Modelling Coach Decisions in Professional Cycling Teams

[1]Maor Sagi, [2]Paulo Saldanha, [1]Prof. Guy Shani , [1]Prof. Robert Moskovitch

[1]Software and Information Systems Engineering – Ben-Gurion University, [2]Israel – Premier tech

ABSTRACT

Cycling racing is a popular field, attracting significant attention in recent years. Assigning a collection of teams' cyclists to specific races may determine whether a team will win or not. While our long term goal is to model the decision making of multiple teams and compare them in the light of their performance in the races, here we propose a model for recommendation of cyclists for a race stage, that consist of binary classifiers, called RaceFit. RaceFit represents a record as features of a race stage and a cyclist's demographical properties, as well as features of his recent weeks of workouts data performance. We evaluated RaceFit on a dataset of Israel Premier Tech's cyclist and race data and found the best performing parameters of the framework, having encouraging results. Additionally, we ranked the most predictive features, which we report here as well.

INTRODUCTION

18 UCI World Teams, **17 UCI Pro teams**, 18-34 cyclists in each team

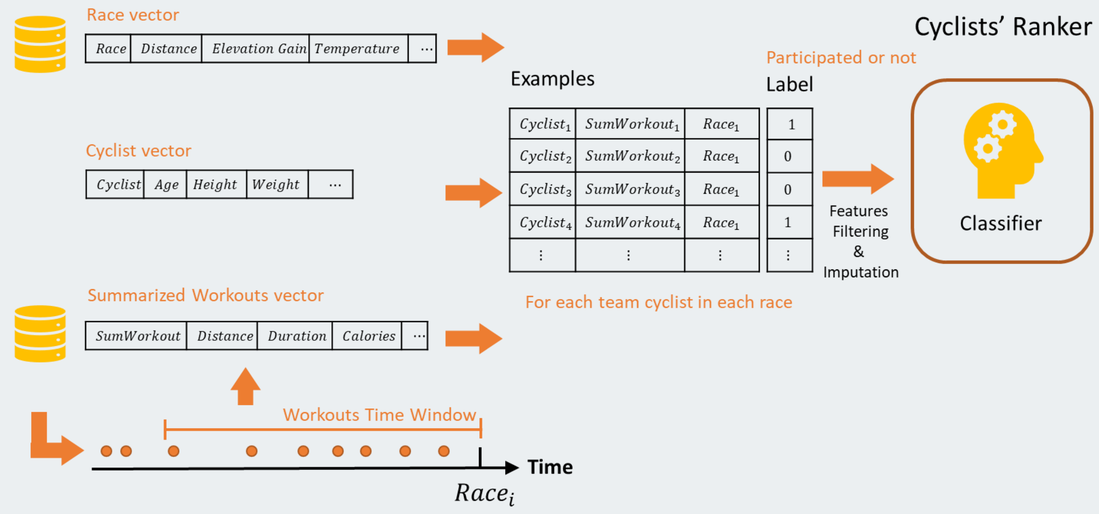
Around **100 races** per year, in which **8 cyclists** race from each team

- Grand Tours – stage races, 21 days, i.e., Tour de France
- Major Tours – stage races, multiple days, i.e., Paris-Nice
- One day races – i.e., Paris-Roubaix

