Power System Load Forecasting Using Machine Learning Algorithms: Optimal Approach

M. Ravindra Babu*, V. Suresh Chintalapudi **‡, Ch. Nagasai Kalyan**, K. Krishna Bhaskar ***

*Department of Electrical and Electronics Engineering, University College of Engineering Narasaraopet, Jawaharlal Nehru Technological University Kakinada, Narasaraopet, Andhra Pradesh, India, 522601

**Department of Electrical and Electronics Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India, 522508

***Department of Mechanical Engineering, University College of Engineering Kakinada, Jawaharlal Nehru Technological University Kakinada, Kakinada, Andhra Pradesh, India, 533003

(raviravi1983@gmail.com, venkatasuresh3@vvit.net, kalyanchallapalli@gmail.com, krisbhaskar22@yahoo.co.in)

‡ Corresponding Author, Chintalapudi V Suresh; Department of Electrical and Electronics Engineering, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, Andhra Pradesh, India, 522508, Tel: +91 9989 254 335, venkatasuresh3@vvit.net

Received: 14.07.2023 Accepted:14.08.2024

Abstract- Accurate forecasting of long-term electricity demand has a prominent role in demand side management and electricity system planning and operation. Demand over-estimation gives rise to over-investment in system assets, driving up the electric power prices, while demand under-estimation may give on to under-investment results in unreliable and insecure electricity. The electrical load on a station varies more dynamically and hence forecasting of load values is more significant. In long-term load forecasting (LF), economic factors, weather conditions, time factors and random effects plays a prominent role. Machine Learning (ML) algorithms often used to estimate the future values constructed on the inferences drawn from the existing values. For these algorithms, the input data is supplied for training and as well testing the algorithm efficacy. Now a days, large amount of data is available everywhere. Therefore, it is essential to analyze this data in order to extract some profitable information and to implement an algorithm based on this analysis. These algorithms works based on historical relationship between the various factors considered in the problem. In this paper, some popular ML algorithms such as Linear Regression, Logistic Regression, Navie Bayes, Random Forest, Decision Tree, Support Vector Machines, and KNN are used to develop the Forecasting model for standard 14 and 30 bus systems. The efficiency of the forecasting procedure is examined based on RMSE, Accuracy and time taken to generate the model. Further, the optimization results encourages for implementing these algorithms for real time operations in order to minimize number of loss violation instances in a given system.

Keywords: Load Forecasting, Machine Learning algorithms, Support Vector Machine, KNN, Regression.

1. Introduction

It is essential to identify the power system parameters optimally within the specified limits prescribed by the power system operators from time to time. As the load is an important parameter which greatly influences the system performance in view of voltage, power flow, and losses. Hence, the power plant operation is one of the challenging tasks to optimally allocate generations among generators so as to meet the future load. There is an important notable constraint is frequency variation. If the load on system is changed, the frequency varies. As per the standards, this frequency variation should not beat 0.5 Hz. It is always signifying that; the load variations should be forecasted in order to manage the congestions against the load variations. The problem in turn manipulates and improves the operation of the power system with regards to demand, losses, efficiency, generation, transmission, and distribution abilities.

There is a significant and most wanted research area is developing advanced load forecasting methods which considers practical constraints into consideration. There are several important factors such as pressure, temperature, losses, weather conditions, etc. for developing mathematical expressions for the algorithm development. The relation between power system planning and input supplied can be analyzed using numerical analysis presented in [1]. Most importantly, various tools to identify an optimal set of operating conditions to operate the power system by considering practical constraints is explained in [2]. The quantity of load is an important parameter to be considered. And also, the duration of the available load on system changes the operating conditions as well. Hence, the duration and type of load variations should be considered to analyse the system operating conditions [3]. Power quality is one of the important considerations with regards to sensitive consumers. The variations in voltage, current, power should be with in tolerable limits [4].

Prediction of load on a system needs expert mechanism and advanced algorithms. It needs more computational power and accurate procedures. Using these algorithms, the load can be estimated for different durations depending on the operator's requirement. It also helps the grid operators to plan the schedules of the generators, transmission, and distribution system parameters. If the generators are scheduled optimally, the generation cost and as well as total system losses in a given network can be optimized so as to obtain minimum cost with decreased power losses. Hence, the load forecasting procedure should estimate the increase in load by considering different factors such as economic, weather, time factors, etc. The uncertainty raised in the problem due to non-correlated factors must be solved using appropriate algorithms [5]. The existing algorithm have reported some of the technical difficulties such as inaccurate load forecasting, dependence of load on temperature parameters, lack of accurate economic and weather conditions data, insufficient computational configurations, lot of expenditure on erecting novel generating plants/systems, transmission, and power distribution networks for meeting the given demand. These problems upturns the researchers to develop machine learning algorithms to forecast the load on a given power system [6].

In practice, the types of load forecasting methods based on duration can be considered as short, medium, and long terms for which time line varies from few hours to days, months, and years.

From the past decades, the application of OPF problem has been continued to solve many optimization problems [7]. From this study, the OPF problem generally identifies the steady-state operating point of the system to optimize certain power system objectives. There are many classical optimization techniques such as convex and non-convex programming approaches, dynamic programming approaches, Interior point and Newton based techniques, etc. Sometimes, these methods unable to handle continuous and discrete control variables, which mean that these methods, cannot able to obtain an optimal result for a given condition/problem with these control variables [8]. Later the concentration is shifted towards the evolutionary algorithms. These methods are developed to overcome the basic problems raised by the classical optimization techniques [9]. Now a day, these methods are applied to solve many problems in regulated and deregulated power systems. Sometime, using these methods, not only single and as well as multi-objective optimization problems are also solved [10]. Normally, the OPF problem was solved subjected to satisfying system constraints such as operational and physical constraints. Sometime the loading conditions, voltage magnitudes, angle constraints, and active, reactive power generations are also considered. From the past decades, many optimization techniques based on linear and interior point, successive and generalized reduced gradient method, Newton method, successive, evolutionary programming methods [11]. All these methods are population based, competitive and cooperative stochastic search algorithms in the area of computational intelligence. There are many intelligent techniques developed in the recent years inspired from nature. Some of them are simulated annealing, particle swarm optimization, harmony search algorithm, etc. are discussed in [12].

