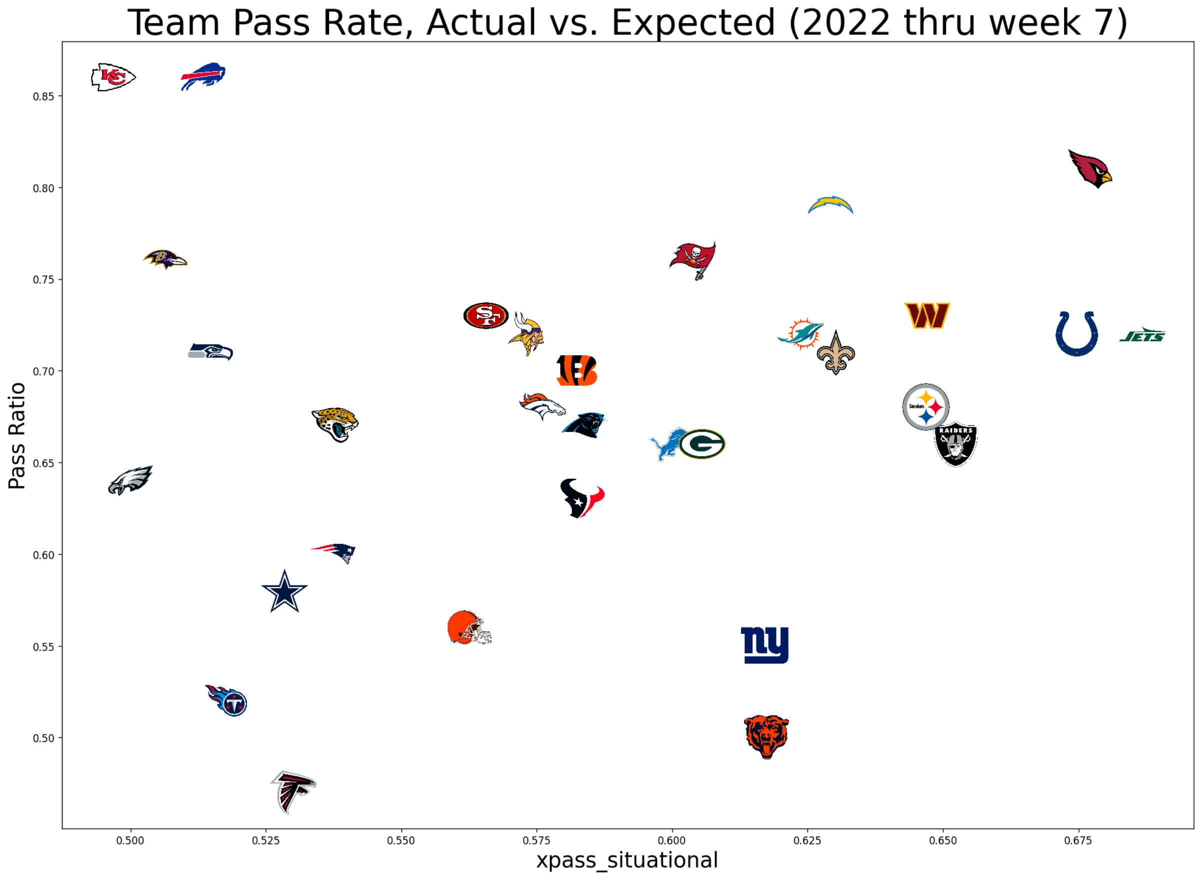
Run-Pass Oracle (RPO)

I. Introduction

This submission aims to build a better methodology for predicting, pre-snap, whether a team will pass the ball. To do this, we created the **Run-Pass Oracle** model, built on top of a LightGBM architecture. We managed to achieve performance that surpasses state-of-the-art literature, while still maintaining interpretability.

Much of this performance is driven by feature engineering, including metrics to track pre-snap factors such as **Motion Intentionality**, **Defensive Congestion**, and **Contextualized Tempo**. These features heavily utilize the provided tracking data, and resultingly show a big boost over extant models, many of which barely eclipse 70% accuracy, those that do coming with major caveats. Perhaps most notable among thesse is Ben Baldwin's expected pass model; however, as the graphic below demonstrates, its predictions are still noisy:



Baldwin's model encompasses important context like down and distance, and knows a little about team strength due to using Vegas odds. However, it doesn't incorporate the rich tracking data utilized in our model, and also misses significantly on teams like the Chiefs and Bills, whose superstart QB's make them far more pass-happy than expected. By incorporating tracking data, as well as other contextual metrics, our model shows nearly a ten-percent boost over earlier literature in both a test and validation set, providing hope for generalizibility while still demonstrating immediate usefulness.

