SMART GRID STABILITY PREDICTION AND EVALUATION USING MODERN MACHINE LEARNING ALGORITHMS

A PROJECT REPORT

Submitted by

SAFNA S

TKM20EEPS13

to

the **APJ Abdul Kalam Technological University**

in partial fulfillment of the requirements for the award of the

Degree of

MASTER OF TECHNOLOGY

in

POWER SYSTEMS

DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

T.K.M COLLEGE OF ENGINEERING

KOLLAM-5

2020 – 2022

DEPARTMENT OF ELECTRICAL & ELECTRONICS ENGINEERING

THANGAL KUNJU MUSALIAR COLLEGE OF ENGINEERING

KOLLAM

CERTIFICATE

This is to certify that the project report entitled "**SMART GRID STABILITY PREDICTION AND EVALUATION USING MODERN MACHINE LEARNING ALGORITHMS**" submitted by **Ms. SAFNA S** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirement for the award of Degree of Master of Technology in power Systems. Electrical & Electronics Engineering is a bonafide record of project work carried out by her under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Prof. VIKI PRASAD

Assistant Professor [Internal Supervisor]

Department of Electrical and Electronics

(EXTERNAL EXAMINER)

Prof. SHANAVAS T N

Associate Professor [PG Coordinator]

Department of Electrical and Electronics

Dr. SABEENA BEEVI K

Associate Professor [HoD]

Department of Electrical and Electronics

ACKNOWLEDGEMENT

I would like to acknowledge the faith granted in me through the unmerited favor poured on me by **God Almighty** who keeps me as target in every step to the progress of the project.

It is my privilege and pleasure to express my gratitude and indebtedness to **Dr.T Shahul Hameed**, Principal of TKM College of Engineering, and **Dr. Sabeena Beevi K**, Head of the Department, Department of Electrical and Electronics Engineering for providing all the required resources for the completion of my project work.

I am greatly obliged to **Prof. Shanavas T N**, Associate Professor, PG Coordinator, Department of Electrical and Electronics Engineering for his encouragement and support.

I express my sincere gratitude to **Prof. Viki Prasad**, Asst. Professor, Project Guide, Department of Electrical and Electronics Engineering for her constructive guidance and advices.

I express my sincere gratitude to **Prof. Jibi P Mathew**, Asst. Professor, project Coordinator, Department of Electrical and Electronics Engineering for all the help and coordination in completing this project successfully and on time.

I will be failing in duty if I do not acknowledge the authors of the references and other pieces of literature referred to in this project

I show my extreme gratitude to all Faculty member and Technical staffs in Electrical and Electronics Dept. for providing all the help and necessary facilities to present this project and my deep hearted cheers to my parents and all my friends who extended their support and co-operation towards the successful presentation of the project.

SAFNA S

ABSTRACT

The global demand for energy is rapidly rising. As a result, energy systems must evolve and be upgraded in order to become more efficient, adaptable, and sustainable. A smart grid reduces workforce while providing consumers with safe, reliable, high-quality, and long-lasting electricity. Smart grids use digital communication technology to enable two-way flow of electricity and data. This vast amount of data needs to be processed for making better decisions for maintaining the grid stability. ML and AI approaches are utilised to acquire, store, and manage this data. This study compares the performance of modern machine learning methods for predicting smart grid stability. The dataset that was chosen contains findings from a smart grid simulation. XGBoost, Adaboost, Gradient Boosting Method (GBM), HistGBM, LightGBM and CatBoost algorithms have been implemented to forecast smart grid stability. Performance of the ML model has been evaluated based on Classification evaluation metrics. The following evaluation metrics such as accuracy, precision, recall, F1-score, MCC, specificity, training time, predicting time, AUC-ROC curve and AUC-PR CURVE are used for classification evaluation.An efficient Stacking Ensemble Classifier (SEC) model developed by using the above mentioned machine learning algorithms and compared the evaluated results of individual machine learning models with the SEC model.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AI	Artificial Intelligence
AUC	Area Under The Curve
CNN	Convolutional Neural Network
DL	Deep Learning
DSA	Dynamic Security Assessment
DT	Decision Tree
EAC	Equal Area Criterion
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FSA	Frequency Stability Assessment
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
HGB	Hist Gradient Boost
HISTGBM	Hist Gradient Boosting Model
HVDC	High Voltage Direct Current
ICT	Information And Communication Technologies
IM	Induction Machine
LGBM	Light Gradient Boosting Model
LSTM	Long Short Term Memory
MCC	Matthews Correlation Coefficient
ML	Machine Learning
MLSTM	Multiplicative Long Short Term Memory
OS-ELM	Online Sequential
PMU	Phasor Measurement Unit
PR	Precision –Recall
PT	Predicting Time
RF	Random Forest

RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristics
SEC	Stacking Ensemble Classifier
SI	Stability Index
SVM	State Vector Model
TDS	Time Domain Simulation
TEF	Transient Energy Function
TN	True Negative
TP	True Positive
TPR	True Positive Rate
TT	Training Time
UCI	University Of California
VSA	Voltage Stability Assessment
WAMS	Wide Area Measurement System
WPP	Wind Power Penetration
XGBOOST	Extreme Gradient Boost

Chapter 1

INTRODUCTION

1.1 GENERAL BACKGROUND

As the need for new stages in the power sector emerges, the necessity of fitting in cannot be overstated. Grids have been transmitting power for a long time, mainly from non-renewable energy sources including coal, oil, and natural gas. Because this power is generated in a scheduled and centralised manner, maintain grid stability is a very tough task for power system engineers in most of the time. Unlike traditional grids, which rely on synchronous generators to maintain stability, renewable energy grids require a lot more effort to maintain stability in the event of outages. Traditional power grids are demand-driven and have a hierarchical structure, with essentially no storage capabilities. The voltage in an energy network is gradually reduced so that electricity can be consumed by these various consumers: voltage levels ranging from transmission through distribution to service voltage levels. Maintaining the power system stability by using conventional grid is difficult. Because they consume more time and have less efficiency for predicting the stability of the system. But the emergence of smart grid made easy for managing the power system.

Smart grid is the type of conventional grids but they are integrated with information and digital communication systems. A smart grid is an electricity network that allows for a two-way flow of electricity and data, as well as the detection, reaction, and prevention of changes in usage and other issues, using digital communications technology. Smart grids are self-healing and allow power users to have an active role in the system. Self-healing capabilities in smart grids enable automatic detection and response to grid problems, as well as speedy recovery following disturbances.

Smart meters are the heart of smart grid. Smart meters help for two way communication in grid. A smart metre is a new type of metering equipment that may digitally send your real energy usage to your utilities supplier for their electricity usage. This implies that consumers will no longer have to rely on anticipated energy bills or enable metre readers to manually read their metres in their homes or yards. Customers can also obtain near real-time data about their own energy usage through the smart metre, allowing them to better understand, evaluate, and manage their energy use, including reducing energy waste and, as a result, the customers can easily manage their electricity usage during peak hours.

AI and ML can help in an improved understanding of consumer behaviour allowing suppliers to gain access to their exact electricity requirements. This will enable generation of the right type of billing information as well. The integration of AI and ML with smart grids will help consumers gain access to their energy consumption and pricing data so that they can actively respond to the requests of cutting down energy consumption during the peak hours of energy demand. This in turn will result in an enhanced operational efficiency of smart grids. The incorporation of Information and Communication Technologies (ICTs) in smart grids allows both consumers and producers to become active agents in maintaining the intended functioning of the smart grid.

1.2 MOTIVATION

The main motive of conducting stability analysis of a smart grid is to have a keen understanding about the operation of the power system and to improve the system performance by arranging the proper maintenance at the critical areas of both consumer and producer side power supply. Different methods are employed for analysing the stability of a smart grid.

Machine learning models used in the AI techniques of smart grid has an important role in maintaining the stability of a smart grid. Modern machine learning algorithms have large amount of data handling capability and they take only less time for predicting the stability of the system than conventional machine learning methods like bagging methods. Modern machine learning algorithms like boosting algorithms will help the smart grid to predict the stability of the system as soon as possible. So we can take apt decisions for maintaining the stability of the system quickly. Stability analysis of a smart grid by using different modern machine learning algorithms have greater significance, because through this study we can find the best machine learning model for the smart grid stability analysis.

1.3 OBJECTIVES OF THE THESIS

In this thesis, the stability analysis of a smart grid is analysed with the help of different modern machine learning algorithms such as gradient boost, XG boost, Ada boost, HistGBM, Catboost, LightGBM and a stacking ensemble classifier. Stability of a smart grid is predicted by using these machine learning models and analysed the correctness of the predictions using classification evaluation metrics such as accuracy, Precision, Recall,

F-1 score, specificity, MCC, predicting time, training time, AUC-ROC curve and AUC-PR curve.

1.4 ORGANISATION OF THE THESIS

The entire thesis is organized as follows. It consists of seven chapters. Chapter 1 is a brief introduction of the thesis, motivation and objectives of the thesis. Chapter 2 deals with the literature review about the several conventional methods of smart grid stability analysis and the technical difference between each of them. Chapter 3 gives idea about different kinds of smart grid stability types .This thesis follows two methodologies one is analysing the stability of a smart grid by using individual modern machine learning algorithms and second methodology is to create a new model combining with the individual models to create a stacking ensemble machine learning algorithm, these all discussed in chapter 4. Chapter 5 gives the idea about the data set used for analysis is under the title case study. Chapter 6 discusses the results obtained during the classification. Finally, chapter 7 gives the conclusion and future research scopes in this area.

Chapter 2

LITERATURE SURVEY

[1] C. Whitefield, A. A. Cardenas, H. Chen, P. Popovski and V. W. S. Wong, "Smart Grids", ***IEEE Wireless Communications*****, vol. 24, no. 2, pp. 8-9, April 2017.**

The global energy demand is rapidly increasing. Therefore, it is necessary to evolve and upgrade the energy systems in order to make them more efficient, flexible and sustainable. The integration of information and digital communication technologies with the conventional power grid systems is known as the smart grid. A smart grid can be defined as a network of self-sufficient systems that enable the integration of power generation sources (both conventional and non-conventional) to the electrical grid causing the reduction of workforce and offering safe, reliable, high quality and sustainable electricity to the consumers

[2] R. Morello, C. De Capua, G. Fulco and S. C. Mukhopadhyay, "A Smart Power Meter to Monitor Energy Flow in Smart Grids: The Role of Advanced Sensing and IOT in the Electric Grid of the Future", ***IEEE Sensors Journal*****, vol. 17, no. 23, pp. 7828-7837, Dec. 2017.**

Smart grids enable two-way flow of electricity and data with the aid of digital communication systems which include smart meters, smart appliances, renewable energy resources and energy efficient resources. The two way communication between the producers and the consumers of electricity and the bidirectional power flow enhances the security, reliability and the efficiency of power systems.

[3] B. K. Bose, "Artificial Intelligence Techniques in Smart Grid and Renewable Energy Systems—Some Example Applications", ***Proceedings of the IEEE*****, vol. 105, no. 11, pp. 2262-2273, Nov. 2017**.

