Christopher Paucar: Predicting Penalty Kicks with ML models

Class: IW Seminar 02
Advisor: Professor Li

1) Motivation & Goal

- Penalty kicks
  - Straightforward way to score
  - Drama & Suspense
  - Decisive
- What if one were able to predict whether a penalty shot would go in?

1) Motivation & Goal

GOAL : Using different ML models, offer the ability to predict whether a penalty kick will go in or not based on certain features.

2) Previous Works

ROBOKEEPER: THE ROBOTIC GOALKEEPER

- Uses image processing to predict where the ball will impact the goal
- Using dynamics and planning to have robot dive

https://www.robokeeper.com/en/facts.html

PREDICTING FOOTBALL PENALTY DIRECTIONS USING IN-MATCH PERFORMANCE INDICATORS

By Lotte Bransen & Jesse Davis

- Used IN-GAME metrics to predict penalty kicks
  - Missed shots
  - Pass Accuracy

https://analytics.scisports.com/research/penalty_predictor

3) Steps

4. Iterative Evaluation of Model Performance
   - Accuracy, precision, etc
   - ***Sklearn, Seaborn, Matplotlib*** Python Libraries

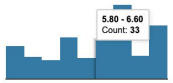
3. Prediction: Binary Classification
   - ***Sklearn*** Python Library
   - ***Imblearn*** Python Library

2. Feature Engineering, EDA
   - Feature Extraction
   - Random Oversampling
   - ***Pandas*** Python Library
   - One Hot Encoding
   - Correlation

1. Datasets
   - World Cup (1984-2018)
   - Champions League (2019-20)
   - **World Cup Misses (2022)**
   - ***Pandas*** Python Library
   - English Premier League(2016-17)
   - **Serie A Misses (2019-20)**
   - **Copa America Misses (2019, 2021)**

4.1) Implementation: Datasets

Scored	Foot	Kick_Direction	Keeper_Direction
Scored 76% Missed 24%	R 80% L 17% Other (3) 3%	L 44% R 37% Other (20) 19%	R 49% L 42% Other (9) 8%
Scored	L	C	R
Scored	R	L	L
Missed	R	L	L
Scored	R	C	L
Scored	R	L	R
Scored	R	L	R
Scored	R	C	L
Scored	R	R	L
Scored	R	C	R
Missed	R	R	R
Missed	L	R	R
Missed	L	L	L
Scored	R	L	L
Scored	R	C	R

Zone	Foot	Keeper	OnTarget	Goal
Zone where the penalty kick was directed.	Foot with which the kick was taken.	Where did the keeper dive.	Whether the shot was on target or not.	Whether the shot was successful or not.
	R 73% L 18% Other (25) 8%	L 42% R 38% Other (63) 21%		
7	R	R	1	1
9	R	C	1	1
6	R	L	1	1
2	R	C	1	1
9	R	L	1	1
4	R	L	1	0
8	L	L	1	0
3	R	R	1	1
9	R	L	1	1

1. GATHER MY DATA POINTS
   a. Kaggle, YouTube

```
path = '/content/drive/MyDrive/IW_Seminar/champions_league_19_20.csv'
champions_league_19_20 = pd.read_csv(path, encoding = ('ISO-8859-1'), low_memory = False)

# I need 3 features
# foot of the kicker
# direction of the ball
# direction of the keeper
dominant_foot_ucl = ['R','R','R','R','R','R','L','R','R','L','L','R','L','R','L','R','R','R','L','R','L','R','L','R','R','R','R','L',
ball_dir_ucl = ['L','L', 'R','L','C','R', 'R','R','C','R','L','L','R','C','R','L','R','C','L','C','R','L','R','R','L','R','C','L','C'
keeper_dir_ucl = ['L','R','R','L','L','L', 'R','R','L','R','R','L','L','C','R','R','L','R','L','L','R','L','L','L','L','L','R','R','L
# Turn into pandas DataFrame
```

4.2) Implementation: Feature Engineering and EDA

DATASETS

```
# ONE HOT ENCODING

dominant_foot_missed = ohe.fit_transform(missed_pens[['Dominant Foot']])
kick_direction_missed = ohe.fit_transform(missed_pens[['Kick Direction']])
keeper_direction_missed = ohe.fit_transform(missed_pens[['Keeper Direction']])
```

Using One Hot Encoding On Features

Features

1. Dominant Foot
   a. Left or Right
2. Kicker Direction
   a. Left, Center, Right
3. Keeper Direction
   a. Left, Center, Right

