Issues in Information Systems
Volume 23, Issue 4, pp. 279-292, 2022

DOI: https://doi.org/10.48009/4_iis_2022_124

Identifying factors that lead to injury in the NFL

Matthew Toner, *Bryant University, mtoner1@bryant.edu*
Suhong Li, *Bryant University, sli@bryant.edu*

Abstract

This study hypothesizes that injury-causing factors can be identified through training machine learning models with NFL injury data. The machine learning process entailed web scraping, pre-processing, cleaning, modeling, and analyzing NFL injury data to identify these factors. The features used to model injuries included the following: games played, games started, weight, height, age, year, years of experience, starting position, and team. The four models used to model NFL injuries were Logistic Regression, Decision Trees, Random Forests, and Gradient Boosted Trees. The model with the best performance was the Gradient Boosted Trees model, with an F1 score of 0.508. In addition, the Gradient Boosted Trees model selected teams, games played, and starting position to be the most influential factors when determining the probability of a player getting injured during the season. The most notable takeaway from this study is that the left guard position is the most injury-prone, followed by defensive secondary player positions such as safety and cornerback.

Keywords: athlete injury prevention, machine learning, undergraduate students

Introduction

Since the 1920s, the National Football League has moved away from being a brute, leather-helmeted game and transitioned into the sport we love today. Nevertheless, this inherently violent sport still poses serious health risks to players. Recent reports of severe brain injuries affecting the long-term mental health of players have placed the NFL under increased scrutiny from fans and players. For the NFL to survive as a sport, the league needs to find ways to minimize the risk of injury without sacrificing the physical contact relished by fans. The scope of this study seeks to address the incidence of injuries in the NFL to determine the factors causing injuries. This study illustrates that accurate data modeling of NFL injuries will prove insightful in identifying factors that cause injuries in the NFL.

The NFL has gained a poor reputation when it comes to protecting its players. In recent years, the league has attempted to regain the trust of its players by implementing rule changes such as increasing the number of yards a team receives for committing a touchback on a kickoff to discourage kick returns. This change was made because the kickoff is the most dangerous play in football. This literature review seeks to uncover other aspects of football that are more dangerous than others to predict player injuries. This review explores injuries to the head and body and identifies other factors such as race that contribute to the likelihood of a player being injured. This study hypothesizes that by incorporating these factors into a dataset and applying machine learning algorithms, factors that contribute to injuries in the NFL can be identified.

Issues in Information Systems
Volume 23, Issue 4, pp. 279-292, 2022

the highest injuries are Jets (72 injuries), Lions (66 injuries), Dolphins (65 injuries), Cardinals (63 injuries) and Texans (62 injuries). While the bottom five teams with the lowest injuries are Titians (30 injuries), Vikings (30 injuries), Falcons (29 injuries), Steelers (24 injuries) and Chargers (15 injuries). This range of injuries enabled the model to decrease the probability a player was injured if they played on the Chargers team.

