International Journal of Scientific Research and Engineering Development-– Volume 6 Issue 1, Jan-Feb 2023

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Data Assessment of the World's Power Plants Using Various Machine Learning Techniques

Rayees Ahmad Wani*, Baljit Kaur**

*(Department of Electrical Engineering
IET Bhaddal, Technical Campus, Ropar, Punjab, India
Email: Rayees2524@gmail.com)

** (Department of Electrical Engineering
IET Bhaddal, Technical Campus, Ropar, Punjab, India
Email: Jeetbaljit025@yahoo.com)

--************************----------------------------------

Abstract:

Nuclear reactors benefits and expenses, and other their associated impacts, are determined by their technology and the amount of electricity they create. Most nations, especially emerging ones where electricity output is expected to grow significantly, do not disclose plant-level generating statistics. The Global Power Plant Database uses this technical information to estimate the yearly energy generation of power plants. Breeze, light, power (hydro), and gas energy plants are examples of diverse fuel sources. We employ different estimating models. Statistical regression and machine learning techniques are used in the process. Plant-level parameters, as with the size of the facility and the liquefied petroleum gas used, as well as Variables at the regional scale, such as the total performance per kilo of installed power, are explanatory variables. We demonstrate this, for wind, solar, and hydro plants, fuel-specific models can deliver more accurate results. Natural gas plant estimates are also improving, although the margin of error remains considerable, especially for smaller facilities

Keywords — *Nuclear Reactors, Electricity, Global Power Plant, Hydro, Wind, Tidal, Solar, Natural Gas, Fossil Fuels.*

--************************----------------------------------

INTRODUCTION

Modern civilization runs on electricity. Records on actually renewable energy by reactors is often kept proprietary get by outsiders, but its value. The world tourism organization (wri) and thus its colleagues created the global hydroelectric repository (gppd) as a fully accessible, fully accessible assessment of the globe's diesel generators (byers et al. 2018). It covers data on coal plant major components such as volume (gigawatt hours), latitude, and hydrogen fuel, which was assembled from numbers of government documents. The database includes the stations' generators (gigawatt-hours) when it is publicly available. As of june 2019, credible methods of stated plant power were available for 33 nations.by factory + scheme controllers and harder to

The maximum electric power rate of a facility is described by its power plant capacity, which is commonly defined in megawatts (MW). A 100 MW plant will create 100 megawatt-hours of power if it works at full capacity for one hour. In other words, capacity refers to the plant's size and potential production rate, whereas generation refers to the plant's actual power output over time. Thermal power plants generate energy using a variety of fire to create a successful turbine, as well as steam to cool devices. (Toasty) water released toward a water body and air which it evaporated are common discharges, as well as contaminants to the air, water, and soil. Energy planners may utilize past plant generating data to track emissions and determine the best way to fulfil changing energy demand over time. The frequency and intensity with which a power plant operates vary by plant type.

Annual power plant generation can be calculated using statistical models or approaches based on electrical grid optimization (also known as optimum dispatch). Grid optimization models take into account each power plant's technical features as well as the cost of raw fuels prior to actually dispatch the modelled stations to lower overall generating costs under the two constraints conditions:

1. Providing the required amount of power at all times

2. Complying with any technical requirements, such as minimal downtime or steepest incline rates

3. Taking into account Concerns with grid connection these types of algorithms estimate "ideal generation" to resemble plant-level generation, albeit they may not represent real historical generation that was produced “non-optimality”.

OBJECTIVE

The objectives below are hopefully achieved in the thesis.

• For its first ever, times series data out of a diversity of ways has been utilized to motivate electric utilities in emerging regions to employ STLF methodologies automated for more reliable voltage regulation.

• To pick the input features for its first times from the fresh unknown set, pattern discovery, visually findings and empirical methods such as engine review, iqr investigation, and block investigation are utilized.

• A complete predicted matrix is generated utilizing predefined controller parameters for linear and non-adjustable STLF model. Potentially, the predictor’s matrix isn't very sophisticated, and data other than histories isn't necessary.

• Using a variety of teaching metrics also including MAPE, RMSE, MSE, D e, plus measure of dispersion, this study compares conventional time-series data sets, linear with non - parametric statistical approaches on a theoretical and practical level. Furthermore, we performed a comprehensive seasonal study to assess and compare the effectiveness of the proposed methods

BACKGROUND AND LITERATURE REVIEW

Velasco, et.al.The load profile modelling results are derived by summing the balanced signals from each DBN using an aggregation teaching approaches. The EMD-based DBN technique, on the other hand, is complex, and it loses the original information about the actual load and power system qualities when choosing IMFs, resulting in considerable forecasts during machine learning algorithms classifier. Unfortunately, it takes a long time to use EMD to partition power system data in to the IMFs and then use Real good to train algorithms on the gathered IMFs . Also, an Adapting Channel Fuzzy Processing Element (ANFIS) boosts the learner at the centre of NNs by merging heuristic algorithm with slightly elevated know about the importance of Boolean controller. The training phase in ANFIS models, on the other hand, is more difficult because like the complex rules, huge proportion of precedents, and shift lags .

