Review Paper Submission

Project title – Electrical Grid Stability

Course Code: INT354

Course Title: **Machine Learning - I**

Submitted To

Dr. Ankita Wadhawan

Id - 23891

Submitted by

Name - T K SANTHOSH RAO

Roll No - 17

Id - 12219295

Sec - KM067

Abstract—

Electrical grid stability is a modern method of collecting the grid data of all the electric meters in each and every point. As the dataset is indicating that this dataset is all about the various parameters and outcomes of the electricity and the includes voltage levels, frequency deviations, power factor, and any other relevant stability indicators. In a smart grid system, consumer demand data is gathered and analyzed centrally in real-time, considering the current supply conditions. By the way, proposed pricing information is back to customers, to give a bigger opportunity them to make informed decisions about their energy usage. The main task of our project is to predict the stability accurately from the past data of the smart grids of the houses To do all this process manually requires a lot of time so adding Any A.I model to this project makes sense and saves a lot of time gives the results so much accuratly. The Dataset was retrieved from the website namely UCI Machine Learning Repository. The Datasetset size is around 10000 attributes and 14 features which indicates a huge and big dataset. Data prepressing was done in different stages and like handling null values dropping the rows etc. which makes the dataset more cleaned. Exploratory Data Analysis (EDA) is also a key step used in this project. Then after continued with working on different types of Classification models which gave the different accuracies like Logistic Regression's accuracy is 99.6 percent, even applied perceptron model with Testing accuracy score of 98.9 percentage then after worked with Support Vector Machine(SVM) algo with 99.7 of testing accuracy continued with Knearest neighbour(KNN) algo with Testing accuracy score of 96.6 percentage which a bit lower tan all of the above. But after fine tuning the parameters of Random Forest(RF) algo achieved an immense accuracy score which is 99.99 percent which is the best at all the accuracies.

Keyword's – **Classification, stability prediction, Random Forests(RF), Support Vector Classifier (SVC), KNearest Neighbour Algorithm(KNN), Logistic Regression, Decentralized Smart Grid Control(DSGC),**

Introduction

Electricity is a major problem of modern technologies, powering everything and everywhere from homes to industries. However, traditional power grids are facing unknown challenges due to problems such as fossil fuel depletion and climate change. The integration of renewable energy sources like wind and solar power is decreasing day-by-day the energy landscape, presenting both opportunities and complexities. Enter smart grid technology, a cutting-edge solution designed to transform the way electricity is generated, transmitted, and consumed. Smart grids leverage advanced technologies such as information communication, computational intelligence, and bidirectional energy flows to enhance stability, reliability, and efficiency. By enabling features like micogrids, electric vehicles, and energy storage, smart grids facilitate the working model of renewble energy sources while supporting market enabling for efficient energy use and economic welfare. But, also ensuring the stability and security of smart grids poses significant challenges. Conventional stability analsis and control approaches may fall short in addrssing the riks managements of modern energy systems. This is where artificial intelligence (AI) steps in, offering great tools to enhance the security and stability of smart grids. With AI algorithms and techniques, smart grids can better perform the rebuilting of renewable energy sources and meet the increasing demands of modern energy systems. As smart grid technology continues to advance, AI will play a vital role in shaping the future of sustainable and efficient energy distribution.

