NFL Big Data Bowl

"Intuitive football related feature engineering to reach top 1%"

Introduction

Main contributions

- Unlike some of the other top teams, that used complex models, we blend two simple models.
- 1. Fully connected neural network and 2. Light gradient boosting multi-class model. (Very common in kaggle competitions)
- However we were able to reach top 1% by creating variations of 2 very important football related features. (In addition to generic features)
- So the benefit of our approach is, we can extract important football related features for actionable strategy decisions.

The 2 important features

- 1. Area/ Distance the rusher is able to create in 0.5 and 1 second after the hand off.
- 2. The Area/ Distance the rusher will be able create in 0.5 and 1 second after the hand off if the rusher moves 1 yard to the right or left,

Definition of Area/ Distance Created

First what is Area / Distance

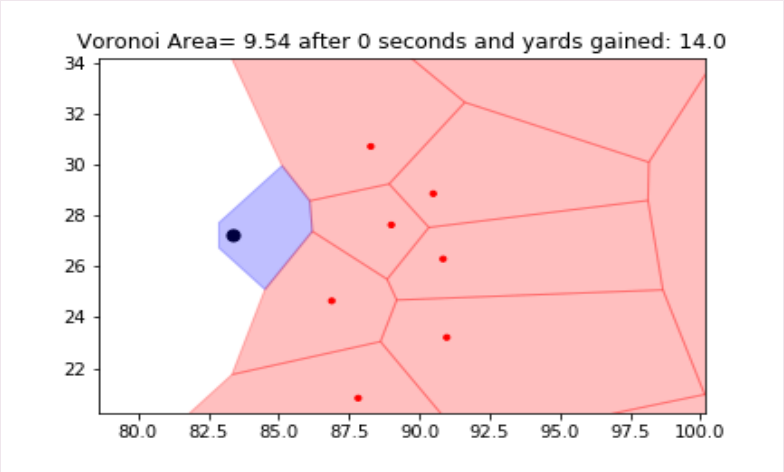
- Area = Voronoi Area of Rusher (excluding other offensive defence players and in front of rusher)
- Distance = Max X of Voronoi Area of Rusher (excluding other offensive defence players)
- Similar to https://arxiv.org/pdf/1906.01760.pdf

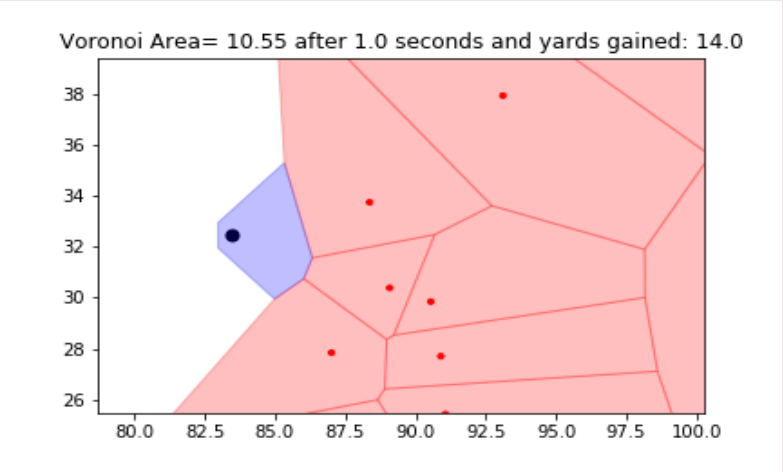
Now what do we mean by Area/ Distance created

- Area Created = Area(after t seconds) - Area(time of handoff)
- Distance Created = Distance(after t seconds) - Distance(time of handoff)
- We derive these feature with t = 0.5 and t=1 seconds.
- In other words area/ distance created after 0.5 and 1 seconds.

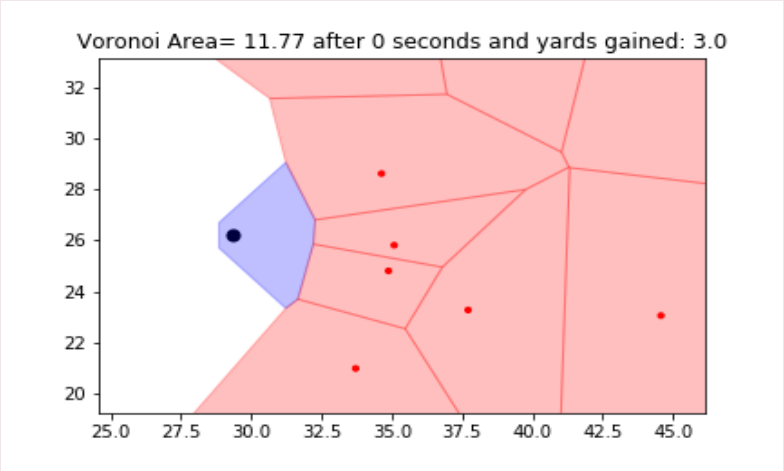
Visualizing Area/ Distance Created

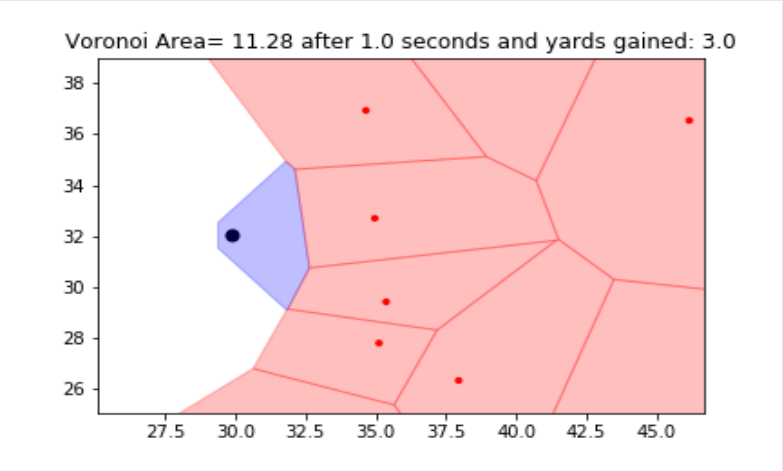
Example Play 1 (long play) with 14 yards gained