II. Engineered Metrics & Other features

a) Feature overview

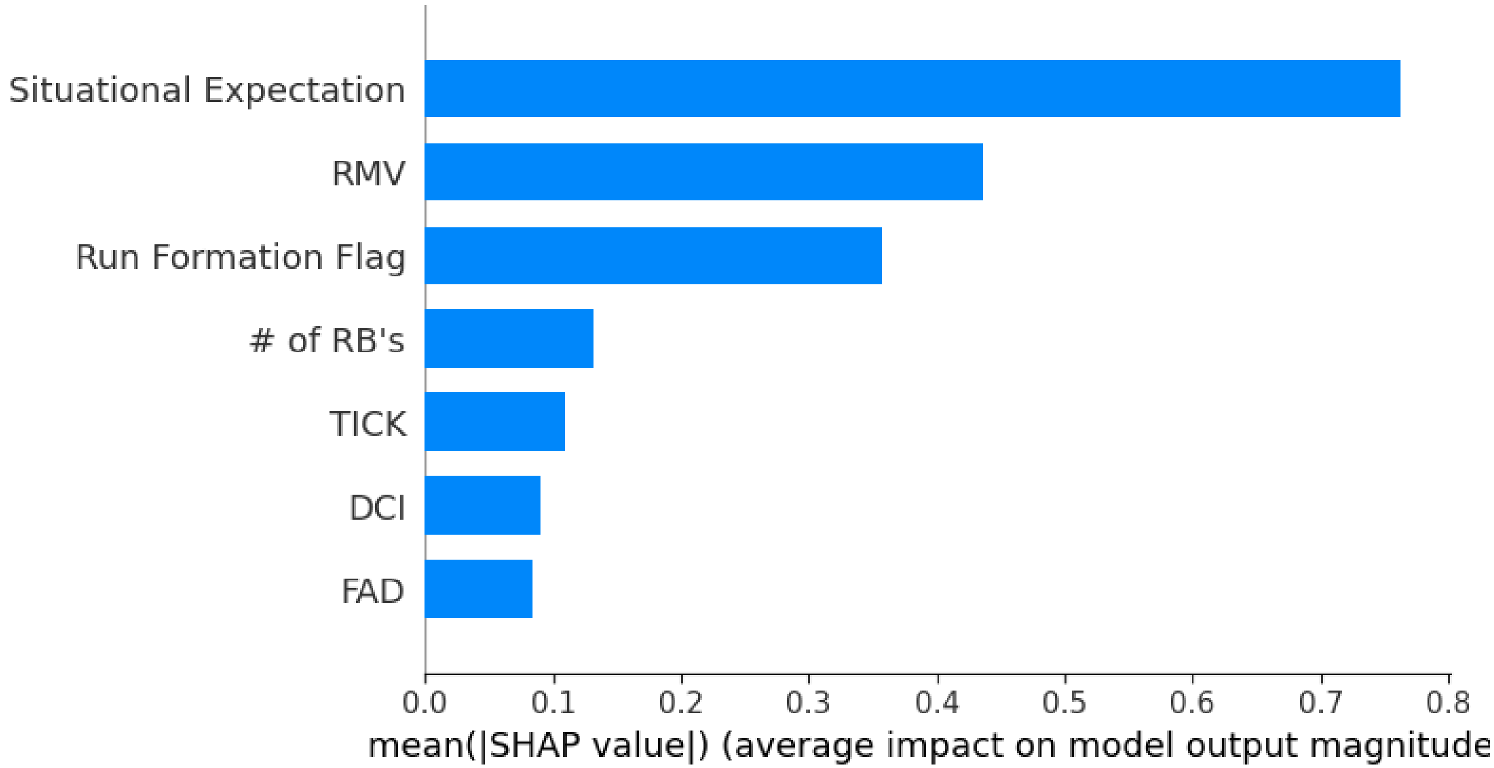
Our model encompasses six features, many of which aggregate less-important features into more informative metrics:

