| **Student Name** | **:** | **MUHAMMAD ABQARI BIN ZULKIFLI** |
| --- | --- | --- |
| **Matric Number** | **:** | **288492** |
| **Supervisor** | **:** | **Suwannit Chareen Chit a/l Sop Chit** |

# Preview

The project's name, SENDA, is taken from the Spanish word for "pathway," which represents the strategic direction and foresight required for monitoring and forecasting energy demand. SENDA is an abbreviation for "Forecasting Sectoral Energy Demand Using AI Approach," which encapsulates the project's primary goal of developing a predictive model for energy consumption across many sectors.

In today's environment, consistent energy availability is critical for supporting daily operations, ranging from domestic activities to large-scale industrial functions. The SENDA project emphasizes the crucial need for efficient energy resource management in the face of problems such as population growth, industrial expansion, and climate change.

By integrating modern artificial intelligence (AI) approaches, SENDA aims to create a clear and exact "pathway" for projecting energy demand, allowing for more robust and sustainable energy management solutions.

The forecasting model created by SENDA takes a hybrid approach, including different machine learning techniques to improve prediction accuracy and robustness. This innovative approach combines Support Vector Regression (SVR), Artificial Neural Networks (ANN), Particle Swarm Optimization (PSO), and Multiple Linear Regression (MLR), resulting in comprehensive and detailed energy consumption projections.

SENDA's framework is built on high-quality, comprehensive historical data gathered meticulously from credible sources such as the Suruhanjaya Tenaga (ST), Department of Statistics Malaysia (DOSM) and the Kementerian Peralihan Tenaga dan Transformasi Air (PETRA). This data ranges from 1978 to 2021, giving a solid foundation for modeling and ensuring that the projections produced are accurate and relevant.

SENDA's objective is consistent with its name, guiding stakeholders down a strategic route towards optimal energy production and distribution. By precisely estimating energy demand, SENDA hopes to reduce wasteful use, improve environmental effect, and contribute to the development of sustainable energy practices. This initiative not only meets the urgent needs of energy management, but also lays the groundwork for future advances in forecasting technology and sustainable resource utilization.

# **1.0** **INTRODUCTION**

In a world without consistent access to energy, humanity would face tough issues affecting everything from basic household functions to large-scale industrial operations. The absence of reliable energy sources would affect critical services such as communication networks, transportation systems, and healthcare facilities. This scenario emphasizes the crucial significance of efficient energy resource management, which is compounded by variables such as population growth, industrial expansion, and the approaching risk of climate change. Addressing these difficulties requires reliable energy demand forecasting across all sectors. Simultaneously, the negative impacts of climate change, such as harsh weather and increasing sea levels, highlight the urgency of transitioning to more sustainable energy.

While a power outage shows our reliance on electricity, the larger issue is how to effectively regulate and optimize energy consumption in the face of rising demand and environmental concerns. Fortunately, today's artificial intelligence (AI) technologies provide potential solutions to these complicated problems. By utilizing AI approaches, we may overcome traditional forecasting limits and adapt to the dynamic nature of energy use trends. These powerful algorithms allow us to integrate multiple datasets, detect complex correlations, and estimate demand changes with higher precision. Indeed, research has demonstrated the potential of AI in improving energy forecasting accuracy and efficiency. By leveraging AI algorithms, we can better anticipate and control energy demand variations, paving the path for more resilient and sustainable energy management strategies.In conclusion, while a power outage highlights the crucial significance of electricity in modern civilization, the larger difficulty is successfully managing energy resources across the board. By precisely estimating energy demand and usage, we may implement strategies to reduce energy consumption while enhancing environmental impact. We can uncover ways to more sustainable and resilient energy management practices by combining AI technology and energy forecasting methodologies, reducing risks and enhancing efficiency in an increasingly complex energy landscape. These ideas provide a glimmer of hope for decreasing natural disasters caused by excessive energy use while also contributing to the preservation of our planet.

## 1.1 PURPOSE

The project's goal is to create a hybrid model for showing energy usage and demand in certain sectors. This project aims to increase the precision and reliability of energy demand projections across multiple industries, resulting in more efficient energy management methods. By improving the precision of these projections, the project hopes to deliver useful insights that will help optimize energy production and distribution, ultimately leading to improved resource utilization and sustainability.

Recent research has shown that hybrid techniques, which incorporate various predictive models, enhance forecast accuracy and robustness. For example, in cloud workload prediction, hybrid models such as Autoregressive Integrated Moving Average - Long Short Term Memory (ARIMA-LSTM) and Convolutional Neural Network - Long Short Term Memory (CNN-LSTM) outperform single models in capturing complicated, non-linear trends. Liu et al. found that a hybrid ARIMA-LSTM model improved prediction accuracy by 6% and 66% when compared to solo LSTM (Long Short-Term Memory) and ARIMA (autoregressive integrated moving average) models (Maiyza et al., 2023)(Fontes & Silva, 2019).

