# **PROJECT REPORT**

|  |  |
| --- | --- |
| **Student Names** | Mark Brackenrig  Priyanka Srinivasa  Roger Yu |
| **Project Name** | MDSI ADSI Assignment 1 Part D |
| **Date** | 2020-03-03 |
| **Deliverables** | 1. Final report: this document 2. Zip file containing project code, model artefacts, instructions or any relevant documents |

# Business Understanding

Decide what we want the business objective to be:

1. for basketball teams to predict who is likely to last beyond 5 years, so the team can pick that player
2. for gambling
3. for a live prediction webapp

# Data Understanding

Not sure there’s much to say here, I (Roger) noted a few things during my experiments:

* The success, attempt, and percentage columns don’t seem to reconcile, i.e. I expected percentage = success / attempt, but this is not the case.
* There are some 0 attempts, which would make the percentages n/a

# Data Preparation

# Modelling

## Experiments

# Evaluation

## Results achieved

* Technical results such as AUC, F1 score, confusion matrix

## Challenges

## Business translation

* How an AUC of 0.70 translates to the business objective.
* Is it useable?
* In the context of drafting, what are the consequences of
  + not picking a player who eventually continues past 5 years (false negative)
  + picking a player who eventually doesn’t continue past 5 years (false positive)

## Improving the model

* Do we need more data?
* Is it possible to get more data?
  + More observations?
  + More features?
* Do we need a richer data set?
  + Reasons for not reaching the 5 year mark

# Deployment

Not sure how much we want to spend here, it depends if the model is suitable for use in production in the first place, but here are some thoughts

* batch learning, we can only learn from our mistakes in long time periods, i.e. to find out if we made the right choice, we would have to wait 5 years. There could be something fundamental changes about the data (people, society) during that time. How do we deal with this?