# Predicting UK Racehorses' Risk of Injury Based on Historical Race and Biometric Data

Sabrina del Rosal

**Topic**: Predicting UK Racehorses' Risk of Injury and Performance Decline

- **Problem**: Racehorses undergo intense physical strain during races, leading to frequent injuries and performance decline.
- **Opportunity**: optimize training schedules, reduce injury risks, and improve performance
- **Affected Stakeholders**:
    - **Horse Trainers**: Better training management.
    - **Jockeys**: Understand horse performance on race day.
    - **Owners**: Reduce costs and increase success rates.

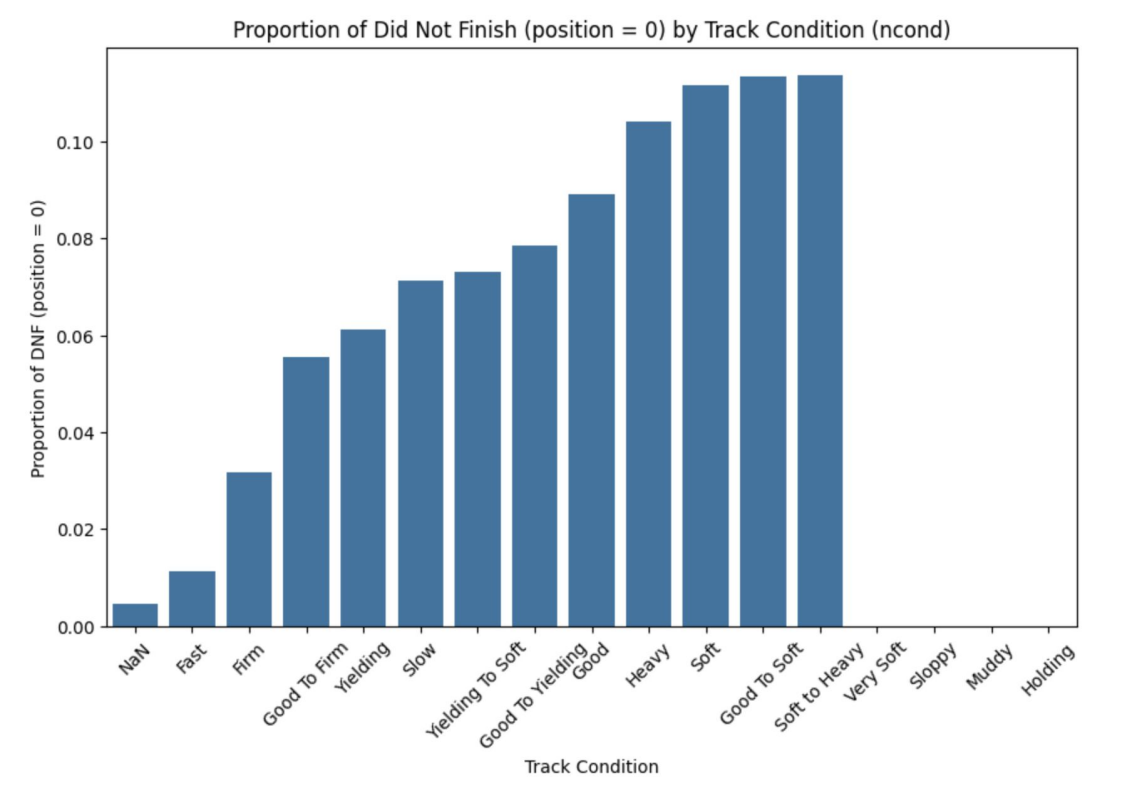
**Introduction to the Dataset, Data Quality Concerns, and Preliminary EDA**

**Dataset Overview**:

1. Age
2. Start Position (saddle)
3. Official Rating
4. Weight
5. Hurdles
6. Distance
7. Track Condition
8. Class
9. Fences
10. Days Rested
11. Races Ran
12. Day of the Year

**Injury Risk**: Yes or No

## **Preliminary Findings from EDA**:



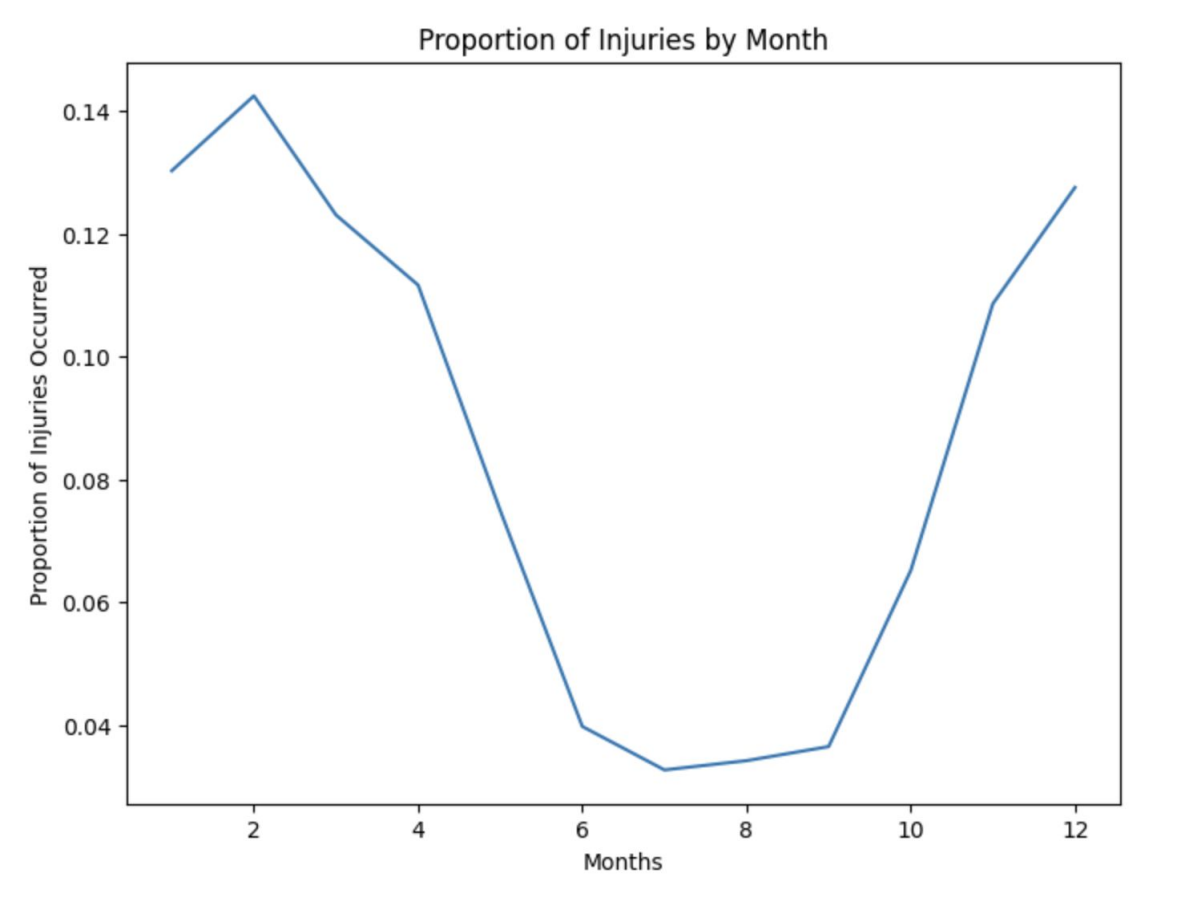

11% DNF for conditions :

12 soft to heavy

10 good to soft

5 soft condition

“Horses with sensitive **tendons**, will feel more at ease on **hard surfaces** because they facilitate their locomotion. Other horses, such as those with **joint** problems, will move more smoothly on **soft surfaces**.”




**Winter Months:**

1. Muscle Stiffness
2. Reduced Flexibility
3. Slippery
4. Longer Warm-Ups
5. Reduced Training

**Baseline Models & Evaluations:**

- **Objective**: Develop a model to classify injury risk by analyzing performance trends and biometric data.

1. XGBoost

2. SMOTE (balancing classes)

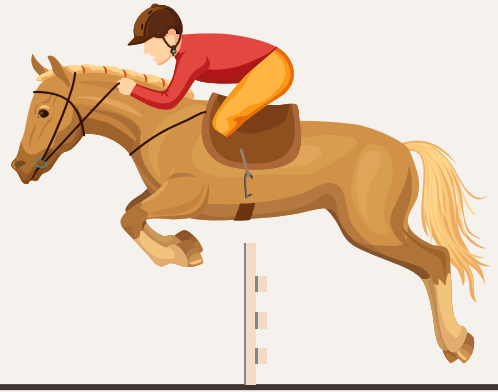
3. Gridsearch (tune parameters)

4. Lowering Threshold (to improve false positives)

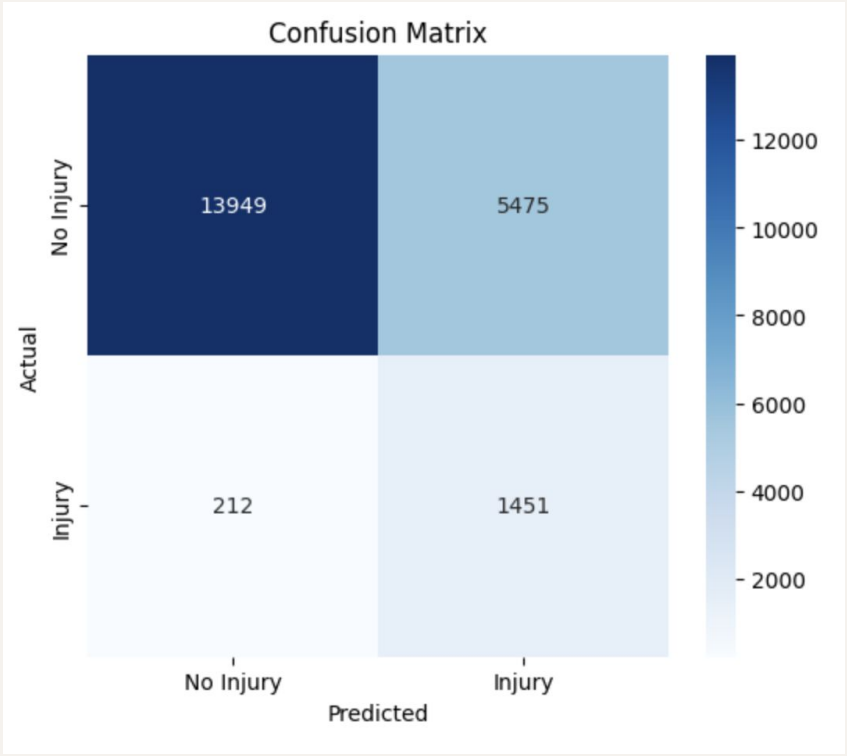
## Accuracy:

Testing Accuracy: **92%**

Training Accuracy: **94%**



# XGBoost with Gridsearch and SMOTE



**Next Steps:**

- **Ensemble Learning :** working with multiple models to get a better accuracy overall

**Main Takeaway:**

Our model does the best in maximizing the prevention of horses racing with the potential that they may get injured. Although there is a bit of overfitting, we can work with other models in the future and incorporate them to have an overall better understanding to minimize economic loss while still ensuring that the horses are able to race and retire healthily.

**XGBoost with Gridsearch & SMOTE**

# Thank You!