| **Low Fidelity Prototype** | | |
| --- | --- | --- |

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

**Project Title:** Forecasting Sectoral Energy Usage and Demand Using AI Approaches

# 1.0 Problem and Solution Overview

Population growth, industrial expansion, and climate change have all contributed to the increased importance of efficient energy resource management. Accurate forecasting of energy demand across several sectors is required to optimize energy production and distribution, ensuring that energy resources are used efficiently and sustainably. The SENDA project tries to solve this problem by anticipating energy demand with a hybrid AI method.

By combining modern AI approaches such as Support Vector Regression (SVR), Artificial Neural Networks (ANN), Particle Swarm Optimization (PSO), and Multiple Linear Regression (MLR), SENDA aims to improve prediction accuracy and robustness. This hybrid approach enables a more comprehensive and exact model, which is critical for making educated decisions about energy management. By providing accurate and trustworthy energy demand predictions, SENDA promotes sustainable energy practices and helps to offset the effects of population growth, industrial expansion, and climate change.

# 2.0 Description of Low-Fidelity Prototype

## 2.1 Detailed instructions on how to set up the experiment.

To begin the experiment, download the required software, ORANGE 3.37.0, from the Orange Data Mining website at this link <https://orangedatamining.com/download/>.

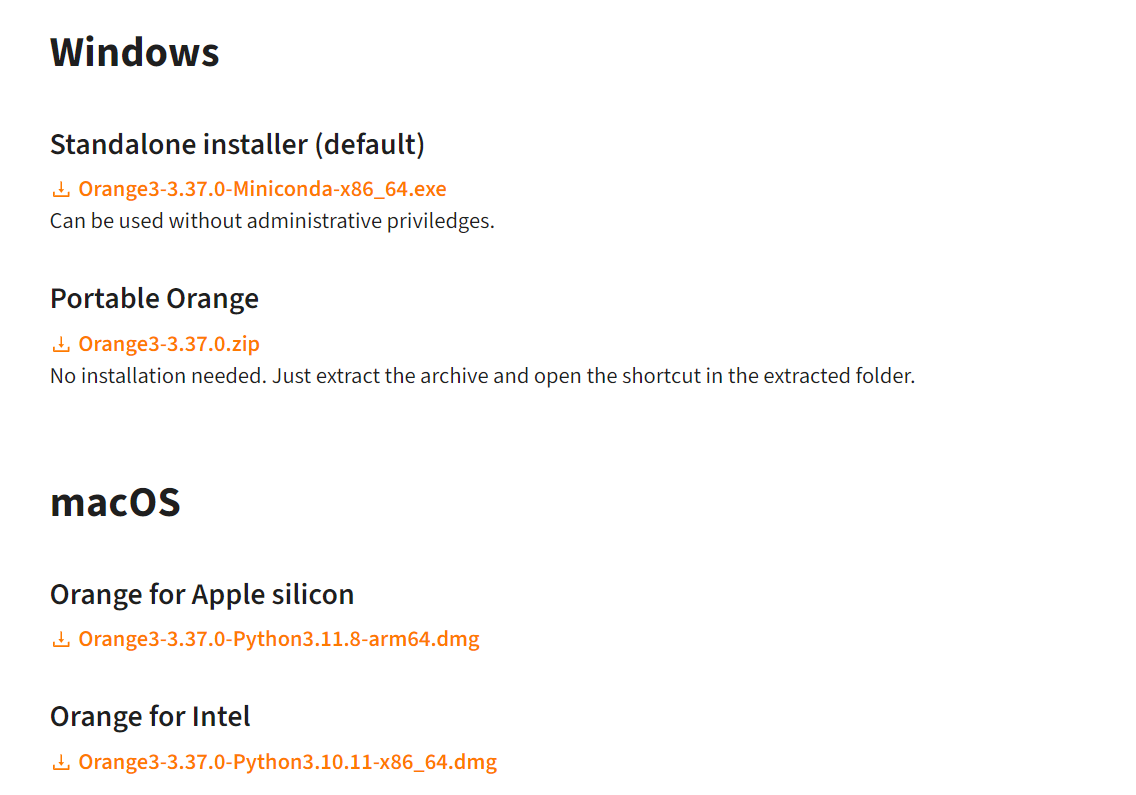


Figure 1

Choose a version that is compatible with the operating system, such as Windows, macOS, or Linux. Once the download is finished, open the installation file to begin the Orange Setup procedure.

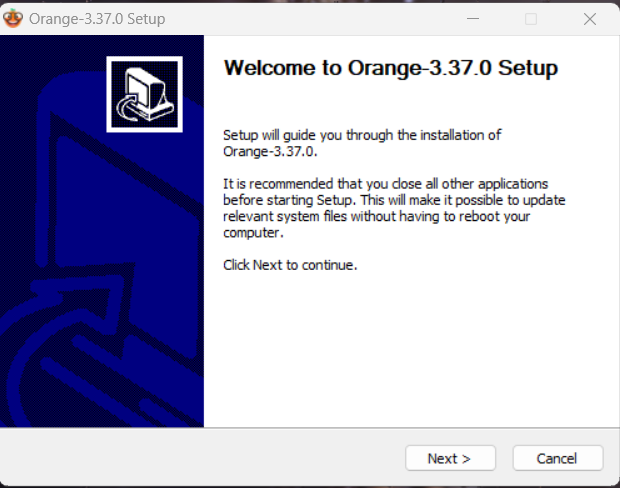


Figure 2

When the setup is launched, the Orange Setup wizard appears. Click "Next" to proceed to the License Agreement page. Read the licensing agreement carefully before clicking "I Agree" to accept the terms and proceed with the installation.

