ARTIFICIAL NEURAL NETWORKS
EHB 420E

Lecturer: Assoc. Prof. Dr. Onur Ergen

Author: Uysal Demirci

Istanbul, 2024

Data Preprocessing

The dataset used in this project initially contained hourly weather data. Since the task required daily temperature prediction, the first step involved aggregating the hourly data to daily data. This aggregation was crucial for aligning the features with the target variable, which was the daily temperature. Once the data was resampled to a daily frequency, the next task was to handle missing values. Missing values can introduce bias and reduce the effectiveness of a machine learning model, so I carefully examined each feature for missing data. Features with a high percentage of missing values were discarded, as they would not contribute meaningfully to the model's performance. Recognizing the significance of temporal patterns in weather data, I created new features based on the date, such as the month and season. These features capture important seasonal variations that influence temperature. Given that months and seasons are cyclical in nature, I applied sine and cosine transformations to encode these cyclic patterns. This transformation allows the model to better understand the continuity of time-related features, as the end of one cycle (e.g., December) is closely related to the beginning of the next cycle (e.g., January).

Modeling Approach

To build a robust predictive model for weather forecasting, I started with an Artificial Neural Network (ANN) structure. The initial goal was to experiment with a flexible architecture that could capture non-linear relationships in the data. Various parameters such as the number of layers, neurons, and activation functions were tuned to enhance model performance. While the ANN provided reasonable results, I sought to explore additional models to improve predictive accuracy and interpretability. Subsequently, I experimented with ensemble models, such as Random Forest and XGBoost, which are well-suited for handling structured data. These models excel at capturing complex patterns and feature interactions due to their tree-based structures. However, despite their advantages, I observed that these models introduced complexity without significant improvement in predictive performance for this task. After testing several approaches, I found that a Linear Regression model yielded the best results for this dataset. Linear Regression is a simple yet effective model for weather forecasting tasks when the relationships between features and the target variable are primarily linear. Given the nature of the dataset, where the features like temperature, wind speed, and seasonal patterns exhibit relatively stable and predictable trends, Linear Regression provided strong generalization with minimal risk of overfitting. Furthermore, its interpretability allowed for a clearer understanding of how each feature influenced the predictions.

Feature Optimization and Interpretation

Once the model was established, I turned my focus to feature optimization to further enhance performance. To understand the relative importance of each feature in the prediction process, I applied Permutation Importance, a model-agnostic interpretation method. This approach is particularly suitable for evaluating feature importance in a simple model like Linear Regression because it measures the impact of randomly shuffling each feature on the model's performance.

Unlike other methods tied to specific models (e.g., tree-based feature importance), Permutation Importance provided a more objective assessment of how much each feature contributed to the predictions. Using the insights from the feature importance analysis, I identified features that had minimal impact on the model's performance. Since the feature set was relatively small, I manually removed less significant features one by one and re-evaluated the model to observe changes in performance metrics. This iterative process helped identify the optimal subset of features that provided the most accurate predictions. Finally, I reapplied Permutation Importance to the refined feature set to confirm the contribution of each feature in the optimized model. This step validated that the selected features aligned well with the model’s predictive power and provided a solid foundation for future improvements or extensions to the model.