Turhanet.al.Technique (PSO) and GA [44] are examples of hybrid Statistical models that integrate evolutionary algorithms techniques or swarm intelligence strategies with biological systems. PSO and GA may be used to find the ideal input data for an ANN in method and non - linear and non ahead STLF frameworks. A GA-based Non-Linear Motor Describe characteristics Network (NARX–NN) manages the loop in the stochastic gradient descent circuit of the NN to either extract irregular rainfall associations in the data database [12,45]. A clustering model like ANN may easily converge to optimal solutions in the varied higher dimensional space, but it fails to diverge to the optimization algorithm.

Jebaraj and Iniyan.A review of the literature is presented in Suganthi and Samuel [68] and recently in Hong and Fan [3] and while, regarding the By employing hybrid ML models, the above-mentioned STLF difficulties may be handled while obtaining SVR, ANN, and KNN are examples of ancient machine learning models with lower classifier.. Every constituent method in hybrid ML models provides resilience and greater accuracy in STLF [36]. The components of graded K-Nearest Peer are optimized using an Optimization Technique (W-KNN)in [33]. (GA). Furthermore, in KNN–ANN and KNN–SVR architectures, the integration of learning algorithms and clustering extracts significant characteristics prior to the personality method, such as altitude, precipitation, and weekdays [37]. Moreover, intricate morphologies and an unlimited amount of ideally clustering limit the usefulness of Support vector machine (svm combination ML models [36].

Ghalehkhondabi et al.The predictions efficacy of neural network models can be improved by incorporating the Stereotype Content Modelling (ELM) also with Haircuts"</h2 (LM) and Contingent Reciprocal Content Classification Techniques (CMIFS). (FF–NN) may be enhanced

Kisi, O.et. al. To increase the performance of the STLF model, a hybrid BA–ELM was developed in . To anticipate the electric load, a Bat Algorithm (BA) optimizes the input weights of an extreme learning machine (ELM) and the bias threshold. The crucial gradient and Hessian computations for the STLF model, on the other hand, increase the temporal complexity of the ELM technique [40]. Furthermore, a deep learning method based on Empirical Mode The EMD method, in which multivariable load demand data is classified into Freestanding Mode Functions (IMFs) and subsequently Pre - trained convolution neural Networking (DBNs) are often used to training the derived IMFs, has really been suggested to increase the classification performance of the STLF framework

.

PROPOSED METHODOLOGY

Our primary aim is to properly estimate yearly past plant-level generation. Models of machine learning are ideally adapted to the task (Olden and colleagues, 2008). We use certified ml models, which define the link between of regression coefficient this case, plant-level annual generation the variable parameters, or forecasters, specified by the authors. Given a set of data that includes both relationship between variables, what should you do? the model is optimized across many iterations with carefully calibrated parameters to minimize the estimation error (called the labelled training data). A huge number of algorithms may be used in machine learning models. As the model algorithm, we used the gradient boosting tree (GBT) regressor. Because of the following, the GBT is an acceptable modeling approach in this scenario.:

1. Regression trees can capture nonlinear interactions (for example, there is no linear connection : daily wind production). As the weather speed is reduced, increases, a turbine will eventually run out of power to create.

2. Tree-based models make it simple to determine which predictors contribute the most to the outcome.

3. Tree-based classifiers forecast by examining equivalent instances in the training phase, assuming that the predictor remains only within training set's targeting variable's range.

4.1.Data Cleaning and Outlier Detection

Although the research is based on official data, inaccurately reported generation or capacity might result in an exaggerated capacity factor for a specific facility. Plants may also be unable to run for a whole year owing to maintenance issues. In this sort of research, we can't foresee extended maintenance intervals, therefore we focus on projecting generation for plants that are constantly accessible throughout the year. Outliers may have an undue impact on the model, resulting in erroneous predictions. To decrease measurement error, we first exclude situations where the generation capacity is greater than or equal to 1. Buildings larger than one are frequently misnamed as generating value or potential. Voltage plants that import more water than they output, such as pressurized storage tanks, can have thermal efficiency that are less than zero. Outlier records with electricity production larger than four standard deviations of the mean (as assessed through all countries for the specified fuel type) are discarded as well. Each component contains details on any further petroleum data purification techniques.

MODEL VALIDATION AND TESTING

Trained, certification, and test sets were created from the specific dataset. The model was applied to the test dataset and, if necessary, iteratedly updated based on the verification set's outcome. After that, the test or mystery data is utilized to evaluate it. The test is 20% of the entire data, split by country to ensure that that match the spatial pattern of the entire particular dataset. We use sorted K-fold stepwise regression to divide the information into testing and levelling. Selection method is used to solve the model over K identically sized subsets, or folds, of the learning algorithm, as shown in Figure 1 for 10 folds. The network is generated on (K-1) folds before being evaluated, or validation, on a sample population that had not been used (Varian 2014). The repetition of the managed to keep random sample is completed. By repeating the experiment K times, we can achieve K certification scores. To identify the highest score is the simple average of the K scores. K-fold analytical method reduces the risk of a new training example to not be inclusive of the whole data and trying to distort the machine (Shulga 2018). It also allows for evaluation during coaching utilizing all of the data, which is very beneficial when data is limited, as it is in our case. To improve trans performances, we transform.

Afterwards, we do the test. set to see how well it performs. Figure 2 depicts the full training, validation, and testing process. Each of the fuel-specific models goes through the same procedure.