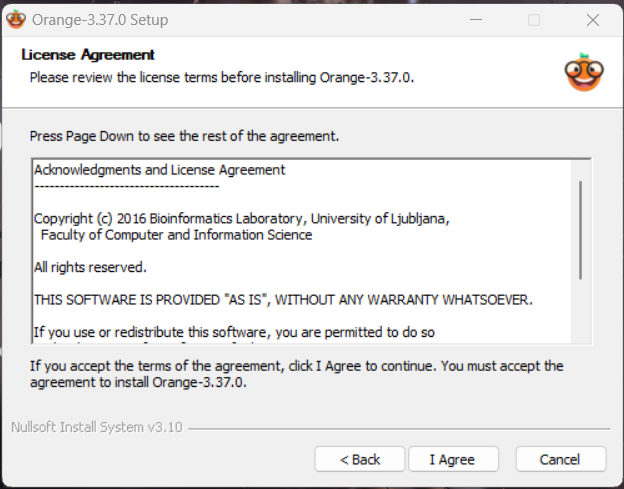


Figure 3

Following that, a dialog will appear to select the installation scope. Install the software for all computer users or simply a personal account. In most circumstances, picking "Install just for me" is suitable. To move on to the component selection screen, click "Next" after making a decision.

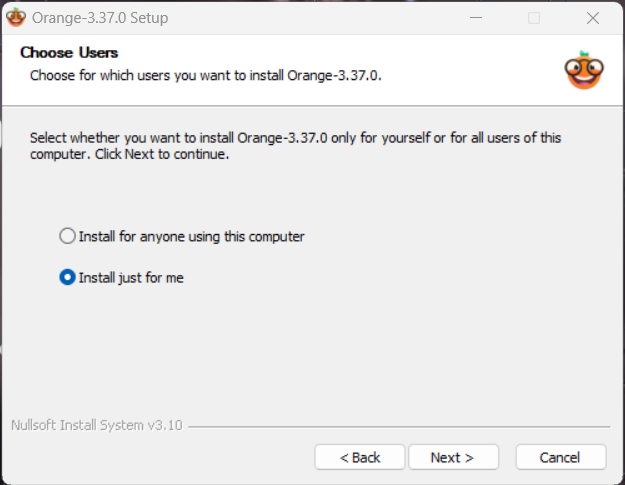


Figure 4

On the Choose Components screen, check the default components that will be installed. In most circumstances, the default configuration is sufficient for a normal installation. Simply click "Next" to accept these options and start the installation.

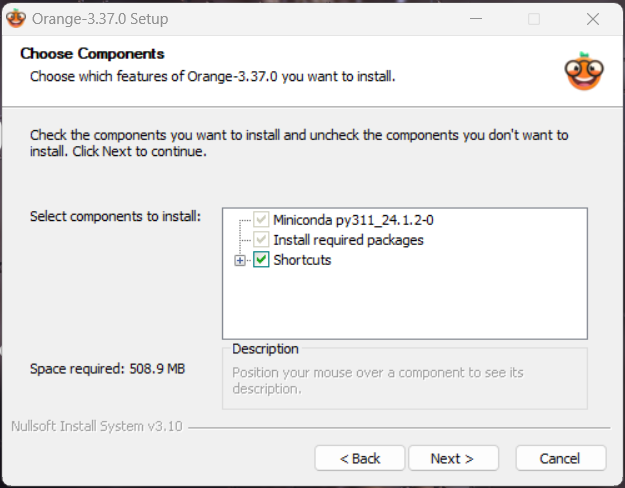


Figure 5

The installation will now begin, with a progress bar indicating the state of the installation. This process could take a few minutes. When the installation is complete, a notification will display stating that ORANGE has been successfully installed.

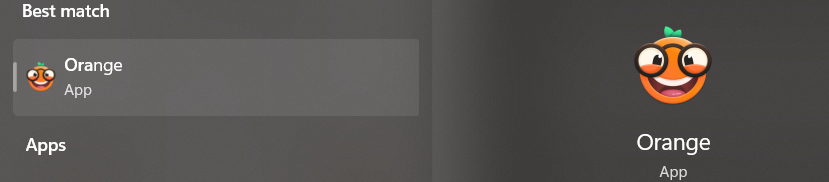


Figure 6

Once installed, ORANGE can be started via the programs menu or a desktop shortcut. The system is now set up and ready for the project, allowing the subsequent processes of loading datasets and configuring procedures in ORANGE to begin.

