Modelling Coach Decisions in Professional Cycling Teams

Maor Sagi[1], Paulo Saldanha[2], Guy Shani[1], and Robert Moskovitch[1,3]

[1] Software and Information Systems Engineering
Ben Gurion University of the Negev, Beer Sheva, Israel
[2] Israel – Premier Tech, Israel
[3] Population Health and Science, Ichan Medical School at Mount Sinai, NYC, USA
maos@post.bgu.ac.il,paulo.saldanha@powerwatts.com
shanigu@bgu.ac.il,robertmo@bgu.ac.il

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.

Keywords: Sports Analytics · Cycling · Recommender Systems

1 Introduction

Road bicycle racing has been an organized competitive team sport for 150 years [18]. In this sport, teams of cyclists are competing in a set of races held throughout the year. In each race, the cyclists from the same team work together. The result of the race is dictated by the first team member to cross the line, and hence, the entire team is working towards bringing one of the team cyclists to the finish line as fast as possible. There are one day races, as well as multiple days races, which are called stage races.

A team is typically composed of about thirty cyclists, but in each race there are only a relatively small number of participants, typically 6 to 8, from each team are allowed. The team's coach determines which cyclists will participate in each race [10], based on the individual recent workouts' performance, but also on the particular race conditions. Races vary in their locations and conditions, some races are in mountain areas, while other races are held in relatively flat conditions. The coach typically provides a scheduling plan for the upcoming races'

season ahead, and then designs a corresponding workouts schedule for each cyclist.
However, this plan often changes before each race, and the coach may decide to send a different cyclist than intended, based on the recent workouts performance of the cyclists. Modern cyclists use various gadgets and wearable devices to monitor a vast amount of data, such as the overall elevation gain and distance, their continuous measurements of heart rate, cadence, power, and estimated energy expenses, overall workout duration, and more. This data is typically uploaded to an online service, such as Training Peaks[4], which can be accessed by the coach. This framework allows the coach to monitor remotely both the physical condition of the cyclist, as well as their workouts' performance. Similar data is also collected during races, thus, all this data is available for the coach in making allocation decisions. Additional source of information about races and cyclists' performance in races, is available through the Pro Cycling Stats (PCS) website[5].

The main long term goal of our work is to build a decision support system that assists the coach in the cyclists' allocation to races decision making process. In this paper, however, we make a first step towards this ambitious goal, by analyzing the decision process of a particular team — the Israel-Premier Tech (IPT) team[6]. We introduce RaceFit, a recommendation system for cyclists' fit for a race, consisting of classification methods and the historical data of the workouts, and race data. We first extract features from the raw data. Then, we construct statistical models that predict, based on the features, the probability a cyclist will participate in a race, and the cyclists having the highest ranking are chosen. Then, we analyze which features contribute the most to the cyclist's fit for a race. Understanding which features are most important to the coach's decisions helps us to understand the coach decision process.

The contributions this paper are the following: we introduce the problem of recommending cyclists for a race stage, based on historical data; we introduce a dedicated dataset that we created for this task; we propose a classification based method for the recommendation of cyclists inclusion in a race stage, and provide a rigorous evaluation on the constructed real data.

2 Related Work

2.1 Sport Analytics

Much sport related research was done in recent years. Some of the research focuses on the physiological domain, such as estimating oxygen uptake during workout [32]. Some studies focus was on prediction tasks; For example, Hilmkil et al. [8] trained a model to predict the heart rate of a cyclist during a workout, and Kataoka et al. [11] focused on predicting power performance at the Tour the

[4] https://www.trainingpeaks.com/
[5] https://www.procyclingstats.com
[6] https://israelpremiertech.com/