From the careful review of the literature, the following objectives are formulated in this paper

i. Development of forecasting model using different ML Algorithms for given 14 and 30 bus systems.

ii. OPF scheduling for given power system

iii. According to the results obtained in MATLAB, thereby system parameters such as power generation and volage magnitude at generators will be optimally scheduled.

iv. The effect of compensators on system performance will be analysed with supporting numerical and illustrations.

2. Factors Affecting Electrical Load

The present electrical utility system was connected with several electrical appliances which consume power as per the requirement. The pattern that is followed by these devices is changes from time to time. Also, the power consumed varies from device to device. Under these circumstances, it is necessary and most important to estimate certain amount of power on the system so as to avoid the contingency conditions [13].

2.1. Economic Factors

These factors gives the importance of considering cost of power equipment and supply system. The expenditure incurred in order to generate, transmit, and distribute power over a network depends on the reliability and type connection. In order to control the power flow, power generation and power distribution, it is necessary to incorporate the suitable controllers. The cost with regards to install these controllers should be considered as one of the important parameters for estimating the future load using suitable algorithms [13]. These factors are greatly influenced by the forecasting duration and demand pattern.

2.2. Time Factors

These factors gives the impact of forecasting duration as per the operator's requirement. There are three different forecasting types such as short, medium, and long term. The variation of load on a system should be monitored and has to be considered as one of the inputs to ML algorithms in order to develop an approximate model. The algorithm should analyze the timely varying load and should generate an approximate mathematical model in terms of the input variables [14].

2.3. Weather Factors

These factors gives the variation of consumption of electricity based on seasons. The electricity consumption is high during summer, moderate during rainy and less during winter seasons. Also, different weather condition related parameters such as humidity, pressure, precipitation, solar irradiance, wind speed, etc. affects the power generated from the respective generating units. It is necessary to consider the variations of those parameters in order to forecast the load as per the requirement [13].

2.4. Other Factors

There are other factors which indirectly influences the quantity of load on supply system. For example, significance of some days, holidays, festival days, etc. Also, the absence of operating personnel, employees, engineers, etc which increases the forecasting procedure more difficult [13]. It is necessary to consider these factors in to account in order to estimate the load.

3. Mathematical Modelling of Forecasting Model Using Machine Learning (ML) Algorithms

ML algorithms are considered as one of the best applications of artificial intelligence (AI). Using these algorithms, various problems related to electrical engineering can be solved. As the procedure involves the learning based on mathematical interpretations and predefined estimations. The intelligence based algorithms requires the necessary data to train the system. Very important and very frequently used algorithms are reported in [15].

There are various problem which in turn used the previous data to infer the necessary information and possibilities to develop profitable mathematical expressions and relations in order to predict/estimate the future values. This needs lot of previous data and as well as effective algorithmic steps to get interpretation in order to get historical relationship. These algorithms can be used to perform image recognitions, intrusion detection, forecasting of various parametric values [16].

There are various classifications in ML algorithms such as supervised, unsupervised, semi supervised, and reinforced learning algorithms. These algorithms were developed to perform regression, classification, clustering, and association operations on the given data. From these processes, it is possible to formulate generalized pattern against the considered hypothesis for the given problem. The segregation and classifications are very frequency used methodologies in data science to solve problem suing stochastic algorithms, rule, logic, and instance based methodologies. The efficacy of these ML algorithms can be measured in terms of speed, complexity, accuracy, and risk against the measured [17].

Linear Regression: This method is implemeted on the data with a set of variation which are wither dependent or independent in nature [18]. The relationship can be drawn by fitting the best line in the form of a curve. The line equation can be represented using the fundamentals can be expressed as

$$y = a * x + b \tag{1}$$

Logistic Regression: This algorithm works based on classification procedure can be used to interpret the relation among the given set of variatbles which independent. The dependent variable's result is discreate. Traditional logistic regression analysis used in binary classification problem, but it has countless iterations and takes a long time to train large amount of data, which is not appreciable [19]. These algorithms are one of the classifications under supervised learning. In this, the relation between variables is discovered using gradient descent methodology and statistical procedures [20]. The mathematical expression for loss function can be expressed as

$$y = \beta_0 + \beta_1 x \tag{2}$$

Decision Tree: This algorithm used a strcture based on tree model. Using this structure, a course of action or procedure has been identified. In this, the occurance, decision and reactions can be considered as a branch of tree. Here, each tree has a set of nodes (patterns drawn from input data) which are internal and leaf nodes (categories considered from patterns). These algorithms can be considered as one of the finest replacements to statistical analysis. Using this, the text and missed data in a data set can be extracted. There are several variations of this algorithm which are used to solve various problems in diversified fields [21, 22].

Naive Bayes: This algorithm works based on classification procedure using Baye's theorem. This algorithm designed based on combinations of different mathematical computations based on probability. The purposed of using these computations is to find the accurate fit of the line with in a given solution domain. The solution is obtained with in the problem constraints. These algorithms are used for solving different problems in diversified fields [23]. The target classifications are identified from the given dataset and different ways.

