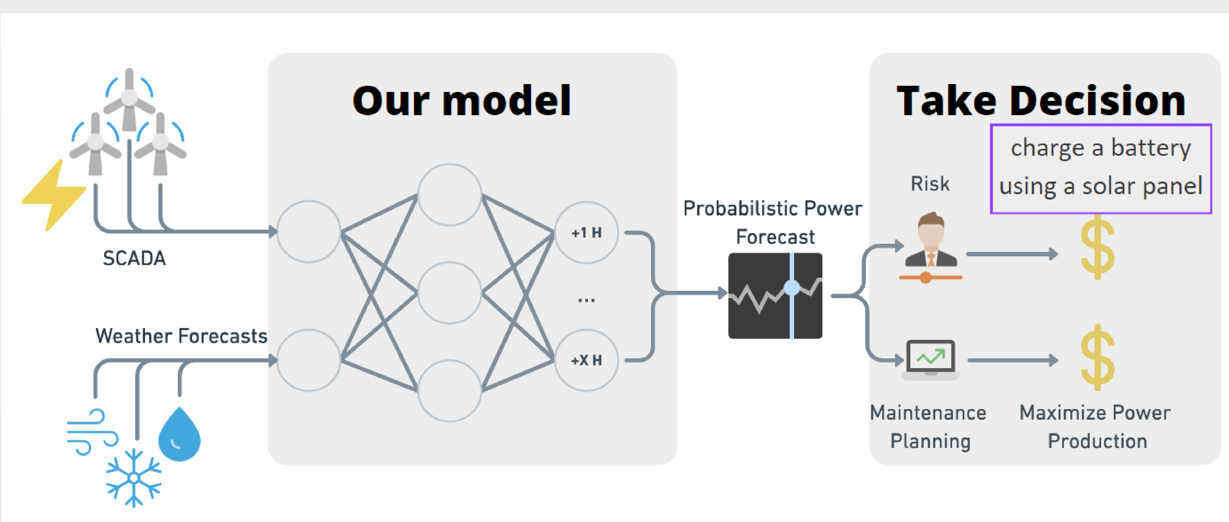
Predicting Energy Consumption using Machine Learning: Optimizing Production and Distribution.

EL-FAILALI-LBALGHITI Mohamed
National School of Applied Sciences Marrakech (ENSA Marrakech)

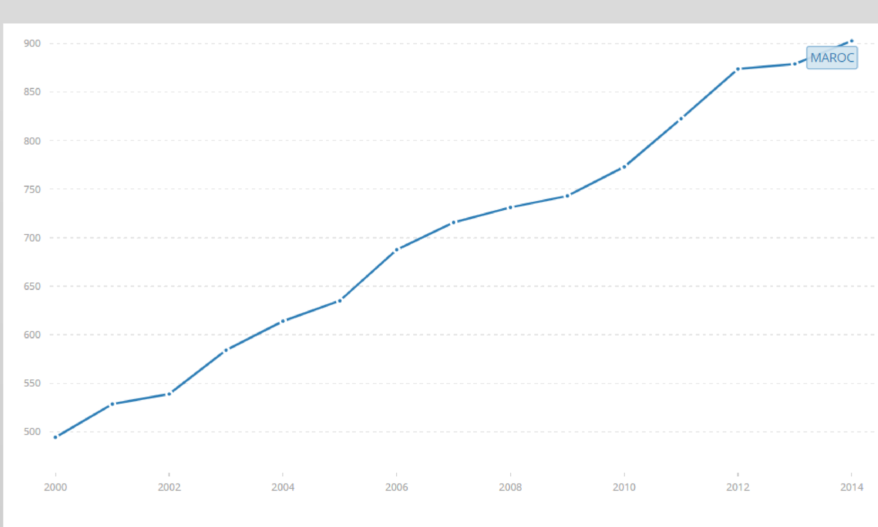
Abstract

This project aims to address the challenge of predicting energy consumption using machine learning techniques to optimize energy production and distribution. The primary goal is to develop an accurate predictive model for energy consumption based on historical data and relevant features. The model will enable energy providers to anticipate demand and improve energy distribution while minimizing waste and reducing costs. The project utilizes various machine learning algorithms such as Linear Regression, Random Forest, and XGBoost to create and evaluate the performance of the predictive model. The results indicate that the XGBoost algorithm provides the most accurate predictions for energy consumption. This project's outcomes contribute to the development of more efficient and sustainable energy systems, reducing environmental impact and improving the overall energy supply chain's efficiency.