Example Play 2 (short play) with 3 yards gained





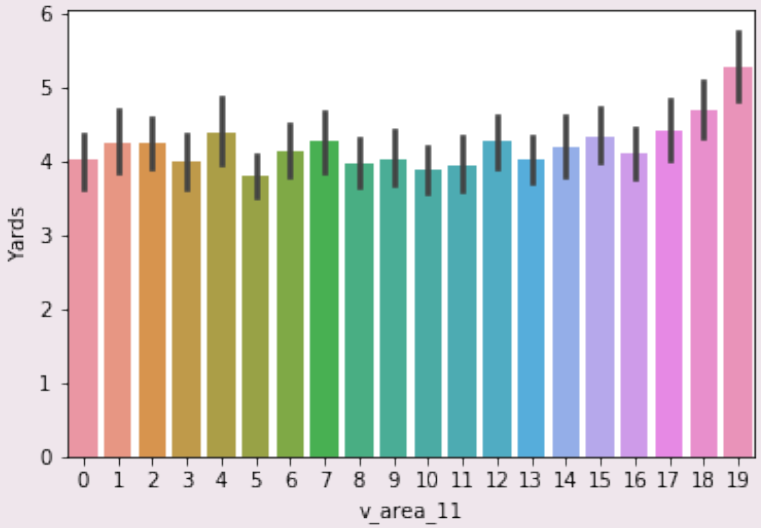
Interpretations of the above graphs

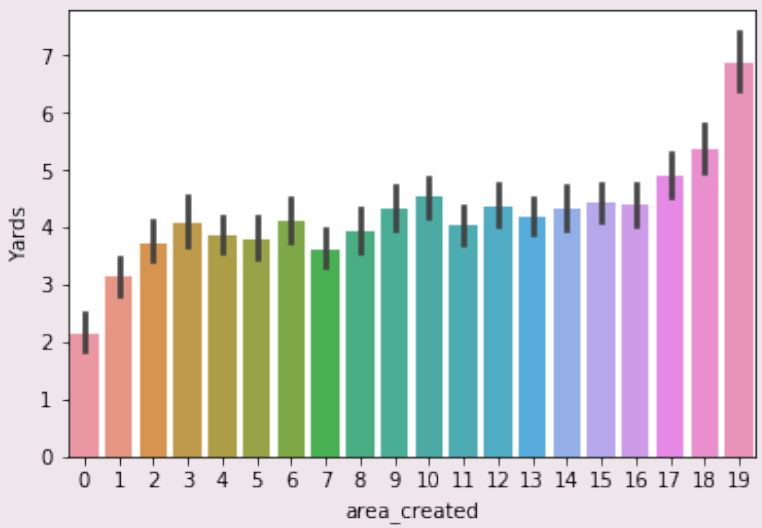
Difference in the above two plays

- In play 1 the voronoi area of rusher was less than the the voronoi area of rusher in play 2 at time of hand-off as well as after 1 second.
- However play 1 resulted in 14 yards because the rusher is projected to increase the voronoi area in play 1
- In contrast for play 2 despite having a larger voronoi area at time of hand the play only resulted in 3 yards because the rusher is NOT projected to increase the voronoi area in 1 second.
- The key difference is that in play 1 increase in expected voronoi area is (+ 1.01) sq-yards whereas in play 2 the expected increase in voronoi area is (- 0.5) sq-yards.

Visual representation of feature vs yards gained relation

Voronoi area at handoff vs difference in voronoi area





- The left graph is voronoi area at handoff split in 20 equal size bins on the x-axis vs average yards gained on the y-axis.
- Similarly, the right graph displays the difference in voronoi graph instead of the voronoi area itself.
- We see a clear relation between the feature in the right graph and not so much in the left one.

Understanding the above analysis

What is important for a long rushing play?

- It not important how much area the rusher has at handoff.
- However, the amount of additional voronoi area that the rusher can create in 0.5 and 1 seconds is very important.
- These findings are consistent for area as well as distance created features (as mentioned above).
- These feature not only increase the model accuracy but can be useful for coaches in designing and understanding plays (discussed in the next slide).

Conclusion

How can coaches use this information?

- Coaches can design plays so that the rusher can create area and distance. Which can be done in 2 ways
- 1. Rusher should be moving to a open area even before taking the ball from the quarter-back.
- 2. The offense players should stop the defence players in manner that allows open area for the rusher.

Main contributions

- We create features that yield very good predictions while keeping the complexity of the model low.
- Our features give insights that can be used by coaches to understand and design rushing plays accordingly.
- In other words, our model is good and at the same time it is not a black-box model and provides football relatable insights.

References

Thank you! Below are references and code guides we used.

- https://www.kaggle.com/cpmpml/initial-wrangling-voronoi-areas-in-python
- https://www.sciencedirect.com/topics/earth-and-planetary-sciences/artificial-neural-network, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.699rep=re