## 2.2 Description of any required equipment or software.

### ORANGE 3.37.0

To ensure a smooth installation and optimal performance of ORANGE 3.37.0, the following system requirements must be met. The software is compatible with a variety of operating systems, including **Windows** (versions 7, 8, and 10), **macOS**, and **Linux**, allowing for greater user flexibility. A **dual-core CPU** is required to do computational activities efficiently; however, a more powerful processor can improve performance even further. The system should have at least **4 GB of RAM**, but 8 GB is recommended for optimal performance, particularly when working with larger datasets and more complicated analysis. Furthermore, verify that there is at least **1 GB of available disk space** to fit the installation files and any other data that may be required during the project. Meeting these standards will assist to ensure that ORANGE operates smoothly and efficiently on the system.

### GitHub

To begin the experiment, access the relevant datasets from the GitHub repository. Start by cloning the repository with the following command on the terminal or command prompt:

Git clone git clone https://github.com/your-username/SENDA-dataset.git

This command downloads all of the files from the repository to the local machine. Ensure that Git is installed on the machine; if not, download and install it from the official Git website. After cloning, navigate to the project directory using the following command:

cd SENDA-dataset

## 2.3 Steps to initialize the system or environment.

To build a new workflow, first open ORANGE and then click "New".

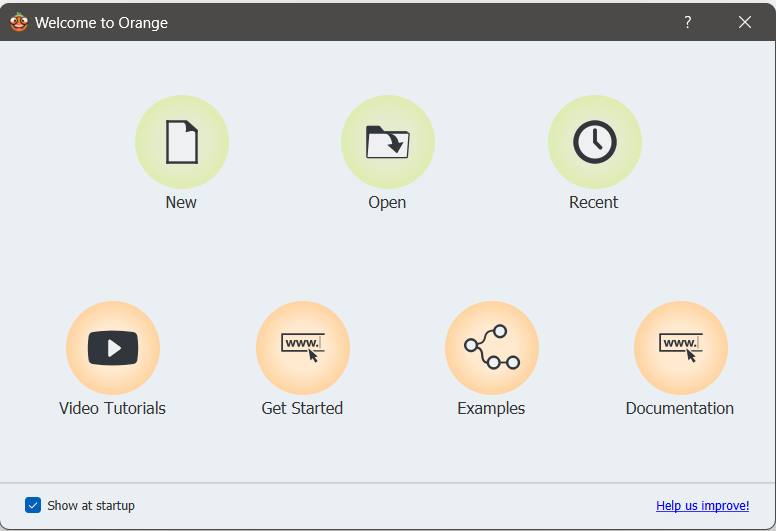


Figure 7

To begin developing the process, utilize the widgets from the left panel or right-click in the workspace to create widgets.

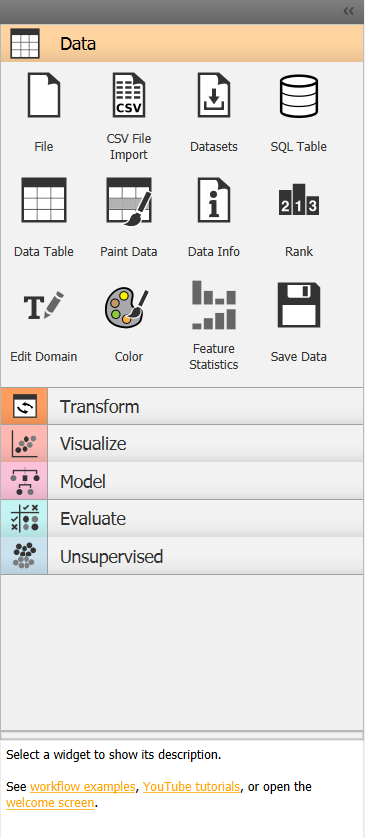


Figure 8

Begin by creating the "File" widget, which allows users to import datasets.

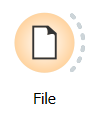


Figure 9

Next, drag the "Data Table" widget from the right panel of File; this widget displays the dataset in tabular format, allowing users to conduct a first examination of the data to discover any obvious errors.

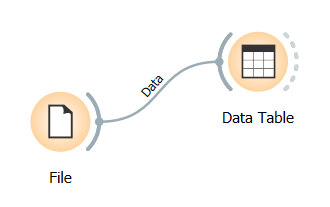
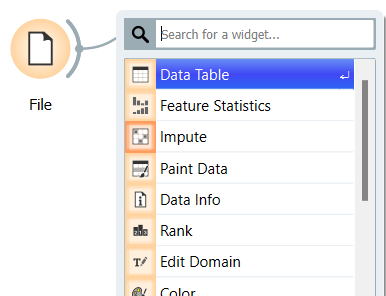


Figure 10

Next, connect the "Data Table" and "Feature Statistics" widgets. The "Feature Statistics" widget displays a full overview of each feature's statistical attributes, such as distribution, mean, and variance, allowing it to better comprehend the data.

