Title: Forecasting Future Energy Consumption for Households and Businesses

TEAM MEMBERS

1. Ousmane OUEDRAOGO
2. Lucien KALMOGO
3. Yann Boris OUEDRAOGO
4. Edene Vasco BENDANE

Table of Contents

Figures

1 Literature Review:

1.1 Introduction:

Energy consumption is a cornerstone of household and business operations. In the era of energy transition, accurate forecasting of energy demand is crucial to:

1. **Optimize available resources**: enabling utility providers to better manage supply and demand.
2. **Reduce energy waste**: preventing unnecessary overproduction that increases costs and impacts the environment.
3. **Support sustainability goals**: contributing to energy systems that are more environmentally friendly by limiting CO_2 emissions.

However, energy forecasting presents significant challenges due to the complexity of factors influencing consumption, such as consumer habits, seasonality, and weather conditions. A literature review is essential to:

- Identify existing models and methodologies.
- Understand their strengths and limitations.
- Justify the relevance of the approaches proposed in this project.

1.2 Organization:

For better readability, the studies have been grouped based on the methodological approaches used.

1.2.1 Classical Statistical Approaches

Early research on energy forecasting relied on statistical models such as **ARIMA (Auto-Regressive Integrated Moving Average)** and **SARIMA** (Seasonal ARIMA). These models are particularly suited for time series with repetitive patterns.

- **Strengths:**
  - Effectively capture linear and seasonal trends.
  - Easy to interpret and implement.
- **Weaknesses:**
  - Limited in handling nonlinear relationships.
  - Struggle to integrate complex exogenous factors like weather or behavioral habits.

Key Example: A study by Box and Jenkins (1970) demonstrated that ARIMA models were widely used in industrial energy planning. However, these models often fail in scenarios with sudden data variations.

1.2.2 Machine Learning Approaches

With the explosion of available energy data, machine learning algorithms have become a preferred tool. Models like **XGBoost**, **Random Forest**, and **simple neural networks (MLP)** excel at analyzing complex relationships.

- **Strengths:**
    - Handle large datasets effectively.
    - Adapt to nonlinear relationships.
    - Automatically select relevant features.
- **Weaknesses:**
    - Require high-quality data (with minimal missing values).
    - Demand significant computational resources.

Key Example: A study by Chou et al. (2018) used XGBoost to forecast energy consumption in residential and industrial settings, achieving a 15% reduction in Mean Absolute Error (MAE) compared to ARIMA models.

1.2.3 Exogenous Factors and Hybrid Models

Recent research integrates external variables (e.g., weather, socioeconomic, or geographic data) into models. This enhances forecasting accuracy, particularly in contexts where energy consumption heavily depends on seasonal factors.

- **Hybrid Models:** Combining statistical techniques with machine learning approaches.
- *Key Example:* Gellert et al. (2020) used a hybrid model combining SARIMA and a neural network to incorporate weather data, resulting in improved energy forecasting accuracy.

1.3 **Summary and Synthesis:**

1.3.1 Commonalities across studies:

- Seasonality and temporal trends are key aspects of energy consumption modeling.
- Modern approaches like machine learning outperform classical models in complex scenarios.

1.3.2 Main differences:

- Statistical approaches, while interpretable, lack flexibility for nonlinear scenarios.
- Machine learning models require richer and well-structured datasets.

1.3.3 Identified gaps:

- Few existing models integrate real-time data for dynamic forecasting.
- Studies rarely address both short-term and long-term energy optimization simultaneously.

1.4 **Conclusion:**

The existing literature shows that energy forecasting is a continuously evolving field. While classical models remain useful for simpler datasets, advancements in machine learning pave the way for more precise and adaptive predictions.

Our project leverages these insights to:

- ✓ Develop an **XGBoost-based model**, integrating historical, climatic, and socioeconomic data.
- ✓ Provide short-term (daily) and long-term (monthly or yearly) forecasts.
- ✓ Offer tools for suppliers and consumers to optimize energy use and reduce costs.