- **Situational Expected Pass Rate** draws Ben Baldwin's pass rate model. While it is the dominant feature in our model, encompassing many other important features, it still is only at about 70% accuracy on its own, and fails to explicitly encompass tracking data & other player info.
- **Motion Favorability**: The delta between number of players in motion post-lineset and the number of players who shifted. (TODO: Do we exclude RB2? Is that worth a small performance dent to snazz up the model?) While this seems fairly simple, it is by far the most informative approach we have found to incorporating motion into our model. (todo: display failed approaches?) This is because it is more important to know how many players are in motion (todo: break out a chart with multiple in motion vs. single vs. none?) vs. other stats like acceleration or distance covered. Similarly, since motion generally implies passing, and simple shifts imply runs, if both occur on the same play, then getting the delta between them concisely informs our prediction.
- **neg_Formations** Captures which formations negatively impact pass rate. (TODO: cite which)
- **TICK**, or more verbosely, **Tempo Including Contextual Knowledge**. This metric combines our **Tempo** metric with other contextual info, such as player weight at key positions, to understand how team personnel and play-calling tendencies combine to predict pass rates (TODO: Figure what exactly I put inside this)
- **Defensive Congestion**, i.e., the mean pairwise distance between players on defense, with more recent frames weighted more heavily. The clear benefit of this is that we can pick up nuanced info that coarser features, such as offensive formations (e.g., 2x2) and men-in-box counts, may miss.

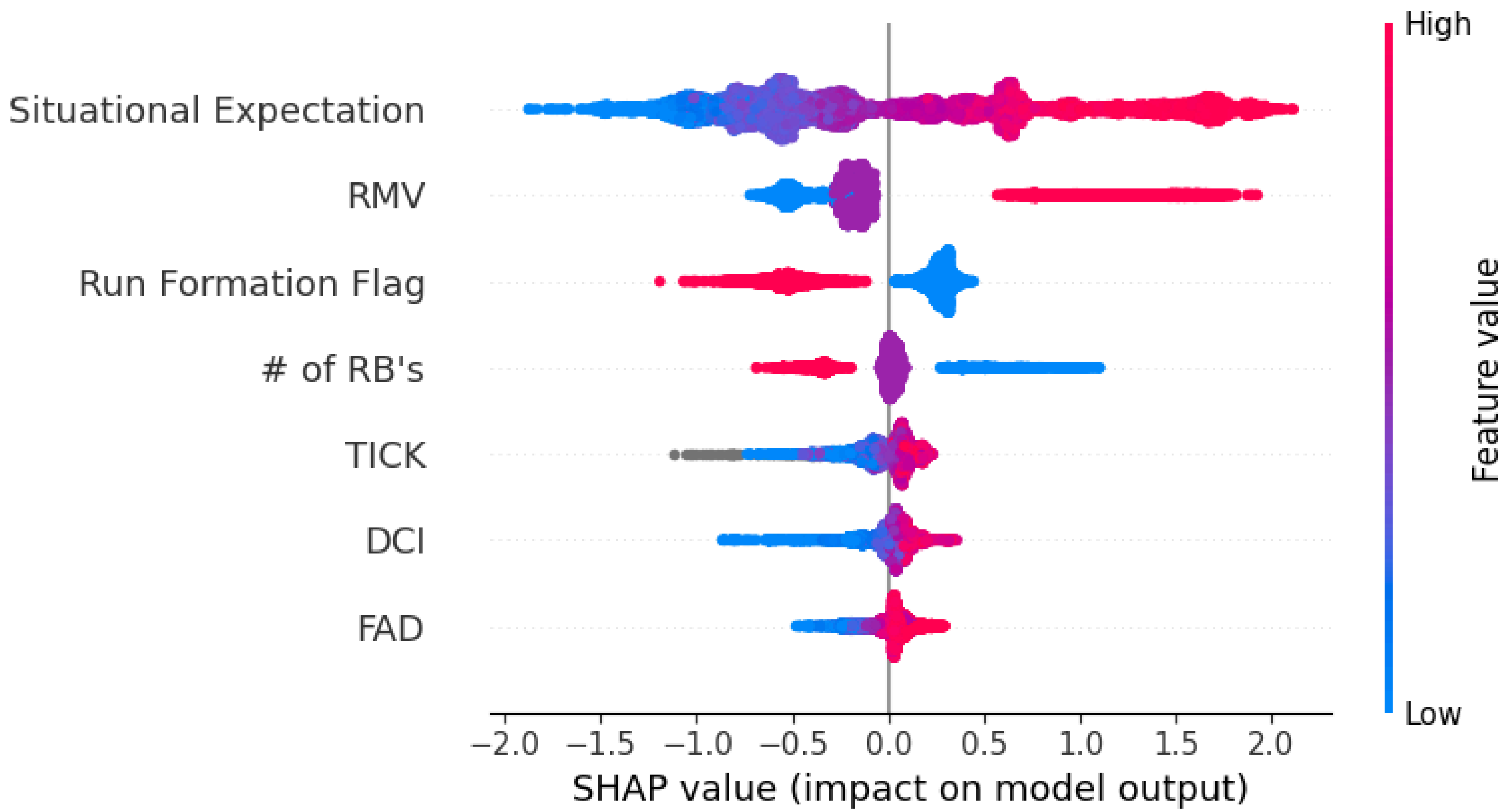
TODO: find graphic for here; maybe dots demonstrating 'defensive congestion index?