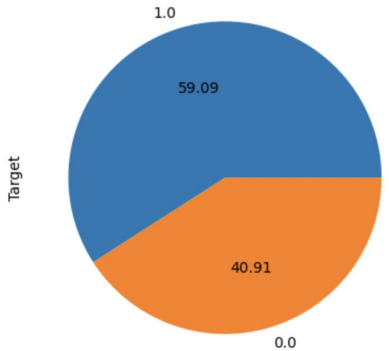
Label

- 0: Penalty Miss
- 1: Penalty Goal

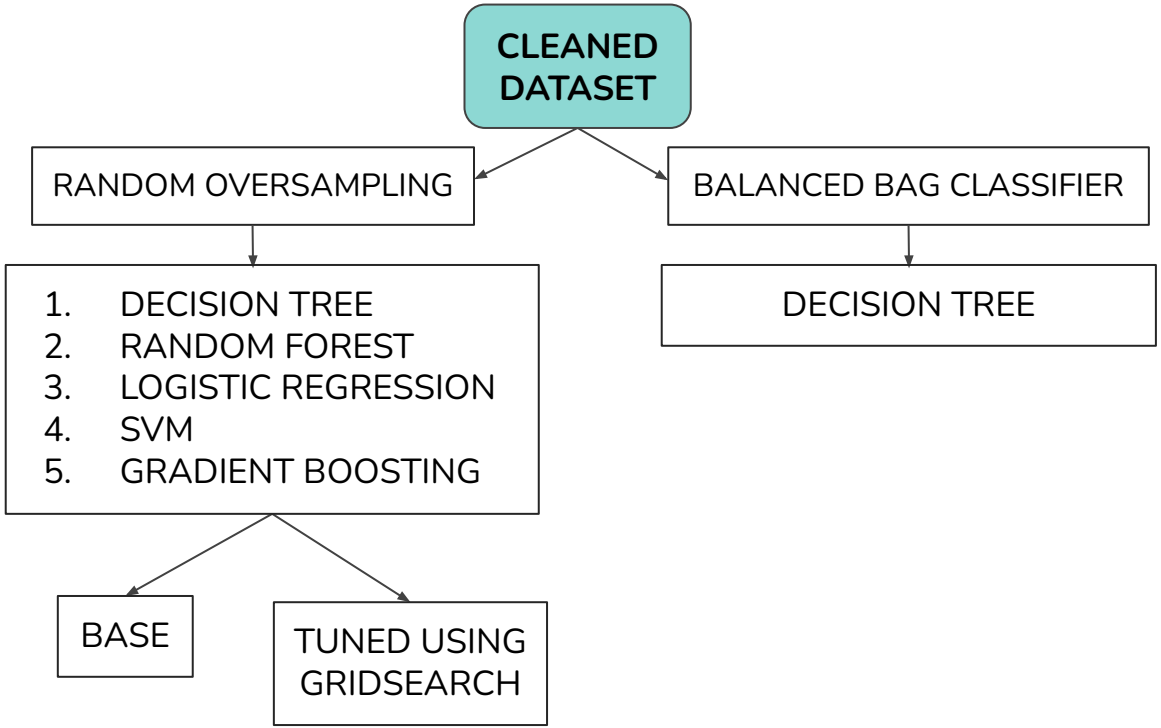
528 x 9 Dataset

Left Footed	Right Footed	Kicked Left	Kicked Center	Kicked Right	Keeper Left	Keeper Center	Keeper Right	Target
0.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	1.0
0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0
0.0	1.0	0.0	1.0	0.0	0.0	0.0	1.0	1.0
1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0
0.0	1.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0

Imbalanced!!



