# Energyconsumptionforecasting byusingmachinelearningandCloudComputingforsmart building:CasestudyinIndia

*by* SAHIL CHAKRABORTY

# “Energy consumption forecasting by using machine learning and Cloud Computing for smart building: Case study in India”

# ABSTRACT

This study focuses on the smart building electric consumption prediction through data-driven methodology. The gist of it is to harness historical utilization data to predict future demand patterns. The project applies Python libraries such as pandas, statsmodels, and matplotlib for data treatment, analysis, visualizations, and modeling.

The process involves:

1. Data Acquisition and Preprocessing: Ingesting data from a CSV file, dealing with blanks, and manipulating dates appropriately.

2. Exploratory Data Analysis: Visualizing the consumption data in time series diagrams, scatter plots, and kernel density estimations to comprehend the distribution and trends.

3. Stationarity Analysis: Detrending the data with seasonal and residual components and applying Dickey-Fuller tests to test stationarity, a precondition for many forecasting models.

4. Data Transformation: Applying logarithmic transformation and calculating exponentially weighted moving averages, probably, to get the data that is stationary.

5. Model Building and Forecasting: Implanting a Simple Exponential Smoothing (ETS) model to track the trends in the data and foresee the consumption of the next 300 days.

6. Evaluation: Illustrating the real thing vs. predicted consumption used to measure the model's performance.

# Chapter 1

# Introduction

The global demand for energy keeps growing, the reason coming from various factors, like population growth, urbanization, and putting much more emphasis on Laws in technology. A portion of the massive energy demand is coming from buildings, where we have living spaces, office areas, and amenities being powered, which constitutes a considerable part of worldwide energy consumption. This has brought a great deal of emphasis to enhancing existing buildings' energy efficiency.

Intelligent building systems are an important instrument for making this dead aim real. These smart sensor monitoring systems employ various actuators, sensor networks, and control systems to increasingly automate and optimize the operations of a building. One of the principal attributes of a building's intelligence is asking about electric consumption and managing it. Accurate forecasting of electricity usage allows building managers to proactively implement energy-saving measures, leading to significant benefits: Accurate forecasting of electricity usage allows building managers to proactively implement energy-saving measures, leading to significant benefits:

• Cost Savings: Knowing energy demand forecasts allows building managers to choose the most optimal energy procurement policies, buy electricity during hours of low demand, and avoid peak demands with pricing requirements.

• Improved Sustainability: Reducing energy usage hence reducing energy transmission and consequently minimizing the carbon issue into a greener built environment.

• Enhanced Comfort: The predictive models are the guiding tools that could pre-heat the premises according to the expected number of occupants and changing weather conditions. This is meant to make consistent thermal comfort for people.

• Preventative Maintenance: Consumption data analytics can prove to help you in identifying anomalies, and even potential device malfunction, which ensures preventive maintenance which in turn prevents the production breakdown that could eventually have disastrous impacts.

This study gives us an understanding of data-driven methodologies in electric building consumption prediction by making use of AWS for data storage, processing, and model generating The most essential aspect is to use the analysis of past usage to come up with hypotheses about future demand. Through the trends analysis and dependence factors identification models can be formed to predict electricity demand with curves under the nature of this process.

Leveraging AWS for This Data-Driven Prediction Modeling

AWS provides a rich set of cloud computing services which are a significant enabler in building a data-driven forecasting system for the electric power consumption of a smart building. undefined

1. Data Acquisition and Storage: Smart meters or building management systems usually produce electricity data that can be gathered and stored in a service like Amazon S3 simple storage (S3).

S3 delivers an integrable and economical object storage option for storing large data sets securely in the cloud.

2. Data Preprocessing and Analysis: AWS services like Amazon SageMaker can be used to pre-process the consumption data that is stored in S3. SageMaker provides data cleaning, transformation, and feature engineering tools, which are the key pre-processing operations before constructing a forecasting model. Furthermore, services such as Amazon QuickSight or Amazon Redshift can be relied upon for data exploration and visualization to make out the nature of the consumption data.

3. Model Selection, Training, and Deployment: AWS has a range of ML services available for building ML-based models, most suitable for forecasting purposes. Amazon Forecast is a dedicated service that works completely automatically for time series forecasting tasks. It enables users to put variables in and pick proper forecasting models, then train and deploy the models without the need for laborious machine learning expertise. Otherwise, Sagemaker is more suitable for creating a custom forecasting model with the help of libraries such as scikit-learn or TensorFlow.

4. Model Evaluation and Visualization: The AWS services such as Amazon CloudWatch can be used for monitoring the performance of the models that are deployed. Metrics regarding model accuracy are available in CloudWatch that can be used to debug the model in case the predictional results are not accurate. Apart from that, tools such as Amazon QuickSight can be deployed to compare the forecasted values with the actual ones. hands-on activities and real consumption data so to test the model's performance by eye.
The features of leveraging AWS for Data-Driven Prediction.

