Prediction on Market Values of FIFA Players

Group 6
Randolph Zhao, Nathan Heller, Ergis Mecaj, Nguyen Luu

Introduction/Insight

FIFA (Federation Internationale de Football Association) is an international non-profit organization that oversees organizations related to football, futsal, and beach soccer.

FIFA oversees six international confederations of association football, with each confederation serving a different continental region of the world.

Each national association within each international confederation is a direct member of FIFA

The FIFA series of video games is meant to be a realistic soccer simulation game that uses real-world players and data collected from those players.

Problem Statement

How closely are the market values of FIFA players predicted by:

- Their **background information**?
- Their **abilities** within different positions and skills?
- Their **performances** over their current careers?

From the trained model, we can

- predict the **values of new players**
- highlight most effective manners to **increase existing players' values**

Initial Dataset

FIFA 19 Complete Player Dataset: over 18000 samples / 89 features

Information regarding personal background, current employment, ratings for each position

- Ratings determined by over 9000 data reviewers and over 300 editors

Crossing	Finishing	Heading Accura	Short Passing	Volleys	Dribbling	Curve
84	95	70	90	86	97	93
84	94	89	81	87	88	81
79	87	62	84	84	96	88
17	13	21	50	13	18	21
93	82	55	92	82	86	85
81	84	61	89	80	95	83
86	72	55	93	76	90	85
77	93	77	82	88	87	86

Preprocessing Data: Feature Engineering

- Remove unit of some features: '$' in 'wages' and ''values', ' 'lbs' in 'weight'
- Separate 'Joined' to 'Joined_month' and 'Joined_year'
- Change 'height' into inch
- Transform text value into numeric value
- Re-map categorical variables into smaller bins
- Rescale, normalize, and eliminate the outliers of the data

# Preprocessing Data: Missing Values & Drop Columns

Missing Values:

- 'Club': fill with 'No Club'
- 'Position': fill with 'Unknown'
- 'Simple_position': fill with 'Unknown'
- 'Body Type': fill with 'Unknown'
- Others: fill with mean

Column dropped:

- Useless:
  - 'Unnamed: 0', 'ID', 'Name', 'Photo', 'Flag', 'Club Logo', 'Real Face', 'Joined', 'Special', 'Joined_month', 'Jersey Number', 'Loaned From', 'Contract Valid Until'
- Transformed to other features:
 - 'Nationality','Work Rate','Position', 'Club'

Preprocessing Data: Multicollinearity Reduction
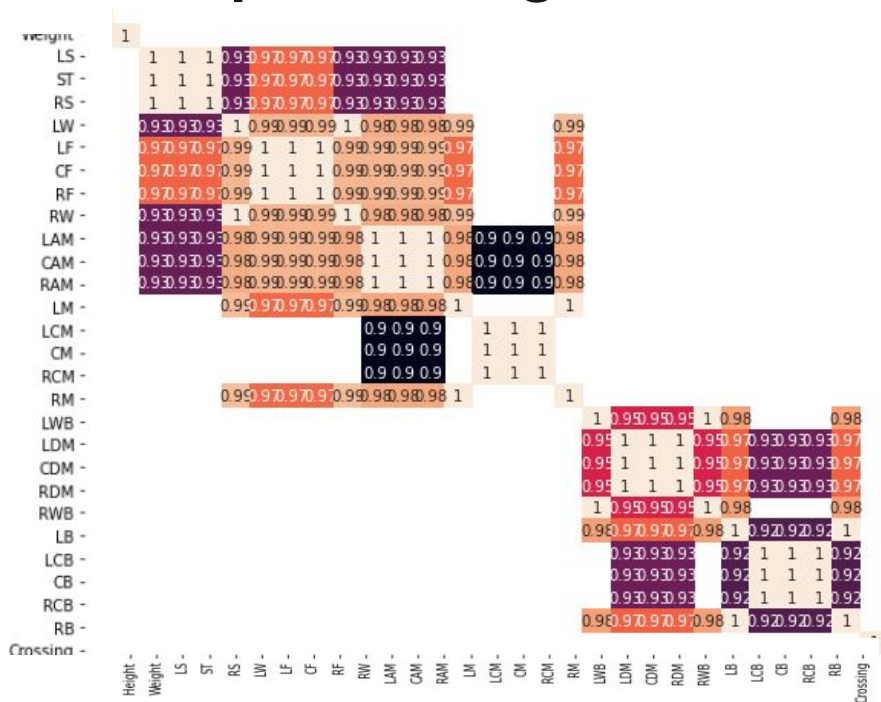
In feature engineering, we looked at features that were highly correlated and combined them

- The standingTackle and slidingTackle features were highly correlated, so they were combined into a single "Tackle" feature
- The various goalkeeping features, indicated with a GK, were all highly correlated with each other, so they were combined into an aggregate goalkeeping feature.
- There were far too many correlated features when looking at the features that described how well a player could play a specific position, so we combined those into general "Forward","Mid", and "Defense" features instead, rather than making a distinction between right wing and left forward.

