

Ucla



T



Men's College Basketball

Ole
Miss

M

H

M

March Madness Seed Prediction

A



P

Y



Project Idea

Predict the March Madness Seed of men's college basketball teams based on data from their season

1. feature engineered our data
2. graphed relationship between features
3. split it into training and testing data
4. used different algorithms to predict the seed using the season stats

Features in Dataset

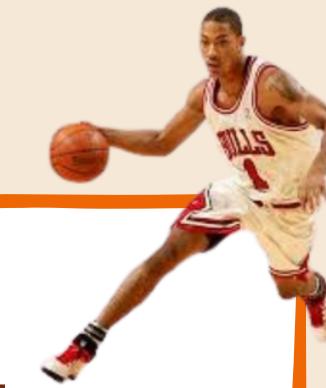
Two-point shooting percentage

Three-point shooting percentage

Two-point shots allowed

Three-point shots allowed

Wins above bubble



Feature Engineering

- Created a new column called win percentage by dividing the number of games won by the number of games played
- Dropped the games won, games played, team, and post season features/columns.

```
df['Win Percentage'] = df['W']/df['G']
df.drop(['W', 'G', 'TEAM', 'POSTSEASON'], axis=1, inplace=True)
```

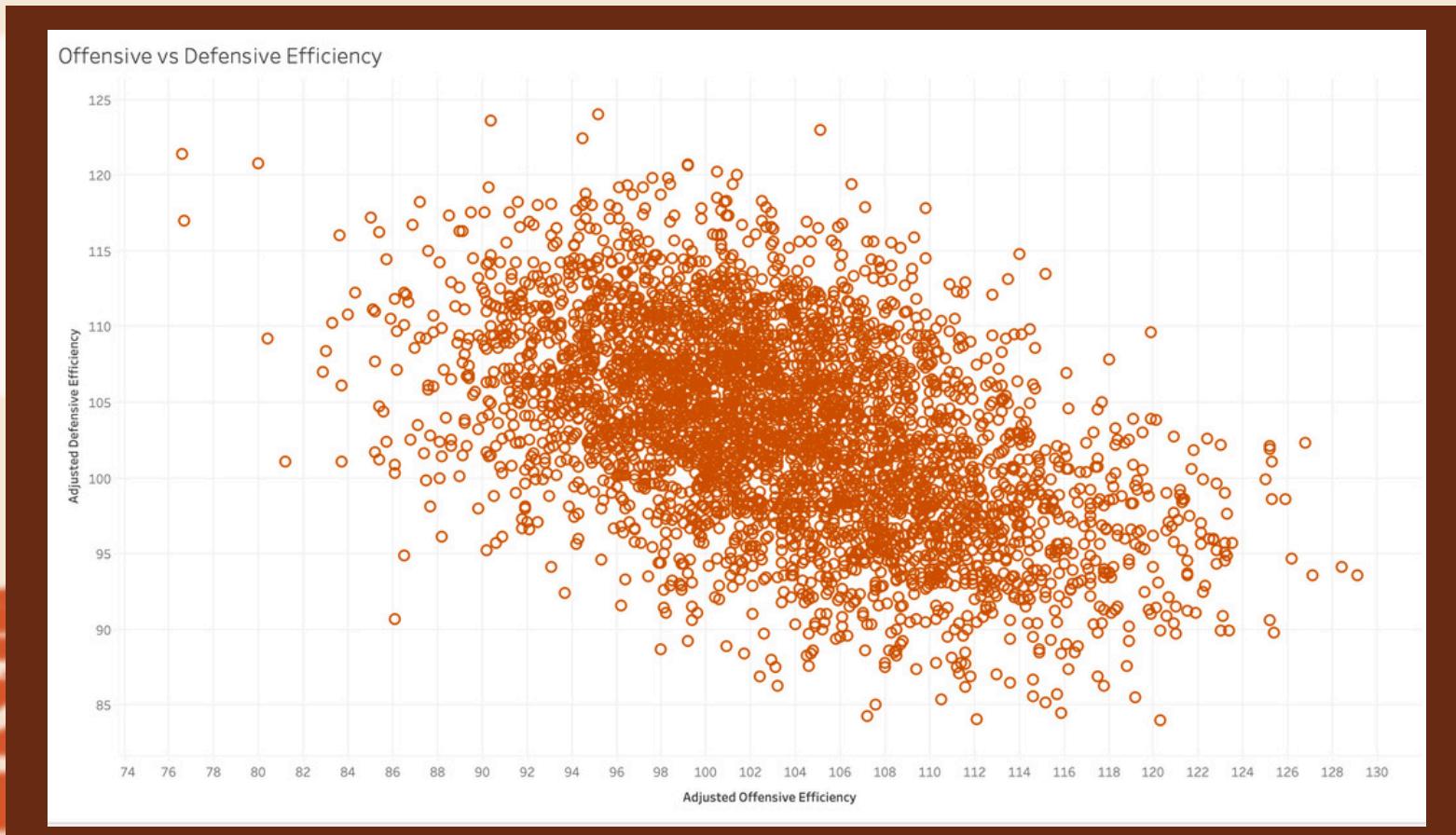


Graphs

Description:

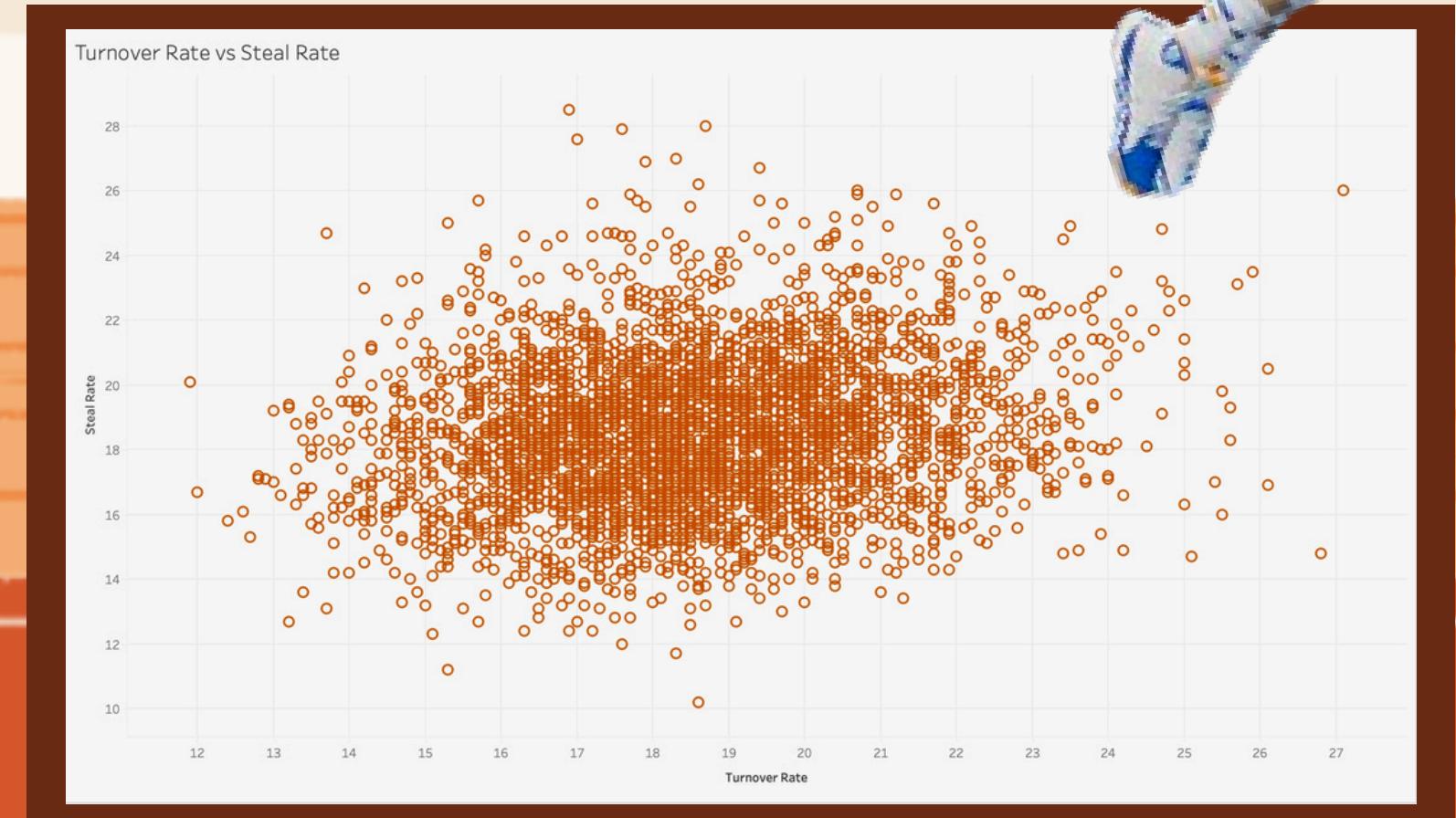
There's a weak negative correlation: teams with stronger offenses tend to have slightly better defenses (lower ADJDE), but the relationship is not strong.