Integrating AWS into the data-driven prediction process offers several advantages: Integrating AWS into the data-driven prediction process offers several advantages:

• Scalability and Flexibility: The AWS cloud infrastructure swings to the requirements for increasing and changing historical data.

• Cost-Effectiveness: Compute resources are paid on-demand, thus customers can pay only for the tools they use.

• Security and Reliability: AWS guarantees data privacy through its robust security solutions and reliable infrastructure to protect built data effectively.

• Ease of Use: An easy way of building and hosting machine learning models which is democratized through managed services like Amazon Forecast is laid out here for users with various expertise in machine learning.

This paper delves into the benefits of data-driven strategies for electric utilization forecasting in intelligent buildings utilizing the repeatable advantages offered by AWS within data storage, processing, and model development. This part will be structured in the following way – each stage of the methods will be discussed as well as the services that AWS provides for successful data management, model deployment, and performance assessment. Furthermore, this report will inform about the received outcomes, which consist of the particular model's effectiveness in terms of forecasting accuracy. Among all, this report will take advantage of informed views on the limitations of the recommended option, and give some constructive ideas to make the approach more powerful in smart building.

# Chapter 2

# Basic Concepts/ Literature Review

## 2.1. Machine learning prediction methodology

Smart buildings can take advantage of machine learning (ML) to forecast electric consumption in the next moments. By studying the data of smart meters or building management systems historically, ML algorithms would be able to identify patterns, seasonal patterns, and external life events affecting energy use. Predictive analytics help in taking energy-saving measures proactively; these translate into cost savings, plus environmental gains and better occupier comfort.

A Framework for Prediction:

undefined

1. Data Acquisition and Preprocessing: We process the obtainment of electric consumption data, clean and normalize then transform it for modeling.

2. Exploratory Data Analysis (EDA): Through data visualization, we can grasp the data's distribution, we can create seasonality line graphs, and we can reveal the outliers. This information is used to make the right ML algorithm selection that can cater to the prediction.

3. Model Selection:

undefined

• Simple Exponential Smoothing (ETS): This is the forecasting model that is mainly used for Time series data that is especially suitable for tracing the trends of time. It is rather easy to implement and can be a useful starting point for an analytical method that can be compared with another one.

• Autoregressive Integrated Moving Average (ARIMA): ARIMA models are very effective in terms of reproducing seasonalities and trends in time series data. They are a good option if you collected your dataset under a clear seasonal pattern, for example, higher consumption during the summer or the winter.

• Gradient Boosting Models (GBMs): This is an approach to design models that amalgamate a few weak decision trees which result in a tremendously powerful ensemble learner. GBMs are very versatile and they can capture complex non-linear relationships within data possibly influenced by environmental conditions (e.g., temperature and occupancy) and other factors.

4. Model Training: For each one of the ML algorithms chosen, that one is going through a training period on a portion of the preprocessed data. During training, the data is fed through the model and the model learns the steco of consumption data. In the process, the model adjusts its internal parameters to minimize the prediction errors.

5. Model Evaluation: All models are thoroughly assessed and subjected to a different test set which aims to curb the probability of generalizing the unseen data. The measurement

model is trained with the Mean Squared Error (MSE) to track the model accuracy of each consumption prediction.

6. Model Refinement and Deployment: According to the model assessment, such refining may be achieved through the use of a hyperparameter tune or the use of the feature engineer. The model that shows the best accuracy on the test set later might be deployed to scrutinize power consumption in the future in the smart building.

Leveraging a Multi-Model Approach:

Indeed, a model chosen can not always capture the whole nuances of construction energy consumption. Creating cross-strengths of the multimodel algorithms, through this we can make gains of their respective capabilities. Similarly, to make an assessment, the ETS can be substantially used while ARIMA is the most efficient tool for seasonal variations. GBMs, as complex models, can also identify latent connections that other models are not suited for.

In the end, the decision on which model to expend for the deployment depends on the project's requirements and how well the model works on the common test set. The evaluation function helps us choose between models that generalize better on the data never seen.
Utilizing Predictions for Smart Building Management: Utilizing Predictions for Smart Building Management:

The predicted electric consumption data can be integrated into building management systems to: The predicted electric consumption data can be integrated into building management systems to:

• Optimize Energy Procurement: Through a realization of future consumption patterns, building managers can become more flexible in their purchasing of energy at off-peak times when prices drop down.

• Implement Demand Response Programs: Sharing the load with demand response programs pays building owners for lowering electricity use during the peak times of higher demand and getting compensated. Correct predictions are vital if one has to follow the action closely.

• Pre-Conditioning for Comfort: Buildings can be curated at optimal thermal comfort based on projected attendance and weather data. This reduces energy used in the building but does not affect thermal comfort.

• Preventative Maintenance: Any anomalies in consumption patterns are a clue that may be associated with equipment breakdowns. Proactive maintenance is possible due to this early detection, and this decreases the wasted money that is spent on very expensive repairs and service interruptions.