The smart grids are interconnected and integrated with advanced metering infrastructure, control, and communication technologies. This leads to the generation of an enormous amount of high dimensional and multi-type data. Such massive amounts of data require collection, storage and management. The existing technologies have many limitations in processing the data. Therefore, it is necessary to implement Artificial Intelligence (AI) techniques in smart grid applications. AI techniques provide an efficient way to analyse data, learn from it and then make appropriate decisions to ensure that the performance of the grid is as intended. AI has the potential to revolutionize the energy industry. AI can act as the brain of the smart grid by continuously collecting and synthesizing overwhelming amounts of data from large number of smart sensors and by making timely decisions to enhance the stability and reliability of the smart grid. AI has an important role to play in making use of the large amount of potentially useful data and generating actionable insights from it. Conventional techniques involve huge and complex data, thereby increasing the computational time and in some cases reduction in accuracy as well, a challenge that can easily be addressed using AI and machine learning (ML)

[4] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, ''A data-driven design for fault detection of wind turbines using random forests and Boost,'' ***IEEE Access*****, vol. 6, pp. 21020–21031, 2018.**

Wind energy has come a long way in the last ten years. The cost of operating a wind turbine is highly dependent on component failure and repair rates, whereas fault detection and isolation will greatly enhance the availability and reliability aspects. The data-driven wind turbine failure detection framework is established in this research using an effective machine learning algorithm, random forests (RFs), in combination with extreme gradient boosting (XGBoost). RF is utilised in the proposed approach to prioritise the features, which are either direct sensor signals or generated variables based on past knowledge. XGBoost then trains the ensemble classifier for each specific defect based on the top-ranking features. The proposed approach is demonstrated to be robust to a variety of wind turbine models, including offshore ones, in a variety of working situations. Furthermore, while dealing with multidimensional data, the suggested ensemble classifier protects against over fitting and produces superior wind turbine defect detection results than the support vector machine method.

[5] Y. Chen, S. M. Mazarin, C. Y. Chung, S. O. Faried, and B. C. Pal, ''Rotor angle stability prediction of power systems with high wind power penetration using a stability index vector ***,''IEEETrans.PowerSyst*****., vol.35,no.6, pp. 4632–4643, Nov. 2020.**

It presents a method for predicting the rotor angle stability of power systems in the case of high wind penetration. First, a new stability metric is developed that takes into account the dynamic behaviour of WPPs. Based on findings from EEAC and PMU investigations, a method is proposed in which the established algorithm is used in parallel to determine SIs for all conceivable IMs layouts; SI vectors are then generated and selected as characteristics for rotor angle stability prediction. Ensemble decision trees on two IEEE test systems with varied wind power penetration levels confirm the usefulness of the suggested approach. The collected results and comparisons show that the proposed approach is superior in terms of accuracy, speed, and resilience.

[6] T. Guo and J. V. Milanovic, ''Online identification of power system dynamic signature using PMU measurements and data mining,'' ***IEEE Trans. Power Syst*****., vol. 31, no. 3, pp. 1760–1768, May 2016.**

Using PMU measurements and data mining, this research offered a two-stage process for determining the dynamic signature of a power system online. The dynamic security evaluation for corrective control often concentrates on transient stability status without dealing with the dynamic behaviour of generators in the event of instability, according to the current literature. This paper's methodology fills that need. In the first stage, it use classic binary classification to determine system transient stability, and in the second stage, it employs both clustering and multiclass classification to identify unstable dynamic behaviour.

[7] Shutang You, Yinfeng Zhao, Mirka Mandich," A Review on Artificial Intelligence for Grid Stability Assessment" ***IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids*** **2020.**

Artificial intelligence offers a simple way to measure power grid stability. Artificial intelligence has the potential to reduce time on model creation and numerical computation in stability evaluation when compared to simulation-based alternatives. The first part of this research looked at the existing literature on utilising artificial intelligence to assess power grid reliability. Then, to analyse power grid transient stability, frequency stability, and tiny signal stability, a machine-learning-based method is described and built. The AI tool's accuracy and usefulness in assessing power grid stability were confirmed by test results.

[8]D.Syed,A.Zannab, S.S.Refaat, H.Abu-rub and O.Bouhali,"Smart Grid Big Data Analytics Survey of Technologies, Techniques and Applications", ***IEEE Access*** **2020.**

Since the previous decade, smart grids have been gradually replacing traditional power systems. This transition is linked to the installation of a large number of smart metres and other data extraction units. This opens up a slew of possibilities for the gathered large data. As a result, the triumph of the factor of big data analytics plays a role in the smart grid energy paradigm. This comprises effective acquisition, as well as big data transmission, processing, visualisation, interpretation, and use the paper goes into great detail. The author provides an overview of several big data technologies and examines big data analytics in the context of smart grids.

[9]S.K.AZMAN, Y.J.ISBEIH, M.S.EI MOURSIA and K.Elbasssioni, "A Unified Online Deep Learning Prediction Model for Small Signal and Transient Stability", *IEEETrans.PowerSyst.* vol.35, no.6, pp. 4585–4598, Nov. 2020.

Using voltage phasor data and deep learning techniques, this research proposes an online model for forecasting rotor angle stability. The primary goal of the provided Models will be used to create a unified strategy to forecasting both. When the system is disturbed, small-signal and transient stability are important. This is accomplished by first conditioning a CNN. A model based on a set of transient responses exhibited by the phase voltages across the system, as well as for wide-range applications of the operating environment when the system has been proven to be stable, the Small-signal stability is expected to reveal information about the low-frequency power system's overall dampening oscillations. An LSTM network and the phase voltages provided by the PMUs are used to train the prediction model for small-signal stability. The proposed model's capacity to give a comparatively accurate and considerably faster transient stability classification is demonstrated by the results obtained using the New England 39-Bus, IEEE 68-Bus, 16-Machines, and IEEE 145-Bus, 50-Machines systems. When the system is subjected to a disturbance, our unified method can also give the small-signal stability response across a broad time scale.

[10]H.Hugmar, L.Tong, R.Eriksson,"Voltage Instability Prediction Using a Deep Recurrent Neural Network," *IEEETrans.PowerSyst.*, vol.36, no.1, pp. 17–27, 2020.

This research provides a new method for predicting voltage instability online using an LSTM network that can use a series of observations to increase classification accuracy. The LSTM network, once trained, can allow system operators to continuously assess and anticipate whether the current system state is stable or will soon evolve into an alert or emergency condition. The network has also been modified to be able to detect when instability arises, allowing system administrators to take more cost-effective control actions.

[11]M. Alzab, S. khan, S.S.R. Krishnan, Q.V. Pham, M.P.K Reddy and T.R. Gadekallu, "A Multidirectional LSTM Model For Predicting The Stability Of A Smart Grid", *IEEE Access*, vol. 8, pp. 85454–85463, 2020.

In this paper, a novel MLSTM model for predicting smart grid stability is presented. The suggested model is tested on the UCI Machine Learning Repository's smart grid dataset. Traditional ML models such as LSTM, GRU, and RNN are compared to the performance of MLSTM. The suggested model outperforms the competition in terms of accuracy, Precision, loss, and ROC curve metrics, according to the comparison analysis. When compared to other typical deep learning models, the suggested model scored 99.07 percentage training and testing accuracy, which is three times higher.

Chapter 3

SMART GRID STABILITY

3.1 INTRODUCTION

There are two primary factors that determine a system's stability in a Smart Grid: First, the generation must always keep up with demand and maintain a reserve (battery storage) for sudden power interruptions. Second, the grid must have enough capacity to maintain voltage stability at all points. For the generation, delivery, and consumption of clean, sustainable, efficient, and dependable energy, smart grid is the latest trend. To ensure stable and secure operation is crucial for the smart grid, which demands effective stability analysis and control. The stability characteristics of the smart grid are much more complex than in the past as a result of its evolution through increasing levels of interconnection, increased incorporation of renewable energy, widespread use of direct current power transmission systems, and liberalisation of electricity markets. These modifications have a number of negative effects on speed, efficacy, and economy for standard stability analysis and control methodologies. Instead, new artificial intelligence (AI) approaches offer strong and promising tools for stability analysis and control in smart grids, and they are gaining more and more attention.

Smart grid uses integrated information, two-way, cyber-secure communication technologies, and computational intelligence across electricity generation, transmission, substations, distribution, and consumption. The smart grid offers numerous additional advantages [14, 15] and enables substantial penetration of stochastic and erratic renewable energy for reduced pollution [16]. In smart grids, where people may both buy and sell energy, bidirectional energy flows are conceivable. Additionally, smart grids can be used to charge energy storage systems, electric vehicles, and micro grids. For more effective energy consumption and increased economic welfare, smart grids also need to enable the market.

3.2 STABILITY ANALYSIS IN SMART GRIDS

3.2.1 NECESSITY

The goal of stability analysis is to identify the type of instability and determine if the electrical grid can sustain stability under disturbances. Only if the precise type of instability can be rapidly provided can timely emergency control actions be performed. The complicated transient features of the system make it increasingly challenging for conventional approaches to handle with the growth of interconnected power grids, the incorporation of renewable energy, and the use of HVDC. Currently, the transient energy function (TEF) approach and time domain simulation (TDS) method are frequently employed. The requirements of an online stability study cannot be met by TDS, despite its excellent Precision. To imitate actual power systems, intricate models with precise characteristics should be created. Although the TEF method does not necessitate lengthy computations, it is challenging to construct an energy function for massive power networks. Additionally, the accuracy of stability analysis may be impacted by the TEF's predominant use of second-order classic models, which overlook the complex transient process.

3.2.2. FEASIBILITY

In general, a new stability analysis technique appropriate for the current smart grid development is required. AI technology has recently experienced fast progress, making it a promising answer. An example of a classification problem that can be tackled using supervised learning algorithms is stability analysis. The wide-area measurement system (WAMS) and phasor measurement units (PMUs) deployed in smart grids allow for the collection of large amounts of real-time electrical data with high sample rates over a uniform time period, which can be used as the database for AI techniques. Data-driven approaches, such as ML and DL, can extract features related to utilising strong learning capacities, to the stability of the input data. In addition, sophisticated modelling and knowledge of electricity grids are not necessary for AI techniques.

3.3 TYPES OF SMART GRID STABILITY ASSESSMENT

3.3.1 DYNAMIC SECURITY ASSESSMENT

Both transient and steady-state time periods are taken into account by dynamic security assessment (DSA). The thermal overloading, voltage and frequency changes (both transient and steady state), and all types of stability (transient stability, small-disturbance stability, voltage stability, and frequency stability) must all be met by a dynamic secure system in order to be considered secure [17, 18]. Pre-contingency DSA and post-contingency DSA are two more divisions of DSA that can be made. Pre-contingency DSA assesses the security of the existing operational situation in light of a contingency that is expected but has not yet materialised [19]. Post-contingency DSA, as opposed to pre-contingency DSA, uses dynamic measures like rotor angle and voltage trajectories to forecast system security in the future after an on-going contingency. Emergency measures to stop the spread of insecurity in the linked power system correspond to post-contingency DSA. Post-contingency DSA, as opposed to pre-contingency DSA, uses dynamic measures like rotor angle and voltage trajectories to forecast system security in the future after an on-going contingency. Emergency measures to stop the spread of insecurity in the linked power system correspond to post-contingency DSA.