1.5 **Proper Citations:**

- Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*.
- Chou, J., et al. (2018). "Machine Learning for Energy Consumption Forecasting in Smart Grids". *IEEE Transactions on Smart Grid*.
- Gellert, Á., et al. (2020). "Forecasting Electricity Consumption and Production in Smart Homes through Statistical Methods".

2 **Data Research:**

2.1 **Introduction**

Understanding energy consumption patterns is critical in the context of increasing global energy demand and the push for sustainability. This project focuses on analyzing household energy usage and identifying key drivers of consumption. Reliable insights will guide the development of predictive models to optimize energy usage, minimize costs, and support sustainable resource management. A thorough exploration of the dataset is essential to uncover meaningful patterns and trends that inform these objectives.

2.2 **Organization**

This report is organized as follows:

- **Data Description**: Overview of the dataset, including its source, structure, and relevance.
- **Data Analysis and Insights**: Summary of key trends, descriptive statistics, and patterns revealed during the data exploration.
- **Conclusion**: Highlights of findings and their significance in achieving the project goals.
- **References**: Proper citation of the data source.

2.3 Data Description

2.3.1 Data Source

The dataset was sourced from **Kaggle**, as part of an energy prediction project.

2.3.2 **Data Format**

- **Format**: CSV file.
- **Columns**: 29, covering household appliance energy consumption, environmental conditions, and indoor climate variables.
- **Rows**: 19,735 entries, each representing hourly measurements.

2.3.3 **Relevance**

This dataset was selected because:

1. **Granular Time Series**: Hourly data enables detailed analysis of consumption trends.

2. **Comprehensive Variables**: The dataset includes both household-specific factors (e.g., temperature, humidity) and external influences (e.g., wind speed, outdoor temperature).
3. **Predictive Potential**: The diversity of features allows for building robust machine learning models.

2.4 Data Analysis and Insights

2.4.1 Key Insights and Patterns

2.4.1.1 Temporal Trends

- **Daily Cycles**: Energy consumption for appliances peaks between **6 p.m. and 9 p.m.**, reflecting typical household routines.
- **Weekly Patterns**: Consumption decreases on weekends, indicating reduced appliance usage during these days.

2.4.1.2 Environnemental Influence

- **Weather Effects**: Energy usage decreases during high wind speeds and low visibility, potentially due to behavioral changes.
- **Temperature Correlation**: Indoor temperatures (T1 to T9) show strong positive correlations with energy consumption, while outdoor temperature (T_out) has a moderate effect.

2.4.1.3 Statistical Relationships

- **High Correlations**: Variables like T1 (living room temperature) and RH_1 (living room humidity) have significant correlations with appliance energy consumption.
- **Low Variance**: Variables such as lights exhibit relatively low variability, contributing minimally to overall energy dynamics.

2.4.2 Descriptive Statistics

Columns	mean	min	25%	50%	75%	max	std
Appliances	97,694958	10	50	60	100	1080	102,52489
lights	3,801875	0	0	0	0	70	7,935988
T1	21,686571	16,79	20,76	21,6	22,6	26,26	1,606066
RH_1	40,259	27,023	37,33	39,65667	43,06	63,36	3,979299
T2	20,341219	16,1	18,79	20	21,5	29,85	2,192974
RH_2	40,42042	20,46333	37,9	40,5	43,26	56,026	4,069813
T3	22,267611	17,2	20,79	22,1	23,29	29,236	2,006111
RH_3	39,2425	28,76667	36,9	38,53	41,76	50,163	3,254576
T4	20,855335	15,1	19,53	20,66667	22,1	26,2	2,042884
RH_4	39,026904	27,66	35,53	38,4	42,156667	51,09	4,341321
T5	19,592106	15,33	18,2775	19,39	20,619643	25,795	1,844623
RH_5	50,949283	29,815	45,4	49,09	53,663333	96,32167	9,022034
T6	7,910939	-6,065	3,626667	7,3	11,256	28,29	6,090347
RH_6	54,609083	1	30,025	55,29	83,226667	99,9	31,149806
T7	20,267106	15,39	18,7	20,03333	21,6	26	2,109993
RH_7	35,3882	23,2	31,5	34,86333	39	51,4	5,114208
T8	22,029107	16,30667	20,79	22,1	23,39	27,23	1,956162
RH_8	42,936165	29,6	39,066667	42,375	46,536	58,78	5,224361
T9	19,485828	14,89	18	19,39	20,6	24,5	2,014712
RH_9	41,552401	29,16667	38,5	40,9	44,338095	53,32667	4,151497
T_out	7,411665	-5	3,666667	6,916667	10,408333	26,1	5,317409
Press_mm_hg	755,5226	729,3	750,93333	756,1	760,93333	772,3	7,399441
RH_out	79,750418	24	70,333333	83,66667	91,666667	100	14,901088
Windspeed	4,039752	0	2	3,666667	5,5	14	2,451221
Visibility	38,330834	1	29	40	40	66	11,794719
Tdewpoint	3,760707	-6,6	0,9	3,433333	6,566667	15,5	4,194648
rv1	24,988033	0,005322	12,497889	24,89765	37,583769	49,99653	14,496634
rv2	24,988033	0,005322	12,497889	24,89765	37,583769	49,99653	14,496634
month	3,101647	1	2	3	4	5	1,3392
hour	11,502002	0	6	12	17	23	6,921953