Conclusion

The ML-based billing and prediction SE-AI system is designed to be able to generate electric consumption predictions in smart buildings. By employing a multi-model technique at last they can find the most reliable one during rigorous evaluation which will give them hints about future energy usage. Such knowledge equips them, resulting in proactive energy-saving strategies. They are now able to build a more stable and economical structure of the environment.

## 2.2. Management of missing data

This project investigates machine learning for predicting energy consumption in smart buildings, focusing on a case study in India. A crucial initial step involves addressing missing data, a common challenge in real-world datasets.
Data Loading and Missing Value Handling:
1. Data Import: Libraries like pandas are used to load the data, typically stored in a CSV file with columns like "Date" and "Consumption."
2. Missing Value Detection: The percentage of missing values is assessed. If the amount is minimal (a few percent), techniques like deletion can be employed. The code utilizes the pdf. drop a () function for this purpose.

Alternative Missing Value Strategies (for Extensive Missingness):
- Interpolation: Techniques like linear interpolation can estimate missing values based on surrounding data points.
- Imputation: Statistical methods like mean/median imputation can replace missing values with the average or median of the existing data. Libraries like sci-kit-learn offer imputation functionalities.

Choosing the Right Approach:
The selection of the missing value handling strategy depends on the data characteristics and the extent of missingness. For a small percentage of missing data, deletion might be acceptable. However, for significant missingness, techniques like interpolation or imputation become crucial to retain valuable information and improve the accuracy of subsequent analyses.
Data Cleaning and Moving Forward:
- Data Type Conversion: The "Date" column is often converted to DateTime format for time-series analysis. This is typically done using pandas functions.
- Setting the Index: The datetime column is frequently set as the index for the data frame, simplifying time-based operations.

By addressing missing data effectively, we prepare the energy consumption data for further exploration, feature engineering, and ultimately, machine learning model training for accurate prediction of energy use in smart buildings.

## 2.3. Employment of cloud-based prediction modeling

Raising energy demand and environmental impact due to energy implies the need for smarter building management principles. Intelligent buildings, such as those provisioned with sensing technology and based on ML algorithms, can cut down on energy consumption and operating expenses. This project is leveraging Machine Learning (ML) on the Amazon Web Services (AWS) platform to predict energy consumption for smart buildings in India.

Data Acquisition and Preprocessing

Data collection in the project starts gathering real-time energy consumption of a smart building in India. undefined

• Date (Years): The following item takes users through the time scale for which the energy consumption is calculated.

• Electricity Consumption: For this characteristic, the exact energy consumption value at the specified moment in time holds significance.

The data may cover a specific period and could be a bit dirty which might require some cleaning and preprocessing before being involved in model training.

Cloud Integration with AWS

AWS platform caters to various cloud-based data storage, processing, and analytics needs. undefined

1. Amazon S3: Such object storage service secures your energy consumption data by uploading them into the cloud in a protected manner. S3 Scalability and Durability are suitable for keeping Time-series information.

2. Amazon SageMaker: Such a managed machine intelligence approach is about the development, education, and deployment of ML models for the energy prediction process. Sagemaker provides staffed algorithms and makes model development simple, which consequently reduces the need for perfecting machine learning skills.

3. Amazon EC2 (Optional): If needed, the resources such as VMs on EC2 are easily scaled to render the data preprocessing and complicated model training tasks. EC2 provides freedom to opt in the amount of computing power that is needed for the project realization.

### Machine Learning for Prediction

After the data has been moved into S3, SageMaker can be employed to train a wide range of ML models for estimating energy consumption. undefined

• Simple Exponential Smoothing (SES): This is quite a simple forecasting technique that can help to reflect current trends for a short period. It is a good way to begin to feel what the data is about and to get acquainted with the initial model.

• ARIMA (Autoregressive Integrated Moving Average): This is a linear model that represents past consumption patterns and statistical properties correlating to future values. ARIMA can be more accurate than SES at better identifying seasonal variations or long-term trends in the data.

• Long Short-Term Memory (LSTM) Networks: They form a type of recurrent neural network that is well-suited to model sequential data such as time series. The advantage of LSTMs is that they can memorize long-term dependencies hence making the predictions precise and accurate even when dealing with various and complex relationships between inputs and outputs.

### Model Training and Evaluation

• The data in S3 had been literately preprocessed, and training and testing sets were created. The training set is used to teach the model, while the test set is used to check its effectiveness on the data that it hasn't seen before.

• The particular ML model is then trained on the training data by SageMaker. SageMaker has all the tools for managing the training itself: hyperparameter tuning to improve model performance.

• The model's performance is measured on the hold-out data using the measures of the mean squared error (MSE) or the root mean squared error (RMSE). Such parameters are a tool for measuring and comparing the actual and predicted amount of consumption.

• The model which gives the best performance is chosen for deployment out of the testing set.

### Deployment and Prediction

• The model has been developed in the form of a web service and has been deployed to SageMaker. SageMaker allows the developers to simplify model deployment resulting in faster execution of predictions by real-time predictions.