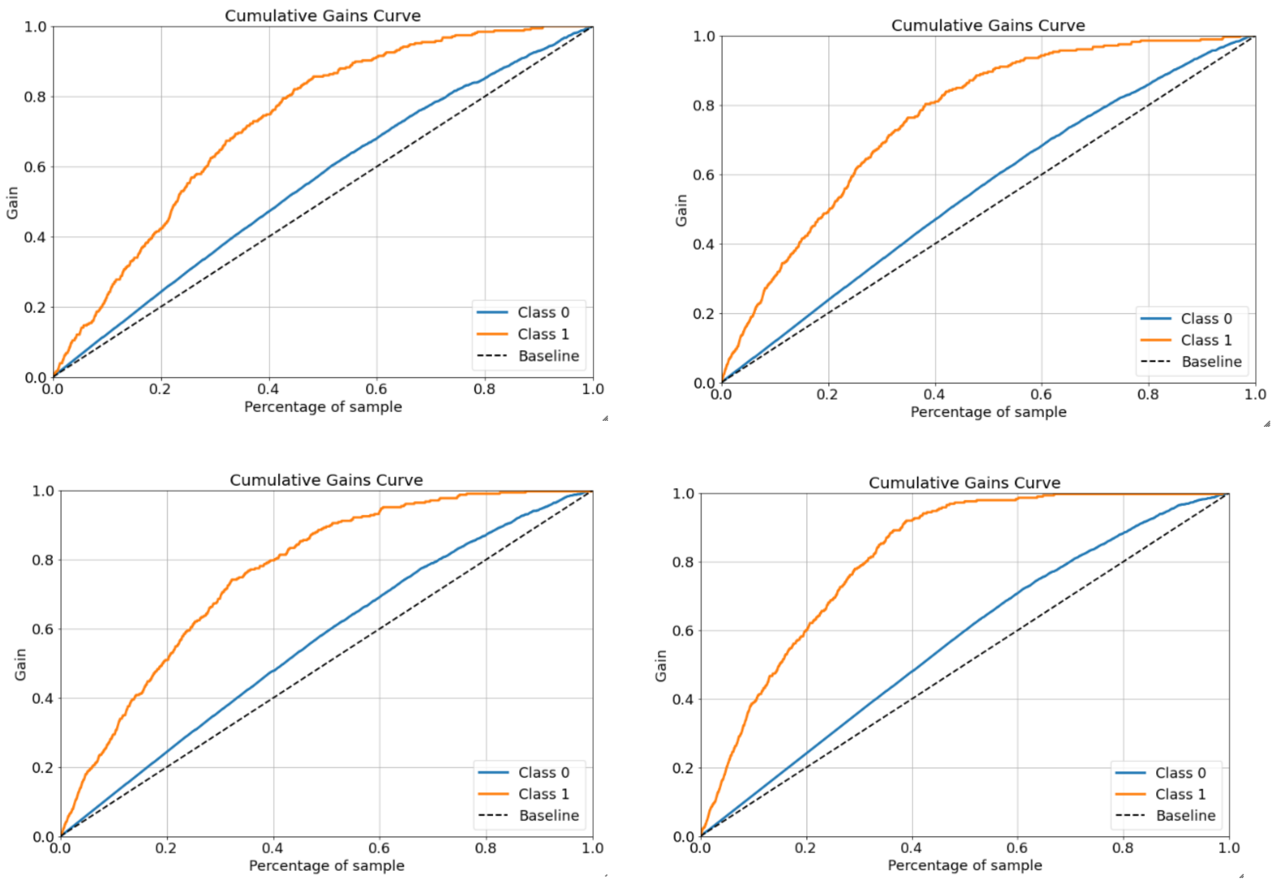

Figure 4 – Cumulative gains curves for each model. Logistic Regression (Top Left), Decision Trees (Top Right), Random Forest (Bottom Left), Gradient Boosted Trees (Bottom Right)

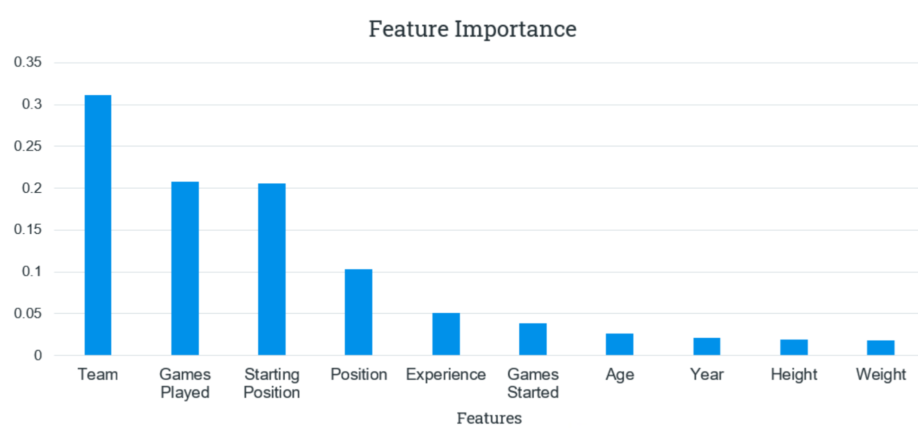

Figure 5 – Feature Importance from the Gradient Boosted Trees model

Games Played: Figure 6 shows the average number of injuries by the number of games played. Each number of games was started as its own column to analyze how the number of games started influenced injury occurrence. The y-axis in Figure 6 is the average of the injury label. The average over the sum was chosen to account for variation in the number of entries for each game played. Injury occurrence decreases as the number of games played increases because the probability of them being injured in that particular season diminishes as fewer games are left. This trend is most apparent after playing in 8 games, the halfway point in the season; this feature appears to follow a normal distribution. Therefore, it makes logical sense that the model could predict that a player was not injured if they played more games in the season than other players.

