# Introduction

The goal of this chapter is to provide a thorough examination of the efforts undertaken in the field of energy prediction. This chapter will cover a variety of subjects all connected to energy forecasting models and the results that determine which method is the best. In the past, best-fit regression equations were utilised for prediction and analysis [14]. This can be thought of as a "go-to" strategy for constructing equations from historical data and predicting future key variables. To better forecast the future, modern machine learning techniques are replacing or collaborating with traditional methodologies such as regression. This chapter will discuss previous work on regression, and machine learning techniques such as decision trees, Facebook Prophet, and RNNs specifically for energy prediction.

# Regression Models

[10, 15] described how regression can be used to predict energy use. According to [10], the amount of energy consumed by buildings supports the need for thorough research and modelling in this field. According to the article, quantitative methods are a particularly convenient option for constructing energy models when its user only has access to historical data but not the multiple values necessary for technical equations. When compared to other quantitative methodology, linear regression was determined to be a comparatively simple and accurate application. Hourly and daily records from a residence dedicated to research were used in this study. Because of the time difference, the researchers were able to investigate the impact of data collection frequency on the model's precision. External temperature and sun radiation were significant variables. The dependent variable was energy consumption. Simple and multiple linear regression, as well as quadratic linear regression, were all examined. It was carried out to check if the quadratic regression's extra depth was supported by a higher quality of performance.

The time interval was discovered to be a critical aspect in determining the model's quality. The model's performance improves with the length of time. To clarify this, the researchers state that when data is obtained over a shorter period, energy usage anomalies reveal large disparities. When a larger amount of data is acquired over a longer period, the errors average out over time. Using a multiple linear regression model, the coefficient of determination was enhanced. However, the RMSE suffered because of this method. When both criteria were considered, it was discovered that multiple linear regression provided the best overall quality of energy prediction. Daily time intervals also yielded the most accurate model parameters.

A multiple regression model for predicting heating energy demand was presented in [15]. Heating energy demand, according to the authors, is a crucial estimate. It is used during the design phase of a building to anticipate how much energy will be required for space conditioning over its lifetime. The global energy loss coefficient of a building, south equivalent surface, and temperature difference were used as independent variables. [15] is a simpler model than [10] because only three variables and a single prediction model were utilised. [15] also adds that in the event of big datasets, such as the one used in this work, a regression can be employed with reasonable success, even though the fact that it is simpler and easier to construct. The model in this article was developed using the "black box" idea. When input and output variables are known, and the user is required to fit in the best curve possible (also known as "black-box" owing to the unknown nature) to build a generalised relationship between dependent and independent factors, this approach is applied. Though least squares estimation is the most popular method, it can occasionally result in errors that are not normally distributed, which is a strong validation of any curve fitting model [15]. As a result, we employed an incrementally reweighted least-squares method. To lessen the impact of a residual anomaly, this approach modifies the weight of the coefficients in the regression model. Ultimately, this method of determining the best fit curve yields a better least square estimate.

The model was evaluated on 17 blocks of apartments once it had been trained. With an  of 0.9744, the model was determined to be quite efficient. Also, it was discovered that 90% of the calculated results exhibited relative errors of 20% or less. [15] also analyses the model against several dynamic solutions, claiming that such a proposed approach is quicker and produces similar outcomes. Regression is the optimum strategy in terms of model quality and efficiency when constructing models for datasets with a lesser set of variables, according to [15].

# Machine Learning Introduction

Machine learning is described as the application of computational methods that have been trained on historical data to assist in the decision-making process for a specific system. These tools are typically used to increase performance or create precise forecasts. Machine learning approaches are currently being utilised in a variety of industries to make predictions using historical data, either as a substitute or in conjunction with traditional regression/statistical models [16]. [16] is a good place to start reading about machine learning's real-world applications. Several machine-learning techniques are now being utilised to estimate energy usage, including but not restricted to RF [17], SVMs [18, 19], and NNs [20, 21, 22]. Several studies have found that ML algorithms are at least as effective as traditional methods. This chapter delves into such publications in-depth.