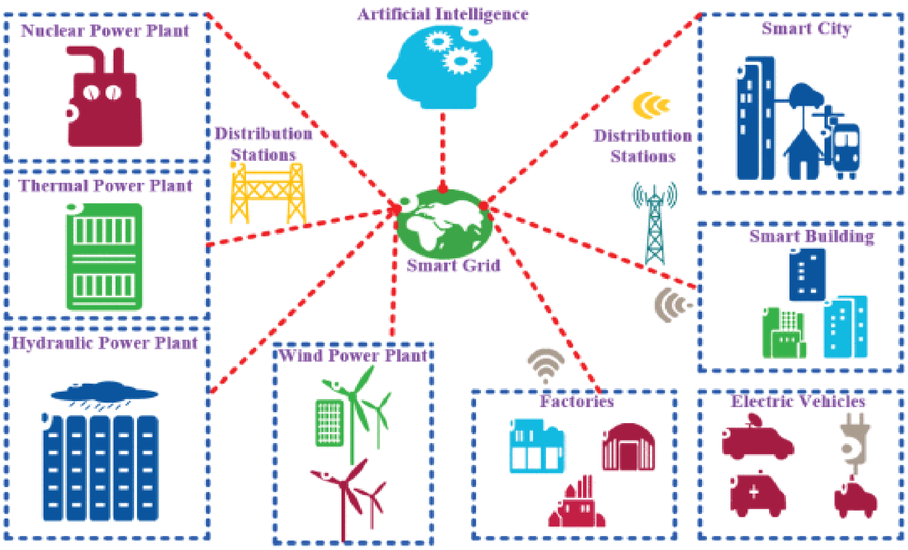

Fig 1 : working of AI with Smart grids

In the figure number 1 also we indicated how the smart grid will work with AI giving the best predictions as how much power is required for the required source. The features of smart grids are vital for their operation, especially focusing on secure and stable performance. But also, these advancements have made smart grid stability more complex compared to traditional power grids, giving the conventional analysis methods not so correctly. Fortunately, artificial intelligence (AI) has emerged as a great addressing solution to address these challenges, making better its calculation power and data-driven insights. AI has seen significant growth since its birth in the human environment, becoming a key player across various industries and domains. In the feild of smart grids, AI innovations have been instrumental in ensuring stability, security, and reliability. Research efforts have explored AI techniques such as machine learning, deep learning, and reinforcement learning for stability analysis and control, and also getting impressive results. However, challenges persist, including the need for robust data, addressing imbalanced learning, ensuring interpretability of AI models, and reducing the issues to communication issues and many diving attacks. This paper aims to review already done research and discuss these open challenges to provide valuable ideas or methodologies for future developments in smart grid stability and AI applications.

The main and important goal of this paper is not only to introduce a model for predicting the stability of a DSGC system but also to get the most effective approach or idea for developing this model. This process involves finding or analyzing different methods used in model generation and researching four models to see which one performs the best. This project is important because it gives a way for developing models on more advanced datasets, which could include simulated datasets based on real user datasets or real data collected from actual DSG networks.

Rest of the paper structured as follows:

- Describe literature review of the research papers
- Describes about proposed methodologies.
- Describes experimental analysis.
- Gives a conclusion and future scope.

Literature Review

In 2023 Borna Franovi´c, Sandi Baressi Šegota, and Nikola Andeli´ worked on the smart Electric grid stability dataset with 10000 attributes and 14 features they worked together with Multilayer Perceptron with hyperparameters as ‘hidden_layer_sizes’, ‘activation’, ‘alpha’ ‘beta_1’, ‘epsilon’ etc. which got an accuracy of 99%

In 2022 Quoc-Viet Pham and Somayaji Siva Rama Krishnan was worked on Grid stability dataset and they proposed Multi -Dimensional LSTM Model and RNN models with achievement of 99.07 accuracy and 0.02 loss. They also used n number of metrices for the model evaluvaltion like precision, f1 score, recall etc.

In 2020 Zhongtuo Shi, Wei Yao, Zhouping Li and Yifan Zhao worked with grid stability dataset with Grdient boosting machine which works with Regression model and applied LDA and worked with preprocessing they also added a hyperparameters like ‘eta’, ’max_depth’, ‘min_child_weight’, and achieved the score as 98.2 % .

In 2017 , X.; Shuai, Tu, and ; Jiang worked on Big data issues in smart grid dataset which was enormous dataset which contains around 50000 attributes and 64 feature they also worked on the classification model Support Vector Machine(SVM) with hyperparameters as “kernal”, “C”, “degree”, “gamma” etc.. they got the accuracy score as 98.6%.

In 2018, N.; Lorencin worked on the stable security dataset with the models like patter Recognition and feed forwarding networks who achived an very huge accuracy score as 99.78% feeding many hypererd parameters to models and also worked on the penalty of the model as the main sources of the model which gave the good accuracy of the model on the dataset.