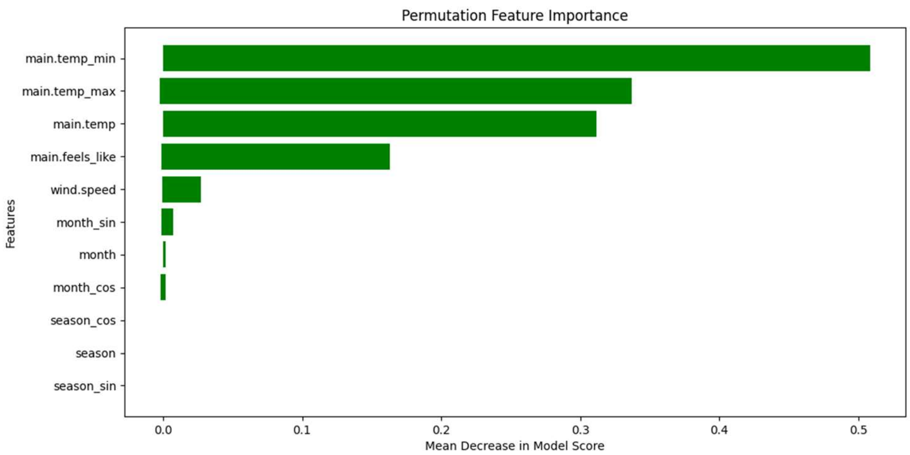

Fig.1: Feature Importance Table

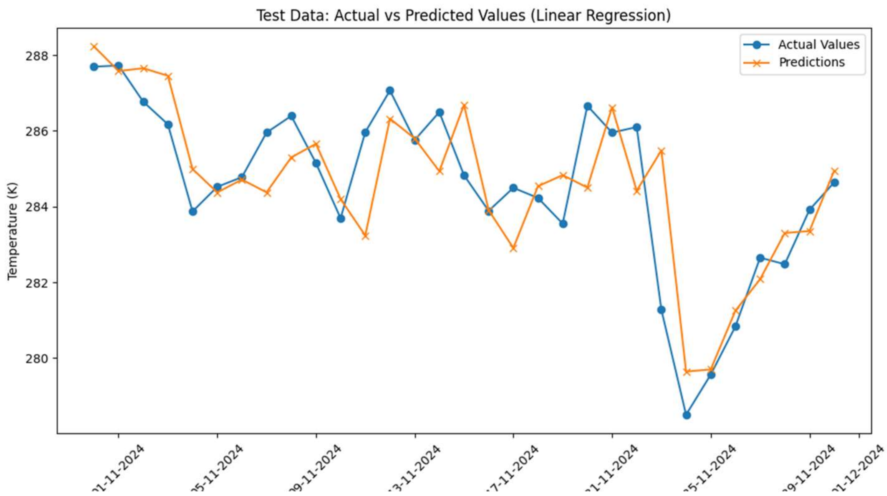

Fig.2: Real vs Predicted Temperature

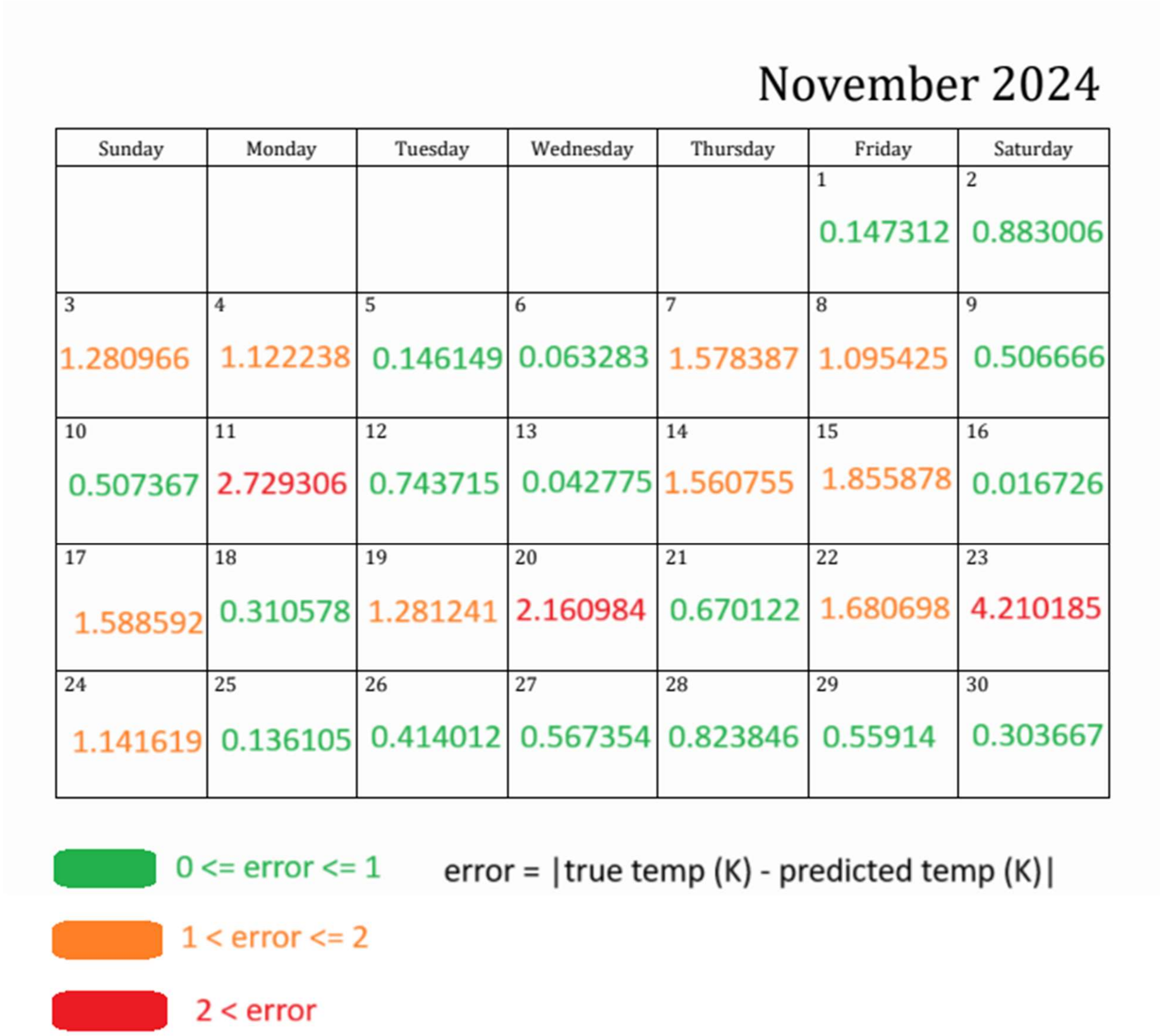

Fig.3: 2024 November Calendar with error values between real and predicted temperatures.