

# Project Proposal Template

**Project Title: Leveraging Machine Learning for NFL Injury Prediction and Draft Success Evaluation**

**Team 5 Members:**

* Kelly He
* Philip Rex Peters
* Samuel Williamson
* Saswata Chakraborty
* Tinkle Jain

**Date:** 01/16/2025

## 1. Project Overview

### Objective

The objective of this project is to leverage machine learning techniques to:

1. **Predict NFL Player Injuries**:
   * Create model from two datasets to identify patterns and factors contributing to player injuries
   * Develop predictive models to assess the likelihood of future injuries with the goal of optimizing player management strategies
2. **Predict Draft Success**:
   * Leverage NFL Combine performance metrics in order to evaluate and predict if a player will be drafted based on their combine performance
   * Assist teams in making more informed decisions during the drafting process by highlighting which factor are most relevant in making these predictions

This dual-focus project is intended to illuminate trends related to both draft success and injury prevention, in order to improve team performance and player safety. Understanding both combine metrics and draft performance should lead to clear conclusions surrounding career longevity.

### Data Sources

1. <https://www.kaggle.com/competitions/nfl-playing-surface-analytics/data>
   1. **Predict NFL Player Injuries**
2. <https://www.kaggle.com/datasets/redlineracer/nfl-combine-performance-data-2009-2019/data>
   1. **Predict Draft Success**

The dataset comprises two parts: Injury Records and Playlist.

\* The Injury Record file (CSV format) contains information on 105 lower-limb injuries sustained during regular season games over two seasons. Each injury can be linked to specific player history records using the PlayerKey, GameID, and PlayKey fields.

\* The Playlist file details the 267,005 player-plays within the dataset. Each play is indexed by PlayerKey, GameID, and PlayKey fields. Game and play details include the player's assigned roster position, stadium type, field type, weather, play type, position for the play, and position group.

\*\*Analysis of Playlist\*\*

1. There are 250 players in the dataset.

2. There are 5,712 games in the dataset.

3. There are 267,005 plays in the dataset.

\*\*Analysis of Injury Record\*\*

1. There are 105 injury records in total.

2. 100 unique players have been injured, indicating instances of multiple injuries to the same player.

3. 28 PlayKey values are missing.

## 2. Project Tools & Technologies

* Scikit-learn, Pandas for data cleaning, Matplotlib

## 3. Project Timeline

### Milestones:

* + Data Collection and Cleaning: 01/21/2025
  + Design and Development:01/25/2025
  + Testing and Feedback: 01/27/2025
  + Finalization and Deployment: 01/29/2025

## 