Figure 1 : Descriptives statistics

The dataset shows that the energy consumption (Appliances) has a wide range, from a minimum of 10 to a maximum of 1080, with a mean of about 97.7. The temperature (T1 to T9) and humidity (RH_1 to RH_9) values are relatively consistent, with minor fluctuations across the rooms. External conditions such as T_out, Press_mm_hg, and Windspeed show a broader range, indicating variability in weather patterns.

2.4.3 Visualizations

- **Energy Consumption Over Time**

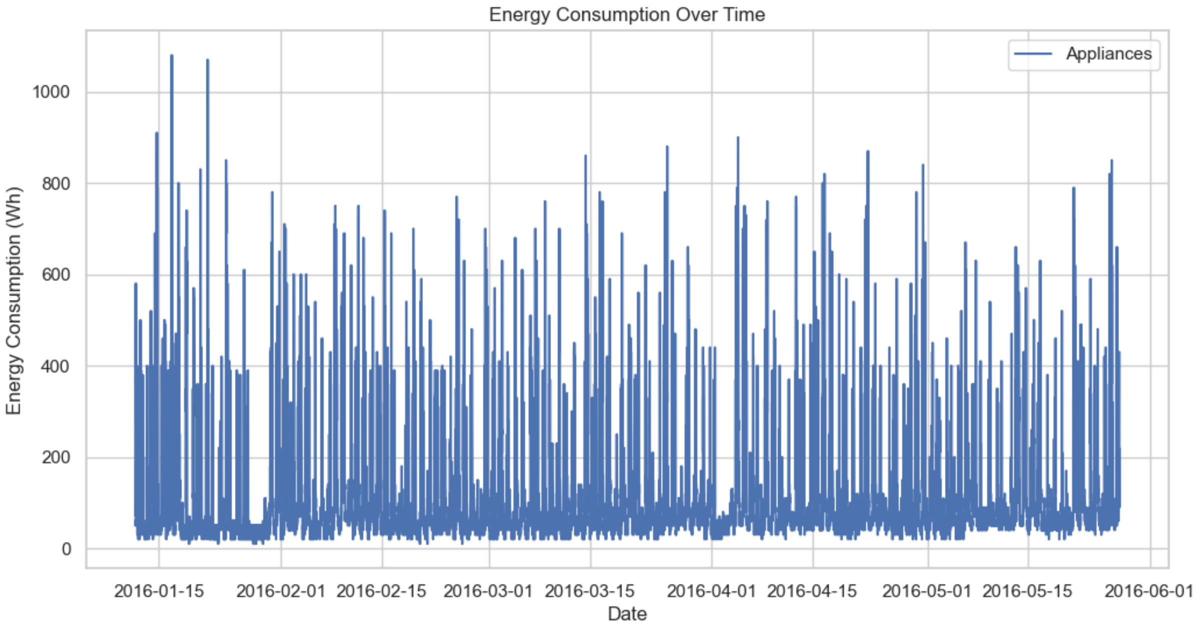

Figure 2 : Energy Consumption Over time

The plot shows fluctuations in energy consumption over time, with sharp peaks and valleys. The consumption appears to vary throughout the months, indicating periods of high and low energy usage. The energy consumption seems to spike periodically, possibly correlating with specific events or changes in usage patterns.