Support Vector Machine (SVM): This algorithm is one of the best algorithms under the classification of supervised learning. Using this methodology, both problems related to classification and regression can be solved. In this methodology, a boundary line/limits are estimated to the n-dimensional solution space. Using this, it is always preferable to separate the set of variables by plotting a plane which bisects the entire data into two parts. This plane is called as hyper plane. In order to plot this plane, it uses formulation of vectors also called as support vectors. The number of these vectors is decided based on number of dimensions considerd for the problem [24].

Random Forest: This method uses construction and formulation of multiple decision trees in order to train the system. After this, the decision can be made based on the decision suggested by majority of trees. The entire procedure works based on the classification of the variables and regression analysis based on deciding the decision trees randomly [25, 26].

K-Nearest Neighbour's Algorithm: This algorithm is one of the easiest algorithms out of the classification of supervised learning methodology. It works based on classification of the variables. The classification mechanism works based on the nature of neighbour. In this algorithm the nature of other control parameters is also taken into consideration. The performance is estimated using RMSE, MAE, and accuracy score values [27].

Some authors have reported the forecasting problem can be solved using various methodologies developed based on Neural Networks [28-32]. These methods solve the forecasting problem under practical considerations such as real time weather conditions, community importance, energy transactions, etc..

4. Optimal Power Flow (OPF) Problem Formulation

The commonly used representation of OPF problem is mathematically expressed in Eq (3). This problem gives the solution within the solution search space. The OPF problem has two different set of variables comprises of state 'x' and control 'u' parameters. For electrical power systems, the generatlized expression for representing OPF problem can be given as

$$\text{minimize } A(x, u) \tag{3}$$

The problem is subjected to satisfy two different set of constraints (equality and inequality). The slack generators' active power generation ($P_{g,slack}$), reactive power generation by the generating units (Q_G), voltages at load buses (V_L) and power flow in transmission lines (S_l) are considered to be the state vector parameters which are capable of changing their values based on the control parameters. Where as the remaining set of self-reliaing parameters also called as control vector parameters are active power generation by the generating units (P_G), voltages at generator buses (V_G), tap settings of transformers (T) and reactive power compensated by the shunt capacitors (Q_{sh}).

The mathematical representation of the above said vectors can be expressed as

$$x^T = \left[P_{G1}, V_{L1}, \dots, V_{L_{NL}}, Q_{G1}, \dots, Q_{G_{NG}}, S_{l1}, \dots, S_{l_{nl}}\right]$$

$$u^T = \left[P_{G2}, \dots P_{G_{NG}}, V_{G1}, \dots, V_{G_{NG}}, T_1, \dots, T_{NT}, Q_{sh1}, \dots, Q_{sh_{NC}}\right]$$

Here in this paper, the total number of generator and load buses, shunt compensators, tap changing transformers and total transmission lines are represented with notations 'NG', 'NL', 'NC', 'NT', and 'nl' correspondingly.

4.1. Transmission Power Losses (TPL)

The power losses objective can be considered to be one of the most important objectives in power system analysis, operation and control aspects. This objective plays a key role to decide the amount of power generation by the generating units. If the power is allowed to flow through a transmission line which offers resistance and reactive for the current flow causes losses which are related to active and reactive powers. These losses are function of voltages at the buses which connected to the transmission lines on both sides. Whenever the current drawn by the load is increased the losses values in turn automatically increased. It is always suggestible to operate the power flow through with in permissible limits i.e. below its thermal limit. Most of the researchers will consider power losses is one of the game changing parameter of the power system operation. The expression for the loss objective can be deduced by using the line conductance (g_i), the respective voltage magnitude ($|V_i|$, $|V_j|$) and angles (δ_i, δ_j) of ith and jth buses is as follows

$$A(x, u) = TPL = \sum_{i=1}^{nl} g_i \left[V_i^2 + V_j^2 - 2V_i V_j cos(\delta_i - \delta_j)\right] MW$$

The constraints considered are explained as follows:

4.2. Constraints

As reported earlier, the OPF problem is solved for the following constraints (equality and inequality)

4.2.1. Equality Constraints

The conventional power flow equaitons which are formulated using total power generation (P_{Gi}, Q_{Gi}), total demand (P_{Di}, Q_{Di}) supplied and total power losses with respect to active and reactive powers. The expressions for representation in terms of magnitude $|Y_{ij}|$ and angle θ_{ij} of bus admittance are as follows

$$P_{Gi} - P_{Di} - \sum_{j=1}^{N_{bus}} |V_i||V_j||Y_{ij}| cos(\theta_{ij} + \delta_i - \delta_j) = 0$$

$$Q_{Gi} - Q_{Di} - \sum_{j=1}^{N_{bus}} |V_i||V_j||Y_{ij}| \sin(\theta_{ij} + \delta_i - \delta_j) = 0$$

4.2.2. In-equality Constraints

Generator bus voltage boundaries: $V_{Gi}^{min} \le V_{Gi} \le V_{Gi}^{max}$

Active Power Generation boundaries: $P_{Gi}^{min} \leq P_{Gi} \leq P_{Gi}^{max}$

Transformers tap setting boundaries: $T_i^{min} \leq T_i \leq T_i^{max}$

Reactive power injected by shunt capacitor boundaries:

$$Q_{sh,i}^{min} \leq Q_{sh,i} \leq Q_{sh,i}^{max}$$

Transmission line flow boundaries: $S_{l,i} \leq S_{l,i}^{max}$

Reactive Power Generation boundaries: $Q_{Gi}^{min} \leq Q_{Gi} \leq Q_{Gi}^{max}$