Traditional simulation techniques for DSA include power flow calculations for voltage constraints or element rating violations, time-domain simulations for transient security and voltage stability, and eigenvalue analyses for small-disturbance stability assessment. These techniques are time-consuming and unsuitable for online application. Since DSA is a common categorization issue, AI is now applied to it thanks to the advancement of WAMS technology and the widespread use of PMUs. A DSA engine, a wind power and load demand forecasting engine, a database generation engine, and a model updating engine make up its structure, all of which are designed to address the influence of wind power. Although the use of AI significantly cuts down the time required for online evaluation, the offline simulation required for labelling samples still takes a lot of time. Stability evaluation or forecast Due to steadily rising load demand, greater use of power electronics equipment, more integration of intermittent renewable energy sources, and the deregulation of the electricity market in smart grids, stability is becoming more and more difficult to maintain in recent years. In order for smart grids to operate securely and steadily, stability assessment, or monitoring and early detection of instability, is crucial. The stability issue can be divided into three categories, according to the IEEE/CIGRE Joint Task Force [18]: rotor angle stability, voltage stability, and frequency stability. Because of its benefits in speed, Precision, and flexibility, AI has been used in all of these fields.

3.3.2 ROTOR ANGLE STABILITY ASSESSMENT

The capacity of synchronous machines in a connected power system to maintain synchronism after being subjected to a disturbance is referred to as "rotor angle stability." Conventional assessment methods (TSA) have limitations for large-disturbance rotor angle stability (i.e., transient stability). Computing power is required for time-domain simulation [20]. Because some state variables are not available, the transient energy functions method is conservative and challenging to employ online [21]. And the classical generator model is the only one for which to extend equal area criterion (EEAC) can be used. ML analyses the relationship between the stability state and the power grid factors. When the necessary measurements are fed into the ML model for an online application, the TSA result is immediately available. Many ML techniques, including shallow ML and deep learning, have been applied in recent years. When the necessary measurements are fed into the ML model for an online application, the TSA result is immediately available. Many ML techniques, including shallow ML and deep learning, have been applied in recent years.

Shallow ML commonly uses manual or algorithmic feature design followed by the construction of a classifier from these features. To choose useful features from the original measurements, binary Jaya (Berjaya) was suggested in [23]. The stability state was then predicted using an ensemble of OS-ELMs (Online Sequential ELMs). In general, the internal features of ML have a significant role in its performance, and creating efficient features calls for deep domain knowledge and rigorous engineering. As previously indicated, the recently developed end-to-end DL does not rely on feature engineering approaches or domain knowledge to extract valuable features. Instead, the DL model can be fed directly with the actual measurement data. The applications listed above are for the large-disturbance VSA in [22], and AI is also utilised for issues with small-disturbance voltage stability.

3.3.3. FREQUENCY STABILITY ASSESSMENT

Frequency stability is the capacity of a power grid to maintain consistent frequency after a significant system upset those results in a significant imbalance between generation and load. The integration of stochastic and volatile renewable energy has an impact on the balance between generation and load in modern smart grids, and the use of power electronic devices tends to reduce system inertia. As a result, frequency stability becomes an apparent and important problem. In these circumstances, frequency stability assessment (FSA) is a helpful technique for guaranteeing stable system performance. Because AI has such incredible Precision and speed, FSA has also used it. [24] Presented an ELM-based real-time FSA method. The predictor trained offline is then utilised live to forecast the frequency stability following possible situations. Instead of a binary categorization, the predictor creates a continuous frequency stability margin index to measure the level of stability. SVM was used to predict stability. In order to re-establish the stability of the electrical system, a new load-shedding mechanism was also introduced.

3.3.4 VOLTAGE STABILITY ASSESSMENT

Voltage stability describes a power grid's capacity to keep constant voltages at all buses in the event of a change from the initial operating condition. Voltage stability is somewhat threatened as a result of power networks being forced to run close to the limits of their transmission capacities due to an ever-increasing load demand. The characteristics of dynamic loads, such as motors and air conditioners, also have a tendency to exacerbate the voltage stability issue in industrial or urban settings. When used with complex controllers or a real-world large-scale power grid, conventional approaches for voltage stability assessment (VSA), such as energy function and bifurcation analysis, can be challenging. Additionally, large-disturbance VSA time-domain simulation requires a lot of computing power [18].For quick assessment or prediction; this is also challenging to apply online. A potent method to VSA has been discovered as the coming AI. The nonlinear link between power system variables and accompanying voltage stability status can be learned via simulation or measured data. As long as the input is given into the AI model at the online application stage, the stability may therefore be evaluated.

Chapter 4

METHODOLOGY

4.1 INTRODUCTION

The evaluation strategies for the smart grid stability analysis are mainly based on the analysis of machine learning models used. Machine learning models used for predicting the smart grid stability can be analysed by using classification evaluation metricses. Classification task for different modern machine learning models are discussed. The methodology consists of a new machine learning algorithm that is created by combining the individual machine learning algorithms to form a stacking ensemble classifier. The evaluation of new model created also done classification. This thesis contains two methodologies first one is analysing the individual modern machine learning models and compare their results, the second methodology is to create a new machine learning model that is stacked together with the individual machine learning models.

4.2 METHODOLOGY 1

In this methodology the smart grid stability are analysed using different modern machine learning algorithms.Fig.4.1 shows the flowchart of this methodology 1.the major steps behind this methodology are collecting the grid data from the source and store it in the personal laptop or computer drive and call them during the first step of the python program using code. Then the data is given for data pre- processing.in this stage the data from the smart grid is analysed carefully that is the behaviour of the data like is there any missing values ,repeated values ,null values or any error in the data will be detected. If there any error then that will be corrected because data is the key of machine learning algorithms. If there any error in the data that will leads to false results in the analysis. The result became fault then the working of smart grid is under danger and creates loss for both producer and consumer. Then split the entire data into train and test set. Here the data is split with 90% of total data for training purpose and 10% of total data is for testing purpose. Next stage is to create machine learning models. Modern machine learning algorithms are used for creating the machine learning models. Gradient boost, Boost. Adaboost, HistGBM, Catboost and LGBM are the modern machine learning algorithms used here. Each model is created individually and analysed individually by classification evaluation matrices, compare the result for finding the best model for smart grid stability analysis. Classification evaluation metrics are accuracy, Precision, Recall, F1-score, AUC –ROC, AUC-PR, MCC, Specificity, training time (TT), and predicting time (PT). Classification metrics are discussed in detail in section 4.2.1.

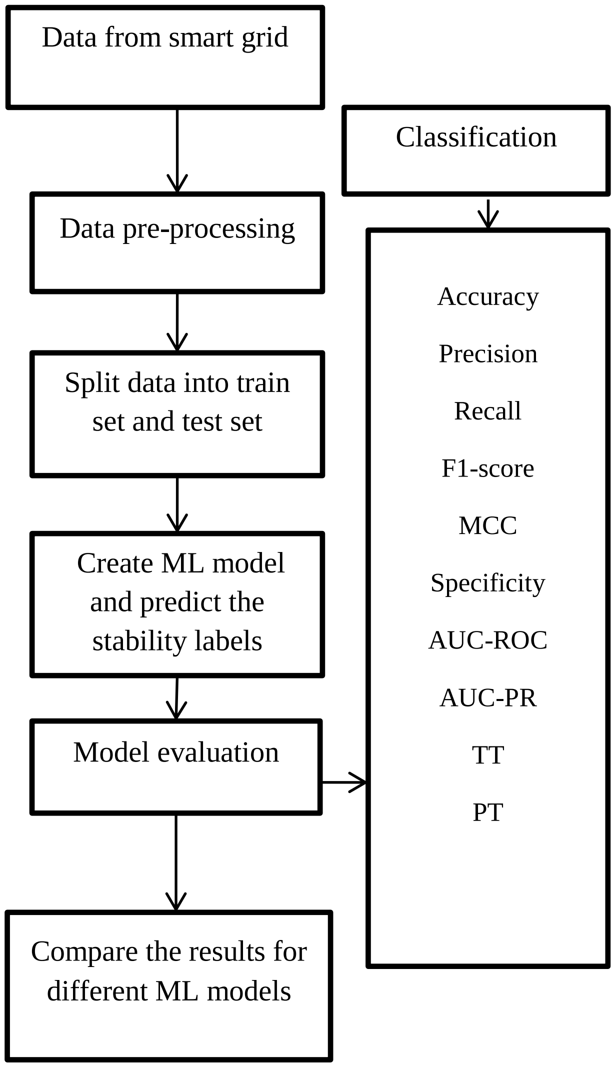

Fig.4.1.Flow chart of evaluation of ML algorithms for stability analysis

4.2.1 CLASSIFICATION EVALUATION METRICS

In machine learning, classification is a supervised learning concept which basically categorizes a set of data into classes. This thesis deals with binary classification .smart grid stability is analysed based on two classes that are both stable and unstable. Stable grid condition is labelled as 1 and unstable grid condition is labelled as 0.the major classification evaluation metrics used are discussed here. Classification evaluation metrics can be found out using confusion metrics. The performance of the classification models for a certain set of test data is assessed using a matrix called the confusion matrix. Only after the true values of the test data are known can it be determined. The matrix itself is simple to comprehend. here the classification is binary type so the number of elements in the confusion metrics of each machine learning algorithm consists of only four elements .first element represent true positive value ,second element represents false negative predictions ,third element shows the number of false positive predictions and fourth element shows total number of rue negative predictions. In the case of grid stability analysis negative prediction means predicting the grid as unstable and positive prediction means predicting the grid condition as stable. The sum of total number of elements in the confusion metrics is equal to the value of observations taken for testing the machine learning models.Fig.4.2 shows that the basics of confusion metrics.

