Final Project

Karagiannis Panagiotis
AEM: 03450, pankaragiannis@uth.gr

Kottarinos Paraskevas
AEM: 03551, pkottarinos@uth.gr

Panopoulos Orestis
AEM: 03269, opanopoulos@uth.gr

January 16, 2025

Introduction - Paper Analysis

The paper revolves around tracking the energy consumption of many different appliances using special sensors in many different locations and under a variety of circumstances, like weather change, day, occupancy, building design etc.

The paper suggests that regression models for energy use can help to understand the relationships between different variables and to quantify their impact. Thus, prediction models of electrical energy consumption in buildings can be useful for a number of applications: to determine adequate sizing of photovoltaics and energy storage to diminish power flow into the grid, to detect abnormal energy use patterns, to be part of an energy management system for load control, to model predictive control applications where the loads are needed, for demand side management (DSM) and demand side response (DSR) and as an input for building performance simulation analysis.

0.1 Dataset Presentation

In this work, the prediction was carried out using different data sources and environmental parameters (indoor and outdoor conditions). Specifically, data from a nearby airport weather station, temperature and humidity in dif- ferent rooms in the house from a wireless sensor network and one sub-metered electrical energy consumption (lights) have been used to predict the energy use by appliances.

- **date:** The timestamp of the record, formatted as `dd/mm/yyyy hh:mm` (e.g., "1/11/2016 17:00").
- **Appliances:** Power consumption of appliances, measured in watts (likely).
- **lights:** Power consumption of lighting, also measured in watts.
- **T1, RH_1:** Indoor temperature (`T1`) and relative humidity (`RH_1`) of the kitchen.
- **T2, RH_2, T3, RH_3, ..., T9, RH_9:** Indoor temperature (`T`) and relative humidity (`RH`) for different zones or rooms in the building. For example:
  - **T2, RH_2:** Temperature and relative humidity in the living room conditions.
  - **T3, RH_3:** Another specific room.

  This pattern continues for nine areas or zones.
- **T_out:** Outdoor temperature.
- **Press_mm_hg:** Atmospheric pressure in millimeters of mercury.
- **RH_out:** Outdoor relative humidity.
- **Windspeed:** Wind speed, possibly measured in m/s or km/h.
- **Visibility:** Likely the visibility distance, in kilometers or meters.

- **Tdewpoint:** Dew point temperature, indicating the temperature at which air becomes saturated with moisture.
- **rv1, rv2:** These could represent residual values, random variables, or sensor readings used for testing purposes. They are less clear without further context.

(a) Floor 1 (b) Floor 2

Figure 1: Top-Down view of the apartment

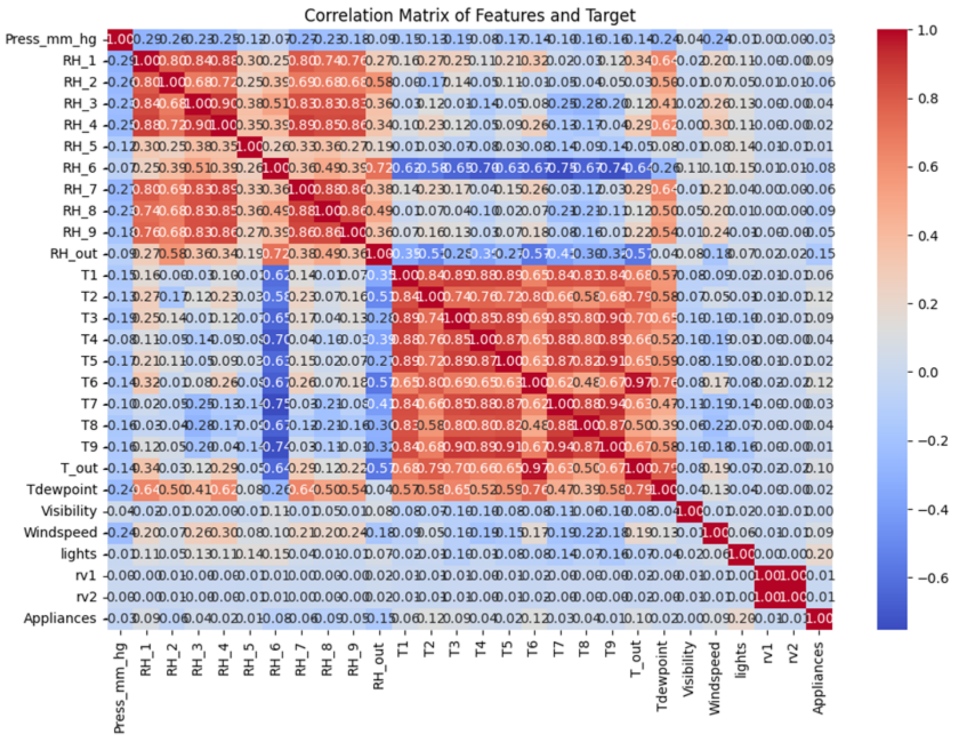


Figure 2: Correlation Matrix of features and targets

0.2 Models Used

To predict the energy usage, the authors of the paper used regression, neural networks, time series, and more.

Specifically, they used 4 models : Random Forests, Support Vector Machines, GBM and Linear Regression.

Machine Learning Models: Summary

1. **Random Forests (RF)**
   A tree-based ensemble method that uses bagging to improve performance.

   - **Strengths:**
     - Handles non-linear relationships and interactions well.
     - Robust to overfitting due to ensemble averaging.
     - Performs well with both categorical and continuous data.
     - Handles missing data and outliers effectively.
   - **Intricacies:**
     - Can be computationally expensive for large datasets.
     - Requires tuning (e.g., number of trees, max depth) for optimal performance.
     - May struggle with extrapolation (e.g., predictions outside the training data range).