Load bus voltage magnitude boundaries: $V_i^{min} \leq V_i \leq V_i^{max}$

As the state vector parameters are dependent, hence it is necessary to transform the optimization problem from constrained conditions to unconstrained conditions. The expression using penalty approach by considering penalty quotients related to active power generation (λ_p), voltage (λ_v) reactive power generation (λ_q) and power flow (λ_s) can be expressed as

$$A_{aug}(x,u) = A(x,u) + \lambda_p\left(P_{G1} - P_{G1}^{limit}\right)^2 + \lambda_v \sum_{m=1}^{NL} \left(v_m - V_m^{limit}\right)^2$$

$$+\lambda_q \sum_{m=1}^{NG} \left(Q_{G,m} - Q_{G,m}^{limit}\right)^2 + \lambda_s \sum_{m=1}^{nl} (S_{lm} - S_{lm}^{max})^2$$

Values for the constrained/limited control parameters can be assigned with the values as per the following representation

$$x^{limit} = \begin{cases} x^{max}; x > x^{max} \\ x^{min}; x > x^{min} \end{cases}$$

5. Proposed Modified Gravitational Search Algorithm (MGSA)

The stepwise process of the proposed MGSA algorithm is specified here:

Step 1: Define the problem control parameters.

Step 2: Initialize the solution search space.

Step 3: Generate initial population for the control parameters.

Step 4: Update system data with new population.

Step 5: Solve the objective function value and calculate respective fitness value for all population.

Step 6: Apply two-stage initialization process.

Step 7: Start the iterative process.

Step 8: Calculate gravitational constant, masses of each of the population.

Step 9: Calculate the number of populations with higher mass.

Step 10: Calculate the total force applied on each of the population, start with higher mass population.

Step 11: Calculate the acceleration of each of the population.

Step 12: Calculate the new velocity and using this values the position can be updated for all the control parameters in each of the population.

Step 13: Update system data with new population.

Step 14: Solve the objective function value and calculate respective fitness value for all new population.

Step 15: Repeat steps from 8 to 14 for the considered set of repetitions.

Step 16: Print the result.

6. Results and Analysis

As per the requirement of realistic operating conditions and scenarios in practical, the optimal set of control paramters of the generating units, transmission systems and distribution networks have to be estimated. For this, it is the most wanted problem to solve or to estimate the future load using certain ML algorithms and methodologies/procedures. As per the algorithms reviewed and reported in section-3, the results and analysis in this paper is presented in two cases. Forecasting load by training the algorithms using the previous data and reporting the accuracy of the predicted/estimated values with validations can be considered to be the first case. Second case is identifying an optimal set of control parameters by considering all equality and inequality constraints. The entire analysis is performed on a computer with 4GB RAM, intel core-i7 processor.

In this paper, standard 14-bus and 30-bus systems are taken with 3 years of historical data as train data and 1 year data as test data. for every 15 minutes all parameters' values are recorded and those data are taken as train and test data. Comparison of results of 14-bus and 30-bus systems is performed by considering two performance features such as Accuracy Score and run-time of the problem. The flow chart for load forecasting is shown in Fig.1.

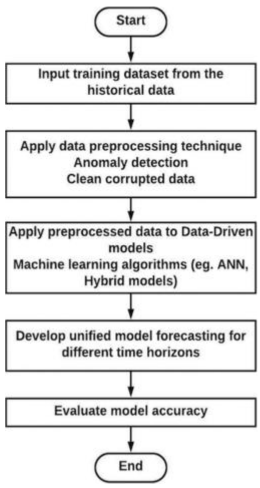

Fig.1. Flow chart of load forecasting using ML Algorithms

All the mentioned Algorithm's forecasting models considered in section-3 are tested on 14 and 30-bus test systems. Initially input training dataset from historical data has been taken for both standard systems for 3years. The train dataset is processed according to our requirement. Next applied preprocessed data to data-driven ML Algorithm models. Developed unified model forecasting for different time zones. Hence evaluated the model accuracy for each ML Algorithm mentioned above. The performance of each Machine Learning algorithm is calculated on basis of RMSE value for Linear Regression, Accuracy-score for all other Machine Learning Algorithms and execution time to build the model for all Machine Learning Algorithms. The RMSE (for linear Regression) values and Accuracy-scores (for all other above mentioned ML Algorithm for 14 and 30-bus systems are listed in Tables 1 and Table 2 respectively.

