THE HONG KONG INSTITUTION OF ENGINEERS 香港工程師學會

A.I. model forecast of building cooling load demand for the reduction of energy consumption to work towards carbon neutrality

Tony Ip[1], Tattwa Darshi Panda[1], Xiaoyu Jia[2], Yiqun Pan[2], Dhanada K Mishra[1], Matthew Yuen[1] and Harris Sun[1]
[1] RaSpect Intelligent Inspection Limited, Hong Kong, People's Republic of China
[2] School of Mechanical Engineering, Tongji University, Shanghai, People's Republic of China

ABSTRACT

The world needs to achieve carbon neutrality or net zero emissions of greenhouse gases (GHG) by 2050. Buildings are major sources of GHG emissions. Applications of the latest innovative technologies of machine learning/A.I. algorithms have opened up new opportunities. The optimal control of cooling plant systems is important to reduce energy consumption and therefore emissions. Knowing the cooling load demand in advance can help facility managers operate cooling plants much more efficiently. This paper presents a real-life application of nine A.I. models for time-series forecasting of the cooling load demand of a commercial building. LSTM neural networks, Facebook Prophet time series model, and DeepAR recurrent neural network models are found to be the most accurate with a Mean Absolute Percentage Error (MAPE) in the range of 15 to 16 with a computing time in the range of 294 to 319 seconds respectively. The LightGBM machine learning model on the other hand proves to be the fastest with a MAPE of 18.96 in just 7 seconds. Thus, different models can be deployed for different requirements. Optimising the operation of cooling systems as per the forecast cooling demand can bring enormous energy savings that are essential for achieving carbon neutrality.

KEYWORDS Buildings energy efficiency; cooling load demand; semantic model; A.I.; forecast; MAPE; prophet
CONTACT Dhanada K Mishra dhanada.mishra@raspect.ai
Received 24 March 2022

1. Introduction

The International Energy Agency (IEA), in a study, has shown that building construction and operations accounted for the largest share of both global final energy use (36%) and energy-related CO_2 emissions (39%) in 2018 (Global Alliance for Buildings and Construction, 2019). Buildings are responsible for close to 40% of the total carbon dioxide emissions in urban areas (Electrical and Mechanical Services Department, 2021). With the rapid global urbanisation, this trend is likely to increase significantly over the coming years. Hong Kong's 42,000 buildings, including 8,000 high-rises, account for more than 90% of the city's electricity consumption. Electricity represents 56% of the total energy end-use as per a report by the Electrical and Mechanical Services Department, EMSD of Hong Kong (2021). As shown in Figure 1, the majority of Hong Kong's electricity comes from fossil fuel-based generation. About 60% of Hong Kong's carbon emissions are accounted for by its buildings. The average annual electricity bill of the heating, ventilation, air-conditioning (HVAC) systems only in buildings is about HKD12.3 billion. Improving building energy efficiency is essential to realise carbon neutrality. The "Energy Saving Plan for Hong Kong's Built Environment 2015~2025+" issued by the government sets the target of 2025 for reducing energy intensity by 40% with 2005 as the base year. According to the latest Hong Kong Energy End-use Data published by the Electrical and Mechanical Services Department (2021), Hong Kong's energy intensity decreased by 32.0% from 2005 to 2019.

There is a worldwide campaign to minimize the energy consumption and carbon footprint of buildings under the following five UN Sustainable Development Goals (SDGs) as shown in Figure 2:

- SDG 7: Affordable and Clean Energy
- SDG 9: Industry, Innovation, and Infrastructure
- SDG 11: Sustainable Cities and Communities
- SDG 12: Responsible Consumption and Production
- SDG 13: Climate Action

A.I.-powered solutions that increase the energy efficiency of buildings will play a key role in achieving these goals.

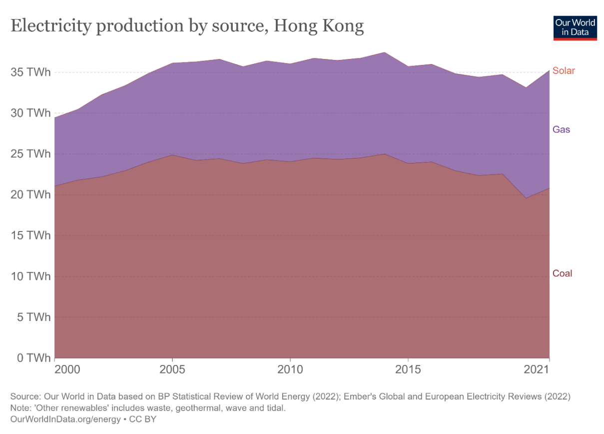


Figure 1. Electricity production by source in Hong Kong: the majority source continues to be fossil fuels.

Figure 2. Sustainable Development Goals (SDG) set by the United Nations (no date)

Recently, tremendous advancements have been reported in the fields of machine learning and artificial intelligence. There have been many successful applications and use cases reported from various sectors of industry, including construction and building operations. There is active research in smart building operations utilising IoT sensors to monitor all aspects including environment (outdoor and indoor), energy consumption and operating parameters of HVAC systems. Use of the data collected in real time allows for optimising energy-intensive processes such as the operation of chiller plant that provides the much-needed cooling of buildings in hot and humid tropical and sub-tropical regions like Hong Kong. In this paper some of the most recent machine learning models are explored and utilised for the forecasting of the cooling load demand of a typical commercial building in Hong Kong.

2. Literature Review

Intelligence is the ability to learn from the surroundings to modify one's behaviour in the face of situations that have not been encountered before. Using mathematical algorithms, computing machines can be programmed to learn certain patterns from data and create complex rules that can be applied to predict future occurrences of a given process or phenomenon. Rapid progress in computing power and the availability of real-time data have led to ever more sophisticated machine learning algorithms to be built and deployed in artificial intelligence applications. Applications and use cases can be seen in all sectors including e-commerce, finance, retail, healthcare, education, etc (Ertel, 2017).

Another major area of application is in smart city management (Kirwan and Fu, 2020). Besides city-level management of energy infrastructure, A.I. can help to make buildings energy efficient. This can be achieved by energy conservation, on-site energy generation, fault detection and predictive maintenance, better building operations and demand-load prediction and balancing (Grueneich, 2015). The role of A.I. in controlling ventilation and integrating air pre-heaters, resulting in about 26% energy savings in commercial buildings in the US has been discussed in a case study by Al Raees et al. (2014). A detailed review of the literature in the field of application of A.I. in energy conservation in relation to buildings is provided by Farzaneh (Farzaneh et al., 2021).