Hybrid weather forecasting models using deep learning techniques like Convolutional Neural Network (CNNs) and Recurrent Neural Networks (RNNs) outperform standard methods by properly managing long-term dependencies in meteorological data (Utku & Can, 2023). Similarly, in the domain of cryptocurrency price forecasting, combining LSTM with attention mechanisms and gradient-specific optimization has proven to boost predictive performance, utilizing the strengths of different deep learning algorithms (Ladhari & Boubaker, 2024).

This study uses a hybrid modeling technique to take advantage of the complimentary characteristics of several algorithms, resulting in more dependable and accurate energy demand estimates. These models will use machine learning techniques such as Support Vector Regression (SVR), Artificial Neural Networks (ANN), Particle Swarm Optimization (PSO), and Multiple Linear Regression (MLR) to provide comprehensive and exact forecasts about energy usage across many industries.

## 1.2 SCOPE

The scope of this research is the use of hybrid models to forecast energy consumption in various sectors. It will use data from open-source platforms, the Suruhanjaya Tenaga(ST), the Department of Statistics Malaysia (DOSM), and the Kementerian Peralihan Tenaga dan Transformasi Air (PETRA). The comprehensive quality of this data will allow for a detailed knowledge of energy demand trends. This will enable the development of models that are not only accurate but also flexible to diverse sector-specific details, ensuring that the forecasts produced are widely applicable and reliable.

//Add specific resources for each of the data

## 1.3 DEFINITIONS, ACRONYMS AND ABBREVIATIONS

* **AI**: Artificial Intelligence
* **SVR**: Support Vector Regression
* **ANN**: Artificial Neural Networks
* **PSO**: Particle Swarm Optimization
* **MLR**: Multi Linear Regression
* **ST**: Suruhanjaya Tenaga
* **DOSM**: Department of Statistics Malaysia
* **PETRA**: Kementerian Peralihan Tenaga dan Transformasi Air
* **SENDA**: Forecasting Sectoral Energy Demand using AI approach
* **Overfitting**: A modeling error that occurs when a model is too closely aligned to the training data, capturing noise along with the underlying pattern, thus performing poorly on new, unseen data.
* **KTOE**: Kilotonnes of Oil Equivalent
* **R2**: Coefficient of Determination
* **MSE**: Mean Squared Error
* **RMSE**: Root Mean Squared Error
* **MAE**: Mean Absolute Error
* **KwJ:** Kilowatt-Joule
* **ARIMA-LSTM**: Autoregressive Integrated Moving Average - Long Short Term Memory
* **CNN-LSTM**: Convolutional Neural Network - Long Short Term Memory
* **RNN:** Recurrent Neural Networks
* **L1**: Lasso regression ; L1 regularization adds the absolute value of the coefficient as a penalty term
* **L2**: Ridge regression ; L2 regularization adds the squared magnitude of the coefficient as a penalty term

# **2.0** **GENERAL DESCRIPTION**

SENDA is a forecasting tool for energy consumption that employs AI approaches to increase prediction accuracy. This means creating and testing a hybrid model that includes many forecasting methodologies. The system's primary users will be energy management experts, regulators, and researchers who require exact energy demand forecasts to optimize energy production and distribution.

The research acknowledges some limitations, including the reliance on data quality and the complexity of hybrid models, which can lead to overfitting if not managed properly. Overfitting happens when a model learns not just the training data's underlying patterns, but also the noise and random fluctuations. This indicates that although the model exhibits remarkable performance on the training set, it is unable to extrapolate to new and unknown data, resulting in poor results in practical applications. Overfitting usually happens when a model is trained for an excessive number of iterations, which causes the model to memorize the training data instead of learning the general patterns, or when highly complex models have an excessive number of parameters in comparison to the amount of training data. Because hybrid models combine numerous methods, each of which adds to the overall complexity, this is particularly important to them.

Overfitting has negative effects on the model's capacity to generalize from training data to unknown data, which reduces the model's usefulness for actual forecasting. The efficacy of energy management measures based on these projections might get compromised as a result, leading to false projections and unreliable forecasts. A split train-test strategy will be used in the project to reduce the risk of overfitting. To do this, split the available dataset into two subsets: a test set for performance evaluation and a training set for training the model. The train-test split facilitates the evaluation of the model's generalization ability to fresh, unobserved input. According to the size and features of the dataset, a typical split might assign 70–80% of the data to training and 20–30% to testing.

Using a train-test split has several advantages, such as validating model performance, spotting overfitting swiftly, and helping with model selection and tuning. Understanding the model's practical relevance requires an objective assessment of its performance on fresh data, which can only be obtained by testing the model on a test set. Overfitting can be identified early on and the model can be adjusted if a model performs significantly better on training data than on test data. The train-test split also makes it possible to compare various models and hyperparameters, which aids in the process of choosing the best-performing model for implementation.