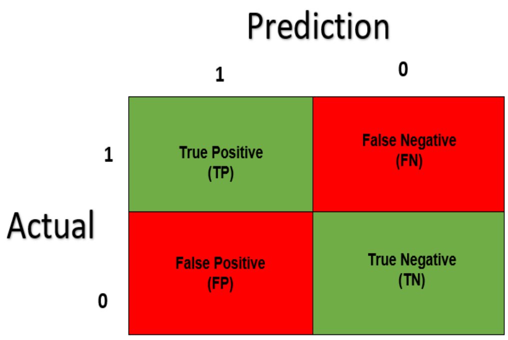

Fig.4.2. Confusion metrics model

Accuracy

The accuracy of a classifier is simply the number of times it predicts accurately. The number of correct predictions divided by the total number of forecasts is known as accuracy.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4.1}$$

Classification-error

There are two sorts of errors in binary classification issues. Type 1 mistakes (false positives) and Type 2 errors are the two types of errors (false negatives). It's common to be able to raise one while decreasing the other through model selection and tuning, and it's also common to have to choose which mistake type is more acceptable.

$$\text{Classification error} = \frac{(\mathrm{FP} + \mathrm{FN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} \tag{4.2}$$

Precision

Precision is one measure of a machine learning model's performance – the accuracy of a model's positive prediction. The number of true positives divided by the total number of positive predictions is known as Precision (i.e., the number of true positives plus the number of false positives).

$$\text{Precision} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FP})} \tag{4.3}$$

Recall

The Recall is determined by dividing the total number of Positive samples by the number of Positive samples accurately categorised as Positive. The Recall is a metric that evaluates how well a model can detect positive samples. The higher the Recall, the greater the number of positive samples found.

$$\text{Recall} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{4.4}$$

Matthews's correlation coefficient (MCC)

Instead, the Matthews correlation coefficient (MCC) is a more accurate statistical rate that yields a high score only if the forecast is correct in all four areas of the confusion matrix (true positives, false negatives, true negatives, and false positives),

$$\text{MCC} = \frac{(\mathrm{TP} * \mathrm{TN} - \mathrm{FP} * \mathrm{FN})}{((\mathrm{TP}+\mathrm{FP}) * (\mathrm{TP}+\mathrm{FN}) * (\mathrm{FN}+\mathrm{TN}) * (\mathrm{FP}+\mathrm{TN})) * 0.5} \tag{4.5}$$

F1 score

It simply counts the number of correct predictions made by a machine learning model. As you've seen, accuracy is a poor statistic to employ when dealing with unbalanced data since it can't distinguish between different sorts of errors (false positives and false negatives). Table 4.1 shows the F1-score values and their interpretations. The F1 score is a mix of Precision and Recall, as stated in the definition. The number of True Positives divided by the total number of True Positives and False Positives equals Precision. Precision can be thought of as a metric for how accurate something is. As a result, low Precision denotes a high number of False Positives.

$$\text{F1 score}= \frac{\text{TP}}{(\text{TP} + (\text{FP} + \text{FN})*0.5)} \tag{4.6}$$

Table 4.1 F1-score values and inferences

F1-Score	Inference
> 0.9	Very good
0.8 - 0.9	Good
0.5 - 0.8	OK
< 0.5	Not good

True positive rate

The true positive rate, also known as sensitivity or Recall in machine learning, is a metric that measures the percentage of actual positives that are accurately identified.

$$\text{True positive rate} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \tag{4.7}$$

False positive rate

The false positive rate (FPR) is a metric for determining how accurate a test is, whether it's a medical diagnostic test, a machine learning model, or something else. The likelihood of incorrectly rejecting the null hypothesis is described as the false positive rate in technical language.

$$\text{False positive rate} = \frac{\text{FP}}{(\text{FP} + \text{TN})} \tag{4.8}$$

Specificity and sensitivity

Sensitivity and specificity are two metrics of a model's performance in machine learning. The proportion of true positives accurately predicted by the model is called sensitivity, whereas the proportion of true negatives correctly predicted by the model is called specificity.

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \tag{4.9}$$

Area under the curve of receiver operating characteristics (AUC-ROC)

The Area Under the Curve (AUC) - ROC curve is a performance statistic for classification issues at various threshold levels. AUC represents the degree or measure of separability, whereas ROC is a probability curve. It indicates how well the model can distinguish between classes. Even in ordinary life, accuracy is a regularly used statistic. The AUC, on the other hand, is only employed when dealing with classification difficulties including probabilities in order to dig further into the forecast. As a result, even a non-technical person may grasp and appreciate accuracy.

Area under the curve of Precision-Recall curve (AUC-PR)

The area under the (Precision-Recall) curve is referred to as AUC-PR. In general, the greater the AUC-PR value, the better the classifier performs for the job at hand. Finding the AP, or average Precision, is one technique to determine AUC-PR. Because good classifiers should outperform chance, the AUC is usually in the range [0.5, 1]. However, in theory, the AUC can be less than 0.5, indicating that the classifier performs worse than a random classifier. The higher your curve is on the y-axis, the better your model's performance. When a classifier produces a score of less than 0.5, it simply signifies that the model performs worse than a random classifier and is hence useless. AUC is higher in a good PR curve (area under curve).

4.3 MODERN MACHINE LEARNING ALGORITHMS

Gradient Boosting classifier

Gradient boosting was inspired by Leo Breiman's remark that boosting can be viewed as an optimization technique on an appropriate cost function. Gradient boosting is a method that stands out for its predictability and speed, especially when dealing with big and complicated datasets. This algorithm has provided the best results in everything from Kaggle competitions to machine learning solutions for businesses. We already know that in any machine learning algorithm, errors play a significant role. Bias error and variance error are the two most common types of error. The gradient boost approach assists us in reducing the model's bias inaccuracy.

AdaBoost

The Ada Boost Algorithm is another boosting approach. This algorithm begins by constructing a decision stump, after which all data points are given equal weights. The weights for all the points that are misclassified are then increased, while the weights for those that are easy to categorise or are correctly classified are decreased. For these weighted data points, a new decision stump is created. The goal is to enhance the accuracy of the first stump's predictions. The key difference between these two techniques is that gradient boosting uses a fixed base estimator, such as Decision Trees, whereas Ada Boost allows us to adjust the base estimator as needed.

Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is a scalable distributed toolbox for gradient-boosted decision trees (GBDT). The best machine learning package available today incorporates parallel tree boosting and is ideal for problems including regression, classification, and ranking. Gradient Boosting is a more regularised version of XGBoost. Advanced regularisation (L1 & L2) is used in XGBoost to improve model generalisation. XGBoost outperforms Gradient Boosting in terms of performance. It has a quick learning curve and can be parallelized across clusters.

Histogram Boosting Gradient Classifier (HBG)

A histogram is a graph that counts or depicts the frequency of data (number of occurrences) over discrete time intervals known as bins. The histogram algorithm is conceptually straightforward, with each bin representing the frequency of the related pixel value. If we have scikit-learn v0.21.0 or later, HGB will be available. In simple terms, we all know that binning is a data pre-processing technique that entails taking VIT University and dividing the students into states such as Tamilnadu, Kerala, and Karnataka, among others. The same binning idea is applied to the Decision Tree (DT) technique after segmentation turns into numerical data. It will be used to boost the algorithm's speed by lowering the number of features. As a result, the HGB classifier, which employs the same concept in DT by grouping with histograms, is known.

In general, we have various parameters for fine-tuning our specialised algorithms to produce the best results for all categories. Learning rate, max iter, max depth, and l2 regularization are all crucial parameters for the HBG classifier. Learning rate deals with shrinkage, max iter with the number of iterations required to achieve a satisfactory result, max depth with multiple trees (Decision tree concepts), and l2 regularization with regularisation concepts to avoid over fitting concerns.

CatBoost

Cat Boost is scalable and has the ability to employ category characteristics directly. The CatBoost machine learning algorithm from Yandex was just released as open source. Deep learning frameworks like Tensor Flow from Google and Core ML from Apple are easy to interface with. It may be used with a range of data types to help organisations with a range of issues. Additionally, it has the highest level of accuracy in the sector. such that it is chosen for analysis of the stability of the smart grid.

It excels in two areas: first, it offers robust out-of-the-box support for the more descriptive data formats that frequently accompany business problems; second, it generates cutting-edge results without the extensive data training that other machine learning techniques necessitate. The words "Category" and "Boosting" are combined to form the name "CatBoost."

LightGBM

LightGBM (Light Gradient Boosting Machine) is a free and open source distributed gradient boosting framework for machine learning that was created by Microsoft. It is used for ranking, classification, and other machine learning applications and is based on decision tree algorithms. LightGBM is a tree-based learning algorithm-based gradient boosting framework. It is intended to be distributed and efficient, and it offers the following benefits:

- Increased training efficiency and speed.
- Reduce your memory utilisation.
- Better Precision.
- Parallel, distributed, and GPU learning are all supported.
- Capable of dealing with massive amounts of data.

4.4 METHODOLOGY 2

This methodology shown in Fig.4.2 deals with a new machine learning algorithm called stacking ensemble model. This methodology can be used for creating stacking ensemble classifier. Otherwise it is called SEC model. Stacking ensemble classifier (SEC) is created by combining the individual classifiers such as XGBoost classifier, Gradient Boost classifier, AdaBoost classifier, CatBoost classifier, HistGBM classifier and Light GBM classifier. In this stacking ensemble classifier Light GBM classifier act as the Meta learner and other classifiers act as the base learners. SEC model use the principle of “intelligence of the crowd”SEC model improves the credibility of the prediction of stability classification of the smart grid.

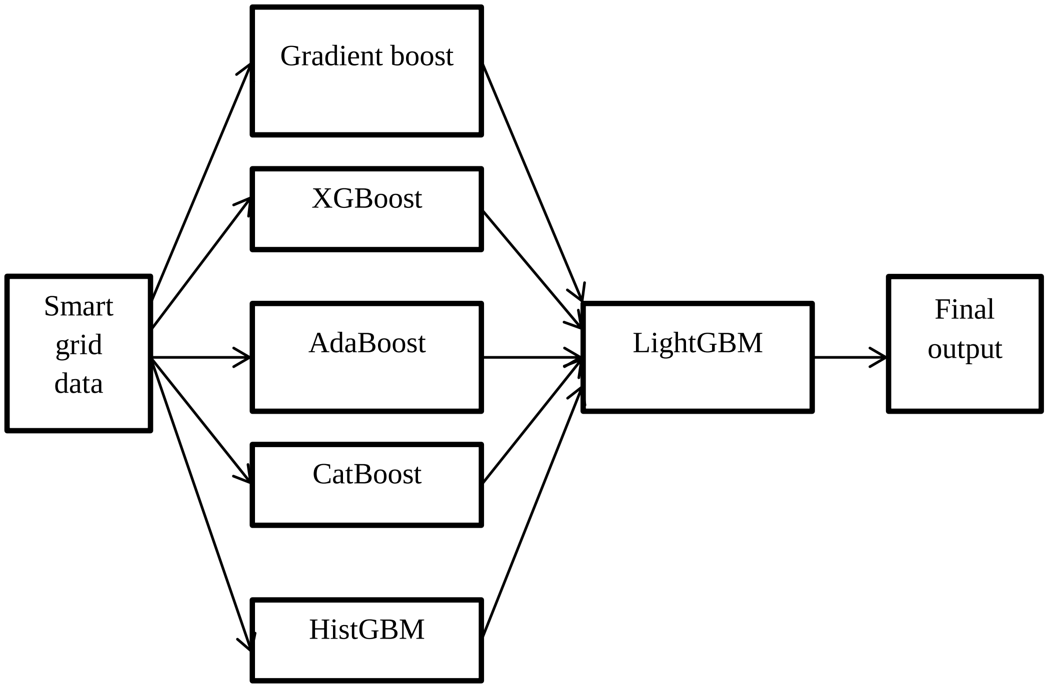

Fig.4.3. Flow chart of stacking ensemble classifier

It has mainly two stages for creation of this model that are creation of stacking model and add a Meta model along with the stacked model. Here the stacked model consist five individual base machine learning models called base estimators that are gradient boost, Histboost, Ada boost, XGboost and Catboost. The Meta model used here is LGBM model. Stacking models will train the data during first level of operation and Meta model will predict the final output from the test set. After creating the stacking classifier analyse the model using classification evaluation matrices. Then compare the results with that of individual machine learning models. This comparison will give the idea about which machine learning model is the best for stability analysis of a smart grid.