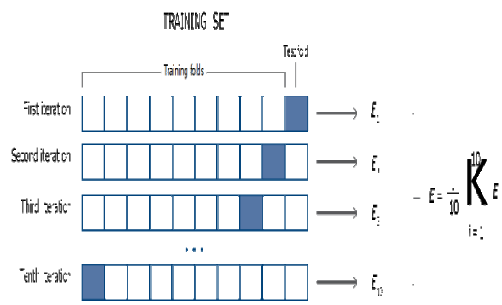


Figure 4.1Cross Validation for Case with K=10 Folds

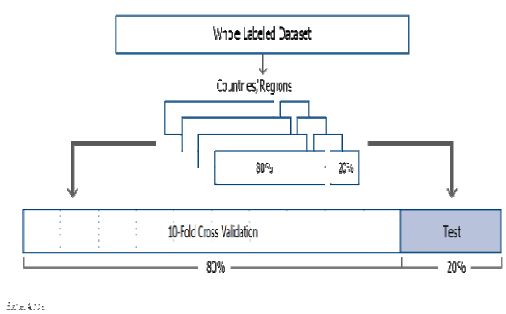


Figure 4.2 Training, Validation, and Test Split

MODEL PERFORMANCE EVALUATION

TO ASSESS MODEL PERFORMANCE, WE USE TWO MEASURES TO COMPARE THE ESTIMATED AND REPORTED CAPACITY FACTORS: THE STANDARD ERROR (MAE) AND RELATIVE ABSOLUTE CONFIDENCE INTERVAL (MAPE) ARE TWO TYPES OF ERRORS (MAPE)

THE MAE RIGHTS IN RESPECT REAL AND ANTICIPATED READINGS AND **ANALYSES** THE PERCENT DIFFERENCE OVER ALL MEASUREMENTS, PROVIDING A SIMPLE ERROR MEASURE, BUT IT WILL NOT EVALUATE THE POSSIBLE ERROR SIGNAL. THE MAPE, FROM THE OTHER HAND, HAS NO UNITS, MAKING IT EASIER TO ASSESS PERFORMANCE OVER A RANGE OF ELECTRICITY PRODUCTION.

TO ASSESS THE EFFECTIVENESS OF THE ALGORITHM, BOTH THE MAE AND MAPE GAUGE THE ACTRESS'S ERROR. THRU OUT THE RESEARCH . THE DATA, WE COMPARE THE SIZE AVAILABLE TO THE INEFFICIENCIES THAT COME FROM THINKING THAT GENERATION PER KILOWATT IS CONSTANT FOR SITES TO THE SAME FUEL TYPE WITHIN AN AREA (THIS SAME FOUNDATION MODEL), AND WE DEMONSTRATE THAT THE APPROACHES WE EMPLOY ARE MORE THAN THE TEST SET.

Data about the model To produce power, energy power plants use a variety of methods, including cogeneration cylinders (CCGTs), open cycle gas jets (OCGTs), and vintage diesel generators (STs).

GENERATING TECHNOLOGY	AVERAGE CAPACITY FACTOR	NUMBER OF PLANTS
Combined cycle gas turbine	0.42	417
Single shaft combined cycle gas turbine	0.45	17
Fuel cell	0.66	32
Combustion gas turbine	0.18	543
Internal combustion engine	0.23	88
Steam turbine	0.13	128

Table 1Natural Gas Average Capacity Factor by Generating Technology, United States (2016)

Plant efficiency is linked to age, with younger Because the plants are more economic, they are deployed more commonly. Because larger facilities are used more frequently and have a greater operational rate, throughput is likewise connected to more rapid delivery. The gradient boosting tree model is given production volume, size, & device type, but also the coefficient of performance per area, to quantify the expected base load for each factory. The approach is depicted in Figure 3. For 1,482 plants for whom 2016 power data is currently lacking in GPPD, we can use the methodology to anticipate productivity. Because the 487 plants by which we do not have a new tech classification are not attached to any facts in WEPP, we will remain to infer capacity.

SYSTEM ARCHITECTURE

Data

Global Database of Power Plants the GPPD includes 2,598 gas-fired power stations that were developed before to 2016, with over 70% of them producing generating. North Asia is home to 66 basis points of all petroleum, but 93 basis points of those who already create power, as seen in Table 3. In parallel to other fuels, gas is used in much electricity production, and many have numerous cylinders with various generating technologies. We only model gas plants with a component of gas-based capacity more than 95% (relatively to all fuels) and a percent of any output technologies (subsumed as CCGT, CS, GT, ST, IC, or FC) bigger than 95% of the power generation for comparison's sake. The set of variables in the training images is reduced from 1,780 to 1,284 as a result of this.