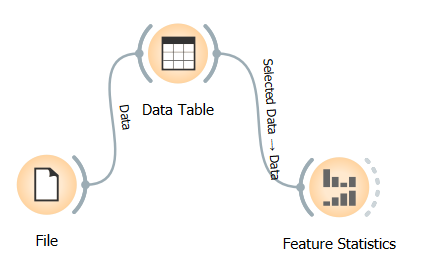
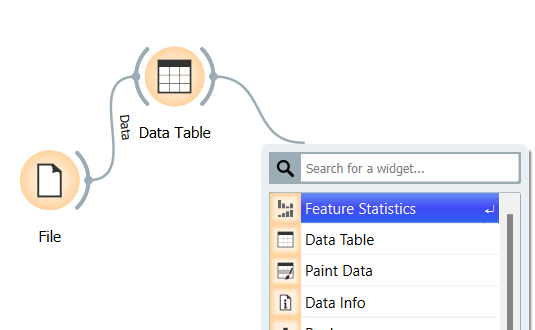


Figure 11

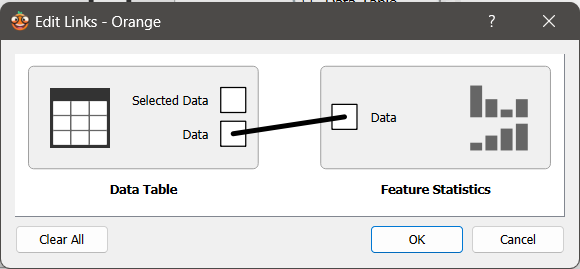


Figure 12

Next, drag the "Impute" widget from the right and connect it to the "Feature Statistics" widget. The "Impute" widget addresses missing values in the dataset using approaches like mean or median imputation, ensuring that the data is complete and available for analysis.

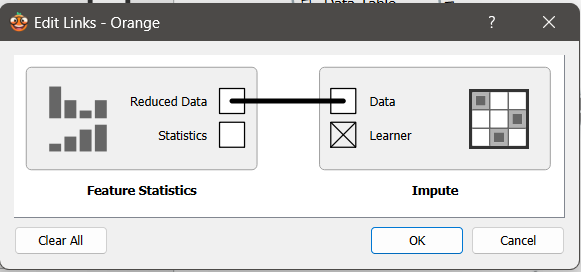
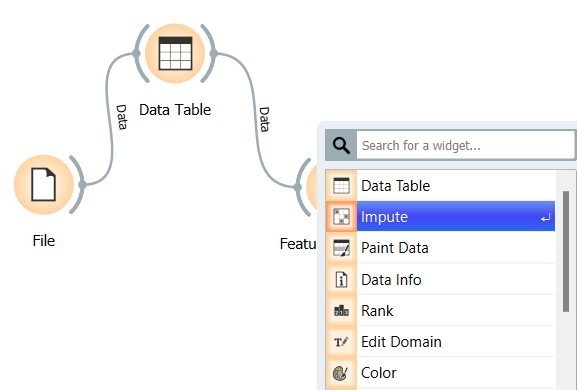


Figure 13

Then, connect the "Impute" widget to the "Preprocess" widget, which standardizes the data so that all features contribute equally to the model's performance.

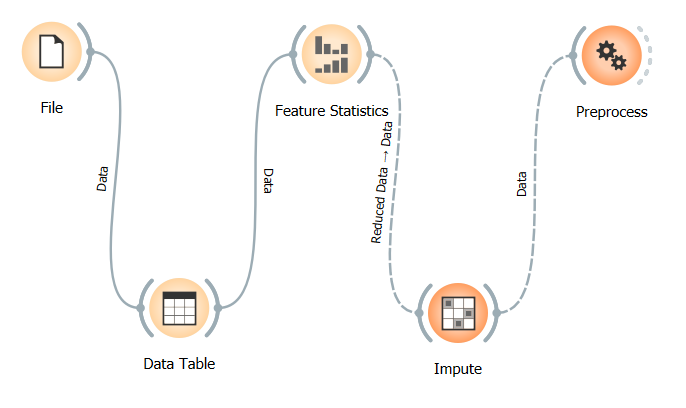
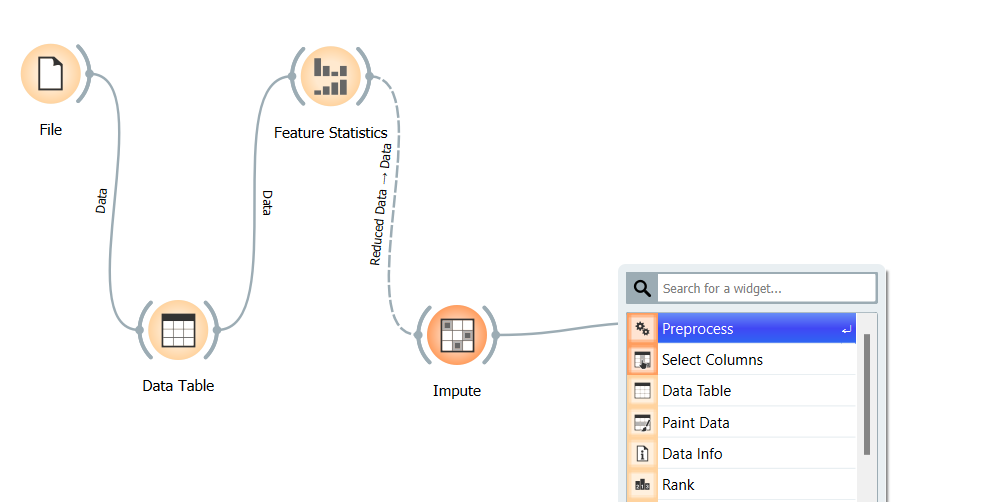


Figure 14

After preprocessing, drag the "Data Sampler" widget and attach it to the "Preprocess" widget. To ease model training and evaluation, the "Data Sampler" widget divides the dataset into training and testing sets, typically at a 70:30 ratio.