Starting Position: Figure 7 shows that the position with the highest ratio of injury occurrence was the left guard position. This offensive linemen position is responsible for blocking the running back on rushing players and protecting the quarterback on passing plays. Considering this position involves physical contact on each play, the selection of this feature suggests that players involved in more contact are more prone to injury. Furthermore, the positions making contact with the left guard are the defensive end (DE), defensive tackle (DT), and occasionally a linebacker (LB). All three of these defensive positions are also within the top 10 positions with the greatest injury occurrence. Figure 7 separates the offensive and defensive positions by color, with offensive positions in red and defensive positions in blue. Out of the top 10 injury-prone positions, two are offensive (LG, TE), whereas the other 8 positions are defensive. This preliminary finding suggests that defensive players making the tackles are more likely to be injured than the players being tackled on the field.

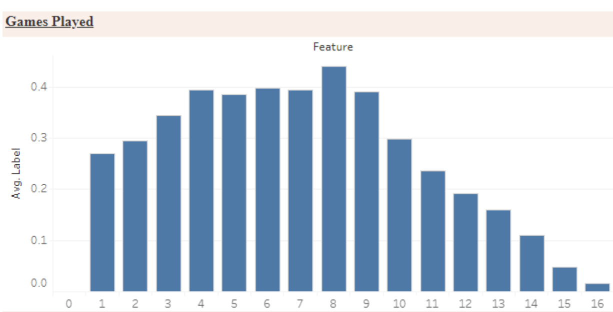

Figure 6 The average injury occurrence based on the number of games played

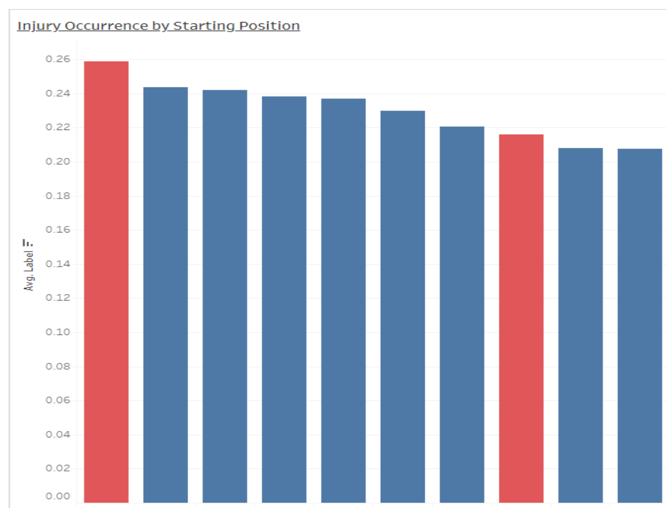

Figure 7 Injury Occurrence by Starting Position

Another important consideration is that 50% of positions with the highest injury occurrence are located in the secondary portion of the field. These positions include safety (S), free safety (FS), defensive back (DB), cornerback (CB), and strong safety (SS). These positions involve tackling players rushing up the field with speed and intercepting passes from receivers sprinting upfield. Considering these plays involve collisions and contact at high speed, these defensive positions initiating contact are more prone to injury than the offensive players being tackled. An interesting future study would be to collect more data to calculate the force of players colliding to see if the magnitude of a collision would be significant in modeling player injuries. These defensive players' injury occurrence suggests that players initiating contact are more likely to be injured than those receiving the tackle.

Conversely, when viewing players with the lowest injury occurrence, the positions include kicker (K), punter (P), and quarterback (QB). These positions were expected to have the lowest injury occurrence because they are football's most well-protected positions. These players receive minimal contact due to league rules restricting how they are tackled. Also, they have numerous players, such as the left guard, whose job assignment is to protect them from contact. Therefore, the rate of injury for well-protected positions supports the idea that more physical contact required by each position plays a role in injury occurrence.