Regularization strategies like lasso regression(L1) and ridge regression(L2) regularization, which reduce the model's complexity, and cross-validation approaches, which further guarantee reliable evaluation by splitting the data into numerous training and validation sets, are further tactics to prevent overfitting. Overfitting can also be avoided by early stopping, which is keeping an eye on the model's performance on a validation set during training and pausing when it begins to decline. By averaging out the mistakes, ensemble approaches, such as bagging or boosting, aggregate the predictions of numerous models and help prevent overfitting.

The project is predicated on the availability of sufficient, high-quality data from sources like ST, DOSM and PETRA. The hybrid models need this data to be trained and validated. In addition, the initiative depends on sufficient computer resources being available to manage the sophisticated algorithms and enormous datasets needed to create and evaluate the hybrid models.

## 2.1 SYSTEM FUNCTION

The system will serve as a forecasting tool for energy consumption, using AI techniques to improve forecast accuracy. The system's goal is to give credible projections of energy consumption across many sectors by constructing a hybrid model that integrates numerous forecasting approaches. These hybrid models will include machine learning algorithms for example Support Vector Regression (SVR), Artificial Neural Networks (ANN), Particle Swarm Optimization (PSO), and Multi Linear Regression (MLR) to ensure accurate forecasting. The ultimate goal is to help stakeholders implement more effective energy management methods by delivering accurate and actionable forecasts.

## 2.2 USER

The system's major users will be energy managers, policy makers, and researchers. Energy management specialists can utilize the forecasts to optimize energy production and distribution, ensuring that supply and demand are met effectively. These projections can help policymakers make informed decisions about energy policies and regulations, with the goal of achieving sustainable and efficient energy usage. Researchers will use the system's rich data and predictive capacities to advance their research on energy consumption patterns and the factors that influence them. Overall, the system is designed to serve as a useful tool for everyone in the energy sector who wants precise and dependable energy demand estimates.

## 2.3 GENERAL CONSTRAINTS

Several constraints could have an impact on the forecasting system's performance. To begin, the accuracy of the forecasts is strongly dependent on the quality and completeness of the acquired data. Inaccurate or inadequate data might result in incorrect forecasts, reducing the system's reliability.

Second, the complex nature of the hybrid models used may result in overfitting if not properly controlled. Overfitting happens when a model is overly customized to the training data, limiting its capacity to generalize to new, unseen data.

Finally, hybrid models' complexity can make them difficult to interpret. Understanding the underlying relationships in the data and how different factors influence energy demand estimates can be difficult, thereby limiting the system's transparency and user trust.

## 2.4 ASSUMPTIONS AND DEPENDENCIES

For the project to succeed, a number of crucial assumptions and dependencies must be fulfilled. It is predicated on the provision of large and superior data by ST, DOSM and PETRA. The models cannot be correctly trained or validated without this data. Moreover, the endeavor is dependent on the availability of sufficient computational resources for the development and testing of the hybrid models. Complex algorithms and large datasets often require significant computing power to be handled by advanced machine learning techniques. These presumptions are important because they directly affect the accuracy and viability of the forecasting model. The availability of these data and resources will be essential to the project's success.

The project also makes the assumption that hybrid models are better than independent ones. By combining the best features of several approaches, hybrid models have the ability to identify and represent deeper correlations and patterns in the data. Hybrid techniques, which combine the complementary strengths of multiple models, can frequently produce forecasts that are more reliable and accurate than those of any one model alone. This benefit is especially significant in situations when the data is complicated and shows both linear and non-linear correlations.

On the other hand, the accuracy and completeness of the datasets are equally crucial. No matter how complex the model is, unfit or inaccurate data might result in large predicting mistakes. Data issues that affect training, such missing values, outliers, or biases, might create models that are not very good at generalizing to new data. For this reason, guaranteeing high-quality datasets is essential to reducing errors and enhancing the forecasting models' dependability.

In conclusion, even though hybrid models combine several modeling approaches to potentially increase prediction accuracy, the project's success depends on the availability of high-quality data from ST, DOSM and PETRA as well as enough processing power. Any compromise in these domains may result in inaccurate projections and harm the project's goals.

# **3.0** **REQUIREMENT**

## 3.1 Research Objectives:

The primary objectives of this research are to develop a sectoral energy usage and demand forecasting model using hybrid approach, evaluate the performance of the developed model, analyze the impact of various factors on sectoral energy demand, and implement and evaluate energy-saving and environmental impact strategies. By achieving these objectives, the project aims to contribute to more efficient energy management practices, and provide valuable insights into energy demand patterns. The development and evaluation of the forecasting model will help determine the best approaches to predicting energy demand accurately.