Chapter 5

CASE STUDY

The grid stability is analysed using a case study that utilises a four node star connected network. First node represents the producer node and the other three nodes represent the consumer nodes respectively. Each consumer nodes consists of thousands of residential and industrial consumers. The producer node consists conventional and non - conventional energy producers the dataset made from this four node network is an upgraded version of Vadis Arzamasov's (Germany) Electrical Grid Stability Simulated Dataset, which was provided to the University of California (UCI) Machine Learning Repository. The grid stability simulation results for a reference 4-node star network are included in the dataset, as shown in Fig.5.1.

The generation node consists of energy producers such as hydro power plant, nuclear power plant, thermal power plant, diesel power plant, wind plants, tidal plants and solar panels from the individual producers etc. The consumption node consists of industrial consumers, domestic consumers and electric vehicles etc. each node is denoted as a single unit but their operations are identical and the consists of several thousands of producers in the producer node and consumers in the consumer node.

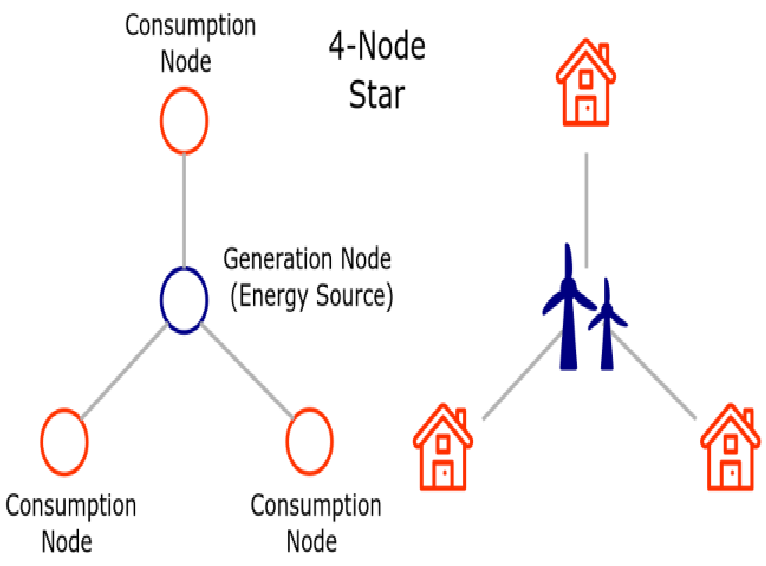

Fig.5.1. Four node smart grid test system

The stability of the system basically consider as the sum of nominal power produced became equal to the nominal power consumed. Otherwise we can say that the sum of power produced by the producer node and the power consumed by the consumer nodes are zero. If the production and consumption of energy became not equal hen we can say that the system is not in synchronism. Production of electricity is directly depends on the consumption of energy because if the consumption is high then we must increase the production and also during the less consumption time we can reduce the production of power also. Nowadays we can store electricity by using battery storage devices.in many houses and industries inverters are used during power outage situations. Inverters will store energy during off peak hours and they will supply power to the system during peak hours. Sign of consumed power is taken as negative and that of produced power is taken as positive. The value power produced and consumed will depend on the energy usage of the customers.

$$P1+P2+P3+P4=0 \tag{5.1}$$

$$P1=-(P2+P3+P4) \tag{5.2}$$

Table 5.1. Dataset parameters used for ML models

Number	Features	Description
1	tau1	Reaction time of the electricity producer
2	tau2	Reaction time of the first electricity consumer
3	tau3	Reaction time of the second electricity consumer
4	tau4	Reaction time of the third electricity consumer
5	p1	Nominal power produced by the producer
6	p2	Nominal power consumed by the first consumer
7	p3	Nominal power consumed by the second consumer
8	p4	Nominal power consumed by the third consumer
9	g1	Gamma coefficient of producer
10	g2	Gamma coefficient of first consumer
11	g3	Gamma coefficient of second consumer
12	g4	Gamma coefficient of third consumer
13	Label	Stable or unstable

There are 60,000 observations in the dataset, with 12 predictive characteristics and two categorical features denoted as label in Table 5.1. The labels are stable and unstable. Grid stable condition is denoted as 1 and grid unstable condition is denoted as 0 in each machine learning algorithm. The model has taken into account input features such as the total power balance (nominal power produced or consumed at each grid node), the reaction time of participants to adjust consumption or production in response to price changes (also known as the reaction time), its value ranges from 0.5 to 10s and the total power balance (nominal power produced or consumed at each grid node. All other factors being equal, energy price elasticity is a measure of the percentage change in energy consumption compared to the percentage change in pricing its value ranges from 0.05 to 1 s^-1.

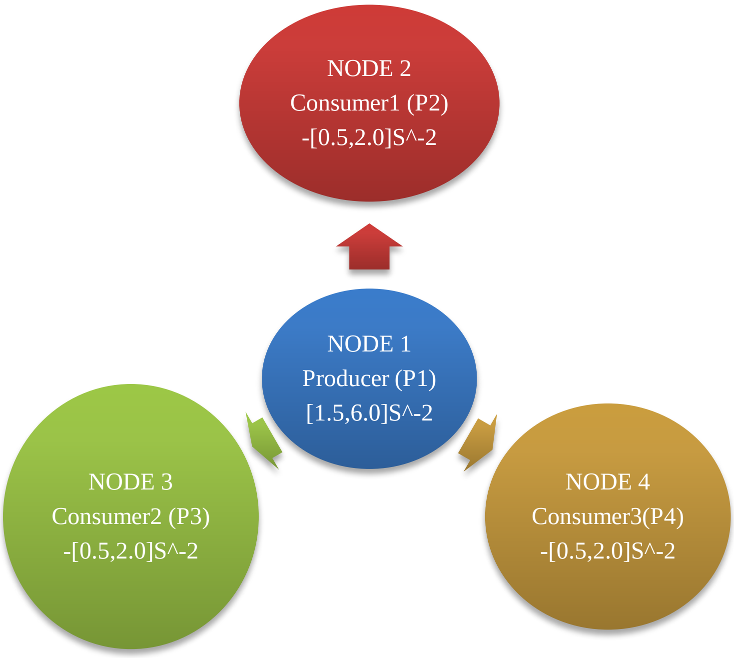

Fig.5.2 Four node star grid simulated with De-central Smart Grid Control

Fig.5.2 shows the nominal power consumed and produced range of the four node star shaped de-centralised smart grid. Energy equilibrium and economic equilibrium are considered for creating the data set. DR is the key element of the simulated dataset. The dataset considered the assumptions and values related with the demand response. Over penetration of renewable energy sources will cause stability problems in the smart grid mainly.so this simulated dataset will consider the demand of energy and its price for predicting the stability of the system. Accurate prediction of stability will help to save money for both producer and consumer.

Chapter 6

RESULTS AND DISCUSSIONS

6.1 CLASSIFICATION RESULTS

The result of each machine learning model is based on the confusion metrics values. Confusion metrics are the necessary part of classification analysis of data. In this work the percentage of test data consider is about 10%.toatal observation are 60000.so the 10% of 60000 is the total number of elements in the confusion metrics. Here it is 6000 elements .by analysing the confusion metrics of each machine learning model we get the behaviour of the model towards predicting the grid stability. That means for a good machine learning model the value of false positive and false negative value should be minimum. For a perfect model it must be zero. That is all the predictions are either true positive or true negative.

Table 6.1.Confusion matrix of ML models

Model	**Label**	**Stable**	**Unstable**
Gradient boost	Stable	1879	225
	Unstable	326	3570
XGBoost	Stable	2131	66
	Unstable	74	3729
AdaBoost	Stable	1589	320
	Unstable	616	3475
HistGBM	Stable	2140	46
	Unstable	65	3749
LightGBM	Stable	2050	84
	Unstable	137	3729
CatBoost	Stable	2179	21
	Unstable	26	3774
SEC	Stable	3780	21
	Unstable	15	2184

In table 6.1 shows the confusion metrics of each machine learning model .the values of the confusion, metrics will give the perfection of the model for predicting the smart grid stability. From the table it is clear that stacking ensemble classifier has very less number of false positive and false negative values .in total 6000 predictions only 15 is predicted as false stable and only 21 is predicted as false unstable. But while considering the individual machine learning models Catboost nearly similar prediction level of SEC model .it also predicted 21 true stable observations as unstable, but 26 unstable observations are predicted as stable .Adaboost model shows poor performance. Because in total 6000 observations 616 observations are false positive and 320 as false negative.

Table 6.2. Classification reports of ML models

Model	**Accuracy**	**Precision**	**Recall**	**F1 score**	**MCC**	**Specificity**	**Classification error**
Gradient boost	0.9082	0.8522	0.8931	0.8721	0.8010	0.9163	0.0918
XGboost	0.9767	0.9664	0.9700	0.9681	0.9497	0.9805	0.0233
AdaBoost	0.8440	0.7206	0.8324	0.7724	0.6586	0.8494	0.1560
HistGBM	0.9815	0.9705	0.9790	0.9747	0.9601	0.9830	0.0185
LGBM	0.9632	0.9374	0.9606	0.9488	0.9202	0.9646	0.0368
CatBoost	0.9922	0.9882	0.9905	0.9893	0.9831	0.9932	0.0078
SEC	0.9940	0.9960	0.9945	0.9952	0.9870	0.9948	0.0060

In Table 6.2 shows the accuracy, Precision ,Recall,F1 score, MCC, Specificity and classification error of modern individual machine learning model and stacking ensemble classifier.acuuracy values of each model shows that the rate of true predictions. Accuracy of SEC model is higher than that of all other individual models. Individual models got maximum 0.9922 accuracy that is for Cat boost classifier, but SEC model got 0.9940.the least accuracy is shown by Adaboost model that is 0.8440. Accuracy of SEC model is higher than that of all other individual models. Individual models got maximum 0.9922 accuracy but SEC model got 0.9940.the least accuracy is shown by Adaboost model that is 0.8440. It indicates that the model is not good for stability analysis of smart grid.

Precision of SEC model is higher than that of all other individual models. Individual models got maximum 0.9882 Precision that is for Cat boost classifier, but SEC model got 0.9960.the least Precision is shown by Adaboost model that is 0.7206.

Recall of SEC model is higher than that of all other individual models. Individual models got maximum 0.9905 Recall that is for Catboost classifier, but SEC model got 0.9940.the least accuracy is shown by Adaboost model that is 0.8440.

F1 score of SEC model is higher than that of all other individual models. Individual models got maximum 0.9893 F1 score but SEC model got 0.9952.the least F1 score is shown by Adaboost model that is 0.7724. It indicates that the model is not good for stability analysis of smart grid.