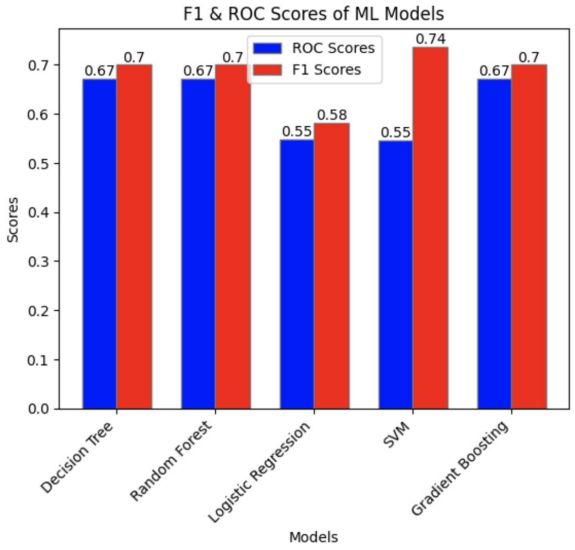
4.3) Implementation: Models



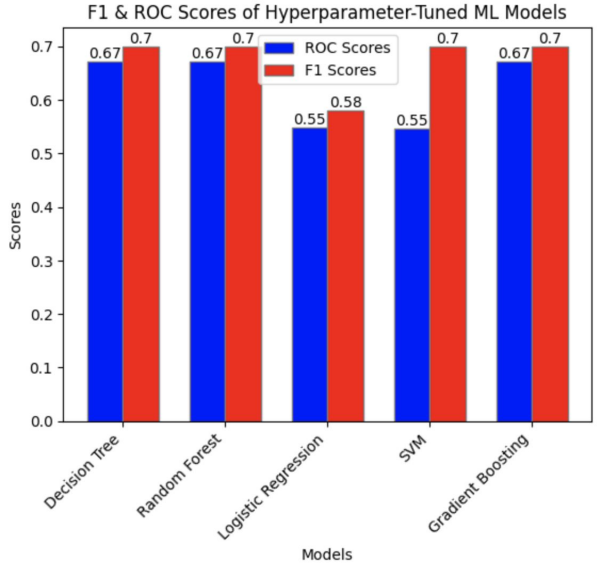
5) Evaluation & Results

RANDOM OVERSAMPLING

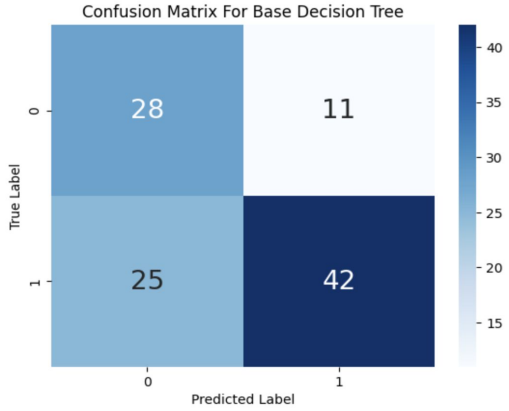
BASE

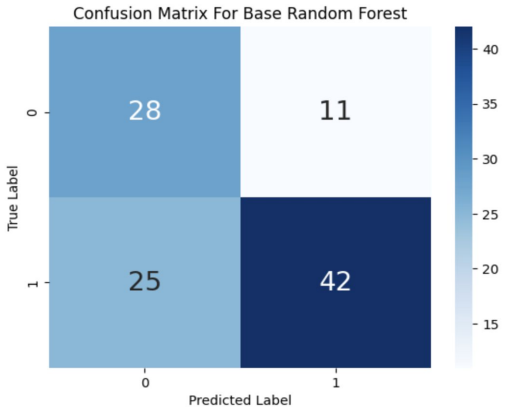


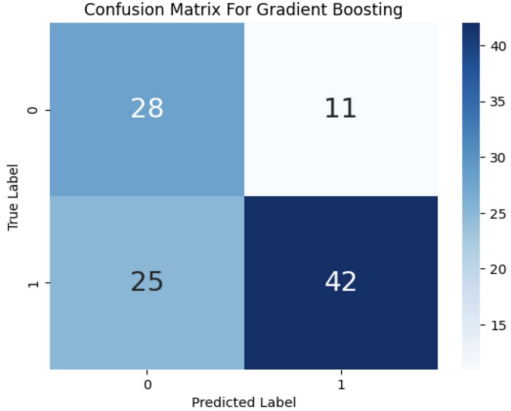
TUNED USING GRIDSEARCH

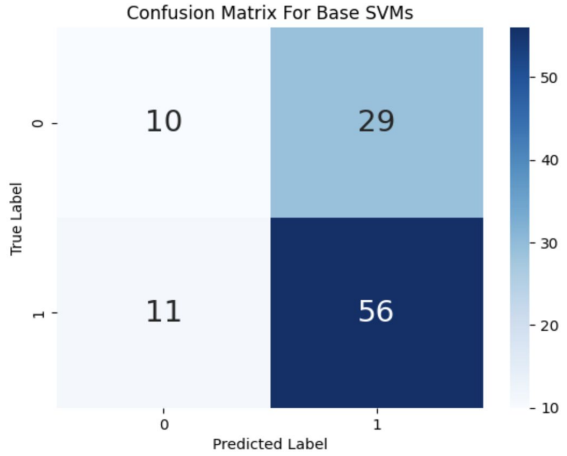