- **Distribution of Relative Humidity in Room 1**

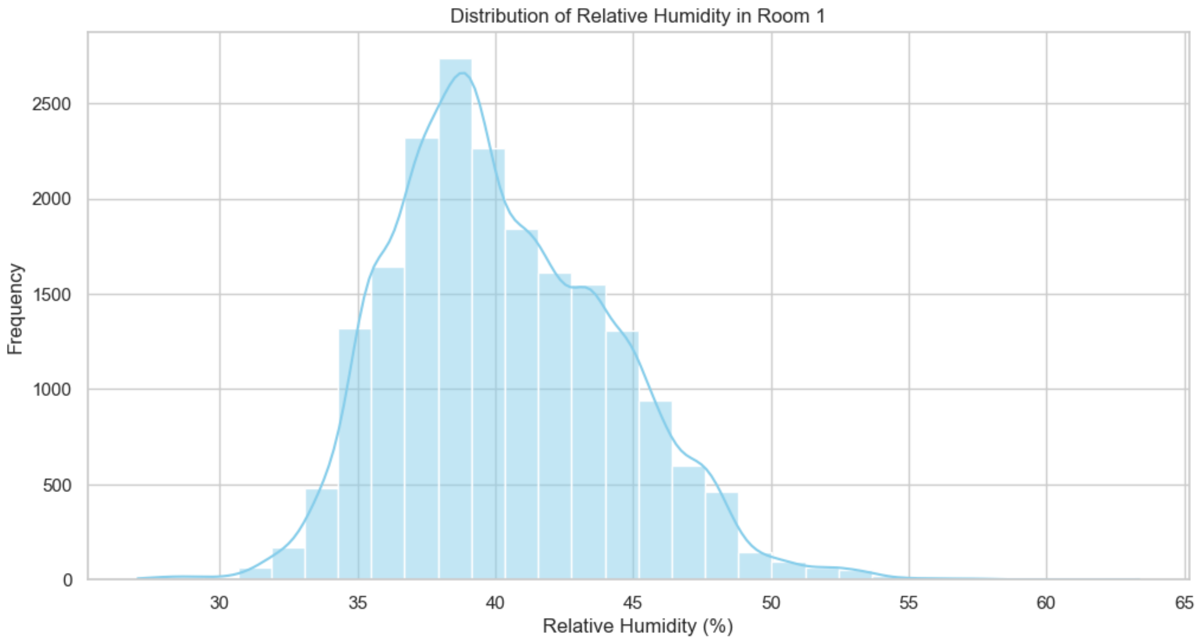

Figure 3 : Distribution of Relative Humidity in Room 1

The graph shows that the relative humidity (RH_1) tends to peak around 40%, with values fluctuating between 30% and 55%. This suggests that the humidity is relatively stable within this range but experiences occasional spikes around the 40% mark, indicating potential changes in environmental conditions or specific events influencing the humidity levels during the observed period.

