Project Proposal

Exploring Machine Learning to Classify Injury Risk in Collegiate Athletes

Jordyn Maurer

7 September 2025

Seminar in Data Analytics

Background

In recent years, data analytics has become an increasingly powerful tool in sports. With the rise of wearable technologies and performance tracking systems, coaches and sports scientists now have access to more information than ever. A growing area of interest within this field is injury prediction, as preventing injuries can impact both athlete health and team success. Recent studies have explored the use of machine learning techniques to predict injuries across various sports. For example, Tsilimigkras et al. (2023) explored how machine learning models enhance injury risk assessment in soccer, while Lövdal et al. (2021) applied similar techniques to predict injuries in competitive runners. While these studies demonstrate the potential of predictive analytics in sports, the methods are still relatively new, and there is a need for further research on different datasets and contexts. At Denison, the force plate has become a central tool in sports performance analysis. A force plate is a device that measures ground reaction forces to track athlete performance (Hawkin Dynamics). Our athletes perform a jump on the plate, and the system returns over seventy metrics related to power, quickness, and fatigue. The data it produces offers deep insights into how an athlete's body responds to training, while also presenting an opportunity to extend current injury prediction research by incorporating a device that has not been widely explored in this context.

The Director of Sports Performance at Denison, Coach Beau Scott, employs student analysts each year to wrangle and interpret the force plate data. As a student analyst myself, I have spent the past year examining force plate metrics across multiple sports, primarily focusing on athlete injuries. We focus specifically on soft tissue injuries, including those to muscles, tendons, and ligaments, as these injuries are often preventable through proper training and proactive interventions. My initial projects have been descriptive, flagging significant changes in force plate metrics leading up to an injury date. These findings have helped inform coaches and trainers by providing a clearer understanding of injury patterns, which can be used to adjust training programs and implement prevention strategies to reduce future injuries. While this descriptive work has been valuable, there is an opportunity to move beyond pattern recognition into predictive analysis. Specifically, by applying classification-based machine learning models, I am

interested in exploring whether **force plate data can be leveraged to proactively identify athletes at higher risk of injury.** If successful, this approach could provide coaches and trainers with an early-warning system to guide interventions, reduce injury rates, and ultimately enhance both athlete safety and team performance.

Ethical Considerations

Because this project uses real athlete performance and injury data, there are important ethical considerations to address to ensure the privacy and well-being of the athletes involved, as well as the responsible interpretation of results. To protect the privacy of the athletes, all data is void of personally identifiable information such as names, student IDs, or jersey numbers. Any reports or visualizations will focus on aggregated trends, rather than individual athletes. At this stage, we want to avoid labeling individual athletes as "injury-prone" without sufficient evidence. Additionally, we will submit the project for IRB approval to ensure the ethical treatment of the athletes involved in the research. Ultimately, the health and safety of the athletes is our top priority, and this research is designed to support and protect it.

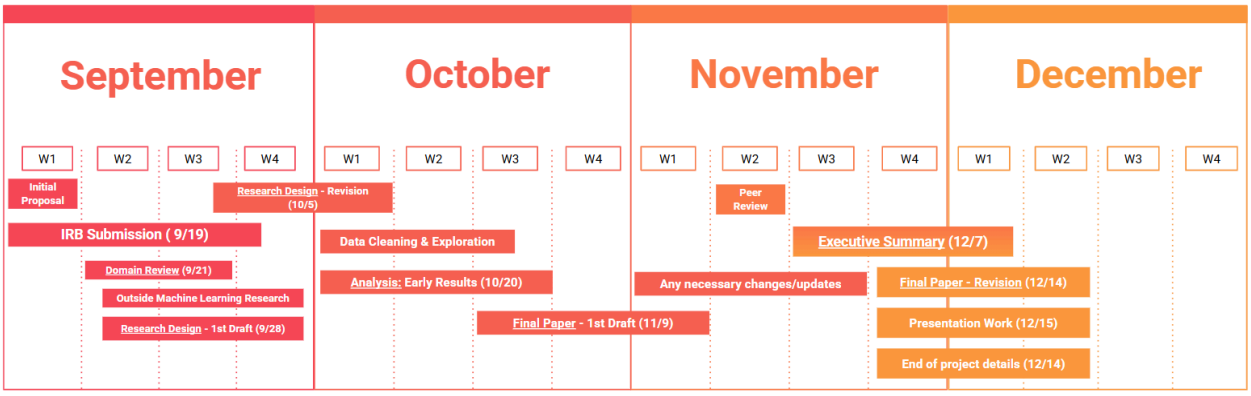
Preparation

My background for this project comes from a combination of coursework, hands-on sports analytics experience, and personal connection to the world of athletics. I have gained direct experience collecting, cleaning, and interpreting force plate data in the past year. Because of my previous work, I currently have access to a year's worth of injury and force plate data which we will use for the project. This role has strengthened my skills in exploratory analysis as well as familiarized me with the variables of interest. Through this work, I have become proficient in R and its associated libraries, which I will use for data cleaning, model building, and visualization throughout the project. Academically, I have built a strong foundation in predictive analytics through my coursework. While I am unsure of the specific method I will be implementing, my Advanced Predictive Methods in Data Analytics course introduced me to a variety of machine learning techniques, including decision trees and random forests. These skills

will directly support my ability to build and evaluate classification models to predict injury risk. However, this project will challenge me to apply these methods outside of the classroom, requiring further research and adaptation in an independent environment. Over the semester, I plan to expand my knowledge and review existing literature on prediction modeling, specifically with imbalanced datasets common in sports injury data. Finally, as a former athlete who has experienced a chronic injury, I bring a personal understanding of athletics that informs my perspective on this research. Thus, I have a stronger ability to interpret the data in context, recognizing and emphasizing the athlete behind each data point.

Timeline

To ensure the completion of the project, I have included a detailed timeline of the semester divided into milestones. However, I understand that working with real-world data can lead to real-world challenges. To address this, I will stay flexible and plan to meet bi-weekly with Coach Scott to keep the project's goals aligned. In the Gantt chart, milestones are underlined with the corresponding due date, and I also included any ad hoc tasks that may come up along the way.



Conclusion

This project aims to use force plate data and machine learning to predict injury risk among varsity athletes. By providing coaches with actionable insights, it has the potential to improve training decisions, reduce injuries, and enhance athlete safety.

References

Hawkin Dynamics. (n.d.). *Hawkin force plate metric database*. Retrieved September 7, 2025, from https://www.hawkindynamics.com/hawkin-metric-database.

Lövdal, S., den Hartigh, R., & Azzopardi, G. (2021). Injury Prediction in Competitive Runners with Machine Learning. International journal of sports physiology and performance, 16(10), 1522–1531.

Tsilimigkras, T., Kakkos, I., Matsopoulos, G. K., & Bogdanis, G. C. (2024). Enhancing sports injury risk assessment in soccer through machine learning and training load analysis. *Journal of Sports Science and Medicine, 23*(5), 537–547.