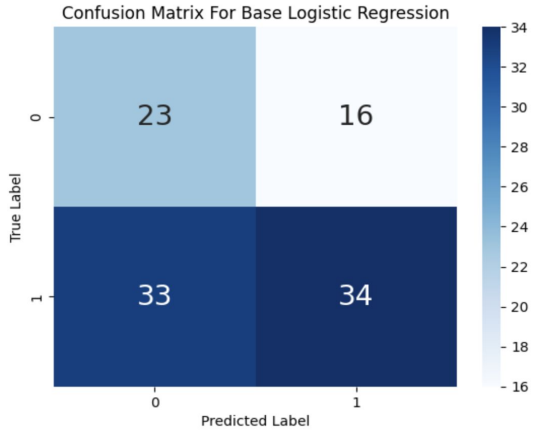
BASE MODELS





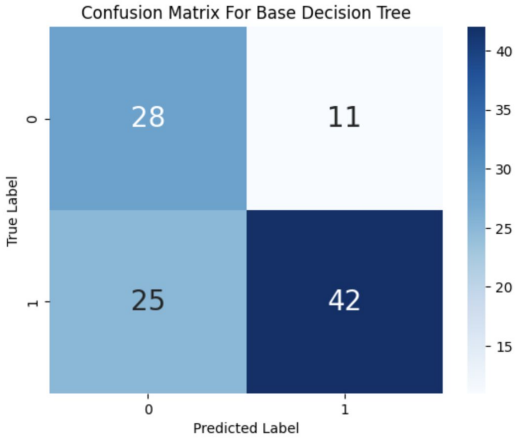


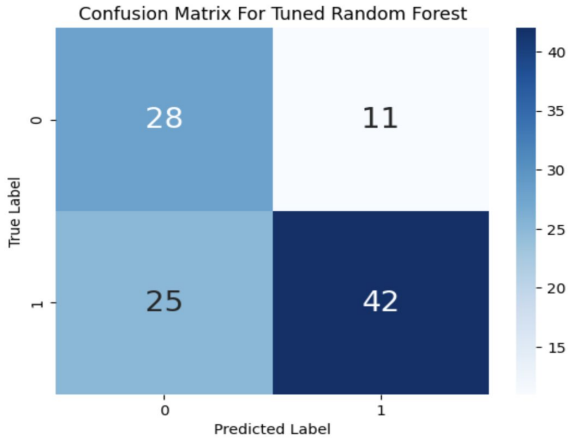


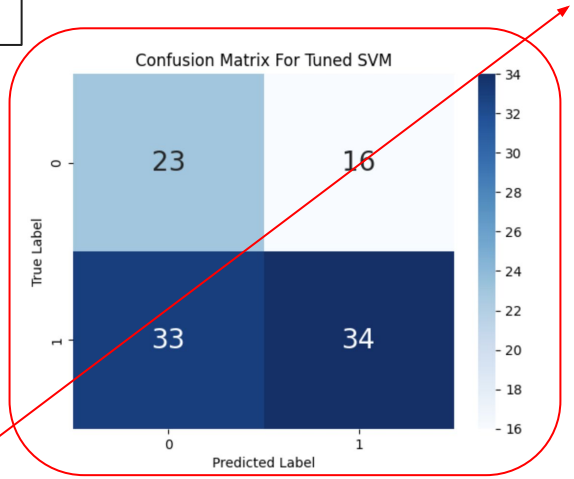


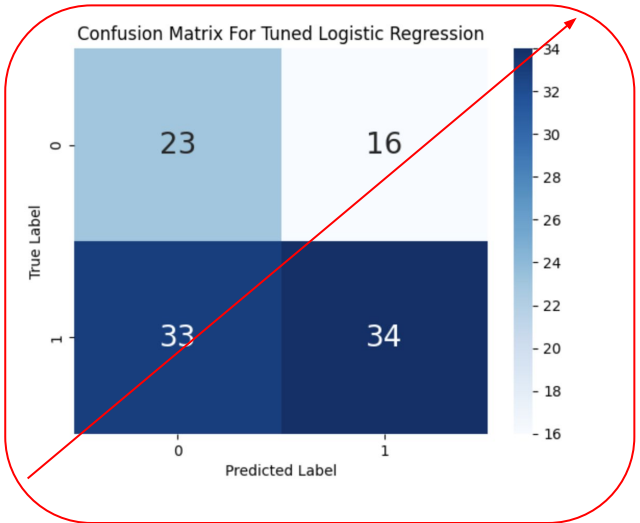
- Highest TN, TP values
  - Decision Tree, Random Forest, Gradient Boosting
- SVM
  - Low TP value, High TN value
- Logistic Regression
  - In the middle