Experience and Age: Figure 8 shows the impact of age and experience on a player's injury. The y-axis lists the years of experience, whereas the x-axis contains the age of players. Figure 7 uses the average of the injury label (1 = injured, 0 = not injured) to visualize injury occurrence depending on age and experience. The combination of age and experience with the highest probability of injury occurs at age 33 with 13 years of experience. Based on the dataset, 66.7% of players that fit this criterion were injured between 2016 and 2019. The general trend appears to be that as a player gets older and gains more experience, their probability of injury increases.

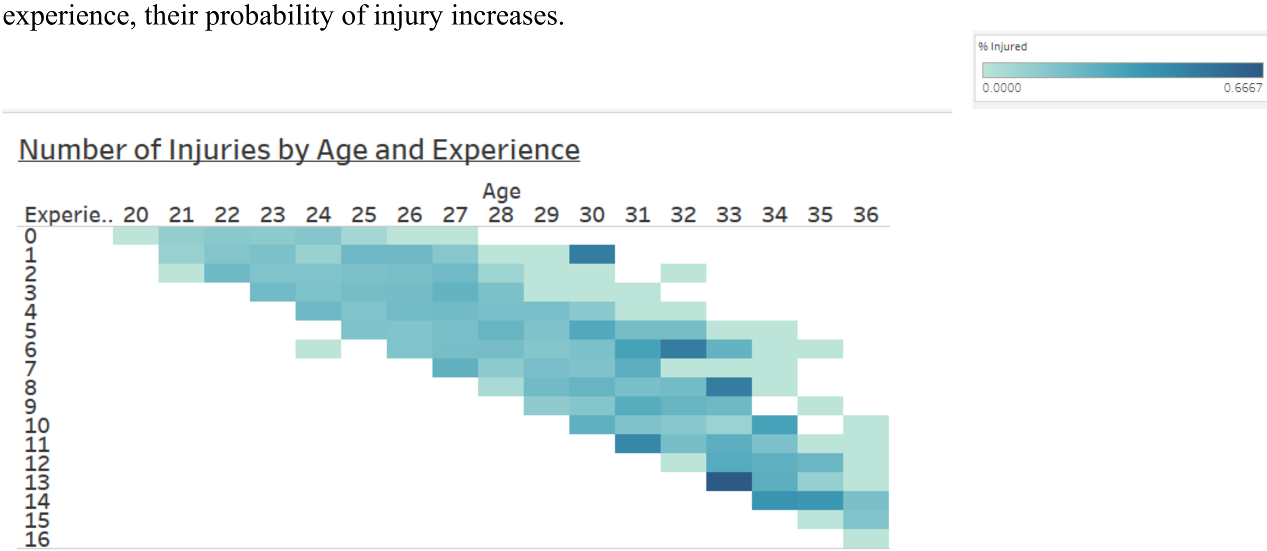

Figure 8 injury occurrence based on age and experience of players. Darker color represents a higher probability of injury

Conclusion, Limitation and Recommendations for Future Study

This study aimed to understand the factors leading to a player's injury using various machine learning models. The model with the best performance was the Gradient Boosted Trees model, with an F1 score of 0.508. In addition, the Gradient Boosted Trees model selected teams, games played, and starting position to be the most influential factors when determining the probability of a player getting injured during the season. The most notable takeaway from this study is that the left guard position is the most injury-prone, followed by defensive secondary player positions such as safety and cornerback. Future NFL rule changes and adaptions should focus on these players in the positions highlighted in the study to improve their safety.

One limitation of this research is that the model only considers a limited number of features due to the limitation of the data and time constraint. The model can be improved by incorporating more features. It is possible that previous injuries of a player and a particular match and range may have an impact on the injury of a player. In addition, further study should assess the team's performance metrics, such as win-toloss ratio and salary cap. Future studies should account for the temperature and weather conditions of the stadium where each game injury occurred. In addition, individual player statistics should also be considered as possible injury-causing variables.