France 2017 grand tour using GPS data, in addition to physical and environmental measurements. Other studies focused on predicting the cyclists' performance in races using their ranking points results [12,14,28]. Others analyzed tactical differences and their impact on the performance of the game; Memmert et al. [17] explored different formations in soccer, Narizuka and Yamazaki [19], explored teams' behavior patterns in football games while exchanging team's structure. Some works related to cycling analysis, focused on analyzing cyclist's physical measurements during a ride. Vogt et al. [30] analyzed such measurements in professional road cyclists, showing intensity and workload estimation during a ride better to be measured with power, rather than heart rate which was common before. Later, Vogt et al. [31] explored the power produced by cyclists during flat and mountain stages, and Lucía et al. [15] engaged in cycling races as well, compared grand tours' difficulty based on physiological loads. There are Erp et al. [7] that compared male and female intensity and load characteristics in training sessions, and more similar studies analyzed performance within cycling grand tours [23,25,24,29,6]. Additional works were done for assisting coaches and athletes in taking tactical decisions during races [21] or predicting race results using machine learning techniques [20,4].

2.2 Sports Recommendation Systems

Sports research studies are motivated typically by winning and best performance. Some works discuss the potential and the importance of improving the performance of athletes using machine learning techniques [2,26]. As far as we know, no studies have been done on cycling road recommendation systems. Therefore, we refer to recommendation systems in other sports, such as Running and Soccer. Soccer players' selection for a specific game seemed relatively resemble to the problem of cyclists allocation to races, even though there are quite significant differences between these tasks. For example, in soccer all the team attends each game, while in cycling multiple races can occur simultaneously, and anyways only several cyclists are required per race, and they can not be replaced during the race. Berndsen et al. [2] presented methods to predict marathon results and suggest assisting athletes to create a personalized training plan to optimize their workouts. Smyth [26] proposed relevant applications using recommendation methods that focus on the endurance sports domain and personalized training plans. Berndsen et al. [3] suggested in race recommendations for marathon runners. The model predicts the finish time of a runner during the race based on real-time data, and helps the runner make strategic decisions. Matthews et al. [16] discussed football results optimization, this work method uses reinforcement learning to choose the ultimate team formation. They presented a feasible approach with top results, but the limitations are the complexity of the algorithm, and it might be hard to explain to the team managers the decisions it takes are reasonable. Al-Asadi [1] discussed on decision support system, that helps football team managers to select the best player to assign to a specific position on the field based on his technical, physical and mental skills. After selecting the ultimate formation of players, the system chooses the available squad (i.e., 4-3-3, 3-5-2 etc.) according to the formation of players that have been chosen before. The system predicts the skills growth of the players and can help them to decide which players should be sold, buy, or keep.

3 Methods

We describe here in relatively high level RaceFit, a classification based recommendation method that we developed, as far as we know for the first time for cyclists' assignment for a race stage. Our data consists of the team's cyclists' workouts and race stages data, in which they participated. Thus, the task is to determine among the team's cyclists who are the best to fit the race stage according to historical decisions of the team's coach.

In order to learn the assignment of cyclists to stages, we learn a binary classifier that is intended to determine whether a cyclist will be assigned or not to a race stage. For that we designed classifier's examples that consist of a pair of a cyclist and stage, whose label is whether the cyclist participated, or not in the stage.

The classification's examples' describe an allocation pair of a stage-cyclist represented by a vector of features includes the stage, described by its properties (i.e., distance, elevation gain, etc.), the cyclist's properties, such as the weight, height, age, and some statistics of the cyclist from the PCS website, as well as features that summarize the cyclist's workout data, based on several recent weeks before the race stage. The detailed list of features appears in the Appendix. Once a classifier is induced, and a stage is given, the classifier will classify each of the cyclists, providing a probability for their participation in the stage. Then the top cyclists, having the highest probability, are recommended for the race stage. When a race has multiple stages (days) then the mean participation probabilities of the stages are used to determine the recommendation for the race.

However, RaceFit has several components that overcome challenges with the data, or enable to operate it with various settings. First, since the data is missing in various stage-cyclist examples, whether missing whole pairs of a stage and a cyclist, or specific features that are missing. For that we determine two thresholds, based on which whole examples, or features, are removed. Thus, the *missing values threshold* is defined by the percentage of missing values in the example, or in a feature. If the ratio of the missing values is above the threshold then the example, or the feature is removed. After the examples, or relevant features were removed, in order to fill the missing values, the framework uses imputation, for which various methods can be used. Additionally, another component of the framework is the classifier that can be used in order to perform the classification, as we demonstrate here with two options. The last parameter in the framework is the number of weeks prior to the given race stage from which features are extracted that summarize the cyclist's workouts' performance.