MCC of SEC model is higher than that of all other individual models. Individual models got maximum 0.9831 MCC but SEC model got 0.9870.the least MCC is shown by Adaboost model that is 0.6586. Specificity of SEC model and Catboost model shows the same value and it is higher than that of all other models 0.9932. The least Specificity is shown by Adaboost model that is 0.8494. It indicates that the model is not good for stability analysis of smart grid.

Classification error of SEC model is very less than that of all other individual models. Individual models got minimum 0.0078 Classification error but SEC model got 0.0060.the least Classification error more in the case shown by Adaboost model that is 0.1560.for a good model classification error is as minimum as possible in ideal case it is 0.0.

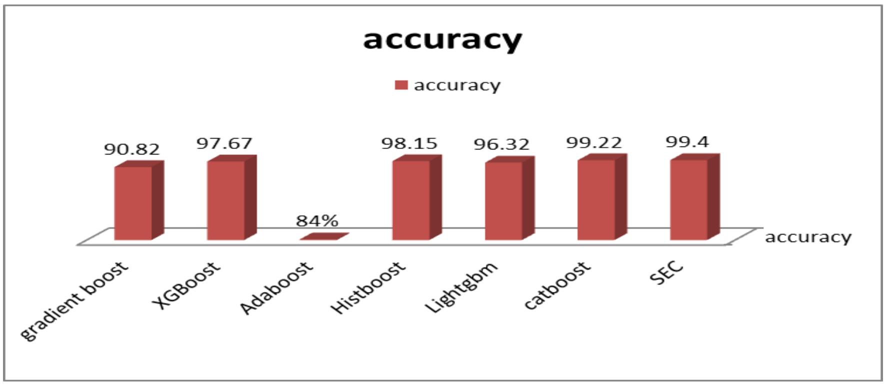

Fig.6.1. Visualisation of Accuracy of ML models

Fig.6.1 is the visualisation of accuracy of smart grid stability prediction using different machine learning algorithms. The accuracy of the prediction is denoted in percentage. From the graph itself it is clear that the Adaboost model has very less accuracy of 84%.gradient boost model has 90.82% accuracy, XGboost model has 97.67%accuracy,histboost model is having nearly similar accuracy range of XGboost that is about 98.15%.LGBM model has less accuracy than Histboost ,Catboost, XGboost models. Proposed SEC model has the highest accuracy range of about 99.4%.Accuracy of a model gives the ability of predicting the true predictions of the classification here it is the ability to predict the true stable condition of the grid or true unstable condition of the grid.so considering the smart grid stability accuracy of ML model highly important.

Fig.6.2 is the visualisation of Precision of smart grid stability prediction using different machine learning algorithms. The Precision of the prediction is denoted in percentage. From the graph itself it is clear that the Adaboost model has very less Precision of 72%%.gradient boost model has 85.22%,xgboost model has 96.64%,histboost model is having nearly similar Precision range of Boost that is about 97.05%.LGBM model has less Precision than Histboost, Catboost, XGboost models. Proposed SEC model has the highest Precision range of about 99.6%.Precision is the ability of the model to predict the true positives in all positive predictions .it means that sometimes the model predict that the system is unstable but actually it is stable and in some the model will predict the grid is stable but actually the grid is not stable.

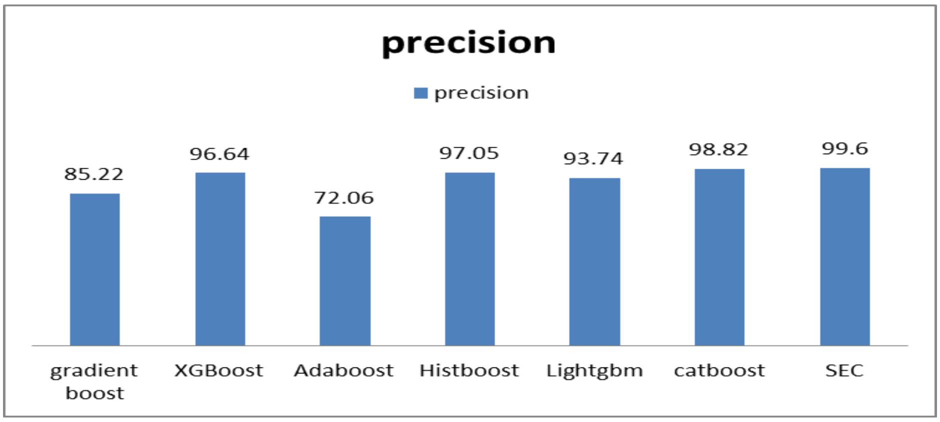

Fig.6.2. Visualisation of Precision of ML models

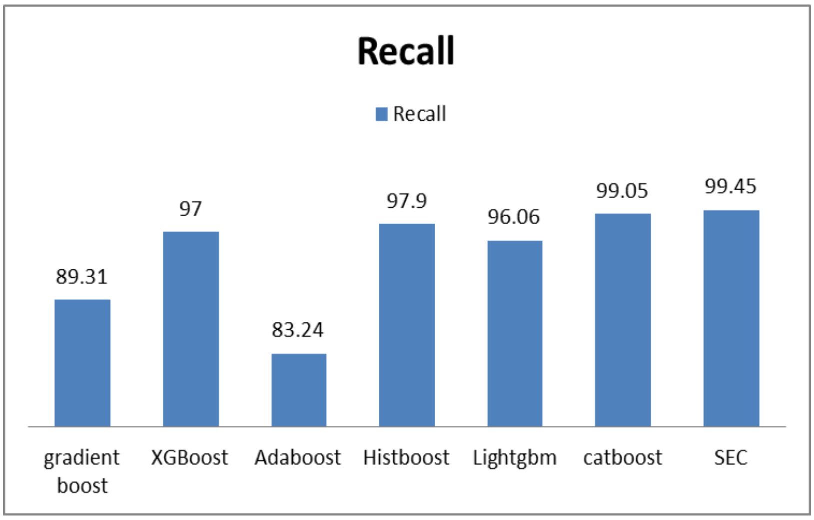

Fig.6.3.Visualisation of Recall of ML models

Fig.6.3 is the visualisation of Recall of smart grid stability prediction using different machine learning algorithms. The Recall of the prediction is denoted in percentage. From the graph itself it is clear that the Adaboost model has very less Recall of 83.24%.gradient boost model has 89.31%,xgboost model has 97% Recall,histboost model Recall range is about 97.9%.LGBM model has less Recall than Histboost, Catboost, XGboost models. Proposed SEC model has the highest Precision range of about 99.52%.

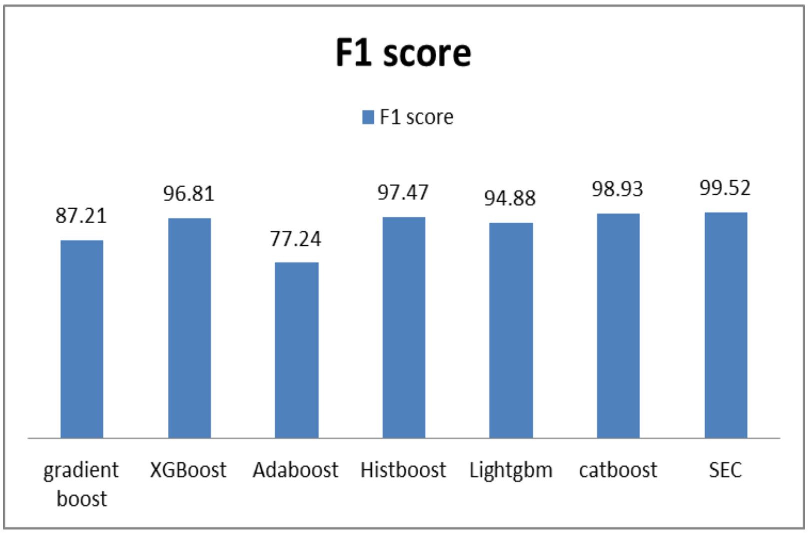

Fig.6.4.Visualisation of F1-score of ML models

Fig.6.4 shows that the f1 –score of various ML models in percentage range. Adaboost model has very poor value of f1 –score that is 77.24%.F1-score indicate the ability of the model to predict the positive predictions of the system. If a system is having high f1-score means the system is good for predicting the positive class labels. The increasing order of the ability of the positive predictions for models are in the order Adaboost,Gradient boost, LGBM, XGboost, Histboost,Catboost and SEC model.

Fig.6.5 shows the visualisation of specificity of the ML models. Specificity means the ability of the model to predict the true negative classes of the predictions. True negative in the case of smart grid stability prediction means that the ability of the model to predict the unstable condition of the grid. Specificity is given in the figure is in percentage range. Specificity is very poor for Adaboost model and it is about 84.94% and high for SEC model 99.48%.xgboost and Histboost has very similar range of specificity that is 98.3 % and 98.5% respectively.

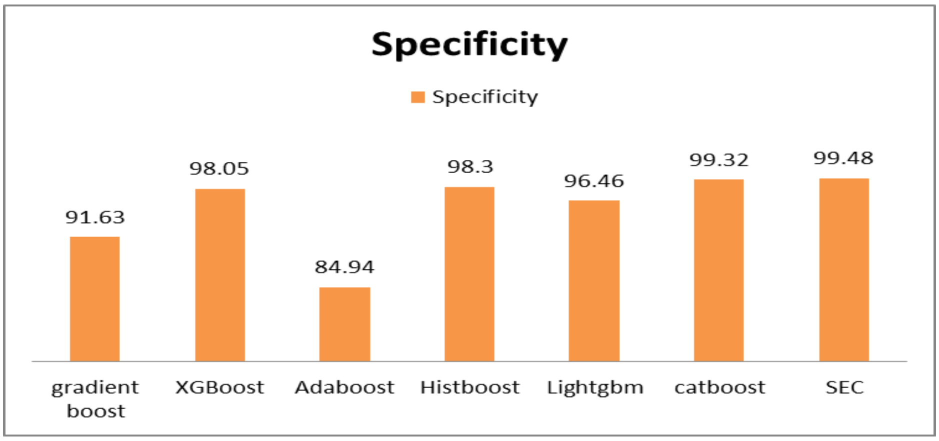

Fig.6.5. Visualisation of Specificity of ML models

Fig.6.6 shows the visualisation of the classification error of different ML models. Classification error is the indication of the false predictions of the events. For a good model the classification error is as minimum as possible. From the figure it is clear that the proposed SEC model has very low value of classification error that is 0.006.clasification error of Ada boost model is high that is 0.156.it indicate that the model classification label predictions are not true in most cases.so we can't use the Adaboost model alone for predicting the smart grid stability.