• Sensor data, such as smart building sensors that contain date and electricity consumption values, can be sent in real-time or as a batch as required to the deployed model for competitive energy consumption prediction.

• The future energy consumption values shown on dashboards, which could be further applied for building control and optimization. For example, such a type of forecast may help to regulate heating, ventilation, air conditioning (HVAC) equipment, and lighting systems to save energy by the expected demand.

Benefits of Cloud-Based Approach

• Scalability: The scaling of services on AWS is easily affected by the project requirements. This thus, provides cost effectiveness and secures against the purchase of resources in excess.

• Cost-Effectiveness: AWS uses pay-per-use pricing that enables lowercase projects to have custom workloads. You pay based on what you use.

• Accessibility: There is no matter where cloud-based platforms can be accessed from, providing remote monitoring and administration of the energy usage in the smart building.

• Security: AWS comes with safe security functions to protect building data with sensitive information, including access control and encryption.

# Chapter 3

# Methodology

The research project would foresee energy use, utilizing the dataset we compiled in-house from January 1991 to January 2021; which was accomplished through the study of reviewed literature. We have output the one occupant dataset projection, perceived as the trend, to be plugged into the model. The predictive methodology incorporates three machine learning algorithms: k-NN, SVM, as well as ANN, with the help of several years of historical data on electricity consumption and actual electricity consumption as feature attributes to determine demand. To generate a prediction model by using Python's basic data processing tools, for example, the Augmented Dickey-Fuller Test (ADFT) would be employed. The data analysis and pre-processing steps will be performed before model training and testing to ease the model training effort. Also, the missing data will be handled properly during these two steps before model training is done. Techniques for the validation process will be applied for each model. undefined

1. Evaluating dataset normalcy

2. Data Preparation

3. The continuous improvement of models for training and development.

4. Model evaluation (testing)

## 3.1. Step 1: Normality testing of dataset

Undoubtedly, normality testing is the first step that helps to draw the distribution graph for consumption of the electricity. The abbreviated augmented Dickey-Fuller test, ADFT, is used to check the non-stationarity. The degree of skewness and the kurtosis levels characterize whether the distribution is normal or not (or out of the norm).

Normalization testing comes next, then you can introduce case machine learning models for consumption prediction. These models include: Normalizaton testing comes next, then you can introduce case machine learning models for consumption prediction. These models include:

• The BEMA (Basic Exponential Moving Average).

• Estimation of time series Holt-Winters with seasonal and trend coefficients (HWES).

• Automatic Interrelated Moving Average (ARIMA).

• LSTM networks and when the problem dealing with the complicated series of patterns is concerned is a good choice.

The model acquisition will be based primarily on attributes of the data and needed target

## 3.2. Step 2: Data pre-processing

This work will teach data prep methods for electrical consumption data. These methods will provide the foundation for a machine learning model, which will be used for predicting the electricity demand of smart buildings. Here we will have a case study of India.

Data Pre-processing

The quality of data, in the meantime, is the foundation of machine learning which determines the effectiveness of the models. Data preprocessing is the process that cleans the raw data for model use through measures such as missing values identification, correction of errors, and stationarity assurance. Here's an overview of the key steps involved: Here's an overview of the key steps involved:

1. Data Loading: The first step shall be to import information from a CSV file with the date and electricity values into the Excel program.

2. Data Cleaning:

o Column Naming: Columns are clarified (for ex., Date' and 'Consumption').

o Handling Missing Values: Missing observables can spoil the learning process of the model. Such ways will help in dropping rows of missing methods or methods of imputations may be also used.

3. Data Formatting:

o Setting Date as Index: Time as a factor is accounted for using the 'datetime' format and designated the index for further time-series analysis. This permits the proper techniques to be applied and the outcome to be analyzed, based on the time.

4. Data Exploration and Visualization:4. Data Exploration and Visualization:

o Visualization: These methods like line plots (plt. plot(pdf)) and scatter plots (df. plot(style='k.')) give clarity in pinpointing trends and outlier points.

o Distribution Analysis: Kernel density estimation, dissected through the KDE plot:(df. plot(kind='kde')), will contribute to the visualization of the distribution of consumption values, giving a possible indication of the presence of skewness or unusual patterns.

5. Trend Analysis: Separating time series data into trend, seasonal, and residual components utilizing appropriate algorithms as in seasonal_decompose from stats models is very useful for understanding underlying patterns and seasonality of energy usage.

6. Stationarity Testing: The thing called stationarity becomes crucial in case the data properties change during the time is kept constant. This allows the forecast to be accurate. A test (ADF test), known as the Dickey-Fuller test, looks for stationarity. The cases of non-stationariness in the data can be altered using differencing or logarithmic transformation techniques. The code displays the stationarity checks.

7. Exponentially Weighted Moving Average (EWMA): It facilitates picking up trends and suppresses noise. The latter situates one in a better position to predict long-term consumption habits.

8. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF): One of the top features of these functions is to expose probable lags for model selection in ARIMA and other forecasting models. The code outputs ACF and PACF to check if any autocorrelated relationships are present in the processed data.

### 3.3. Step 3: Model development (training)

Following the data scrubbing and the inspection, model development is moved to the next stage to deliver a model that is a solid predictor for future consumption.

One key step in conducting this analysis is to check for the series stationarity. A curve that is not stationary has a range of trend and seasonal patterns which limit model accuracy. One of the Dickey-Fuller test's merits is its potential to detect whether a series is stationary (or not); and, if needed, strategies like differencing or transforming a logarithm may help in achieving that goal.

It is the implementation of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to detect the potential relationship and lags for forthcoming modeling. These measures point out how the degree of past consumption values correlates with the present values and the future values. Lagging behind the selected model shows the level of our knowledge in machine learning.

Thus, we analyze the stationarity, automatic correlation function, and partial auto-correlation function(ACF/PACF) and choose a better model for the case study(e.g., Simple Exponential Smoothing). The selected model employs historical data to figure out the actual purchase patterns and eventually can be applied to predict future consumption for a specified duration say a month or year.

This project leverages the power of Amazon Web Services (AWS) to facilitate efficient model development and deployment: This project leverages the power of Amazon Web Services (AWS) to facilitate efficient model development and deployment:

• Amazon SageMaker: The ease of work is demonstrated in this managed service where the models are built trained, and deployed without a machine learning expert. SageMaker can train the model automatically on the large datasets that it can handle from services such as Amazon

S3 and SageMaker can also publish the trained model as a web service for real-time prediction in consumption.

• Amazon EC2: On the procedurally heavy chores like training the models with complex algorithms, Amazon EC2 provides virtual machines highly customizable and scalable, hence you can get the right proportions that suit the nature of the job.

Through the utilization of AWS, the project is then capable of scaling the system to promptly analyze huge amounts of data from several smart buildings, thereby developing a model that is reliable and inclusive. This paradigm can be continuously advanced by integrating fresh data streams and anomaly training on AWS which invariably translate into better energy efficiency and cost optimization in smart buildings throughout the country.

### 3.3.1. k-Nearest Neighbour (k-NN)

k-Nearest Neighbors (k-NN) is a machine learning algorithm where regression and classification tasks can be carried out. Picturing a place where, there exist lots of data boxes with houses on a map, with parts like size and price of each house, can be better understood. If the k-NN feature happens, you introduce a new unseen house in the map as a data point. It comes to an estimated price or assigning class (like expensive or affordable) by considering the k nearest houses on the map based on their features.

The exact magic is achieved by feeding a well-tuned k while considering only the very closest neighbors – that makes the model both responsive to outliers and highly global. On the other hand, a high k means digging deep into the parameters more remote, and, possibly, it can end up with having too many data points that are noisy. Through the adoption of the best k and by the identification of common features in data points, k-NN provides an uncomplicated but surprisingly high performance in most cases of machine learning models. Yet, it should be remembered that k-NN requires to be used in the whole dataset for the prediction, which makes it more computationally costly in the case of the large datasets scale.

### 3.3.2. Support Vector Machine (SVM)

Functioning:

One can have attribute data as points that classify the different categories of a given instance, for instance, spam or not-spam emails. undefined

1. Mapping Data: First, the SVMs can predict whether the data points can generate a higher dimensional space instead and the separation will be better than before.

2. Finding the Optimal Hyperplane: The SVM classifier searches for a hyperplane with the highest distance from the closest points of the classes (called support vectors). There is even the possibility of a single degree over the absolute zero being enough for the differentiation.

3. Classification: Basically, the new data points that are unknown are projected on the same high-dimensional space. The predicate category of the new data points is decided on the side of the hyperplane where they are mapped.

Proper Use:

SVMs are powerful tools, but they are best suited for specific scenarios: SVMs are strong weapons, but they work perfectly in some cases.

• Classification Tasks: SVMs are awesome at detecting data in various groups which helps maximize the goal of the categorization. This can be the first choice of the tasks where they search for the spambots, visually identify objects on an image, or grammatically detect the attitude of sentences: pros, cons, or neutral.

• High-Dimensional Data: SVMs with higher dimensional data features enable them to deal effectively with them and thus they are the apt algorithms for complex datasets that are multidimensional with many features.

• Limited Data: Besides, the SVMs can produce good results even though the data number is small because they emphasize less on the less important vectors.

Things to Consider:

• Computational Cost: SVMs are computationally exhaustive when trained for the datasets with too many instances.

• Interpretation: The SVM method in line with its non-linearity nature being more complex compared to linear models, the exploration process that leads to the prediction could be difficult.

### 3.3.3. Artificial Neural Network (ANN)

ANNs represent one branch of ML algorithms that adopt the human brain nature and the cognitive process. Similar to brains, ANNs can pick up different types of patterns, and predict new data as well.

