**Integrative Demand Forecasting for Energy Consumption: A Machine Learning Approach with Seasonal Decomposition and Climate Change Impact Analysis in New South Wales (NSW)**

**Team Members:**

* **Tom Woodley (z5450185) - role**
* **Tariq Khan (z5414837) - role**
* **Eugene Ho (z5497345) - role**
* **Majuwana Kariyawasam (z5398970) - role**

**Date: 30/08/2024**

**Table of Contents**

0.1 Introduction and Motivation  
0.2 Brief Literature Review  
0.3 Methods, Software, and Data Description  
0.4 Activities and Schedule  
0.4.1 Main tasks and activities  
0.4.2 Activities and Schedule  
0.5 Prediction Models and Techniques  
1 References

**1 Introduction and Motivation**

Energy demand forecasting is crucial for balancing supply and demand in the energy market of New South Wales (NSW), ensuring that energy supply remains profitable and efficient. The volatility of energy demand, influenced by various factors including weather patterns, holidays, and daily or seasonal variations, presents a complex challenge for energy suppliers in NSW. Accurate demand prediction is essential for maintaining grid stability, reducing costs, and optimizing resource allocation.

Climate change further complicates this task by altering weather patterns, which in turn affects energy consumption habits in NSW. For instance, rising temperatures increase the use of air conditioning, while colder winters can lead to higher heating demand. Understanding these relationships is crucial for making informed decisions about energy production and distribution within the state.

This project aims to explore the intricate relationships between weather patterns, climate change, and energy demand specifically in NSW through the application of machine learning models. We will analyze how temperature, seasonal changes, and special events impact energy consumption in NSW and develop a predictive model that incorporates these variables. The outcome will be a robust forecasting tool that can support energy suppliers in NSW in optimizing their operations and planning for future demand.

**2 Brief Literature Review**

In recent years, numerous studies have explored the relationship between weather conditions and energy demand. For instance, studies have shown that temperature is a key driver of energy consumption, particularly in regions with extreme climates like NSW. Research has also highlighted the importance of accounting for seasonal variations and special events in demand forecasting models.

One significant study by Smith et al. [1] employed a multiple linear regression model to forecast energy demand, incorporating temperature and day-of-week variables which demonstrate the effectiveness of combining traditional statistical methods with machine learning approaches.

Our work builds on these foundations by integrating additional factors such as climate change projections, population growth, and air quality into our models. We also plan to use advanced machine learning algorithms, including Neural Networks and Gradient Boosting Machines, to enhance the accuracy and robustness of our forecasts specifically for NSW.

**3 Methods, Software, and Data Description**

**Methods**:

* **Time Series Decomposition**: To break down energy demand data into trend, seasonal, and residual components, allowing for a clearer understanding of underlying patterns specific to NSW.
* **Multiple Linear Regression**: To model the relationship between energy demand and key predictors such as temperature, population, air quality, and holiday effects in NSW.
* **Machine Learning Algorithms**: Including Random Forests, Gradient Boosting Machines, and Neural Networks, to capture non-linear relationships and interactions between variables in the NSW context.
* **Climate Change Impact Assessment**: To quantify the effects of climate variables on energy demand over time in NSW.

**Software and Libraries**:

* **Python**: For data processing, analysis, and model development.
* **Pandas**: For data manipulation and cleaning.
* **Matplotlib & Seaborn**: For data visualization.
* **Scikit-Learn**: For implementing machine learning models.
* **TensorFlow/Keras**: For developing Neural Networks.
* **Jupyter Notebooks**: For interactive data analysis, visualization, and report generation.