Table 1. Machine Learning Algorithm results for IEEE 14-bus System

Data	**Linear Regression (RMSE)**	**Logistic Regression**	**Decision tree**	**Naïve Bayes**	**SVM**	**Random Forest**	**KNN**
Traindata2	4.73997	0.10764	0.312678	0.05698	0.046890	0.312678	0.14568
Testdata2	4.33609	0.00483	0.134699	0.06567	0.025768	0.134699	0.04568
Traindata3	2.15844	0.05689	0.245679	0.00587	0.015704	0.245679	0.07867
Testdata3	18.8666	0.09436	0.114579	0.95767	0.035067	0.114579	0.07854
Traindata4	8.78608	0.09425	0.467976	0.06789	0.045079	0.467976	0.04657
Testdata4	9.52793	0.08345	0.354567	0.05680	0.015967	0.354567	0.04956
Traindata5	1.34721	0.07246	0.475540	0.14685	0.035479	0.475540	0.06988
Testdata5	1.51549	0.05485	0.248870	0.08959	0.094768	0.248870	0.09560
Traindata6	2.83232	0.09356	0.178548	0.09864	0.028960	0.178548	0.06896
Testdata6	2.24035	0.08354	0.397647	0.04970	0.019765	0.397647	0.04679
Traindata9	5.86831	0.07345	0.425685	0.08678	0.058897	0.425685	0.07896
Testdata9	5.91049	0.09356	0.115678	0.07967	0.085967	0.115678	0.04678
Traindata10	1.92632	0.04689	0.358909	0.05795	0.022208	0.358909	0.05678
Testdata10	1.79349	0.09357	0.094768	0.09566	0.037497	0.094768	0.04577
Traindata11	6.72194	0.08357	0.289605	0.08067	0.095683	0.289605	0.04687
Testdata11	0.70402	0.06064	0.197653	0.05699	0.070796	0.197653	0.06789
Traindata12	1.90968	0.09358	0.588976	0.09075	0.024567	0.588976	0.05679
Testdata12	1.21539	0.09248	0.085967	0.06794	0.011457	0.085967	0.05678
Traindata13	3.44125	0.08235	0.222086	0.06789	0.046797	0.222086	0.04568
Testdata13	2.69817	0.06890	0.374978	0.09964	0.035456	0.374978	0.05456
Traindata14	3.71945	0.04798	0.956854	0.05679	0.092456	0.956854	0.09654
Testdata14	2.91462	0.89684	0.070796	0.00975	0.056788	0.070796	0.07895

Table 2. Machine Learning Algorithm results for IEEE 30-bus System

Data	**Linear Regression (RMSE)**	**Logistic Regression**	**Decision tree**	**Naïve Bayes**	**SVM**	**Random Forest**	**KNN**
Traindata2	4.73997	0.04057	0.312678	0.06068	0.057878	0.312678	0.06784
Testdata2	4.33609	0.05680	0.134699	0.05875	0.008517	0.134699	0.07896
Traindata3	7.15944	0.06794	0.245679	0.05698	0.024959	0.245679	0.09756
Testdata3	16.8766	0.09495	0.114579	0.06567	0.016979	0.114579	0.04579
Traindata4	11.7860	0.00405	0.467976	0.00587	0.049766	0.467976	0.00456
Testdata4	7.52793	0.04567	0.354567	0.95767	0.037907	0.354567	0.05678
Traindata5	7.34721	0.07804	0.475540	0.06789	0.046890	0.475540	0.05687
Testdata5	9.51549	0.04575	0.248870	0.05680	0.025768	0.248870	0.04588
Traindata7	4.84322	0.10764	0.178548	0.14685	0.015704	0.178548	0.14568
Testdata7	6.54935	0.00483	0.397647	0.08959	0.035067	0.397647	0.04568
Traindata8	9.97831	0.05689	0.425685	0.09864	0.045079	0.425685	0.07867
Testdata8	11.9104	0.09436	0.115678	0.04970	0.015967	0.115678	0.07854
Traindata10	3.96732	0.09425	0.358909	0.08678	0.035479	0.358909	0.04657
Testdata10	4.86349	0.08345	0.094768	0.07967	0.094768	0.094768	0.04956
Traindata12	7.73454	0.07246	0.289605	0.05795	0.0289605	0.289605	0.06988
Testdata12	2.72432	0.05485	0.197653	0.09566	0.0197653	0.197653	0.09560
Traindata14	3.73568	0.09356	0.588976	0.08067	0.0588976	0.588976	0.06896
Testdata14	4.34589	0.08354	0.085967	0.05699	0.085967	0.085967	0.04679
Traindata15	1.63465	0.07345	0.222086	0.09075	0.0222086	0.222086	0.07896

Testdata15	3.59877	0.09356	0.374978	0.06794	0.0374978	0.374978	0.04678
Traindata16	9.59875	0.04689	0.956854	0.06789	0.0956854	0.956854	0.05678
Testdata16	7.36086	0.09357	0.070796	0.09964	0.070796	0.070796	0.04577
Traindata17	4.40703	0.08357	0.245679	0.05679	0.0245679	0.245679	0.04687
Testdata17	5.45809	0.06064	0.114579	0.00975	0.0114579	0.114579	0.06789
Traindata18	8.67938	0.09358	0.467976	0.09754	0.0467976	0.467976	0.05679
Testdata18	17.4396	0.09248	0.354567	0.08965	0.0354567	0.354567	0.05678
Traindata19	18.5670	0.08235	0.475540	0.00974	0.0475540	0.475540	0.04568
Testdata19	8.56893	0.06890	0.248870	0.09764	0.0248870	0.248870	0.05456
Traindata20	9.56989	0.09257	0.178548	0.09567	0.0178548	0.178548	0.04568
Testdata20	10.5157	0.08367	0.397647	0.04568	0.0397647	0.397647	0.07868
Traindata21	6.50895	0.00853	0.425685	0.07899	0.0425685	0.425685	0.04568
Testdata21	7.56790	0.04079	0.115678	0.09567	0.0115678	0.115678	0.04568
Traindata23	7.07978	0.04368	0.358909	0.09853	0.0358909	0.358909	0.06789
Testdata23	11.7789	0.08345	0.094768	0.09457	0.094768	0.094768	0.05979
Traindata24	6.95672	0.07468	0.289605	0.05580	0.0289605	0.289605	0.06790
Testdata24	4.86349	0.14689	0.197653	0.14579	0.0197653	0.197653	0.1679
Traindata26	8.73494	0.08546	0.588976	0.08996	0.0588976	0.588976	0.05678
Testdata26	2.59302	0.04885	0.085967	0.07890	0.085967	0.085967	0.06789
Traindata29	7.70568	0.13578	0.222086	0.16794	0.0222086	0.222086	0.16784
Testdata29	8.35679	0.04067	0.374978	0.05896	0.0374978	0.374978	0.07859
Traindata30	3.65346	0.04676	0.956854	0.08905	0.0956854	0.956854	0.04589
Testdata30	5.56677	0.16079	0.707965	0.14577	0.070796	0.707965	0.14580

In this paper, for both systems, 3 years of historical data is used to train system/algorithm and 1 year of future data to test the accuracy of the algorithm. Hence Decision Tree has given best accuracy score and execution is less than other Machine Learning Algorithm. Later on, load flow analysis is performed on 14 and 30 bus systems are plotted the ploss with respect to time in hours shown in Figs. 2 and 3 respectively. The main inference to be drawn from the results is to minimize the number of loss violation instances.