2. **Support Vector Machines (SVM)**
   A model that constructs hyperplanes or decision boundaries to classify data or make regressions.

   - **Strengths:**
     - Works well for high-dimensional data.
     - Effective for small to medium-sized datasets.
     - Kernel trick allows for flexible modeling of complex patterns (e.g., linear, RBF kernels).
     - Can be robust to outliers if using soft-margin parameters.
   - **Intricacies:**
     - Computationally intensive for large datasets.
     - Requires careful tuning of hyperparameters (e.g., C, kernel, gamma).
     - Less interpretable than simpler models like linear regression.

3. **Gradient Boosting Machines (GBM)**
   An ensemble method that builds models sequentially, with each step reducing residual errors.

   - **Strengths:**
     - Highly flexible and accurate, often state-of-the-art for tabular data.
     - Handles non-linear relationships well.
     - Works effectively with missing data and various variable types.
     - Feature importance metrics aid interpretability.
   - **Intricacies:**
     - Prone to overfitting if not tuned properly (e.g., learning rate, max depth).
     - Computationally expensive compared to simpler models.
     - Sensitive to outliers unless handled during preprocessing.

4. **Linear Regression**
   A simple model that assumes a linear relationship between inputs and the target variable.

   - **Strengths:**
     - Fast, interpretable, and easy to implement.
     - Works well for linearly separable data or datasets with few features.
     - Provides insight into feature importance through coefficients.
   - **Intricacies:**
     - Assumes linearity, independence, homoscedasticity, and normality of errors.
     - Sensitive to outliers and multicollinearity.
     - Limited in capturing non-linear relationships.

0.2.1 Random Forests

0.3 Paper Results Explanation

Some important observations include some energy use patterns like for example that appliances contribute significantly to the residential energy load, reaching 30 percent of total usage in some cases. Furthermore, while devices like refrigerators show a stable and consistent load in the network, other devices like dishwashers for example vary depending on the day or time of the day.

Some additional observations are that (as maybe expected) the number of current occupants in the household is connected to higher energy consumption, the number of appliances owned (entertainment systems, tv, etc.) seem to also affect it. Buildings that are insulated to a good degree underscore the importance of thermal gains from appliances in energy balance.

We also notice High Pairwise Correlations between some variables. Like for example T1 and T2 seem to have close values in many cases indicating possibly that the measurement of just one variable could suffice for modelling.

The paper results indicate a positive correlation between :

- **Appliances–Lights:** Correlation coefficient = 0.19.
- **T1–T2:** Correlation coefficient = 0.84.
- **RH_1–T2:** Correlation coefficient = 0.27.
- **RH_1–RH_2:** Correlation coefficient = 0.8.
- **T2–T3:** Correlation coefficient = 0.74.
- **RH_1–RH_3:** Correlation coefficient = 0.68.

Those results indicate redundancy of some available data such as RH2 (living room) and RH1 (kitchen), meaning that we can skip the inclusion of some, in order to save resources.

The same goes for T1,T2,T3, etc. as suggested by the authors of the paper.

In the third picture, we can see that the same goes for `T_out` and `Tdewpoint`, `T_out` and `NSM`, `T_out–T6`, `Tdewpoint–T6`.

- Room-specific humidity levels correlate with appliance consumption. For instance:
  - **Positive correlations** in high-use areas (e.g., `RH1`/kitchen: 0.06; `RH3`/laundry: 0.04) likely reflect activity such as cooking or washing.
  - **Negative correlations** in the bedrooms and the living room can indicate lower energy usage during rest or low-activity periods.

Weather and Energy Patterns: Atmospheric pressure negatively correlates with wind speed (-0.23), and both factors influence occupancy and appliance usage. Lower pressure and higher wind speeds often correspond to higher energy use, likely due to increased indoor occupancy during poor weather

0.4 Some additional notes

The GBM and RF models improve the RSME and R2 of predictions compared to the SVM-radial and mul- tiple linear regression lm. For all the models, the time information (NSM) was ranked as the most important to predict the appliances' consumption.

The weather data from the nearby weather station was shown to increase the prediction accuracy in the GBM models. The GBM models with only weather data ranked the pressure as the most important weather variable, followed by the outdoor temperature, dew point temperature, outdoor relative humidity, wind speed and visibility.

The paper suggest that the possible explanation for why the pressure has a strong prediction power may be related to its influence on the wind speed and higher rainfall probability which could potentially increase the occupancy of the house. Research found that atmospheric pressure is highly correlated with the cooling degree minutes (CDM) and heating degree minutes (HDM). Also, pressure has direct effects on air humidity ratio, density and enthalpy.

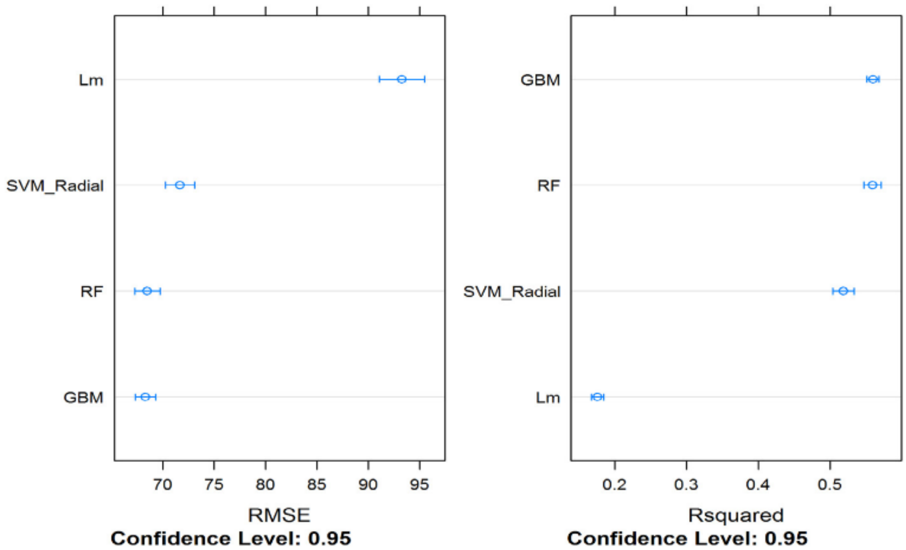


Fig. 14. Trained models comparison of RMSE and *R*-squared values.

