EXPERIMENT REPORT

Student Name	Brilliant Jepkogei Kiptoo
Project Name	Predicting if a college basketball player will be drafted to join the NBA league.
Date	18/08/2023
Deliverables	<Google Colab> <Random Forest> <https://github.com/Brilliantkiptoo/AdvMLA-Kaggle-competion-assingment/blob/main/NBA_Prediction%20_playerFinal_notebook_Billiant_Kiptoo_24699314_AdvMLA_Lab1_Kaggle_asignment_Solutions_(2).ipynb >

EXPERIMENT BACKGROUND	
Provide information about the problem/project such as the scope, the overall objective, expectations. Lay do 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 primary business objective is to develop a predictive model that effectively identifies basketball players who have a high likelihood of being selected in the NBA draft. The mod aims to assist NBA teams, sport commentators, and fans in making informed decisions and predictions about players' draft prospects based on their current season statistics and poten contextual factors.

1.b. Hypothesis	Incorporating advanced player performance metrics such as PER (Player Efficiency Rating WS (Win Shares) will lead to improved predictive accuracy in identifying potential NBA draftees. Including historical player performance data from previous seasons will enable the model capture player development trends and enhance draft predictions.
1.c. Experiment Objective	Detail what will be the expected outcome of the experiment. If possible, estimate the goal are expecting. List the possible scenarios resulting from this experiment. We expect the outcome to come out as the players with the best features, like having the hi height, having many throws will be selected for the league.

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	Describe the steps taken for preparing the data (if any). Explain the rationale why you had perform these steps. List also the steps you decided to not execute and the reasoning behin Highlight any step that may potentially be important for future experiments The steps i have taken to prepare the data include; Data Preprocessing and feature engineering- Here I have loaded the data, converted categ data into numerical using the encoding technique and split the data into features and targe variables. Training the model using random forest classifier

	Model Evaluation by predicting the probabilities of the players being drafted on the test s then compute the AUROC score.
2.b. Feature Engineering	I identified the most relevant features that could impact a players status, after that I create Categorical features for a player positions and consider one-hot encoding There was a string in my dataset, and I had to perform encoding to remove the unwanted. Also I had to identify the column containing non numeric values and then convert the categorical string values to numeric values using hot encoding.
2.c. Modelling	Random Forest classifier is popular ensemble machine learning for both classification an regression, Random state was set to 42 to ensure reproducibilitymodel from sklearn librar been trained in my experiment. The model is trained using the training data, represented by X train _scale and y train The Y _ train contains the corresponding target labels indicating whether a player was dr (1) or not (0)

EXPERIMENT RESULTS	
Analyse in detail the results achieved from this experiment from a technical and business perspective. Not o report performance metrics results but also any interpretation on model features, incorrect results, risks iden	
3.a. Technical Performance	Score of the relevant performance metric(s). Provide analysis on the main underperform cases/observations and potential root causes. The model I used was evaluated using the AUROC metric. The function computes. The ROC curve using the y_true and predictions, it then calculates area under the Roc c AUROC values range between 0 and 1 AUROC is used for evaluating binary classification models, and it allows you to unders their ability to differentiate between classes and make informed decisions about model selection and performance improvement
3.b. Business Impact	Interpret the results of the experiments related to the business objective set earlier. Estin the impacts of the incorrect results for the business (some results may have more impact compared to others)
3.c. Encountered Issues	List all the issues you faced during the experiments (solved and unsolved). Present solut workarounds for overcoming them. Highlight also the issues that may have to be dealt v future experiments. The dataset is imbalanced, with significantly fewer positive class samples (drafted playe compared to negative class samples. We should Continuously monitor class balance and explore more advanced techniques to handle imbalance. Complex models might be challenging to interpret and explain to stakeholders. Explore advanced techniques for model interpretability and communicate findings effec

FUTURE EXPERIMENT	
Reflect on the experiment and highlight the key information/insights you gained from it that are valuable fo overall project objectives from a technical and business perspective.	
4.a. Key Learning	Reflect on the outcome of the experiment and list the new insights you gained from Provide rationale for pursuing more experimentation with the current approach or if you think it is a dead end.
4.b. Suggestions / Recommendations	Given the results achieved and the overall objective of the project, list the potential steps and experiments. For each of them assess the expected uplift or gains and ran accordingly. If the experiment achieved the required outcome for the business, recommend the steps to deploy this solution into production. The prioritization of next steps and experiments should consider the potential gain feasibility, and alignment with the project's objectives. Once a model demonstrates desired performance, deployment steps involve thorough validation, integration, an continuous monitoring to ensure successful deployment and ongoing improvement