NFL Big Data Bowl 2024: Predicted Tackle Probability

By Matthew Rabin

Introduction

The National Football League has been working in partnership with Amazon Web Services since 2017 to capture "real time location data, speed and acceleration for every player, every play on every inch of the field". NFL's Next Gen Stats have enabled new insights and analytics in football but have primarily focused on offensive metrics. This project uses machine learning to create a new defensive metric for tackle probability.

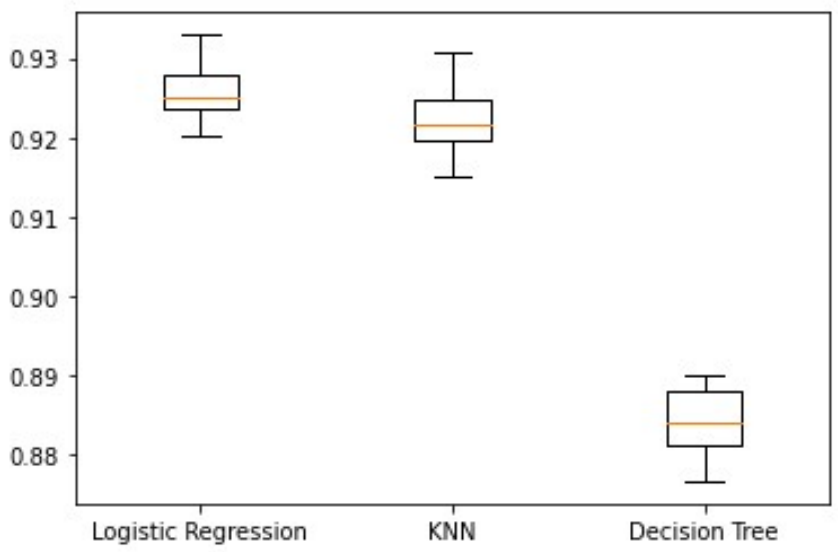
Data

Tracking data from the ball and the defensive players was used from each event on each play. Game event data was used to identify the team on defense. Tackles data was used as the outcome variable in a model to predict the probability of a defender making a tackle.

Methodology

Python was used to prepare the data and build the predictive model (see appendix). The features used for the prediction were player speed, player acceleration, distance from the player to the football, orientation of the player relative to the football and direction of the player relative to the football. A standard scaler was applied to convert the features to the same scale. The outcome was a binary classification for whether the defender made or did not make a tackle. The data was randomly split into two groups for training (80%) and testing (20%). The training data was further divided ten non-overlapping times into 90% training and 10% validation groups. This allowed for 10-fold cross validation to optimize model parameters and prevent overfitting.

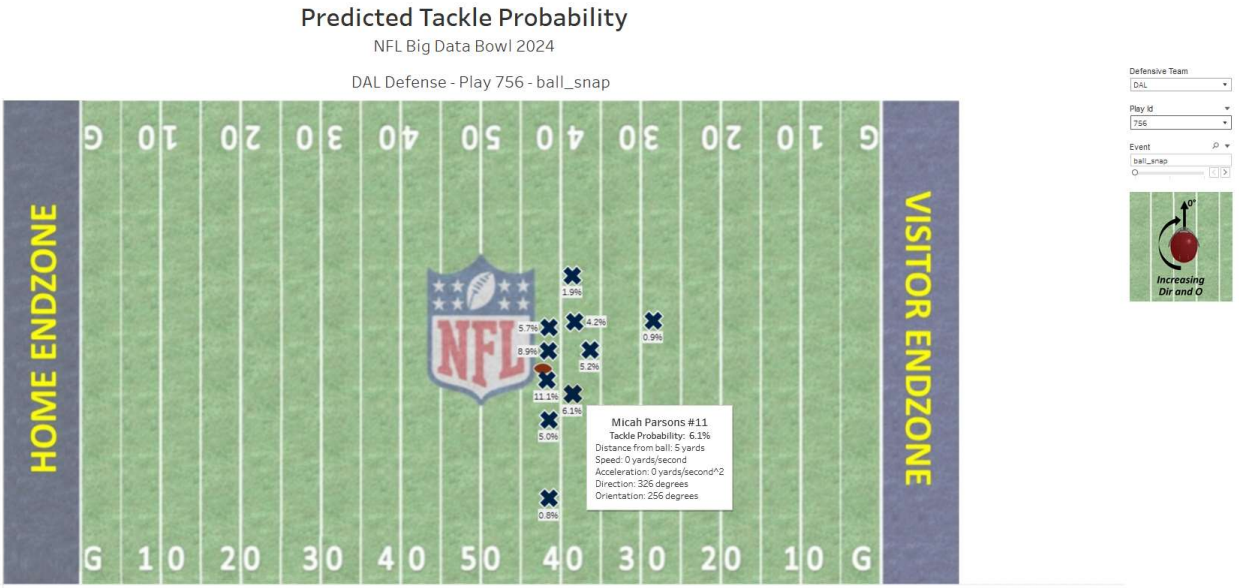
Three different machine learning models were built and tested to determine which one had the best performance. Models were built using Logistic Regression, K-Nearest Neighbors and Decision Tree classification. The Logistic Regression model was selected after returning the best results. The predict_proba method was used in python to estimate the probability a tackle based on the Logistic Regression model. The tackle probability metric was added to the original data and then visualized in a Tableau dashboard.




```
Logistic Regression Test Set Accuracy: 0.9274815290716351
KNN Test Set Accuracy: 0.9248313523931898
Decision Tree Test Set Accuracy: 0.884998393832316
```

Discussion/Conclusions

This research provides the framework for a new metric called “tackle probability” percentage. The Logistic Regression model performed best among the three classification models tested with 93% accuracy for tackle predictions. The tableau dashboard demonstrates insights that could be drawn from these predictions. It includes a graph of the football field with tackle probability shown for each defender. If the user hovers their mouse over a player then a tooltip will display with the features used to predict tackle probability. Users can filter by defensive team, play and play event. This new metric could be added to the Next Gen Stat library and used by NFL teams and fans.



https://public.tableau.com/app/profile/matthew.rabin/viz/PredictedTackleProbability/PredictedTackleProbability

Improvements/Future Work

Speed and acceleration of the football could be added to improve the accuracy of the tackle prediction. These were not used because directional data for the football was not available but the values could be inferred from other fields in the tracking dataset. Other classification models could also be tested such as Random Forest classification.

Acknowledgements

Professor Brian Hall and the Artificial Intelligence and Machine Learning course in Certificate of Sports Analytics at New York University.

References

Next Gen Stats quote: https://nextgenstats.nfl.com/glossary

Python resources: https://app.datacamp.com/learn/courses/supervised-learning-with-scikit-learn

Football image for dashboard: https://www.cleanpng.com/png-nfl-duke-blue-devils-football-miami-dolphins-ameri-6211659/

Team color codes for dashboard: https://teamcolorcodes.com/nfl-team-color-codes/

Appendix

Python code:

https://github.com/matthew-rabin/NFL-Big-Data-Bowl-2024/tree/main

Tableau dashboard:
https://public.tableau.com/app/profile/matthew.rabin/viz/PredictedTackleProbability/PredictedTackleProbability