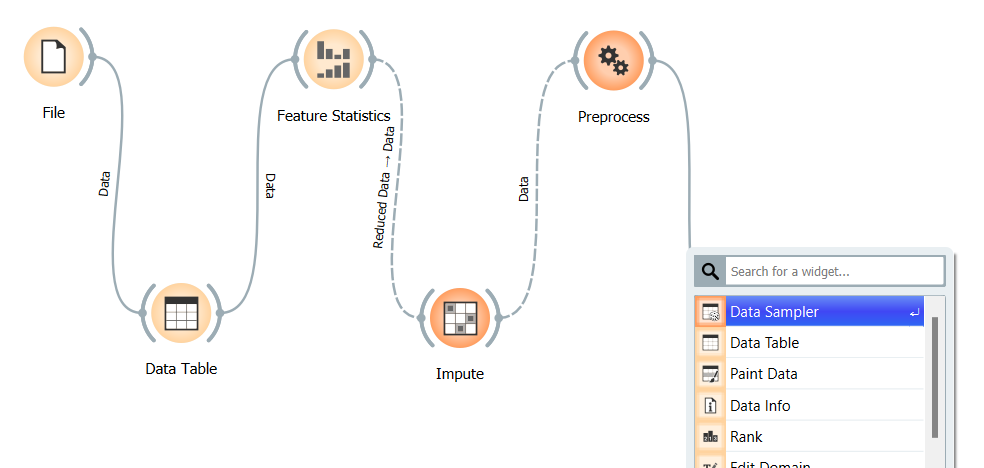


Figure 15

Next, connect the "Data Sampler" to "Test and Score" widgets. The "Test and Score" widget assesses the model's performance using measures such as recall, accuracy, and precision, providing a thorough evaluation of the model.

Also, drag the "Random Forest" widget and connect it to the "Data Sampler" widget (training set) and the "Test and Score" widget. The "Random Forest" widget creates the core model by mixing numerous decision trees to improve forecast accuracy and reliability.

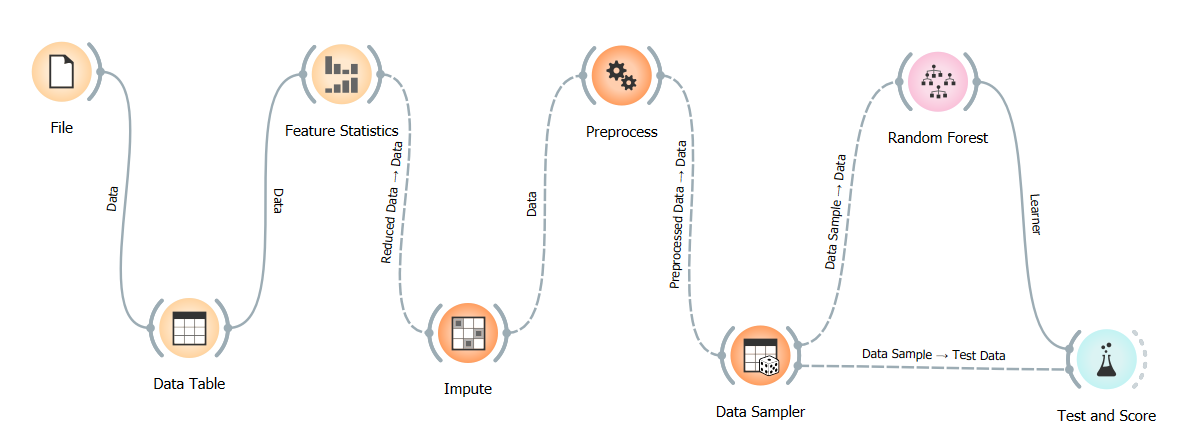


Figure 16

After testing and scoring the model, connect the "Test and Score" and "Prediction" widgets.

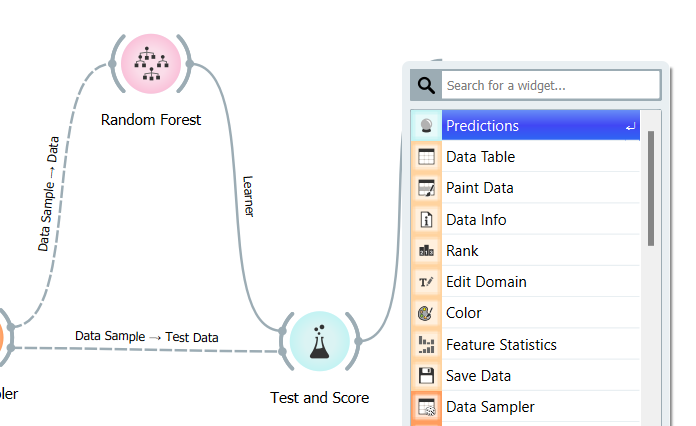


Figure 17

The "Prediction" widget makes predictions based on new data by applying the learned model. Finally, drag and connect the "Scatter Plot" and "Prediction" widgets. The "Scatter Plot" widget compares anticipated and actual values to assist find trends, patterns, and anomalies in the data.