Smart buildings utilise a combination of the physical and digital world to take advantage of data-driven operation of a building's various systems such as HVAC, transport (elevator/escalator), security, water supply and drainage, lighting, etc. With the objective of increasing the safety and comfort of occupants, smart building management systems (BMS) help increase energy efficiency which leads to a lower emission of greenhouse gases (GHG) and cost of operation (Farzaneh et al., 2021).

A.I. technology has been widely used for cooling load prediction in buildings with the development of smart building management systems (Panchalingam and Chan, 2019). Short-term load prediction (1 hour to 1 week) plays a crucial role in optimising the operation of energy systems and has received extensive attention from scholars in recent years. Time series models such as the Summed Auto-Regressive Integrated Moving Average (SARIMA) model (Kawashima et al., 1995) and its derivative ARX (Sarwar et al., 2017) are widely used for short-term forecasting of building energy consumption. Shallow learning models such as Support Vector Machines (SVMs) (Tao et al., 2019) and Extreme Gradient Boosting (XGBoost) (Fan et al., 2014) have proven to be more computationally efficient. With the rapid development of deep learning and deep neural networks have been introduced as a tool for short-term load forecasting. Deep learning algorithms based on Long Short-Term Memory (LSTM) (Wang et al., 2019) have received extensive attention and research due to their high accuracy.

A semantic A.I. platform has been designed and developed by the Electrical and Mechanical Services Department (EMSD) of the Hong Kong Government to perform big data processing and analytics (Pang, 2014). It is a novel approach of using knowledge graphic ontology for describing the components of buildings and the relationships between them. It represents buildings as directed labelled graphs using the Resource Description Framework (RDF) data model and describes the relationship between spaces, equipment, sensors, data, etc. It helps make buildings and their systems machine-readable. Semantic A.I. has been deployed to perform E&M facility performance analysis and prediction modelling of buildings. Such A.I. models that are developed for a particular building can be used in other buildings and deployed at a regional or city scale to drive energy efficiency.

3. Problem Statement and Research Significance

Chilling plant operation is key to the efficiency of the cooling system in a building and hence affects the energy efficiency significantly. Large commercial buildings with multiple chiller units need to decide which chillers should be on or off and their operation parameters. Facility management needs to make assumptions such as the upcoming cooling load demand and external environmental factors. These assumptions may or may not reflect real conditions. They may waste electricity when the cooling load is assumed to be too high. If assumed to be too low, it may result in bad occupants' ratings. The problem discussed in this paper involves the building of an A.I. model based on 15 months of actual operational data from a commercial building located in Hong Kong to predict its future cooling load demand. The 15-month data was supplemented by a further three months of subsequent data obtained with the help of the semantic model of the building's HVAC system (Figure 3). These data were used for evaluation of the model. Finally, the model accuracy was tested against a seven-day period for which input data were given. The data available required pre-processing and cleansing to make them suitable for visualisation and model training. Several models from the literature were chosen and the evaluation results were compared.

Figure 3.Screenshot of the Semantic model of the Building HVAC system developed by the EMSD

4. Exploratory Data Processing

The training data set contains cooling load data related to a commercial building complex located in Hong Kong, for a period of 18 months from 1 April, 2020, to 30 September, 2021. While the first 15 months of data were available in the form of a csv file, the remaining three months of data were calculated from the raw data downloaded from the semantic model online. The variables in this CSV file are Timestamp, Temperature, Humidity, UV Index, Rainfall, Cooling Load for South Tower, Cooling Load for North Tower, and the Combined Cooling Load of these two towers (i.e., summation of values in the Cooling Load for North Tower and South Tower). The remaining three-month cooling loads of the South Tower and North Tower were calculated from their respective chilling plant flow rate and differential temperature of inlet and outlet chilled water taken from the appropriate sensors, and the formula is shown as Equation (1). The combined cooling load is the sum of the cooling load for South Tower and North Tower.

$$CL = m\,C\,\Delta T, \tag{1}$$

where,

- CL is Cooling Load
- m is the flow rate (L/s) x density of water (1kg/L)
- C is the heat capacity (4.19kJ/kg °C)
- ΔT is the difference between the inlet (return) and outlet (supply) chilled water temperature (°C)

A single seven-day testing data item (in CSV format) from 24 September, 2021, to 30 September, 2021, is provided to test the model's prediction.

The original 15-month training data, three-month evaluation data and seven-day test data that are in 15 min intervals, are converted into hourly intervals by averaging. Figure 4 shows the entire data set with average hourly and daily cooling loads. Weekly variations in terms of the difference between weekdays and weekends are also clear. Figure 5 shows the daily variation of the cooling load. The working hours for the target building run from 09:00 to 17:00, and the cooling load begins to increase two hours before the working time. Likewise, most equipment is turned off after 17:00, causing cooling loads to decline. The cooling loads on working days are not much different and significantly larger than those on weekends. This might be caused by working schedules. Figure 6 shows the seasonal variation indicating the significantly higher cooling load in summer compared to winter. The trend of cooling load is consistent with the trend of outdoor air temperature. Figure 7 depicts the monthly daily trend of the cooling load, and the green, red, yellow, and blue curves represent spring, summer, autumn, and winter, respectively. The daily trend is similar for each month of the year. The cooling load in summer is the largest. June, July, August, and September are the months with the highest load in the year, while the cooling load in winter, especially January, has the lowest cooling load. The daily load remains the same at night (19:00 - 06:00). The load from May to September has an inflexion point at 9:00, which might be caused by the fact that the chiller is not turned on at night. The load for October-April has an inflexion point at 8:00, and the cooling load gradually increases with the increase of outdoor temperature. The cooling load declines after 17:00 due to factors such as staff starting to leave.

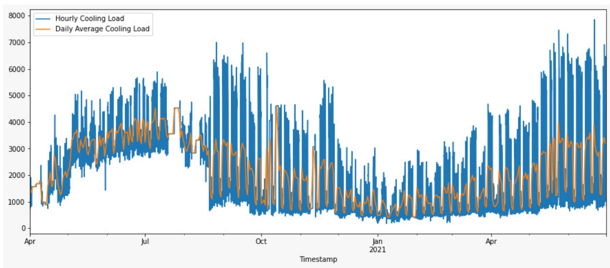


Figure 4. Raw data after cleansing for 15 months as hourly data and daily average.

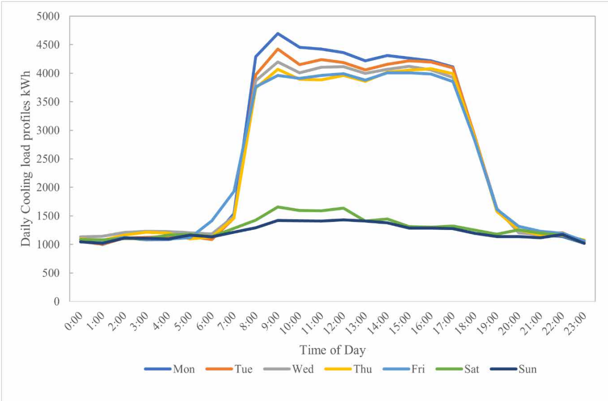


Figure 5. Daily variation in cooling load over an 18-month period.