b) Feature impact analysis

Below are our shapley values for our dataset. Given the breadth of features that our **Situational Expected Pass** metric encompasses, it makes sense why it would be so important. It does not, however, render the other features entirely useless, as they each, on average, still makes significant contributions to our prediction values.



The below graph explores Shapley values from a different angle, looking at a broader distribution of impact (vs. just average):



Given most of our features are positively correlated with passing, we see an expected result where higher feature values (red) imply higher a pass rate (more red points to the right). The inverse—i.e., low (blue) values depressing pass rate—is also generally true, excepting our **negative formations** feature, which, when flagged as True, should lower expected pass rate (see next section). One key thing to note is how while the mean Shapley values paint a meaningful disparity between our four lower features, they each exhibit similar distributions in our violin plot, and are thus closer to equal importance than the prior graph would imply.

c) Examining Formational Impact

Our model heavily relies on the concept of "negative formations", i.e., formations that heavily favor runs. This feature is an important example of how necessary it is to encode concepts that, however obvious they may seem to humans, are still necessary to teach our models about. One of these formations is **I-Formation**, a heavier personnel grouping used almost exclusively for runs or play-action that is perhaps *the* stereotypical run formation. Similarly self-explanatory is **Pistol**, used to great effect by teams like the Ravens to take advantage of QB Lamar Jackson's running prowess. Less explicable is the **Single-back** formation's inclusion (TODO: Explain it)

In [1]:
```
import plotly
plotly.io.read_json('image/cincy_drive.json')
```

The above graph shows just how impactful these 'negative' formations are in depressing our expected pass rate. This is an example from a **Cincinnati Bengals** drive (todo: cite which week?), one of the most pass-happy teams over the course of the 2022 season at an average 62% pass rate. Our model thinks that, given all the other information we provide it, most of the Bengals are more likely to pass than not on almost all the plays of the drive.