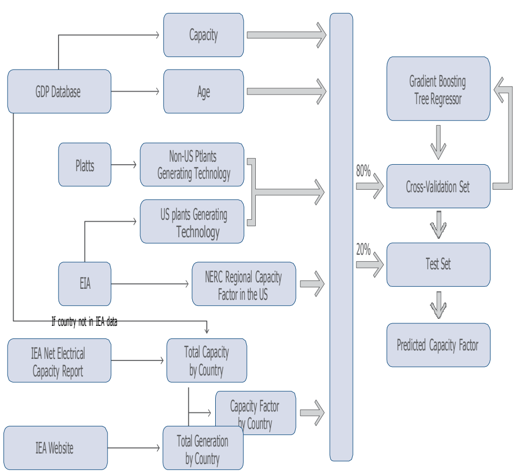


Figure 5.1Gas-Fired Power Plant Generation Estimation Workflow

	NUMBER OF GAS PLANTS BY REGION	GAS PLANTS BY REGION (% OF WORLD PLANTS)	NUMBER OF GAS PLANTS WITH REPORTED GENERATION DATA (PLANTS THAT CAN BE USED FOR TRAINING)	GAS PLANTS WITH REPORTED GENERATION DATA (% OF WORLD PLANTS WITH REPORTED GENERATION)
NORTH AMERICA	1,708	65.7%	1,651	92.8%
SOUTH AMERICA	128	4.9%	0	0.0%
EUROPE	267	10.3%	73	4.1%
AFRICA	68	2.6%	1	0.0%
ASIA	425	16.4%	55	3.1%
AUSTRALIA/OCEANIA	2	0.1%	0	0.0%
TOTAL	2,598	100%	1,780	100%

Table 2Gas-Fired Plants and Plants with Reported Generation by Region (2016) in the GPPD

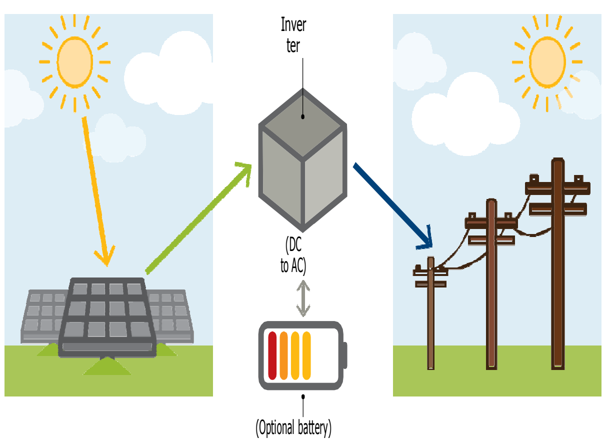
SOLAR PHOTOVOLTAIC

This subsection is about pv solar plants. Rooftop solar plants grow up only 1.6 percent of the national solar capacity, hence they aren't included in this analysis. Characteristics of the model (6.1) Figure 11 depicts the essential features of a grid-connected photo - voltaic (PV) system. In Pv array, the solar cell turns entering energy from the sun. After energy is created, aspects of the consistency (BoS) help regulate it so that map output is produced. And some of the most important BoS component is the rectifier, which transforms electricity from dc current (DC) to pulsating dc (AC) so that it may be distributed.When weather conditions such as rainfall exist, the Solarpv may include a tank or some other kind of storage method, helping it to provide accurate outcomes rather than erratic output. One of the important components in the number of calories provided per MW of annual production is the proportion of solar rays produced by the photovoltaic panels, which may be approximated using the seasonal average world perpendicular incident (GHI) somewhere at surface level. Heat is also a role, since heatwaves diminish the efficiency of the pv (Dubey et al. 2013). As a function, the research includes the yearly average water temperature. PV detection rate with time, decreasing by 0.5 percent to 1% each year and (Jordan and

Kurtz 2013). The PV may very well be able to obtain this impact. .Other factors, such as solar photovoltaic systems, that develop over time but aren't evident in the data, might potentially be connected to age. We also looked at designs that didn't contain age since it's not generally documented. We acquire interfacial GHI and surrounding heat flux from MERRA-2 at a scale of 0.5° by 0.625°. . Data on irradiation level should always be calculated from nearby voxels in order to create.

To summarize, the thermal efficiency is the photovoltaic (pv model's response variable, while the factors are land annually actual average laterally irradiance; annual total heat flux at the solar farm's location; and per year average world downward insulation at the photovoltaic (pv farm's site.

age of the plant, power of the plant, andby state solar generation capacity.



Notes: DC stands for direct current, AC for alternating current.

Figure 5.1Simplified Grid-Connected Photovoltaic System

The model outputs are more consistent than the basis estimates based also on holdout set of 263 trials, as seen in Table 9 and Figure 13. A generating farm's volume figure is expected to be fewer than 15% of its true rated capacity on estimate accurate test set. The datasets on the top of figure 13 clump in around closed interval, demonstrating that the determinants are effective at interpreting solar farm variation within regions. As observed on the best side of the image, the royalty payments by each exemplar in the test set are proportionally and important indicator, showing that the model has stable enough to produce consistent judgments.

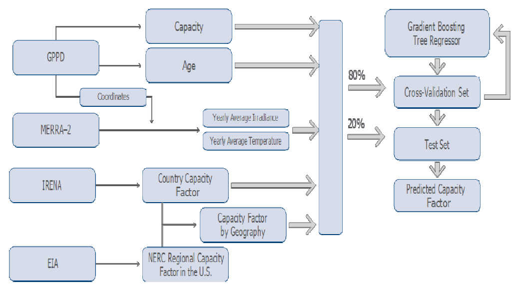


Figure 5.2 Workflow of solar PV

Hydropower plant

Hydro power includes hydroelectric energy projects using walls, local mill plants, and back - ordered. This study excludes piped storage hydropower (PSH) plants and they're storage instead of just a producing device. PSH plants have a negative yearly gdp because water flowing to an upper reservoir consumes more energy than the amount produced once the water is returned to the engine.