# Long Short-Term Memory (LSTM)

Given the availability of studies that use alternative networks, the far more effective methods for forecasting electricity usage are LSTM networks. As a result, the research presented in [27] investigated the usage of multiple LSTM designs for projecting electricity use in the near to medium term. The ideal number of hidden layers with time lags was determined using a GA. The approach's appropriateness was tested using data from France's usage. Bedi et al. [26] provided a methodology for analysing past data's long-term connections as well as short-term trends in segmented data. Later, utilizing electric load data from India, LSTM was implemented with a shifting frame. DRNN, ANN, and SVR were all surpassed by the model constructed. [28] presents the analysis of power prediction using temperature as an independent factor. They combined power demand data with some other factors such as temperature, air pressure, and humidity to predict power demand usage data.  For the short to medium range  (24 and 48 hours, 7 and 30 days), the prediction is used. Performance measurements like RSME and MAPE were used to compare the proposed model to other conventional methods. The analysis indicated that LSTM surpasses competing algorithms when it comes to boosting predictive performance. Using a MATLAB toolkit, the LSTM system was autonomously tuned. For many prediction horizons, the results are then compared to that of ARMA, SARIMA, and ARMAX.

External variables were also given to the LSTM by Kwon et al. [29]. These hyperparameters were configured using a heuristic methodology. The electricity system administrator in Korea utilised 2 years to analyse the model, with only an error of 1.5%. The LSTM presented by [30], on the other hand, provided a data dimension reduction to reduce computing costs. To evaluate the approach's effectiveness, the authors devised two groups of trials. The suggested technique outperformed ANN, ARMA, and ARFIMA in comparisons. The suggested LSTM technique was shown to have an RMSE of 19.7% lesser than that of the benchmark feed-forward neural network. Subsequently, [31] suggested the CVOA, which has been utilised to improve the hyper-parameters of an LSTM network. The findings provided surpassed a large variety of deep learning techniques combined with well-known optimization strategies. As a baseline, data on Spanish power use has been employed. [32] proposed a multi-layer bi-directional RNN centred on LSTM and GRU to estimate electricity usage. The researchers beat the findings of ANN and SVR when they looked at peak usage and seasonality individually. Pegalajar et al. introduced in recent years three variants of RNN to anticipate Spanish energy demands, comparing the results against a wide range of ML models and finding that they outperformed every one of them [33].

Utilizing the LSTM algorithm and raw data from a housing complex, Kong et al. [34] established a framework for estimating short-term load. For most datasets, the suggested LSTM architecture had the best prediction performance. Based on the LSTM method, Jiao et al. [35] created a model for forecasting the demand for non-residential buildings. The suggested framework was built using real data from 48 non-residential buildings, which included energy consumption statistics. Kim and Cho [36] used a mixture of LSTM and CNNs to design a model for precisely estimating the power usage of a housing complex for a steady electrical supply. CNNs were designed to obtain the features of variables that affect power use in the research, which were initially designed for image analysis. LSTM may detect details in time-series data that have an abnormal pattern. Through this work, a CNN–LSTM model was created by merging the properties of the two algorithms, and electricity consumption was forecasted almost accurately, which was formerly hard to forecast. Even with minute, hourly, daily, and weekly data, a high level of predictive performance was discovered.

Khafaf et al. [37] suggested an LSTM neural network model for forecasting energy consumers' 3-day energy use clusters. Wang et al. [38] used the LSTM neural network to estimate power usage and identify grid irregularities. Regarding energy consumption forecasting in residential and commercial buildings, Khan et al. [39] used a hybrid CNN with an LSTM autoencoder. Regarding everyday natural gas consumption forecast, Wei et al. [40] developed a hybrid single spectrum analysis and an LSTM model. Singaravel et al. [41] examined the effectiveness of LSTM in the planning phase of a building. With four alternative versions of the LSTM model, 201 design scenarios were analysed. LSTM models are found to have superior accuracy and processing time than ANN models. To estimate the power usage of air-conditioning systems, Zhou et al. [42] presented an LSTM model. The number of iterations, the size of the time series feeding the LSTM, and the learning rate were all tested repeatedly to identify the ideal hyper-parameters.

