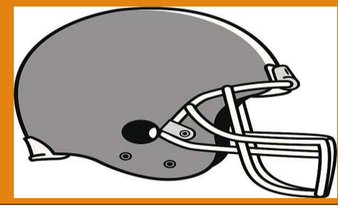
# Applying Machine Learning Techniques to Generate Fantasy Football Player Predictions




## Background

- Fantasy football is a game in which the participants serve as owners and general managers of virtual professional American football teams. The competitors select their rosters by participating in a draft in which all relevant National Football League players are available.
- Following the draft, each week the virtual manager selects their players to play in the starting lineup or to sit on their bench. Throughout the season the lineup needs to be adjusted to account for bye weeks, poor performance and injuries.
- Waiver wires are when new players get selected to play on the virtual teams. How do new players get prioritized? They do so by a ranking, which can be based on a combination of sportscaster advice, rankings and tribal knowledge.

## Data Science Development Process

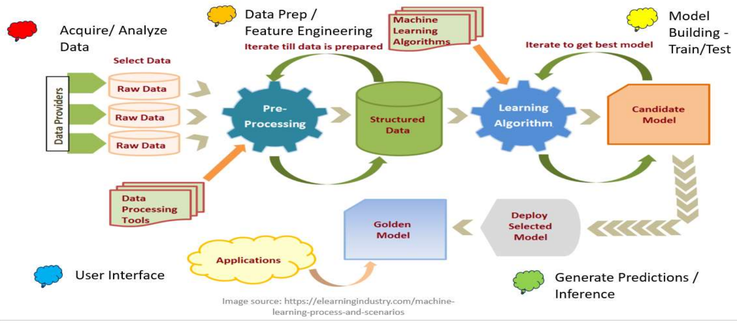



Image source: https://elearningindustry.com/machine-learning-process-and-scenarios

## Development Process Applied to Generate NFL Player Predictions

1. Data Acquisition/Analysis: Download player Stats at the end of each week. Join the data with the previous weekly stats.
   - (Can also include Analyze Data, reinspecting for changes, quality issues, etc.)
2. Feature Engineering: Calculate Cumulative Sum for each player statistic.
3. Model Building: Train/Test Model Using newest acquired data
4. Inference: Generate forecasted stats for the next week.
   - Decompose into raw stats.
   - Calculate forecasted Fantasy points using player formulas.

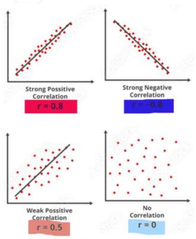
QB Actual Points Formula:

$$(passingyards * 0.04) + (passingtouchdowns * 4) + (interceptions * -2) + (receivingyards * 0.1) + (fumbles * -2) + (receivingtouchdowns * 6)$$

5. Deployment: (Aspirational) Deploy final model predictions into format for other users to interact with.

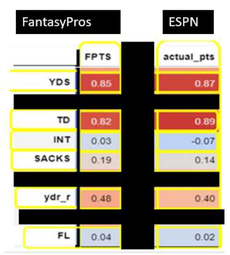
## Data Acquisition & Data Analysis: Correlations between KPIs and Fantasy Points

Correlation Key:




Images source

Correlation Matrix for QB




KPIs

**Key:**
YDS = Passing Yards

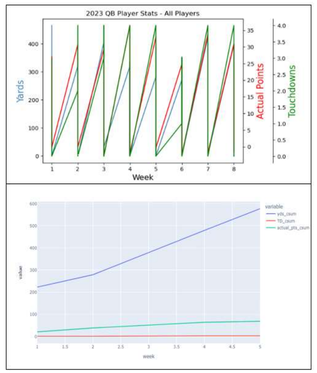
TD = Touchdowns
INT = Interceptions
SACKS= Quarterback is tackled by the defense

Ydr_r = Receiving Yards
FL = Fumbles Lost

FPTS = FantasyPros.com Total Score
actual_pts = Actual Score of the players as per ESPN.com

## Pre-processing & Feature Engineering

- Transformation of raw data into relevant variables.
- The goal of pre-processing and feature engineering is to improve the performance of machine learning (ML) algorithms.
- How this concept was applied to the project: Create cumulative sum of statistics for better accuracy in forecasting of player stats



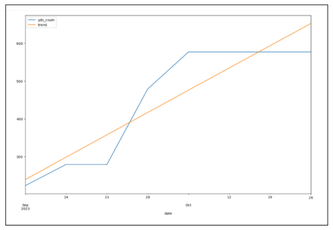

## ML Models Tested & Rated Performance

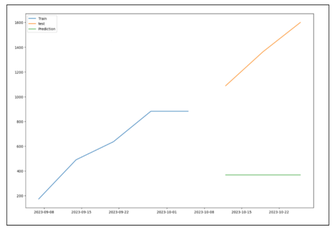
| Purpose | Model | Feasibility | Performance* |
|---|---|---|---|
| Smoothing | Holt-Winters Exponential-Smoothing | Worked as expected – similar results to cumulative sum | Smoothed inputs to the model |
| Determining Stationarity | Augmented Dickey-Fuller Test | Worked as expected – Test for stationarity | Helpful to determine if trends are time dependent |
| Forecasting | Holt-Winters Simple Exponential Smoothing | Did not perform well, but room for optimization | AIC: 2707 |
| Forecasting | ARIMA | Used Auto ARIMA for parameter optimization & performed well | Possible benefit over long term data acquisition. AIC=36.989 - 3354 |
| Forecasting | SARIMA | Did not perform well, since no seasonality | RMSE: 1421, AIC: 3353 - 4636 |
| Forecasting | Holt's Linear Trend Method | Has potential, but so far has not performed as good as ARIMA | Bias against spikes in performant stats, but performed well on smoothed data. AIC: 93.204 |

*(**AIC**) is an estimator of prediction error and thereby relative quality of statistical **models** for a given set of data.

## Inference & Model Evaluation

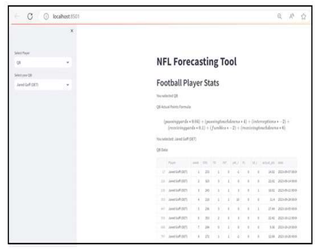
- Use the model to generate predictions on specific player statistics, for example Passing Yards.
- Evaluate the model's ability to handle previously unseen data

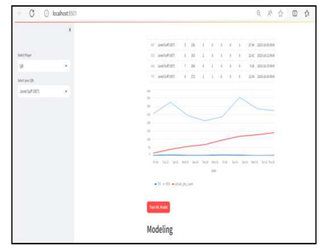




## Deployment to User Interface

- Makes the data & models more accessible
- Streamlit python package - enables website creation w/o html experience
- Users can select a player type and forecast a specific players' future points
- Adding the ARIMA models allow the models to be rebuilt each week and can be used to generate predictions & create new rankings.




## Data Issues & Challenges

Injuries
- Ranges from very hard to impossible to predict
- Can happen to great players with no previous history of injuries
- Turf vs grass theory – not tested (some propose turf should be banned)

Short Term Data
- Is not efficient for modeling (normally forecast models require 1 year of data to be accurate)
- BYE weeks introduce inconsistencies

Long Term Data
- Not necessarily representative season-to-season due to changes in coach, team dynamics, skill progression, etc.

Other Considerations
- Variables associated with environment at the time of play are not considered (temp/humidity, indoor vs outdoor stadium, etc.)
- Missing coaching strategy, actual plays, who will be on the field handling the ball during game day ahead of time

**Jessica C. Arket - jca1802@rit.edu**