In 2023 Breviglieri, P was worked on sample properties of simulations using Lating hypercube sampling as he on the dataset . Smart Grid Stability used a regression model and well tuned used the Genetic Programming (GP) for getting the good f1 score and also he achievd around 92.0 as f1 score.

Table 1:- Summarized Table of Literature Review

Authors	Dataset	MODEL	Accuracy of f1 score	Model Evaluation
Franovi ´c, Sandi Baressi Šegota, Nikola Andeli ´	Electric grid stability	Multilayer Perceptron	99%	classification
Quoc-Viet Pham Somayaji Siva Rama Krishnan	Grid stability	Multi -Dimensional LSTM Model and RNN	99.07 f1 score and 0.02 loss.	regression
X.; Shuai, Tu	Grid stability and security	Grdient boosting machine which	98.6	regression
N.; Lorencin	stable security	Support Vector Machine(SVM)	99.78	classfication
Breviglieri, P	Smart Grid Stability	Genetic Programming (GP)	92.0	regression

Proposed Methodology

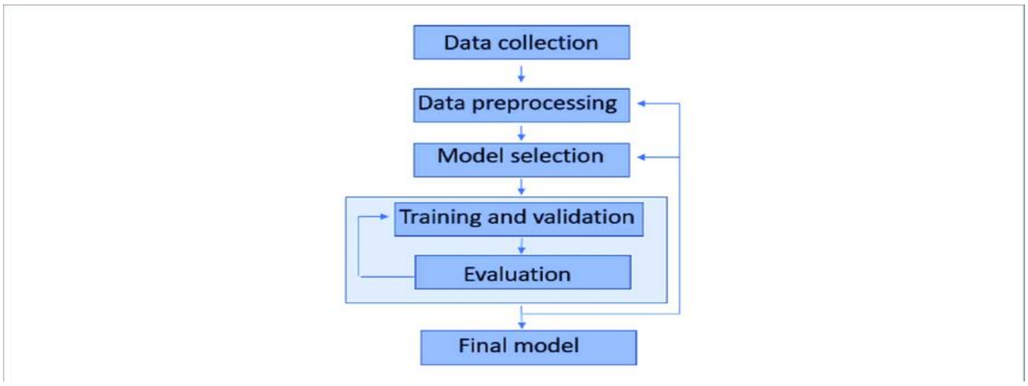

Fig 2 : Flowchart of the methodology

Dataset -

In this project we had worked with electrical grid stability simulated data Dataset which consist of 10000 Attributes and 14 features which was publicly available in the Kaggle Website **.** we splitted this dataset in the ratio of (70: 30) which one is globally uses the dataset consists of the columns with column names as tau, tau2, tau3, tau4, p1, p2, p3, p4, g1, g2, g3, g4, stab, stabf the last stabf column is considered as the target variable and remaining all other are the input variables. The Target variable consists of 2 unique classes Stable and Unstable in the categorical form and this dataset doesn't consists of null values so there is no requirement of handling of the null values. Classification and Regression both can be used on this dataset.

Data Preprocessing -

In this project we had worked with N number of data preprocessing techniques like using Label Encoding which converted the categorical values to numerical values with this label encoding we converted "unstable" and "stable" to "0" and "1" . In this dataset we also worked with null values handling and even dropped a row which consists of null values. Plotted Boxplot for the outlier checking then we come know that there are no outliers in the dataset the used splitted that dataset into training and testing of input and target variable datasets as above mentioned into 70:30 ratio. We had used many libraries in this projects like some are numpy, pandas, matplotlib, saborn, metrics etc... We also plotted a piechart to check wethear the data is balanced or not and correlation matrix for the dependencies of the features on the target variable as a part of data preprocessing for the model .