Furthermore, to increase LSTM's predictive performance, hybrid evolutionary optimisation approaches such as GA and PSO are merged with it. To find the optimum weight matrix or component of the LSTM hyper-parameters, the majority of GA and PSO are used. He et al. [43] presented a hybrid short-load forecasting system combining variational mode decomposition and LSTM networks, with the LSTM network's hyper-parameters tuned by using the Bayesian Optimisation algorithm. For home energy usage forecasting, Kim et al. [44] developed an LSTM network. To determine the best hyperparameters, including learning rate, layer size, and dropout rate, the PSO is used. Yang et al. [45] suggested a hybrid prediction model that combined extreme learning, RNNs, and SVMs. To choose the best weight matrix between these networks, PSO is used. Guo et al. [46] presented an LSTM neural network-based short-term prediction method that captures real-time responses into account. To get the appropriate weight matrix for LSTM, the enhanced GA is applied. The LSTM network's architecture, on the other hand, is set based upon that developer's expertise. Regarding hourly natural gas demand forecasting, Su et al. [47] developed a hybrid wavelet transform and an LSTM model. Trial-and-error is used to determine the amount of LSTM layers, whereas GA is used to optimise the number of neurons for each LSTM layer.

# Bidirectional-LSTM (BLSTM)

Deep neural networks based on BLSTM have been extensively used in speech recognition and text classification, but they are seldom used in time series prediction or stock market forecasting. RNNs aid in the detection of relationships in sequential data. Bi-LSTMs are a type of RNN that can model data forward as well as backwards. This incorporates several previous stock prices as well as potential future stock values. M. Jia et al. [25] established a methodology for predicting the future price of a stock using a BLSTM neural network. The writers used the GREE stock's statistical data. From January 2017 to May 2019, they compiled data for 568 days. There were 14 different features in the data, which included open, high, close, volume, and so on. The data was pre-processed and standardised. For the forecast, the close price was used as a baseline. A one-way and two-way LSTM were used to process the pre-processed data. Overfitting can be avoided by using a technique called dropout. The neural network was subjected to the dropout approach. The suggested BLSTM was assessed to both the ARIMA and a standard LSTM model. The RMSE, MAE, and deviation result of the proposed system were determined to assess its accuracy. The suggested approach surpassed both the ARIMA and the LSTM models, as evidenced by the findings. The RMSE and MAE also were lowered by 24.2% and 19.4%, correspondingly.

[23], on the other hand, adopted BLSTM for energy load prediction. The performance of various architectures was examined, including BLSTM, multilayer LSTM, and decoder-encoder architecture. The BLSTM structure was found to have the best results. [24] also used BLSTM for network-wide traffic speed forecast. Both linear SLSTM and BLSTM neural networks were examined in this study. It investigated several LSTM structures and found that two layers of stacked BLSTM surpass other LSTM-based designs. Employing multiple tests to assess BLSTM with SLSTM, we study and assess the integration of a suggested methodology using BLSTM into financial time series forecasting.

# Gradient Boosting Regression Trees

GBRT variants have been pitted against a slew of newer models. Random Forest [56] , XGBoost [56][57], CatBoost [58], AdaBoost [59], SVM [60], MLP [61], and CNN and GRU framework [62] have all outperformed GBRT. [63] advocated evaluating 7 algorithms per day during the forecast, including GBRT and XGBoost, and then using the most effective for the next day's forecast. One suggested incorporation of the model is to use it as a foundation for a stacking model that includes RF, XGBoost, SVR, and kNN to increase performance and generalisation [65]. Investigating feature selection approaches such as Mutual Information, F-regression, Elastic Net, and Recursive Feature Elimination in tandem with the GBRT model [64] is another field of research that is beyond the scope of this paper.

## XGBoost

[65] developed a new method for anticipating electricity demand. They took daily power load data and translated it to weekly power demand data. This expands the number of features that can be used to forecast load for a lag parameter. For selecting features from the transformed data, the XGBoost method was utilised. After that, the model was trained, yielding a minimum MAPE of 10%, 97% accuracy, and an MAE of 88.90%.

When used in the literature, the XGBoost approach has had a lot of success.   XGBoost or XGBoost hybrid systems surpassed other methods [57], [66], [67], [68]. With its capacity to execute parallel processing, XGBoost is incredibly effective [67]. Before training, XGBoost organises the data, stores it in a block structure, and then uses it in later phases. This increases performance by drastically decreasing processing, as grouping the values of the attributes are amongst the most time-consuming phases in decision tree learning. As a result, in the Wang et al. [56] analysis, XGBoost had the optimum performance. Another aspect of XGBoost is its built-in capacity to choose and save important features; as a result, additional feature selection algorithms aren't always necessary by the XGBoost algorithm and can even degrade XGBoost efficiency [70].