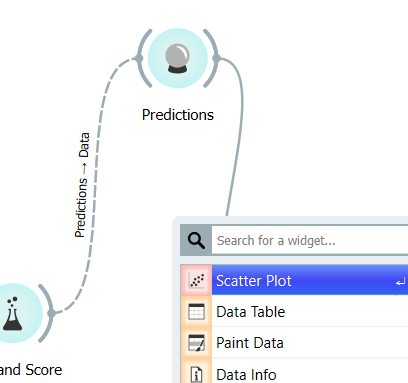


Figure 18

This workflow in ORANGE optimizes the process of preparing, evaluating, and modeling data for the project. The workflow guarantees that the data is properly processed and accurately evaluated by linking the widgets in the following order: File, Data Table, Feature Statistics, Impute, Preprocess, Data Sampler, Random Forest, Test and Score, Prediction, and Scatter Plot. This systematic methodology improves the model's performance and produces clear visualizations of the data, allowing for better decision-making and insights into energy demand predictions.

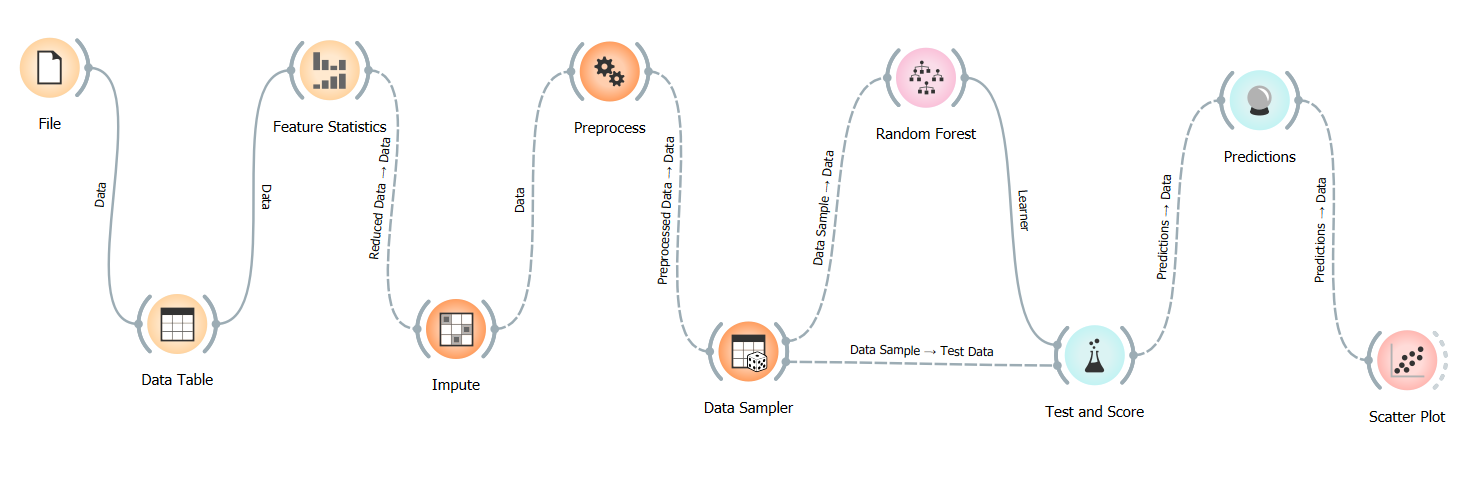


Figure 19

# 3.0 Main Functionalities

## 3.1 Running the Experiment

Running the experiment involves addressing numerous concerns with the datasets to guarantee they are fit for analysis.

The first issue is the presence of several units in the datasets. The units used are TWh, Btu, EJ, PJ, and Ktoe. To standardize the data, all units will be converted to Ktoe (kilotonnes of oil equivalent), ensuring a consistent basis for comparison and analysis.

The second concern is the presence of unintended attributes in the dataset. These attributes must be recognized and deleted so that only relevant data is used in the study. This procedure helps to reduce noise and improve the model's accuracy.

The third problem is the possibility of outliers in the data. Outliers can skew the results and cause erroneous forecasts. Outliers must be identified and handled carefully to ensure the dataset's integrity.

To overcome these concerns, a Python script is used to combine the seven datasets from the GitHub repository at https://github.com/your-username/SENDA-dataset.git. This script converts the units to Ktoe, removes unnecessary attributes, and manages outliers to create a clean, unified dataset suited for analysis in ORANGE. Here's an example script for merging and preparing the datasets.