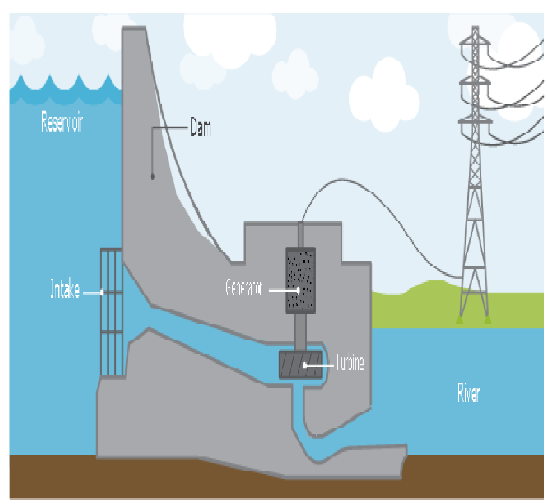


Figure 5.3 Simplified Hydro **power plant.**

The stored energy can be quickly utilised to meet utilities or district consumption. Although pumped containers are certainly essential parts of various energy circuits, their yearly activity as essentially electricity middlemen is tough to track. Description of the model (7.1) Small hydropower facilities uses water to drive turbines that create electricity (Figure 16). In major hydroelectric projects, structures and ponds are often used. Cheaper particles are mainly take plants with only a weir to pool the runoff rather than just a dam. The weir redirects a water from the main large rivers to a turbine (Paish 2002). Pumped hydro is generated by accumulating rain and water in the basin region where the product is produced. to be prepared

1. Plant output, which refers to the highest quantity of energy that may be generated at any one time.
2. Runoff median for the nuclear reactor location
3. The size of the river running into the pond is determined by the river pattern..

Longer grades are used to refer to rivers flowing. Order 1 refers to the internodes river spanning source to destination; order 2 refers to all streams falling in to the first river; order 3 refers to all headwaters feeding into a first river; and order 0 refers to composite of tiny coastal districts (Lehner and Grill 2013).

4. Per year combined cycle by country, which uses data from additional nations or areas that we don't observe or monitor direct. The administration of hydro power plants is frequently limited by filled to the brim and legislative compliance, which may include pollution controls (Niu and Insley 2013).

Well over period of the year, weekly demands fade away, and runoff volume becomes the most important component in generating (Kao et al. 2015). When stormwater replies swiftly to rainstorms, but deep, or **stream flow**, discharge responses very slowly, we divide surface runoff in the model (Kao et al. 2015). Tracking flow at the reservoirs location provides less evidence since water gathers across a river zone.

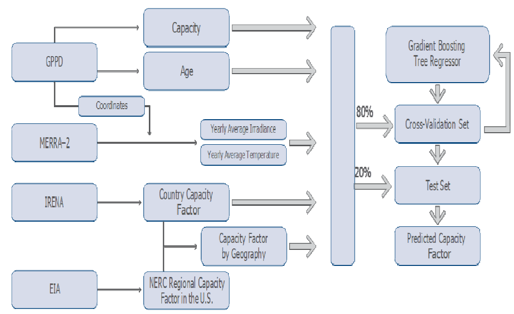


Figure 5.4 Workflow of Hydro power plant

We used total recordings over the critical river territory as a generator of formation, as suggested by Kao et al (2015). Using ERA5 global temperature data and Hydro BASINS, the flood area is covered and important wind speed within the region is gathered. Hydro BASINS is a set of geometric layering that illustrates catchment and comment section boundaries on a global scale. Each triangle has its own ID and is connected to the one ahead. Using the flower data, we discover the polygon where every plant is located (e.g., polygon 1 in Figure 17), and then reverse all upward polygons to discover whole headwaters. The shafts that cross squares represent actual water flow and accumulation. .Each polygon's upper hexagon is determined by Hydro BASINS. We find 3 by looking above in the upper circle of 2. We'll keep looking until we come across a region with no adjacent quads . The runoff area of these plants is determined using the aggregate of all circles acquired in this technique.

RESULTS AND DISCUSSION

Electricity plants in order to generate a specified amount of energy over a timeframe, but once they are turned off, it are no longer providing energy

The Capacity Factor

Buildings can be used by energy fanatics to evaluate the dependability of alternative power plants. It essentially tracks the number of times a device operates at full capacity. A unit with a rated capacity of 55 percent produces power all of the time.

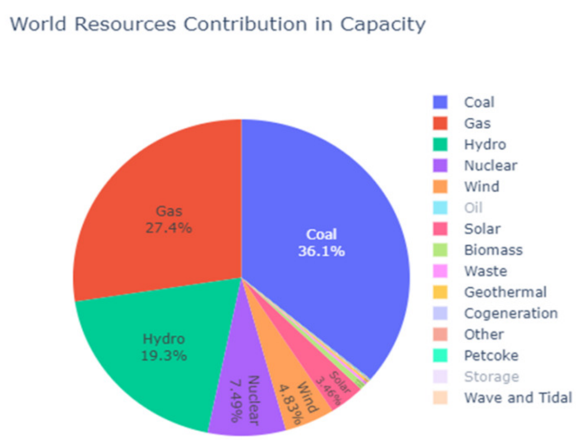


Figure 6.1World Resources Contribution in capacity

```
df_generation = df.groupby('country')['estimated_generation_gwh_2020'].sum().sort_values(ascending=False)
.to_frame()

px.bar(df_generation,title='Electricity Generation Per Country')

from matplotlib import pyplot as plt

import pandas as pd

from collections import Counter

data=pd.read_csv('../input/global-powerplants/powerplants
(global) - global_power_plants.csv')


countries=data['country']
country=[]
fori in countries:
country.append(i)

cnt=Counter(country)

arr=cnt.most_common(15)
x_values=[]
y_values=[]
forelement inarr:
x_values.append(element[1])
y_values.append(element[0])

y_values.reverse()
x_values.reverse()

plt.barh(y_values,x_values,color='r')
plt.style.use('fivethirtyeight')
```

```
plt.title(' Global Power Plant Count')
plt.xlabel('Number of plants')

plt.show()
defget_country_mix(country):

country_data=data['country']
source_data=data['primary_fuel']
cap_data=data['capacity in MW']

source=[]
cap=[]

fori in range(len(country_data)):
ifcountry_data[i]==country:
source.append(source_data[i])
cap.append(cap_data[i])

plot_data={}
t=0

fors in source:
fori in range(len(source)):
ifsource[i]==s:
t+=cap[i]

plot_data[s]=t
t=0
plt.style.use('seaborn')

xval=[]
yval=[]

plot_data=plot_data.items()

foritem inplot_data:
xval.append(item[0])
yval.append(item[1])

plt.barh(xval,yval,color='#ffcc00')
plt.tight_layout()

plt.title('Energy Mix of {}'.format(country))
plt.xlabel('Capacity in MW')

plt.show()

c='India'
print()
get_country_mix(c)
```

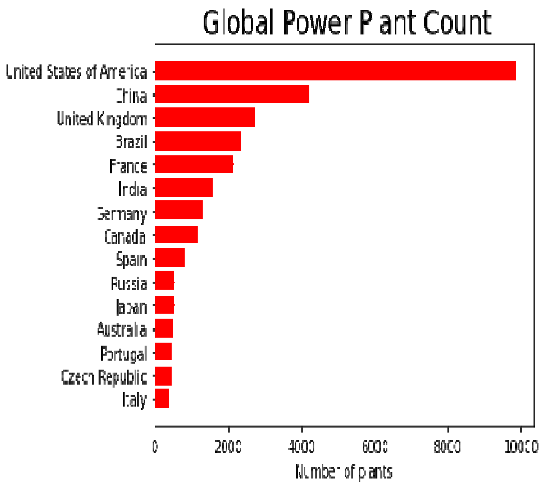


Figure 6.2 Global Power plant estimated count

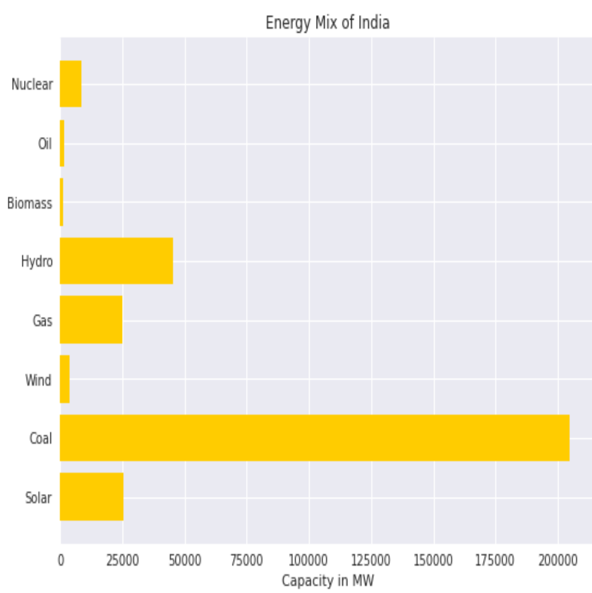


Figure 6.3 Energy mix of india estimated

	Solar accuracy	Biomass accuracy	nuclear accuracy	coal accuracy	gas accuracy	Hydro accuracy
Raw data	81.2	82.4	87.5	83.8	83.5	88.6
Correlation analysis	81.1	82.7	88.6	82.0	83.1	90.2

table 3 finial result

CONCLUSION AND FUTURE SCOPE

Conclusion

Based on geographic, environmental, and system characteristics, this experiment assessed the accuracy of yearly generation estimates for wind, solar, hydro, and natural gas facilities.The approach aims at evaluating each factory's variation from the average population of seeds of its species, using detailed info on plant-level with enviro data. The findings shows that monthly photovoltaic systems production can be predicted relatively well using knowledge on how much wind blasts but how much sun shines first at factory. Due to the sheer low caliber of wind and solar until previously, operational constraints have been limited in actuality. Wind and solar resources are typically sent as they become available. Annual generation from hydropower facilities is less predictable and is highly dependent on water runoff. Natural gas plants were the most hardest to forecast yearly generation for, demonstrating the importance of system characteristics for which we have limited data in deciding when and how they are dispatched. Prospective study should focus on a few countries to improve energy rates. Only when training dataset varies from either the test dataset or when unidentified samples outnumber tagged trials by a considerable margin, the first technique utilizes techniques to boost method robustness (Kostopoulos et al. 2018). The second possibility is the seek for evidence that is closely linked to manufacturing. One potential for power sources is laser metadata that identifies how well a plant is on or off, similar to the convolution layer utilized by the Dioxide Tracking system for coal fired power stations (Gray et al. 2018).