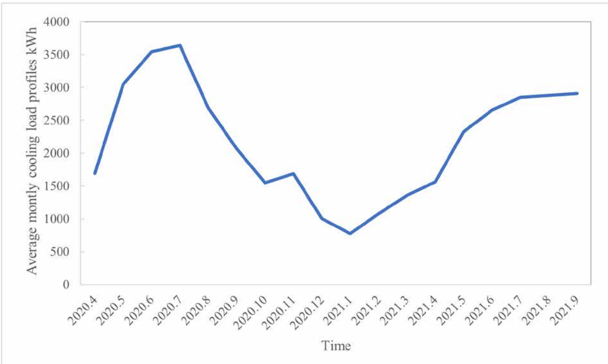


Figure 6. Seasonal variation of average daily cooling load.

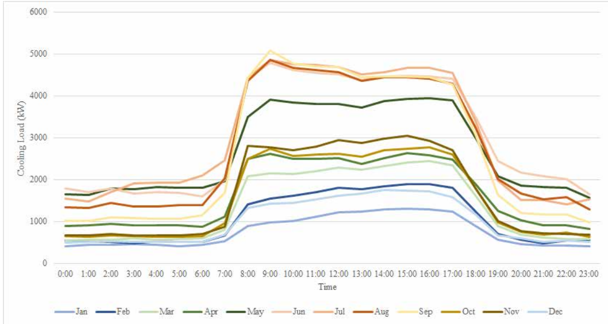


Figure 7. Average daily cooling load for each month.

5. A.I. Model Training and Predictions

Three types of algorithms are used to predict cooling demand as shown below in Figure 8.

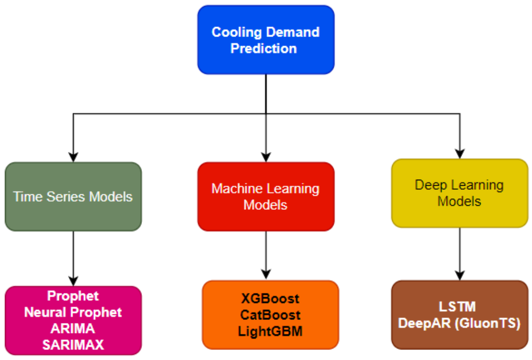


Figure 8. Three types of models considered for the cooling demand forecasting problem.

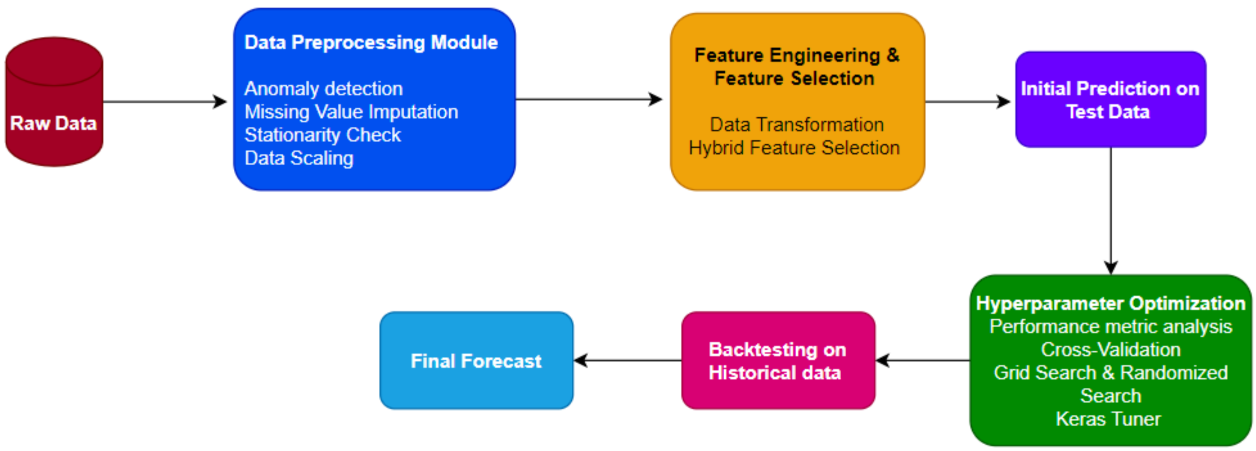


Figure 9. End-to-end prediction pipeline.

Time series models involve considering date as the index parameter. An initial set of data is used to train the model and then predictions are generated based on the remaining data to evaluate the model's performance. The time series models considered in this paper are ARIMA, SARIMAX, Prophet, and Neural Prophet.

Another way of solving a time series problem is to convert it into a machine learning problem and solve it using the standard machine learning procedure (Mitchell, 2007). Here date is not directly considered as a parameter. The date variable is further exploited to obtain the day, month and hour parameters. In the current work, three machine learning models are considered: XGBoost, LightGBM, CatBoost. These three models take the least training and inference time as compared to all other A.I. models presented in the paper.

Deep learning models are sequence-to-sequence models (Yu et al., 2019). A sequence of data (time periods) is used to train a deep learning model and then the prediction for the next single point in time is generated. Deep learning models use neural networks to recognise patterns in the time series data. Two advanced deep learning models (Multivariate LSTM and DeepAR) are presented in the paper.

For each of the algorithms, an end-to-end pipeline is developed as shown below in Figure 9.

The raw data are passed through a sequence of steps before obtaining the final forecast. In the first step, extensive data pre-processing is required. Few data pre-processing steps are mentioned in the previous section. In this section two important steps for time series are presented: stationarity check (Priestley and Rao, 1969) and data scaling (Rosenfeld, 2021).

Performing stationarity checks ensures that the statistical properties of the time series do not change drastically over a period. The Augmented Dickey-Fuller test (ADF Test) (Priestley and Rao, 1969) and Kwiatkowski-Phillips-Schmidt-Shin test (KPSS test) (Priestley and Rao, 1969) are used for checking the stationarity of the cooling demand time series data.

In the second step, extensive feature engineering (Mitchell, 2007) is performed. For machine learning algorithms, different parameters like months, days, holidays, and weekends are obtained from the date variable. Feature engineering also ensures that the distribution of the data is not changed. For the current problem, holiday and weekend parameters have a significant impact on the final forecast.