## 3.2 Hypotheses:

The hypotheses for this research are twofold. First, it is hypothesized that hybrid models will provide more accurate energy demand forecasts than single-method models. By combining different machine learning techniques, hybrid models can leverage the strengths of each method, resulting in more robust and precise predictions.

Second, it is hypothesized that different factors, such as economic growth and climate change, significantly impact energy demand in various sectors. Understanding these factors' influence will help tailor the forecasting models to account for specific sectoral needs and variations.

## 3.3 Experimental Design:

The experimental design will employ a hybrid approach that combines machine learning algorithms with domain knowledge integration. This approach aims to create robust and interpretable forecasting models. The methodology will involve data collection, model development, and evaluation phases. During the data collection phase, data from ST, DOSM, PETRA, and other reliable sources will be gathered. The model development phase will involve training hybrid models using this data, and the evaluation phase will assess the models' performance using various metrics to ensure their accuracy and reliability. To put into perspective, the several technique in AI is summarized as below:

**Table 1: Comparison of AI Techniques**

| Feature | SVR | ANN | PSO | MLR |
| --- | --- | --- | --- | --- |
| Full Form | Support Vector Regression | Artificial Neural Network | Particle Swarm Optimization | Multiple Linear Regression |
| Type | Regression | Machine Learning | Optimization Algorithm | Regression |
| Use case | Regression tasks | Classification and regression | Optimization problems | Regression tasks |
| Algorithm Complexity | Moderate | High | Moderate | Low |
| Training Time | Moderate | High | High(if used in conjunction) | Low |
| Model Interpretability | Moderate | Low | Not applicable | High |
| Handling Non-linearity | Good | Excellent | Depends on objective function | Poor |
| Scalability | Moderate | High | High | High |
| Parameter Tuning | Complex (kernel selection) | Complex (architecture tuning) | Complex (parameter tuning) | Simple |
| Overfitting Tendency | Moderate (depends on C, epsilon) | High (depends on architecture) | Depends on objective function | High |
| Data Requirements | Moderate | High | Depends on the problem | Low |
| Computational Resources | Moderate | High | Moderate | Low |
| Implementation | Moderate | Complex | Moderate | Simple |
| Convergence | Guaranteed (convex problems) | Not guaranteed (depends on network) | Not guaranteed (stochastic nature) | Guaranteed |
| Common Applications | Financial modeling, forecasting | Image recognition, NLP | Function optimization | Economic forecasting, data analysis |
| Strengths | Effective in high-dimensional spaces | Can model complex patterns | Efficient global search | Simplicity, ease of interpretation |
| Weakness | Sensitive to kernel choice | Requires large datasets, overfitting | May converge to local minima | Assumes linear relationships |

Data preprocessing is an essential phase in the experimental design, as it ensures the data is in an appropriate form for analysis and modeling. In this step, missing value handling, feature selection, and normalization are all part of converting raw data into a clear and useable format. The performance and accuracy of the forecasting models are improved when the input data is of higher quality, which is achieved by proper data preparation.

Preprocessing data is crucial for a number of reasons. By standardizing and normalizing the data, it first greatly enhances the performance of the model by guaranteeing that every feature contributes equally to the model's performance and avoiding any bias towards characteristics with bigger scales. This is especially crucial for distance-based algorithms that can affect the outcomes, like SVM and K-means.

Second, by streamlining the data, data preprocessing shortens computation times, accelerating training and improving model performance. Preprocessing also improves the quality of the data by spotting and fixing mistakes or inconsistencies, which helps to avoid misleading results and guarantees the accuracy of the forecasts.

Moreover, methods like imputation efficiently manage absent values, whereas feature selection concentrates on the most pertinent features, enhancing model precision and diminishing overfitting. Through the implementation of extensive preprocessing techniques, the project guarantees that the dataset is of superior quality and appropriate for constructing resilient forecasting models.