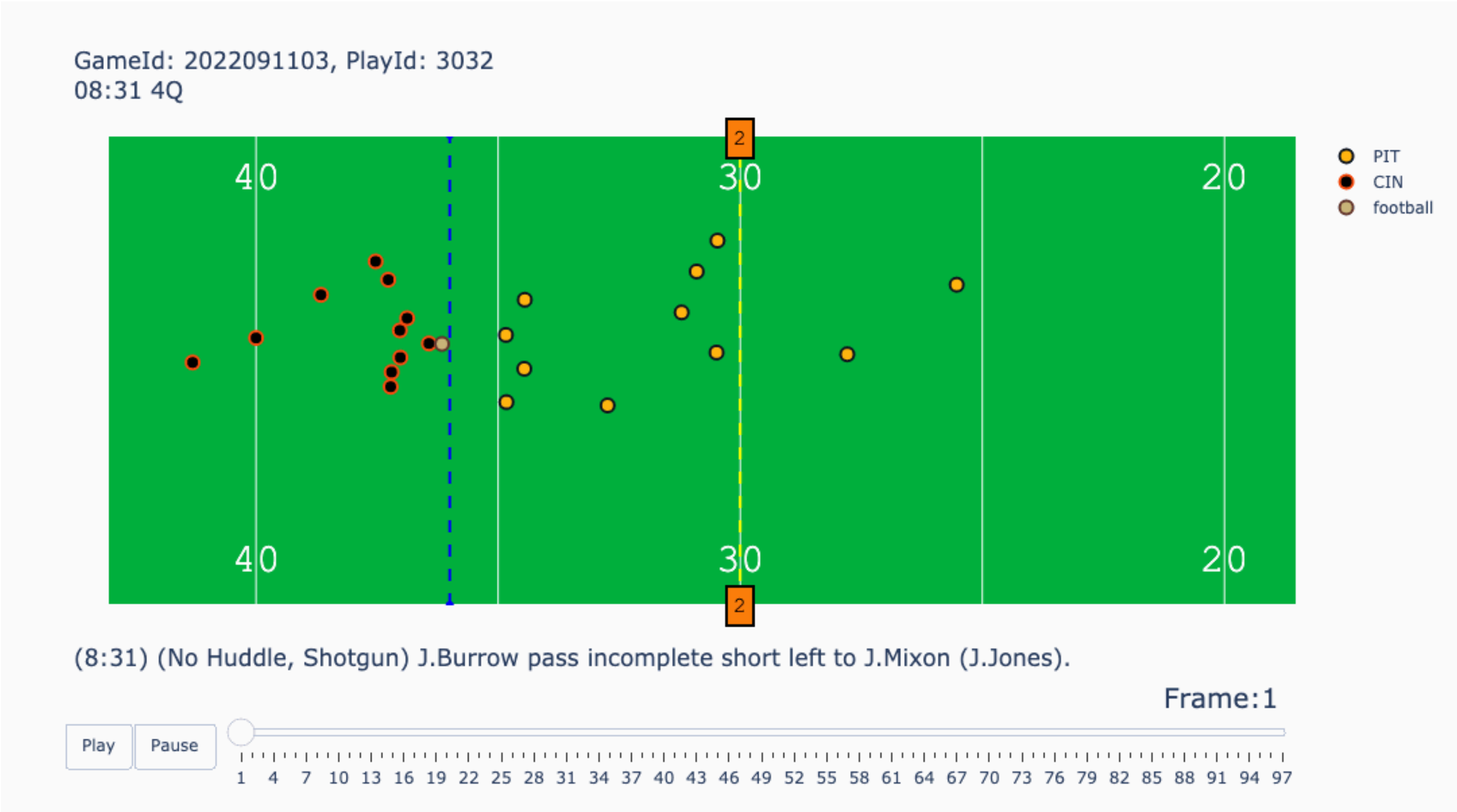
This of course tracks with the common understanding that Bengals QB Joe Burrow strongly prefers to pass from Shotgun, perhaps implying a departure from this could significantly depress the expected pass rate. Yet the extent to which knowing a formation is negatively related to passing depresses our expected pass rate is remarkable, and clearly informative.

TODO: maybe find a way to line up what actually happened here? find a way to cleanly present as a series of dots or something what actually happened vis a vis run pass?

d) Motion

Below is a typical example of pre-snap motion from a Week 6 matchup between Miami and Minnesota. In it, we see Miami (left, offense) running back Raheem Mostert move across the formation right before the snap. Such an example is where our **FAD (Final Acceleration Difference)** metric proves most useful, since it tells us how much faster the likely motion player (i.e. whoever has the highest acceleration in the last few frames pre-snap) is going than the second-fastest player. Since only one player can be in-motion when the ball snaps—though many teams push this rule to its limits—this intuitively helps us suss out whether a player went in motion on a play.

The whole purpose of **FAD**, however, is to clean up inconsistencies in **RMV (Rectified Motion Value)**, one of our most useful features. RMV is seemingly simple, subtracting the number of shifted players from the number in motion. Yet it it's also surprisingly effective, in large part because it encompasses the main things we want to know about pre-snap movement. The biggest benefit is that it essentially removes some noise from our data, since many of the players shifting pre-snap are erroneously market as "in motion", due to the speed at which they shift. Thus, while it seems coarse, **motion-momentum** helps us differentiate shifts from actual motion plays.



Above is such an example, where a quick exchange on the left side results in two shifting players erroneously being marked as "in motion". We see at the end of our graphic, however, that there's a real motion occurring, which we want our metric to reflect. Thus, by subtracting what our data says are 2 shifted players from our 3 in motion—as RMV does—we get the actual correct number of players in motion, which is 1.