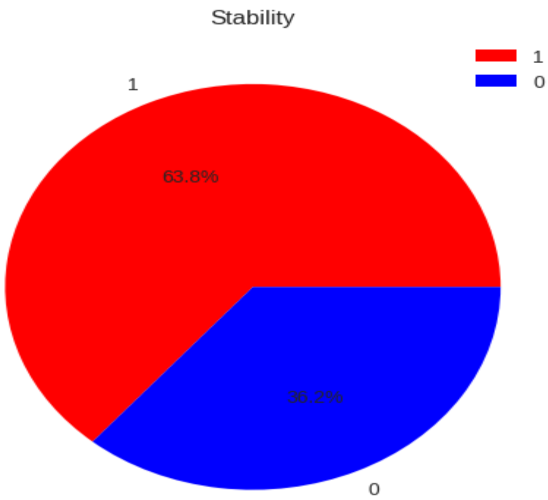

Fig 3 :Plotted Pie chart of the target variable

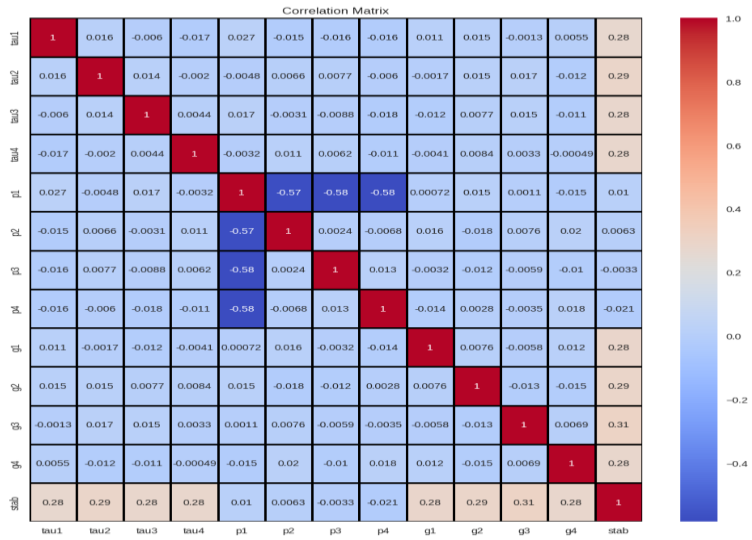

Fig 4 :Plotted Correlation matrix of the dataset

Model Training -

Logistic Regression :

Logistic regression is a commonly used statistical method for solving problems of binary classification, i.e., when the output variable can have only one or two values, usually 0 and 1. The fundamental idea of logistic regression is comparatively straightforward based on a logistic function, it calculates the probable fact that the presented data belongs to one of the two classes . The transformed output of a linear aggregation of input data and their coefficients is combined in a logistic function that generates a value between 0 and 1, representing the probability of having the presented input in a positive class . The parameters of logistic regression are the coefficients of the input variables while the model is trained by the machine learning algorithm to fit the optimal values that minimize the difference between the predicted probabilities and actual labels of items . The performance of this model is usually calculated by measures including accuracy, precision, recall and f1 etc.. After training this model I was achieved around 99.6 accuracy score.

Perceptron :

A perceptron is a binary classifier based on neural networks. At the input are the signs of the object, they are multiplied by the corresponding weights, and then the resulting products are added. This amount is passed through the activation function, which is the step function. It returns a 0 or 1: which class the case belongs to. During the learning process, the perceptron corrects the weights depending on the error between the predicted class and the correct one. Also, in the process of adjusting the weights, it uses the learning rate parameter, which affects the step of adjustments. This process is done until the model reaches the required accuracy level, a satisfactory classification result. This model is suitable when we have linearly separable data when the two classes form an easily separable plane. After training this model I was achieved around 98.9 percent of accuracy score.

