Meter Reading Regressor using Random Forest & MLP

Powering Predictive Insights into Energy Consumption

Edwin Thomas, Mahalakshmi Muthuvel, Remond Xie, Sathvika Balumuri

INTRODUCTION

Goal: *To predict the Meter Reading based on climatic and temporal features.*

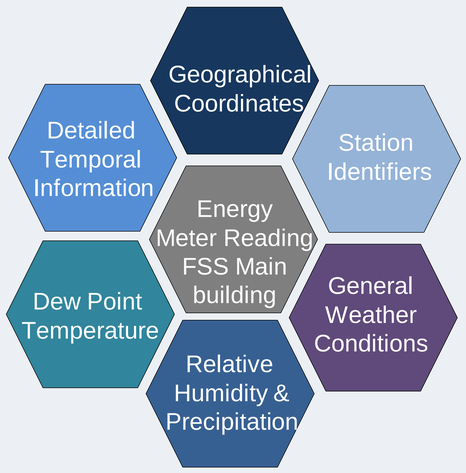
The machine learning models used in this project are:

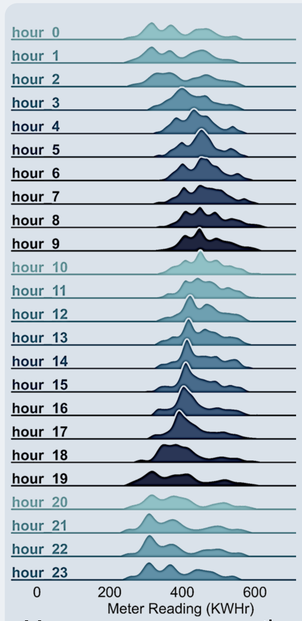
- The Multilayer Perceptron (MLP)
- Random Forest

The models are trained on a dataset that includes features like temperature, humidity, wind speed alongside temporal markers such as date and time. Analyzing how these variables influence energy consumption patterns, to provide accurate meter reading forecasts.

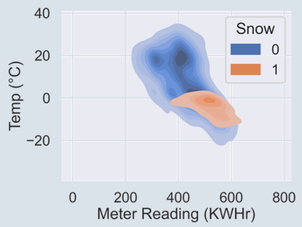
DATA DESCRIPTION

The dataset included both climate and energy consumption data.

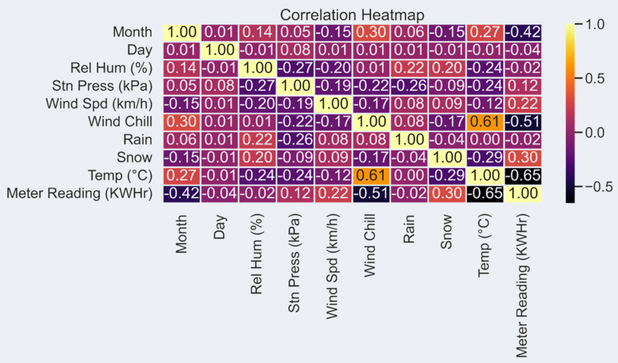




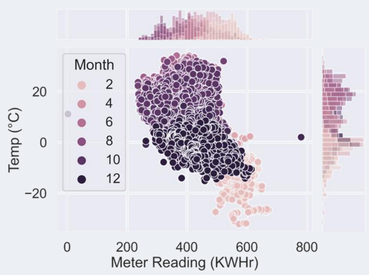
More energy consumption during working hours 9 AM - 5 PM.



- Meter Readings Highest when temperatures are lowest and mostly when there is no snow
- The contours suggest a trend where meter readings are higher in colder temperatures. This could indicate increased energy consumption for heating purposes during colder weather.



EXPLORATORY DATA ANALYSIS



- Highest Meter Reading around coldest months of the year.
- Strong correlation between temperature and energy usage

DATA PREPROCESSING

MERGE FEATURES WITH TARGET
- Merge by Primary Key (Date Time)
- Slide Target by T-1 hr.

