EXPERIMENT REPORT

|  |  |
| --- | --- |
| Student Name | Brilliant Jepkogei Kiptoo |
| Project Name | Predicting if a college basketball player will be drafted to join the NBA league using Logistic\_ Random Forest Classifier |
| Date | 25/08/2023 |
| Deliverables | <Google Colab>  <Random Forest>  <https://github.com/Brilliantkiptoo/AdvMLA-Kaggle-competion-assingment/blob/main/notebook\_Kiptoo\_\_\_Brilliant\_\_\_24699314\_week\_week\_2\_\_\_logistic\_random\_forest.ipynb> |

|  |  |
| --- | --- |
| 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 objective is to leverage NBA player performance data and analytics to enhance team performance and increase fan engagement. The organization aims to achieve this by making data-driven decisions in player selection, game strategy, and training programs. The objective is specific (improving team performance and increasing fan engagement), measurable (through performance metrics and fan engagement metrics), achievable (using advanced analytics tools and techniques), relevant to the NBA team's goals, and time-bound (ongoing optimization throughout the season). |
| 1.b. Hypothesis | The performance statistics of college basketball players during their current season can be used to predict whether they will be drafted to the NBA league.  This hypothesis suggest that there is a correlation between the performance statistics of college basketball players and their likelihood of being drafted to the NBA league. The hypothesis is tested by training an Random Forest model on the provided data and evaluating its performance using metrics like AUROC score. |
| 1.c. Experiment Objective | We expect the outcome to come out as the players with the best features, like having the highest 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 to perform these steps. List also the steps you decided to not execute and the reasoning behind it. 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 categorical data into numerical using the encoding technique and split the data into features and target variables.  Training the model using Random Forest Model  Model Evaluation by predicting the probabilities of the players being drafted  on the test set and then compute the AUROC score. |
| 2.b. Feature Engineering | I identified the most relevant features that could impact a players status, after that I created  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 | 1. I did Data Preprocessing by handling missing data, encoding categorical variables 2. Identified the most important features that contribute significantly to the model's predictive power. Techniques like feature importance analysis 3. RandomForestClassifier was the algorithm for the NBA draft prediction task.   Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It's well-suited for classification tasks and handles complex relationships in the data. |

|  |  |
| --- | --- |
| 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 | Score of the relevant performance metric(s). Provide analysis on the main underperforming 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 curve. AUROC values range between 0 and 1  AUROC is used for evaluating binary classification models, and it allows you to understand 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. Estimate 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 solutions or workarounds for overcoming them. Highlight also the issues that may have to be dealt with in future experiments.  The dataset is imbalanced, with significantly fewer positive class samples (drafted players) 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 effectively. |

|  |  |
| --- | --- |
| 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 | Reflect on the outcome of the experiment and list the new insights you gained from it. Provide rationale for pursuing more experimentation with the current approach or call out 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 next steps and experiments. For each of them assess the expected uplift or gains and rank them accordingly. If the experiment achieved the required outcome for the business, recommend the steps to deploy this solution into production.  We should Continuously gather feedback from users and stakeholders to refine and improve the model's predictions and usability.  Explore collaborations with NBA teams to validate the model's predictions against actual draft selections.  Integrate historical player data, external data sources, the expert opinions to further enhance prediction accuracy.   The prioritization of next steps and experiments should consider the potential gains, feasibility, and alignment with the project's objectives. Once a model demonstrates the desired performance, deployment steps needs thorough validation, integration, and continuous monitoring to ensure successful deployment and ongoing improvements. |