Feature selection (Li et al., 2017) is also performed in the second step. In feature selection, the most important features affecting the final forecast are selected. These features are given as input to machine learning and deep learning models.

To ensure the consistency in performance of each model time series cross-validation is performed. This is called forward chain cross-validation which is very different from cross-validation used in machine learning. Time series cross-validation has been illustrated in the Prophet model section.

5.1. ARIMA & SARIMAX

The first model deployed is one of the most 'seasoned' techniques for time series forecasting – the ARIMA (Ho and Xie, 1998) – which is the acronym of Auto Regressive Integrated Moving Average. ARIMA is a combination of: Autoregressive (AR), Integrated (I), and Moving Average (MA).

Autoregressive property means that the ARIMA model uses the dependent relationship between an observation and a few lagged observations. It predicts future values based on previous values. For the integrated part, the difference between an observation and some previous observation is used. Moving average captures the change in average of all data points over a particular time range. For ARIMA, a user needs to explicitly mention the values for AR, I and MA.

The input to the ARIMA mathematical model is given in the form of p (lag order), d (degree of differencing) and q (length of moving average window). Each of p, d and q are non-negative integers. These three parameters are necessary for ARIMA to work.

Often some kind of statistical methodologies are used to generate the values for p, d and q which is a time-consuming and error-prone task. Hence, the Auto ARIMA package is used to perform ARIMA and Seasonal Auto Regressive Integrated Moving Average with eXogenus factors (SARIMAX) on cooling load demand data. In Auto ARIMA, there is an inbuilt internal treatment for the model to generate optimal p, d, and q values which would be helpful for achieving better forecasts.

For the current data, there are two external parameters: temperature and humidity. These parameters have seasonal variations and are data with repeated cycles. Even though ARIMA has shown evidence of working well, it has the major pitfall of not being able to handle seasonality (Ho and Xie, 1998). The ARIMA model is not most well-suited for the problem because of the strong domination by weekly, monthly and seasonal patterns. Training and inference using ARIMA exhibits a Mean Absolute Percentage Error or MAPE of 53.

The second model trialled was the SARIMAX model (Elamin and Fukushige, 2018), which stands for Seasonal Auto Regressive Integrated Moving Average with eXogenus factors. Accordingly, SARIMAX represents an 'upgrade' which is performed for adding seasonality and exogenous factors such as outdoor temperature and humidity. The predictions using the SARIMAX model that adopts automatic optimisation of model parameters resulted in a mean absolute percentage error (MAPE) of around 43.

The downside with ARIMA and SARIMAX is their complexity. They are computationally expensive and not scalable processes.

The accuracy of the Hourly Cooling Load Prediction model is measured by the Root Mean Square Error (RMSE) as given by Equation (2):

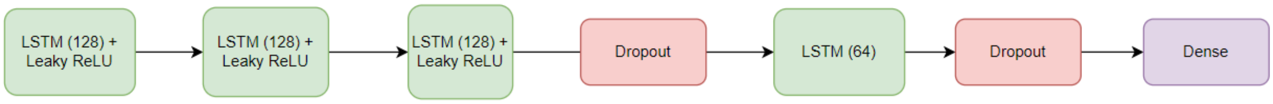


Figure 10. LSTM Architecture.

$$\epsilon = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - a_i)^2}\,, \tag{2}$$

where

- $\mathcal{E}$ is the RMSE value (score),
- n is the total number of forecasts, which is the next 3 months of data in the dataset
- p_i is the prediction value of the cooling load demand in kW, and
- a_i is the actual cooling load in kW

Another measure of model prediction accuracy is known as Mean Absolute Percentage Error (MAPE) given by

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| * 100\,, \tag{3}$$

where A_t is the actual value, F_t is the forecast value, n is the sample size and t is the time step.

5.2. Long Short-Term Memory (LSTM) Networks

The current time series problem is solved using a multivariate time series model based on Long Short-Term Memory (LSTM) Networks (Yu et al., 2019). LSTM is an advanced form of recurrent neural networks (RNNs). A RNN is a type of deep learning neural architecture used specifically for sequence modelling. For LSTM, a time series generator is required for data preparation for input into the multi-layer LSTM model.

For the current problem, one of the advantages of using LSTM is its ability to capture the non-linear relationships in a dataset.

To account for the outliers in the dataset, a minmax scaler (Bisong, 2019) is used. The minmax scaler scales the values of all the data observations between 0 & 1. Scaling is an important requirement for gradients to converge faster. The data are scaled for the training data and then the scaled values are used for the test data by not varying the data distribution.

An optimal neural network architecture of LSTM layers is developed for the forecasting as given below in Figure 10. It involves a sequence of LSTM layers followed by two dropout layers and one dense layer. Leaky ReLU (Yu et al., 2019) is the activation function used for all the layers. Leaky ReLU is a mathematical function that speeds up the overall training process of a neural network.

For hyperparameter tuning, Keras Tuner is used. Keras Tuner is an automatic hyperparameter tuning framework to find the best values of hyperparameters that give the most accurate prediction. A grid search method is used to list all the hyperparameters and combined with Keras Tuner to find the optimal values for each parameter.

The hyperparameters for the LSTM network are the number of epochs, batch size, dropout values and number of neurons. Four epochs were enough to obtain the desired accuracy. A batch size of 32 was found to be more computationally efficient in terms of training time and accuracy. For the first three LSTM layers, using 128 neurons for each neuron was found to be very effective. The prediction using multivariate LSTM is shown in Figure 11. The MAPE obtained is 15. LSTM is found to be 71.6% more accurate than ARIMA and 65.1% more accurate than SARIMAX. The model training and prediction take 294 seconds.

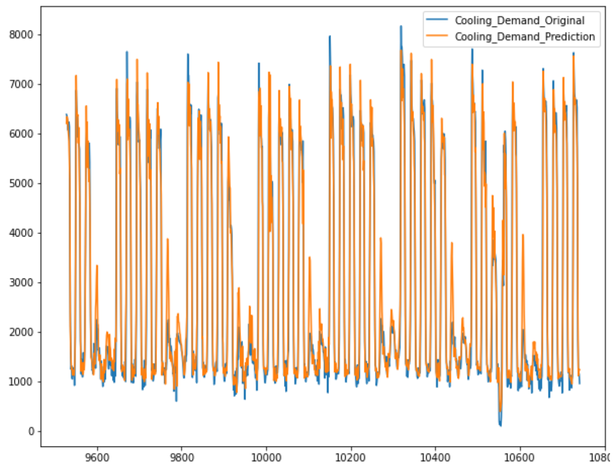


Figure 11. Original Data vs. LSTM prediction.