Since input parameters are critical, this might improve prediction performance, but it would be dependent on the presence of those data.

Secondly, we continue to add new learning data and it will become free. While the additional data will have no effect on our estimation, it will help to enhance prediction. They'll also help us construct time series, whether it's by augmenting year productivity projections for oher years or providing higher spectral data. We'll also add new species to the list of genera by which the processes specified in this minor note can be employed as updates are available.

Future Scope

Considering the accuracy, complexity, and computation amount, the combination of the RF and physic model-based approach, hybrid of different time-scale methods, accurate, and fast simulations are the future trends for the other networks .

Compared with the physic model-based and reliability-based techniques, the Linear Regression have superior advantages in the trade-off between the safety, reliability, and economics of the RF. With the advancement of the information technologies and ML algorithms, together with the hybrid of the various approaches in different time scales, the LR is to be a promising technique for the advanced modeling in the future.

ACKNOWLEDGEMENT

First of all, I would like to express my gratitude to MISS BALJIT KAUR, Assistant Professor, Electrical Engineering Department IET BHADDAL TECHNICAL CAMPUS ROPAR for her patient guidance and support throughout this thesis work. I am truly very fortunate to have the opportunity to work with her.

I am also thankful to our Head of the Department, as well as entire faculty and staff of Electrical Engineering Department for their encouragement and cooperation. Without the wise counsel and able guidance, it would have been impossible to complete the thesis in this manner.

My greatest thanks are to all who wished me success especially my parents. Above all I render my gratitude to the Almighty who bestowed self-confidence, ability and strength in me to complete this work for not letting me down at the time of crisis and showing me the silver lining in the dark clouds.

RAYEES AHMAD

REFERENCES

1. Abu-Shikhah, N.; Elkarmi, F. Medium-term electric load forecasting using singular value decomposition. Energy 2011, 36, 4259–4271
2. Ahmed, W.; Ansari, H.; Khan, B.; Ullah, Z.; Ali, S.M.; Mehmood, C.A.A. Machine learning based energy management model for smart grid and renewable energy districts. IEEE Access 2020, 8, 185059–185078.
3. Amjady, N.; Keynia, F. A new neural network approach to short term load forecasting of electrical power systems. Energies 2011, 4, 488–503.
4. Arora, S.; Taylor, J.W. Short-term forecasting of anomalous load using rule-based triple seasonal

methods. IEEE Trans. Power Syst. 2013, 28, 3235–3242.

5. Bianchi, F.M.; Santis, E.D.; Rizzi, A.; Sadeghian, A. Short-term electric load forecasting using echo state networks and PCA decomposition. IEEE Access 2015, 3, 1931–1943.
6. Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. Renew. Sustain. Energy Rev. 2017, 74, 902–924.
7. Dodamani, S.; Shetty, V.; Magadum, R. Short term load forecast based on time series analysis: A case study. In Proceedings of the IEEE International Conference on Technological Advancements in Power and Energy (TAP Energy), Kollam, India, 24–26 June 2015; pp. 299–303.
8. Dudek, G. Short-term load forecasting using neural networks with pattern similarity-based error weights. Energies 2021, 14, 2334
9. Edigera, V.Ş.; Akarb, S. ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy 2007, 35, 1701–1708
10. Fallah, S.N.; Deo, R.C.; Shojafar, M.; Conti, M.; Shamshirband, S. Computational intelligence approaches for energy load forecasting in smart energy management grids: State of the art, future challenges, and research directions. Energies 2018, 11, 596.
11. Fan, G.-F.; Guo, Y.-H.; Zheng, J.-M.; Hong, W.-C. Application of the weighted K-nearest neighbor algorithm for short-term load forecasting. Energies 2019, 12, 916.
12. Hussain, I.; Ali, S.M.; Khan, B.; Ullah, Z.; Mehmood, C.A.; Jawad, M.; Farid, U.; Haider, A. Stochastic wind energy management model within smart grid framework: A joint bi-directional Service Level Agreement (SLA) between smart grid and wind energy district prosumers. Renew. Energy 2019, 134, 1017–1033.
13. Jahan, I.S.; Snasel, V.; Misak, S. Intelligent systems for power load forecasting: A study review. Energies 2020, 13, 6105.
14. Jallal, M.A.; González-Vidal, A.; Skarmeta, A.F.; Chabaa, A.; Zeroual, A. A hybrid neuro-fuzzy inference system-based algorithm for time series forecasting applied to energy consumption prediction. Applied Energy 2020, 268, 114977.
15. Jawad, M.; Ali, S.M.; Khan, B.; Mehmood, C.A.; Farid, U.; Ullah, Z.; Usman, S.; Fayyaz, A.; Jadoon, J.; Tareen, N.; et al. Genetic algorithm-based non-linear auto-regressive with exogenous inputs neural network short-term and medium-term uncertainty modelling and prediction for electrical load and wind speed. J. Eng. 2018, 2018, 721–729.
16. Jawad, M.; Qureshi, M.B.; Khan, M.U.S.; Ali, S.M.; Mehmood, A.; Khan, B.; Wang, X.; Khan, S.U. A robust optimization technique for energy cost minimization of cloud data centers. IEEE Trans. Cloud Comput. 2021, 9, 447–460.
17. Jawad, M.; Rafique, A.; Khosa, I.; Ghous, I.; Akhtar, J.; Ali, S.M. Improving disturbance storm time index prediction using linear and nonlinear parametric models: A comprehensive analysis. IEEE Trans. Plasma Sci. 2019, 47, 1429–1444
18. Jun-long, F.; Yu, X.; Yu, F.; Yang, X.; Guo-liang, L. Rural power system load forecast based on principal component analysis. J. Northeast. 2015, 22, 67–72.
19. Khan, K.S.; Ali, S.M.; Ullah, Z.; Sami, I.; Khan, B.; Mehmood, C.A. Statistical energy information and analysis of Pakistan economic corridor based on