Choosing Models

Models: XGBoost, GRU, Transformers

1. **XGBoost (Extreme Gradient Boosting)**
   A highly efficient and scalable implementation of gradient boosting for structured/tabular data.
   - **Key Features:**
     - Uses ensemble learning by building decision trees sequentially.
     - Optimized for speed and performance through parallel processing and regularization techniques.
     - Supports missing data handling and built-in cross-validation.
   - **Strengths:**
     - High accuracy for tabular datasets.
     - Robust to overfitting when properly tuned (e.g., learning rate, tree depth).
     - Provides feature importance metrics for interpretability.
   - **Use Cases:** Tabular data, competitions (e.g., Kaggle), structured datasets.
2. **GRU (Gated Recurrent Unit)**
   A type of recurrent neural network (RNN) designed to handle sequential data by solving the vanishing gradient problem.
   - **Key Features:**
     - Uses gating mechanisms (update and reset gates) to decide what information to keep or discard.
     - Simpler and computationally less expensive than LSTMs (Long Short-Term Memory).
   - **Strengths:**
     - Efficient for processing sequential data like time series or text.
     - Captures long-term dependencies in sequences.
   - **Use Cases:** Time-series forecasting, natural language processing (NLP), speech recognition.
3. **Transformers**
   A neural network architecture designed to process sequential data without relying on recurrence.
   - **Key Features:**

- Uses attention mechanisms (especially self-attention) to focus on relevant parts of the input sequence.
  - Highly parallelizable, allowing faster training than RNNs.
  - Basis of large-scale models like BERT, GPT, and other state-of-the-art NLP systems.
- **Strengths:**
  - Handles long-range dependencies in sequences effectively.
  - Scalable and flexible for various domains (NLP, vision, etc.).
- **Use Cases:** NLP (e.g., language translation, summarization), computer vision (e.g., Vision Transformers), time-series forecasting.

Table

Model	Key Features	Strengths
XGBoost	Gradient boosting, ensemble learning	Accurate, interpretable, scalable
GRU	Sequential gating, no recurrence issues	Efficient, captures long-term dependencies
Transformers	Self-attention, parallelizable	Handles long dependencies, scalable

Data Collection and Preprocessing

A smart way to know what features from the dataset is the least usefull and dont have predictive power is by using the Boruta Package.

The Boruta package compares importance of attributes with importance of shadow attributes that are created by shuffling orig- inal ones. As can be seen in Fig. 11, the Boruta algorithm is capable of detecting the two random variables (boxplots in red) that have no predicting power for the appliances' energy consumption.

The algorithm also ranks the variables in order of importance starting with the NSM variable, to the least important, the Week- Status variable.

To test how many variables are optimal to minimize the RMSE the recursive feature elimination (RFE) is used to select the optimal input.

Running these packages produces this :

Feature Selection Results

Selected Features by Boruta

`lights, T1, RH_1, T2, RH_2, T3, RH_3, T4, RH_4, T5, RH_5, T6, RH_6, T7, RH_7, RH_8, RH_9, Press_mm_hg, RH_out, Windspeed, Tdewpoint`

Running RFE for Feature Selection

Optimal Features by RFE

`lights, T1, RH_1, T2, RH_2, T3, RH_3, T4, RH_4, T5, RH_5, T6, RH_6, T7, RH_7, RH_8, RH_9, Press_mm_hg, RH_out, Windspeed, Tdewpoint`

Model Performance

RMSE with Optimal Features: **67.81**

Transformers Presentation

Core Features

- **Self-Attention Mechanism:** The self-attention mechanism allows the model to weigh the importance of different words (or tokens) in a sequence relative to one another. This capability enables the model to capture long-range dependencies more effectively than recurrent architectures.

- **Scalability:** Transformers parallelize computation, making them faster to train compared to RNNs or LSTMs. This scalability has been pivotal in training large models like GPT and BERT.
- **Flexibility:** While initially designed for text, Transformers have been adapted for a wide range of tasks:
  - **NLP:** Models like GPT, BERT, and T5 dominate language-related tasks like translation, summarization, and question answering.
  - **Vision:** Vision Transformers (ViTs) apply the architecture to image classification and related tasks.
  - **Multimodal Learning:** Models like CLIP and DALL-E combine text and vision.

Key Strengths

- **State-of-the-Art Performance:** Transformers have set benchmarks in a variety of applications, including conversational AI and generative tasks.
- **Pretrained Models:** Transfer learning with large-scale pretrained transformers (e.g., BERT, GPT) reduces the data and computational burden for downstream tasks.
- **Versatility:** Transformers are effective across tasks and modalities due to their flexible architecture.

Limitations

- **Resource Intensive:** Training transformers demands substantial computational power and memory, making them less accessible for smaller organizations.
- **Data-Hungry:** Large datasets are often required to unleash their full potential.
- **Interpretability:** Despite advances, their decision-making processes can be opaque, raising challenges for explainable AI.