6.1. Load Flow Analysis Results

6.1.1. For 14-bus System

For this system, the minimum and maximum power losses ($Ploss_{min}$ and $Ploss_{max}$) are 4.999873 MW and 19.99645 MW. The mean loss value ($Ploss_{mean}$) is 12.51937 MW. From the load flow results, it is notified that, there are 6748 instances where the loss values are more than the mean value. Which requires the identification of optimal set of control parameters to minimize the number of power loss violation instances.

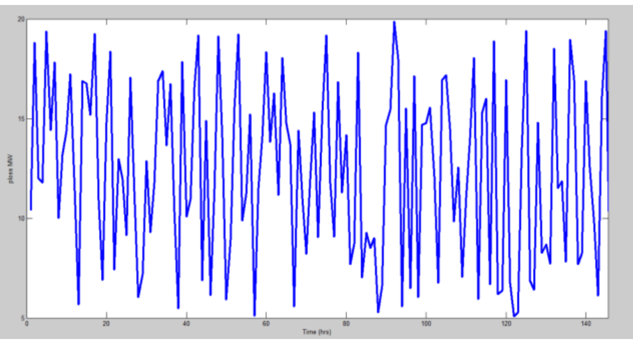

Fig.2. Variation of power loss of 14-bus system

6.1.2. For 30-bus System

For this system, the minimum and maximum power losses ($Ploss_{min}$ and $Ploss_{max}$) are 3.984753 MW and 29.97845 MW. The mean loss value ($Ploss_{mean}$) is 23.76537 MW. From the load flow results, it is notified that, there are 5773 instances where the loss values are more than the mean value. Which requires the identification of optimal set of control parameters to minimize the number of power loss violation instances.

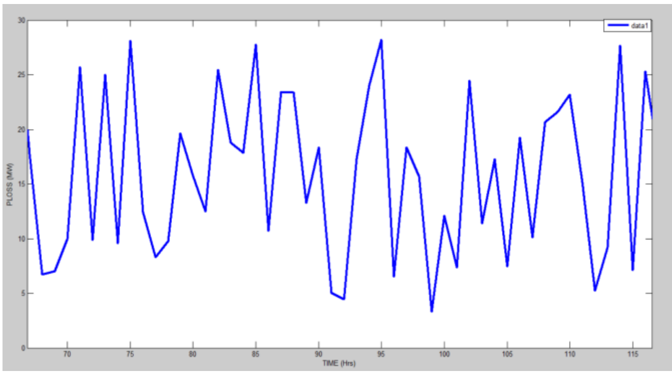

Fig.3. Variation of power loss of 30-bus system

6.2. Result and Analysis of OPF

The proposed OPF method is tested on the considered test systems. Initially, Formulated OPF problem is solved to minimize the transmission losses as per the parameter constraints. The number of iterations considerd for the OPF problem is 100. The number of trails perfomed to confirm the OPF results is considered as fifty. Out of all trails, the best OPF results are identified and presented for all the considered control parameters.

6.2.1. For 14-bus System

OPF results of 14 bus system system are tabulated in Table 3 and corresponding convergence characteristics are shown in Fig.4. From these results, it is inferred that, the control paramerters are adjusted within the given limits to obtain minimum power losses. After this, at these control parameters are set as the base operating conditions and load flow is executed to identify the number of loss violation instances. For this system, after executing this procedure, it is notified that, there are 1240 which is less than 5508 instances. Hence, it is always suggestible to perform optimal power flow to identify an optimal set of control parameters for the forecasted load conditions.

Table.3 OPF (with MGSA) results of IEEE 14-bus system

S.No	Control Parameters		With MGSA
1	Real power Generation (MW)	P_{G1}	10.171
		P_{G2}	54.753
		P_{G3}	60.000
		P_{G4}	50.000
		P_{G5}	31.155
2	Generator Voltages (p.u.)	V_{G1}	1.053
		V_{G2}	1.048
		V_{G3}	1.038
		V_{G4}	1.064
		V_{G5}	1.067
3	Transformer Tap setting (p.u.)	Tap_1	0.973
		Tap_2	0.932
		Tap_3	0.977
4	Shunt compensator (MVar)	Q_{c1}	8.883
5	Total generation (MW)		206.079
6	Total power loss (MW)		1.102
7	Time(sec)		30.240