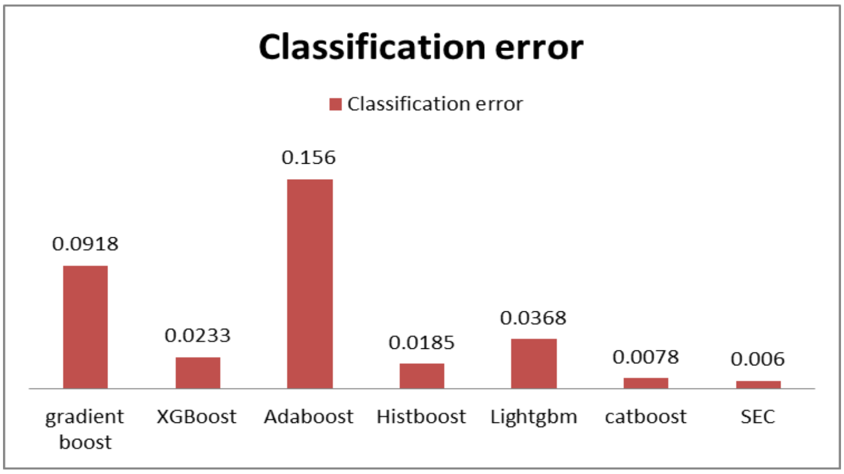

Fig.6.6.Visualisation of Classification error of ML models

Table 6.3.Shows the true positive and false positive rate of each machine learning models. And Fig.6.7 shows the visualisation of the TPR and FPR of the ML models. True positive rate (TPR) of SEC model is higher than that of all other individual models. Individual models got maximum 0.9905 True positive rates that are for Catboost classifier, but SEC model got 0.9945.the least True positive rate is shown by Adaboost model that is

0.8324.False positive rate (FPR) of SEC model and Catboost model shows the same value and it is less than that of all other models 0.0068. The least false positive rate is shown by Adaboost model is higher than that of all other models that is 0.1506. Fig.6.7 shows the visualisation of TPR and FPR of the models. For a good model true positive rate must be maximum and false positive rate should be minimum as possible.

Table 6.3.True positive and false positive rates of ML models

Model	**TPR**	**FPR**
Gradient boost	0.8931	0.0837
XGboost	0.9700	0.0195
AdaBoost	0.8324	0.1506
HistGBM	0.9790	0.0170
LightGM	0.9606	0.0354
CatBoost	0.9905	0.0068
SEC	0.9945	0.0068

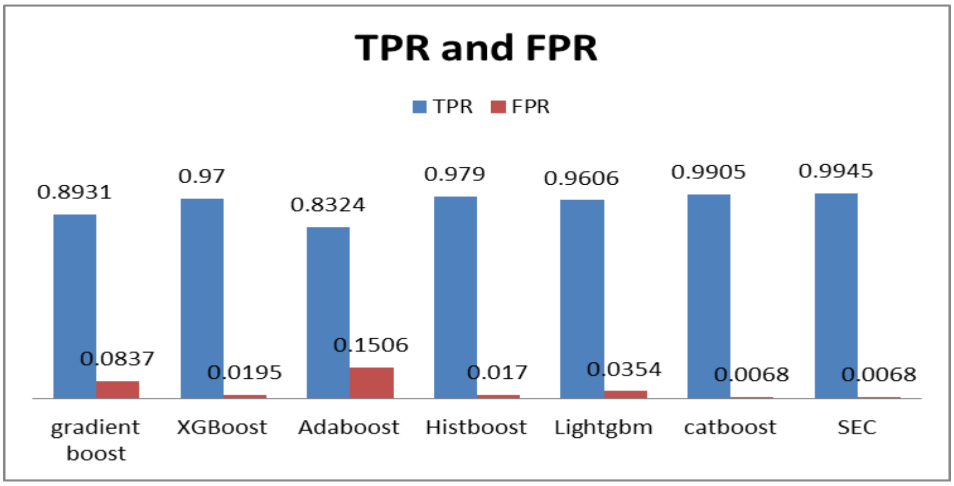

Fig.6.7.Visualisation of TPR and FPR of ML models

Table 6.4.Training time and Predicting time of ML models

Model	T,T	P.T
Gradient boost	8.4730	0.0156
XGboost	9.1373	0.0156
AdaBoost	8.1150	0.0625
HistGBM	4.4746	0.0937
LightGM	0.5936	0.0156
CatBoost	10.231	0.1560
SEC	51.352	0.1710

Table 6.4.Shows that the training and predicting time of each machine learning models. Training time depends on the number of training set in the dataset. Here we choose 90% data for training so the training time is more. The most modern ML model LGBM shows very less training time and predicting time .due to the complexity of the SEC model it takes high time for both training and prediction of the models.

Fig.6.8 to 6.13 shows AUC-ROC curve of each individual machine learning models.

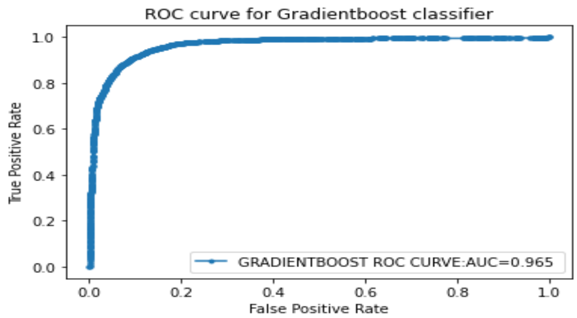

Fig.6.8. Gradient boost classifier ROC curve

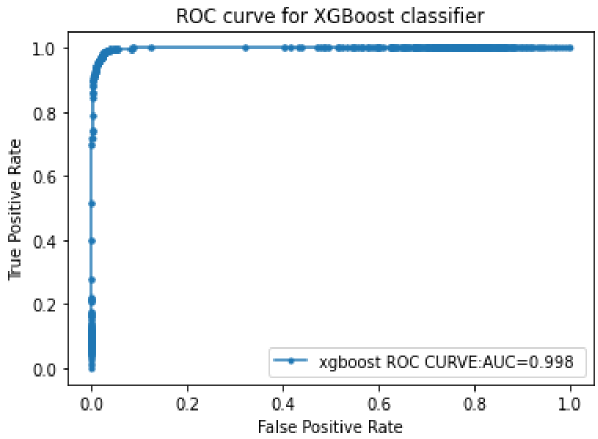

Fig.6.9.XGboost classifier ROC curve

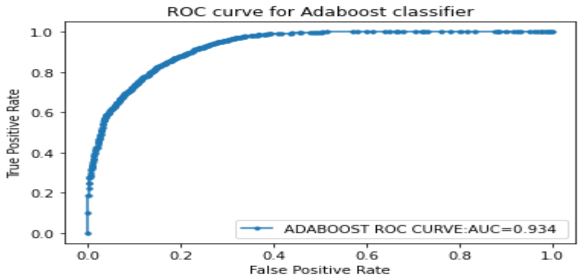

Fig.6.10.Adaboost classifier ROC curve

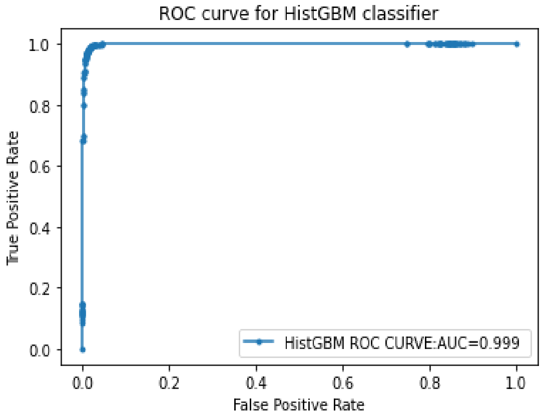

Fig.6.11.HistGBM classifier ROC curve

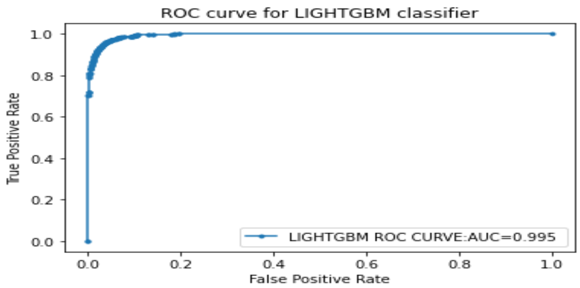

Fig.6.12.LightGBM classifier ROC curve

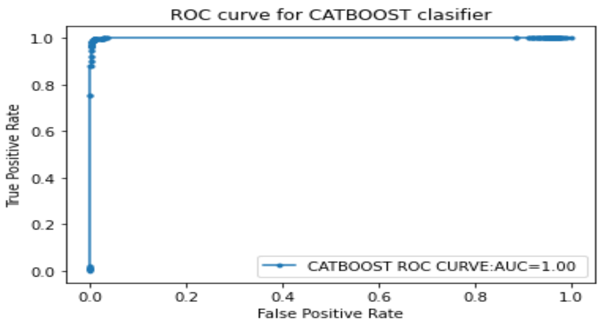

Fig.6.13.Catboost classifier ROC curve

Table 6.5.AUC-ROC values of ML models

Model	**AUC-ROC**
Gradient boost	0.965
XGboost	0.998
AdaBoost	0.932
HistGBM	0.999
LightGM	0.995
CatBoost	1.000

In Table 6.5 all the AUC values of the ROC curve of each ML model is entered. From this table it is clear that Catboost classifier shows maximum value of AUC. That shows the perfection of the model and Adaboost classifier shows very less value of AUC, that is 0.932.

Fig.6.14 to 6.19 shows AUC-PR curve of each individual machine learning models.PR curve is the plot between Precision and Recall. Precision along Y-axis and Recall along X-axis. Maximum value of AUC of a PR curve 1.

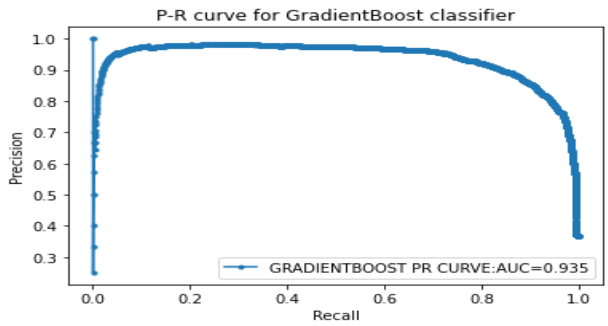

Fig.6.14. Gradient boost classifier PR curve

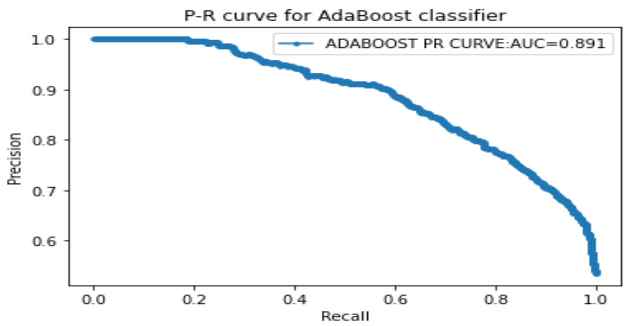

Fig.6.15.Adaboost classifier PR curve

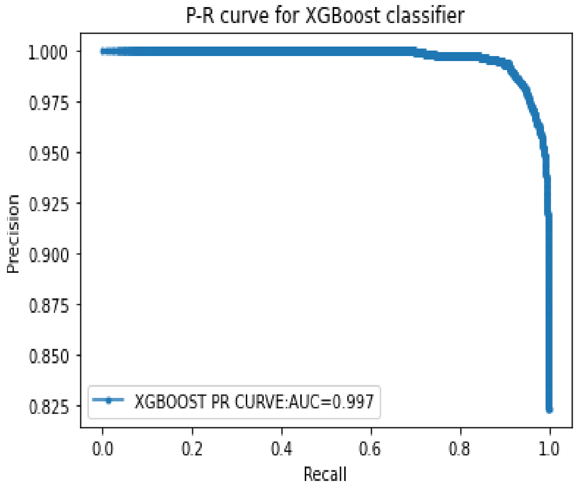

Fig.6.16.XGboost classifier ROC curve

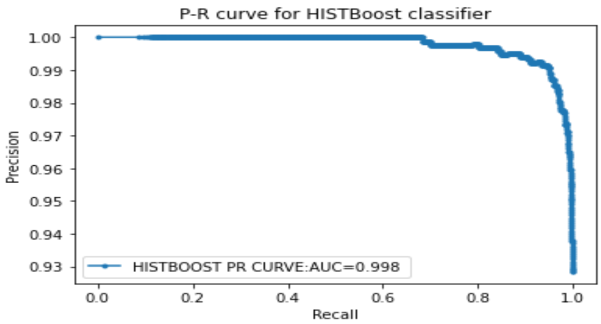

Fig.6.17.HistGBM classifier PR curve

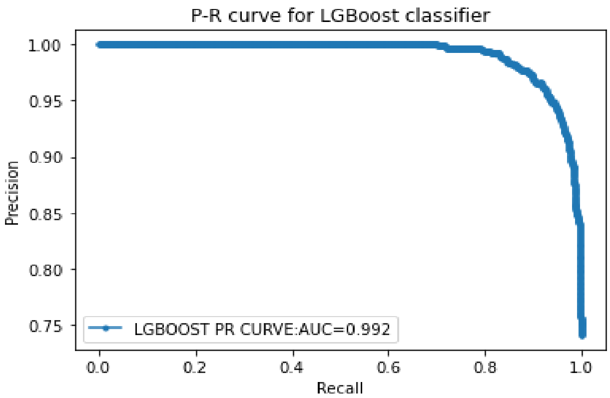

Fig.6.18.LightGBM classifier PR curve

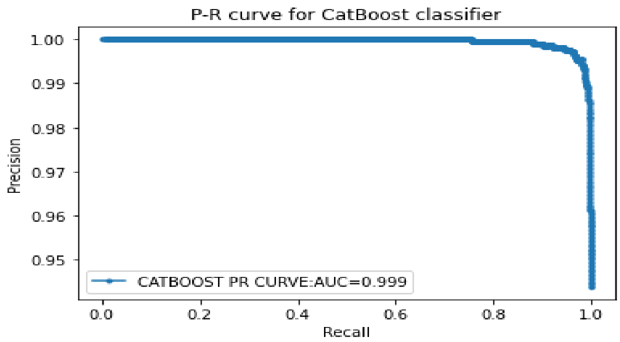

Fig.6.19.Catboost classifier PR curve

In Table 6.6 all the AUC values of the PR curve of each ML model is entered. From this table it is clear that Catboost classifier shows maximum value of AUC in PR curve that is 1.00.that shows the perfection of the model and Adaboost classifier shows very less value of AUC, that is 0.885 it indicate that the model is not good for stability analysis of smart grid.

Table 6.6.AUC-PR values of ML models

Model	**AUC-PR**
Gradient boost	0.935
XGboost	0.997
AdaBoost	0.885
HistGBM	0.998
LightGBM	0.992
CatBoost	1.000

Chapter 7

CONCLUSION

Analysed the performance of the individual ML models such as Boost, GradientBoost , AdaBoost, CatBoost, HistGBM and Light GBM classifiers for predicting the smart grid stability using accuracy,Precision,Recall,F-1 score, specificity, MCC, predicting time ,training time, AUC-ROC curve and AUC-PR curve. Created a SEC model by combining the individual classifiers such as XGBoost classifier, Gradient Boost classifier, AdaBoost classifier, CatBoost classifier, HistGBM classifier and Light GBM classifier , then find the results of classification evaluation metrics. From all these classification evaluation matrics clearly shows that stacking ensemble model (SEC) is the best while comparing with other individual modern machine learning algorithms. Considering only the individual machine learning models Catboost classifier is the best one. Ada boost model shows very poor performance in all classification evaluation metrics.

Accuracy, Precision, Recall and F1–score of SEC model is higher than that of any other individual ML models. It indicates that the SEC model is effective for predicting the positive behaviour of the system. That is the SEC model is good for predicting the stable condition of the system and also the specificity value of the SEC model is very much higher than that of all other models. It indicates that the model is good for predicting the instability of the smart grid also. While considering the individual ML models Cat boost classifier has nearly similar performance of SEC model in some evaluation metrics. But considering the credibility of the system stability prediction SEC model results are better.

The future research scope of this thesis are analysing the performance of the smart grid stability using different kinds deep learning methods and can evaluate the performance of different smart grid stability dataset based on different ML models and the developed SEC model. Stability of the smart grid can be analysed by considering different kinds of features other than nominal power balance, price elasticity coefficient and reaction time. We can consider the features such as frequency, voltage, rotor angle, load demand, various faults etc.in this thesis binary classification is used for predicting the stability of the smart grid. But can use multi classification of smart grid stability based on transient nature, dynamic nature steady state nature, voltage stability, frequency stability and rotor angle stability.

REFERENCES

[1] **C. Wietfeld, A. A. Cardenas, H. Chen, P. Popovski and V. W. S. Wong**, "Smart Grids", *IEEE Wireless Communications*, vol. 24, no. 2, pp. 8-9, April 2017.

[2] **R. Morello, C. De Capua, G. Fulco and S. C. Mukhopadhyay**, "A Smart Power Meter to Monitor Energy Flow in Smart Grids: The Role of Advanced Sensing and IOT in the Electric Grid of the Future", *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7828-7837, Dec. 2017.

[3] **B. K. Bose,** "Artificial Intelligence Techniques in Smart Grid and Renewable Energy Systems—Some Example Applications", *Proceedings of the IEEE*, vol. 105, no. 11, pp. 2262-2273, Nov. 2017.

[4] **D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si,** ''A data-driven design for fault detection of wind turbines using random forests and XGboost,'' *IEEE Access*, vol. 6, pp. 21020–21031, 2018.

[5] **Y. Chen, S. M. Mazhari, C. Y. Chung, S. O. Faried, and B. C. Pal,** ''Rotor angle stability prediction of power systems with high wind power penetration using a stability index vector,*IEEETrans.PowerSyst*.,vol.35,no.6, pp. 4632–4643, Nov. 2020.

[6] **T. Guo and J. V. Milanovic**, ''Online identification of power system dynamic signature using PMU measurements and data mining,'' *IEEE Trans. Power Syst*., vol. 31, no. 3, pp. 1760–1768, May 2016.

[7] **Shutang You, Yinfeng Zhao, Mirka Mandich**, ''A Review on Artificial Intelligence for Grid Stability Assessment" *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids* 2020.

[8]**D.Syed, A.Zannab, S.S.Refaat, H.Abu-rub and O.Bouhali,** ''Smart Grid Big Data Analytics Survey of Technologies, Techniques and Applications", *IEEE Access* 2020.

[9]**S.K.Azman,Y.J.Isbeih,M.S.Ei Moursia and K.Elbasssioni**, ''A Unified Online Deep Learning Prediction Model For Small Signal and Transient Stability," *IEEETrans.PowerSyst*.,vol.35,no.6, pp. 4585–4598, Nov. 2020.

[10]**H. Hugmar, L.Tong, R. Eriksson**, ‘‘Voltage Instability Prediction Using a Deep Recurrent Neural Network,” *IEEETrans.PowerSyst*., vol.36, no.1, pp. 17–27, 2020.

[11]**M.Alzab, S. khan, S.S.R. Krishnan, Q.V.Pham, M.P.K. Reddy and T.R Gadekallu**, ‘‘A Multidirectional LSTM Model For Predicting The Stability Of A Smart Grid”,*IEEE Access*, vol. 8, pp. 85454–85463, 2020.

[12] **Gharavi H, Ghafurian R.** ‘‘Smart grid: The electric energy system of the future”. *IEEE* 2011; 99:917–21.

[13] **Bose BK.** ‘‘Power electronics, smart grid, and renewable energy systems. ” *IEEE* 2017; 105:2011–8.

[14] **Sun K, Yao W, Fang J, Ai X, Wen J, Cheng S**. ‘‘ Impedance modelling and stability analysis of grid-connected DFIG-based wind farm with a VSC-HVDC”. *IEEE J Emergy Sell Top Power Electron* 2020; 8:1375–90.

[15] **Peng X, Yao W, Yan C, Wen J, Cheng S**. ‘‘Two-stage variable proportion coefficient based frequency support of grid-connected DFIG-WTs”. *IEEE Trans Power Syst* 2020; 35:962–74

[16] **Wang Q, Yao W, Fang J, Ai X, Wen J, Yang X, et al.** ‘‘ Dynamic modelling and small signal stability analysis of distributed photovoltaic grid-connected system with large scale of panel level DC optimizers”. *Appl Energy* 2020; 259:114132.

[17] **Morison K, Wang L, Kundur P**. ‘‘Power system security assessment”. *IEEE Power Energy Mag.* 2004; 2:30–9.

[18] **Kundur P, Paserba J, Ajjarapu V, Andersson G, Bose A, Canizares C, et al**. ‘‘Definition and Classification of Power System Stability”, Vol. 19. 2004, p. 1387–40

[19] **Xu Y, Dong ZY, Zhao JH, Zhang P, Wong KP.** ‘‘A reliable intelligent system for real-time dynamic security assessment of power systems. ” *IEEE Trans Power System* 2012; 27:1253–63

[20] **Gupta A, Gurrala G, Sastry PS.** ‘‘An Online power system stability monitoring system using convolutional neural networks. ” *IEEE Trans Power System* 2019; 34:864–72

[21] **Gomez FR, Rajapakse AD, Annakkage UD, Fernando IT**. ‘‘Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements. ” *IEEE Trans Power System* 2011; 26:1474–83.

[22] **Yu JJQ, Hill DJ, Lam AYS, GU J, Li VOK**. ‘‘Intelligent time-adaptive transient stability assessment system. ” *IEEE Trans Power System* 2018; 33:1049–58.

[23] **Li Y, Yang Z**. ‘‘Application of EOS-ELM with binary Jaya-based feature selection to real-time transient stability assessment using PMU data. ” *IEEE Access* 2017; 5:23092–101

[24] **Xu Y, Dai Y, Dong ZY, Zhang R, Meng K.** ‘‘Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems. ” *Neural Computer Application* 2013; 22:501–8.