XGBoost was used in conjunction with other techniques by several academics. CEEDMAN would be used initially to de-noise the original data by dividing it into 12 samples, according to Lu et al. XGBoost surpassed SVM with PSO, Least-squares SVM, CEEMDAN Random Forest, and Radial basis function NN in the 12 samples, and the forecast results were then averaged and denormalized. [73]. Pairing EMD and ARIMA with XGBoost [74], along with solely using ARIMA [71], and merging XGBoost with k-means on Similar Days selection [69] are two further instances. All the XGBoost-based models in the prior articles surpassed other algorithms in contrast. Different architectures such as the CNN and GRU frameworks [62], as well as an ANN [72], were able to best XGBoost in a few articles.

## LightGBM

A concurrent IDS built on a PID method and LightGBM was proposed by Jin et al. [78]. To accomplish response time outside of reducing attack detection performance, the suggested framework employs two main methods. To begin, the intrusion sensor is a LightGBM. Next, traffic data is properly analysed using PID. Rapid IDS is built on PID methods that have expenses in terms of collaboration and communication. Furthermore, with a connection rate of up to 1.26 Gbit/s, the suggested framework is stable.

When dealing with non-linear traffic data, machine learning is more precise than mathematical statistics. Nonetheless, the lack of understandability of prediction outputs limits ML. A variety of boosting techniques inspired by the GBDT have recently been discovered. XGBoost [75] and LightGBM [76] are two examples. Due to its parallel learning, robustness to anomalous values, and adaptability to missing values, they have been used and coupled with other models in a variety of domains. To predict ultra-short-term wind energy, Ju et al. [77] suggested a hybrid model based on CNN and LightGBM, and the findings confirmed that the inclusion of LightGBM surpassed effectiveness and precision.

# Facebook Prophet

[48] present Prophet as a young technique for predicting that has a lot of promise for use in power demand prediction. In recent years, various applications of this concept have been discovered. In the case of Bitcoin prediction, Yenidogan et al. [49] examined two methods: ARIMA and Prophet. The results suggest the Prophet model had a precision of around 94.5%, making it significantly higher than ARIMA's 68%. Furthermore, Ashwini Chaudhari [50] forecasted the costs of currencies like Bitcoin, Litecoin, and Ethereum using three models: ARIMA, Prophet, and LSTM networks. The findings show that using LSTM and Prophet yielded very accurate results for the 3 cryptocurrencies, ranging from 93% to 99% while using the ARIMA model yielded just 82% to 66% accuracy. Furthermore, Bianchi et al. [51] used raw data from an Italian utility business to compare thermal short-term demand forecasting methodologies utilising the ARM, NARM, and Prophet. In terms of short-term predicting, the ARM outperformed the other models. Das [52] conducted a comprehensive study in which five alternative forecasting models (SES, Dynamic Harmonic Regression, NN,  ARIMA, and Prophet) were utilised for wind speed prediction in two Indian states (Tamil Nadu and Maharashtra). The greatest results came from the neural network. The Prophet framework, on the other hand, produced encouraging results and was suggested for future purposes.

The evaluation of random forests and Facebook's Prophet in predicting daily flow up to 7 days in advance in a river in the United States is examined by Papacharalampous & Tyralis [53]. These prediction systems employ historical streamflow data, with random forests also using historical rainfall data. They employ a naive approach based on previous streamflow observations and an MLR model that incorporates identical data as random forests. The findings demonstrate that random forests surpass the naive technique overall, whereas Prophet surpasses it over prediction timeframes greater than 3 days.

Earlier studies [54] employed it to forecast sales, and the findings were presented using the MAPE level for sales forecasting of various categories of items. On a quarterly forecast, they were able to attain a MAPE of little less than 30% for 70% of the goods. For this study, standard seasonality trends were applied. The WMS manages a complicated series of processes known as warehousing. The idea of smart WMS is outlined in [54]. The implementation of Facebook's Prophet algorithm for sales forecasting was outlined by Zunic et al. [55] as part of the smart WMS notion and improvement of supply companies. In several of Bosnia and Herzegovina's major facilities, the notion of smart WMS and sales forecast has been proven in actual scenarios and with real data. According to the preceding published papers, no study has yet been done on the Prophet model's accuracy in long-term energy load prediction.