Model Metrics file

Purpose of modeling activities

Predict a winner from the 2021 Super Bowl season, by predicting the scores of each team in the Super bowl and comparing their points.

Data needed and/or available:

[NFL Team Stats 2002-2021 (ESPN) via Kaggle](#)

Data Timeframe:

09/05/2002–12/27/2021

Modeling Response:

Prediction - actual result from model = accuracy

Score for each team in the Super Bowl during the game.

Type of Model:

Regression Model

What Deliverables will be generated:

- PDF outlining modeling process from data exploration to best model results
- Model Metrics File detailing the model and its scope

Data Preprocessing steps of note

- Checked for missing values
- Created basic model data frame
- Created advanced model data frame
- Dealt with missing values

Model Description

- Input data size = 5357 observations
- Total features = 39
    - Model Algorithms and Parameters:
        - Ridge Regression Model with GridSearch
        - Elastic Net Regression Model with GridSearch
        - Random Forest Regressor with GridSearch

Model Metrics file

Model Performance

- *Ridge Regression*:
  - Best hyperparameter value for the Ridge model is"{'alpha': 1}
  - Mean Squared Error: 34.572282765535036
- *Elastic Net Regression*:
  - Best hyperparameter values for the ElasticNet model is: {'alpha': 0.05, 'l1_ratio': 0.05}
  - Mean Squared Error: 34.84804385229307
- *Random Forest Regressor*:
  - Best hyperparameter values for the Random Forest Regressor model is: {'n_estimators': 400}
  - Mean Squared Error: 5.0775927984794444

Model Findings

- Feature Importances include "Score" which was the best indicator for the target.

Next Steps

- New hyperparameters were tested and different modeling methods chosen.
- The Random Forest Regressor Model showed the best performance with an emphasis on high dimensional/complexity for the data.