Transformers and Time series

Why use transformers for time series prediction? Time series prediction involves forecasting future values based on historical data. Traditional models like ARIMA, LSTMs, and GRUs struggle with, for instance, long-term dependencies, multi-variate high-dimensional data and complex non-linear patterns

Transformers for time series are adapted to handle numerical data by modifying their architecture. The raw time series data is converted into a sequence of vectors using techniques like positional encoding or time embedding whilst capturing categorical or auxiliary features like seasonality. The encoder captures patterns and relationships in the historical data, and the decoder predicts future time steps based on the encoded representations. The self-attention mechanism identifies dependencies within the time series and cross-attention (in the decoder) links past and future data.

Transformers in Time Series can be used in forecasting, for example in finance or weather predictions, in anomaly detection in data sequences, in traffic flow prediction like optimizing transport and logistics by forecasting traffic patterns.

They offer a revolutionary approach to time series prediction, unlocking new possibilities in accuracy and scalability. With their ability to handle complex dependencies, they are rapidly becoming a go-to choice for challenging forecasting tasks.

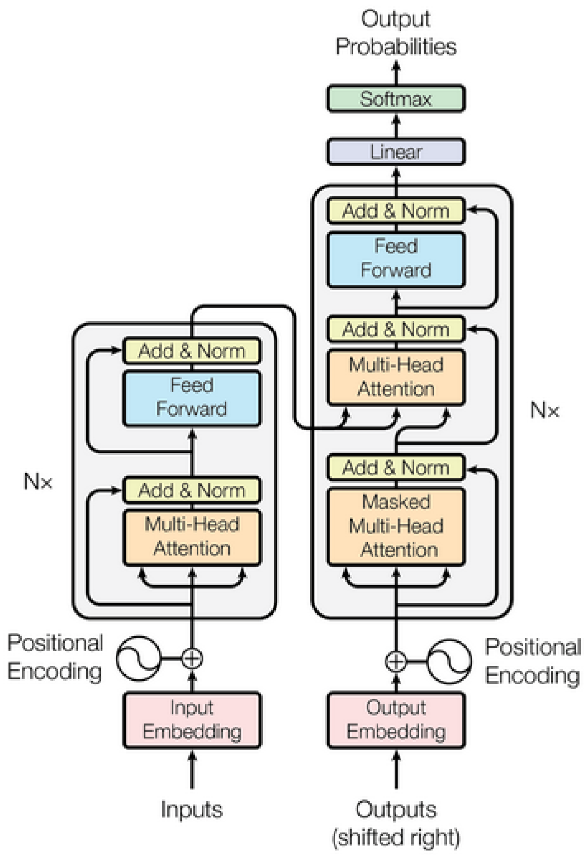


Figure 3: Transformer Architecture

Results of our Study

What metrics are mostly used and why?

- **Mean Absolute Error (MAE)** is a metric that measures the average magnitude of the errors between the predicted and actual values. It's a simple way of measuring the accuracy of a model.
- **Mean Squared Error (MSE)** is a metric that measures the average squared magnitude of the errors between the predicted and actual values. Squaring the difference between those two values helps with emphasizing the impact of large prediction errors, as they are highly affected by this process.
- **Root Mean Squared Error (RMSE)** is a metric that measures the square root of Mean Squared Error. It measures the standard deviation of residuals.

The lower value of these metrics implies higher accuracy.

Out of these 3, the most convenient seems to be RMSE mainly because of the better interpretable scale of the result.

XGBoost seems to be performing the best of the three even with changing parameters to improve RMSE.

Even when using the features selected by boruta the result seems to be the same (with better performances in general, but XGBoost remains the best of the three).

