Phase 4 Report: Model Evaluation

Artur Zagitov, Ildar Zalialiev, Artem Nazarov

Contents

1 Chapter 5: Model Evaluation

1.1 Model Validation Report

The trained MLP model was evaluated on a test dataset (10,000 samples) not used during training. The MLP model achieved the following performance:

- **MAE:** 0.1622

- **MSE:** 0.0586

- **RMSE:** 0.2420

These results indicate that the model meets the project's initial success criterion of achieving an RMSE of 0.25 or lower.

1.1.1 Giskard Validation and Vulnerability Analysis

While the overall performance is promising, Giskard validation revealed some areas for potential improvement within the model's predictions. These areas highlight specific data slices where the model's performance slightly deviates from the global average:

1. Category-Specific Performance Insights:

- **Children's Products and Toys:** The model exhibits a moderately higher MSE (+91.21%) for ads classified as "Children's products and toys" compared to other categories. This suggests difficulty capturing the nuances of pricing and demand within this category, possibly due to factors like brand loyalty, seasonality, or a wider price range.

- **Transport:** Similarly, the "Transport" category displays a higher MAE (+53.56%) compared to the global average. This discrepancy might be due to the model's inability to account for variations in vehicle condition, brand reputation, or fluctuations in the used car market.

2. Feature Engineering Refinement Opportunities:

- **PCA Components:** Minor performance variations are observed within specific ranges of certain PCA components derived from the title and description data (`title_pca_7`, `description_pca_13`, `description_pca_9`, `description_pca_12`, `description_pca_2`, `description_pca_3`, and `description_pca_15`). This sensitivity to engineered features suggests that the PCA transformation might not be optimally capturing the underlying information or that these components are correlated with other features in a way that introduces bias. A more complex approach to text data may be needed, such as usage of RNNs, to fully capture the underlying information.

3. Further Exploration of Missing Value Imputation:

- **Missing Parameters:** Ads with missing values in `param_2` and `param_3`, imputed as 'missing,' exhibit slightly higher MSE values (+35.37% and +28.39% respectively) compared to ads with complete data. This suggests that the model might be interpreting the 'missing' category as a strong signal, potentially leading to biased predictions for ads lacking this specific information. Maybe an alternative method of imputation might solve the issue.

- **Item Sequence Number:** A minor increase in MSE (+30.1%) is observed within a specific range of the `item_seq_number` feature. This issue requires further investigation to understand if there is a genuine relationship between the order in which a user posts ads and ad demand, or if it represents a false correlation.

1.2 Discussion

The model validation results demonstrate that the MLP model achieves the pre-defined success criteria for RMSE, indicating its potential for providing valuable demand predictions on the Avito platform. The Giskard validation, while highlighting areas for potential improvement, does not present significant concerns.

1.3 Deployment Decision

Based on the overall positive model validation results and the manageable vulnerability issues, we recommend proceeding with the deployment of the MLP model.