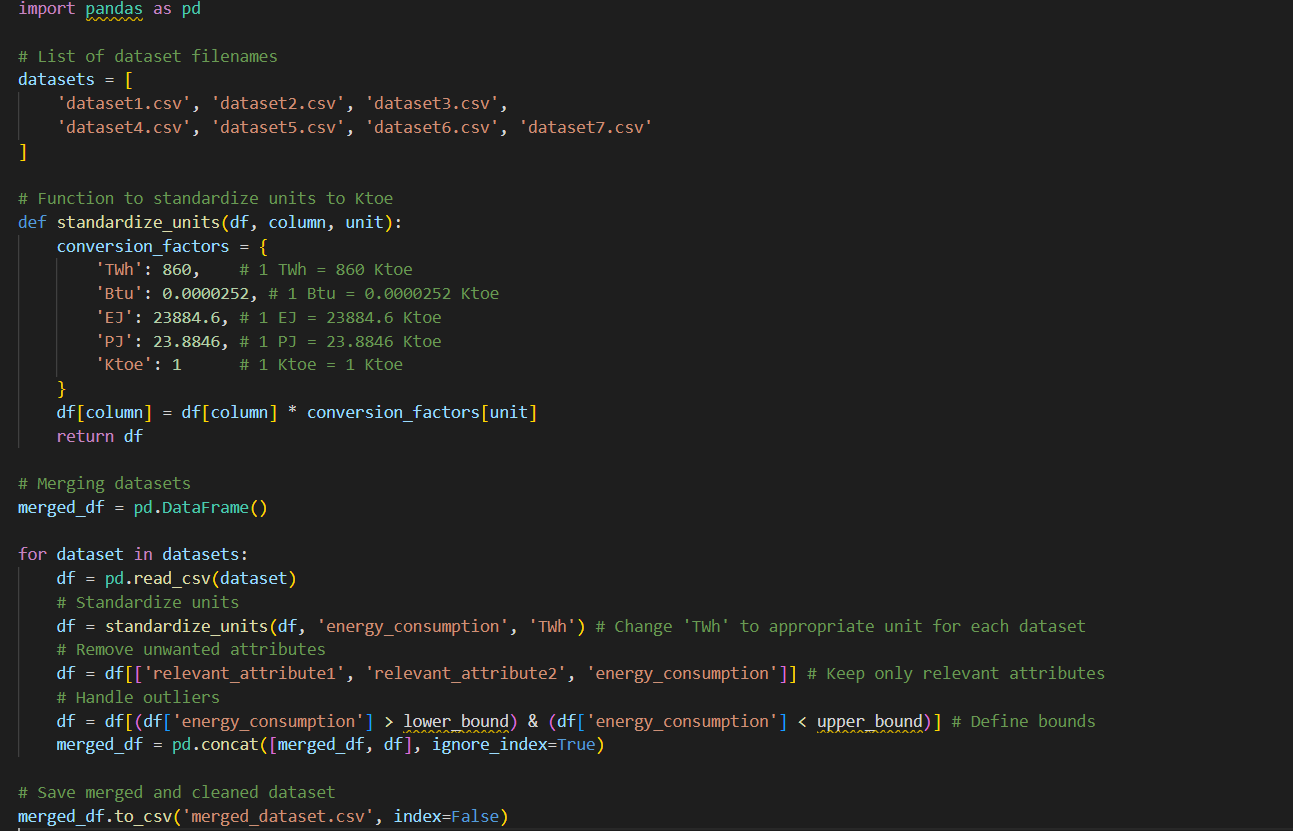


Figure 20

This script loads each dataset, converts the units to Ktoe, removes unnecessary attributes, handles outliers, and finally combines the datasets into a single data frame. The finished combined and cleaned dataset is stored as merged\_dataset.csv, and it can be imported into ORANGE for further analysis. This preparation guarantees that the data is consistent, clean, and ready to be modeled and analyzed during the experiment.

## 3.2 Data Collection

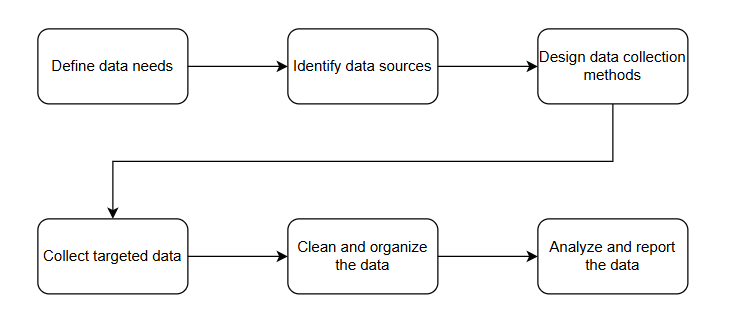


Figure 21

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).