5.3. Boosting Algorithms

Boosting algorithms are found to be very effective for the current time series problem. Three algorithms, XGBoost, CatBoost & LightGBM (Daoud, 2019) are applied to the current problem. These algorithms are found to be fastest in training and prediction among all the algorithms used in this paper. The mathematical formulations behind these algorithms are studied in detail in (Daoud, 2019).

All the three boosting algorithms can run on multiple languages. They are portable on different platforms and integrable with various cloud computing systems. These algorithms are implemented on top of gradient boosting which is a simple boosting algorithm in ensemble learning. The core functionality of these algorithms uses the parallelisation concept. These algorithms use all the cores of the CPU available. Cache optimisation is an important characteristic of these frameworks. In cache optimisation all the intermediate calculations are kept in the cache for further optimisation. In these three boosting algorithms, there is a provision of optimising the memory in such a way that it can work on data having a size larger than the size of the RAM.

All the three algorithms are built on top of a boosting algorithm. The decision tree creation process is different in these three algorithms. In CatBoost, decision trees are symmetric trees. In LightGBM, the decision tree grows leafwise and in XGBoost the decision tree grows normal depth wise.

Machine learning algorithms offer an additional advantage of computing the feature importance. This metric is useful in indicating which features contribute most to the prediction. The feature importance obtained using XGBoost is shown below in Figure 12. As shown below, temperature is the most important feature in obtaining accurate prediction.

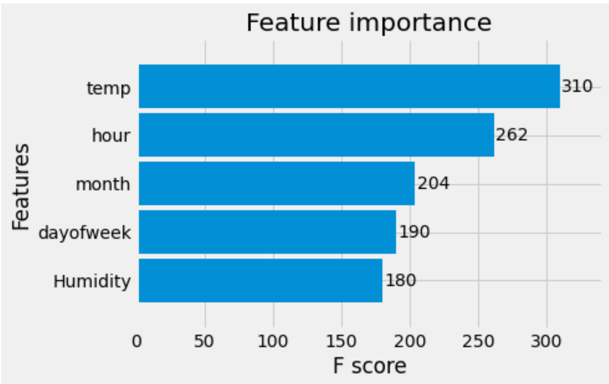


Figure 12. Feature importance for cooling load demand prediction.

The original data is divided into training (80%), validation (20%) and testing (20%) data sets. Three models are trained on a training data set and initial evaluation is performed on the validation data set. The hyperparameter tuning is carried out based on the performance of each model on the validation data set. The final prediction is done on the prediction data set and shown in Figures 13, 14 and 15.

The most important hyperparameters considered are the number of estimators (decision trees) and the learning rate for each model. A grid search algorithm is used to find the most optimal values for each of the hyperparameters.

The MAPE obtained is 21.23 for XGBoost, 18.96 for LightGBM and 19.52 for CatBoost.

The time required for training and prediction is 10 seconds which is 284% faster than for LSTM.

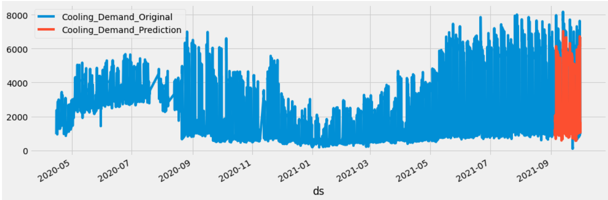


Figure 13. Cooling Prediction using XGBoost.

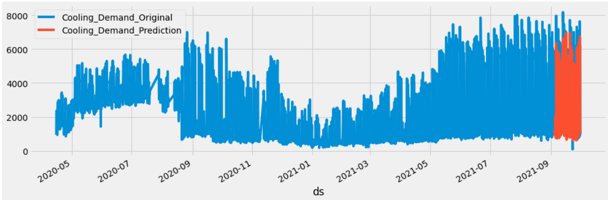


Figure 14. Cooling Prediction using LightGBM.

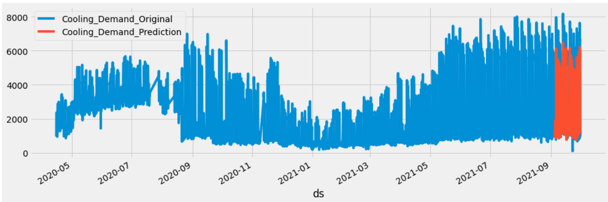


Figure 15. Cooling Prediction using CatBoost.

5.4. DeepAR Algorithm

DeepAR (Salinas et al., 2020) is a supervised learning algorithm for time series forecasting that uses recurrent neural networks to produce both point in time as well as probabilistic forecasts. DeepAR internally trains an autoregressive recurrent neural network for producing probabilistic forecasts. It can work seamlessly on multiple time series. The best thing observed about DeepAR is that it has inbuilt special treatment for cases when the magnitude of time series varies widely. DeepAR can handle varying time series by developing a single global model. GluonTS (Salinas et al., 2020) is an open-source package for time series and internally uses Apache Mxnet to develop deep learning models. DeepAR estimator is a part of the GluonTS package.

MAPE obtained using DeepAR is 16.34. The time involved for training and prediction is 329 seconds. The model performance during backtesting is shown below in Figures 16 and 17. The green-shaded region shows a 95% confidence interval.

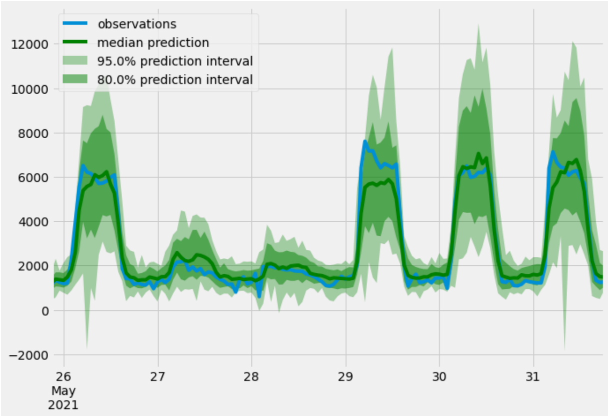


Figure 16. Deep AR prediction on back testing period (26 May 2021 – 1 June 2021).

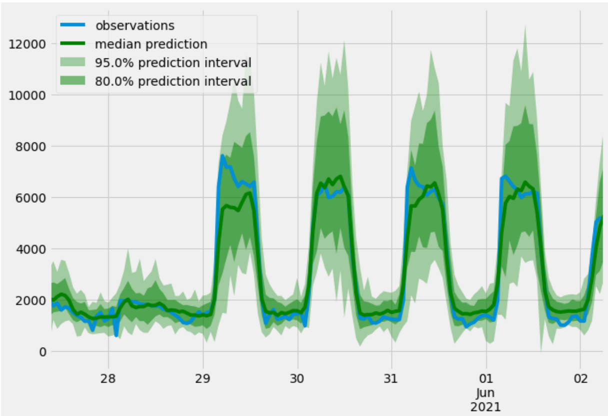


Figure 17. Deep AR prediction on back testing period (27 May 2021 – 2 June 2021).

