**Forecasting Natural Gas Consumption and Assessing Environmental Impact in the USA**

**Abstract**

The efficient management of energy resources and the mitigation of environmental impacts are critical challenges in the United States, particularly in the context of natural gas consumption. This study employs advanced statistical and machine learning methodologies to forecast natural gas consumption and assess its environmental and economic drivers across various states. Utilizing a comprehensive dataset from 2014 to the present, which includes variables such as monthly gas consumption in million cubic feet (MMCF), geographic locations, economic indicators, and seasonal trends, this research aims to provide a detailed analysis of consumption patterns. Key objectives include the development of accurate predictive models using ARIMA, SARIMA, and Machine Learning algorithms like Random Forests to project future consumption rates. Furthermore, the study explores the relationship between gas consumption and CO2 emissions to assess the environmental impact, particularly focusing on the commercial sector. The outcomes are intended to inform policymakers and stakeholders in crafting effective energy policies and initiatives that align with sustainability goals. The findings are expected to contribute significantly to the discourse on energy management, highlighting regional consumption trends, the influence of economic cycles, and the effectiveness of current energy policies in reducing carbon footprints across the United States.

**Introduction**

The United States stands as one of the world's largest consumers of natural gas, utilizing this resource extensively across residential, commercial, and industrial sectors. As environmental concerns grow and the economy evolves, understanding and predicting natural gas consumption patterns become crucial for energy planning and sustainability efforts. Advances in data analytics, particularly through machine learning, offer potent tools for enhancing our ability to forecast and manage energy use effectively. This project seeks to leverage a comprehensive dataset detailing gas consumption across various states to forecast future demand and analyze consumption trends. By employing sophisticated machine learning techniques alongside traditional statistical models, this study aims to provide actionable insights that could influence policy decisions, optimize resource allocation, and minimize environmental impacts.

**Aims**

The primary aim of this project is to develop robust predictive models that can accurately forecast natural gas consumption across different regions of the USA. By integrating machine learning with traditional forecasting methods, the project seeks to enhance the precision of consumption predictions and provide a deeper understanding of the factors influencing these patterns.

**Objectives**

1. **Develop Predictive Models**: Construct and refine predictive models using both machine learning techniques (such as Random Forest and Neural Networks) and traditional statistical methods (like ARIMA and SARIMA) to forecast monthly and annual gas consumption.
2. **Evaluate Environmental Impact**: Assess the environmental implications of current and forecasted gas consumption patterns. This includes estimating greenhouse gas emissions associated with projected gas consumption and analyzing potential mitigation strategies.
3. **Compare Forecasting Models**: Evaluate the effectiveness of various forecasting models in terms of accuracy, reliability, and computational efficiency. Identify the strengths and limitations of each model in different scenarios and regions.

**Dataset Description**

The dataset contains the following columns:

1. **Year**: Represents the year when the data was recorded

**Usage**: Crucial for time series analysis. Helps in identifying trends and patterns over multiple years, allowing for the examination of long-term changes in natural gas consumption. Essential for building forecasting models like ARIMA and SARIMA that rely on temporal sequences.

1. **Month**: Represents the month when the data was recorded

**Usage**: Essential for detecting seasonal variations in gas consumption. Allows for the identification of peak and low consumption periods within a year, critical for seasonal forecasting models like SARIMA. Aids in understanding monthly consumption patterns and their correlation with climatic conditions.

1. **Duo area**: Abbreviated state code

**Usage**: Used to group data by state for geographical analysis. Helps in comparing natural gas consumption across different states and regions, identifying states with higher or lower consumption levels, and tailoring energy policies and strategies to specific areas.

1. **Area-Name**: Full state name

**Usage**: Similar to Duoarea, used for geographical analysis and visualization. Makes the dataset more readable and can be used in reporting and presenting findings to stakeholders.

1. **Product**: Product code (e.g., EPG0 for natural gas).

**Usage**: Distinguishes different types of energy products if the dataset includes multiple products. For this project, since the focus is on natural gas, it can be used to filter and confirm that the analysis is specifically on the correct product.

1. **Product-Name**: Product name (e.g., Natural Gas).

**Usage**: Provides a clear identification of the product being analyzed. Ensures that the analysis is consistent and focused on natural gas, which is the subject of this study.