- **Average Daily Energy Consumption by Day of the Week**

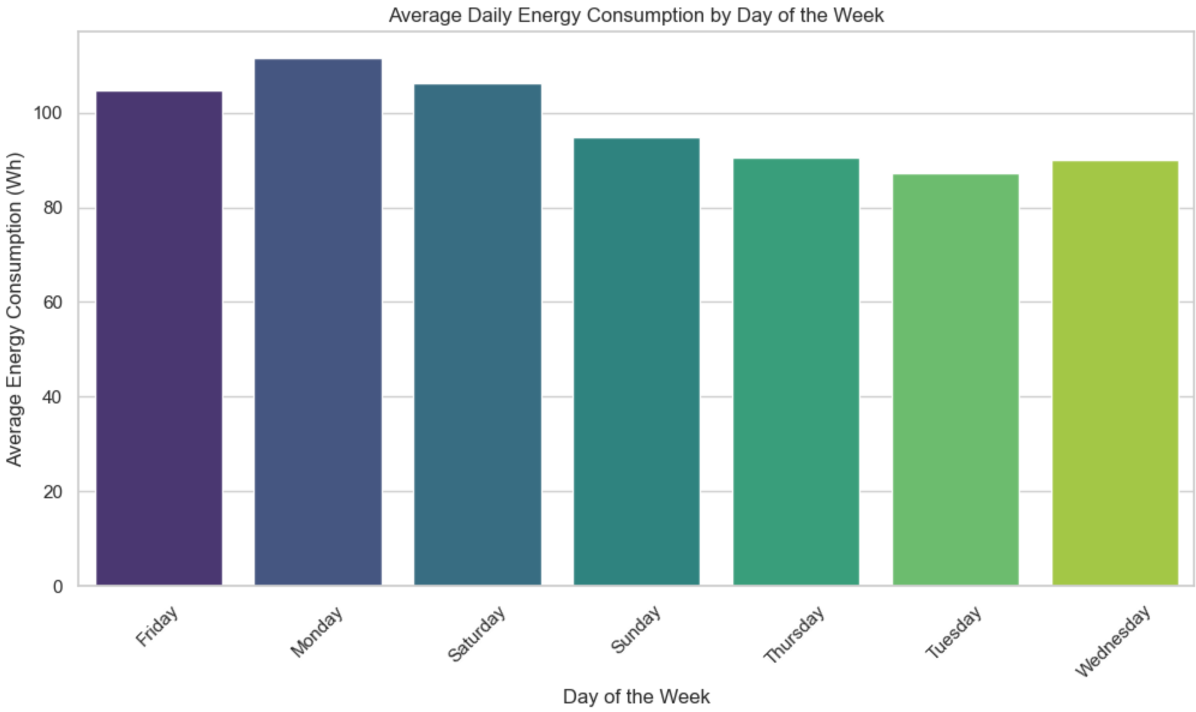

Figure 4 : Average Daily Energy Consumption by Day of the Week

The graph showing average daily energy consumption by day of the week indicates that energy consumption is significantly higher on Mondays and Saturdays compared to other days. This could suggest that these days involve higher activity or specific events that lead to increased energy usage, such as people being at home more on weekends or starting the week with higher demand.