Introduction

As the world's population and economies continue to expand, energy consumption has elevated to a top global concern.



Morocco's electricity usage from 2000 to 2014

Inefficient use of energy not only results in high costs for consumers but also has negative environmental impacts. Therefore, there is a need for effective management of energy resources to ensure their optimal use while reducing the environmental footprint.
This poster aims to present a machine-learning approach for predicting energy consumption, with a focus on optimizing energy production and distribution. The goal is to provide stakeholders with accurate and reliable information to aid decision-making processes (better manage purchases and stock) and ultimately contribute to sustainable energy management.
Understanding the energy consumption and needs of the country, starting with a medium-sized city (in this case, Tetouan), could be a step further in planning these resources.

Methodology

1. Data Collection:

The Supervisory Control and Data Acquisition System (**SCADA**) of Amendis, a public service operator, is responsible for collecting and providing the project's power consumption data. The distribution network is powered by 3 zone stations, namely: Quads, Smir and Boussafou. The 3 zone stations power 3 different areas of the city, but in our study we will focus on the **first zone** named **Quads**, this is why we have one potential target variable.

Dataset Description

The data has observations in 10-minute windows starting from January 1st, 2017, all the way through December 30th of the same year. Some of the features are:

- Date Time: Time window of ten minutes
- Temperature: Weather Temperature in °C
- Humidity: Weather Humidity in %
- Wind Speed: Wind Speed in km/h
- General Diffuse Flows: The general diffuse flows in the city, which may include air or water flows.
- Diffuse Flows: This attribute represent a specific type of diffuse flow in the city.
- Quads Zone 1 Power Consumption in KiloWatts (KW)

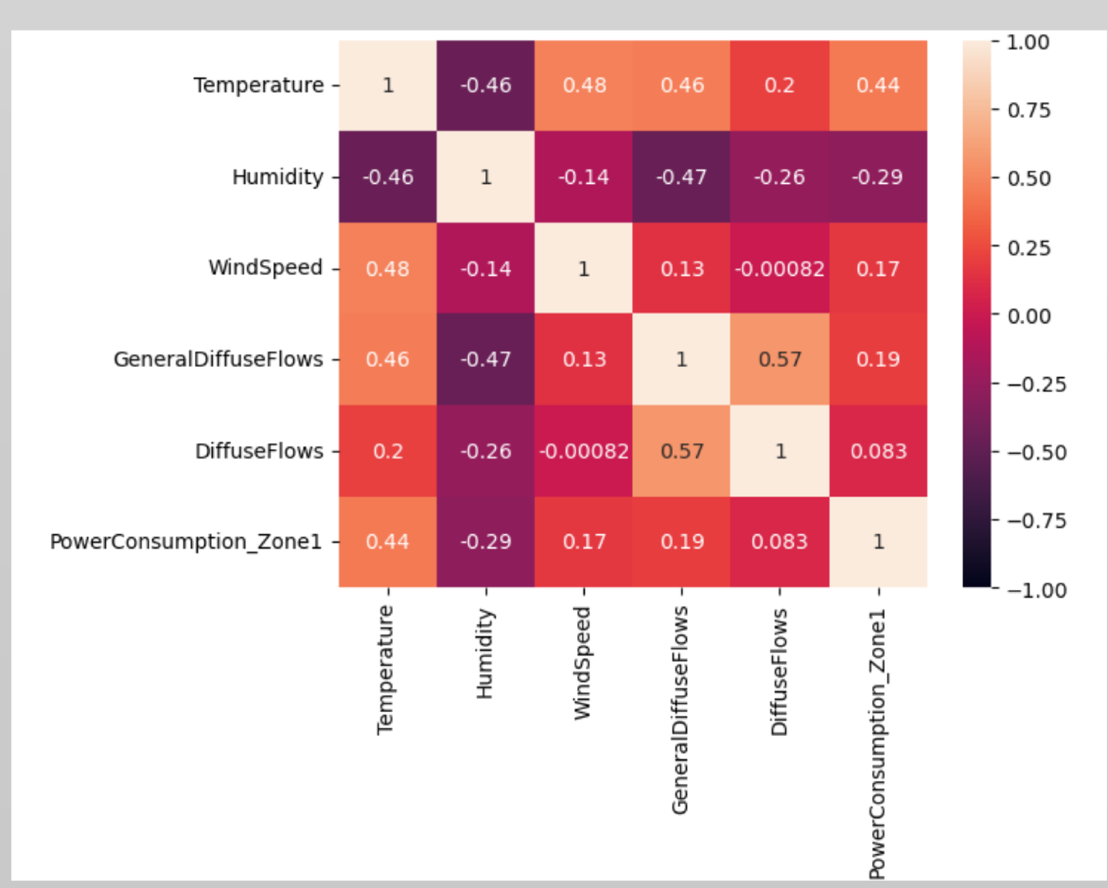
2. Data Cleaning and Preparation:

- Once we had the data, it was cleaned and prepared for analysis. So we don't need to remove any missing or inconsistent values, etc.

```
Number of NaN values present:
Temperature                0
Humidity                   0
WindSpeed                  0
GeneralDiffuseFlows        0
DiffuseFlows               0
PowerConsumption_Zone1     0
```

3. Feature Selection:

We performed feature selection using correlation analysis and domain knowledge. We selected features that had a strong correlation with zone 1 energy consumption and were known to influence our target

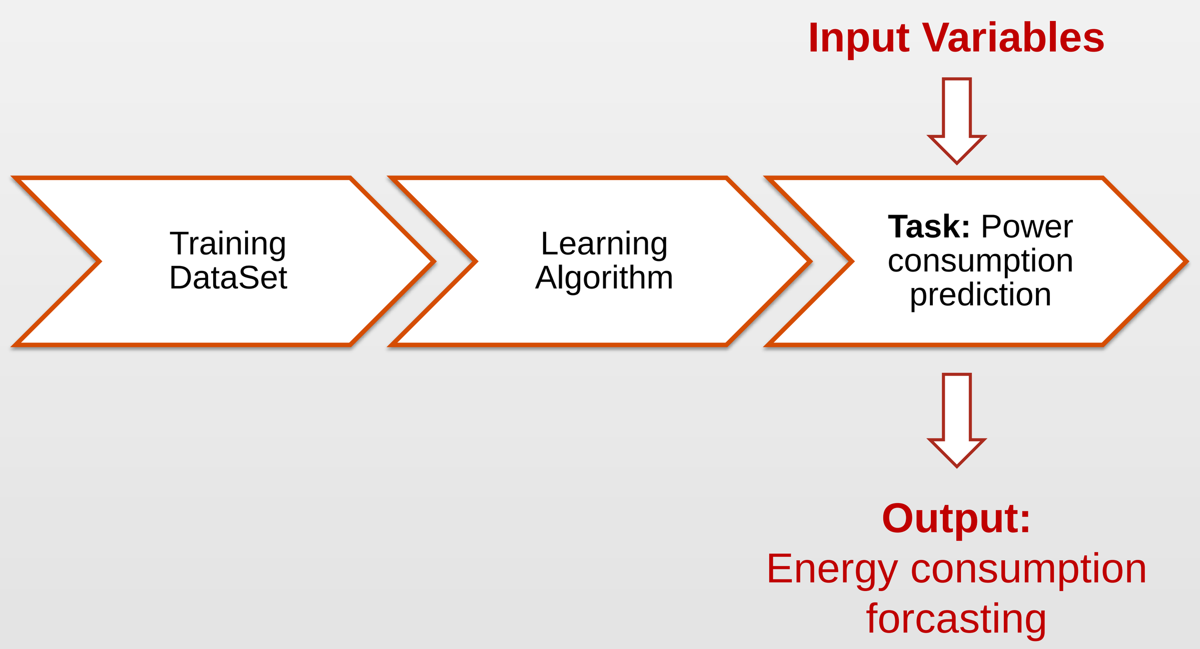


- **Feature Creation**

In time series analysis and forecasting, In time series analysis and forecasting, we can extract a lot of information from the "time" feature. we used the DateTime index to add new columns (features) for this reason, like month, hour, day of week, etc.

4. Model Selection:

We tested several regression models, including **Linear Regression**, **XGBoost**, **Decision Tree Regression**, and **Random Forest Regression**, to identify the best model for predicting energy consumption.



5. Model Evaluation:

We evaluated each model's performance using various metrics, including mean squared error (**MSE**), and R-squared (**R2**) score. We selected the model with the lowest MSE and highest R2 score.

	XGBOOST	DECISION TREE REGRESSION	RANDOM FOREST REGRESSION	LINEAR REGRESSION
MSE	**696654.87**	1579785.42	764674.96	17740113.69
R-SQUARED	**0.9863**	0.9690	0.9850	0.6524

Results

1. Model Performance

The xgboost model showed a **mean squared error** of **696,654.87** and an **R-squared** value of **0.986**. This indicates that the model has a good ability to predict the energy consumption values based on the given input features.

