and scikit-learn libraries are used in project development.

There are several challenges we faced during the implementation of the project are: Visualizing the model performance we tried to visualize the best fit using other than a scatterplot Few features in the data are anonymous and the units of the features are also unknown to better understand the data.

VII. RESULTS ANALYSIS

A. Evaluation metrics

To check the model's robustness different evaluation metrics are deployed. R2: R squared metric. This is also known as the coefficient of determination. This gives the understanding of how well-observed outcomes are predicted by the model in percentage. This is calculated as the fraction of the total variation described by the model Mean squared error: MSE measures the goodness or correctness of the estimator in predictions. It is always a positive value. The value of MSE approaches zero in ideal cases. The error or correctness
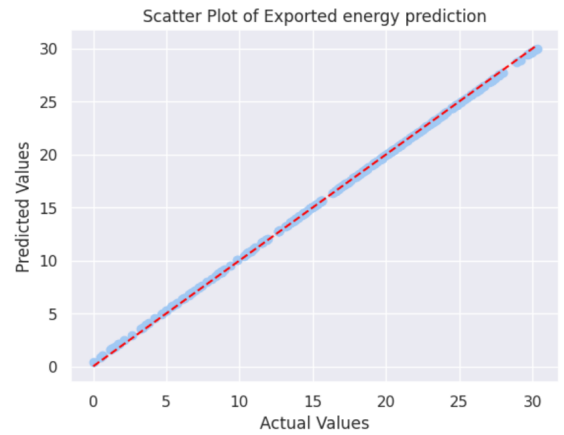

Figure 8. Exported energy best fit line

B. Comparative analysis

In this project in the step of stacking regressor weak learners are the KNN regressor, Support Vector Regressor, Random Forest regressor and Linear regression models. To make final
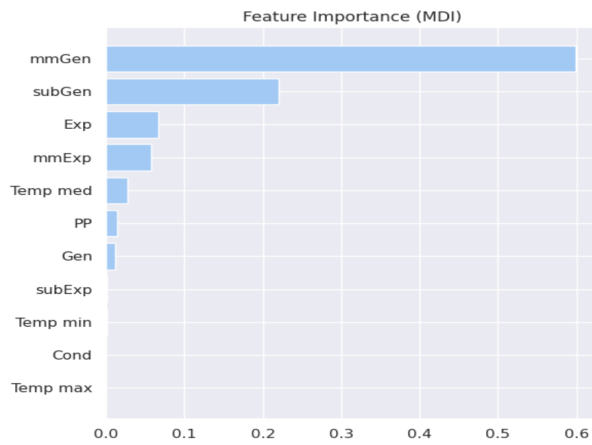
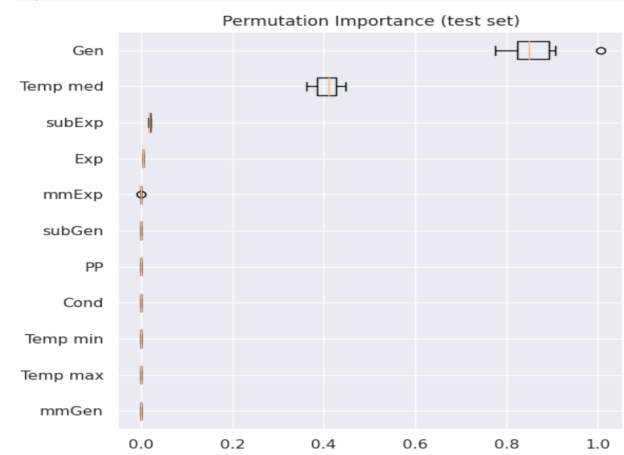Figure 9. feature importance score of generated energy