DATA

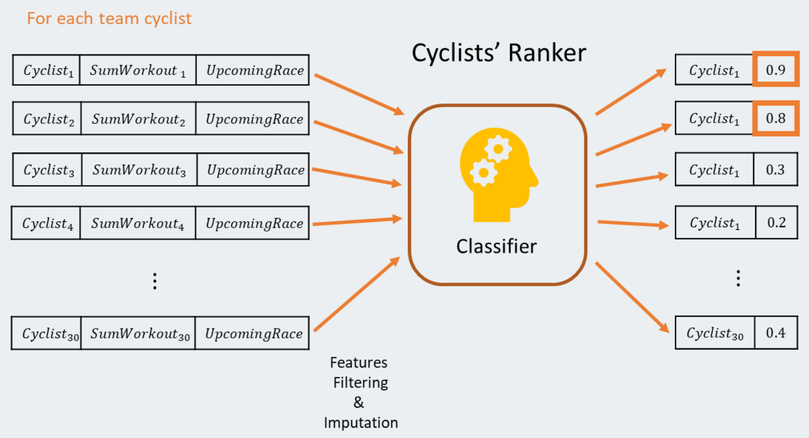
- **583 Cyclists from 10 teams**
  - Age, weight, height, cyclist's statistics and specialties (i.e., sprint, climbing)
- **1025 Races, 3100 Stages during 2017-2022**
  - Race classification, distance, elevation gain, results and rankings
- **65k Training Peaks Workouts of 37 cyclists during 2017-2022**
- **302k STRAVA Workouts of 364 cyclists during 2017-2022**
  - Workout's duration, distance, elevation gain, calories, power, speed, heart rate, temperature, cadence, etc.

RACE FIT



Training Phase

Given $Race$, $Cyclist$ and $Workout$ summarized vector of the last weeks workouts, we train a classifier while the labels are represented by participation of $Cyclist$ in $Race$.



Recommendation Phase

Given $Upcoming\ Race$, and the teams cyclists along with their last workouts. we are using the classifier to predict for each $Cyclist$ in the team, the matching score of $Cyclist$ and $Upcoming\ Race$ using the classifier probabilities. Then, top cyclists recommended.

EVALUATION & RESULTS

RaceFit Coach Decision Modelling Performance

- Evaluation Metrics:

$$Precision@i = \frac{|Cyclists\ Raced\ \cap\ top\ i\ Recommended|}{|top\ i\ Recommended|}$$

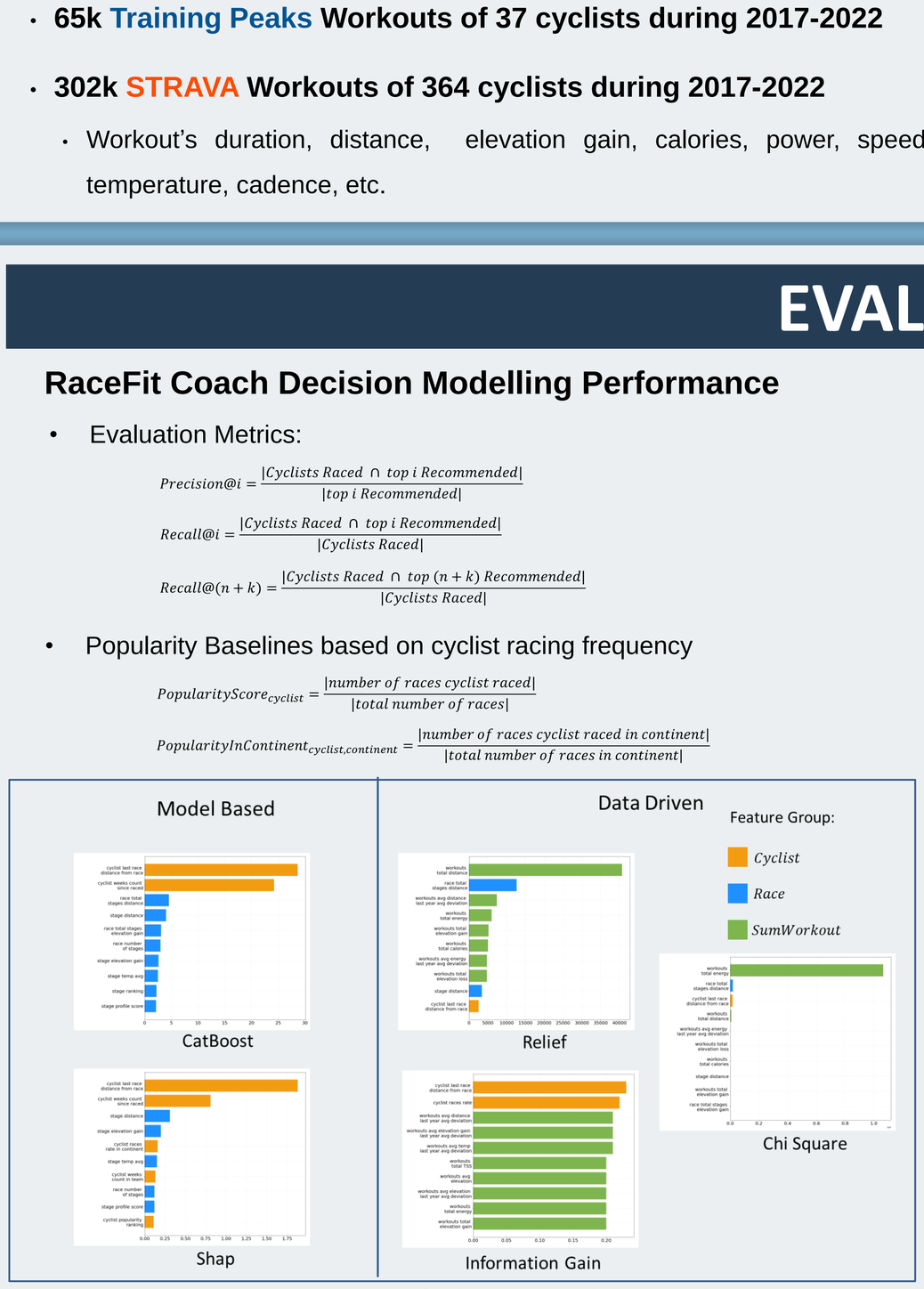
$$Recall@i = \frac{|Cyclists\ Raced\ \cap\ top\ i\ Recommended|}{|Cyclists\ Raced|}$$

