Model Metrics for Online Activity Prediction

Model Overview

The online activity prediction model aims to predict the online activity hours of users based on their historical activity and demographic information. The model uses the LightGBM algorithm, and we performed hyperparameter tuning using Optuna to optimize its performance.

Features Used

The model was trained on the following features:

- Age
- Gender (encoded as numerical values)
- Number of Kids
- Month
- Day of the Month
- Day of the Year
- Week of the Year
- Day of the Week
- Year
- Is Weekend (binary indicator)
- Quarter
- Is Month Start (binary indicator)
- Is Month End (binary indicator)
- Is Quarter Start (binary indicator)
- Is Quarter End (binary indicator)
- Is Year Start (binary indicator)
- Is Year End (binary indicator)
- 7-Day Rolling Window Mean of Online Hours (for various time periods)
- Lag Features for Historical Online Hours (up to 21-day lag)

Model Parameters

The LightGBM model was configured with the following parameters:

- Number of Estimators: 1000
- Learning Rate: Varies based on Optuna optimization
- Maximum Depth: Varies based on Optuna optimization
- Number of Leaves: Varies based on Optuna optimization
- Subsample for Bin: Varies based on Optuna optimization
- Minimum Child Weight: Varies based on Optuna optimization

- Minimum Child Samples: Varies based on Optuna optimization

Hyperparameter Tuning

I used Optuna to perform hyperparameter tuning for the LightGBM model. Optuna performed a series of trials to find the best combination of hyperparameters that minimized the Root Mean Square Error (RMSE) on the validation data. The optimal hyperparameters were selected based on their performance during the tuning process.

Best params:
{'n_estimators': 2200,
 'learning_rate': 0.013645872924300499,
'max_depth': 11,
 'num_leaves': 50,
 'subsample_for_bin': 273000,
'min_child_weight': 8.245740340142923e-05,
 'min_child_samples': 30}

Model Performance

The final model's performance was evaluated using RMSE on the test dataset. The optimized LightGBM model achieved an RMSE of 2.50068 on the test data. This improved performance compared to the baseline model, demonstrating the effectiveness of the engineered features and optimized hyperparameters.

Recommendations

Based on the model's performance and the features used, I recommend the following:

Deploy the optimized LightGBM model to predict online activity hours for users in real-time.
Continuously collect data and update the model to adapt to changing user behaviors.
Consider incorporating additional features, such as user engagement metrics or external factors, to further enhance the model's accuracy.
Conduct A/B testing to assess the impact of the model's predictions on business outcomes and user engagement.
Monitor the model's performance over time and reevaluate hyperparameters periodically to maintain optimal predictive capabilities.

Conclusion

The online activity prediction model using LightGBM with optimized hyperparameters successfully predicts users' online activity hours. By leveraging historical activity data and demographic features, the model provides valuable insights for resource planning and user engagement strategies. Continuous improvement and refinement of the model will ensure its continued relevance and accuracy in the dynamic online landscape.