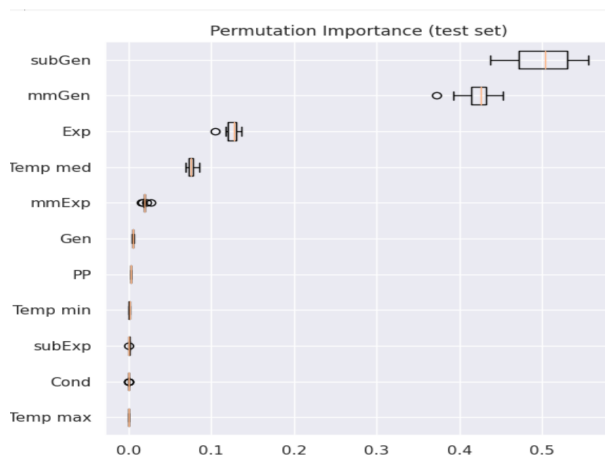Figure 12. permutation importance score of exported energy

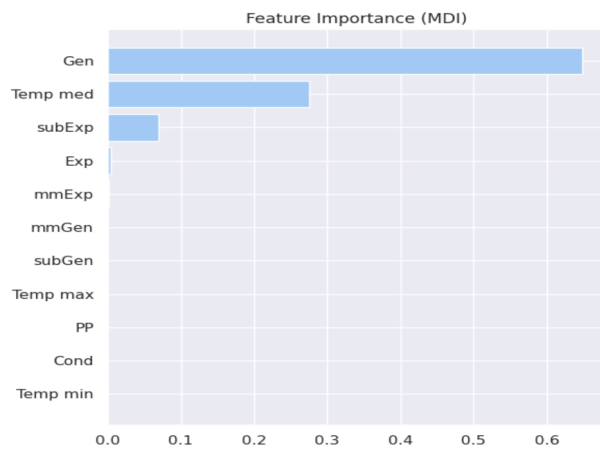Figure 10. permutation importance score of generated energy

are 10 and the bagging regressor scores the best fit by 99 percent. for boosting regressor, Gradient Boosting regressor is used. The score r2 in this case is 97 percent. After the comparative analysis stacked model performed the best in terms of overall accuracy.

In conclusion, the stacked model performed well in terms of overall accuracy with 0.99, 0.15,0.02,0.12,0.17 with respective R2, RMSE,MSE,MAE,MAPE metrics.

metric	Stacking regressor	Bagging regressor	Boosting regressor
R2	0.99	0.99	0.97
RMSE	0.15	0.42	0.94
MSE	02	0.18	0.899
MAE	0.12	0.25 0.69	
MAPE	0.17	0.16	0.03

Table II

SUMMARY OF MODEL PERFORMANCES

Figure 11. feature importance score of exported energy

predictions a meta learner is used that is Lasso regression.The stacked regression model scores best fit by 98 percent of the data. For the bagging regression method, the RandomForest regressor is used as a estimator and the number of estimators

REFERENCES

[1] Lukas Böhning, Mathias Herget, and Ulf Schwalbe. Investigation of energy storage systems - improvement of utilization by use case combination. In *2021 Ural-Siberian Smart Energy Conference (USSEC)*, pages 175–180, 2021.

[2] Yue Chen, Tongxin Li, Changhong Zhao, and Wei Wei. Decentralized provision of renewable predictions within a virtual power plant. In *2021 IEEE Power Energy Society General Meeting (PESGM)*, pages 1–1, 2021.

[3] Xiaoyun Deng, Chao Ma, Ke Yin, Ziwei Li, Weiting Xu, and Youbo Liu. Multi-agent smart control of virtual power plant energy storage in active distribution network. In *2023 Panda Forum on Power and Energy (PandaFPE)*, pages 1802–1807, 2023.

[4] Sambalaye Diop, Papa Silly Traore, Boubacar Niang, and Mamadou Lamine Ndiaye. Using multilayer neural network to increase the prediction accuracy: application in the taïba ndiaye power plant. In *2022 IEEE International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, volume 4, pages 1–6, 2022.

[5] Rui Fan, Huimin Wu, Xiao Chang, Chaoying Yang, Shifeng Zhang, and Jun Zhao. A new power prediction accuracy evaluation method of renewable energy plant. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 610–612, 2019.

[6] Tao Fang, Sirui Huang, Ya Zhou, and Huibing Zhang. Multi-model stacking ensemble learning for student achievement prediction. In *2021 12th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*, pages 136–140, 2021.

[7] Muhammad Khaifa Gifari, Kemas M Lhaksmana, and P. Mahendra Dwifebri. Sentiment analysis on movie review using ensemble stacking model. In *2021 International Conference Advancement in Data Science, E-learning and Information Systems (ICADEIS)*, pages 1–5, 2021.

[8] Zhao He and Han Liu. A novel soft sensor model based on stacking ensemble learning framework. In *2022 IEEE 11th Data Driven Control and Learning Systems Conference (DDCLS)*, pages 507–512, 2022.

[9] Fahima Hossain, Linta Islam, and Mohammed Nasir Uddin. Phishrescue: A stacked ensemble model to identify phishing website using lexical features. In *2022 5th International Conference of Computer and Informatics Engineering (IC2IE)*, pages 342–347, 2022.

[10] Ivan Izonin, Roman Tkachenko, Pavlo Vitynskyi, Khrystyna Zub, Pavlo Tkachenko, and Ivanna Dronyuk. Stacking-based grnn-sgtm ensemble model for prediction tasks. In *2020 International Conference on Decision Aid Sciences and Application (DASA)*, pages 326–330, 2020.

[11] Prof. Priyanka G. Jaiswal, Amey Dhote, Aniket Dubey, Harshal Patil, Neha Madavi, Nidhi Rahangdale, Shailesh Patil, and Swati Kale. A stacking ensemble learning model for rainfall prediction based on indian climate. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, pages 1–6, 2023.

[12] Zhen Lei, Lizhi Mu, Dawei Su, Yuchen Hao, and Mingming Shi. Reactive power control strategy based on electrochemical energy storage power plant to resist the risk of commutation failure. In *2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2)*, pages 95–99, 2019.

[13] Seethalakshmi Perumal and Kola Sujatha P. Stacking ensemble-based xss attack detection strategy using classification algorithms. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, pages 897–901, 2021.

[14] Ghulam Qadir, Irfan Bhacho, and Naeem Ahmed Mahoto. Predicting the energy efficiency of thermal power plant. In *2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pages 1–6, 2020.

[15] Yidan Tang, Jinjin Dong, Yixuan Guo, Yihan Zhou, Feifan Lu, and Bo Zhang. A malicious pdf file detection method based on improved ensemble learning stacking. In *2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, pages 475–478, 2022.

[16] Vinayak Vijaya Chandran and Roopa Adepu. Reduced order modeling of a heat exchanger with a stacking ensemble to reduce computational inefficiencies. In *2022 IEEE International Symposium on Systems Engineering (ISSE)*, pages 1–5, 2022.

[17] Wei Yang, Li Jia, Yong Chen, Yue Xu, and Chengyu Zhou. Dynamic linear prediction model based on energy storage system compensating prediction error for wind power. In *2022 6th International Conference on Power and Energy Engineering (ICPEE)*, pages 59–66, 2022.

[18] Zhang YanQi, Zhou Qiang, Zhao Long, Ding Kun, Wang Dingmei, and Zhang Ruixiao. The key technology for optimal scheduling and control of wind-photovoltaic-storage multi-energy complementary system. In *2020 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1517–1522, 2020.

[19] Yu Zhang, Xuechun Liang, and Xuebin Lü. Short term water level prediction based on c-stacking ensemble model. In *2021 7th International Conference on Hydraulic and Civil Engineering Smart Water Conservancy and Intelligent Disaster Reduction Forum (ICHCE SWIDR)*, pages 116–121, 2021.

[20] Wenjing Zu, Peng Li, Shiqian Wang, Huixuan Li, Yihan Zhang, Hongkai Zhang, and Yida Du. Research on distributed energy virtual power plant based on spatial prediction model. In *2022 International Conference on 3D Immersion, Interaction and Multi-sensory Experiences (ICDIIME)*, pages 60–64, 2022.