

NFL Combine + NCAA Football Data Analysis



Kevin Gross

October 13, 2022

Agenda



- Overview
- Business & Data Understanding
- Modeling
- Evaluation
- Recommendations & Next Steps

NFL Scouting Combine and NCAA Football Stats

- NFL Scouting Combine

- Data from 2009 to 2019.
- Week-long showcase occurring every February at Lucas Oil Stadium in Indianapolis.
- College football players perform physical and mental tests in front of National Football League coaches, general managers, and scouts.
- ~320 athletes attend per year by invitation only.

- NCAA Football Stats

- Data from 2009 to 2013.
- 57 quantifiable data points per player.
 - Rush yards, rush attempts, pass completions, solo tackles, sacks, forced fumbles, etc.
- Used ~1,700 player's data who also attended the combine.

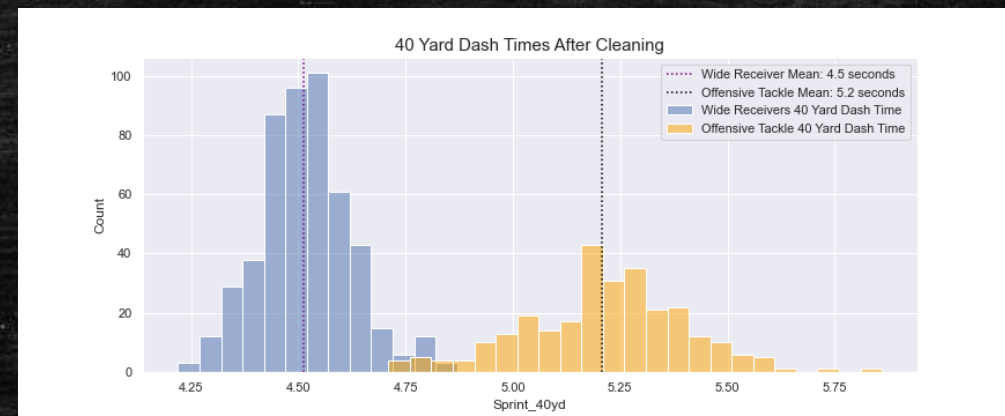
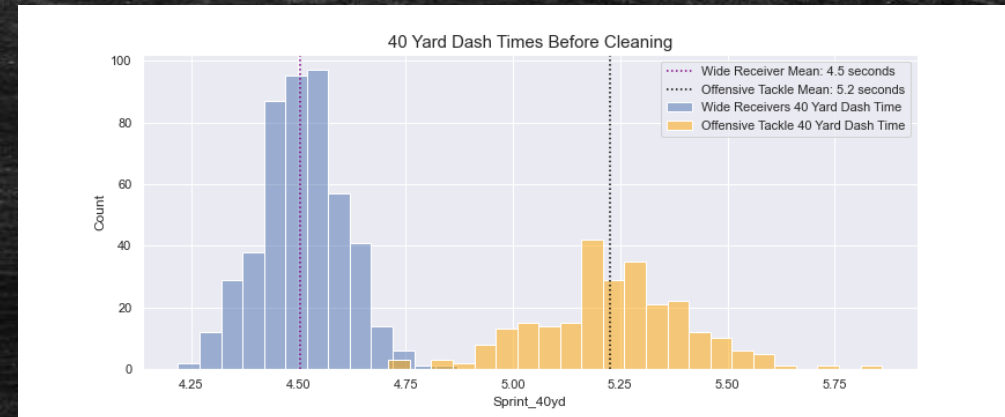
Business Understanding



- The NFL is a big money, high stakes business and every year the schedule kicks off with their marquee event, the **NFL Draft**.
 - The NFL draft is where collegiate players get selected to one of the 32 teams in the National Football League.
 - Over the course of 3 days, 259 players are selected based on a multitude of factors.
- The rigorous selection process utilizes millions of dollars and countless hours of analysis all to figure out who will bring the most amount of talent to a team.
- Why is this selection process so intense?
- Due to this high stakes business, **the stakeholders for this project are all 32 of the teams in the National Football League.**

Modeling

- Sci-Kit Learn:
 - Iterative Imputer
 - Grid Search CV
- 4 Types of Models:
 - Logistic Regression
 - K-Nearest Neighbors
 - Decision Trees
 - Random Forest