4 Evaluation

Our main goal in the evaluation of this paper was to evaluate the performance of RaceFit, given its parameters, and analyze which features of the cyclists, whether personal, or that were extracted from the workouts data, and the races' properties are most important in the prediction of the cyclists collection for a given race.

4.1 The IPT's Cyclists' Workouts and Races Dataset

To perform the analysis and evaluate the proposed model, we created our database based mainly on IPT's Training Peaks (TP) storage of raw workout and race data, and PCS's data. TP's data includes the team's cyclists' workout and race data that is uploaded from wearable devices, such as the heart rate monitor, and cycling computers. The physical measures include continuous measurements of the heart rate, cadence, power produced, speed, estimated calories burned, and energy consumption. Another source of information is the PCS data including cyclists' personal details – age, height, weight, and nation, and information about the races' stages, including their location, date, UCI race classifications, stage type – Classic Stage Race, Prologue, One day race, or Time Trial. There are 69 riders in our IPT's cyclists dataset, including 22.5K TP workouts from late 2016, and 92 Races from 2017 to 2022.

4.2 RaceFit Parameters and Baselines

We report here the results based on RaceFit's parameters' settings that performed best, however, we report all the parameters that we tried. Additionally, we describe here the baseline methods that we used for comparison, to make sure that the model learns something that is more meaningful than anything basic that could be used, such as the most frequent cyclist.

RaceFit parameters that we experimented with include: the thresholds for the removal of examples (cyclist in a stage); the thresholds for the removal of features; imputation methods; number of weeks prior to the race the summary of workouts consist of; type of classification method. We tried with missing values thresholds of 30%, 50% ,70% for the removal of examples, and for features, and using 30% performed slightly better than without.

After the removal, we tried the following imputation methods from the scikit-learn [22] package: SimpleImputer, KnnImputer and IterativeImputer, in which SimpleImputer performed best. All Imputers were used with the default parameters that are given in the package. The imputation method of SimpleImputer is based on imputing missing values with the mean value. KNNImputer is based on the k-nearest neighbors method, and the value imputed is the mean value of 5 nearest neighbors. The IterativeImputer imputes using all other features to estimate the current feature missing values with Bayesian Ridge Regressor [27]. We experimented also with 1 to 7 weeks prior to the race for the summarized workouts data, which performed quite similarly, while the use of 5 weeks performed better. We experimented with several classifiers, including XGBoost,CatBoost, RandomForest, AdaBoost, DecisionTree, and LogisticRegression, while we report here only the results of two classifiers that performed best — Catboost [5], and RandomForest [9], which is a popular and typically very successful classifier.

We defined two baselines that are based on the "popularity" of the cyclists, thus, a cyclist that was assigned to races most frequently till the current race (the number of races the cyclist participated in divided by the number of races the team participated in) is highly recommended. We had two versions of this baseline, one in general, and the second within the continent of the current race, but they both performed quite the same, and we show in the charts the continent based popularity baseline, which performed slightly better.

4.3 Experimental Plan

In the experiments below we use the following protocol: for each race r, in order of occurrence, all the data prior to r is used as the training set, including both the workouts data and the previous races data. We construct a separate model for each race. The model for the first race hence has no historical training data, and selects cyclists randomly. We expect the model accuracy to improve for later races. We train the various models and then predict which cyclists will be chosen for the race r. Following preliminary experiments, for the workout data, we focus on a 5 weeks window prior to the race, ignoring the final week before the race. This is because we assume that the final coach decisions were made prior to that week, to allow the cyclists time to prepare and travel to the race location. We evaluate each model separately and refer to the mean results over all models.