**Table 2: Comparison of Preprocessing Technique.**

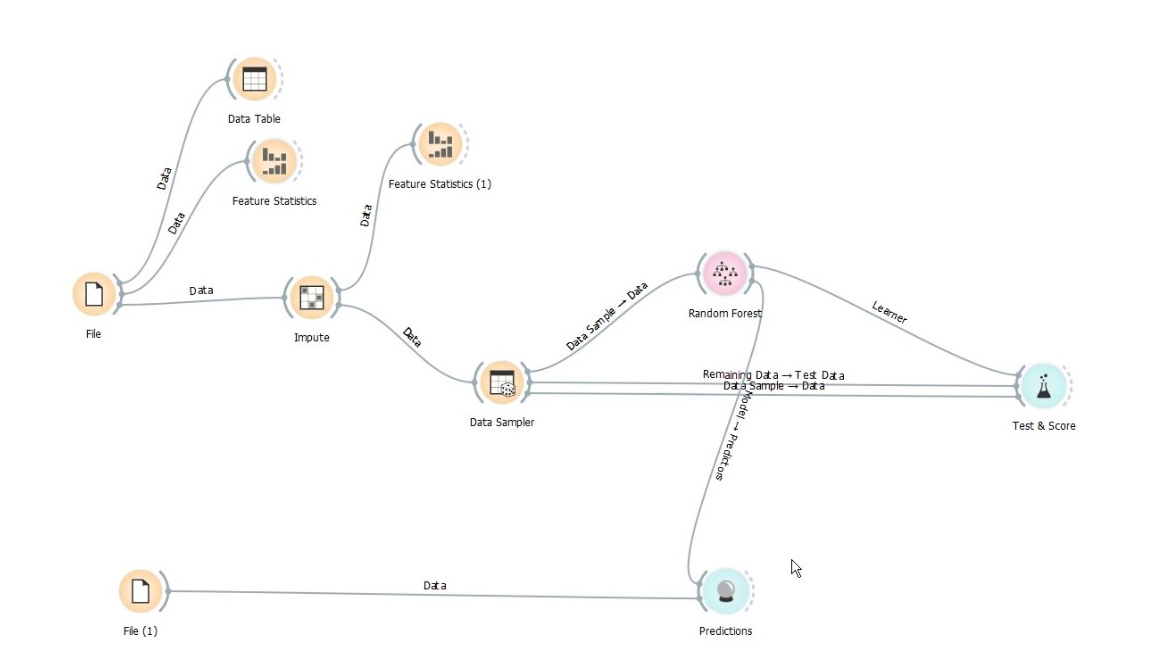
| Feature | Description | Importance |
| --- | --- | --- |
| Normalization | Adjusting the range of data values to a standard scale, typically [0,1] or [-1,1] | Essential for algorithms that rely on distance calculations, like SVM and K-means |
| Standardization | Transforming data to have a mean of 0 and a standard deviation of 1 | Important for algorithm for algorithms assuming normally distributed data |
| Imputation | Filling in missing values using strategies like mean, median, or mode | Critical for handling incomplete datasets |
| Discretization | Converting continuous features into discrete bins | Useful for algorithms requiring categorical inputs |
| Feature Selection | Selecting the most relevant features for the model | Improves model accuracy and reduces overfitting |

In this project, normalization was selected as the most suitable method because it ensures every characteristic contributes equally to the model’s performance, particularly for algorithms sensitive to feature scales.

Errors and inconsistencies in the dataset must be eliminated or corrected as part of data cleaning. Ensuring the quality and integrity of the data used for modeling requires taking this crucial step.

**Table 3: Comparison of Cleaning Technique.**

| Technique | Description | Importance |
| --- | --- | --- |
| Missing Value Treatment | Replacing missing values with statistical measures (mean, median) or using algorithms to predict missing values | Ensures completeness of the dataset |
| Outlier Detection | Identifying and handling outliers that may distort the model | Prevents skewed results and improves model robustness |
| Consistency Checking | Ensuring all data entries follow the same format and rules | Maintains data integrity |
| Duplicate Removal | Identifying and removing duplicate entries | Avoids redundancy and bias in the dataset |



**Figure 1: Experimental Design**

This project will use ORANGE, a data mining and machine learning tool, to conduct critical tasks like data cleaning, preprocessing, and model construction. ORANGE's user-friendly interface enables users to design workflows using a drag-and-drop method, making data analysis simple and efficient.

The presented graphic illustrates a sample workflow that demonstrates the steps we could take. This workflow consists of importing the dataset using the "File" widget, visualizing the data in a tabular fashion using the "Data Table" widget, and evaluating the feature statistics using the "Feature Statistics" widget. Missing values are handled via the "Impute" widget to guarantee that the data is clean. The "Data Sampler" widget then divides the dataset into training and testing sets. A "Random Forest" model is created for prediction, evaluated using the "Test & Score" widget, and then applied to new data to generate predictions using the "Predictions" widget.

This sample workflow was inspired by Saif Kabir Asif's lesson. It acts as a framework for future adjustments and refinements based on the project's individual requirements and subtleties. Orange Data Mining Examples provides more thorough examples and help on implementing similar workflows. These materials include detailed instructions and examples to help users use ORANGE for data analysis and machine learning activities (Saif, 2019).

## 3.4 Variables:

The variables in this research will include independent variables such as economic growth, population growth, and climate change factors. These variables are expected to influence energy demand across different sectors. The dependent variable will be energy demand, which the forecasting models aim to predict accurately. Control variables will include data quality and model parameters, ensuring that the models are trained and evaluated under consistent and controlled conditions to produce reliable results.