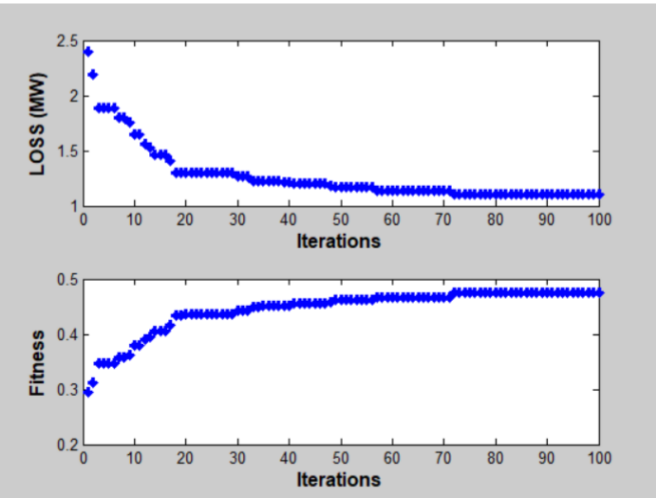

Fig.4. Convergence characteristics of 14-bus system

6.2.2. For 30-bus System

OPF results of 30 bus system system are tabulated in Table 4 and corresponding convergence characteristics are shown in Fig.5. From these results, it is inferred that, the control paramerters are adjusted within the given limits to obtain minimum power losses. After this, at these control parameters are set as the base operating conditions and load flow is executed to identify the number of loss violation instances. For this system, after executing this procedure, it is notified that, there are 2341 which is less than 3432 instances. For this system, the instances value is less when compared to previous system. This is due to change in total load value, network configuration, generation capacities, etc. Hence, it is always suggestible to perform optimal power flow to identify an optimal set of control parameters for the forecasted load conditions.

Table.4 OPF (with MGSA) results of IEEE 30-bus system

S.No	Control Parameters		With MGSA
1	Real power Generation (MW)	P_{G1}	50.122
		P_{G2}	28.514
		P_{G3}	50.000
		P_{G4}	35.000
		P_{G5}	30.000
		P_{G6}	40.000
2	Generator Voltages (p.u.)	V_{G1}	1.050
		V_{G2}	1.043
		V_{G3}	1.034
		V_{G4}	1.043
		V_{G5}	1.050
		V_{G6}	1.050
3	Transformer Tap setting (p.u.)	Tap_1	1.016
		Tap_2	0.997
		Tap_3	0.992
		Tap_4	0.973
4	Shunt compensator (MVar)	Q_{c1}	17.173
		Q_{c2}	11.466
5	Total generation (MW)		233.636
6	Total power loss (MW)		1.879
7	Time(sec)		66.7384

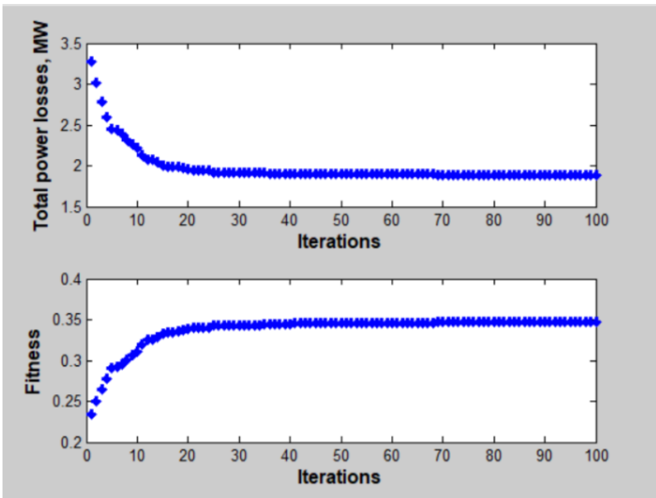

Fig.5. Convergence characteristics of 30-bus system

7. Conclusion

In this paper, popular ML Algorithms such as Linear Regression, Logistic Regression, Navie Bayes, Random Forest, Decision Tree, Support Vector Machines, and KNN have been used to estimate the forecasted load for 14 and 30 bus systems. The performance of the forecasting model is examined on basis of RMSE, Accuracy-Score and time taken to generate the model. By comparing all above-mentioned Machine Learning Algorithm, Decision Tree provides best Accuracy- Score in less time. From load forecasting results, it has been noticed that, load forecasting using Decision Tree have given best results when compared to other Machine Learning Algorithms From the load flow analysis on the test systems, it has been noticed that, there are more number of loss violation instances. This is due to the operation of system control parameters at fixed operating conditions. Further, the number of loss violation instances has been minimized by solving OPF without violation of system constraints. It has been also noticed that, MGSM optimization algorithm along with step-by-step procedure has been presented in detail. From the OPF results it has been noticed that, the proposed algorithm identifies optimal set of control variables (P_G, V_G, T, Q_{sh}) for minimizing power loss violation instances and total power losses.

References

[1] Dharamjit and D.K. Tanti, “Load flow analysis on IEEE 30 bus system”, International Journal of Scientific and Research Publications, Vol.2, Issue 11, Nov, 2012.

[2] Afaneen Anwer, Ali H. Almosawi and Ghazwan Alshabbani , “Optimal power flow for Iraqi power system”, 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA).

[3] Al Mamun M., and Negasaka K., “Artificial neural networks applied to long-term electricity demand forecasting,” Proceedings of the Fourth International Conference on Hybrid Intelligent Systems (HIS'04), pp. 204-209, Dec. 2004.

[4] Dang Khoa T. Q. and Oanh P. T., “Application of Elman and neural wavelet network to long-term load forecasting,” ISEE Journal, track 3, sec. B, No. 20, pp. 1-6, 2005.

[5] Al-Hamidi H. M. and Soliman S. A., “Long- term/mid-term electric load forecasting based on short-term correlation and annual growth,” Electric Power System Research (Elsevier), Vol. 74, No. 3, pp. 353-361, June 2005.

[6] Negasaka K. and Al Mamun M., “Long-term peak demand prediction of 9 Japanese power utilities using radial basis function networks”, IEEE Power Engineering Society General Meeting, Vol. 1, pp. 315-322, 6-10 June 2004.

[7] M. R. AI Rashidi, M. E. EI-Hawary, “Hyrid particle swarm optimization Approach for solving the discrete OPF problem considering the valve loading effects”, IEEE Transactions on power systems, 2007, Vol.22, No.4, pp.

