ANALYZING TEAM SUCCESS IN FIFA WORLD CUP

Ryan Cullen, Alec Zerona, Warren Paintsil, Will Hensley

ISSUE: US MEN'S SOCCER HAS PERFORMED POORLY AT HIGH LEVEL

<u>Past:</u> 12 World Cup appearances: 5-6-2 (W-L-D)

Best "modern" result: reached 2022 quarterfinal

Best result**:** Reached Semi-final in 1930

<u>Present (2024 stats)</u>: The US Men's soccer team has a losing record and negative goal differential

This stat was updated November 15, 2024

DATA ANALYTICS STRATEGIES

01. Logistic Regression

- Predict winning outcomes using features with selection using Elastic Net, LASSO, and Ridge

02. Random Forest

- Decision trees with bootstrap sampling to improve accuracy and reduce variance

03. Boosting

- Build strong learners from weak ones, using lambda for slow learning to optimize insights.

DATA TRANSFORMATION

Feature Categories

Offensive Opportunities:

- Shot accuracy, shot attempts, crosses, corners, free kicks, passing the defensive line.
- **New Feature**: `team1_win` (1 = win, 0 = loss).

Defensive Opportunities:

- Turnovers, pressure, defensive line break efficiency, fouls, offsides.
- Relative features: Difference between Team 1 and Team 2 values.

Tactical Movement & Passing Dynamics:

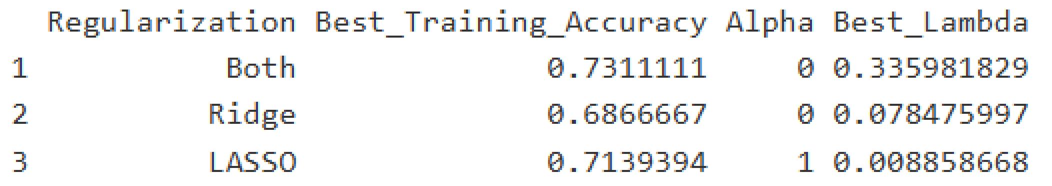
- Possession, passing strategies, field switching, ball movement, preferred positioning.
- Relative features: Difference between Team 1 and Team 2.

Data Cleaning Checks: Duplicates, nulls, infinites

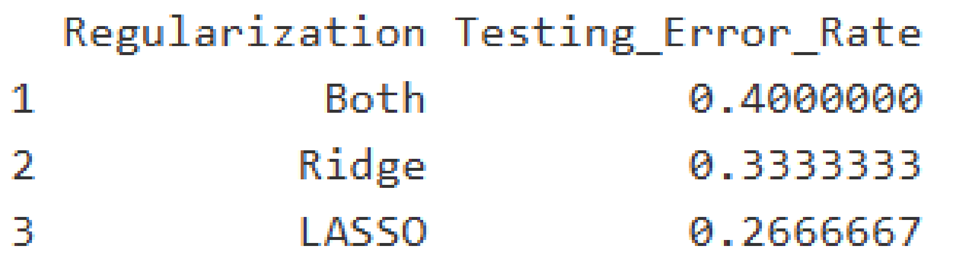
LOGISTIC MODEL RESULTS

Confusion Matrices

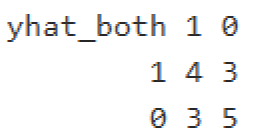
Comparing Models on Training and Testing Data



```
  Regularization Best_Training_Accuracy Alpha Best_Lambda
1           Both              0.7311111     0 0.335981829
2          Ridge              0.6866667     0 0.078475997
3          LASSO              0.7139394     1 0.008858668
```

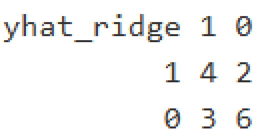


```
  Regularization Testing_Error_Rate
1           Both          0.4000000
2          Ridge          0.3333333
3          LASSO          0.2666667
```



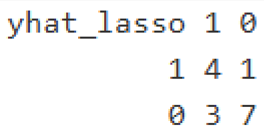
```
yhat_both 1 0
        1 4 3
        0 3 5
```

Accuracy = 0.60
Precision = 0.57



```
yhat_ridge 1 0
         1 4 2
         0 3 6
```

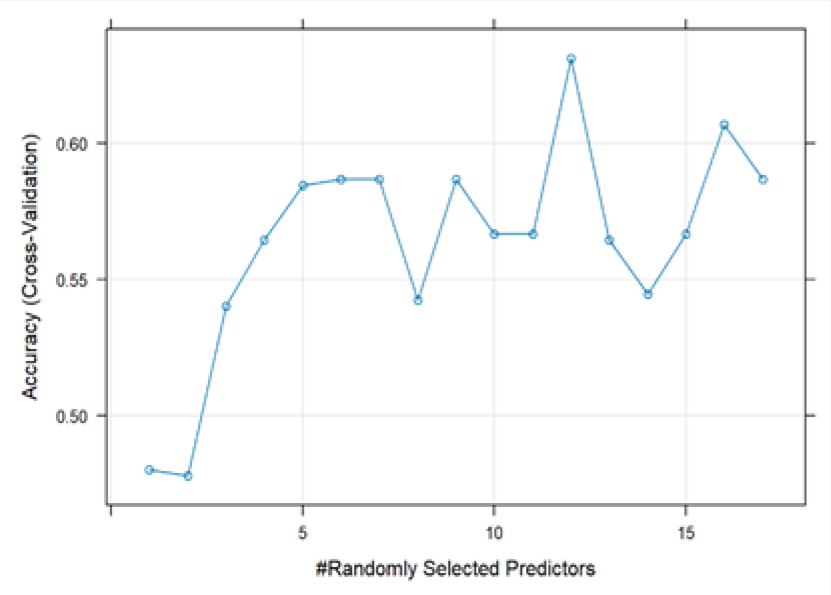
Accuracy = 0.67
Precision = 0.67

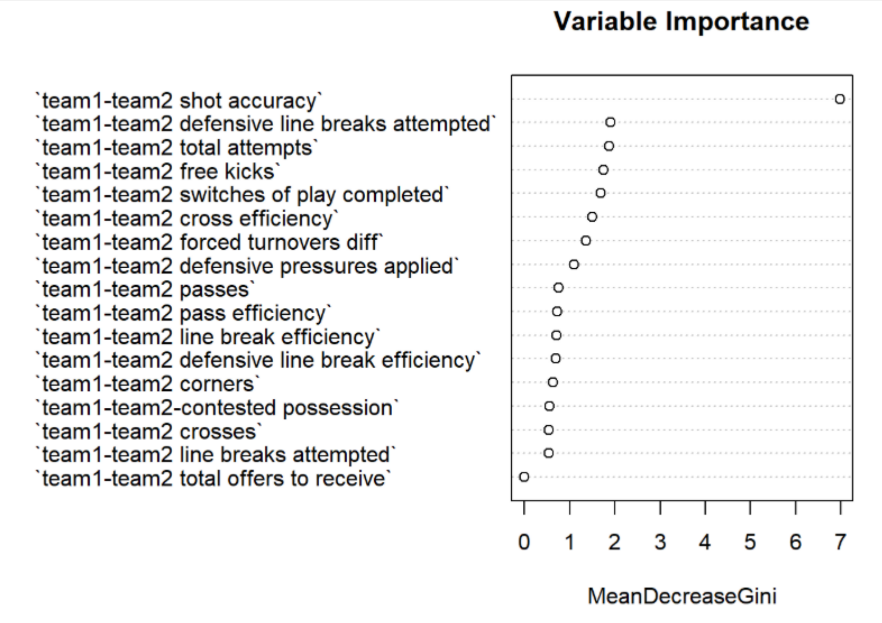


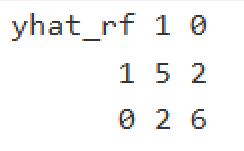
```
yhat_lasso 1 0
         1 4 1
         0 3 7
```

Accuracy = 0.73
Precision = 0.80

RANDOM FOREST MODEL RESULTS



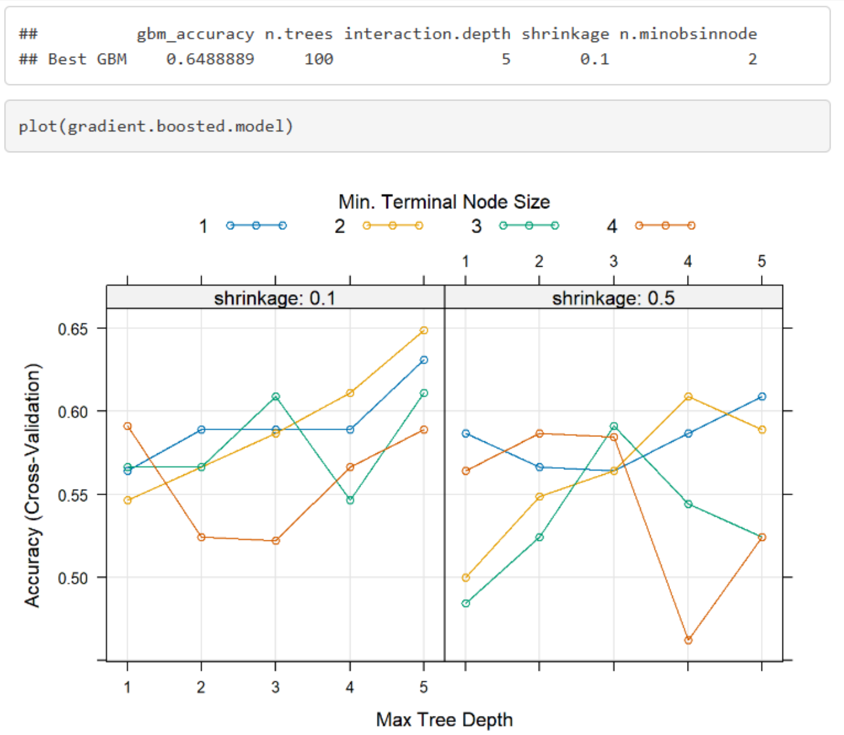




```
yhat_rf 1 0
      1 5 2
      0 2 6
```

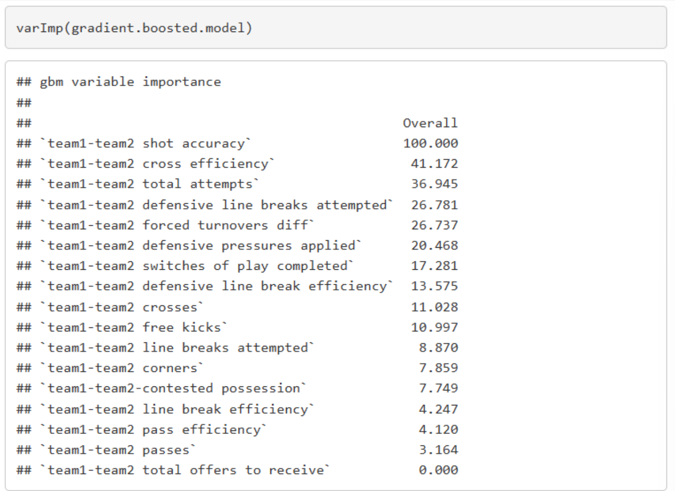
Accuracy = 0.73
Precision = 0.71

BOOSTING MODEL RESULTS



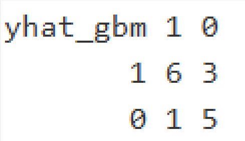
```
##          gbm_accuracy n.trees interaction.depth shrinkage n.minobsinnode
## Best GBM    0.6488889     100                 5       0.1              2
```

```
plot(gradient.boosted.model)
```



```
varImp(gradient.boosted.model)
```

```
## gbm variable importance
##
##                                                      Overall
## `team1-team2 shot accuracy`                          100.000
## `team1-team2 cross efficiency`                        41.172
## `team1-team2 total attempts`                          36.945
## `team1-team2 defensive line breaks attempted`         26.781
## `team1-team2 forced turnovers diff`                   26.737
## `team1-team2 defensive pressures applied`             20.468
## `team1-team2 switches of play completed`              17.281
## `team1-team2 defensive line break efficiency`         13.575
## `team1-team2 crosses`                                 11.028
## `team1-team2 free kicks`                              10.997
## `team1-team2 line breaks attempted`                    8.870
## `team1-team2 corners`                                  7.859
## `team1-team2-contested possession`                     7.749
## `team1-team2 line break efficiency`                    4.247
## `team1-team2 pass efficiency`                          4.120
## `team1-team2 passes`                                   3.164
## `team1-team2 total offers to receive`                  0.000
```



```
yhat_gbm 1 0
       1 6 3
       0 1 5
```

Accuracy = 0.73

Precision = 0.67

WHAT'S NEXT?

Further Model Optimization

Continue fine-tuning models: Test additional hyperparameters, consider alternative models, and refine existing models for higher accuracy and predictive power.

Expanded Feature Engineering

Incorporate more features: Include advanced features like player-level statistics and match dynamics to improve predictions.

Real-World Application

Focus on actionable insights: Translate findings into tactical recommendations for the U.S. Men's National Team, focusing on improving both offensive and defensive strategies.

Future Testing and Validation

Test on new data: Continuously validate models with new match data to ensure robustness and adaptability over time.