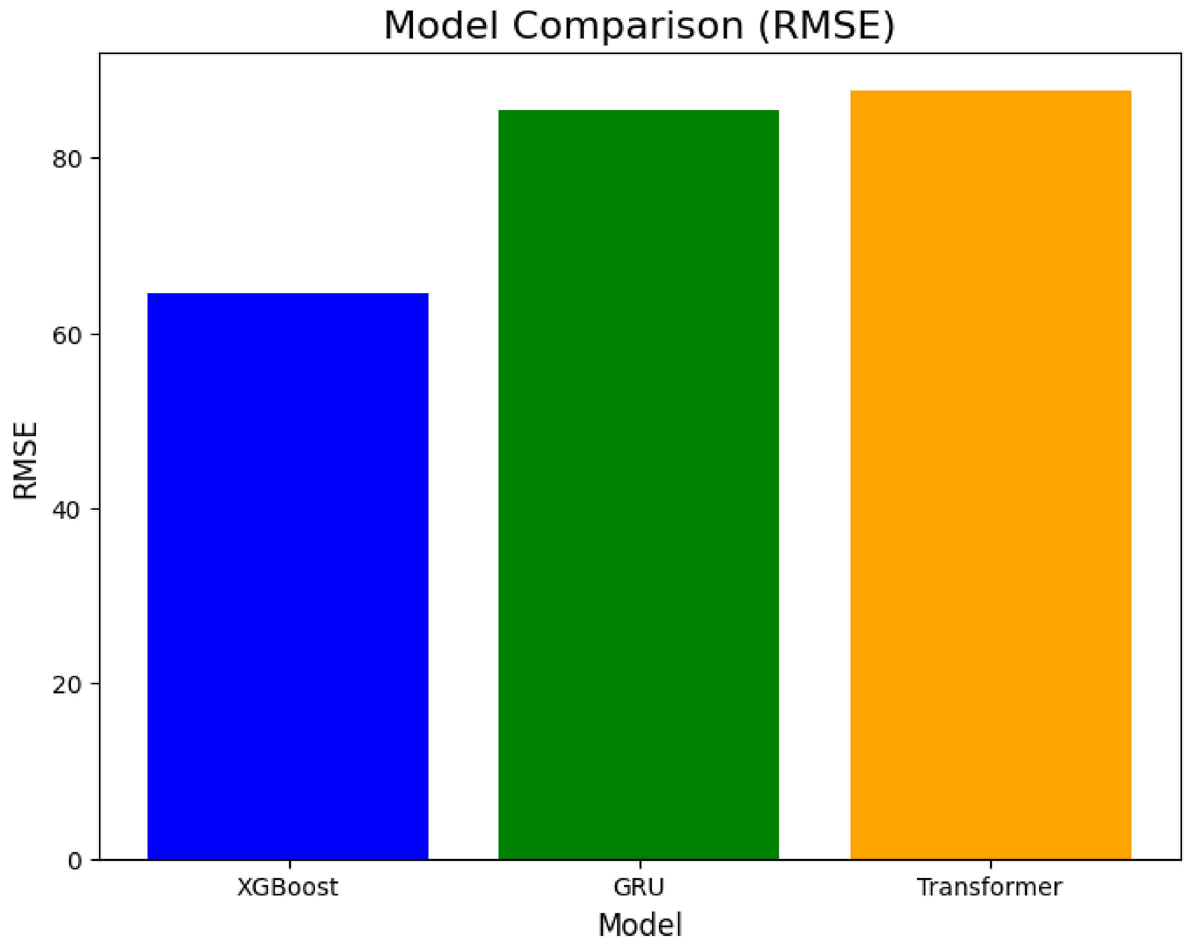


Figure 4: Performance with only Boruta selected Features used in the models

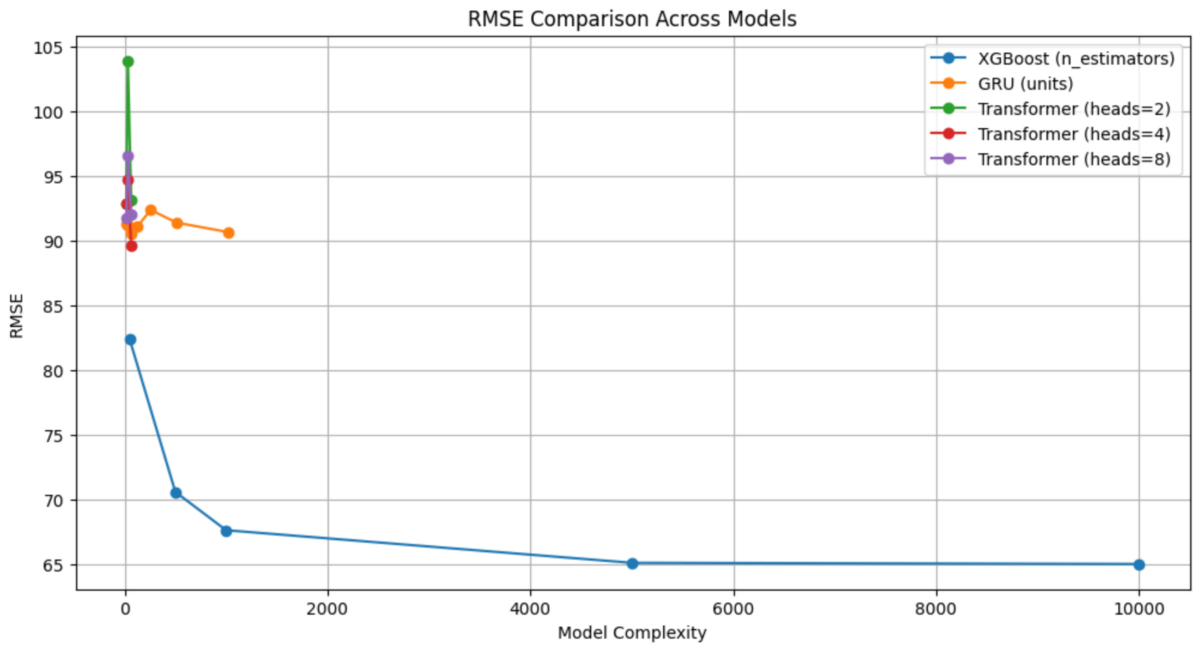


Figure 5: Performance metrics comparison with the appropriate changes in order to maximize performance

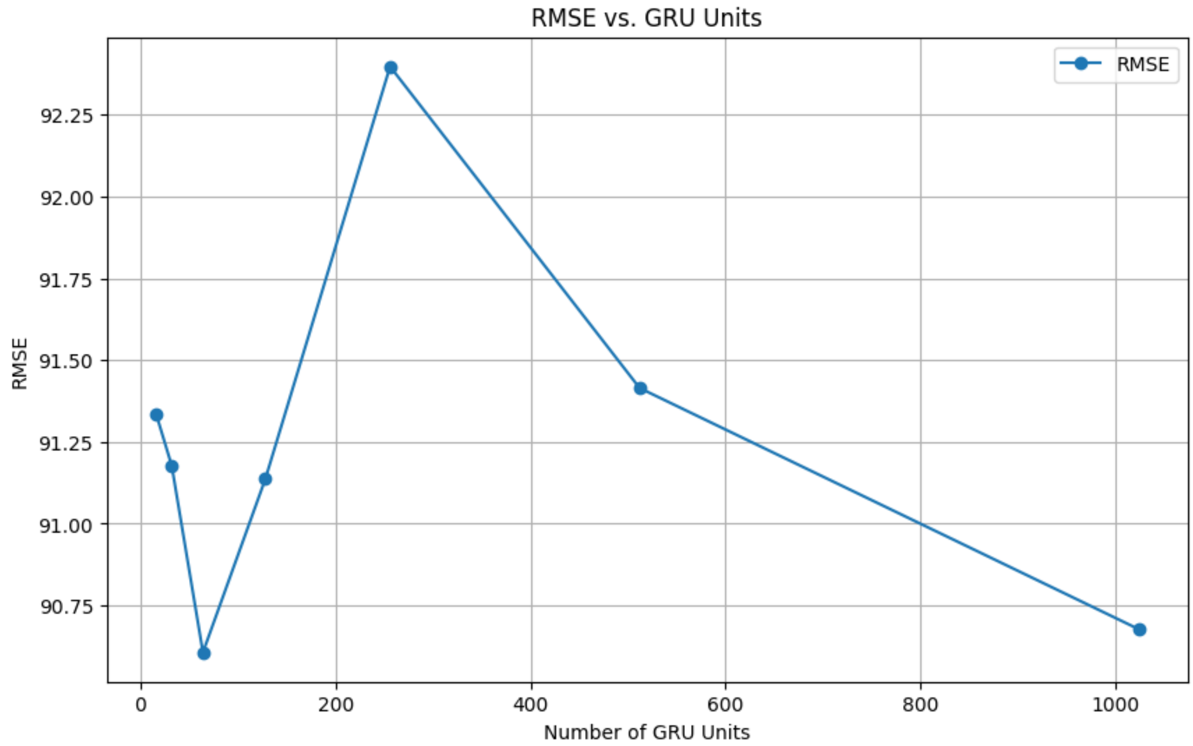


Figure 6: GRU Performance

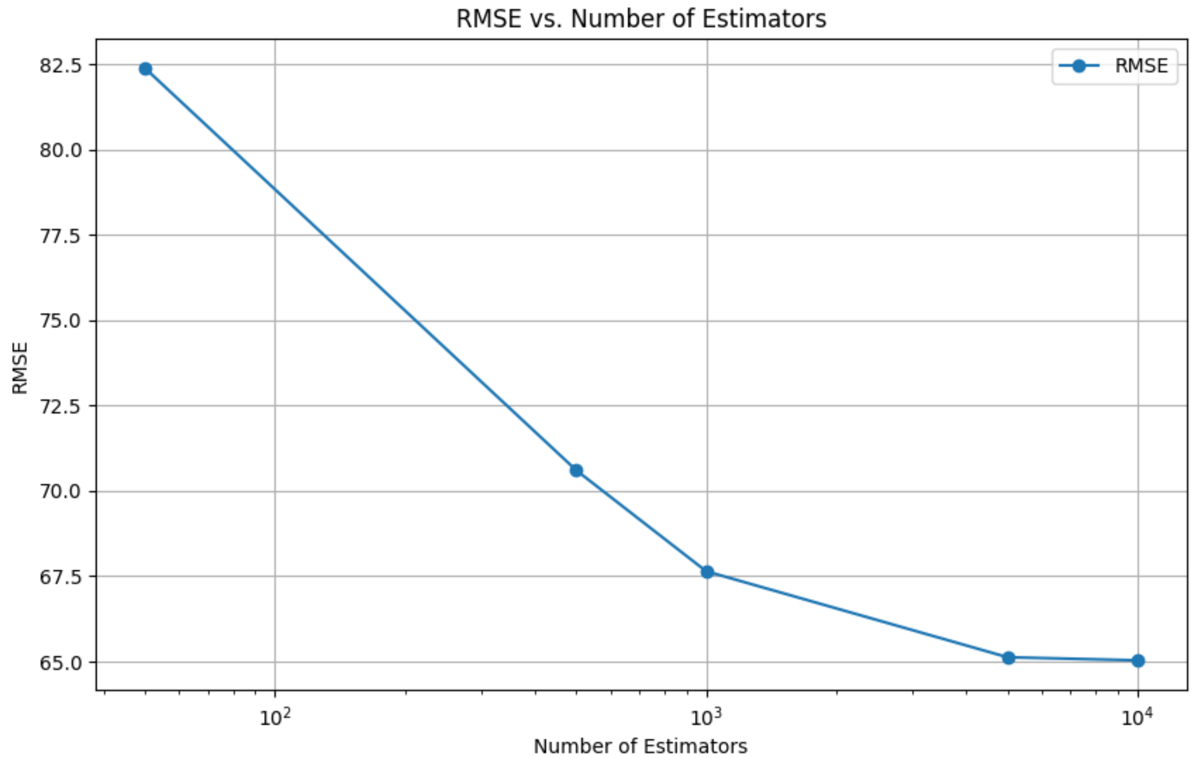


Figure 7: XGBoost - Changing the number of Estimators

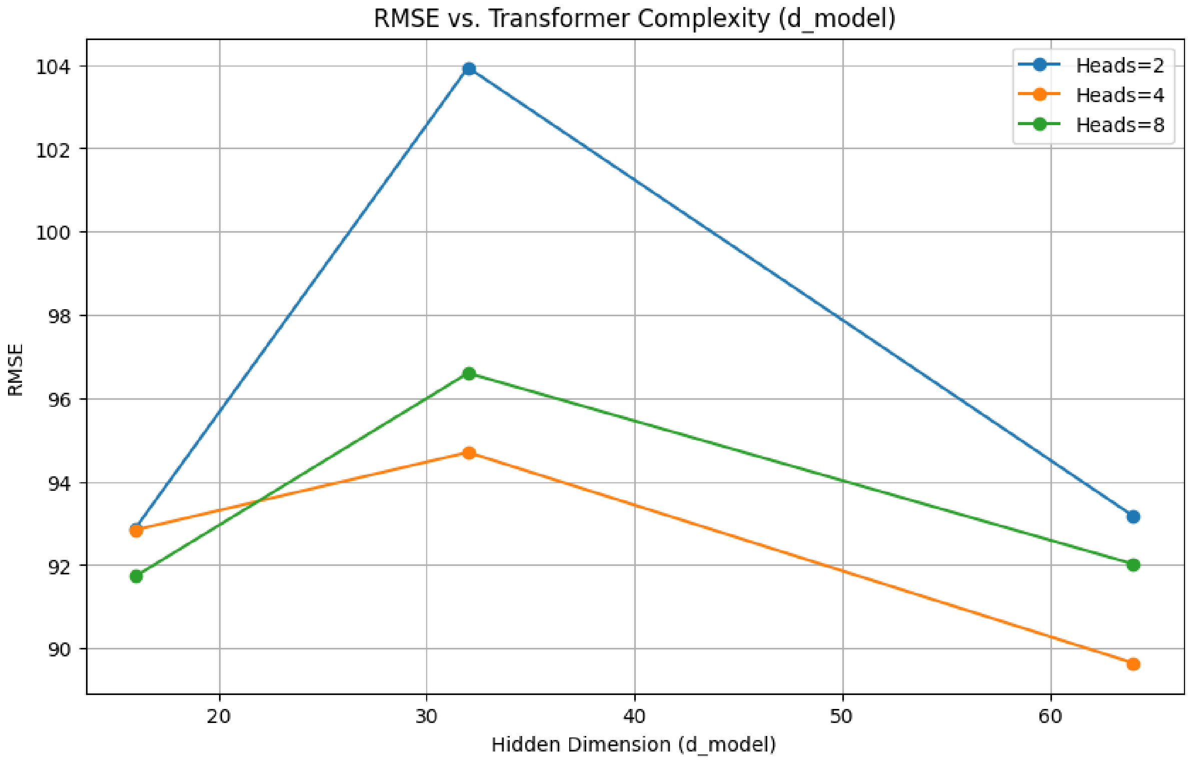


Figure 8: Transformers Performance - Changing due to Heads

Conclusion

The results suggest that RMSE provides a better understanding of model performance due to its interpretable scale. Based on the comparison of metrics, Transformer appears to perform better than other models for this dataset.

In order to perform a better test and evaluate the XGBoost performance further, we run a test by changing the number of estimators each time and monitoring the result.

Our results and judging from the SHAP plot seem on par with the papers results. Quoted below

”The weather data from the nearby weather station was shown to increase the prediction accuracy in the GBM models. The GBM models with only weather data ranked pressure as the most important weather variable, followed by the outdoor temperature, dew point temperature, outdoor relative humidity, wind speed, and visibility. The possible explanation for why the pressure has a strong prediction power may be related to its influence on the wind speed and higher rainfall probability which could potentially increase the occupancy of the house. Research by [67] found that atmospheric pressure is highly correlated with the cooling degree minutes (CDM) and heating degree minutes (HDM). Also, pressure has direct effects on air humidity ratio, density and enthalpy. Data from a wireless sensor network that measures humidity and temperature has been proven to increase the prediction accu- racy. The data analysis showed that data from the kitchen, laundry room, living room and bathrooms had the most important con- tributions. Data from the other rooms also helps in the prediction. When looking at the appliances in each room in Table 1, it can be seen that the laundry, kitchen and living rooms would be expected to have the highest contributions because of the equipment present (see also Fig. 1). The prediction of appliances’ consumption with data from the wireless network indicates that it can help to locate where in a building the main appliances’ energy consumption con- tributions are found. When using all the predictors the light consumption was ranked highly. However, when studying different predictor subsets, removing the light consumption appeared not to have a significant impact. This may be an indication that other features are correlated well with the light energy consumption. This study has found curious relationships between variables. Future work could include considering weather data such as solar radiation and precipitation.”

Below you can see the SHAP Plot and Bar plot showing the mean absolute SHAP values for each variable.

