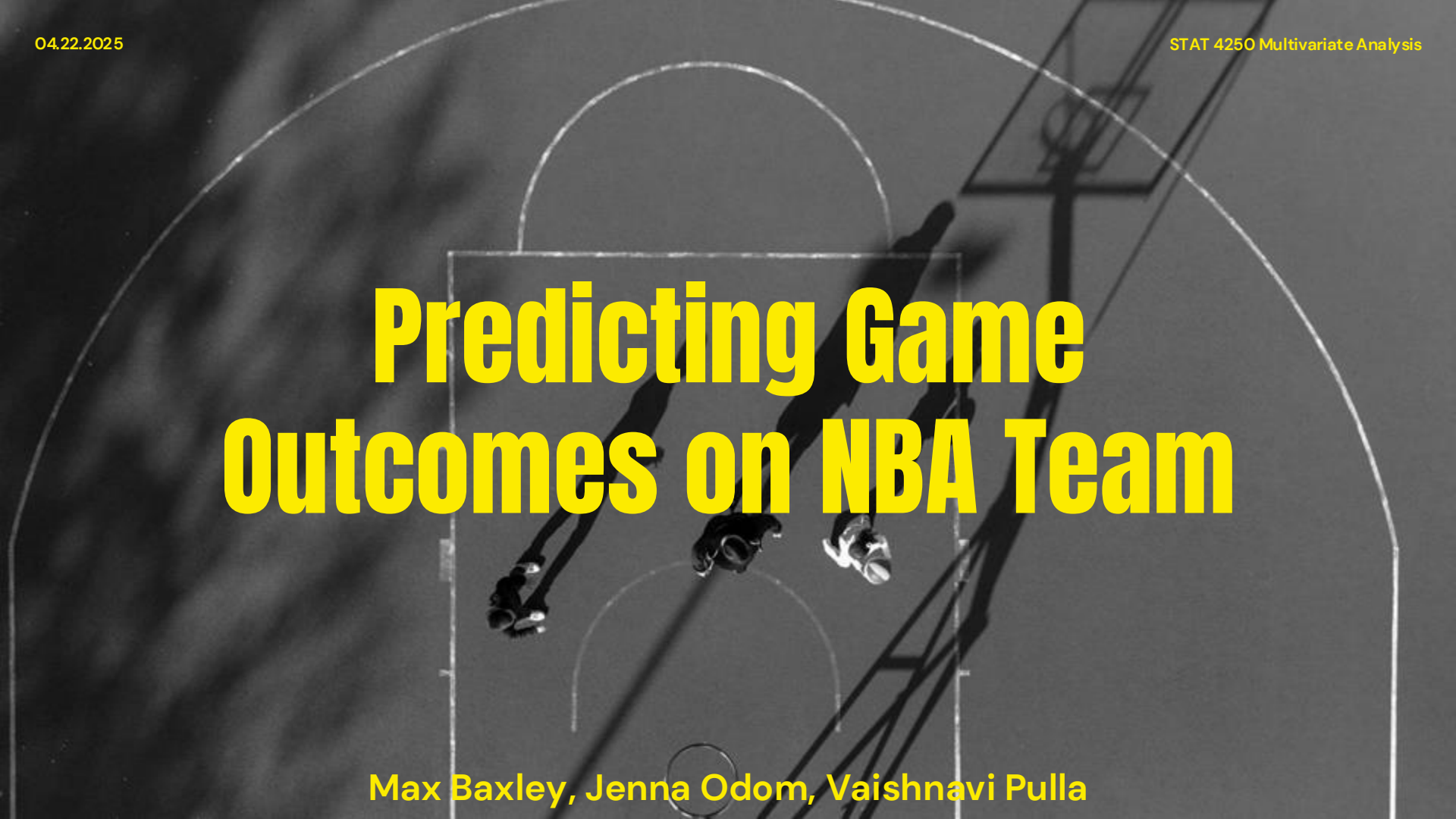


Predicting Game Outcomes on NBA Team

Max Baxley, Jenna Odom, Vaishnavi Pulla

Regular Season:

- 30 teams (2 conferences - Eastern and Western)
- Each team plays 82 games during the regular season
- Play-In Tournament - Teams finishing 7th - 10th in each conference participate in a playin for the final two playoff seeds.

The Playoffs:

- 4 rounds, best of seven-series
- Teams are seeded based on their regular season performance.
- The winners of eastern and western conference face of in the finals.
- Winner is the NBA Champion

The NBA Cup:

- In-season Tournament
- One off games similar to Playoffs in the middle of the regular season.