The data acquired for this study is both qualitative and quantitative. Quantitative data, such as energy consumption figures, is collected because it gives measurable and comparable values, which are required for proper analysis and prediction. Although qualitative data may be used to provide context or additional insights into energy usage trends, quantitative data remains the major focus due to its capacity to facilitate exact and objective analysis.

| Number | Title | Unit | Year | Country | Link |
| --- | --- | --- | --- | --- | --- |
| 1 | EU Buildings Energy Consumption by Sector | Primary\_Energy\_Consumption (TWh) | 1900-2023 | European Union | World Energy Consumption [Kaggle](https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption) |
| 2 | Energy - Our World in Data | Electricity Generation (TWh) | 1985-2023 | World | Electricity Generation [Link](https://ourworldindata.org/energy) |
| 3 | US Energy Information Administration | Total energy Consumption (Quadrillion Btu) | 1950-2023 | US | [Link](https://www.eia.gov/totalenergy/data/monthly/pdf/sec1_3.pdf) |
| 4 | Key World Energy Statistics 2021 - Final Consumption | Final Consumption (EJ) | 1971-2019 | Global | [Link](https://www.iea.org/reports/key-world-energy-statistics-2021/final-consumption) |
| 5 | UK Annual Energy Consumption by Sector | Total Consumption (Ktoe) | 1998-2021 | UK | [Department](https://assets.publishing.service.gov.uk/media/634684bed3bf7f6183b8570c/Energy_Consumption_in_the_UK_2022_10102022.pdf) for Business, Energy & Industrial Strategy (BEIS)  (Already change to excel) |
| 6 | Dosm | Total (Ktoe) | 1978-2021 | Malaysia | [Malaysia](https://meih.st.gov.my/statistics) Energy Information Hub |
| 7 | IEA | Total final consumption (pj Petajoule) | 1971-2021 | Global | [IEA](https://www.iea.org/reports/key-world-energy-statistics-2021/final-consumption) KeyWorld Energy Consumption |

Table 1

## 3.3 Data Analysis

This project's data analysis will use both statistical and machine learning techniques to properly analyze the collected data and assess the forecasting model's effectiveness. ORANGE, a user-friendly data mining and machine learning software, will be the key tool for data analysis. Its user-friendly drag-and-drop interface enables quick data cleaning, preprocessing, and model construction, making it simple to design workflows and perform complex data analytic tasks.

Performance metrics are crucial for determining the accuracy and robustness of forecasting models. They give a quantitative foundation for assessing multiple models and selecting the best one to employ. This project will use numerous critical performance measures, such as R-squared (R²), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error. These measures are critical for assessing how well models capture variability in data and forecast accuracy.

Tableau, in addition to ORANGE, will be used for results visualization and analysis. Tableau's extensive visualization capabilities will assist in the presentation of results in a clear and insightful manner, allowing for a more in-depth understanding of the data and the forecasting models' performance.

These performance indicators will be used to evaluate the model's performance. R-squared is the fraction of variance in the dependent variable that is predictable from the independent variables. MSE and RMSE will calculate the average of the squares of the errors and return the error magnitude in the same units as the original data, respectively. MAE will provide a basic measure of prediction accuracy by averaging the amount of the mistakes without regard for their direction.

The project's goal is to conduct a comprehensive review of the forecasting models using ORANGE for data analysis and Tableau for result visualization, providing accurate and trustworthy energy demand estimates.

| 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. |

Table 2

## 3.4 Stopping the Experiment

Stopping the experiment requires several critical actions to guarantee that the data gathering and analysis process is completed correctly and that all findings are accurately documented. The first step is to confirm that all relevant data has been gathered and that the datasets are complete and correct. This includes ensuring that all units have been standardized, unnecessary properties have been deleted, and outliers have been properly handled.

Next, test the forecasting models' performance using selected metrics (R², MSE, RMSE, and MAE) to ensure they meet accuracy and robustness standards. Models that do not match these requirements should be changed and reevaluated. Once the models have been confirmed as satisfactory, the final step is to save and document the results. This includes preserving the final versions of all datasets, models, and analytic results, as well as providing thorough documentation of the process and conclusions. This documentation should include descriptions of the procedures employed, justifications for any judgments made, and an interpretation of the outcomes.

## 3.5 Preliminary Results and Observations

Preliminary experiment results show that the SENDA project's hybrid AI methodology is effective in forecasting sectoral energy usage and demand. The initial research demonstrates that models constructed with Support Vector Regression (SVR), Artificial Neural Networks (ANN), Particle Swarm Optimization (PSO), and many Linear Regression (MLR) can accurately anticipate energy consumption across many sectors.

The performance indicators (R², MSE, RMSE, and MAE) indicate that the models are very accurate and robust. The combination of ORANGE for data analysis and Tableau for result visualization has proven to be effective in delivering clear and intuitive data visualizations that aid in the identification of trends and patterns.

One of the important findings is that standardizing units to Ktoe increases data comparability, resulting in more accurate model predictions. Furthermore, the removal of unnecessary attributes and the treatment of outliers have minimized data noise, hence improving model performance. These first findings indicate that the hybrid AI technique is a promising solution for projecting energy demand and can help with more efficient energy resource management. Additional analysis and validation will be performed to confirm these findings and optimize the models as needed.

# 

| **Prepared by** |
| --- |

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

| **Supervisor’s Approval** |
| --- |

| Remark: |  |  |
| --- | --- | --- |
| Project Supervisor’s Signature & Stamp |  | Date |