How ANNs Learn: An Imitation of Human Interaction

Visualize a neural network with numerous processing units linked together into dense layers like neurons in the brain. These are the basic building blocks that build neural networks. An individual from the group can easily understand this signal, perform a calculation, and simply follow the group's output. The network's network topology as perfectly demonstrated by these interconnected sagittal sections allows the system to complete those complex operations.

Backpropagation is the key to machine learning. The way it does it will be discussed later. In the course of training, the algorithm will be presented with data as well as the desired target. Hence, it analyzes its accomplishments based on irrationally expected results and modifies the links between the units, similarly to what happens in the human brain. In the loop of this process, the network accuracy is enhanced and a good end-point is achieved.

When to Use ANNs: Besides the Difficult Tasks of Moderating

They are good at mapping out latent patterns that the human brain fails to detect. undefined

• Image Recognition: This type of technology can be taught to be able to identify with a high level of accuracy any object in an image. On the other hand, these very smart cameras come in small sizes, can be easily installed, and usually consume lower power, hence making them suitable for facial recognition in photos and product identification in self-checkout systems.

• Speech Recognition: Are you now satisfied with voice commands on your cell phone, right? At the back end, the AI may convert the spoken voice into text. They precisely facilitate the units of humans and machines to be in a union.

• Natural Language Processing: Human language processing and generation is a sequential composition or sequence of tasks. Nowadays, ANNs are developed to ensure that the machines understand to translate many languages, produce various kinds of content and deal with natural questions & queries.

Important Considerations: Humanization and Nationalization are the two sides of a Coin that can surface as we are captivated by the destiny of the characters.

While ANNs offer immense potential, some factors need to be kept in mind: According to this, ANN is a magical technology which even though it has many difficult drawbacks to be improved.

• Computational Demands: The processing expense would have a rapid increase during the phase of the training of gradient-based ANN which would be achieved by the system of high seasons requiring computers with huge memory capacity and the speed of processing the data.

• The Black Box Problem: The ANN form of decision-making does not operate in the classical rule of order of human decision which used to be a linear process and fit in with the given set of rules only. The community remains overcome by ignorance and lack of understanding of art and even reckoning the sense of it. Nevertheless, this point of view can be considered disadvantageous, especially in such situations when participants will feel passive and the only possible thing that can be nice is when details are not lost.

• Fine-tuning the Learning Process: Networks additionally provide the hyperparameters features as managerial roles over training processes. Completely to the contrary, do not fail to understand that the quality and the suitableness of implementing our strategy so that it will work best for us is not anything we will come up with at once. While have to try a lot of methods as well as get knowledge about rather what goes for the compeers who have succeeded.

A network of Artificial Neurons taking into account machine learning approaches is very desirable these days. This may address their talent in detecting ambiguous meanings in patterns they experience and skills essential for the solution of prediction tasks. Only the performance and the configurations of the machine aren't as those stated earlier, for example, higher power efficiency performance, more monitoring, and tweaking the parameters of the machine according to wattage requirements.

1

## 3.4. Step 4: Model evaluation (test)

Through this assessment, the whole process of electricity consumption forecasting is assessed to determine the most effective models. Through the method with the Simple Exponential Smoothing (SES) model as part of the code, AWS has more sophisticated applications for the more difficult issues.

Machine Learning Models on AWS for Energy Consumption Prediction: Machine Learning Models on AWS for Energy Consumption Prediction:

Simple Exponential Smoothing (SES): Using this basic model will work well when series aren't linear or swayed by seasonal factors. With the AWS SageMaker, modifying, training, and deploying SES models for a hasty baseline forecasting is one possible step.

Exponential Smoothing with Trend/Seasonality (Holt-Winters): This type of SES is being extended by including factors of trend and seasonal hence it can be used for data that has the nature of trend and season. SageMaker helps get the models out for deployment as well as do the hyperparameter fine tuning so that the models are optimally accurate.

ARIMA (Autoregressive Integrated Moving Average): The previously discussed model, which is one of the commonly used ones, mainly establishes such temporal patterns in complex time series. SageMaker covers all the steps of building, training, and evaluating ARIMA models from its built-in algorithms.

Support Vector Machines (SVM): Although they do not have the time-series forecasting for the purpose originally intended, SVMs can be customized for that task. SVM architectures for energy consumption prediction can be prototyped and even unconventional connections between data elements can be uncovered with SageMaker.

Artificial Neural Networks (ANNs): Such imposing models can pick out complex patterns from available information. SageMaker enables the training of ANNs for energy usage prediction by offering tools for building, training, and deploying those models. One of the aspects of SageMaker's tuning process is to configure the network architecture in terms of the number of layers and set of weights, which are intended to make no error by choosing the best ANN model for this task.

K-Nearest Neighbors (KNN): This approach constructs potential based on gradual changes in time series by tracking prior patterns. SmartMaker provides a framework for utilizing KNN models to predict energy consumption. One needs to set the number of neighbors (K) adequately in SageMaker before training the model to obtain accurate results and results.