Evaluation

NFL Combine Models

Model Name	Accuracy Score	Precision Score	Recall Score	F1 Score
Baseline LogReg	0.675	0.701	0.858	0.772
LogReg L1	0.678	0.704	0.858	0.773
LogReg L2	0.675	0.701	0.858	0.772
Baseline KNN	0.626	0.666	0.837	0.741
KNN Model 2	0.652	0.651	0.984	0.783
KNN Model 3	0.647	0.662	0.916	0.769
KNN Model 4	0.623	0.658	0.856	0.744
Baseline DTC	0.602	0.690	0.688	0.689
DTC Model 2	0.639	0.641	0.991	0.779
DTC Model 3	0.654	0.653	0.982	0.784
DTC Model 4	0.654	0.671	0.903	0.770
Baseline RF	0.679	0.691	0.903	0.783
RF Model 2	0.651	0.650	0.986	0.783
RF Model 3	0.659	0.657	0.977	0.786

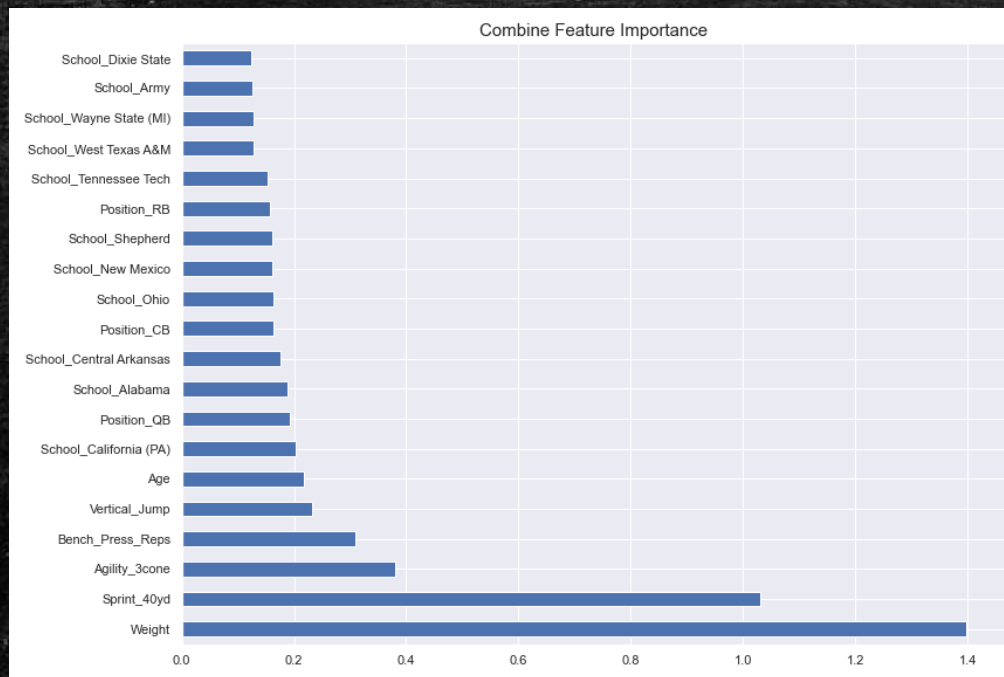
NCAA Football Stats with NFL Combine Models

Model Name	Accuracy Score	Precision Score	Recall Score	F1 Score
Baseline LogReg	0.710	0.758	0.843	0.798
LogReg L1	0.710	0.750	0.860	0.801
LogReg L2	0.698	0.754	0.826	0.788
Baseline KNN	0.650	0.709	0.823	0.761
KNN Model 2	0.689	0.700	0.949	0.806
Baseline DTC	0.666	0.763	0.737	0.750
DTC Model 2	0.664	0.733	0.795	0.763
DTC Model 3	0.636	0.710	0.785	0.746
Baseline RF	0.696	0.711	0.932	0.806
RF Model 2	0.696	0.698	0.976	0.814