$$Recall@(n+k) = \frac{|Cyclists\ Raced\ \cap\ top\ (n+k)\ Recommended|}{|Cyclists\ Raced|}$$

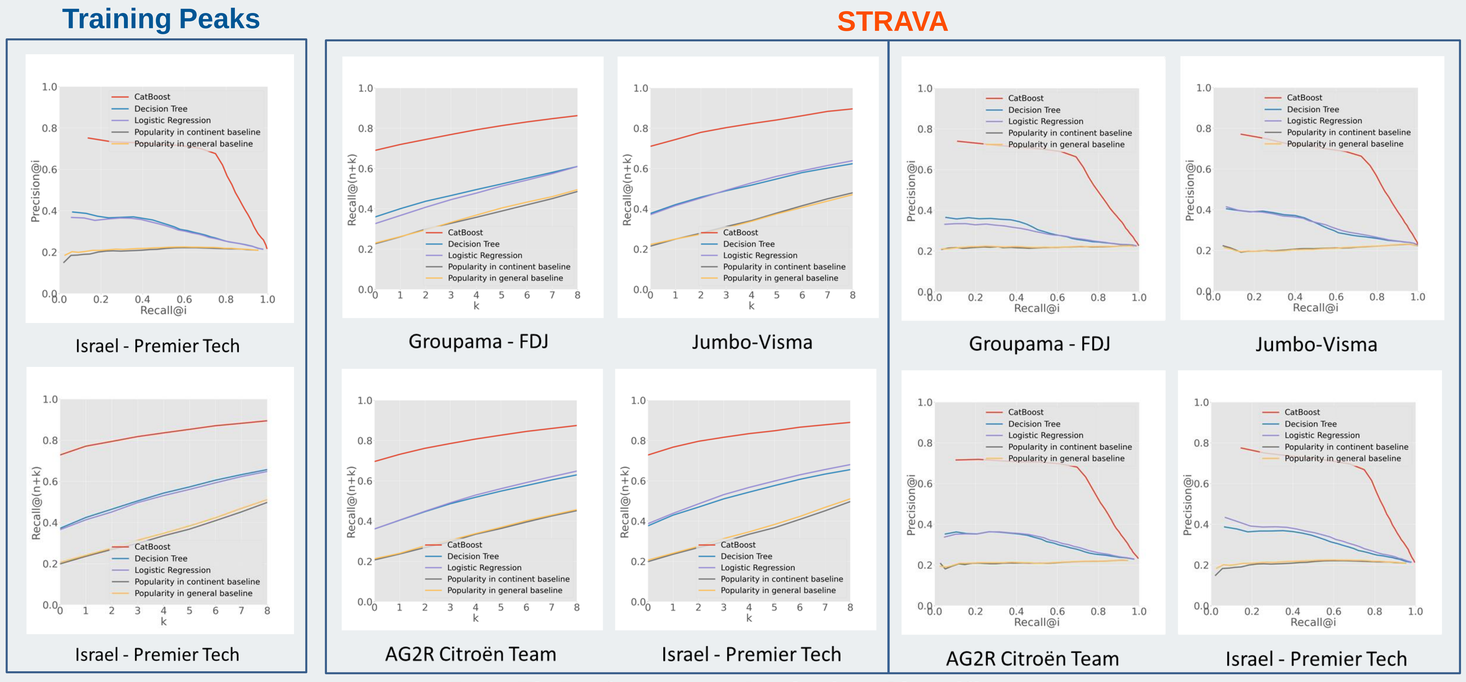
- Popularity Baselines based on cyclist racing frequency