## 3.5 Sampling Strategy:

A well-structured sampling method is essential for reducing risks and ensuring the accuracy of forecasting models. Various sampling procedures can be used to ensure comprehensive and representative data collecting. Stratified sampling divides a population into different subgroups or strata that have comparable characteristics. For example, when projecting energy demand, different sectors (industrial, residential, and commercial) can be treated as separate strata. By ensuring that each stratum is appropriately represented, this strategy decreases sampling bias and boosts forecast accuracy. However, this strategy requires accurate demographic statistics for proper stratification.

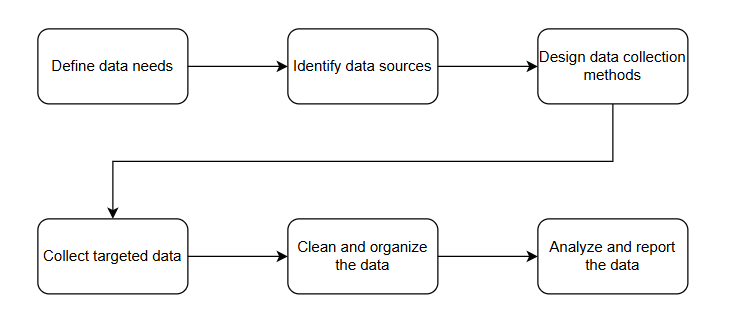
Systematic sampling involves selecting elements at regular intervals from an ordered sampling frame. For example, if we had annual energy usage data from 1978 to 2021, we may choose the fifth year for in-depth examination. This method is simple and straightforward to implement, but it runs into the risk of periodicity, which may bias the sample if the population exhibits periodic patterns. Cluster sampling divides the population into clusters and picks some of them at random for further investigation. This method is especially beneficial when the population is vast and spread, such as treating regions or states in Malaysia as clusters. While cost-effective and logistically practical, this strategy may produce a greater sample error if clusters are not similar.

Random sampling ensures that every member of the population has a fair chance of being chosen. This strategy reduces bias and is simple to implement, but it may not be practicable for large populations without sufficient resources. Using historical data from the Suruhanjaya Tenaga (ST), Department of Statistics Malaysia (DOSM) and Kementerian Peralihan Tenaga dan Transformasi Air (PETRA) between 1978 and 2021 guarantees that the forecasting models are trained and validated on a large dataset. This data set includes numerous economic indicators, climatic data, and sectoral energy consumption statistics, providing a solid platform for study.

The most important factors for using this data include assuring data accuracy and quality, covering all relevant sectors and variables across the whole period, and using the most recent data available to capture current trends and patterns. In addition to ST, DOSM and PETRA, adding data from additional reputable sources can improve the predictability of forecasting models. These sources may include International Energy Agency (IEA) studies on global energy trends and comparative analysis, World Bank data on economic indicators and development statistics, and meteorological departments that provide detailed climate and weather data on the consumption of energy.

Implementing a planned sampling method is critical for reducing risks and increasing the accuracy of energy demand projections. It can create solid forecasting models by integrating stratified, systematic, cluster, and random sampling methods, as well as using complete historical data from ST, DOSM, PETRA, and other credible sources. These models will be not only accurate, but also adaptive to diverse sector-specific features, ensuring broad application and reliability in anticipating energy demand trends.

## 3.6 Data Collection:



**Figure 2: Data Collection Procedures**

To effectively collect data from multiple sources, start by defining clear objectives to guide the process, ensuring the data gathered aligns with the goals. Next, identify reliable and relevant sources such as databases, websites, and surveys. Develop a detailed plan, including timelines, responsibilities, and methods for data collection. Begin gathering data while adhering to ethical guidelines and maintaining consistency in methods. Organize the collected data in a structured format, using tools like spreadsheets or databases. Finally, clean the data by checking for inconsistencies, errors, or missing values to ensure its quality and reliability (How Do You Collect Data From Multiple Sources?, 2023).

## 3.7 Data Analysis Plan:

The data analysis plan will include statistical and machine learning techniques for analyzing the acquired data and evaluating the forecasting model's effectiveness. This project will use ORANGE, a user-friendly data mining and machine learning software, as its primary data analysis tool. ORANGE's intuitive drag-and-drop interface enables rapid data cleaning, preprocessing, and model development, making it easy for users to design workflows and conduct difficult data analytic tasks.

Performance measures are critical for determining the accuracy and robustness of forecasting models. They give a quantitative foundation for comparing many models and choosing the optimal one for execution. This project will take advantage of several important performance indicators, including R-squared (R²), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These measures help in determining how well models capture variability in data and their predicted accuracy.