Evaluation Metrics:

After deployed on SageMaker, these models can be produce so many results that you can evaluate them by using so many metrics. MSE (Mean Squared Error) and MAPE (mean absolute percentage error) are the typical choices, as they serve as the different siZe of the error between the predicted and actual consumption. The model having the lowest error percentage is taken as the most precise model for this data so far.

Visualization of Results:

SageMaker brings up a window where the performance of models can be visualized. Another way can be to plot reality vis-a-vis actual data consumption above the model's forecasts. This portrays how well the model of the underlying trends and patterns captures it.

Based on SageMaker's palette of machine learning models, as well as assessment tools, this project equips researchers to carry out different prediction approaches and reach the top solution for the important task of predicting energy consumption in smart buildings.

# Chapter 4

# Results and discussion

The results of the experimentation were discussed in sections based on the steps of the prediction framework. The pre-processing method and imputation of missing data using basic concepts. The findings regarding energy consumption prediction were reviewed for each month and performance comparison was provided for the prediction result of Artificial Neural Network (ANN), Support Vector Machine (SVM) and ADT.

## 4.1 Normality of dataset

1. Visualize the Distribution: Create a histogram to see if the data roughly follows a bell curve. Look for skewness (tailing off to one side) or outliers that could affect normality.
2. Descriptive Statistics: Examine summary statistics like mean, median, standard deviation, and quartiles (25th and 75th percentiles). If the mean and median are close and the distribution is relatively symmetrical, normality is more likely.
3. Normality Tests: Conduct statistical tests like Shapiro-Wilk or Kolmogorov-Smirnov to assess normality quantitatively. These tests provide a p-value, which indicates the probability of observing the data if it were truly normal. A high p-value (generally $> 0.05$) suggests normality.

Possible Scenarios:

- Normal: If the electric consumption exhibits minimal seasonality or trend, and the distribution appears symmetrical in the histogram, the data might be close to normal.
- Non-Normal: If seasonality or trends are evident, or the histogram shows skewness or outliers, the data might not be normal.

Addressing Non-Normality:

- Transformations: In some cases, data transformations (e.g., logarithm, square root) can make the distribution more normal.
- Non-Parametric Tests: If normality is rejected, consider using non-parametric statistical tests that don't rely on the normality assumption.

## 4.2 Data preprocessing

Before analyzing the electric consumption data, it's essential to assess its normality. This can be done by visualizing the distribution with a histogram, examining descriptive statistics (mean, median, standard deviation), and potentially conducting normality tests like Shapiro-Wilk. Knowing the normality of the data will help determine if transformations (e.g., logarithm) are necessary or if non-parametric statistical tests should be used for further analysis.

## 4.3 Performance

While the provided information (month-year, comma-separated electric consumption) doesn't directly assess ADT (Anomaly Detection Tree) model performance, here's a discussion on how you might approach it:
ADT Model and Anomaly Detection:

- ADT models are specifically designed for anomaly detection. They excel at identifying data points that deviate significantly from the expected pattern.

- In your case, electric consumption data likely exhibits seasonality and trends. These regular variations wouldn't be considered anomalies by the ADT model.

Performance Evaluation for ADT on This Dataset:

- It would be challenging to directly evaluate ADT performance on this dataset because the expected behavior (seasonal and trend changes) wouldn't be labeled as anomalies.

Alternative Approaches:

1. Simulate Anomalies:

You could artificially introduce anomalies into the data (e.g., sudden spikes, drops) and then evaluate how well the ADT model identifies them. This allows you to gauge its ability to detect unexpected deviations.

2. Focus on Specific Anomalies:

If you have external knowledge about potential anomalies (e.g., equipment malfunctions, outages), you could assess how effectively the ADT model picks them up compared to other methods.

Additional Considerations:

- Feature Engineering:

Consider creating additional features from the month-year data (e.g., month number, season) to help the model capture seasonality.

- Model Comparison:

13

Compare the ADT model's performance to other anomaly detection techniques (e.g., Isolation Forest, Local Outlier Factor) on your data (potentially with simulated anomalies) to see which one excels.

## 4.4 Comparision with other models

While the provided dataset (month-year and electric consumption) offers valuable information, directly assessing the performance of an Anomaly Detection Tree (ADT) model on this raw data proves challenging. This is because the ADT excels at identifying deviations from the expected pattern, and in this case, the expected pattern includes seasonal variations and potential trends in consumption. The model wouldn't flag these regular changes as anomalies.
However, we can explore how ADT compares to other models for anomaly detection in electric consumption data, considering potential approaches and considerations:

Alternative Techniques:

- Isolation Forest: This isolates anomalies by randomly partitioning the data. Points requiring a shorter path for isolation are considered more anomalous. It's robust to outliers and efficient 17 large datasets.
- Local Outlier Factor (LOF): This calculates the local density deviation of a data point compared to its neighbors. Significant deviations are flagged as anomalies. LOF works well for identifying local anomalies and clustering normal data points.