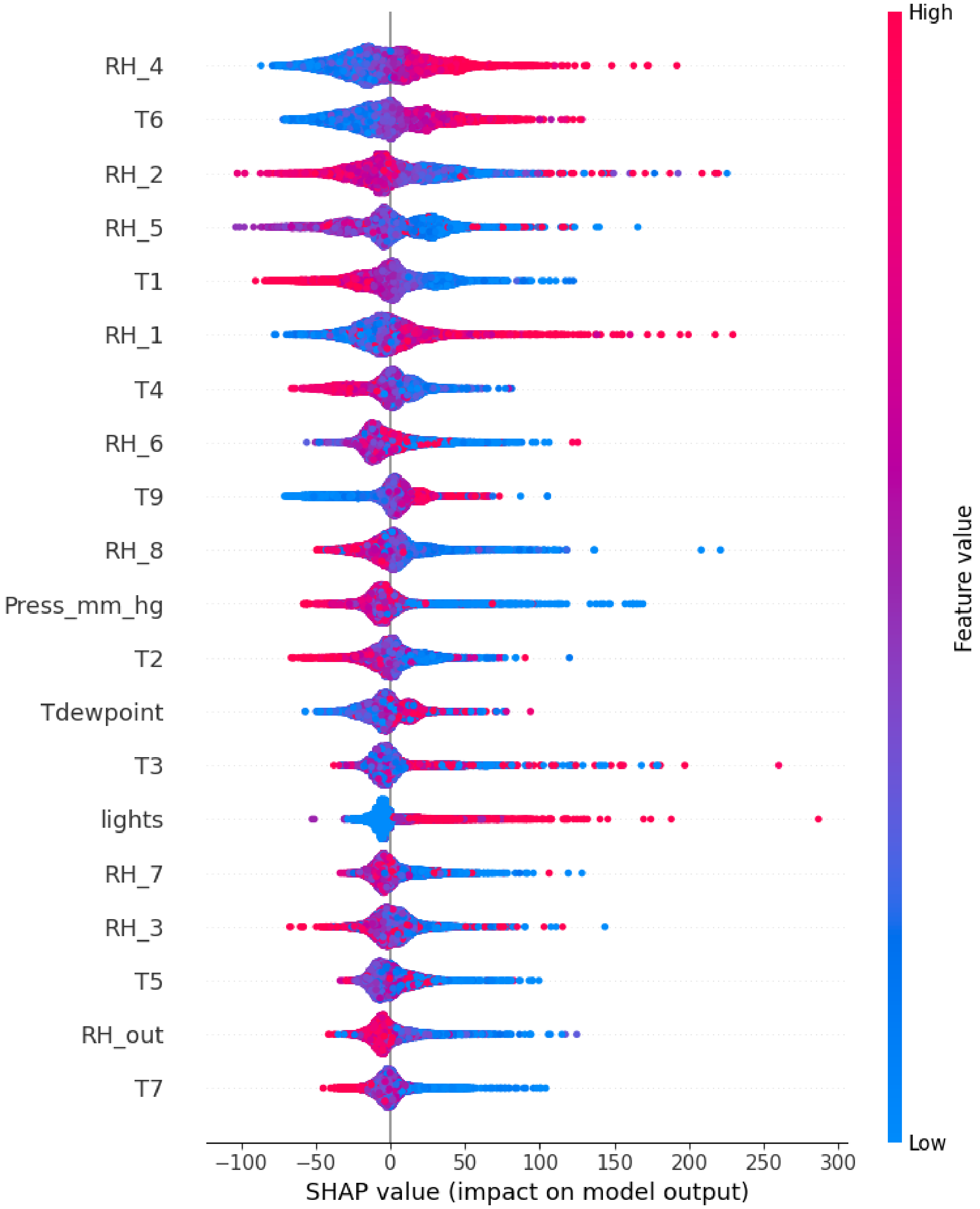
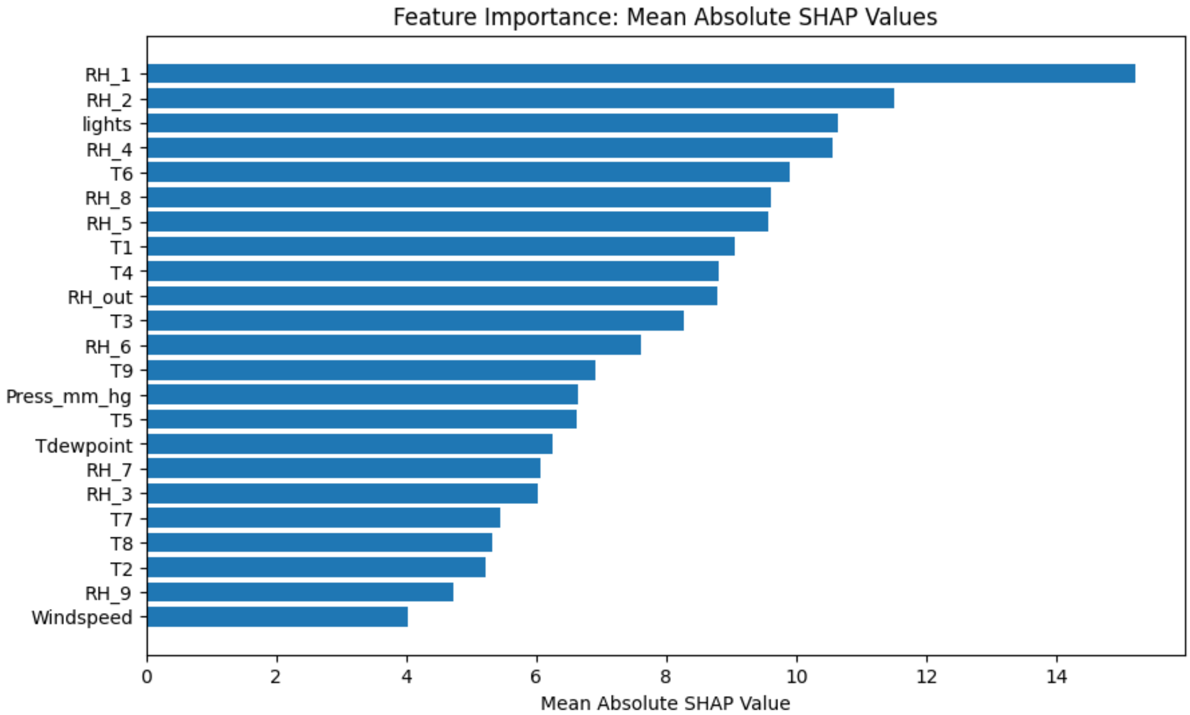


Figure 9: Feature Importance from SHAP Plot



References

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.*

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need.*

[3] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Amodei, D. (2021). *Learning Transferable Visual Models From Natural Language Supervision (CLIP).*

[4] Stevens, E., Antiga, L., & Viehmann, T. (2020). *Deep Learning with PyTorch.* Manning Publications.

[5] Tunstall, L., von Werra, L., & Wolf, T. (2022). *Natural Language Processing with Transformers.* O'Reilly Media.

[6] Hugging Face. *Transformers Documentation.*

[7] PyTorch Team. *PyTorch Documentation.*

[8] Hugging Face Blog. *Fine-tuning BERT for Sentiment Analysis.*

[9] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). *IMDb Dataset for Sentiment Analysis.*

[10] Hugging Face. *Transformers GitHub Repository.*

[11] Scikit-learn. *Precision, Recall, and F1-Score Metrics.*

[12] PyTorch Team. *Mixed Precision Training with PyTorch.*