**Data Description**:

* **Energy Demand Data**: Collected at 5-minute intervals from 1 January 2010 to 1 August 2022, specific to NSW, and stored in CSV format.
* **Temperature Data**: Collected at 30-minute intervals for the same period, with necessary cleaning steps applied to remove invalid entries, and specific to NSW.
* **Meteorological Variables**: Additional weather-related data specific to NSW will be downloaded and integrated into the analysis to ensure comprehensive coverage of factors influencing energy demand:
  + **Air Temperature**: Even though temperature data is already available, obtaining this from a consistent source ensures uniformity across datasets.
  + **Relative Humidity**: This variable affects energy demand, particularly for cooling and heating in NSW.
  + **Global Solar Radiation**: Influences energy demand for cooling during daytime hours in NSW.
  + **Wind Speed**: Affects temperature perception and consequently heating/cooling needs in NSW.
  + **Rainfall**: Weather conditions like rainfall can significantly influence energy usage patterns in NSW.
* **Air Quality and Wind Data**: This data will be sourced from the NSW Air Quality Data Explorer available at [NSW Air Quality Data Services](https://www.airquality.nsw.gov.au/air-quality-data-services/data-explorer). This will include variables such as particulate matter (PM2.5, PM10), ozone levels, and wind speed/direction, which are relevant for understanding how air quality and wind patterns impact energy demand in NSW.
* **Population Projections**: Sourced from the Australian Bureau of Statistics (ABS), population projection data specific to NSW will be integrated to account for the impact of population growth on energy demand. This will be particularly relevant for long-term demand forecasting in NSW.

The data is stored in CSV format and requires significant preprocessing to handle missing values, outliers, and inconsistencies. The complexity of the data, with its high frequency and multiple influencing factors, makes it a suitable candidate for advanced machine learning techniques.

**4 Activities and Schedule**

**4.1 Main Tasks and Activities**

To achieve the project's objectives, the following tasks will be undertaken:

* **Literature Review**: Conduct an in-depth review of existing studies on energy demand forecasting, focusing on the use of weather, climate, and demographic variables in NSW.
* **Data Assessment**: Evaluate the provided datasets specific to NSW to ensure they meet the project’s requirements and identify any gaps that need addressing.
* **Algorithm Research**: Identify and select the most appropriate machine learning algorithms for the analysis specific to NSW.
* **Data Cleaning**: Remove irrelevant or incorrect data, such as invalid temperature readings, specific to NSW.
* **Data Enrichment**: Integrate additional datasets, including population projections, air quality data, and climate forecasts, specific to NSW, to enhance the model's predictive capabilities.
* **Data Integration**: Combine all relevant datasets into a unified database for analysis specific to NSW.
* **Model Development**: Implement machine learning models to forecast energy demand based on the identified variables in NSW.
* **Output Analysis and Recommendations**: Analyze the model outputs and provide actionable recommendations for energy suppliers in NSW.
* **Visualization and Communication**: Create visualizations to effectively communicate the analysis results to stakeholders in NSW.
* **Report Writing**: Compile the project’s findings into a comprehensive report.
* **Video Presentation**: Develop a video presentation summarizing the project and its outcomes.

**4.2 Activities and Schedule**

| **Activity** | **Week 1** | **Week 2** | **Week 3** | **Week 4** | **Week 5** | **Week 6** |
| --- | --- | --- | --- | --- | --- | --- |
| Literature Review | X | X |  |  |  |  |
| Data Assessment | X |  |  |  |  |  |
| Algorithm Research | X | X |  |  |  |  |
| Data Cleaning & Enrichment |  |  | X |  |  |  |
| Data Integration |  |  | X |  |  |  |
| Model Development |  |  | X | X |  |  |
| Output Analysis & Recommendations |  |  |  | X | X |  |
| Visualization |  |  |  | X | X |  |
| Report Writing |  |  |  |  | X | X |
| Video Presentation |  |  |  |  | X | X |

**5 Prediction Models and Techniques**

This section outlines the models and techniques that will be employed in forecasting energy demand in NSW.

**5.1 Time Series Decomposition Models**

* **Seasonal Decomposition of Time Series (STL)**: To decompose energy demand data into trend, seasonal, and residual components, providing a clearer understanding of underlying patterns specific to NSW.

**5.2 Regression Models**

* **Linear Regression**: To establish the relationship between energy demand and various predictors, such as temperature and day of the week, in NSW.
* **Multiple Linear Regression**: Extends linear regression by incorporating multiple independent variables to predict energy demand in NSW.

**5.3 Machine Learning Models**

* **Decision Trees**: Used to capture non-linear relationships and interactions between variables in NSW.
* **Random Forests**: An ensemble method that averages the predictions of multiple decision trees to improve accuracy in NSW.
* **Gradient Boosting Machines (GBM)**: Sequentially builds trees, with each tree correcting the errors of the previous one, specifically for NSW data.
* \*\*XGBoost