4.4 Evaluation Metrics

To measure the accuracy of RaceFit, given the list of recommended cyclists ordered by the decreasing predicted likelihood of participating in the race, we take the top i cyclists in the cyclists' list recommended by RaceFIt, and compute the $precision@i$ and $recall@i$ metrics, defined in equations 1 and 2 respectively. $Cyclists\ Raced$ refers to the actual cyclists that participated in the tested race, and the $top\ i\ Recommended$ refers to the top i recommended cyclists that were ordered by RaceFit.

$$precision@i = \frac{|Cyclists\ Raced \cap top\ i\ Recommended|}{|top\ i\ Recommended|} \tag{1}$$

$$recall@i = \frac{|Cyclists\ Raced \cap top\ i\ Recommended|}{|CyclistsRaced|} \tag{2}$$

Precision and recall are most appropriate here because this is a one class prediction problem, i.e., we are only interested in correctly predicting which cyclists will participate in the race.

While in many applications precision — the portion of correct predictions in the predicted list is most important, in this application the amount of correct predictions, cyclists that participate in the race, is very low, and hence recall seems more expressive. Moreover, it is likely that when the upcoming race requires n cyclists, the coach will ask the framework for at least n recommended cyclists, however, the coach typically will require additional k recommended cyclists spare in case any of the top favored cyclists will not be able to participate. We hence also compute $recall@(n+k)$ for $k = 0, 1 \ldots$.

4.5 Results

We now report the results of our empirical evaluation over the data of the IPT team. First we present the performance with or without imputation, then the performance when using the classifiers, and eventually we show the performance with the expanded evaluation metric of the recommendation of n + k cyclists.

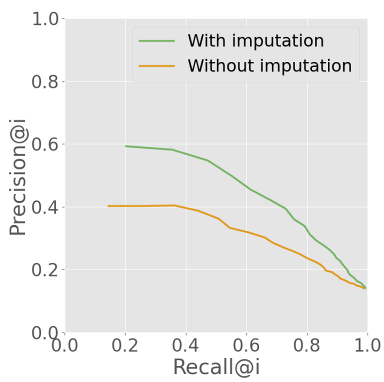

Fig. 1. The use of imputation outperforms without imputation.

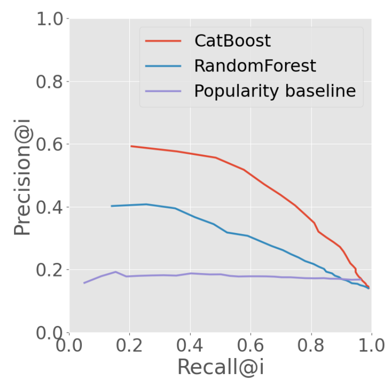

Fig. 2. Comparison of the $precision@i$ given $recall@i$ of the CatBoost, RandomForest classifiers, and the popularity baseline.

Figure 1 and Figure 2 show the $Precision@i$ and $Recall@i$, where $i = 1, 2, \ldots 30$ is the number of cyclists selected. Both figures report the mean performance over all the models in the experiment, where each model is trained on a different number of races. Initially the models are trained on no previous data, or only few, while increasingly the models are trained on previous races, which obviously become larger. Figure 1 shows that the mean performance of both classifiers when using the SimpleImputer imputation method performed better than not using imputation for all values of i. While we had experimented with more classifiers, due to limited space, Figure 2 shows only the results of the CatBoost and the RandomForest, in comparison to the popularity baseline (based on favoring the most frequently assigned cyclists), in which the use of the CatBoost classifier performed best for all values of i. Thus, in cycling assignment of athletes, unlike other sports like basketball for example, in which the same five players often open the games, here due to various constraints, popularity, or frequent cyclists are not repetitively being sent to races. For CatBoost, which was the best classifier, the precision for the first recommended cyclist is on average 0.6. That is, in 60% of the races, CatBoost correctly identifies, of the thirty cyclists in the team, the cyclists that participated in the race.

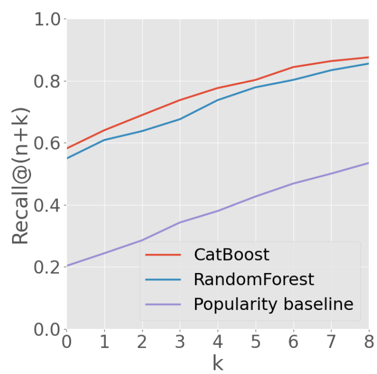

Fig. 3. Comparison of the $recall@(n+k)$ given k for CatBoost, RandomForest, and the popularity baseline.

Looking at the $recall@(n+k)$ (Figure 3) we see that when requesting the exact number of cyclists that are required for the race, both CatBoost and RandomForest recommend about 60% of the cyclists that participated in the race. Thus, for example, when allowing for an additional 5 names, CatBoost identifies 80% of the cyclists. However, identifying the final 20% of the cyclists is difficult for the classifiers. This shows that the classifiers do not fully capture the coach's decisions, and fail to correctly identify 20% of these decisions. We believe that the above results demonstrate that a significant portion of the coach decisions can be potentially identified using RaceFit. We now move to discuss which features in the data influence the decisions the most.

Figure 4 shows the most important features computed by the Relief feature importance, or ranking, algorithm [13]. The most important feature is the total distance that the cyclists passed during the workouts in the last 5 weeks before the coming race. It seems that this feature captures the number of recent workouts of the cyclist, which indicates cyclists that train harder. Also, this feature can identify cyclists that suffer from injuries (that are not directly reported in the personal reports), and are hence not available for the upcoming race. The next feature is the race's total distance, which appears to be highly indicative of the features of the race, and hence, helps to decide which cyclist should participate in the race. Next, the workouts total energy, which can let us an indication of how hard the last workouts were for the rider, which may indicate the riders' fitness. The next two features measure the difference in the cyclist performance in recent workouts, relative to the annual mean performance. This can indicate, on the one hand, cyclists that increased their performance recently, but also help again identify cyclists that reduced their performance due to a recent injury. Cyclist workouts energy is the power cyclist produced in Watts integrated with workouts duration average during the recent weeks. It is reasonable to assume that the high values of energy that the cyclist produced, together with long distances, would make the cyclist appropriate for the upcoming race, and vice versa. Elevation gain and elevation loss both describe the elevation changes during the

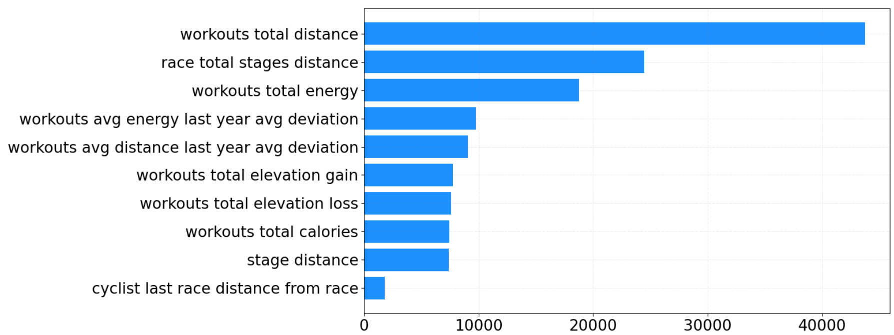

Fig. 4. Feature Importance - Relief

workouts. As some races are conducted in mountain areas, while others are conducted in the plain, it is reasonable that elevation changes during the workouts would influence the decision of which cyclist would be chosen. Calories burned during workouts can imply the intensity of the last workouts, it has a related meaning to the energy produced. The stage distance can identify races with very long distances, that might require higher endurance, and demand that cyclists spend more energy than others. The last feature is a geographic constraint, the distance of the last location the cyclist raced to the current race location. Again, to reduce travel, the coach tends to assign cyclists to races that are relatively nearby.

5 Discussion and Conclusions

In this paper, we describe the problem of pro cyclists' assignment to stages, which as far as we know, is investigated for the first time in a data driven approach. For that, we created a meaningful dataset of the Israel Premier Tech team that consists of races' stages, and the cyclists and their participation and performances, as well as the cyclists' workouts data. We propose RaceFit, a recommendation system that consists of a binary classifier that is trained on whether a cyclist will be assigned or not to a race stage. Our long term goal in our analysis is to model coaches decision making processes in assigning cyclists to races, and compare different professional cycling teams assignment strategies and relate them to the team's ongoing performance in races. We evaluated the proposed RaceFit framework, and saw in our experiments that the use of imputation is effective, while the CatBoost classifier performed best, also relative to the popularity baseline. Referring to the limitations of our study, it is worth mentioning that time trial stages were not included in the dataset in this analysis, due to their different nature of cycling style, which we intend to include in our future analysis. Additionally, we intend to create a larger database of multiple professional cycling teams and compare their decision making modeling, and important features. It will be desirable to understand whether successful teams make decisions in a similar way and different than less successful teams. Recommendation systems are designed to influence the user behavior, in our case, the coach. In many recommendation domains this may bias the information that is collected for future models, creating a feedback loop. However, our goal is to use data from teams, and learn their decision making to learn recommendations that lead to successful performance, or less, as we intend to do as future work.

Acknowledgements

This project was funded by a donation from Israel Premier Tech, and SYLVAN ADAMS FAMILY FOUNDATION ISRAEL.

References

1. Al-Asadi, M.A.M.: Decision support system for a football team management by using machine learning techniques. Xinyang Teachers College **10**(2), 1–15 (2018)
2. Berndsen, J., Lawlor, A., Smyth, B.: Running with recommendation. In: HealthRecSys@ RecSys. pp. 18–21 (2017)
3. Berndsen, J., Smyth, B., Lawlor, A.: Pace my race: recommendations for marathon running. In: Proceedings of the 13th ACM Conference on Recommender Systems. pp. 246–250 (2019)
4. De Spiegeleer, E.: Predicting cycling results using machine learning
5. Dorogush, A.V., Ershov, V., Gulin, A.: Catboost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363 (2018)
6. van Erp, T., Sanders, D.: Demands of professional cycling races: influence of race category and result [published online ahead of print july 16, 2020]. Eur J Sport Sci
7. Erp, T., Sanders, D., de Koning, J.: Training characteristics of male and female professional road cyclists: A four-year retrospective analysis. International journal of sports physiology and performance (07 2019). https://doi.org/10.1123/ijspp.2019-0320
8. Hilmkil, A., Ivarsson, O., Johansson, M., Kuylenstierna, D., van Erp, T.: Towards machine learning on data from professional cyclists. arXiv preprint arXiv:1808.00198 (2018)
9. Ho, T.K., et al.: Proceedings of 3rd international conference on document analysis and recognition. In: IEEE. pp. 278–282 (1995)

10. Internationale, U.C.: Uci cycling regulations part i - general organisation of cycling as a sport (2022), https://assets.ctfassets.net/761l7gh5x5an/wQympSG6EWlKq6o6HKw9E/d4f7039ce3bf3963b457ae35cf5449bd/1-GEN-20220301-E.pdf
11. Kataoka, Y., Gray, P.: Real-time power performance prediction in tour de france. In: International workshop on machine learning and data mining for sports analytics. pp. 121–130. Springer (2018)
12. Kholkine, L., Servotte, T., De Leeuw, A.W., De Schepper, T., Hellinckx, P., Verdonck, T., Latré, S.: A learn-to-rank approach for predicting road cycling race outcomes. Frontiers in sports and active living **3** (2021)
13. Kira, K., Rendell, L.A., et al.: The feature selection problem: Traditional methods and a new algorithm. In: Aaai. vol. 2, pp. 129–134 (1992)
14. Leeuw, A.W.d., Heijboer, M., Hofmijster, M., Zwaard, S.v.d., Knobbe, A.: Time series regression in professional road cycling. In: International Conference on Discovery Science. pp. 689–703. Springer (2020)
15. Lucía, A., Hoyos, J., Santalla, A., Earnest, C., Chicharro, J.L.: Tour de france versus vuelta a espana: which is harder? Medicine & Science in Sports & Exercise **35**(5), 872–878 (2003)
16. Matthews, T., Ramchurn, S., Chalkiadakis, G.: Competing with humans at fantasy football: Team formation in large partially-observable domains. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 26 (2012)
17. Memmert, D., Raabe, D., Schwab, S., Rein, R.: A tactical comparison of the 4-2-3-1 and 3-5-2 formation in soccer: A theory-oriented, experimental approach based on positional data in an 11 vs. 11 game set-up. PloS one **14**(1), e0210191 (2019)
18. Mignot, J.F.: The history of professional road cycling. In: The economics of professional road cycling, pp. 7–31. Springer (2016)
19. Narizuka, T., Yamazaki, Y.: Clustering algorithm for formations in football games. Scientific reports **9**(1), 1–8 (2019)
20. Ofoghi, B., Zeleznikow, J., MacMahon, C., Dwyer, D.: A machine learning approach to predicting winning patterns in track cycling omnium. In: IFIP International Conference on Artificial Intelligence in Theory and Practice. pp. 67–76. Springer (2010)
21. Ofoghi, B., Zeleznikow, J., MacMahon, C., Dwyer, D.: Supporting athlete selection and strategic planning in track cycling omnium: A statistical and machine learning approach. Information Sciences **233**, 200–213 (2013)
22. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
23. Rodríguez-Marroyo, J.A., Villa, J.G., Pernía, R., Foster, C.: Decrement in professional cyclists' performance after a grand tour. International journal of sports physiology and performance **12**(10), 1348–1355 (2017)
24. Sanders, D., Heijboer, M.: Physical demands and power profile of different stage types within a cycling grand tour. European journal of sport science **19**(6), 736–744 (2019)
25. Sanders, D., Heijboer, M., Hesselink, M.K., Myers, T., Akubat, I.: Analysing a cycling grand tour: Can we monitor fatigue with intensity or load ratios? Journal of sports sciences **36**(12), 1385–1391 (2018)
26. Smyth, B.: Recommender systems: A healthy obsession. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, pp. 9790–9794 (2019)

27. Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. Journal of machine learning research **1**(Jun), 211–244 (2001)
28. Van Bulck, D., Vande Weghe, A., Goossens, D.: Result-based talent identification in road cycling: discovering the next eddy merckx. Annals of Operations Research pp. 1–18 (2021)
29. Van Erp, T., Hoozemans, M., Foster, C., De Koning, J.J.: Case report: load, intensity, and performance characteristics in multiple grand tours. Med Sci Sports Exerc **52**(4), 868–875 (2020)
30. Vogt, S., Heinrich, L., Schumacher, Y.O., Blum, A., Roecker, K., Dickhuth, H.h., Schmid, A.: Power output during stage racing in professional road cycling. Medicine and science in sports and exercise **38**(1), 147 (2006)
31. Vogt, S., Schumacher, Y.O., Blum, A., Roecker, K., Dickhuth, H.H., Schmid, A., Heinrich, L.: Cycling power output produced during flat and mountain stages in the giro d'italia: a case study. Journal of sports sciences **25**(12), 1299–1305 (2007)
32. Zignoli, A., Fornasiero, A., Ragni, M., Pellegrini, B., Schena, F., Biral, F., Laursen, P.B.: Estimating an individual's oxygen uptake during cycling exercise with a recurrent neural network trained from easy-to-obtain inputs: A pilot study. PloS one **15**(3), e0229466 (2020)

6 Appendices

The list of features while the firsts are the races features and the lasts are the cyclists' features. In Table 1 races' features from PCS. Table 2 presents the features of cyclists taken from PCS. Later, Table 3 present features of Training Peaks.

Table 1. Pro Cycling Stats Races' features

Feature	**Description**
race difficulty level	The race difficulty level (e.g., Hard, Intermediate, Easy)
race continent	Race's continent
stage ranking	Ranking of stage difficulty level
stage profile score	Score of the stage that indicates the race profile hardness
stage profile type	The race profile type (i.e., hilly, mountains, flat)
stage distance	Stage average distance
stage elevation gain	Stage average elevation gain
stage temp avg	Stage average temperature
race classification	The race class (i.e., WorldTour 1.WT)
race total stages distance	The race stages distance sum
race total stages elevation gain	The race stages elevation gain sum
race number of stages	The number of stages in race

Table 2. Pro Cycling Stats Cyclists' features

Feature	**Description**
cyclist weight	
cyclist height	
cyclist age	
cyclist weeks count in team	Number of weeks since the cyclist joined the team
cyclist total races count	Number of races cyclists raced
cyclist total races count last year	Number of races a cyclist raced during the last year
cyclist total races count in race continent	Number of races a cyclist raced in the current race's continent
cyclist races rate	Number of races cyclist raced divided by the number of races of the team
cyclist races rate in continent	Number of races cyclist raced in race's continent divided by the number of races of the team in the current race's continent
cyclist popularity ranking	Ranking of cyclist races rate compared to the other cyclists in the team
cyclist popularity ranking in continent	Ranking of cyclist races rate in continent compared to the other cyclists in the team
cyclist weeks count since raced	The number of weeks since cyclists last raced
cyclist last race distance from race	The distance in kilometers from the last race cyclist raced to the current race

Table 3. Training Peaks Cyclists' features

Feature	**Description**
workouts total distance	Last weeks workouts distance sum
workouts total duration	Last weeks workouts duration sum
workouts avg hr	Average heart rate of the last workouts
workouts total calories	Last workouts claries burned sum
workouts total TSS	Last workouts TSS sum, TSS is training intensity measure, consider normalized power, intensity factor duration
workouts avg IF	Average intensity factor (IF) of the last workouts, indicates how hard the workout was in relation to the cyclist fitness
workouts avg speed	Average speed of the last workouts
workouts avg power	Average power produced (in Watt) of last workouts
workouts avg norm power	Average normalized power of the last workouts
workouts total energy	Last workouts energy sum, consider power produced and workout duration
workouts total elevation gain	Last workouts elevation gain sum, the elevation gain is the total amount cyclist climb in a workout
workouts total elevation loss	Last workouts elevation loss sum, the elevation loss is the total amount cyclist descend in a workout
workouts avg elevation	Average elevation of the last workouts
workouts avg temp	Average temperature of the last workouts
workouts avg cadence	Cadence average of last workouts, cadence is the rate which cyclist pedals
workouts count	The number of workouts in the last weeks
workouts count $>$ 28 deg	The number of workouts in the last weeks, temperature was above 28 degrees
workouts avg calories last year avg deviation	Deviation of last weeks' calories mean from annual mean
workouts avg elevation gain last year avg deviation	Deviation of last weeks' elevation gain mean from annual mean
workouts avg power last year avg deviation	Deviation of last weeks' power mean from annual mean
workouts avg elevation loss last year avg deviation	Deviation of last weeks' elevation loss mean from annual mean
workouts avg elevation last year avg deviation	Deviation of last weeks' elevation mean from annual mean
workouts avg duration last year avg deviation	Deviation of last weeks' duration mean from annual mean
workouts avg temp last year avg deviation	Deviation of last weeks' temperature mean from annual mean
workouts avg distance last year avg deviation	Deviation of last weeks' distance mean from annual mean
workouts avg TSS last year avg deviation	Deviation of last weeks' TSS mean from annual mean
workouts avg IF last year avg deviation	Deviation of last weeks' IF mean from annual mean
workouts avg speed last year avg deviation	Deviation of last weeks' speed mean from annual mean
workouts avg cadence last year avg deviation	Deviation of last weeks' cadence mean from annual mean
workouts avg hr last year avg deviation	Deviation of last weeks' heart rate mean from annual mean
workouts avg norm power last year avg deviation	Deviation of last weeks' normalized power mean from annual mean
workouts avg energy last year avg deviation	Deviation of last weeks' energy mean from annual mean