Table 4 presents the comparison between these performance metrics:

**Table 4: Comparison of Performance Metrics**

| Metric | Description | Importance | Best for Project |
| --- | --- | --- | --- |
| R2 | Measures the proportion of the variance in the dependent variable that is predictable from the independent variables. | Indicates the goodness of fit. A higher R² indicates a better fit of the model to the data. | Useful for understanding the overall fit of the model. |
| MSE | Measures the average of the squares of the errors or deviations, i.e., the difference between the estimator and the estimated. | Penalizes larger errors more significantly. Lower MSE indicates better model performance. | Important for assessing model accuracy. |
| RMSE | The square root of the MSE, providing error magnitude in the same units as the original data. | Penalizes larger errors more significantly. Lower MSE indicates better model performance. | Good for understanding model performance in practical terms. |
| MAE | Measures the average magnitude of the errors in a set of predictions, without considering their direction. | Provides a straightforward measure of prediction accuracy. Lower MAE indicates better performance. | Useful for understanding the average error magnitude. |

In the context of this project, RMSE is probably the ideal metric to use because it provides an error measure in the same units as the original data, making it easier to grasp the practical implications of forecasting errors. RMSE is also sensitive to big mistakes, which is critical for ensuring that the forecasting model does not ignore substantial deviations.

The study's goal is to obtain actionable insights and assess the usefulness of hybrid forecasting models by using these performance criteria. The detailed analysis facilitated by ORANGE will allow for the detection of relationships and trends within the data, ensuring that the models developed are accurate and adaptive. This approach will help to improve energy management practices and provide useful insights into energy consumption patterns across industries.

## 3.8 Ethical Considerations:

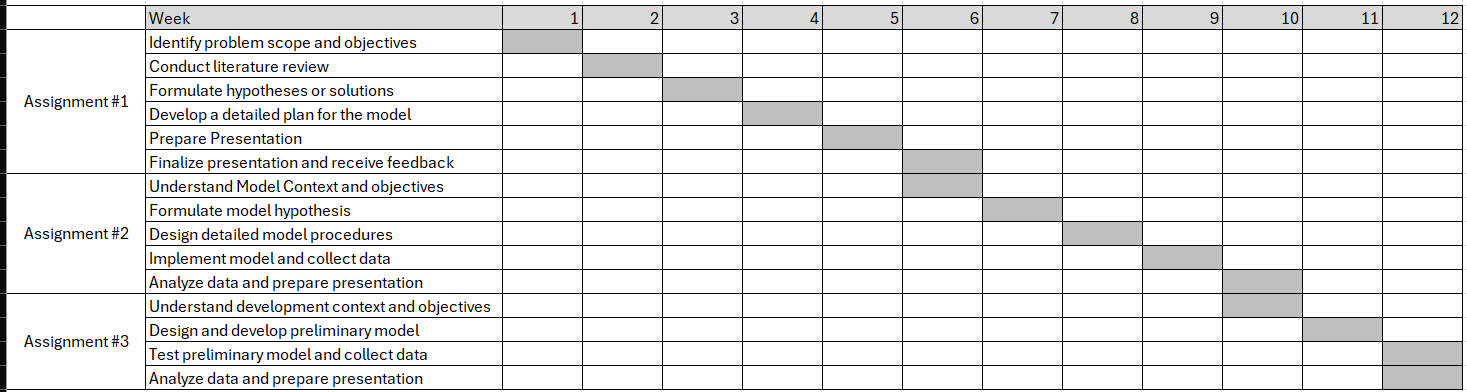
Ethical issues will be prioritized throughout the study process. Ensuring data privacy and security will be a top priority, protecting the confidentiality of any sensitive information used in the study. Data usage approvals will be obtained from relevant authorities to ensure compliance with legal and ethical norms. Furthermore, the study will address any potential harm caused by the findings, ensuring that they are used properly and ethically to benefit society.

The majority of this project's data sources will be public datasets from the Suruhanjaya Tenaga (ST), Department of Statistics Malaysia (DOSM) and the Kementerian Peralihan Tenaga dan Transformasi Air (PETRA). Using public datasets helps to ensure that the research is transparent and sustainable. Because the data is already publicly available, there are fewer issues regarding data privacy and permission, as long as it is utilized responsibly and in compliance with any applicable terms of service.

Furthermore, this study does not use animals or living beings, which eliminates any ethical concerns about animal welfare or human subject research. This avoids concerns about ethical issues that are often related with trials using living subjects.

To contribute to greater transparency and ethical research practices, the workflow of the ORANGE model utilized in this project, including methodologies, processes, and findings, will be made public. Sharing the workflow and results allows other researchers and stakeholders to verify, reproduce, and build on the findings, contributing to the larger scientific community and encouraging the practice of open and ethical research.

## 3.9 Timeline:



**Figure 3: Project Timeframe**

This project has a 12-week timeframe that is organized into three primary assignments, each with specified duties and goals to ensure systematic development and completion. To lay the foundation for the research, the first two weeks will be spent outlining the problem scope and objectives, as well as completing a thorough literature review. By week three, hypotheses or solutions will be developed based on the initial results and ideas from the literature review. In week four, a formal plan for the model will be created, outlining the approaches and technologies that will be used. Week five will be spent creating a presentation that summarizes the findings, proposed techniques, and project plan, which will be finalized in week six after getting comments from supervisors or stakeholders.