- Top right: High offense, weak defense
- Bottom left: Weak offense, strong defense
- Bottom right: Well-rounded, elite teams
- Top left: Struggling on both ends



Description:

The data points are widely scattered with no strong visible correlation, suggesting that while steals contribute to turnovers, most turnovers are likely caused by other factors (e.g., bad passes, shot clock violations). The dense, rectangular cluster of points indicates that most teams fall within similar ranges for both metrics, but there are outliers with particularly high turnover or steal rates.



Graphs

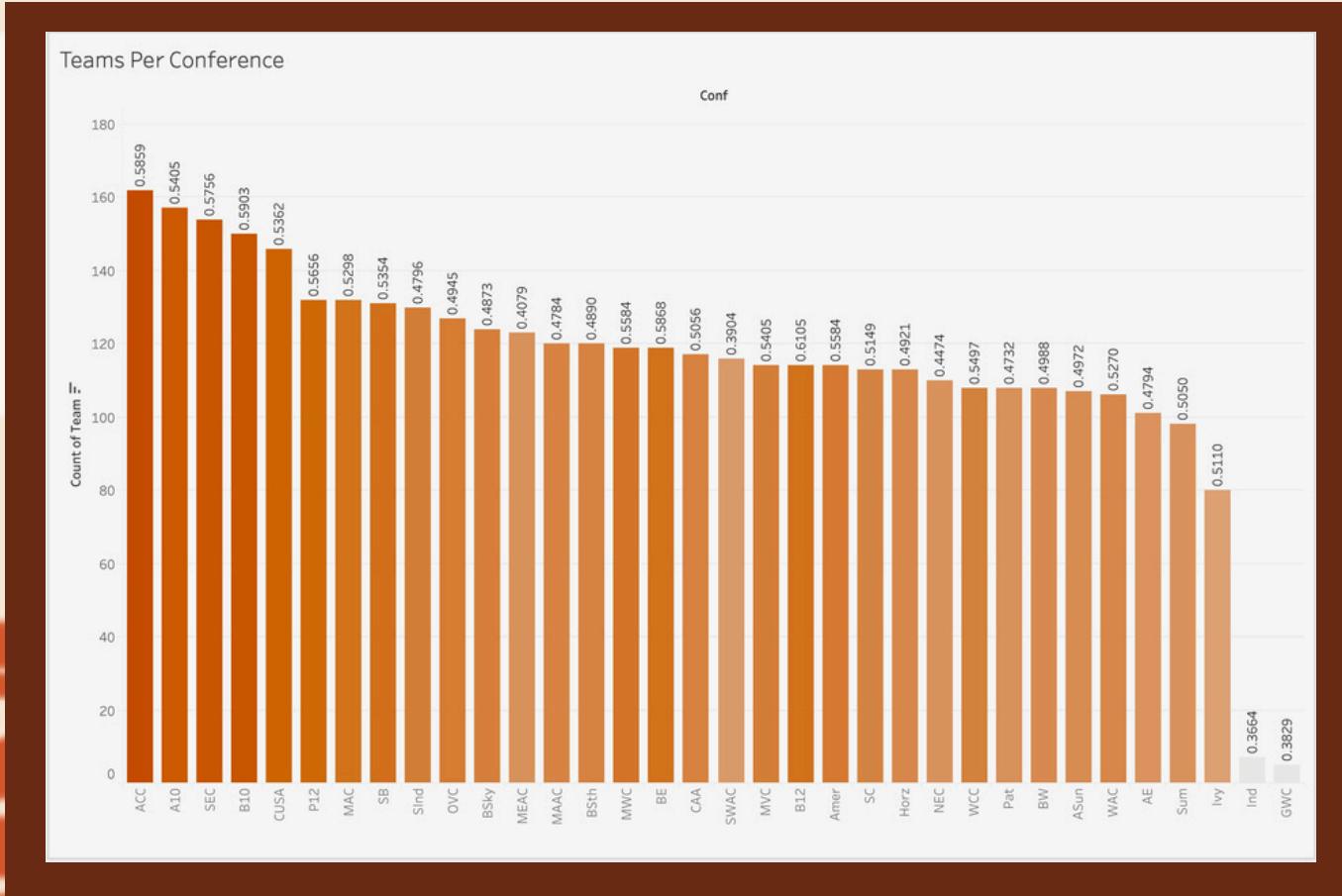
Description:

Each bar represents a different athletic conference, sorted from the highest to the lowest number of teams.

The ACC (Atlantic Coast Conference) has the most teams, followed closely by the A10, SEC, and Big East.

On the lower end, smaller conferences like the Ivy League, SWAC, and Great Midwest have the fewest teams represented.

The chart highlights the distribution disparity across conferences, which can impact competitiveness, scheduling, and postseason representation.

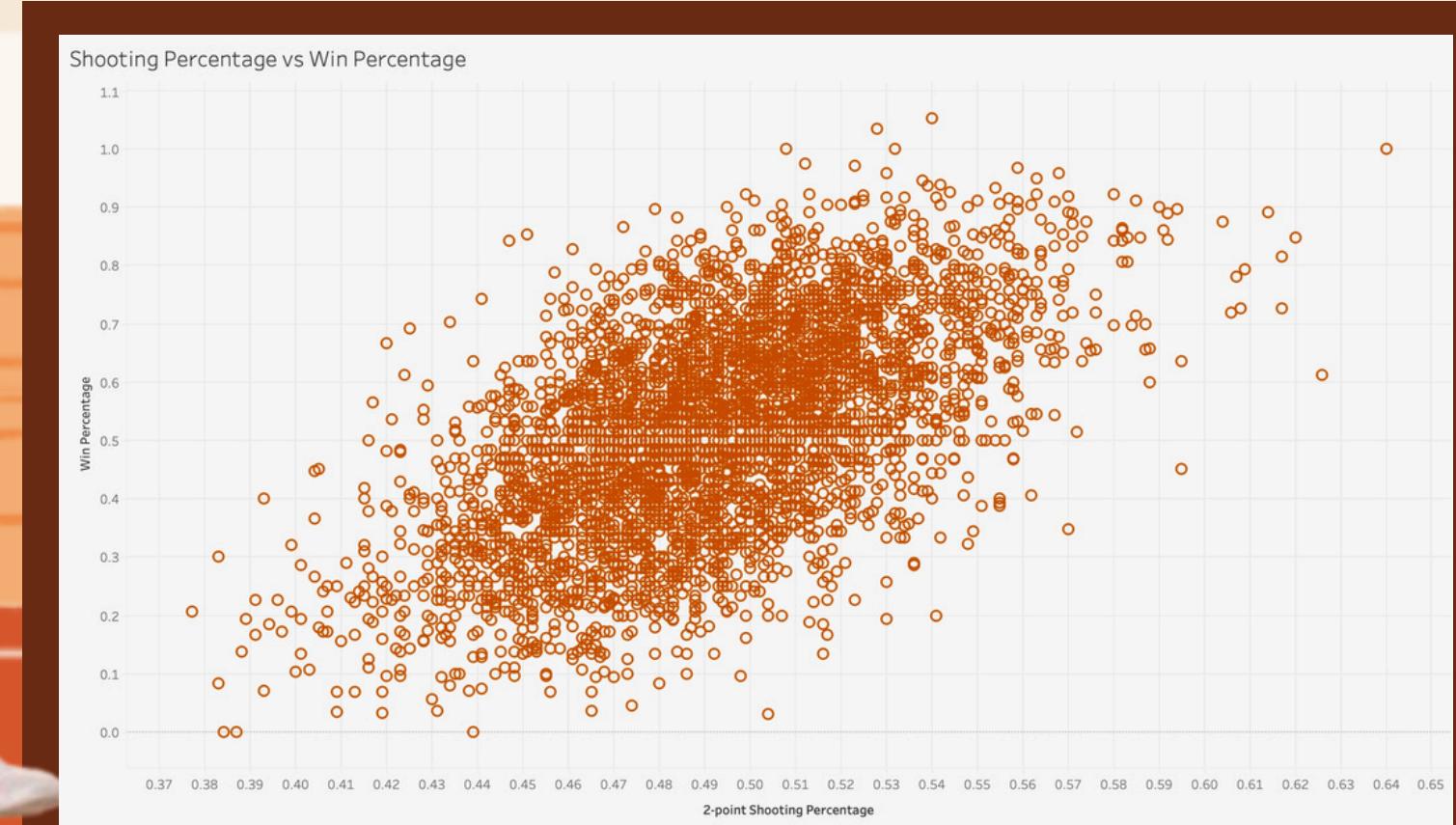


Description:

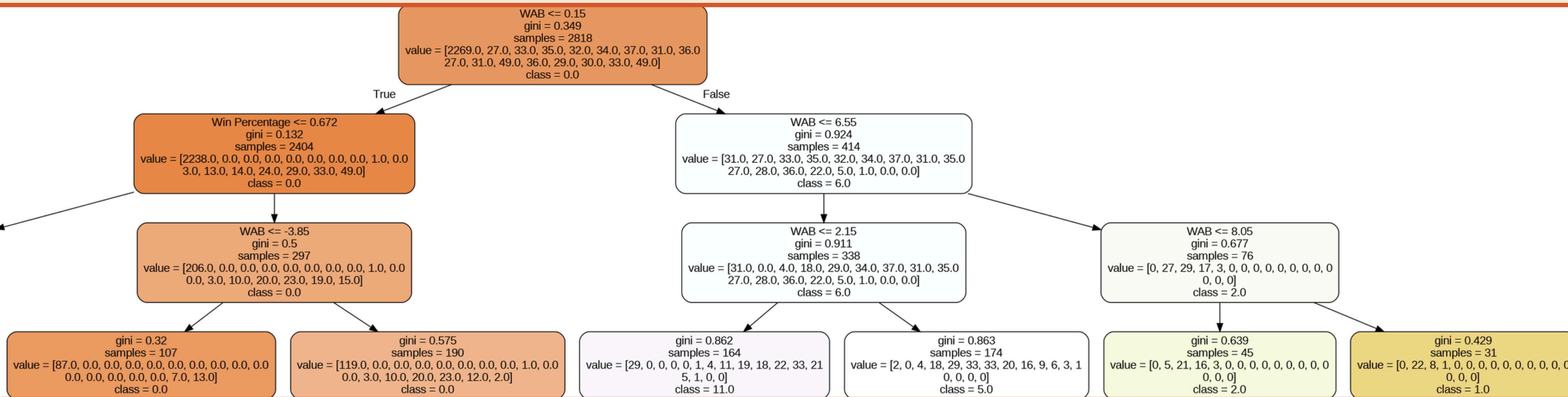
There is a moderate positive correlation: teams that shoot more efficiently inside the arc tend to win more games. As 2-point shooting percentage increases, win percentage generally rises as well.

The clustering of points suggests most teams fall in the mid-range for both metrics, while a few outliers show extremely high shooting or win rates.

This visualization highlights the importance of efficient scoring in driving team success.



Decision Tree



Accuracies for Each Algorithm



- Decision Tree: **0.8326241134751773**
- K-Nearest Neighbors: **0.8312056737588652**
- Naive Bayes: **0.7205673758865249**
- Random Forest: **0.8368794326241135**

Seed Predicted !!