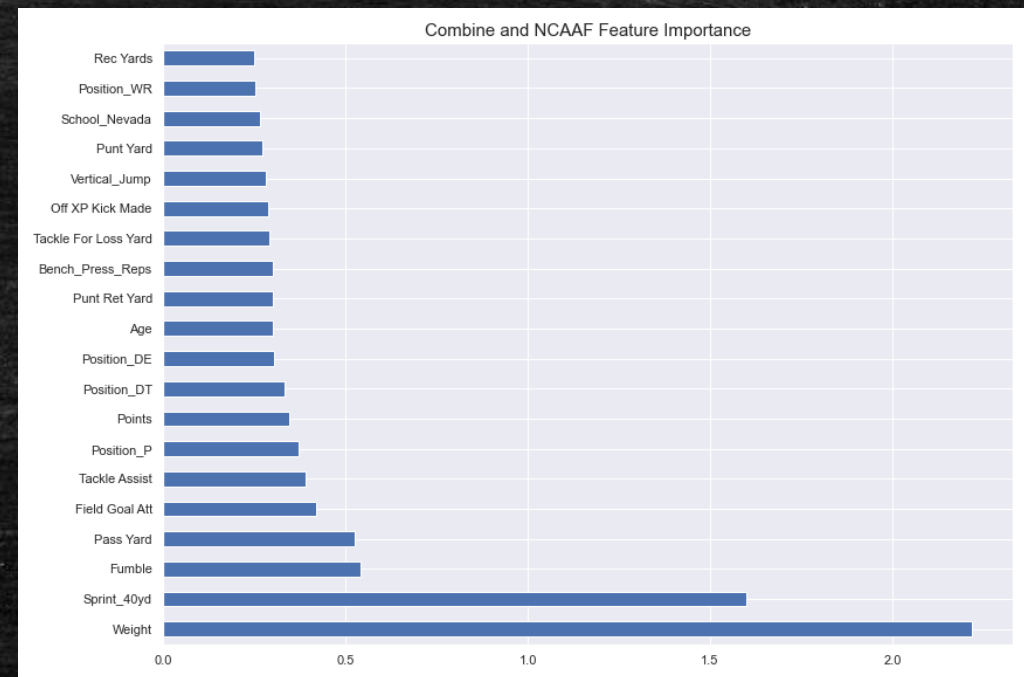


Evaluation

NFL Combine Models



NCAA Football Stats with NFL Combine Models



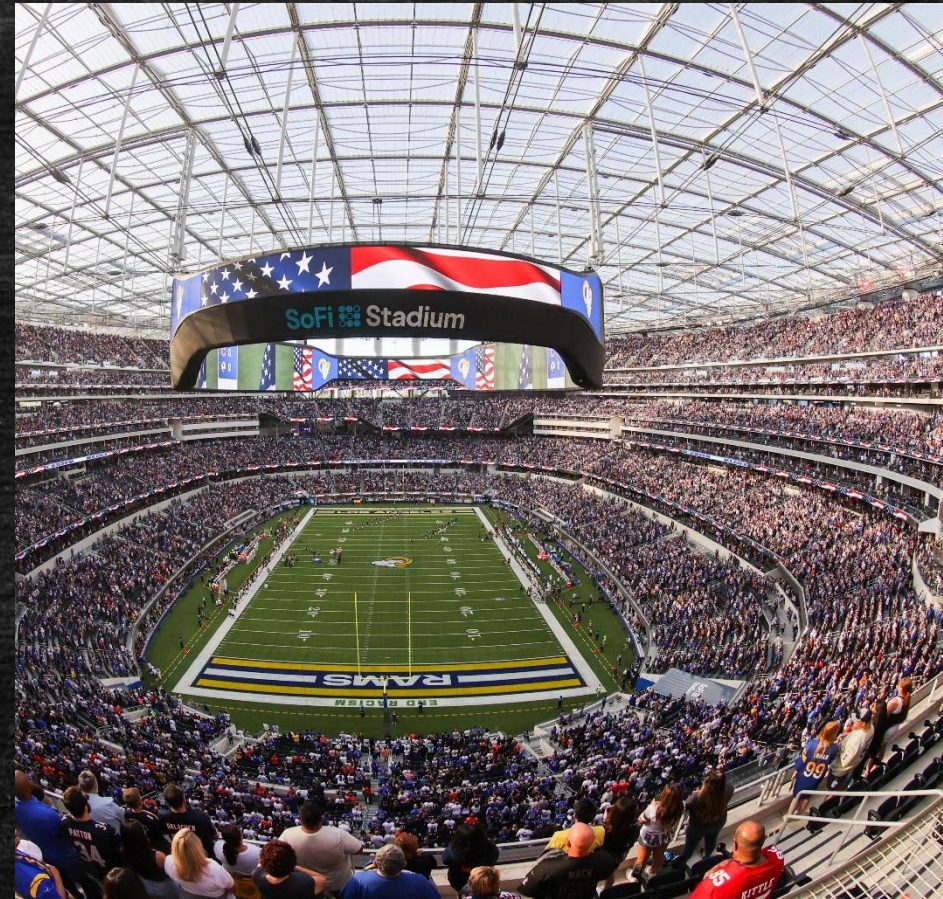
Results



- Dummy Model: 64.8% accuracy
- NFL Combine Model: 67.8% accuracy
- NCAA Football Stats with NFL Combine Model: 71.0%

Recommendations & Next Steps

- Hard to predict whether a player will be drafted based purely on combine and collegiate stats:
 - More info than physical attributes
 - Every position is extremely different
 - Combine is the best of the best
- Avoid multi million-dollar draft mistakes



Thank you!



Questions?

- Kevin Gross
- kevgross89@gmail.com
- github.com/kevgross89