Comparison Points:

- Sensitivity: How well each model detects actual anomalies (e.g., equipment malfunctions, outages) compared to true negatives (normal fluctuations).
- Specificity: How effectively each model avoids flagging normal seasonal variations as anomalies. This reduces false positives and unnecessary alarms.
- Computational Efficiency: Training and running time of each model for scalability, especially with large datasets.

Considerations:

- Feature Engineering: Extracting additional features from the month-year data (e.g., month number, season) can improve model performance by explicitly capturing seasonality.
- Simulated Anomalies: Artificially introducing anomalies into the data (e.g., sudden spikes or drops) allows for evaluating how each model identifies them compared to baseline behavior.
- Dataset Size and Distribution: Different models might have varying performance depending on the dataset size and the distribution of anomalies (sparse vs. frequent).

# Chapter 5

# Conclusion

Here, the research probed into the possibility of using machine learning to predict energy usage in smart buildings. Among the algorithms, efficiency estimates were made and compared. The generated findings will be a guide to selecting suitable models for practical implementations of smart buildings in India.

The opportunity for additional study in this area is immense. undefined

• High-Performance Computing on Cloud Platforms: Research into high-performance computing or AWS instances on cloud platforms for faster training of demanding models like SVMs. It opens the possibility to analyze more realistic models with possibly higher predictive power.

• Data Enrichment: Enrich the dataset by including extra attributes that might determine energy usage. The weather data, occupancy status, building attributes such as insulation and window sizes as well as appliance operation patterns may be the ones to include. An inclusive dataset can result in models that can capture the subtleties of smart buildings’ energy consumption and use.

• Hybrid or Ensemble Models: Discuss the benefits of blending different machine learning models. Hybrid models that consist of various techniques or methods where the output of different trained models is averaged can attain higher accuracy in comparison to separate models.

• Explainable AI (XAI): When it comes to the question of the accuracy of machine learning models, the understanding of the way how they reach their predictions becomes a necessary condition for the development of trust in their results. Different methods from Explainable AI (XAI) can be used to illuminate the decision-making variety of such systems, which consequently can lead to better understanding and possible further model upgrades.

• Real-Time Feedback and Optimization: Create a system, that incorporates the prediction model with real-time energy usage data for effective achievement of the goal. Continuous feedback and optimization would be available constantly thanks to this. The model will always get its place through an adjustment with the actual energy figure, which may create a more precise future prediction.

• Integration with Building Management Systems (BMS): Through the day, raise awareness on the energy predictive model and its impact. The end game is to establish a link between the energy model and the Building Management System (BMS). It will be thus possible to have power consumption adjusted there in smart buildings. The BMS, for instance, may automatically trim down heating, ventilation, and air conditioning (HVAC) settings or engage in demand response programs with utility firms on account of knowing peak consumption hours.

• Smart Grid Integration: In smart cities in the future, smart buildings will be linked to smart grid energy systems. A smart grid that has energy prediction models can be used to set renewable energy trading strategies that would be capable of optimizing energy distribution over the grid, which would contribute to an energy infrastructure that is sustainable and efficient in nature.

By addressing these research gaps and improving the accuracy and robustness of the energy consumption prediction models for smart buildings in India, we can develop forecasting models that better capture the complex dynamics of energy usage in Indian buildings. These enhancements will give way to better energy management, lower energy costs, and a lesser ecological footprint for infrastructure in the future.

# Energyconsumptionforecastingbyusingmachinelearningan... building:CasestudyinIndia

ORIGINALITY REPORT

| 4% | 3% | 2% | 2% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | umpir.ump.edu.my<br>Internet Source | 1% |
| 2 | Mel Keytingan M. Shapi, Nor Azuana Ramli, Lilik J. Awalin. "Energy consumption prediction by using machine learning for smart building: Case study in Malaysia", Developments in the Built Environment, 2021<br>Publication | <1% |
| 3 | Submitted to Swinburne University of Technology<br>Student Paper | <1% |
| 4 | Submitted to Liverpool John Moores University<br>Student Paper | <1% |
| 5 | fastercapital.com<br>Internet Source | <1% |
| 6 | Submitted to Brunel University<br>Student Paper | <1% |
| 7 | Submitted to globsyn<br>Student Paper | |

| | | |
|---|---|---|
| | | <1% |
| 8 | Submitted to The University of the West of Scotland<br>Student Paper | <1% |
| 9 | www.mdpi.com<br>Internet Source | <1% |
| 10 | Submitted to University of Nottingham<br>Student Paper | <1% |
| 11 | apps.dtic.mil<br>Internet Source | <1% |
| 12 | dergipark.org.tr<br>Internet Source | <1% |
| 13 | digitalgurus.online<br>Internet Source | <1% |
| 14 | medinform.jmir.org<br>Internet Source | <1% |
| 15 | www.york.ac.uk<br>Internet Source | <1% |
| 16 | open.alberta.ca<br>Internet Source | <1% |
| 17 | unsworks.unsw.edu.au<br>Internet Source | <1% |