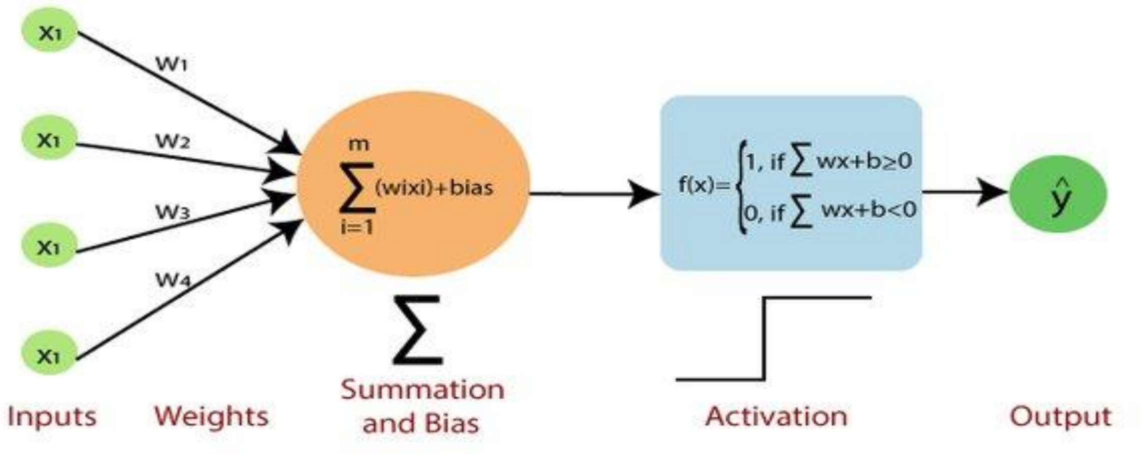

Fig 4 : working of Perceptron Model

Random Forests (Decision tree Classifier) :

A Random Forest is a robust machine learning algorithm commonly used in classification tasks. This model creates many decision trees during training and outputs the mode of the classes in classification for a certain feature, or the mean predictions for the trees in regression. Individual trees are trained on a random subset of the data, with the final prediction being the aggregate of all the trees. Such an ensemble approach ensures that the final model is not only accurate but also less likely to over-fit. Random Forest is a flexible model: you can vary the number of trees in the forest, the maximum depth of each tree, and the number of features allowed for a split at each decision node. It is also known to be noise-immune and effective in the presence of large high-dimensional data. After training this model I was achieved around 100 percent of accuracy score.

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2$$

Fig 5 : Formula to find Gini value in the Perceptron Model

Support Vector Classifier (SVC) :

The Third algorithm I conducted is the Support Vector Classifier which is a type of classification algorithm. It aims to determine the optimal decision boundary to separate each class within the data space. Its operational principle depends on identifying the optimal hyperplane that achieves the maximum margin between classes. The margin is defined as the distance between the closest members of each class and the hyperplane. The important parameters in SVC include the type of kernel function and the value of regularization C, which is equivalent to the trade-off approach or the weighting factor between the margin of separation and classification error minimization. In conclusion, SVC serves as a reputable classifier in machine learning and may be applied in various machine learning tasks due to its capacity to define complex decision boundaries and resistance to overfitting. After training this model I was achieved around 99.7 percent of accuracy score.

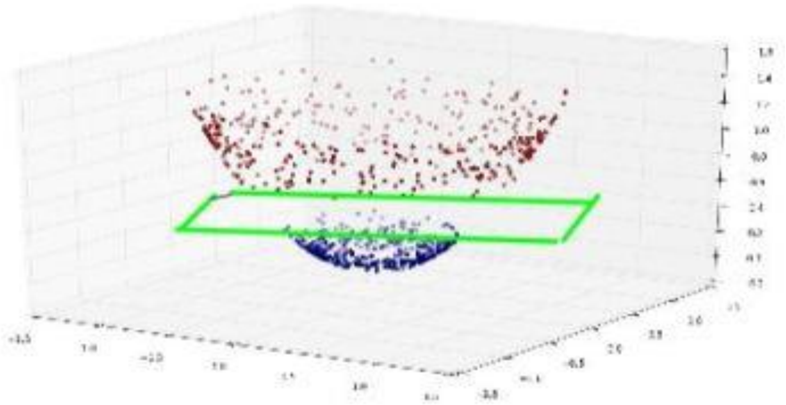

Fig 6 : Diagrammatic Representation of Hyperplane Model

KNearest Neighbour (KNN) :-

K-Nearest Neighbors is a widely used simple yet effective algorithm for classification problems. The idea behind its work is clear: upon receiving a new data point, KNN finds K points from the training set that are situated closest to the new point, based on some distance metric. These points are referred to as neighbors. The class of the new point is chosen based on the majority class of its neighbors. The algorithm's main parameter is K—the number of neighboring points to put into consideration. Choosing the right value for K matters: a small K makes predictions noisy, while a large K will not allow learning local patterns and help in smoothing decision boundaries. Moreover, the choice of the distance metric matters as well. The most commonly used is Euclidean distance. After training this model I was achieved around 96.6 percent of accuracy score.

$$Manhattan\ Distance = \sum_{i=1}^{d} |x_{1i} - x_{2i}|$$

$$Euclidean\ Distance = \left(\sum_{i=1}^{d} (x_{1i} - x_{2i})^2\right)^{\frac{1}{2}}$$

$$Minkowski\ Distance = \left(\sum_{i=1}^{d} |x_{1i} - x_{2i}|^p\right)^{\frac{1}{p}}$$

Fig 7 : Formulas of Distances of all KNN Model

Experimental Analysis

Table 2 :- Summarized Table of Tuned Parameters and accuracy scores of all models

Classification		
AI Algorithm	Average Score	Chosen Hyperparameters
Logistic Regression	99.6	'C' = [0.001, 0.01, 0.1, 1, 10, 100], 'penalty' = ['l1', 'l2'], 'solver' = ['liblinear', 'saga']
Perceptron	98.9	hidden_layer_sizes = (50,50,50,50,50),activation = logistic
Random Forest	100	n_estimators = [50, 100, 150], max_depth =[6, 8,10,20] , min_samples_split = [4, 6, 8], min_samples_leaf = [2, 4],
Support vector Classifier(SVC)	99.7	C = [0.1, 1, 10,15] , kernel = ['linear', 'rbf', 'poly'], gamma = ['auto', 'scale']
KNearst Neighbour(KNN)	96.6	n_neighbors = np.arange(1,100,2) , weights = "uniform", "distance"], metric = ["euclidean","manhattan"]

In Table-2 we are trying to include all the parameters used in all the models which makes us to to improve our accuracy. We also plotted the bar the Bar graphs to show the results of all the models like f1 score, precision, recall, and accuracy score. Due to the Hyperparameter tuning method used in the model named GridSearch which commonly includes cross validation as a parameter to interate n number of times of dividing the dataset into the n number of folds and analyzing all the accuracies and considering the best parameters using the ‘best_parameters’ .

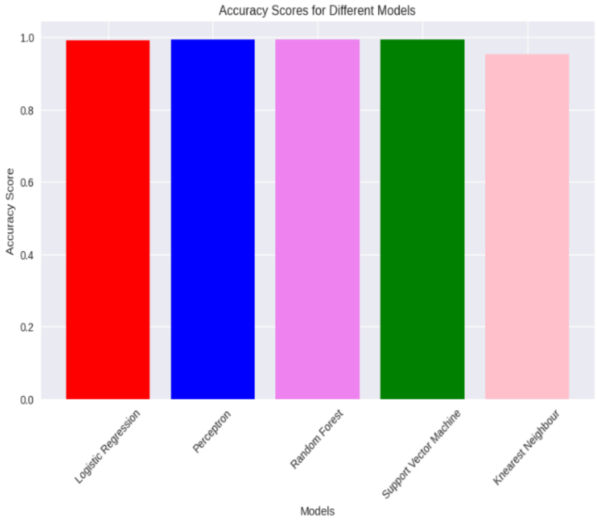

Fig 8 :Barplot of all models Accuracy Scores

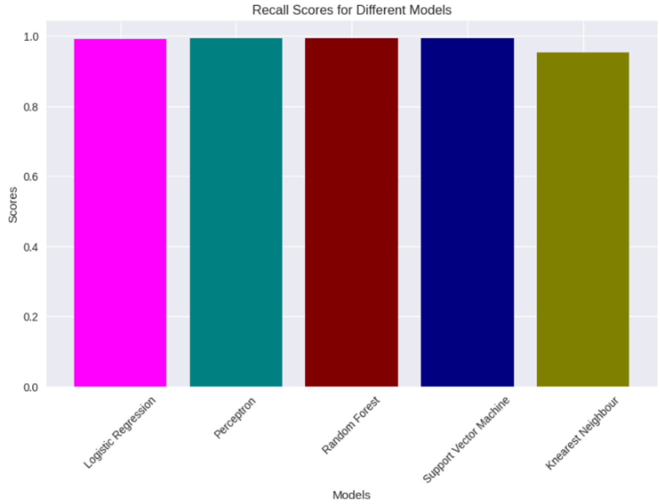

Fig 9 : Barplot of all models Recall Scores

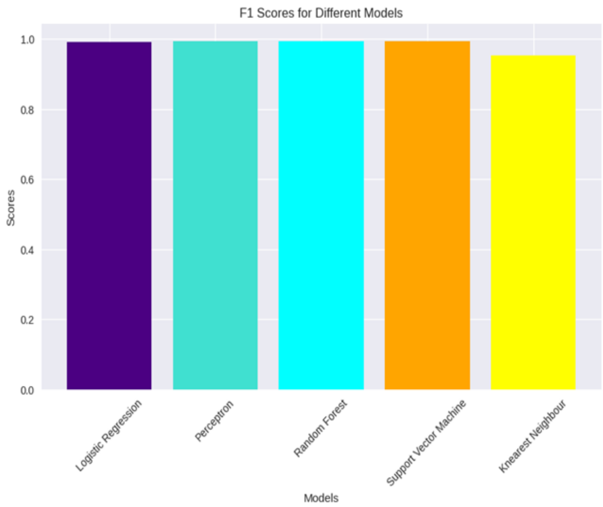

Fig 10 : Barplot of all models F1 Scores

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$Precision = \frac{T_p}{T_p + F_p}$$

$$Recall = \frac{T_p}{T_p + T_n}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Fig 11 : Formulas of all metrices used in the project

In Fig 8 the Bar graph represents the accuracy scores of all models like Logistic Regression,perceptron, Random Forests, SVM,KNN . In Fig 9 the Bar graph represents the Recall scores of all models like Logistic Regression,perceptron, Random Forests, SVM,KNN. In Fig 10 the Bar graph represents the F1 scores of all models like Logistic Regression,perceptron, Random Forests, SVM,KNN .In Fig 11 image represents the formulas of the metrics where TN means TrueNegatives ,TP means TruePositives, FN means FalseNegatives,and FP means FalsePositives In this way the model will calculate the metrics values of all the metrics in our project.

Conclusion and Future Scope

Electrical grid stability simulated data is a dataset which we was worked upon and applied many preprocessing techniques like handling null values, converted from categorical to the numerical data, handled outliers and plotted boxplot and correlation matrix of the feature scaling, data analysis and model training.we also worked with hyperparameter tuning named as GridSearch which gives our project boosting in the accuracy score. we also even worked with many Evaluvation metrics which used to evaluvate the models working prediction like accuracy score,F1 Score, Recall, Precision Score, etc.. If I conclude the project with the Random Forest a decision tree model achieved the highest accuracy as 100 percentage of the testing data using some specific tuned parameters.It means that Model is predicting wether the target is stable or unstable very much accurately. I also plotted many to visualize the data and the accuracy of the model.

Future Scope—

- **Advanced Machine Learning Techniques**: Further developments of machine learning algorithms, commonly known as deep learning, reinforcement learning, and ensemble methods, might be used to increase the accuracy and efficiency of grid stability prognostic models. They are designed to manage high-dimensional, complex data and, thus, recognize and perfectly catch nonlinear relationships than traditional approaches .

- **Integration of Big Data and IoT**: Big data analytics and the Internet of Things may be discovered and used to gather and analyze huge amounts of real-time data from the various sensors, smart meters, and other sources placed in the grid composition. Machine learning-based models can then cultivate this data to identify regularities, problems, and prognostic insights to boost grid stability.

- **Preparing a working Model :** After getting accurate predictions this project can also ready to work with real time projects like it can accurately predict the stability of and the required resources of a home or an industries etc.. sources..

- **Auto voltage controller :** If the gets into working in future smart auto grids will auto increase or decrease the voltages or make power stable with the help of our A.I Model.

References

- Bashir, Ali Kashif, Suleman Khan, B. Prabadevi, N. Deepa, Waleed S. Alnumay, Thippa Reddy Gadekallu, and Praveen Kumar Reddy Maddikunta. "Comparative analysis of machine learning algorithms for prediction of smart grid stability." *International Transactions on Electrical Energy Systems* 31, no. 9 (2021): e12706.

- Franović, Borna, Sandi Baressi Šegota, Nikola Anđelić, and Zlatan Car. "Decentralized smart grid stability modeling with machine learning." *Energies* 16, no. 22 (2023): 7562.

- Shi, Zhongtuo, Wei Yao, Zhouping Li, Lingkang Zeng, Yifan Zhao, Runfeng Zhang, Yong Tang, and Jinyu Wen. "Artificial intelligence techniques for stability analysis and control in smart grids: Methodologies, applications, challenges and future directions." *Applied Energy* 278 (2020): 115733.

- Alazab, Mamoun, Suleman Khan, Somayaji Siva Rama Krishnan, Quoc-Viet Pham, M. Praveen Kumar Reddy, and Thippa Reddy Gadekallu. "A multidirectional LSTM model for predicting the stability of a smart grid." *IEEE Access* 8 (2020): 85454-85463.

- Hasan, Md Rokibul. "Revitalizing the Electric Grid: A Machine Learning Paradigm for Ensuring Stability in the USA." *Journal of Computer Science and Technology Studies* 6, no. 1 (2024): 141-154.

- Alsirhani, Amjad, Mohammed Mujib Alshahrani, Abdulwahab Abukwaik, Ahmed I. Taloba, Rasha M. Abd El-Aziz, and Mostafa Salem. "A novel approach to predicting the stability of the smart grid utilizing MLP-ELM technique." *Alexandria Engineering Journal* 74 (2023): 495-508.

- Önder, Mithat, Muhsin Ugur Dogan, and Kemal Polat. "Classification of smart grid stability prediction using cascade machine learning methods and the internet of things in smart grid." *Neural Computing and Applications* 35, no. 24 (2023): 17851-17869.

- Risco, Adriana Babilonia, Renzo Iván González Salinas, Alhiet Orbegoso Guerrero, and Daniel Leonardo Barrera Esparta. "IoT-based SCADA system for smart grid stability monitoring using machine learning algorithms." In *2021 IEEE XXVIII International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1-4. IEEE, 2021.

- Ghosh, Ankit, and ALOK KOLE. "A comparative analysis of enhanced machine learning algorithms for smart grid stability prediction." *Authorea Preprints* (2023).

- Vu, Thanh Long, and Konstantin Turitsyn. "A framework for robust assessment of power grid stability and resiliency." *IEEE Transactions on Automatic Control* 62, no. 3 (2016): 1165-1177.

- You, Shutang, Yinfeng Zhao, Mirka Mandich, Yi Cui, Hongyu Li, Huangqing Xiao, Summer Fabus et al. "A review on artificial intelligence for grid stability assessment." In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1-6. IEEE, 2020.

- Liu, Wentao, Tamas Kerekes, Tomislav Dragicevic, and Remus Teodorescu. "Review of grid stability assessment based on AI and a new concept of converter-dominated power system state of stability assessment." *IEEE Journal of Emerging and Selected Topics in Industrial Electronics* 4, no. 3 (2023): 928-938.

- Aygul, Kemal, Mostafa Mohammadpourfard, Mert Kesici, Fatih Kucuktezcan, and Istemihan Genc. "Benchmark of machine learning algorithms on transient stability prediction in renewable rich power grids under cyber-attacks." *Internet of Things* 25 (2024): 101012.

- Rashed, Muhammad, Iqbal Gondal, Joarder Kamruzzaman, and Syed Islam. "Assessing Reliability of Smart Grid Against Cyberattacks using Stability Index." In *2021 31st Australasian Universities Power Engineering Conference (AUPEC)*, pp. 1-6. IEEE, 2021.