References

Alizadeh, H. V., Fanton, M. G., Domel, A. G., Grant, G., & Camarillo, D. B. (2019, October). Prevention of Traumatic Brain Injury with Liquid Shock Absoprtion. *Journal of Biomechanical Engineering*, 1-9. Retrieved March 7, 2021

Arthur Lazarus, M. M. (2011, Jan/Feb). NFL Concussions and Common Sense. *Physician Executive, 37*(1), 6-7. Retrieved March 7, 2021

Arulanantham, A., Anzar, A. A., Rajalingam, G., Ingran, K., & Haddela, P. S. (2019, September 25). Effective Mechanism to Mitigate Injuries During NFL Plaus. *Conference on Industrial and Information Systems*. Retrieved March 8, 2021

Baker, H. P., Lee, K., Dayton, S., Terry, M., & Tjong, V. (2019, March 8). Thursday Night Football's impact on all-cause injuries in NFL players during 2012-2017. *The Physician and Sportsmedicine*, 350-352. Retrieved March 7, 2021

Cai, Y., Wu, S., Zhao, W., Li, Z., Wu, Z., & Ji, S. (2018). Concussion classification via deep learning using whole-brain white matter fiber strains. *PLOS ONE, 13*(5). doi:10.1371/journal.pone.0197992

Cohen, G. I., Lynch, H. F., & Deubert, C. R. (2016, Nov/Dec). A Proposal to Address NFL Club Doctors' Conflicts of Interest and to Promote Player Trust. *Hastings Center Report, 46*, 2-24. Retrieved March 7, 2021

Dreyer, N., Mack, C., & Anderon, R. (2019, September 1). Lessons on Data Collection and Curation from the NFL Injury Surveillance Program. *Sports Health: A Multidisciplinary Approach, 11*(5), 440-445. Retrieved March 8, 2021

Krill, M. K., Borchers, J. R., Hoffman, J. T., & Hewett, T. E. (2017). Effect of position, time in the season, and playing surface on Achilles tendon ruptures in NFL games: a 2009-10 to 2016-2017 review. *The Physician and Sportsmedicine, 45*(3), 259-264. Retrieved March 7, 2021

Mack, C. D., Kent, R. W., & Coughlin, M. J. (2020, June 2). Incidence of Lower Extremity Injury in the National Football League: 2015 to 2018. *The American Journal of Sports Medicine*, 1. Retrieved March 7, 2020

Mall, N. A., Buchowski, J., & Zebala, L. (2012, May 30). Spine and Axial Skeleton Injuries in the National Football League. *The American Journal of Sports Medicine*, 1. Retrieved March 7, 2021

Ruestow, P., Duke, T. J., Finley, B. L., & Pierce, J. (2015, Decemeber). Effects of the NFL's Amendments to the Free Kick Rule on Injuries during the 2010 and 2011 Seasons. *Journal of Occupational and Environmental Hygiene*, 875-882. Retrieved March 8, 2021

Siegel, B. (2019). Concussions and Capital: Tom Brady, CTE, and the NFL's Crisis of Identity. *Journal of Sport and Social Issues, 43*(6), 551-574. Retrieved March 7, 2021

Villwock, M. R., Meyer, E. G., Powell, J. W., & Fouty, A. (2009, March). Football playing surface and shoe design affect rotational traction. *Am J Sports Medicine*, 518-525. Retrieved March 7, 2021

Wang, D., Weiss, L., Abrams, M., Barnes, R., Warren, R., Rodeo, S., & Taylor, S. (2018, December). Athletes With Musculoskeletal Injuries Identified at the NFL Scouting Combine and Prediction of Outcomes in the NFL: A Systematic Review. *Orhopedic Journal of Sports Medicine*. Retrieved March 7, 2021

Woods, A., Chen, S. S.-C., Chen, C.-y., & Street, E. (2018). A Case of Racial Tasking and Prejudice: Analyses of Race, Playing Position, Rate of Injuries, and Salaries in the NFL. *KAHPERD, 55*(2), 28. Retrieved March 7, 2021

Zvijac, J., & Toriscelli, T. (2013, May 28). Isokinetic Concentric Quadriceps and Hamstring Strength Variables From the NFL Scouting Combine Are Not Predictive of Hamstring Injury in First-Year Professional Football Players. *The American Journal of Sports Medicine*, 1. Retrieved March 7, 2021