5.5. Facebook Prophet & Neural Prophet

The next model trialled was the Prophet model (Taylor and Letham, 2018) which is an open-source library published by Facebook that is based on decomposable (trend + seasonality + holidays) models.

Prophet is a linear combination of four mathematical functions. It is based on the following mathematical formula:

$$A(t) = G(t) + S(t) + H(t) + E(t), \tag{4}$$

where,

G(t): This function is a piecewise linear or logistic curve responsible for modelling non periodic changes in the time series

S(t): This function models periodic changes (weekly, monthly, quarterly & yearly seasonality)

H(t): This function models holidays or external events which have some effect on the time series

E(t): This function accounts for some other abnormal changes in the time series data

Applying Prophet without hyperparameter tuning resulted in a much-improved MAPE of around 23 initially as compared to ARIMA. To improve the accuracy, hyperparameter tuning, and time series cross-validation were performed.

Three hyperparameters were found to be very effective for the Prophet model. These are holiday, seasonality and weightage of external variables (temperature and humidity). Each of these hyperparameter scales varies between 0.1 and 10. The greater the scale, the better the ability of the model to capture differential patterns in the data. However, sometimes choosing large values for hyperparameters can lead to overfitting. Hence, cross-validation was performed to ensure that the model can perform consistently well for each window. The cross-validation process is illustrated below in Figure 18.

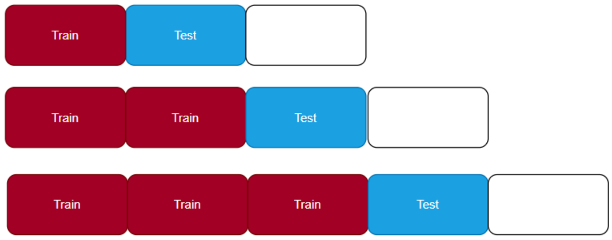


Figure 18. Cross-validation.

In cross-validation, an initial set of training data is first chosen and the prediction is conducted on the test set with the performance metrics evaluated. In the next cycle the previous test set is included in the training set and the model is again trained with the new set of training data that includes the previous training set and testing set with its performance is tested on the next testing set. This procedure is repeated for multiple cycles. Cross-validation shows that the model's performance is consistent over the entire time period.

Optuna, an open source hyperparameter optimisation framework (Bravo, 2021) is used to automate hyperparameter search. After hyperparameter tuning, MAPE obtained was 16. An advanced version of Prophet, Neural Prophet (Triebe et al., 2021), was also trialled; it is a hybrid combination of neural network and Facebook Prophet. MAPE obtained using Neural Prophet was 20. In terms of efficiency, compared to SARIMAX, Prophet is much better for parameter tuning (Merwe, 2018).

The results are summarised in Table 1 to give a comparison between the different models, their key characteristics, and the accuracy of the results obtained. The MAPE score is improved to 16 which is considered good, while below 10 would be excellent.

Table 1. Comparison of different models used for time series prediction of cooling load demand.

Sl No	Name	Key characteristics	MAPE score	Remarks
1	ARIMA	**AutoRegressive Integrated Moving Average** is a class of models that capture a spectrum of different standard temporal structures present in time series data. It is described by the parameters *p*, *d* and *q* for lag observation, number of differentiation and size of moving windows respectively.	53	Does not have the ability to learn seasonality; therefore, it gave poor results.
2	SARIMAX	**SARIMAX** is an updated version of the ARIMA model that includes seasonal effects and eXogenous factors. It needs four additional parameters, e.g., Seasonal AR specification, Seasonal Integration order, Seasonal MA, Seasonal periodicity.	43	Can give better results using Fourier series transformation on time series data.
3	Prophet	**Prophet** is an opensource algorithm for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.	16	Can give better results with further hyper-parameter tuning.
4	DeepAR	DeepAR is probabilistic time series forecasting model which trains an autoregressive recurrent neural network.	16.24	Significant accuracy can still be achieved by increasing the complexity of neural networks.
5	LSTM	End-to-end multivariate LSTM uses a sequence-to-sequence model for time series modelling.	15	The model's training time is high in LSTM.
6	XGBoost LightGBM CatBoost	All the three algorithms are built on top of gradient boosting. They are the fastest algorithms of all the algorithms considered in this paper.	21.23 18.96 19.52	Accuracy can be increased by increasing the depth of the decision trees.

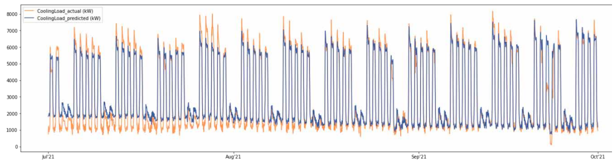


Figure 19. Evaluation results of the final A.I. model based on the tuned Prophet algorithm.

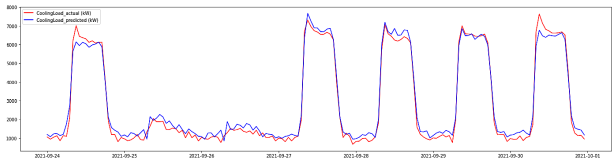


Figure 20. Forecast for one week in September (24th - 31st 2021) using the trained tuned Prophet model.

5.6. Anomaly Detection

An anomaly detection module is also included in the current paper. This functionality will help to avoid risks by identifying abnormalities in the data set.

For the current problem, anomaly detection is performed using Isolation Forest (Liu et al., 2008). Isolation Forest is an unsupervised learning algorithm for anomaly detection. It works on the principle of isolating anomalies. It builds an ensemble of isolation trees for the data set and anomalous points have shorter average lengths on the isolation trees.

Isolation Forest isolates the observation by randomly selecting some features and starts finding anomalies by splitting the value between a minimum and maximum range.

The core hyperparameter for Isolation Forest is contamination level. It represents the amount of contamination in the data set. Contamination represents the percentage of outliers that need to be detected in the data set. The best value of contamination level is obtained by multiple experiments and domain knowledge.

In Figure 21, the global outliers are illustrated for the current problem. Global outliers are abnormal datapoints in the data set. These data points are different from the rest of the data points in a quantitative sense. Figure 22 shows contextual outliers, which represent abnormal contextual datapoints. These data points are not very different from other data points in the quantitative context but are in the business context.