About
Brief
Concept
Design
Next Steps

The Data

information to predict win/loss

- We have stats such as steals, blocks, assists, and points scored.
- We have the plus or minus values for the outcomes of each game and we define a win or loss as 1 or -1.
- Sign of the +1 and -1 for the feature Plus_MINUS depicts whether the game was won or lost

Data Features

'SEASON_ID', 'TEAM_ID', 'TEAM_ABBREVIATION', 'TEAM_NAME',
'GAME_ID',
'GAME_DATE', 'MATCHUP', 'WL', 'MIN', 'PTS', 'FGM', 'FGA',
'FG_PCT',
'FG3M', 'FG3A', 'FG3_PCT', 'FTM', 'FTA', 'FT_PCT', 'OREB', 'DREB',
'REB', 'AST', 'STL', 'BLK', 'TOV', 'PF', 'PLUS_MINUS'],

	SEASON_ID	TEAM_ID	TEAM_ABBREVIATION	TEAM_NAME	GAME_ID	GAME_DATE	MATCHUP	WL	MIN	PTS	...	FT_PCT	OREB	DREB	REB	AST	STL	BLK	TOV	PF	PLUS_MINUS
2	22024	1610612737	ATL	Atlanta Hawks	0022401186	2025-04-13	ATL vs. ORL	W	241	117	...	0.600	9	35	44	32	8.0	2	15	15	12.0
3	22024	1610612737	ATL	Atlanta Hawks	0022401173	2025-04-11	ATL @ PHI	W	240	124	...	0.750	18	35	53	28	9.0	5	14	15	14.0
4	22024	1610612737	ATL	Atlanta Hawks	0022401169	2025-04-10	ATL @ BKN	W	238	133	...	0.667	7	44	51	36	11.0	1	15	19	24.0
5	22024	1610612737	ATL	Atlanta Hawks	0022401149	2025-04-08	ATL @ ORL	L	241	112	...	0.818	14	27	41	25	7.0	6	18	26	-7.0
6	22024	1610612737	ATL	Atlanta Hawks	0022401136	2025-04-06	ATL vs. UTA	W	239	147	...	0.850	11	31	42	43	7.0	3	11	19	13.0

IMPORTANT STEPS

Defining

Scale/ Normalize Variables

Develop

Fitting The Models

Test

Calculating Score Value

Tune

Cross Validation

Basic Models

Linear, Ridge, and Lasso Models were fitted

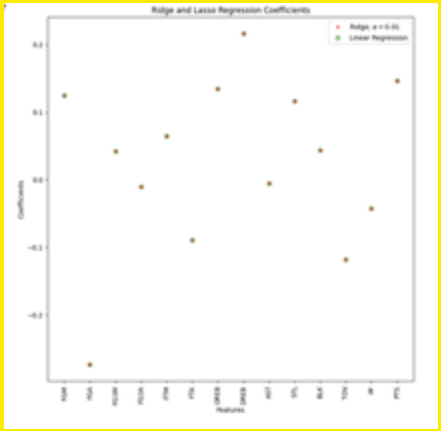
The coefficient of determination was = low.

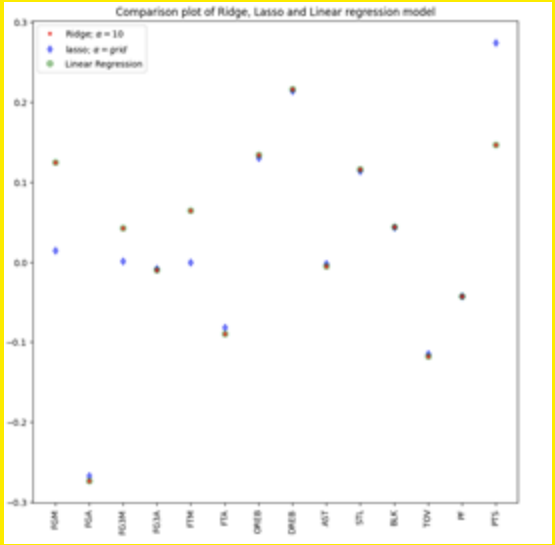
The graphs can be seen below.

The optimized alpha values of the model were also determined.

Fitting the Linear Regression

The coefficient of determination was also low.





52.67% Linear Regression Score

52.68% RidgeCV Regression Score

52.67% LassoCV Regression Score

The above values looked at the values for the testing dataset

THE $$R^2$$ VALUE FOR LINEAR, RIDGE, AND LASSO IS OKAY?

NO, IT'S NOT

In the context of determining, if a team is projected to win or lose in a matchup. Being able to explain around 50% of the variation in the data with a model is not a valuable insight to provide a company.

We need to know:

the most important factors in winning a game

To help inform operational decisions that the basketball office can make.

94.7

Highest Accuracy that we were able to achieve with new methods

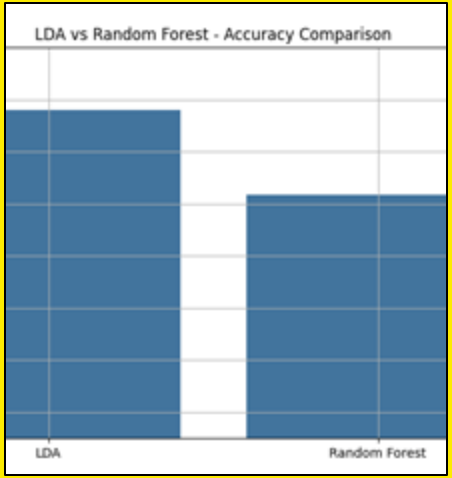
This shows the impact that utilized dimension reduction methods can have on making models for effective.

Over the next couple of slides we will go over the specifics of our methods.

Looking at LDA and Random Forest

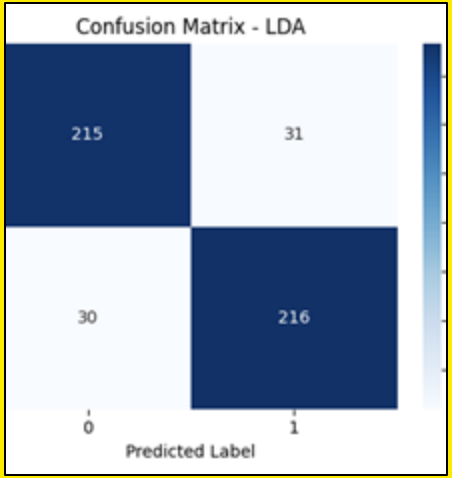
This is look at some of the comparison measures.

LDA & Random Forest



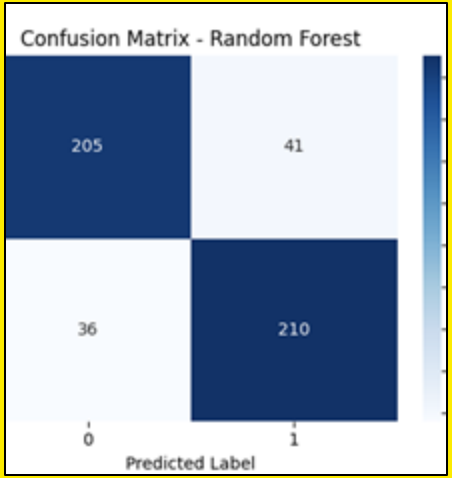
Comparing Accuracy

Confusion LDA



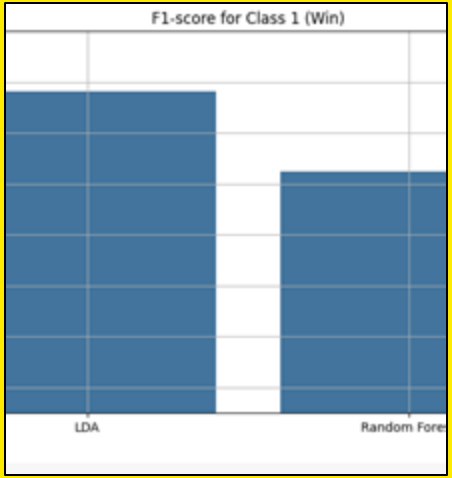
Confusion Matrix

Random Forest Confusion



Confusion Matrix

F1 Score for Class One Win



F1 Score

LDA

```
LDA Results:
Accuracy: 0.8760162601626016
Confusion Matrix:
 [[215  31]
 [ 30 216]]
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.87      0.88       246
           1       0.87      0.88      0.88       246

    accuracy                           0.88       492
   macro avg       0.88      0.88      0.88       492
weighted avg       0.88      0.88      0.88       492
```

Linear Discriminant Analysis

- Models data distribution for the wins and losses
- Uses Bayes Theorem to classify new data points(games)
- Accuracy - 87.60%

Random Forest

Random Forest

- Combines output of multiple decision trees into reaching a single result
- This is a case of a classification problem
- Extension of Bagging and Feature randomness to create uncorrelated trees (feature randomness means that it only considers a subset of features – as a opposed to all possible features in LDA)
- Accuracy - 84.35%

```
Random Forest Results:
Accuracy: 0.8434959349593496
Confusion Matrix:
 [[205  41]
 [ 36 210]]
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.83      0.84       246
           1       0.84      0.85      0.85       246

    accuracy                           0.84       492
   macro avg       0.84      0.84      0.84       492
weighted avg       0.84      0.84      0.84       492
```

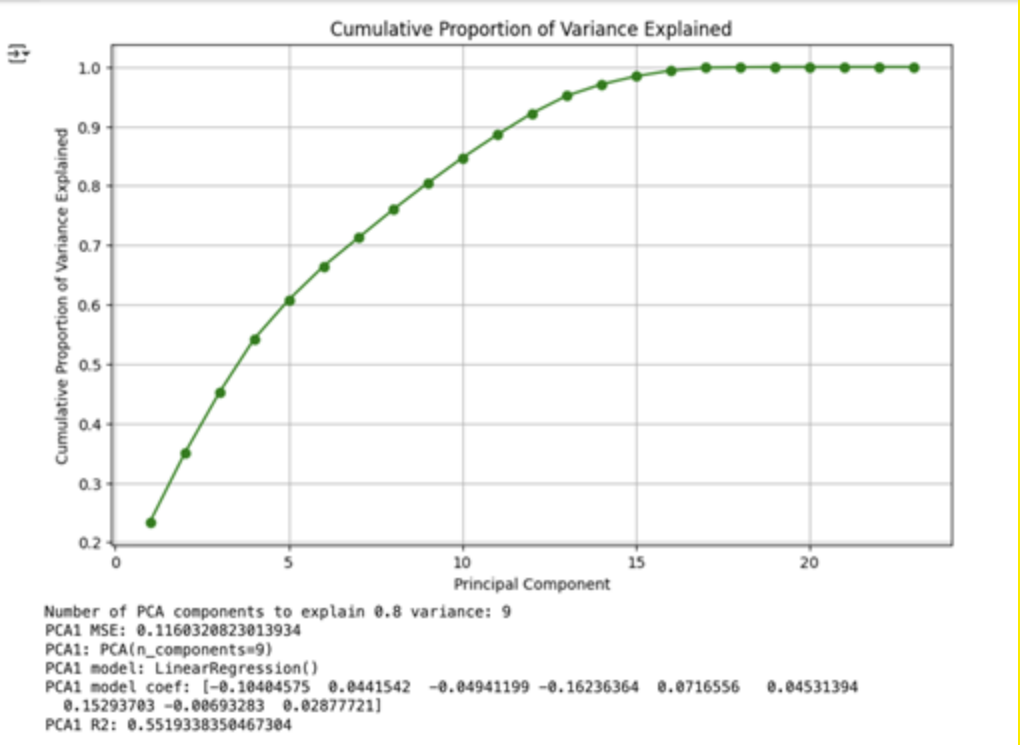
LOOKING AT METHODS

FOR CLOSE GAMES


OUR PROCESS

Principal Component Analysis Linear Model

Principal Component Analysis (PCA) is dimensionality reduction method to make a large dataset smaller while making maintaining patterns and trends