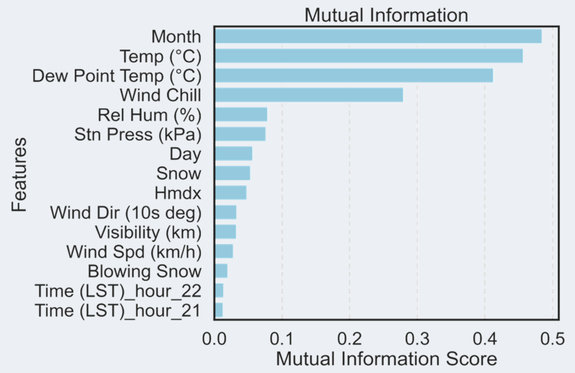
FEATURE PREPROCESSING
- Drop empty and constant feature columns.
- Fill missing values based on Mode.
- Decompose concatenated feature columns (E.g.: Weather column)

FEATURE ENCODING
- Categorical to 1-hot encoding

FEATURE NORMALISATION
- Min-Max scalar

FEATURE SELECTION

Mutual Information – Measure correlation of individual variables to the target variables and decide feature importance



- Attributes Month and Temperature have more importance according to mutual information with target variable.
- Total Number of Features: 57

EXPERIMENTAL SETUP

5-Fold Cross- Validation

- 4 Folds of Training
  - 80% of data
- 1-Fold of Testing
  - 20% of data

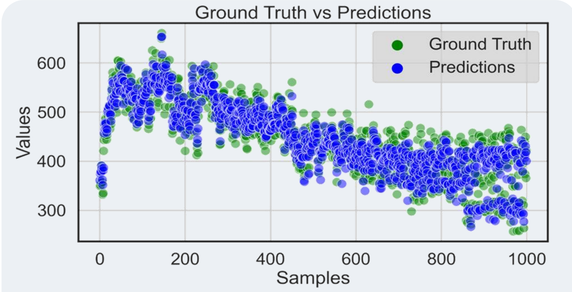
- Average the results of the **5** Cycles
- Developed a Baseline Model (Random Predictor) to compare results

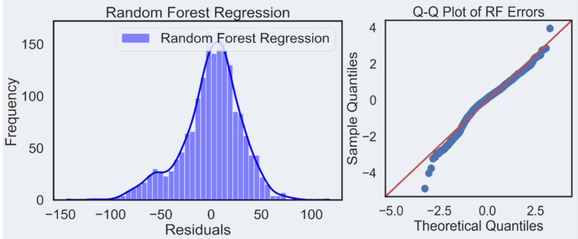
MODELS USED
Random Forest
Multi Layer Perceptron

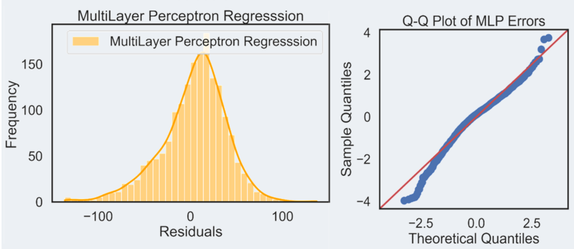
RESULTS

Average Mean Square Error

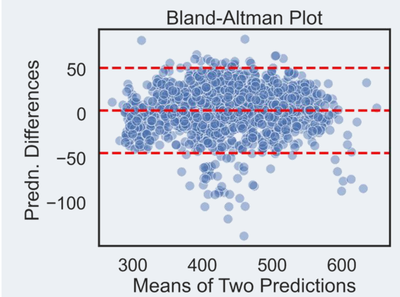
	Baseline	Random Forest	MLP
Fold 1	19369	866.12	1267.91
Fold 2	50029.06	944.44	1318.21
Fold 3	21276.68	939.66	1385.07
Fold 4	36821.93	1011.76	1371.93
Fold 5	20215.07	875.02	1291.57
Average	29542.35	927.4	1326.94







- Test set R^2 (Random Forest Regression): **0.84**
- Test set R^2 (MLP Regression): **0.78**



A scatter plot where the differences between two measurements are plotted against their averages.

Accessing the consistency between the prediction results of both the models

CONCLUSION

- The Random Forest Model outperformed the Multilayer Perceptron in terms of prediction accuracy, with a lower MSE and a higher R2 score, indicating a stronger correlation between predicted and actual meter readings..
- Stakeholders can better anticipate energy needs and implement proactive measures for energy optimization and conservation.