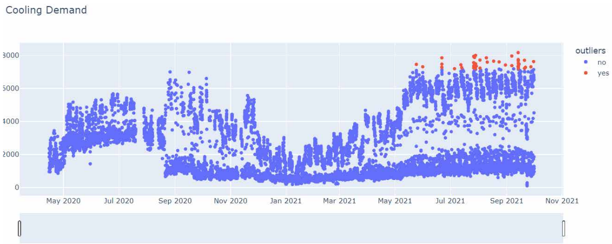


Figure 21. Global outliers.

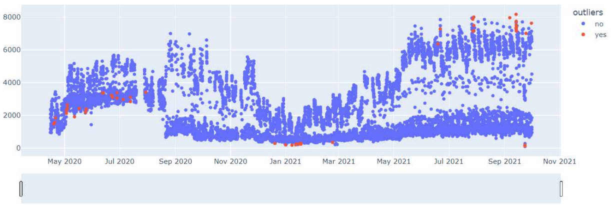


Figure 22. Contextual outliers.

5.7. Application of the Cooling Load Prediction Model

The cooling load prediction model developed is adopted in the RaSpect Smart Building Management System®. This integrated building management system (iBMS) offers a one-stop solution for the property management industry. Its modular design enables extensive compatibility for retrieving data from different building services systems, while channelling the data through an automated cleansing process and allowing complex computational analysis to be performed. The System is adopted in a commercial building in Hong Kong as a proof-of-concept project, demonstrating its capability in monitoring and controlling various building services systems with improved efficiency.

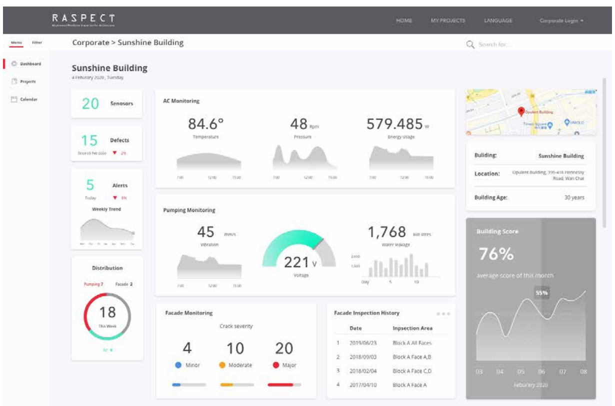


Figure 23. Illustration of RaSpect Smart Building Management System®, which incorporates a cooling load demand prediction model.

6. Benefits

Adoption of artificial intelligence in predicting the power consumption loading of a building can vastly improve its energy efficiency by adjusting the output of the building services systems, according to the predicted usage. A non-exhaustive matrix of environmental and human factors is considered during the prediction, thereby achieving high accuracy. More importantly, the A.I. model can learn from the discrepancy between its prediction and the actual data, further improving the responsiveness and agility of the model.

In this paper, the use-case spotlights the cooling load prediction, while the technology is applicable and transferrable to other different building services systems, including lighting, pumping & plumbing and other E&M facilities. Forecasting the usage and controlling the output of the systems automatically prior to the assumption can also improve the experience of the occupants in the building, by offering an optimised indoor environment.

7. Limitations

Big data analytics requires training data, and usually the more the better. The existing study incorporated a limited amount of data, hindering the prediction accuracy. Application of iBMS would require contribution of data from different building services systems. Establishing the data pipeline and feedback control mechanism is practically difficult as different manufacturers have their own individual hardware setups and software configurations. The initialisation cost of iBMS, especially for aged buildings, is comparatively high, thus lowering the incentive for the industry to adopt this technology.

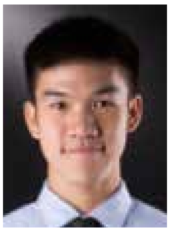
8. Conclusions

In its report titled "From Thousands to Billions", the World Green Building Council (WGBC) (2017) defines a net zero carbon building as a highly energy-efficient one that uses all its operational energy from renewable sources preferably produced on site (Merwe, 2018). It calls for all new buildings to be carbon net zero from 2030 and all buildings to be carbon net zero from 2050. A.I. can take advantage of such smart building models for data acquisition, processing, and model training for the forecasting of buildings' energy demand. LSTM neural networks are found to be the most accurate with a MAPE of 15 in 294 seconds. The Facebook Prophet time series model and DeepAR recurrent neural network model achieved the second-lowest MAPE of close to 16 in 319 seconds. The LightGBM machine learning model on the other hand proved to be the fastest algorithm with a MAPE of 18.96 in just 7 seconds. An anomaly detection model was developed to identify abnormalities in the data set in a continuous manner to avoid potential risks. Eventually it is planned to integrate the model into a smart building management platform.

Acknowledgements

The work described in this paper was expanded on the research foundation for the entry into the Global A.I. Challenge competition during 2021 and 2022. Jayant Gangwar and Pushpesh Kumar, senior students of IIT Kanpur and interns at RaSpect, are thanked for their contribution.

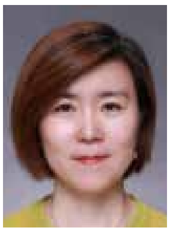
Notes on contributors



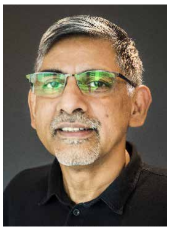
Mr Tony Ip is a Mechanical Engineer and founder of RaSpect AI. He is a professional SUA pilot with an advanced rating issued by CAD. Tony has a background in mechanical engineering and is specialised in the application of SUA for building inspection and conditional survey.

Mr Tattwa Darshi Panda has an MPhil in Mathematics and Computing from The Chinese University of Hong Kong (CUHK) and currently serves as a product manager and data scientist at RaSpect AI. He was a former visiting researcher at Tata Institute of Fundamental Research (TIFR), Bangalore, India.

Ms Xiaoyu Jia is a Doctor of Philosophy (Ph.D.) candidate in the School of Mechanical Engineering at Tongji University, Shanghai, China. She received her Bachelor of Science (B.Sc.) major in built environment and energy engineering from Beijing University of Technology, Beijing, China. Her current research area includes building performance simulation, building load prediction, occupant behavior modelling and building operation optimization. She had published 2 scientific journal papers in the past 3 years.



Prof Yiqun Pan is a key professor of Heating Ventilation and Air Conditioning in the School of Mechanical Engineering at Tongji University, Shanghai, China. Her research focuses on building performance simulation and district energy planning. She has been the principal leader of more than 30 research projects and participated in many internationally collaborated research projects including the development of an EnergyPlus Interface in Chinese (VisualEPlus), collaborating with LBNL, the development of a co-simulation model of occupant behavior, IAQ and building energy – subtask 2 of the kernel development of whole building performance simulation platform (China National Key R&D Plan). She has authored and co-authored more than 5 books and published over 110 scientific journal papers about building energy analysis and related topics.