$$PopularityScore_{cyclist} = \frac{|number\ of\ races\ cyclist\ raced|}{|total\ number\ of\ races|}$$

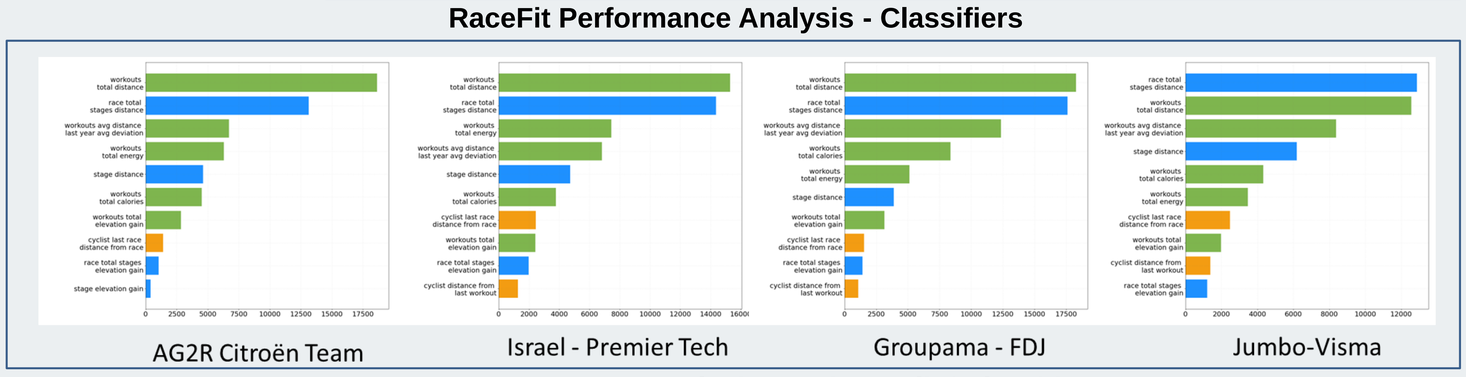
$$PopularityInContinent_{cyclist,continent} = \frac{|number\ of\ races\ cyclist\ raced\ in\ continent|}{|total\ number\ of\ races\ in\ continent|}$$



Feature Importance Analysis – IPT Training Peaks Data



RaceFit Performance Analysis - Classifiers



Feature Importance Analysis – Using Relief Algorithm