III. Model

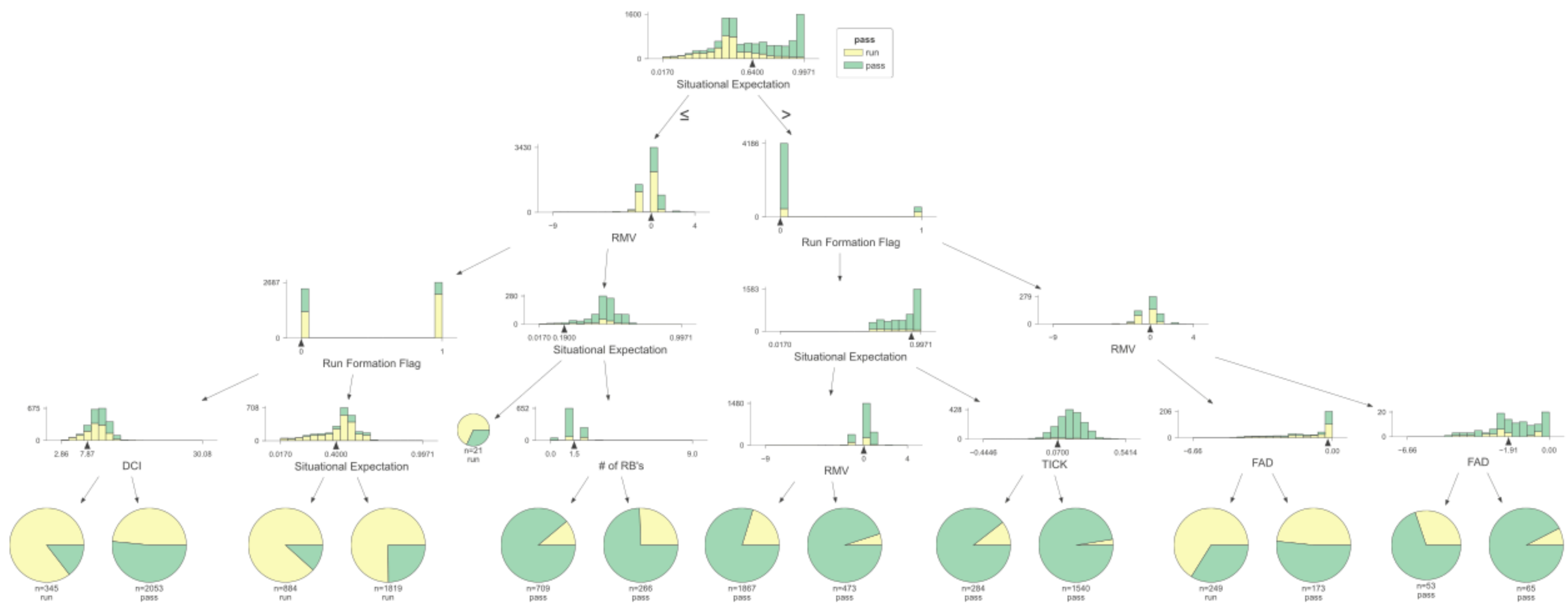
a) Model design

Our model itself was a LightGBM model, with a maximum depth of 5 as to prevent overfitting. For the 9 weeks of tracking data provided, we trained from weeks 3-7, using week 8 as a validation set, then week 9 as test proper.

The first two weeks are excluded for their heavy reliance on prior-year information, which is much noisier than current-season info. This is an issue acknowledged by members of the football analytic community such as FTN's Aaron Schatz, who provides weighted version of his DVOA metric that weighs recent weeks more highly as the season goes on, and thus becomes more predictive.

Thus, given that our model is set up to predict on mid-season plays, fitting on more recent data is empirically sound. The benefit of this is seen in how the model gets a 3% performance boost in test over validation after re-fitting, showing the likely benefit of additional data to our model.

(TODO: maybe find better visual aid here. tree feels out of place/underwhelming)