1. **Process**: Process code (e.g., VCS for Commercial Consumption).

**Usage**: Identifies the type of consumption process. For example, VCS stands for Commercial Consumption. Useful for differentiating between various consumption types (e.g., commercial, residential, vehicle fuel) and focusing the analysis on commercial consumption specifically.

1. **Process-Name**: Process description (e.g., Commercial Consumption).

**Usage**: Similar to the Process code but provides a more readable description. Ensures clarity in the analysis and helps in reporting the findings accurately.

1. **Series**: Series code (e.g., N3020FL2).

**Usage**: Uniquely identifies each data series. Useful for data management and can be used to ensure that the correct data series is being analyzed and referenced in the study.

1. **Series-Description**: Description of the series (e.g., Natural Gas Deliveries to Commercial Consumers in Florida).

**Usage**: Provides detailed information about what the data series represents. Helps in understanding the context and scope of the data, ensuring that the analysis is correctly focused on the intended metric.

1. **Value**: The volume of gas consumed (e.g., 6605 MMCF).

**Usage**: This is the dependent variable (DV) in the study. Represents the actual volume of natural gas delivered to commercial consumers. This feature is the primary focus of the analysis, and all models will aim to predict this value based on the independent variables.

1. **Units**: The units of measurement (e.g., MMCF).

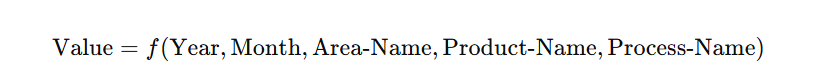
**Usage**: Ensures consistency and clarity in the measurement of gas consumption values. Crucial for accurate analysis, comparison, and interpretation of the data.

**Preliminary Research Model**

The preliminary research model includes the following key variables:

* **Dependent Variable (DV)**: Volume of natural gas delivered to commercial consumers (Value in MMCF).
* **Independent Variables (IVs)**: Year, Month, Area-Name, Product-Name, Process-Name.

The relationship between the dependent and independent variables can be represented as:

****

**Research Questions**

1. How do seasonal patterns affect gas consumption in different states?
2. Can machine learning models accurately forecast future gas consumption in the USA based on historical data?
3. Can regional comparisons within the USA provide insights into targeted strategies for reducing gas consumption?

**Methodology**

1. **Time Series Analysis**:
   * Utilize Year and Month to identify trends and seasonal patterns in gas consumption.
   * Build ARIMA and SARIMA models to forecast future consumption based on historical data.
2. **Geographical Analysis**:
   * Use Duoarea and Area-Name to compare gas consumption across different states.
   * Identify regions with higher or lower consumption levels and investigate potential causes.
3. **Machine Learning Models**:
   * Develop predictive models using Random Forests and Neural Networks.
   * Evaluate model performance based on accuracy, reliability, and computational efficiency.
4. **Environmental Impact Analysis**:
   * Correlate gas consumption (Value) with emissions data to assess the environmental impact.
   * Estimate greenhouse gas emissions associated with forecasted gas consumption.

**Literature Review**

A comprehensive literature review will be conducted to contextualize the research within the existing body of knowledge. Key references include:

* Bala, D. A., & Shuaibu, M. (2024). Forecasting United Kingdom’s energy consumption using machine learning and hybrid approaches. Energy & Environment, 35(3), 1493-1531. <https://doi.org/10.1177/0958305X221140569>’
* AL-Musaylh, M. S., Al-Daffaie, K., & Prasad, R. (2021). Gas consumption demand forecasting with empirical wavelet transform based machine learning model.

These studies provide insights into the methodologies and techniques that can be adapted and applied to the current research.

**Ethics Considerations**

The research will adhere to ethical guidelines to ensure the integrity and reliability of the findings. Key ethical considerations include:

* Ensuring data privacy and confidentiality.
* Accurately reporting and interpreting the data and results.
* Avoiding any form of data manipulation or bias.

**Conclusion**

This project aims to develop robust predictive models for natural gas consumption and assess its environmental impact in the USA. By leveraging advanced statistical and machine learning techniques, the research will provide valuable insights into consumption patterns, regional differences, and the effectiveness of energy policies. The findings are expected to inform policymakers and stakeholders in crafting effective energy strategies that align with sustainability goals.