[8] M. R. AIRashidi, M. E. EI-Hawary, “Applications of Computational Intelligent Techniques for Solving the revived optimal power flow problem”, Electric power system research, 2009, Vol.79.

[9] A.G.Bakirzis, Pandel N. Bikas, C.E.Zoumas, V.Petridis, “Optimal power flow by Enhanced Genetic Algoritm”, IEEE Transactions on power systems, 2002, Vol. 17, No.2.

[10] M.Varadarajan, K.S..Swarup, “Solving multi-objective optimal power flow using differential evolution”, IET Generation, Transmission & Distribution, 2008.

[11] Hueatult M, Galiana FD., “A survey of the optimal power flow literature”, IEEE Trans power syst, 1991, Vol.6, No.2, pp.762-770.

[12] Roa Sepulveda CA, Pavez Lazo BJ., “A solution to the optimal power flow using simulated annealing”, International Journal of Electrical power energy systems, 2003, Vol.25, pp.47-57.

[13] G. Gross and F.D. Galiana, “Short-Term Load Forecasting,” Proc. IEEE, vol. 75, pp. 1558-1573, Dec. 1987.

[14] T. Hong, M. Gui, M. Baran, and H.L. Willis, “Modeling and forecasting hourly electric load by multiple linear regression with interactions,” Power and Energy Soc. General Meeting, Minneapolis, MN, 2010, pp. 1-8.

[15] Susmita ray, “Quick review on machine learning algorithms,” IEEE 2019 International Conference on Machine Learning , Big Data, cloud and parallel computing COMIT Con.

[16] Sheena Angra and Sachin Ahuja, “Machine learning and its applications: Review,” 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC).

[17] Amanpreet Singh, Narina Thakur and Aakanksha Sharma, “A Review of supervised Machine Learning Algorithms,” 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).

[18] Hyun-Il Lim, “A Linear Regression approach to modelling software characteristics for classifying similar software,” 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC) .

[19] Xiaonan Zou; Yong Hu; Zhewen Tian; Kaiyuan Shen, “ Logistic Regression model optimization and Case Analysis,”2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT).

[20] Kaixin Zhou; Wei Nai; Shuoxian Zhu; Shupei Zhang; Ying Xing; Zan Yang; Dan Li, “Logistic Regression based on Bat inspired Algorithm with Gaussian Intialisation,” 2021 IEEE 5th Advanced

Information Technology, Electronic and Automation Control Conference (IAEAC).

[21] Arundhati Navada; Aamir Nizam Ansari; Siddharth Patil; Balwant A. Sonkamble, "Overview of use of Decision Tree algorithms in Machine Learning,"2011 IEEE Control and System Graduate Research Colloquium.

[22] Philip H. Swain; Hans Hauska, " The Decision Tree Classifier: Design and Potential," IEEE Transactions on Geoscience Electronics (Volume: 15, Issue: 3, July 1977).

[23] Feng-Jen Yang, "An Implementation of Naïve Bayes classifier," 2018 International Conference on Computational Science and Computational Intelligence (CSCI).

[24] Sourish Ghosh; Anasuya Dasgupta; Aleena Swetapadma, "A Study on Support Vector Machine based Linear and Non-Linear pattern Classification," 2019 International Conference on Intelligent Sustainable Systems (ICISS).

[25] M. Suresh Kumar, V. Soundarya, S. Kavitha, E.S.Keerthika, E. Aswini, "Credit Card Fraud Detection using Random Forest Algorithm," 2019 3rd International Conference on Computing and Communication Technologies.

[26] Shivam Mishra; Aakash Shukla; Sandeep Arora; Himandhu Kathuria; Mandeep Singh, "Controlling Weather Dependent Tasks using Random Forest Algorithm," 2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC).

[27] Metodija Atanasovski; Mitko Kostov; Blagoja Arapinoski; Mile Spirovski, "K-Nearest Neighbor Regression for forecasting electric demand," 2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST).

[28] T. Fujiwara and Y. Ueda, "Load forecasting method for Commercial facilities by determination of working time and considering weather information," 2018, 7th International Conference on Renewable Energy Research and Applications (ICRERA), Paris, France, 2018, pp. 336-341, doi: 10.1109/ICRERA.2018.8567019.

[29] Eya Kalboussi, Nadia Ndhaief, Nidhal Rezg, "Deferred Supplier Energy Amount Prediction Using Neural Network based on Switching Strategy for Resilient Smart Grid," 2023, 11th International Conference on Smart Grid, June 04-07, Paris.

[30] Hussein Abubakar Bakiri, Hellen Maziku, Nerey Mvungi, Ndyetabura Hamisi, Massawe Libe, "A Novel Load Forecasting Model for Automatic Fault Clearance in Secondary Distribution Electric Power Grid using an Extended-Multivariate Nonlinear Regression," 2021, International Journal of Smart Grid, Vol.5, No.2, pp.103-112.

[31] Hussein Abubakar Bakiri, Hellen Maziku, Nerey Mvungi, Libe Massawe, Hamisi Ndyetabura, "Towards the Establishment of a Robust Load Forecasting Mechanism in Tanzania Grid: A Study to Investigate Effect of Air Temperature and Daytime on Electricity Consumption in Residential Buildings," 2021, International Journal of Smart Grid, Vol.5, No.1, pp.23-36.

[32] A. Shahid, "Smart Grid Integration of Renewable Energy Systems," 2018 7th International Conference on Renewable Energy Research and Applications (ICRERA), Paris, France, 2018, pp. 944-948, doi: 10.1109/ICRERA.2018.8566827.