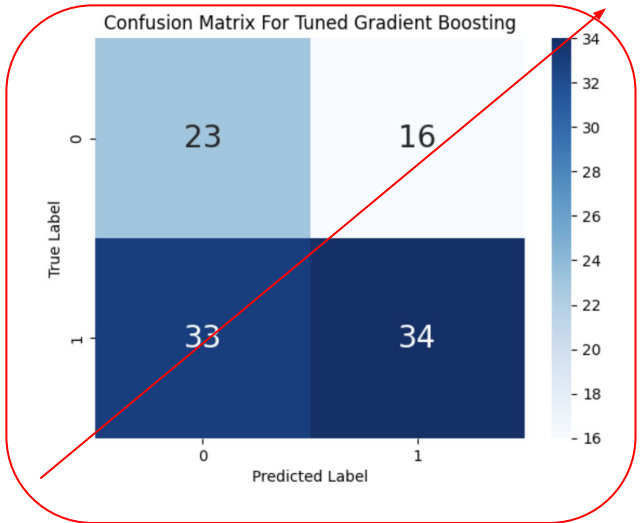
TUNED MODELS



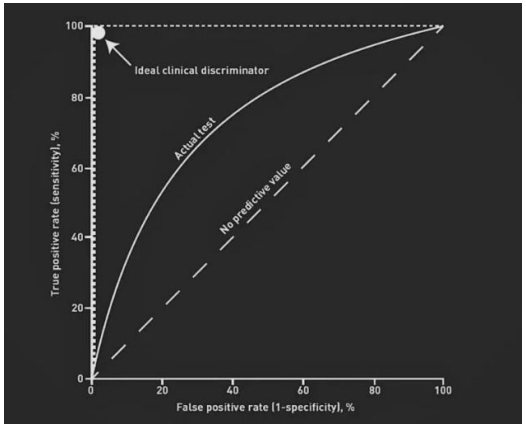


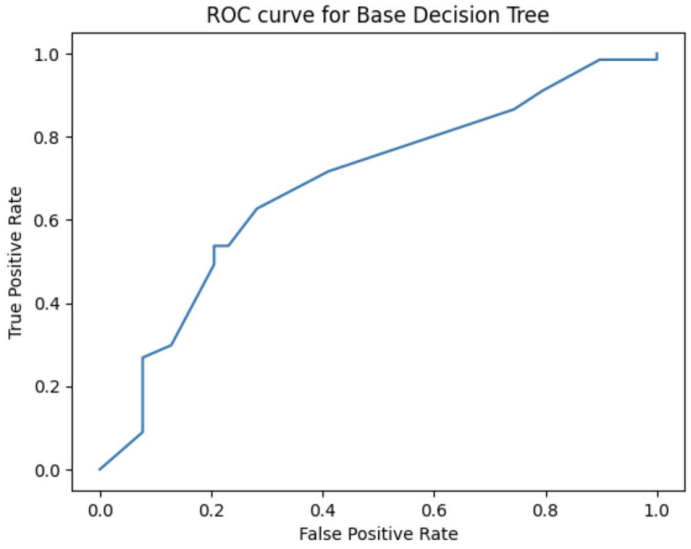




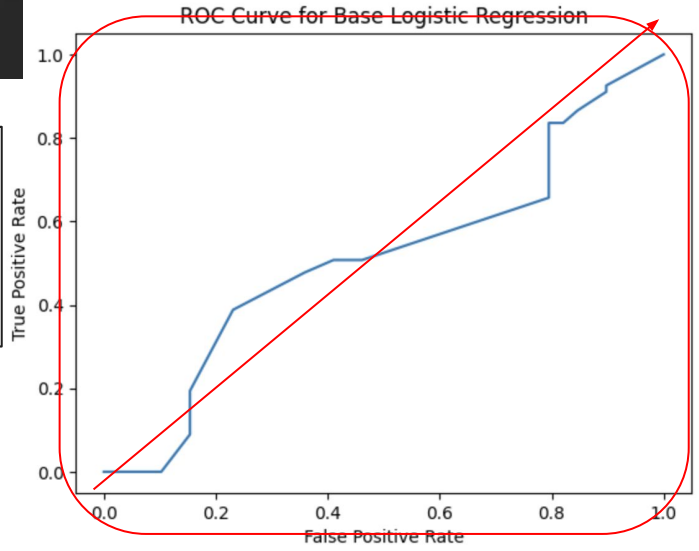


ROC CURVES



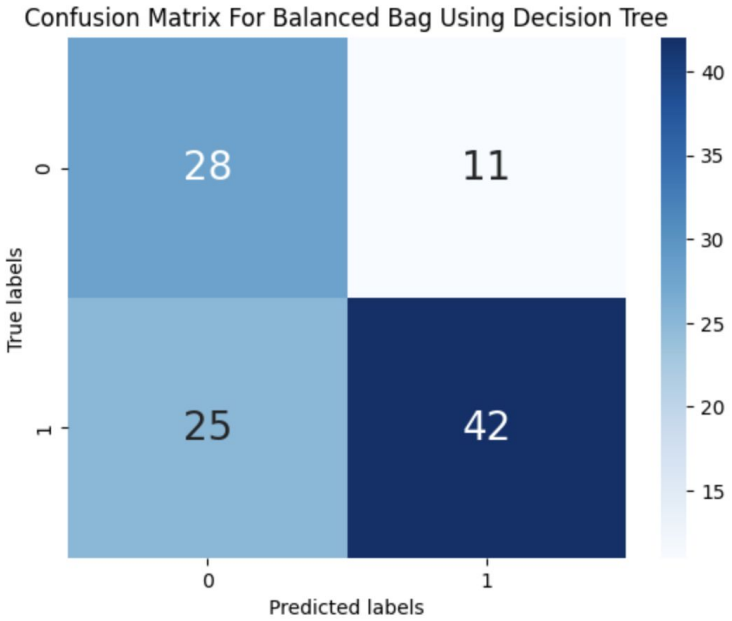


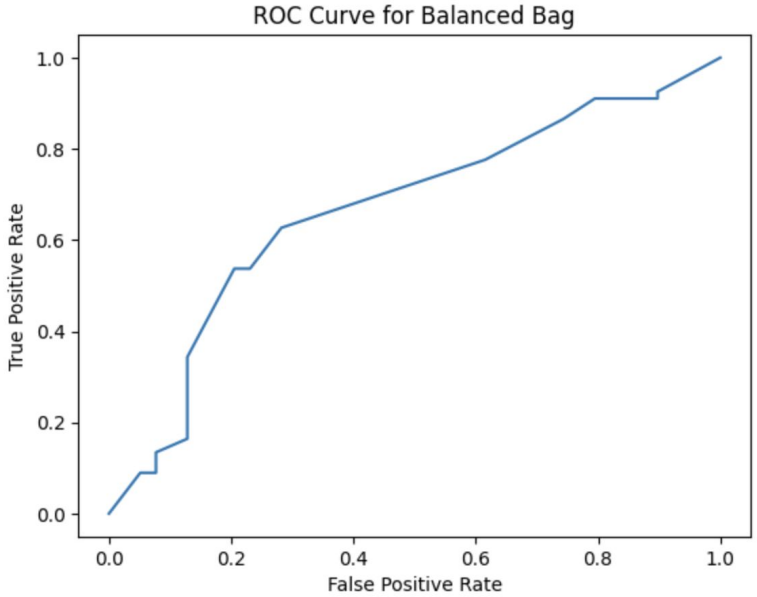
- Logistic Regression ROC Curve much closer to the “No predictive value” line



5) Evaluation & Results

BALANCED BAG CLASSIFIER





6) Conclusions & Future Work

- **CONCLUSION**
  - At a standstill in terms of performance
    - F1: 0.7 ; ROC: 0.67
  - More deciding factors in predicting penalty kicks
  - Following advice from Professor Li, will extract more features in order to enhance performance

- **FUTURE WORK**
  - Penalty Predictor Using Images & Clips of In-Moment Penalty Kicks

7) Acknowledgements

- Big Thanks To...
  - Professor Li
  - Meet Patel
  - Seminar Classmates
  - Camila & Natalie, for their advice and fun times throughout this seminar:)