EXPERIMENT REPORT

Student Name	Rohan Rocky Britto
Project Name	Assignment 1 Week 3
Date	01/09/2023
Deliverables	Britto_Rohan-24610990-week3_adaboost_optimization.ipynb Adaboost Algorithm Github Link: https://github.com/rohanbrit/Adv_ml_asgn1

---

1. EXPERIMENT BACKGROUND	
Provide information about the problem/project such as the scope, the overall objective, expectations. Lay down the goal of this experiment and what are the insights, answers you want to gain or level of performance you are expecting to reach.	
1.a. Business Objective	The NBA draft is an annual event in which teams select players from their American colleges as well as international professional leagues to join their rosters. The task is to build a model that will predict if a college basketball player will be drafted to join the NBA league based on his statistics for the current season. An incorrect prediction can impact the reputation of the organization that uses them.
1.b. Hypothesis	Now that we know that AdaBoost is performing the best for the given dataset, I will be optimizing its performance by using feature selection and automated hyperparameter tuning.
1.c. Experiment Objective	My main objective for this experiment is to try and reduce the overfitting of the model so that we can achieve better results with Kaggle test dataset.

2. EXPERIMENT DETAILS	
Elaborate on the approach taken for this experiment. List the different steps/techniques used and explain the rationale for choosing them.	
2.a. Data Preparation	The data preparation steps were performed in the previous experiments and the saved dataset has been used in this experiment. Hence, no data preparation steps were performed.
2.b. Feature Engineering	I have checked the most important features for the model using the feature_importance attribute and eliminated those that have least importance. Other than this, features like 'Rec_Rank', 'dunks_ratio', 'pick' could be important from a future perspective, and we should check if we can get the actual values for these features from the business.
2.c. Modelling	AdaBoost performed the best in the previous experiment with slight overfitting. For this reason, I have decided to perform automated hyperparameter tuning and feature selection to improve its performance and reduce overfitting.

3. EXPERIMENT RESULTS	
Analyse in detail the results achieved from this experiment from a technical and business perspective. Not only report performance metrics results but also any interpretation on model features, incorrect results, risks identified.	
3.a. Technical Performance	I used AUROC as the evaluation metric for this experiment as specified in the task description. With feature selection, I was able to achieve similar scores as compared to using the original dataset, i.e., 0.9965 for the training set and 0.9960 for the validation set. Hence, I decided to drop the features with zero feature importance. With hyperparameter tuning, I was able to achieve a score of 0.9864 on the testing set.
3.b. Business Impact	The results achieved after feature selection and hyperparameter tuning are quite good and we can try and implement this solution in production. We will need to monitor its performance on a timely basis.
3.c. Encountered Issues	The presence of null values in potentially important features.

---

4. FUTURE EXPERIMENT	
Reflect on the experiment and highlight the key information/insights you gained from it that are valuable for the overall project objectives from a technical and business perspective.	
4.a. Key Learning	After feature selection and hyperparameter tuning, we were able to achieve very good scores with the AdaBoost model.
4.b. Suggestions / Recommendations	As we have achieved very good scores and the overfitting is also very low, we can go ahead with this model.