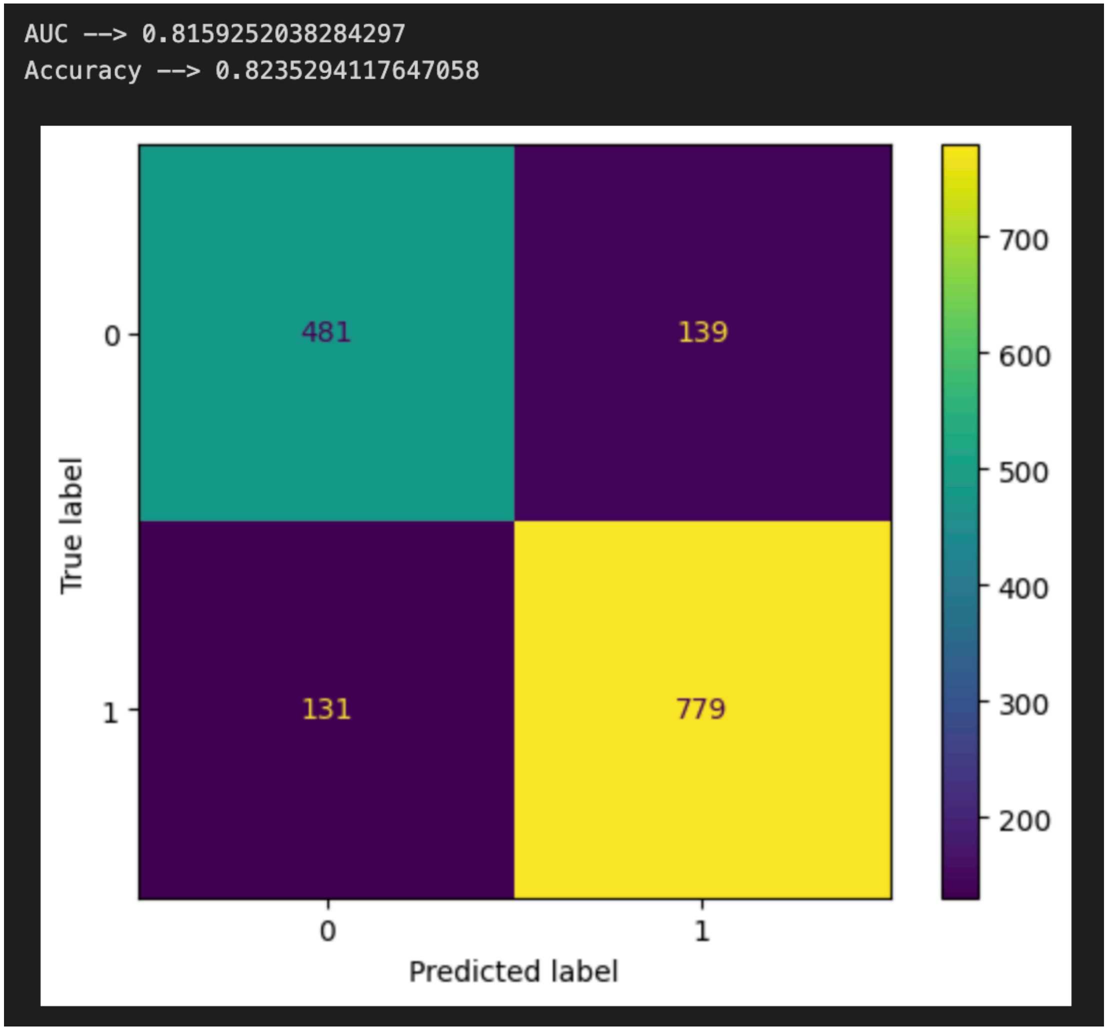


The above graphic demonstrates what a typical tree for our LightGBM model looks like. Though formational and situational (down, distance, etc.) info are quite important, it's crucial to note that all our created metrics are utilized. That we can incorporate all these features to a meaningful extent while limiting our maximum depth to 4 (to prevent overfitting) bodes well for our model design.

b) Model performance

As previously discussed, our model beats most play-level benchmarks in the public literature for play-level pass prediction. The model we present is, in our opinion, the best blend of interpretability and generalization we've achieved.

Our model shows a slight bias toward predicting passes; however, this is to be expected, since the general tendency is for teams to pass more frequently. Notably, its misses are essentially equal between events that were runs and passes, suggesting a decent generalizibility.



IV. Conclusion

V. Citations

Joash Fernandes, Craig et al. 'Predicting Plays in the National Football League'. 1 Jan. 2020 : 35 – 43.

Goyal, Udgam. (2020). Leveraging machine learning to predict playcalling tendencies in the NFL.

Marius Ötting, Predicting play calls in the National Football League using hidden Markov models, IMA Journal of Management Mathematics, Volume 32, Issue 4, October 2021, Pages 535–545, https://doi.org/10.1093/imaman/dpab005

In [ ]: