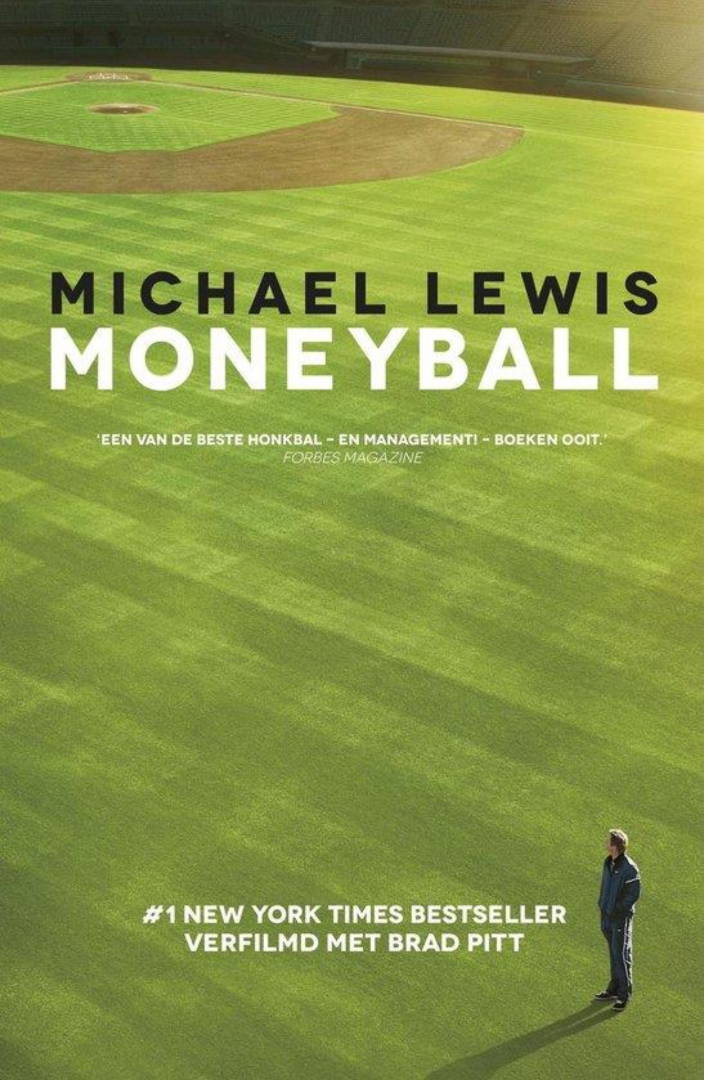
A Machine Learning approach for Race Cycling Performance Prediction

Leonid Kholkine, Thomas Servotte, Arie-Willem de Leeuw, Tom De Schepper, Tim Verdonck and Steven Latré

IDLAB, IMEC RESEARCH GROUP AT GHENT UNIVERSITY AND UNIVERSITY OF ANTWERP



Predicting Sports Outcomes

- Mostly in team sports
  - Football, NBL, NBA, NHL, ...
- A range of techniques
  - Probabilistic models
  - Monte Carlo
  - Machine learning
- Classification problem
- External conditions are similar or easy to model

Predicting Sports Outcomes

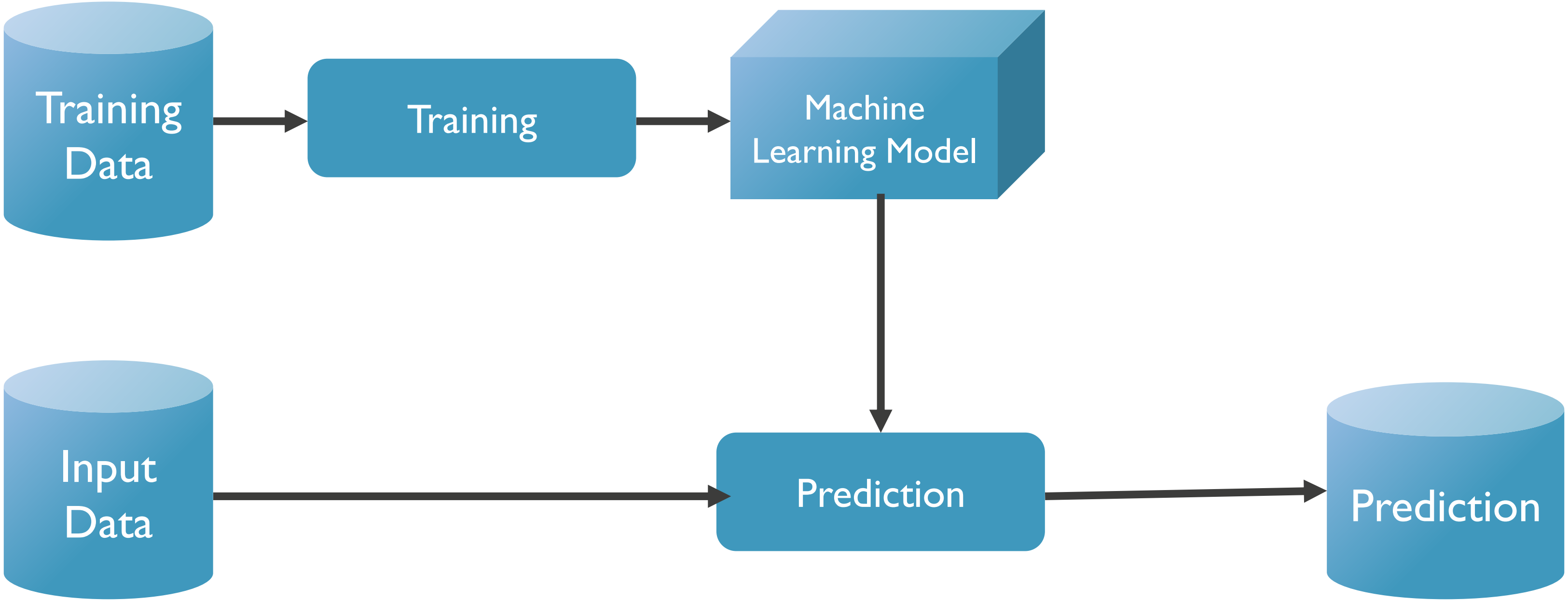
- Mostly in team sports
  - Football, NBL, NBA, NHL, ...
- A range of techniques
  - Probabilistic models
  - Monte Carlo
  - Machine learning
- Classification problem
- External conditions are similar or easy to model

Is it possible to predict top 10 riders of a race based on public data?

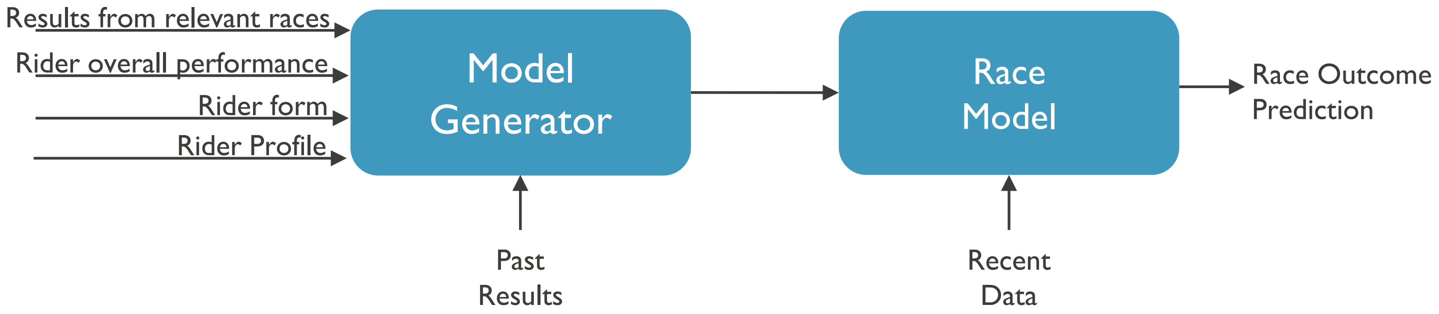
Motivation

- Fans, journalists and coaches try to estimate how a certain race might unfold;
- Predicting the outcomes of a race needs domain-expertise;
- Because of the many factors involved and the available historical data, it is an interesting challenge for machine learning.
- Goal: Demonstrate the potential of Machine Learning Techniques

Machine Learning Basics

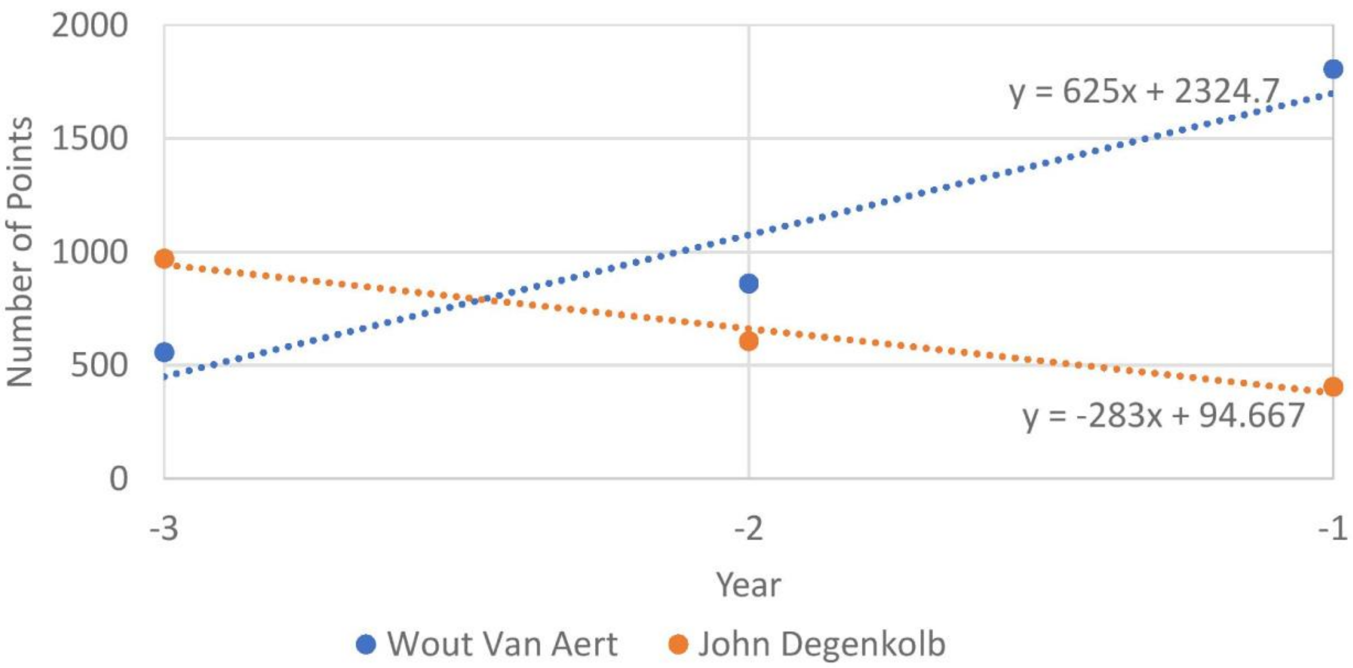


Proposed Framework



Features

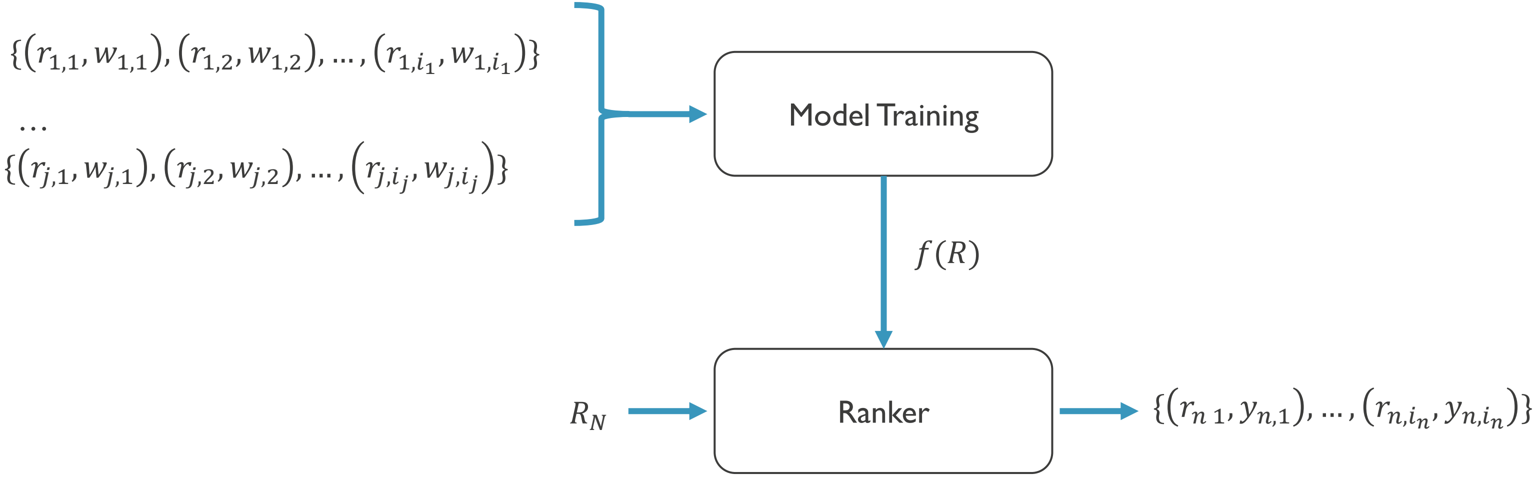
- Selected races
  - What are similar races? (e.g. classics)
  - Results from current and previous years
- Average points in:
  - Different types of GC stages
  - Different types of one-day races
- Form: Results in the past 6 weeks
- Evolution of the past years

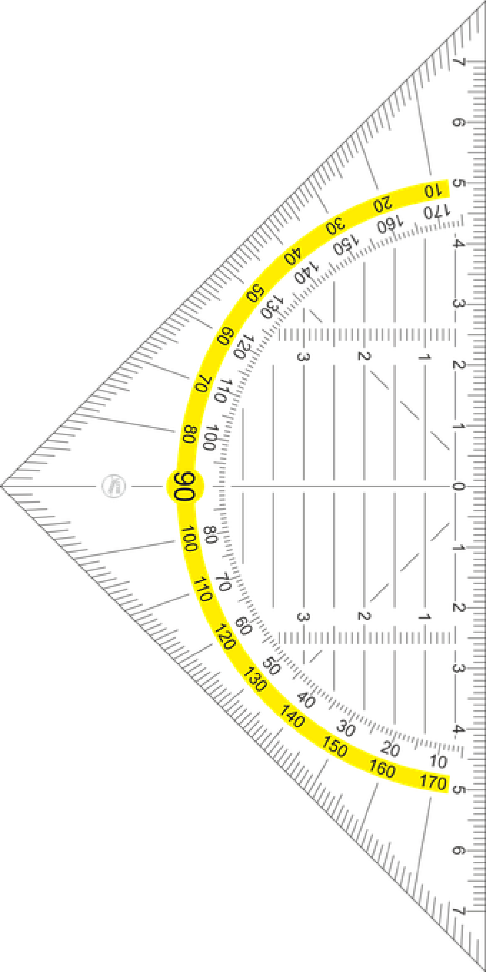

2000
1500
1000
500
0
Number of Points
y = 625x + 2324.7
y = -283x + 94.667
-3
-2
-1
Year
Wout Van Aert
John Degenkolb

Features

- Selected races
  - What are similar races? (e.g. classics)
  - Results from current and previous years
- Average points in:
  - Different types of GC stages
  - Different types of one-day races
- Form: Number of points gained in 6 weeks leading to the race
- Evolution of the past years
- Best result and time since best result
- Rider profile (age and career length)

Learn-to-Rank

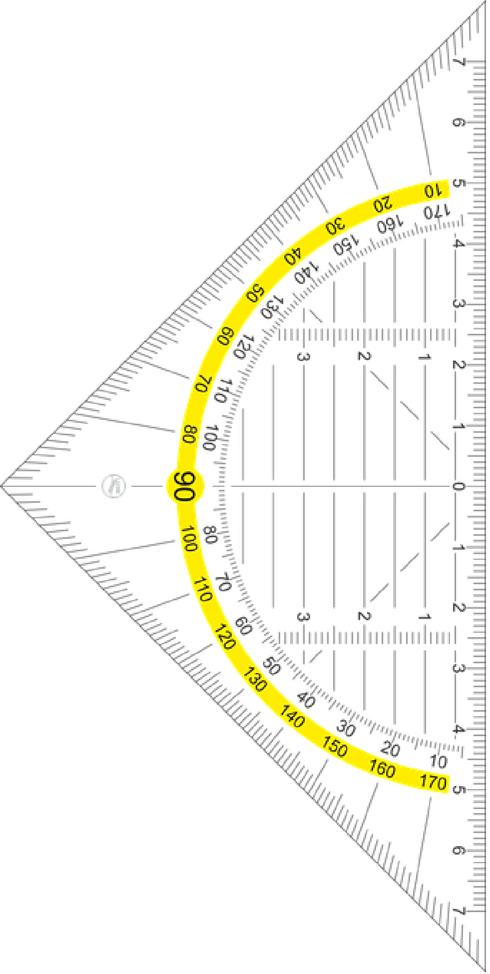




Metric: NDCG

$$DCG@k = \sum_{i=1}^{k} \frac{w_i}{\log_2(i+1)}$$

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

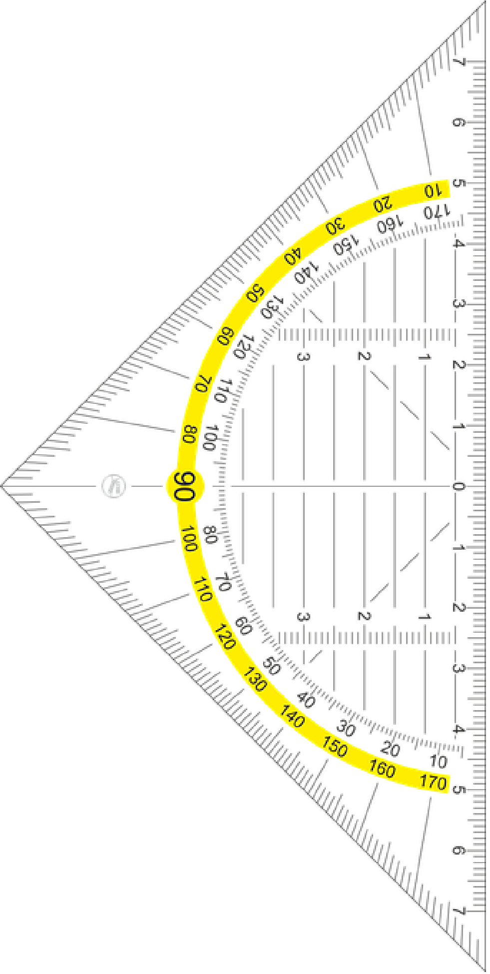


Metric: NDCG

$$DCG@k = \sum_{i=1}^{k} \frac{w_i}{\log_2(i+1)}$$

$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

RANK	WEIGHT
1	10
2	9
3	8
4	7
5	6
6	5
7	4
8	3
9	2
10	1



Metric: NDCG

$$DCG@k = \sum_{i=1}^{k} \frac{w_i}{\log_2(i+1)}$$

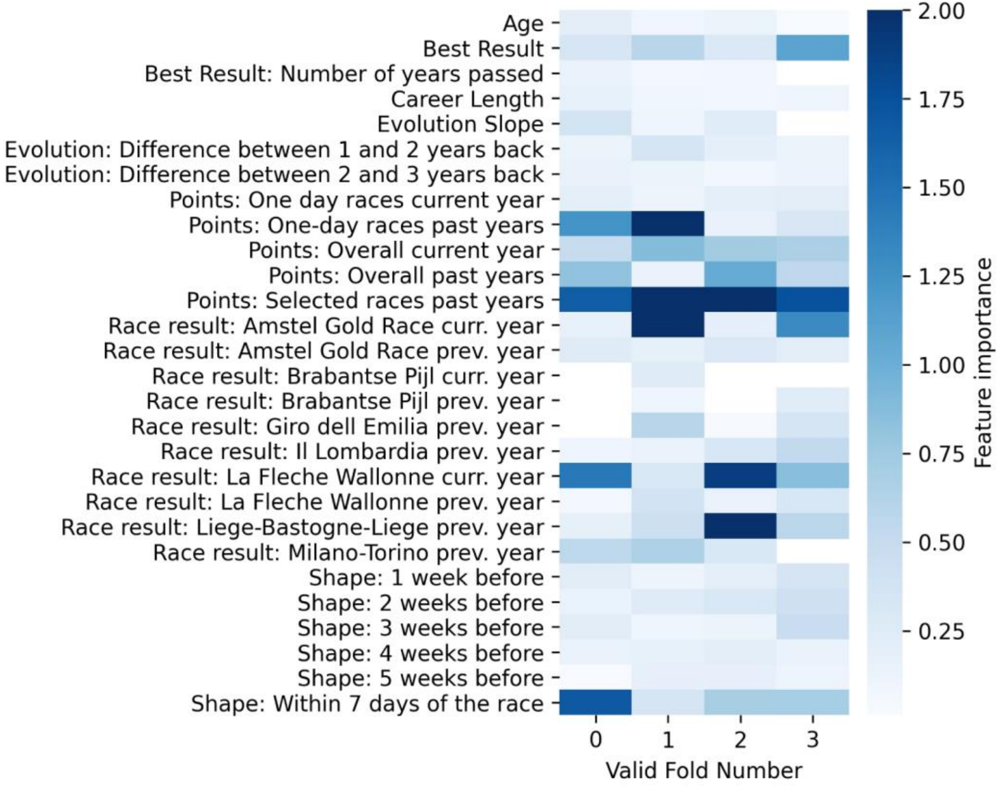
$$NDCG@k = \frac{DCG@k}{IDCG@k}$$

	NAME	REAL RANK	WEIGHT	DISCOUNTED GAIN
1	MATHIEU VAN DER POEL	2	9	9.0
2	GREG VAN AVERMAET	3	8	5.0
3	WOUT VAN AERT	6	5	2.5
4	OLIVER NAESSAN	33	0	0
5	JASPER STUYVEN	4	7	2.7
6	MATTEO TRENTIN	57	0	0
7	MICHAEL MATHEWS	21	0	0
8	DYLAN VAN BAARLE	10	1	0.3
9	FLORIAN SENECHAL	9	2	0.6
10	ANTHONY TURGIS	8	3	0.8

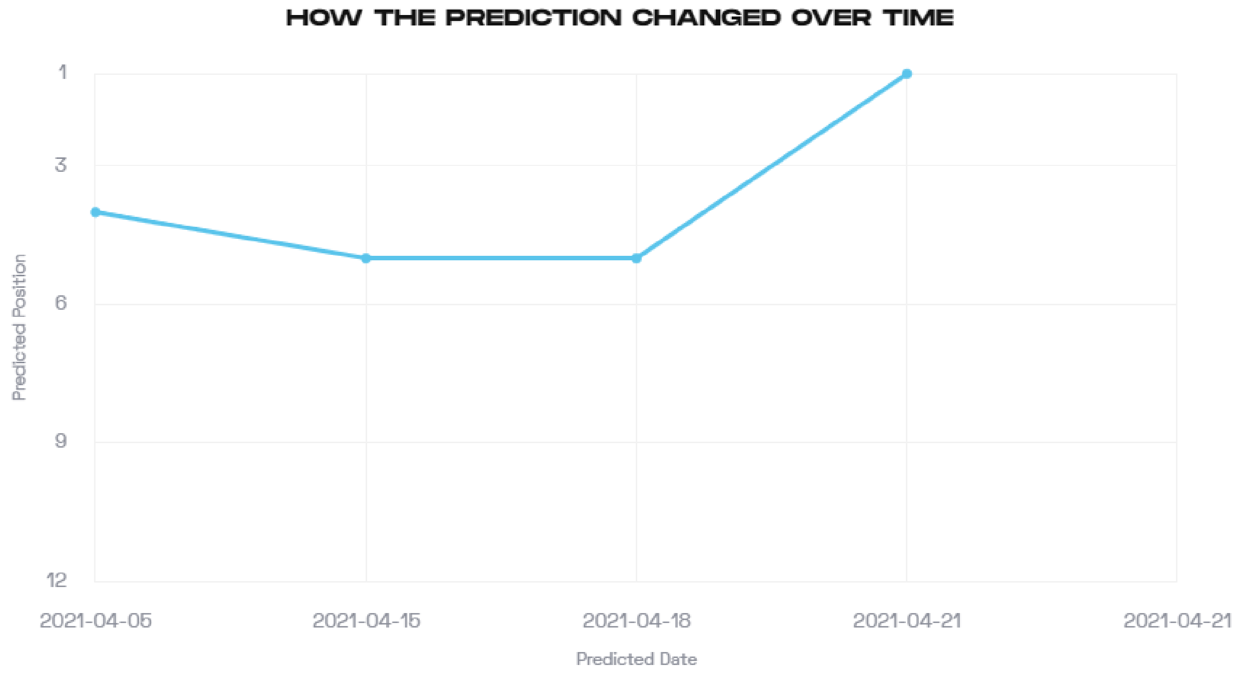
Results

Race	Year	Fan NDCG	Model NDCG	Model Correct	Difference between Model and Fans
E3 Saxo Bank Classic	2018	0.58	0.54	6	0
E3 Saxo Bank Classic	2019	0.50	0.54	5	0
Ghent-Wevelgem	2018	0.68	0.62	6	-2
Ghent-Wevelgem	2019	0.23	0.32	3	0
Tour of Flanders	2018	0.62	0.67	6	-1
Tour of Flanders	2019	0.27	0.21	4	-1
Paris-Roubaix	2018	0.77	0.74	6	0
Paris-Roubaix	2019	0.35	0.44	4	0
La Flèche Wallonne	2018	0.57	0.60	5	2
La Flèche Wallonne	2019	0.55	0.61	5	1
Liège-Bastogne-Liège	2018	0.28	0.38	5	1
Liège-Bastogne-Liège	2019	0.43	0.31	3	-1
E3 Saxo Bank Classic	2021	0.32	0.37	3	-1
Ghent-Wevelgem	2021	0.41	0.63	5	3
Tour of Flanders	2021	0.69	0.69	7	0
La Flèche Wallonne	2021	0.84	0.76	6	-1
Liège-Bastogne-Liège	2021	0.69	0.81	8	1
Average:		0.52	0.55	5.12	0.12

Liege – Bastogne – Liege Feature Importance



Prediction evolution: Liege – Bastogne – Liege
Julian Alaphilippe



imec IDLab UNIVERSITEIT GENT Universiteit Antwerpen

Conclusions

- Prediction NDCG similar to mass fan
- Applications:
  - Fans
  - Journalists

www.wiewintdekoers.be

www.whowillwintherace.com

Wie wint de Ronde? Artificiële intelligentie geeft het antwoord

01/04/21 om 07:51 Bijgewerkt om 13:46

Robben Scheire
Medewerker van Sport/Voetbalmagazine.

Onderzoekers van ID-lab, een Imec-onderzoeksgroep aan de Universiteit van Antwerpen, maken een website waarop ze via Artificial Intelligence (AI) de koers kunnen voorspellen.

Mathieu Van der Poel wint Ronde van Vlaanderen ... volgens onderzoekers UAntwerpen

ANTWERPEN De Nederlander Mathieu van der Poel wint zondag de Ronde van Vlaanderen. Dat voorspelt een nieuw ontwikkeld computersysteem aan de hand van artificiële intelligentie (AI). Op basis van historische prestatiedata tracht het systeem, dat door onderzoekers van imec en de Universiteit Antwerpen ontworpen is, de koersuitslagen van eendagswedstrijden te voorspellen. De voorspellingen zijn vanaf donderdag te raadplegen op wiewintdekoers.be.

imec IDLab UNIVERSITEIT GENT Universiteit Antwerpen

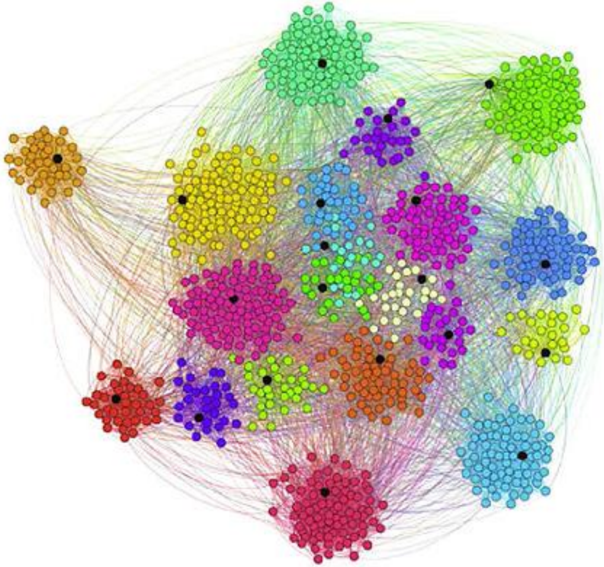
Conclusions

- Prediction NDCG similar to mass fan
- Applications:
  - Fans
  - Journalists
- Further development:
  - Identification of future talent
  - Additional insight to understand how a rider compares to others

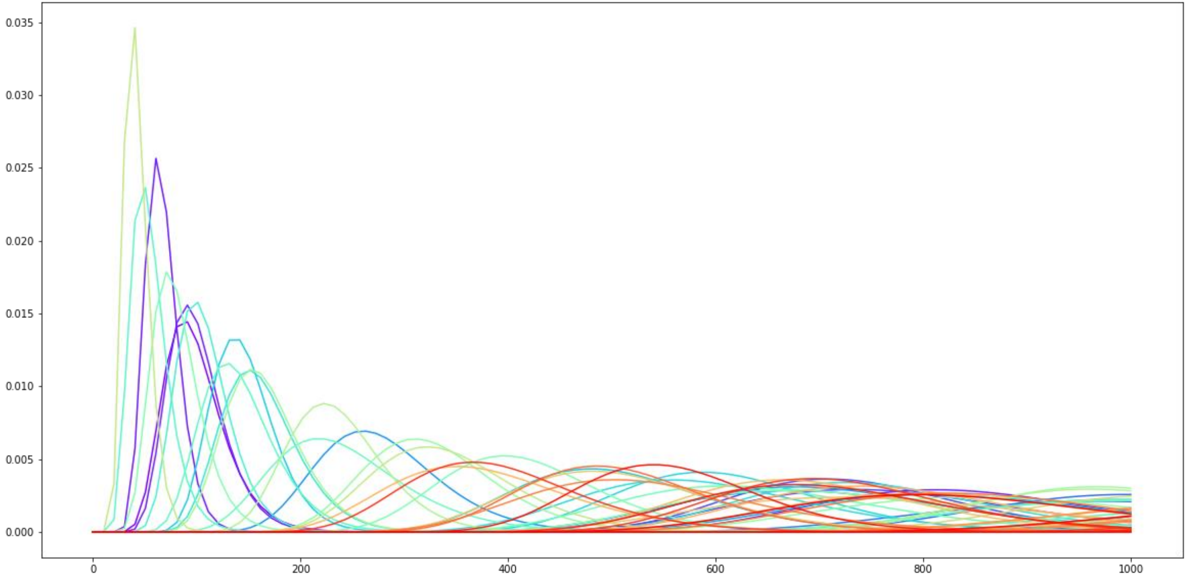
Conclusions

- Prediction NDCG similar to mass fan
- Applications:
  - Fans
  - Journalists
- Further development:
  - Identification of future talent
  - Additional insight to understand how a rider compares to others
- Demonstrates the power of Machine Learning applied to sports
  - Tool which can help make sense of all the data

Future Work



Clustering and automatic selection of related races



Conditional Distribution Estimation

ORIGINAL RESEARCH article

Front. Sports Act. Living | doi: 10.3389/fspor.2021.714107

A Learn-to-Rank Approach for Predicting Road Cycling Race Outcomes

Provisionally accepted
The final, formatted version of the article will be published soon

Notify me

Leonid Kholkine[1*], Thomas Servotte[2], Arie-Willem De Leeuw[1], Tom De Schepper[1], Peter Hellinckx[1], Tim Verdonck[2] and Steven Latré[1]

[1]Department of Computer Science, Faculty of Sciences, University of Antwerp, Belgium
[2]Department of Mathematics, Faculty of Sciences, University of Antwerp, Belgium

Professional road cycling is a very competitive sport, and many factors influence the outcome of the race. These factors can be internal (e.g.psychological preparedness, physiological profile of the rider and the preparedness or fitness of the rider) or external (e.g. the weather or the team's strategy) to the rider, or even completely unpredictable (e.g. crashes or mechanical failure). This variety makes perfectly predicting the outcome of a certain race an impossible task and the sport even more interesting. Nonetheless, before each race, journalists, ex-pro cyclists, websites and cycling fans try to predict the possible top 3, 5 or 10 riders.

In this article, we use easily accessible data on road cycling from the past 20 years and the Machine Learning technique Learn-to-Rank to predict the top 10 contenders for one-day road cycling races. We accomplish this by mapping a relevancy weight to the finishing place in the first 10 positions. We assess the performance of this approach on the 2018, 2019 and 2021 editions of 6 spring classic one-day races. In the end, we compare the output of the framework with a mass fan prediction on the Normalised Discounted Cumulative Gain (NDCG) metric and the number of correct top 10 guesses. We found that our model, on average, has a slightly higher performance on both metrics than the mass fan prediction. We also analyse which variables of our model have the

Frontier in Sports and Active Living

Special issue on Using Artificial Intelligence to Enhance Sport Performance

QUESTIONS?

leonid.kholkine@uantwerpen.be

Who will win the World Championship?

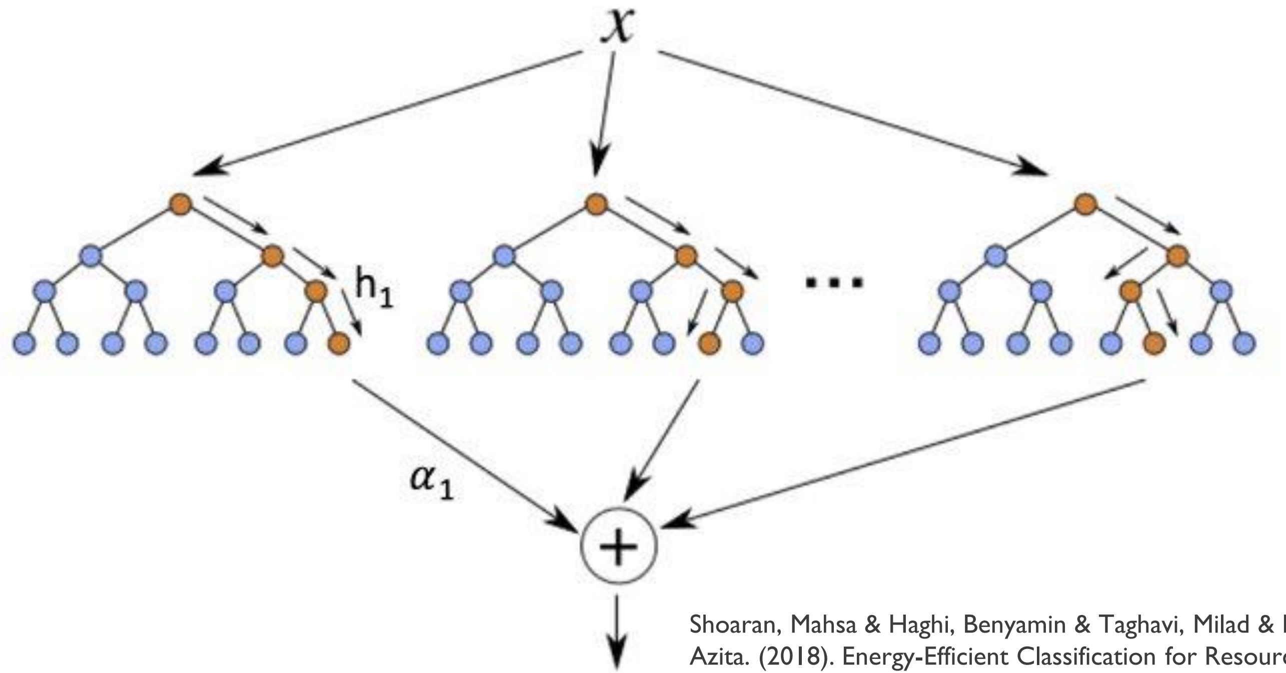
	NAME
1	WOUT VAN AERT
2	JULIAN ALAPHILIPPE
3	MATHIEU VAN DER POEL
4	TADEJ POGACAR
5	SONNY COLBRELLI
6	JASPER STUYVEN
7	MICHAEL MATTHEWS
8	PRIMOZ ROGLIC
9	MATEJ MOHORIC
10	ALEXANDRE KRISTOFF

Do you agree?

Vote for your top 5:

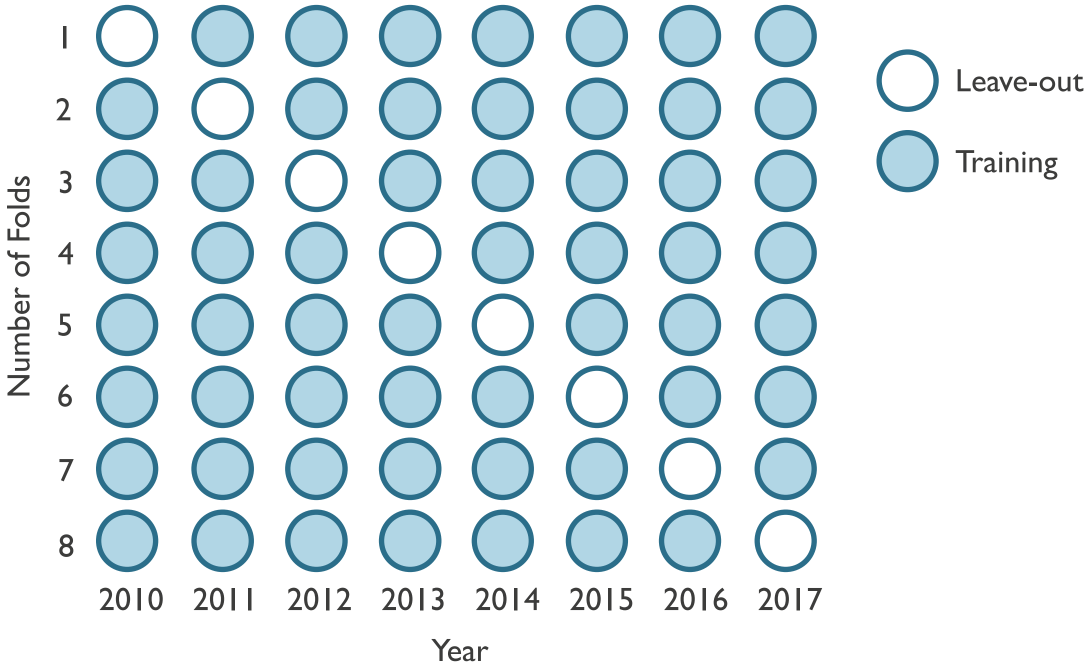
https://poll.whowillwintherace.com

Boosting Trees

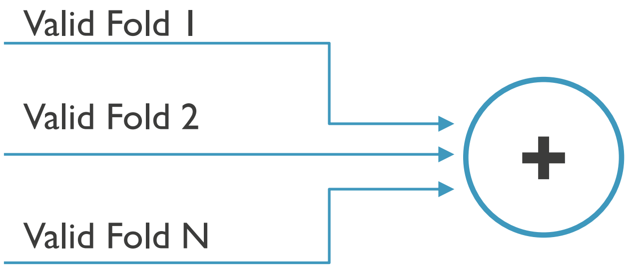


Shoaran, Mahsa & Haghi, Benyamin & Taghavi, Milad & Farivar, Masoud & Emami, Azita. (2018). Energy-Efficient Classification for Resource-Constrained Biomedical Applications. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. PP. 1-1. 10.1109/JETCAS.2018.2844733.

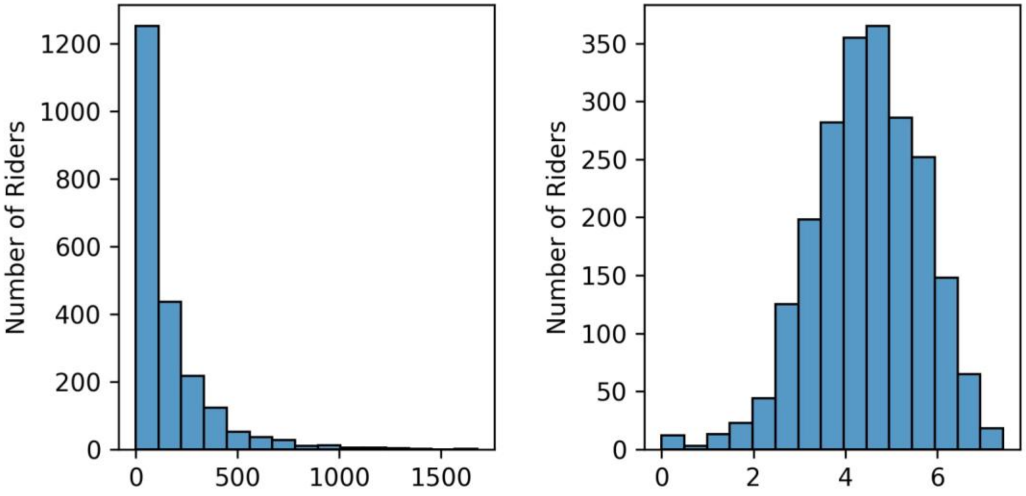
Yearly Cross Validation



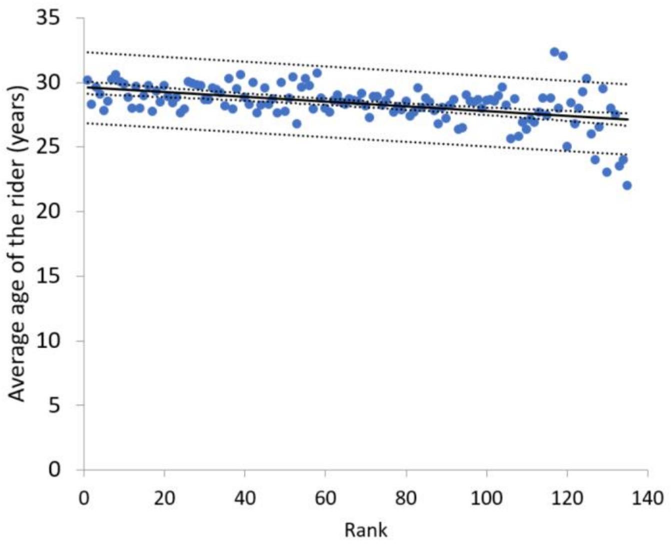
Fold Ensemble

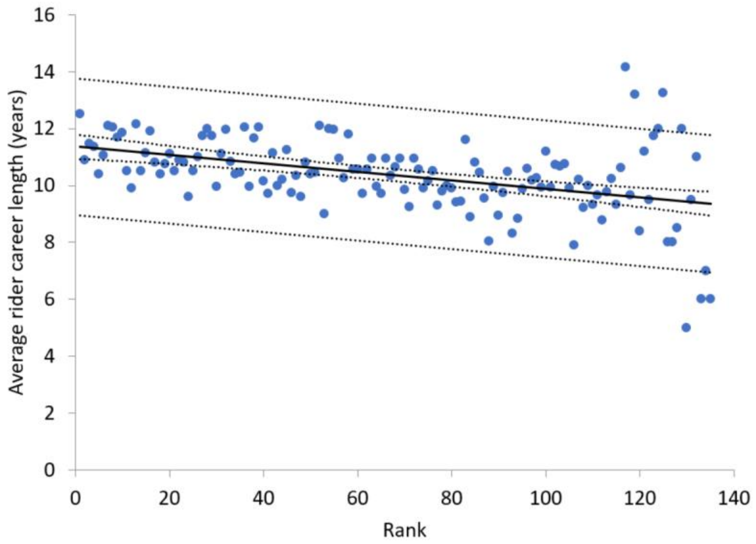


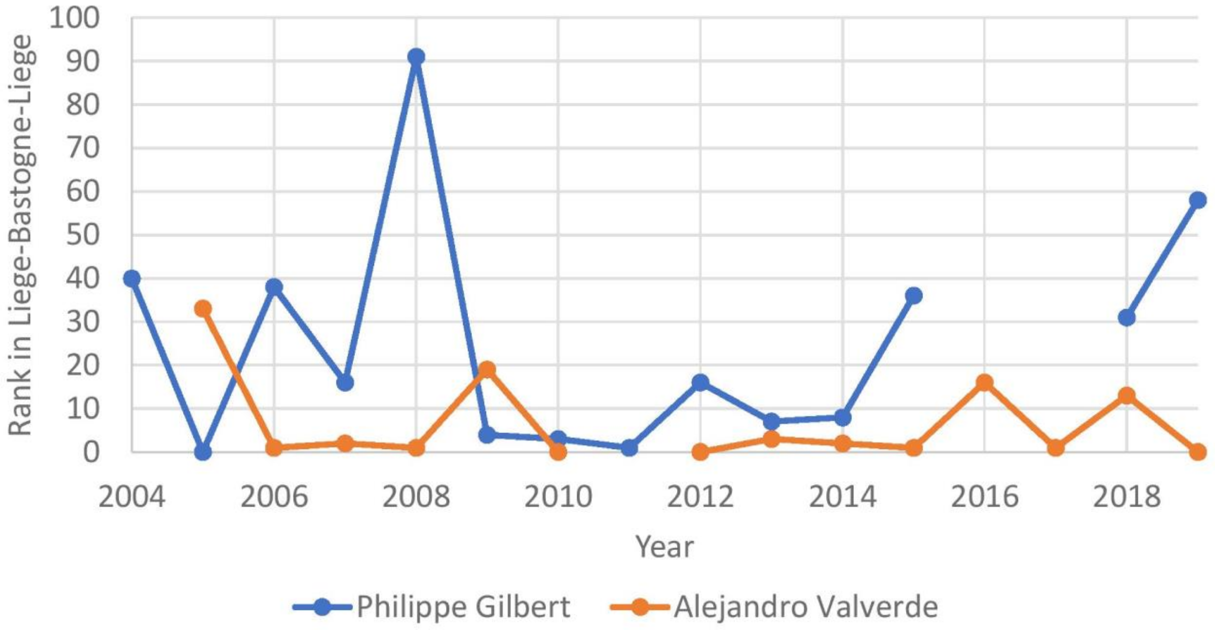
- Valid fold: Minimum number of iterations
- Valid ensemble: Minimum number of valid folds



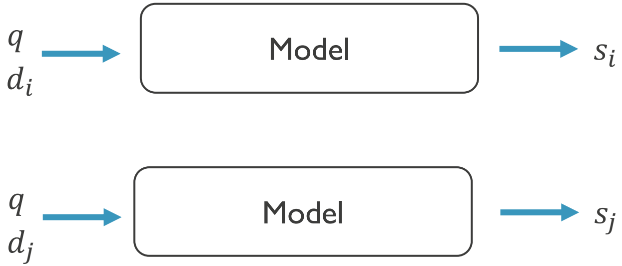
Number of Riders
1200
1000
800
600
400
200
0
0
500
1000
1500
Number of Riders
350
300
250
200
150
100
50
0
0
2
4
6


Average age of the rider (years)
Rank
0
5
10
15
20
25
30
35
0
20
40
60
80
100
120
140


Average rider career length (years)
Rank
0
2
4
6
8
10
12
14
16
0
20
40
60
80
100
120
140

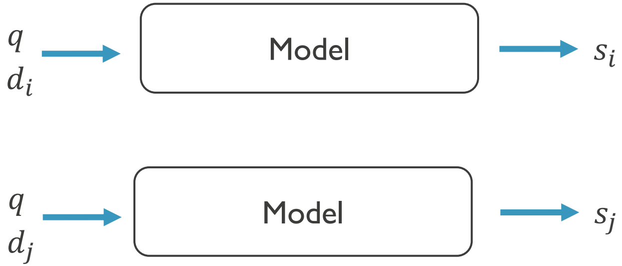

Rank in Liege-Bastogne-Liege
100
90
80
70
60
50
40
30
20
10
0
2004
2006
2008
2010
2012
2014
2016
2018
Year
Philippe Gilbert
Alejandro Valverde

Algorithm: LambdaMART



Burges, Christopher. (2010). From ranknet to lambdarank to lambdamart: An overview. Learning. 11.

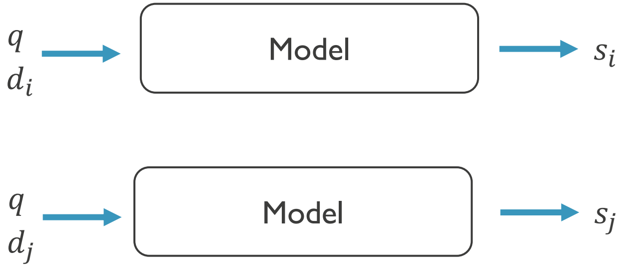
Algorithm: LambdaMART



$$C = \frac{1}{2}(1 - S_{ij})\sigma(s_i - s_j) + \log\left(1 + e^{-\sigma(s_i - s_j)}\right)$$

Burges, Christopher. (2010). From ranknet to lambdarank to lambdamart: An overview. Learning. 11.

Algorithm: LambdaMART



$$C = \frac{1}{2}(1 - S_{ij})\sigma(s_i - s_j) + \log\left(1 + e^{-\sigma(s_i - s_j)}\right)$$

Burges, Christopher. (2010). From ranknet to lambdarank to lambdamart: An overview. Learning. 11.

Algorithm: LambdaMART

$$\frac{\partial C}{\partial w_k} = \frac{\partial C}{\partial s_i}\frac{\partial s_i}{\partial w_k} + \frac{\partial C}{\partial s_j}\frac{\partial s_j}{\partial w_k} = \sigma\left(\frac{1}{2}(1 - S_{ij}) - \frac{1}{1+e^{\sigma(s_i - s_j)}}\right)\left(\frac{\partial s_i}{\partial w_k} - \frac{\partial s_j}{\partial w_k}\right)$$

$$= \lambda_{ij}\left(\frac{\partial s_i}{\partial w_k} - \frac{\partial s_j}{\partial w_k}\right)$$

Burges, Christopher. (2010). From ranknet to lambdarank to lambdamart: An overview. Learning. 11.

Algorithm: LambdaMART

$$\frac{\partial C}{\partial w_k} = \frac{\partial C}{\partial s_i}\frac{\partial s_i}{\partial w_k} + \frac{\partial C}{\partial s_j}\frac{\partial s_j}{\partial w_k} = \sigma\left(\frac{1}{2}(1-S_{ij}) - \frac{1}{1+e^{\sigma(s_i-s_j)}}\right)\left(\frac{\partial s_i}{\partial w_k} - \frac{\partial s_j}{\partial w_k}\right)$$

$$= \lambda_{ij}\left(\frac{\partial s_i}{\partial w_k} - \frac{\partial s_j}{\partial w_k}\right)$$

$$\lambda_{ij} = \frac{\partial C(s_i - s_j)}{\partial s_i} = \frac{-\sigma}{1+e^{\sigma(s_i-s_j)}}|\Delta_{NDCG}|$$

Burges, Christopher. (2010). From ranknet to lambdarank to lambdamart: An overview. Learning. 11.