EXPERIMENT REPORT

Student Name	April Nommesen
Project Name	NBA Career Prediction
Date	23/11/2022
Deliverables	notebook names: • nommesen_april-week3-randomforest2.ipynb • nommesen_april-week3-randomforest3.ipynb • nommesen_april-week3-svc1.ipynb • nommesen_april-week3-svc2.ipynb model name: • svc1 • svc2 • randomforest2 • randomforest3 GitHub repo: https://github.com/aprilgum/adv-dsi-2022-at1-grp4/tree/master/nba-career-prediction

---

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

Explain clearly what is the goal of this project for the business. How will the results be used? What will be the impact of accurate or incorrect results?

Rookie players joining in the first round of NBA draft are given four-year contracts where the first two years of the contract are guaranteed with the NBA team and the third and fourth years can change. Rookie players joining in the second round of NBA draft and undrafted players can sign contracts that can be anything from one year to four years and that are either fully guaranteed or not guaranteed at all.

The value of rookie players contract is tied to the salary cap and the team they will play for.

A rookie player who lasts for at least five years are considered successful and a player

	who does not last at least five years is considered risky. Being able to have a data-based decision to decide whether a rookie is a potential success or a potential risk greatly impacts the team's budget to pay the players' salary and also impacts the team's performance.
1.b. Hypothesis	Present the hypothesis you want to test, the question you want to answer or the insight you are seeking. Explain the reasons why you think it is worthwhile considering it, I would like to begin with the question, who are the rookie players who are likely to last at least 5 years?
1.c. Experiment Objective	Detail what will be the expected outcome of the experiment. If possible, estimate the goal you are expecting. List the possible scenarios resulting from this experiment. To try different hyperparameters for SVC and Random Forest Classifier and to correct mistakes I made in the last two weeks. I am not sure how much I can improve the model performance this week.

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

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

I have reused last week's data preparation for standardizing or not standardising data. All the SVC and Random Forest Classifier have handling of imbalanced sampling.

2.b. Feature Engineering

Describe the steps taken for generating features (if any). Explain the rationale why you had to perform these steps. List also the feature you decided to remove and the reasoning behind it. Highlight any feature that may potentially be important for future experiments.

For this week, I used all the features/independent variables. I did not transform any of the variables except resampling the target variable or standardizing.

2.c. Modelling

Describe the model(s) trained for this experiment and why you choose them. List the hyperparameter tuned and the values tested and also the rationale why you choose them. List also the models you decided to not train and the reasoning behind it. Highlight any model or hyperparameter that may potentially be important for future experiments

1. SVC All Features (SVC1)
No standardization of features
Using SVC to handle imbalanced data

```
SVC(kernel='poly', class_weight='balanced', probability=True)
```

SVC All Features (SVC2)
Standardised features – this is a "correction" for the previous experiment where I did not standardise the features
Using SVC to handle imbalanced data

```
SVC(kernel='linear', class_weight='balanced', probability=True)
```

2. Random Forest Classifier with Balanced Weights (RANDOMFOREST 2)
Standardised features
Using Random Forest Classifier to handle imbalanced data and to feature select
I experimented with the hyperparameter max_depth and settled at 7. I also experimented with criterion. There was not any difference between using log_loss and entropy.

```
RandomForestClassifier(n_estimators=2000,
                                                class_weight = 'balanced',
                                                criterion = 'log_loss',
                                                random_state=7,
                                                max_depth=7
                                               )
```

3. Random Forest Classifier with Balanced Weights (RANDOMFOREST 3)
No standardization of features
Using Random Forest Classifier to handle imbalanced data and to feature select

```
RandomForestClassifier(n_estimators=2000,
                                                class_weight = 'balanced',
                                                criterion = 'log_loss',
                                                random_state=7,
                                                max_depth=7
                                               )
```

I logged all the experiments I had in this spreadsheet: *experiments_list.xlsx*

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

Score of the relevant performance metric(s). Provide analysis on the main underperforming cases/observations and potential root causes.

Model	Accuracy	AUROC	
SVC All Features (SVC0)	0.63	0.706	This is my best output last week.
SVC All Features (SVC1)	0.57	0.706	
SVC All Features (SVC2)	0.62	0.706	My best this week

Using SCV has so far given me the best result. But I am still not satisfied with the performance so I need to do my best for the last round next week.

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.

---

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

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.

Manually experimenting with hyperparameters is time-consuming. Need to apply ways to make the process of repetition faster and less prone to making handling mistakes.

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. Next week: I would try XGBoost and hyperopt. I would also review everything from the top to eliminate the hypotheses that I do not want to test again. Analyse the impact of hyperparameter changes.