strengths, availabilities, and future roadmap. IEEE Access 2020, 8, 169701–169739.

20. Kiprijanovska, I.; Stankoski, S.; Ilievski, I.; Jovanovski, S.; Gams, M.; Gjoreski, H. HousEEC: Day-ahead household electrical energy consumption forecasting using deep learning. Energies 2020, 13, 2672.
21. Li, S.; Wang, P.; Goel, L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. IEEE Trans. Power Syst. 2016, 31, 1788–1798.
22. Li, W.; Shi, Q.; Sibtain, M.; Li, D.; Mbanze, D.E. A hybrid forecasting model for short-term power load based on sample entropy, two-phase decomposition and whale algorithm optimized support vector regression. IEEE Access 2020, 8, 166907–166921.
23. López, M.; Sans, C.; Valero, S.; Senabre, C. Empirical comparison of neural network and auto-regressive models in short-term load forecasting. Energies 2018, 11, 2080.
24. Madrid, E.A.; Antonio, N. Short-term electricity load forecasting with machine learning. Information 2021, 12, 50.
25. Mamun, A.A.; Sohel, M.; Mohammad, N.; Sunny, M.S.H.; Dipta, D.R.; Hos, E. A comprehensive review of the load forecasting techniques using single and hybrid predictive models. IEEE Access 2020, 8, 34911–134939
26. Musbah, H.; El-Hawary, M. SARIMA model forecasting of short-term electrical load data augmented by fast fourier transform seasonality detection. In Proceedings of the IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 5–8 May 2019; pp. 1–4.
27. Oprea, S.-V.; Bâra, A. Machine learning algorithms for short-term load forecast in residential buildings using smart meters, sensors and big data solutions. IEEE Access 2019, 7, 177874–177889.
28. Qiu, X.; Ren, Y.; Suganthan, P.N.; Amaratunga, G.A.J. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. Appl. Soft Comput. 2017, 54, 246–255.
29. Román-Portabales, A.; López-Nores, M.; Pazos-Arias, J.J. Systematic review of electricity demand forecast using ANN-based machine learning algorithms. Sensors 2021, 21, 4544.
30. Shabbir, N.; Kütt, L.; Jawad, M.; Amadiahanger, R.; Iqbal, M.N.; Rosin, A. Wind energy forecasting using recurrent neural networks. In Proceedings of the Big Data, Knowledge and Control Systems Engineering (BdKCSE), Sofia, Bulgaria, 21–22 November 2019; pp. 1–5
31. Shabbir, N.; Kutt, L.; Jawad, M.; Iqbal, M.N.; Ghahfarokhi, P.S. Forecasting of energy consumption and production using recurrent neural networks. Adv. Electr. Electron. Eng. 2020, 18, 190–197
32. Shah, I.; Iftikhar, H.; Ali, S.; Wang, D. Short-term electricity demand forecasting using components estimation technique. Energies 2019, 12, 2532.
33. Shi, H.; Xu, M.; Li, R. Deep learning for household load forecasting—A novel pooling deep RNN. IEEE Trans. Smart Grid 2018, 9, 5271–5280.
34. Shirzadi, N.; Nizami, A.; Khazen, M.; Nik-Bakht, M. Medium-term regional electricity load forecasting through machine learning and deep learning. Designs 2021, 5, 27.

35. Sun, W.; Zhang, C. A Hybrid BA-ELM model based on factor analysis and similar-day approach for short-term load forecasting. Energies 2018, 11, 1282.
36. Tayaba, U.B.; Zia, A.; Yanga, F.; Lu, J.; Kashif, M. Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform. Energy 2020, 203, 117857.
37. Turhan, C.; Simani, S.; Zajic, I.; Akkurt, G.G. Performance analysis of data-driven and model-based control strategies applied to a thermal unit model. Energies 2017, 10, 67
38. Velasco, L.C.P.; Estoperez, N.R.; Jayson, R.J.R.; Sabijon, C.J.T.; Sayles, V.C. Day-ahead base, intermediate, and peak load forecasting using k-means and artificial neural networks. Int. J. Adv. Comput. Sci. Appl. 2018, 9, 62–67.
39. Wood, A.J.; Wollenberg, B.F.; Sheblé, G.B. Power Generation, Operation, and Control; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; Chapter 12; pp. 566–569.
40. Yildiz, B.; Bilbao, J.; Sproul, A. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew. Sustain. Energy Rev. 2017, 73, 1104–1