Dr Dhanada K Mishra has a Ph.D. in Civil Engineering from the University of Michigan, USA. He has over 25 years of diverse work experience in the areas of structures, non-destructive testing, and structural health monitoring. Prior to joining RaSpect, he was a visiting research scholar at HKUST. He is a member of the Institution of Engineers (India), Fellow of the Hong Kong Concrete Institute (HKCI) and member of the American Society of Civil Engineers (ASCE). He has over 50 publications to his credit.

Prof Matthew Ming Fai Yuen is a Professor Emeritus of Mechanical and Aerospace Engineering at The Hong Kong University of Science and Technology. He is the Managing Director of Praxis Green Consulting Limited. He serves as a Specialist Professor of Wuyi University in China and Director of its Advanced Material and Equipment Research Center. He has published more than 250 papers in international journals and conferences. He holds 12 US patents, 5 Chinese patents, and has 6 US patents pending. He is a Fellow of the Institution of Mechanical Engineers (UK), a Fellow of the Hong Kong Institution of Engineers, and was a Council member of the Chinese Society of Mechanical Engineers.

Mr Harris Sun is the CEO and Founder of RaSpect Intelligence Inspection Ltd. He graduated with an M.Sc. in computer science from the Chinese University of Hong Kong and B.Sc. in Computer Science and Information System from the University of Hong Kong. He was the one of founding Members of the Innovation Centre of Cathay Pacific Airways.
He has won multiple start-up competition awards and operations, from academia research projects to private sector projects.

References

[1] Bisong E (2019). Introduction to Scikit-learn. In: Bisong E (ed). *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, pp. 215-229.

[2] Bravo P S (2021). *Time series parameters finding using Prophet and Optuna Bayesian optimization*. [online]. Medium. Available at: <https://medium.com/spikelab/time-series-parameters-finding-using-prophet-and-optuna-bayesian-optimization-e618614bd8b7>.

[3] Daoud E A (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering, 13*(1), pp. 6-10.

[4] Elamin N and Fukushige M (2018). Modeling and forecasting hourly electricity demand by SARIMAX with interactions. *Energy*, 165, pp. 257-268.

[5] Electrical and Mechanical Services Department (2021). *Hong Kong Energy End-use Data 2021*. Hong Kong: Hong Kong SAR government.

[6] Ertel W (2017). *Introduction to Artificial Intelligence*. Cham Switzerland: Springer Cham.

[7] Fan C, Xiao F and Wang S (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Applied Energy*, 127, pp. 1-10.

[8] Farzaneh H, Malehmirchegini L, Bejan A, Afolabi T, Mulumba A and Daka P P (2021). Artificial Intelligence Evolution in Smart Buildings for Energy Efficiency. *Applied Sciences*, 11(2).

[9] Global Alliance for Buildings and Construction (2019). *2019 Global Status Report for Buildings and Construction*. Paris: International Energy Agency.

[10] Grueneich D M (2015). The Next Level of Energy Efficiency: The Five Challenges Ahead. *The Electricity Journal*, 28(7), pp. 44-56.

[11] Ho S L and Xie M (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, 35(1-2), pp. 213-216.

[12] Kawashima M, Dorgan C E and Mitchell J W (1995). Hourly thermal load prediction for the next 24 hours by ARIMA, EWMA, LR and an artificial neural network. *ASHRAE Transactions*, 101(1), pp. 186-200.

[13] Kirwan C and Fu Z (2020). *Smart Cities and Artificial Intelligence*. Amsterdam Netherlands: Elsevier.

[14] Li J, Cheng K, Wang S, Morstatter F, Trevino R P, Tang J and Liu H (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, 50(6), pp. 1-45.

[15] Liu F T, Ting K M and Zhou Z H (2008). *Isolation forest*. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 413-422.

[16] Merwe R (2018). *Implementing Facebook Prophet efficiently*. [online]. Medium. Available at: <https://towardsdatascience.com/implementing-facebook-prophet-efficiently-c241305405a3/>.

[17] Mitchell T M (2007). *Machine learning*. New York: McGraw-Hill.

[18] Mushailov J (2021). LSTM Framework for Univariate Time-Series Prediction. *Towards Data Science*. [online]. Available at: <https://towardsdatascience.com/lstm-framework-for-univariate-time-series-prediction-d9e7252699e>.

[19] Panchalingam R and Chan K C (2019). A state-of-the-art review on artificial intelligence for Smart Buildings. *Intelligent Buildings International*, pp. 203-226.

[20] Pang J P (2021). *Address to The International Conference on Intelligent and Green Building Technology*. Hong Kong. 22 July 2021. Hong Kong: The Open University of Hong Kong.

[21] Raees N A, Nassif N and Rifaie F A (2014). The effect of ventilation and economizer on energy consumptions for air source heat pumps in schools. *American Journal of Engineering and Applied Sciences*, 7(1), pp. 58-65.

[22] Priestley M B and Rao T S (1969). A test for non-stationarity of time-series. *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(1), pp. 140-149.

[23] Rosenfeld J S (2021). *Scaling laws for deep learning*. Ph.D. Massachusetts Institute of Technology.

[24] Salinas D, Flunkert V, Gasthaus J and Januschowski T (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting, 36*(3), pp. 1181-1191.

[25] Sarwar R, Cho H, Cox S J, Mago P J and Luck R (2017). Field validation study of a time and temperature indexed autoregressive with exogenous (ARX) model for building thermal load prediction. *Energy*, 119, pp. 483-496.

[26] Tao Y, Yan H, Gao H, Gao H, Sun Y and Li G (2019). Application of SVR optimized by Modified Simulated Annealing (MSA-SVR) air conditioning load prediction model. *Journal of Industrial Information Integration*, 15, pp. 247-251.

[27] Taylor S J and Letham B (2018). Forecasting at scale. *The American Statistician*, 72(1), pp. 37-45.

[28] Triebe O, Hewamalage H, Pilyugina P, Laptev N, Bergmeir C and Rajagopal R (2021). Neuralprophet: Explainable forecasting at scale. Available from: arXiv preprint arXiv:2111.15397.

[29] United Nations (no date). *Department of Economic and Social Affairs Sustainable Development*. [online]. Available at: <https://sdgs.un.org/>.

[30] Wang Z, Hong T and Piette M A (2019). Predicting plug loads with occupant count data through a deep learning approach. *Energy*, 181, pp. 29-42.

[31] World Green Building Council (2017). *FROM THOUSANDS TO BILLIONS*. London: World Green Building Council.

[32] Yu Y, Si X, Hu C and Zhang J (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, 31(7), pp. 1235-1270.