MLOps Group No: 92

Group Members Names:

1. PEYALA SAMARASIMHA REDDY - 2023AA05072
2. PEGALLAPATI SAI MAHARSHI - 2023AA05924
3. ANIRUDDHA DILIP MADURWAR - 2023AA05982
4. K VAMSIKRISHNA - 2023AA05209

M3: Model Experimentation and Packaging

Report on Hyperparameter Tuning Results for Titanic Survival Prediction Model

Introduction:

In this segment of the project, we performed **Hyperparameter Tuning** using **Optuna** to optimize the performance of the **Random Forest Classifier** for predicting Titanic survival. The dataset consists of various features, including passenger class, sex, age, siblings/spouses aboard, parents/children aboard, fare, and embarkation port. Our goal was to improve the model's accuracy by optimizing key hyperparameters through systematic experimentation.

We employed **Optuna**, an open-source optimization framework, to tune the hyperparameters and improve the model's performance on the test data. Hyperparameter tuning is a critical part of machine learning as it helps in adjusting the model's parameters to find the best-fit configuration that yields the highest predictive accuracy.

Hyperparameters Tuned:

The following hyperparameters of the **Random Forest Classifier** were optimized:

1. **n_estimators**: The number of trees in the forest (range: 50 to 200).
2. **max_depth**: The maximum depth of each tree (range: 5 to 20).
3. **min_samples_split**: The minimum number of samples required to split an internal node (range: 2 to 10).
4. **min_samples_leaf**: The minimum number of samples required to be at a leaf node (range: 1 to 4).

These hyperparameters control the complexity and overfitting/underfitting of the Random Forest model, and tuning them ensures better generalization on unseen data.

Optimization Process:

The optimization process was carried out in the following steps:

1. **Objective Function**: We defined an objective function that:
    - Trains the Random Forest model with hyperparameters selected by Optuna.
    - Evaluates the model's accuracy on the test set.
    - Returns the accuracy as the objective value to be maximized.
2. **Optuna Study**: An Optuna study was created with the goal of finding the hyperparameters that maximize accuracy. The study ran for **10 trials** in total.
3. **Best Hyperparameters**: After completing the 10 trials, we identified the best hyperparameters for the Random Forest model. The following hyperparameters were found to provide the best results:

- **n_estimators**: 70
- **max_depth**: 18
- **min_samples_split**: 8
- **min_samples_leaf**: 1

Hyperparameter Tuning Trials and Results:

Here is a summary of the results from the **10 trials** conducted by Optuna:

Trial	n_estimators	max_depth	min_samples_split	min_samples_leaf	Accuracy
1	79	13	2	3	0.8379
2	64	12	5	1	0.8436
3	125	18	10	4	0.8156
4	70	18	8	1	0.8547

5	143	13	9	3	0.8212
6	124	17	4	4	0.8156
7	73	13	5	1	0.8436
8	84	12	7	3	0.8268
9	130	12	2	3	0.8380
10	124	9	3	1	0.8380

Best Hyperparameters:

After the 10 trials, the best hyperparameters found were:

- **n_estimators**: 70
- **max_depth**: 18
- **min_samples_split**: 8
- **min_samples_leaf**: 1

These values were selected because they resulted in the highest accuracy **(0.8547)** among all the trials.

Final Model Accuracy:

Using the best-found hyperparameters, we trained the **Random Forest Classifier** and evaluated it on the test set. The accuracy of the best model was:

Best Model Accuracy: **0.8547**

Conclusion:

This indicates a significant improvement in model performance compared to the default settings of the Random Forest model. The hyperparameter tuning process has helped us achieve a more accurate model for Titanic survival prediction.