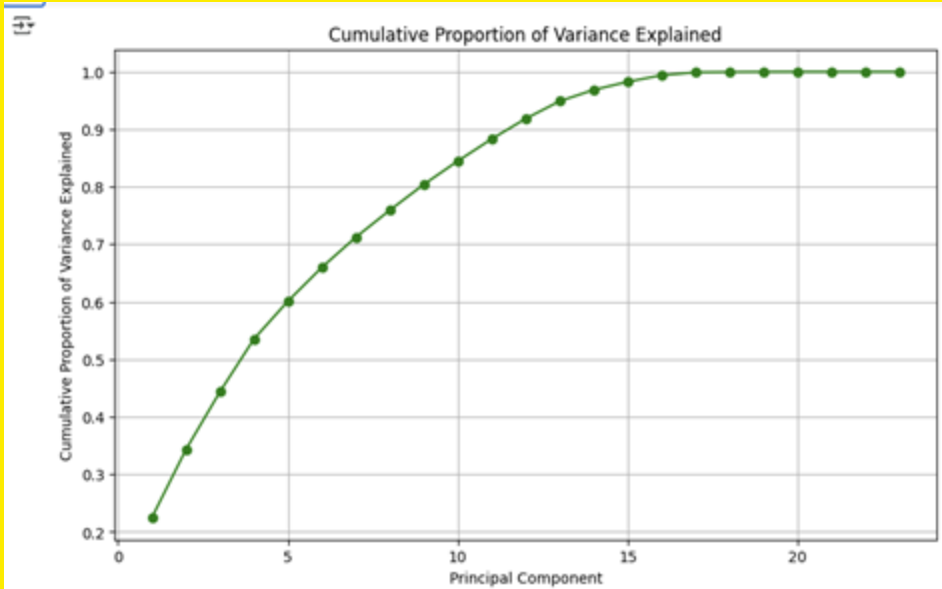


Utilizes PCA values to create Linear Model
R2 value is 55.19%

SVM using Principal Component Analysis

SVM - Support Vector Machines Aided by Principal Component Analysis

The accuracy of the model is around 82.94%.



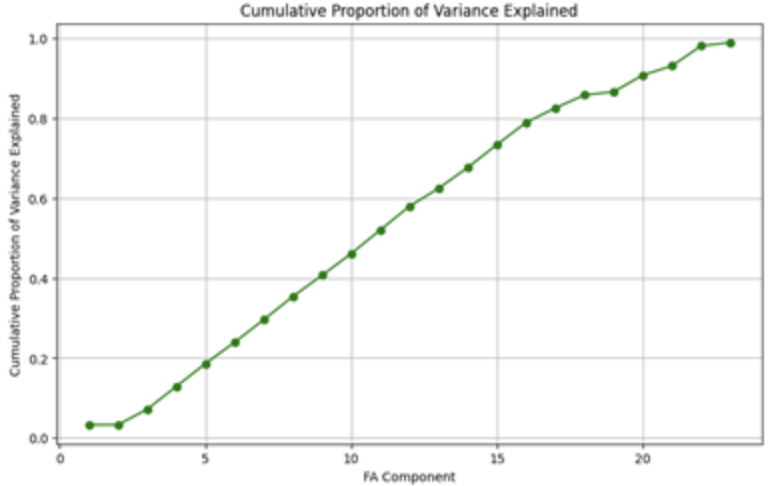

```
Number of PCA components to explain 0.8 variance: 9
PCA SVM Accuracy: 0.8294117647058824
Confusion Matrix:
 Predicted->Positive(PP), Negative(PN)
 true Positive (P): [79 13]
 true Negative (N): [16 62]

 [[79 13]
 [16 62]]
Classification Report:
              precision    recall  f1-score   support

          -1       0.83      0.86      0.84        92
           1       0.83      0.79      0.81        78

    accuracy                           0.83       170
   macro avg       0.83      0.83      0.83       170
weighted avg       0.83      0.83      0.83       170
```

SVM using Factor Analysis

SVM – Support Vector Machines Aided by Factor Analysis

The accuracy of the model is around 94.71%.

```
Variance explained by each factor: [3.30871469e-02 1.79573090e-43 3.86098571e-02 5.72318553e-02
 5.72448014e-02 5.33416010e-02 5.66717480e-02 5.79338969e-02
 5.31079890e-02 5.42262091e-02 5.95079120e-02 5.89987811e-02
 4.51570136e-02 5.19591617e-02 5.74354781e-02 5.51976393e-02
 3.62390511e-02 3.24188470e-02 8.12863560e-03 4.16957263e-02
 2.30864760e-02 5.01514739e-02 8.33000983e-03]
```




```
Number of FA components to explain 0.8 variance: 17
Factor Analysis SVM Accuracy: 0.9470588235294117
Confusion Matrix:
 Predicted->Positive(PP), Negative(PN)
 true Positive (P): [88  4]
 true Negative (N): [ 5 73]

 [[88  4]
 [ 5 73]]
Classification Report:
              precision    recall  f1-score   support

          -1       0.95      0.96      0.95        92
           1       0.95      0.94      0.94        78

    accuracy                           0.95       170
   macro avg       0.95      0.95      0.95       170
weighted avg       0.95      0.95      0.95       170
```

The Solution

Research

The model that performed the best was the SVM using Factor Analysis Model.

Optimize

After going through multiple iterations, this gave us the best outcomes.

Experiment

We working through multiple models taking into consideration a variety of factors and estimating how these values can better predict outcomes and allow office to make more informed decisions.

Let's talk about what's next!

Expanding the Dataset to take into consideration player vs player interactions in predicting game outcomes.

Address

STAT 4250 Multivariate FInal Project

Contact

Max Baxley
Jenna Odom
Vaish Pulla

Thank You