- **Temperature vs Relative Humidity in Room 1**

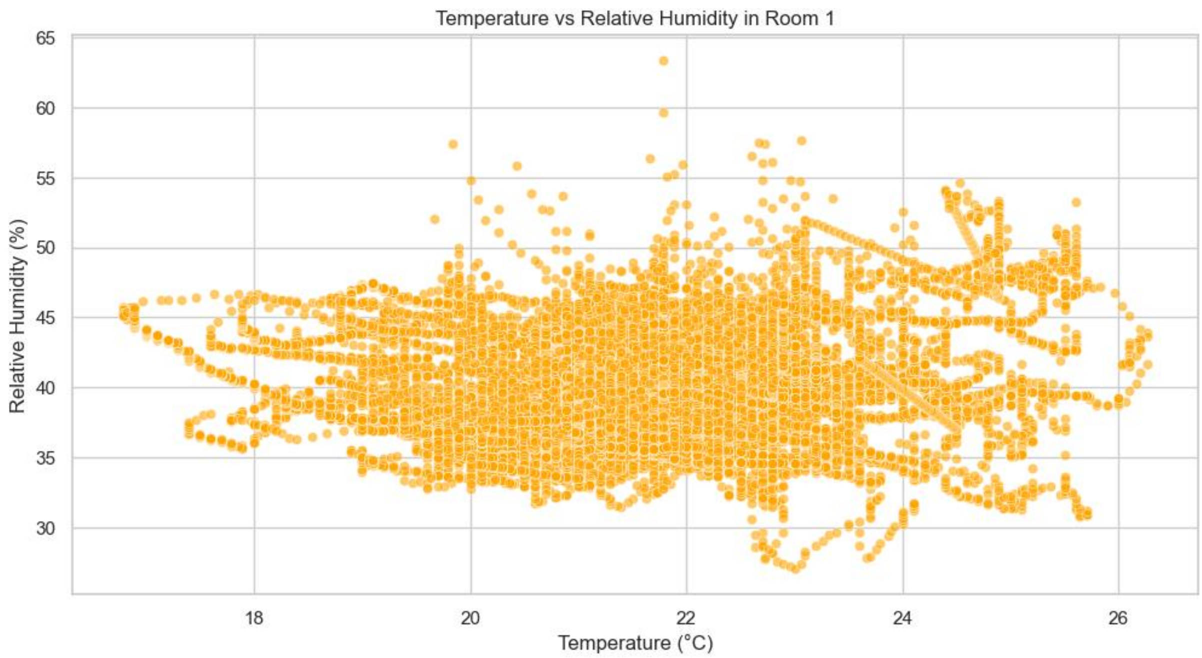

Figure 5 : Temperature vs Relative Humidity in Room 1

The graph showing the relationship between temperature and relative humidity in Room 1 reveals that the temperature fluctuates between 18°C and 26°C, with a concentration of values between 20°C and 24°C. This indicates that the temperature tends to stay within a moderate range, and relative humidity levels may be more stable or consistent within this temperature range, suggesting comfortable room conditions during this period.

2.5 **Conclusion**

This research has uncovered significant patterns in household energy consumption, with notable correlations between environmental factors such as temperature and humidity and energy demand. The analysis of daily and seasonal consumption variations, along with temperature-humidity relationships, highlights the potential for predictive models to optimize energy usage more effectively. These insights contribute to the goal of enhancing energy efficiency and serve as a valuable foundation for developing sustainable energy solutions, offering actionable guidance for both consumers and policymakers.

2.6 **Proper Citations**

- Kaggle, *Energy Prediction Dataset*: Kaggle Link.

3 Technology Review

3.1.1 Introduction

This technology review explores the tools used for developing an energy consumption forecasting model based on time series data. The goal is to select the best technologies for data processing, model deployment, and result visualization.

Importance: Analyzing the relevant technologies is crucial for ensuring the model's performance, scalability, and accuracy.
Relevance: Time series models, Django, and Power BI are the key tools chosen for this project, each addressing specific challenges.

3.1.2 Technology Overview

- **Time Series Models**: Using methods like ARIMA, XGBoost, and LSTM for energy consumption forecasting. ARIMA is simple but limited, while XGBoost and LSTM are more flexible and perform better on large datasets.
- **Django for Deployment**: A Python framework for creating an API accessible via a web interface, facilitating model integration.
- **Power BI for Visualization**: A business intelligence tool that enables the creation of interactive dashboards for analyzing and visualizing energy consumption forecast results.

3.1.3 Relevance to the Project

These technologies are crucial for:

- **Time Series Models**: Predicting energy consumption based on historical data.
- **Django**: Deploying the model and facilitating interaction with end-users.
- **Power BI**: Visualizing data dynamically for informed decision-making.

3.1.4 Comparison and Evaluation

- **Time Series Models**: ARIMA is simple but limited. XGBoost and LSTM provide more flexibility but require more resources.

- **Django**: Easy to deploy and well-integrated with Python, but can be limited in handling large loads.
- **Power BI**: Ideal for data visualization with an intuitive interface but has limitations for real-time analysis of large data volumes.

3.1.5 Use Cases

- **Time Series**: Electricity demand forecasting in smart cities demonstrates the effectiveness of XGBoost.
- **Django**: Used for deploying predictive models in energy management.
- **Power BI**: Used in energy management companies for visualizing forecasts.

3.1.6 Gaps and Opportunities

- **Gaps**: ARIMA/SARIMA models struggle with sudden changes, and Power BI may have limitations with large datasets.
- **Opportunities**: Improving forecasts with hybrid models and integrating real-time data.

3.1.7 Conclusion

The chosen technologies (time series models, Django, Power BI) are well-suited for this energy consumption forecasting project. They enable efficient modeling, simple deployment, and clear visualization for end-users. These tools are essential to ensure the performance and accuracy of the predictive model.