For Assignment #2, the seventh week will be dedicated to thoroughly understanding the model context and objectives, as well as developing a precise model hypothesis. By week eight, specific model processes will be developed, outlining the steps and methodologies for model implementation. The ninth week will be spent implementing the model and gathering the necessary data from designated sources. In week ten, the collected data will be examined, and a presentation will be developed to showcase the early findings and model performance.

In Assignment #3, week eleven will center on understanding the expansion context and targets, followed by developing and establishing a test model based on the data analysis. The final week, week twelve, will be spent testing the preliminary model, gathering more data as needed, and analyzing the results to create a final presentation. This planned calendar ensures that each phase of the project receives appropriate attention and resources, allowing for a complete and systematic approach to creating, testing, and validating the energy demand forecasting model. By sticking to this plan, the project hopes to complete its objectives within the timeframe specified while maintaining high standards of research and analysis.

## 3.10 System Requirements

The successful execution of this project requires specific hardware and software components to handle the complex tasks associated with data collection, preprocessing, model development, and analysis.

**Hardware**

The primary hardware to be used includes a high-performance laptop or desktop computer with the following specifications:

Device name KaleidoHeart

Processor Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.60 GHz

Installed RAM 16.0 GB (15.8 GB usable)

System type 64-bit operating system, x64-based processor

Computer ACPI x64-based PC

Micro-Star International Co,. Ltd GF65 Thin 10UE

Disk Driver KINGSTON OM8CP3512F-AI1

Display adapters INTEL(R) UHD Graphics

NVIDIA GeForce RTX 3060 Laptop GPU

Network adapters Intel(R) Wi-Fi 6 AX201 160MHz

Phantom TAP-Windows Adapter V9

Realtek PCIe GbE Family Controller

These specs ensure that the computational resources are sufficient for managing the project's advanced algorithms and massive datasets. The combination of a strong CPU and enough RAM allows for efficient processing, while the sophisticated GPU capabilities meet the graphics requirements of data visualization and machine learning workloads.

**Software**

For software, the project will use one essential tools and platforms to support data analysis and model development:

Orange: A complete data mining and machine learning tool with a simple drag-and-drop interface. Orange will be used for important activities including data cleaning, preprocessing, and model creation, making complicated data analysis easier and more efficient.

# 

# 

# 

# 

# 

# 4.0 Reference

1. Fontes, X., & Silva, D. C. (2019). Hybrid approaches for time series prediction. In *Advances in intelligent systems and computing* (pp. 146–155). https://doi.org/10.1007/978-3-030-14347-3\_15
2. *How do you collect data from multiple sources?* (2023, August 31). www.linkedin.com. https://www.linkedin.com/advice/0/how-do-you-collect-data-from-multiple-sources
3. Ladhari, A., & Boubaker, H. (2024). Deep Learning Models for Bitcoin Prediction Using Hybrid Approaches with Gradient-Specific Optimization. *Forecasting*, *6*(2), 279–295. https://doi.org/10.3390/forecast6020016
4. Maiyza, A. I., Korany, N. O., Banawan, K., Hassan, H. A., & Sheta, W. M. (2023). VTGAN: hybrid generative adversarial networks for cloud workload prediction. *Journal of Cloud Computing*, *12*(1). https://doi.org/10.1186/s13677-023-00473-z
5. Saif, K. A. (2019, May 25). *Exploring Orange : a beginner friendly data mining toolkit*. https://saifkabirasif.com/. Retrieved June 29, 2024, from https://saifkabirasif.com/2019/05/25/Orange3\_Tutorial/
6. Utku, A., & Can, U. (2023). An efficient hybrid weather prediction model based on deep learning. *International Journal of Environmental Science and Technology*, *20*(10), 11107–11120. https://doi.org/10.1007/s13762-023-05092-4

| **Declaration** |
| --- |

I hereby declare that this submission is **my own work** and to the best of my knowledge it contains no materials previously published or written by another.

|  |  |  |
| --- | --- | --- |
| Student’s Signature |  | Date |

| **Supervisor’s Approval** |
| --- |

| **X** | Approve without modification |
| --- | --- |
|  | Approve with modification |
|  | Reject |

Remark:

Abqari is showing good progress in his work. I am so far satisfied with his work both professionally, and individually.

Over the past few weeks, Abqari has shown his ability to work independently, and produce results which I strongly feel he is on the right track to completing this project by the 2nd half of next semester.

|  |  | **30 June 2024** |
| --- | --- | --- |
| Project Supervisor’s Signature & Stamp |  | Date |