EXPERIMENT REPORT: XGBoost Classifier

Student Name: Brilliant Jepkogei Kiptoo

Student ID:24699314

Project Name: NBA Draft Prediction

Date: 1st September 1, 2023

1. EXPERIMENT BACKGROUND

1. a. Business Objective

The NBA Draft's anticipation and excitement drive the need for predictive models to forecast college players' selection into the league. The objective is to build an accurate model that predicts whether a college player will be drafted, providing insights to basketball enthusiasts and teams alike.

1. b. Hypothesis

The hypothesis examined is whether an XGBoost model can leverage college basketball player statistics to accurately predict their NBA draft selection status. This advanced ensemble technique is hypothesized to effectively learn from the performance metrics and discriminating patterns between future draftees and non-draftees. Testing this machine learning approach is worthwhile given XGBoost's widespread success in modeling complex data and delivering state-of-the-art predictions across domains. Its robust generalization capabilities may aptly capture statistical differences between drafted and undrafted players.

1. c. Experiment Objective

The experiment's goal is to develop an XGBoost model with a high AUROC score, demonstrating its potential for predicting NBA draft outcomes. The expected outcome is a model that surpasses the random guessing baseline and provides valuable insights for predicting draft selections.

2. EXPERIMENT DETAILS

2. a. Data Preparation

During this particular stage, the data underwent preparation through the processes of loading and cleansing the datasets. To ensure high-quality data, any rows containing missing values in numeric features were eliminated. Only relevant attributes were retained for analysis, while the 'player_id' attribute was excluded as it did not contribute to the overall analysis. Additionally, to maintain consistency within the dataset, it was divided into training and validation subsets and standardized using StandardScaler on its numeric features.

In order to enhance future iterations of data completeness, exploring techniques such as imputation for handling missing values in categorical features is highly recommended. Furthermore, potential enhancements can be explored by experimenting with feature engineering techniques and subsets of attributes in order to gain deeper insights into predicting draft outcomes.

2. b. Feature Engineering

Feature engineering for the XGBoost model followed the same minimalist approach as the previous experiment. Existing player statistics were utilized, and no new features were introduced to keep the model's interpretability intact.

2. c. Modeling

XGBoost, a robust gradient boosting algorithm, was chosen for its ability to capture complex relationships and handle non-linear patterns effectively. Hyperparameters, including learning rate, maximum depth, and the number of estimators, were tuned for optimal performance.

3. EXPERIMENT RESULTS

3. a. Technical Performance

The optimized XGBoost Classifier model attained an AUROC score of 0.74 on the holdout validation set for predicting draft selection outcomes. This result demonstrates the model's proficient separation between the drafted and undrafted classes based on input college statistics. An AUROC exceeding 0.75 indicates that XGBoost could extract predictive signals from the performance metrics despite the challenging modeling task.

3. b. Business Impact

The achieved AUROC score suggests that the XGBoost modeldo not possesses predictive capabilities for NBA draft outcomes. The business impact of this model lies in its potential to guide player selection strategies and inform fans and commentators about draft predictions.

3. c. Encountered Issues

During the experiment, challenges included optimizing hyperparameters to avoid overfitting and ensuring model interpretability. Additionally, the model's predictions might have been influenced by unaccounted variables, leading to incorrect classifications.

4. FUTURE EXPERIMENT

4. a. Key Learning

The XGBoost Classifier's performance underscored its effectiveness in NBA draft prediction tasks. The model's ability to capture nuanced relationships in player statistics highlighted the importance of leveraging advanced algorithms for complex tasks.

4. b. Suggestions / Recommendations

To enhance the XGBoost model, further hyperparameter tuning, considering additional relevant features, and addressing potential overfitting could lead to improved performance. A comparative analysis between the Random Forest Regressor and the XGBoost Classifier could provide insights into their respective strengths and weaknesses.