Model Selection & Feature Selection
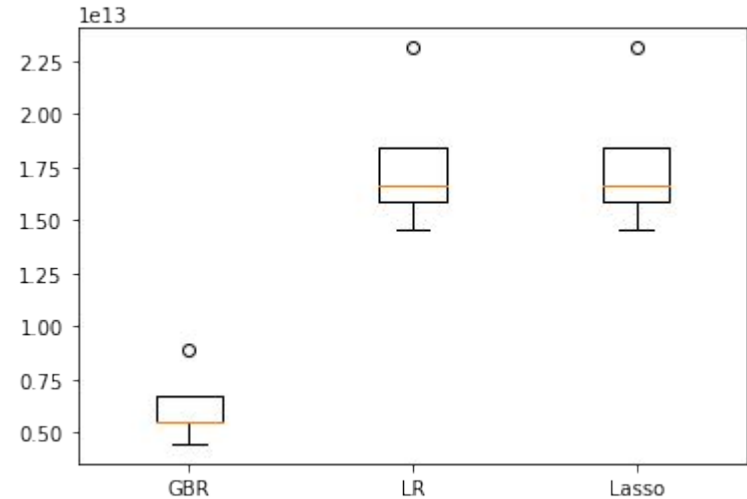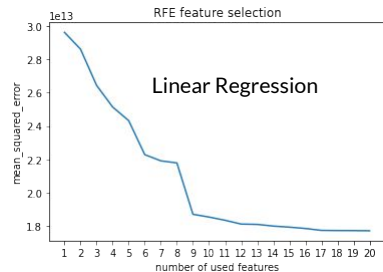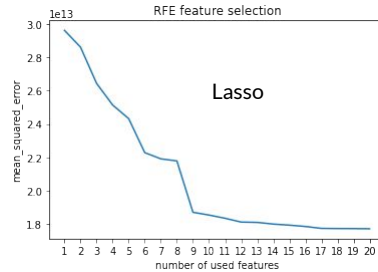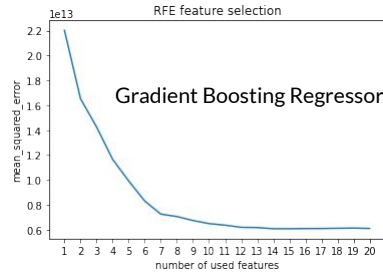

```
GradientBoostingRegressor:      6217628050008.035      + 1509903485396.85
LinearRegression:               17715346477530.023     + 2968663261528.3794
Lasso:                          17715346557704.79      + 2968672219817.735
```
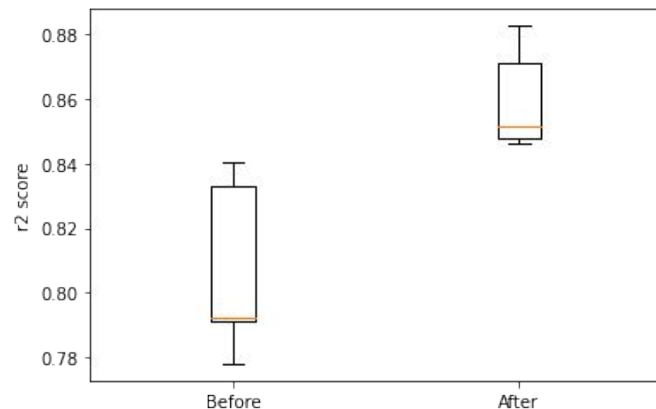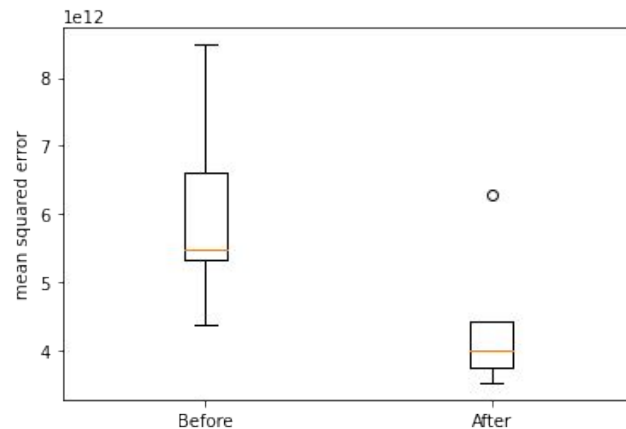
Tuning Hyperparameters

```
{'learning_rate': 0.1,
 'loss': 'squared_error',
 'max_depth': 5,
 'n_estimators': 240}
```

```
{'max_features': 13,
 'min_samples_leaf': 20,
 'min_samples_split': 25,
 'subsample': 1.0}
```

Model Results & Conclusion

- MSE cross validation score:
  - mean = 4,399,891,465,323.725
  - std = 984,225,798,043.3031
- R2 cross validation score:
  - mean = 0.8598939706399216
  - std = 0.014417427062726796

- Based on R2 score, the model performs well in the regression of values of FIFA players.
  - i.e. we can have an approximation on values of each player based on their personal data.