Discussion:

- The mean squared error (MSE) value of 696,654.87 suggests that the average **difference between the predicted energy consumption values and the actual values** is 696,654.87. This value is relatively low, indicating that the model has good predictive capabilities.

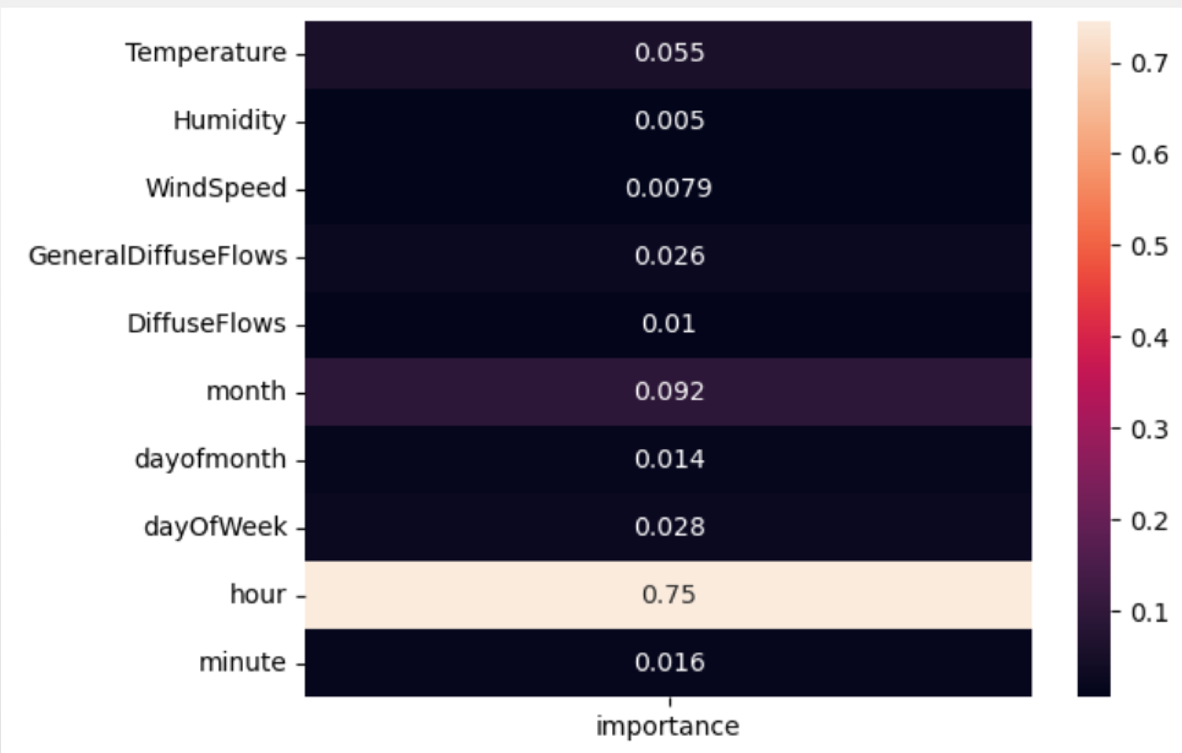
$$MSE = \frac{1}{n}\Sigma\left(y - \hat{y}\right)^2$$

- The R-squared value of 0.986 indicates that the model explains 98.6% of the variability in the energy consumption data. This suggests that the model is a good fit for the data and has a high level of accuracy in predicting energy consumption values.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

2. Feature Importance

We examined the feature importance of the model to determine which features were most crucial in predicting the building's energy consumption.
The results showed that the most important features were the **hour**, **month**, and the **temperature**.



Conclusion

This model aimed to **predict energy consumption** using **supervised machine learning**, with the goal of optimizing production and distribution.
Through the implementation of various regression models, including linear regression and XGBoost, we were able to accurately predict energy consumption with an R-squared value of 0.986.
Our analysis also revealed that the hour, month, and temperature were the most important features in predicting energy consumption. This information can be used by stakeholders to better manage energy purchases and stocks, as well as plan for the production and distribution of energy from renewable sources.
Overall, the successful application of machine learning in predicting energy consumption showcases the potential for technology to be utilized in the energy industry.
As Morocco aims to reduce its dependency on non-renewable sources and increase **production from renewable sources**, accurate predictions of energy consumption will be crucial in achieving this goal.

Furthermore, a possible future project could focus on predicting the energy production using solar panels. This would allow for even more precise planning and optimization of renewable energy sources, ultimately contributing to a more sustainable and efficient energy system.