Predicting Cycling Performance Using Strava Cycling Data

Rhishabh Hattarki

Abstract—This research delves into predicting cycling performance, specifically Functional Threshold Power (FTP), using comprehensive Strava activity data. Embracing a seven-year cycling journey, the study leverages heart rate monitors, power meters, and sensors to construct a nuanced dataset. Traditional FTP tests, often impeded by factors like overfatigue and mental stress, serve as benchmarks for improvement. The primary goal is to develop a predictive model circumventing the need for these tests, exploring alternative approaches based on Strava data. Research objectives encompass trend analysis in training zones, performance changes, FTP prediction, model defense, and causal relationship evaluation. A critical literature review highlights the research gap in FTP prediction using Strava data, setting the stage for this novel contribution.

Index Terms— cycling training, performance improvement, Strava data analysis

1 Introduction

Embarking on a seven-year cycling journey, my Strava activity log encapsulates a meticulous collection of data through heart rate monitors, power meters, and sensors. Engaging in max power tests and bike races has further enabled an objective assessment of performance.

Within the realm of cycling, training styles are often distinguished by the time allocated to specific heart rate or power zones during training cycles. Existing research sheds light on their varied effects on performance, with traditional metrics like Functional Threshold Power (FTP), derived from a demanding one-hour or a 20-minute test, serving as benchmarks for improvement [1]. However, the feasibility of such tests is hindered by factors such as overfatigue and mental stress.

The core objective of this project is to develop a predictive model for cycling performance, specifically targeting the estimation of Functional Threshold Power (FTP) without the need for traditional FTP tests.

Considering the challenges posed by conventional testing methods, the project aims to explore alternative approaches for estimating a cyclist's FTP using their training data, leveraging the structured information provided by the Strava platform.

The research questions guiding this endeavor encompass describing trends within the time spent in training zones and performance changes over training periods, predicting FTP using Strava data from a training block, defending the proposed FTP prediction model, and evaluating the causal relationships implied by the model.

Predicting FTP from training data not only streamlines evaluation but also alleviates stress, enabling cyclists to optimize their training more efficiently. While basic data analysis tools provided by Strava and add-on extensions exist, the novel contribution of FTP prediction addresses a significant knowledge gap within the field.

1.1 Background Knowledge

Cycling, a sport embraced globally by participants at various skill levels, offers a nuanced journey of skill development. Progressing through these levels involves multifaceted strategies for improvement beyond merely increasing mileage. Strava, a popular platform among cyclists, plays a pivotal role in this journey by enabling meticulous tracking of rides. Cyclists upload comprehensive data, including GPS, speed, elevation, heart rate, and power, onto Strava, leveraging the social aspects of the platform to enhance discipline and motivation [2].

For cyclists utilizing power meters, tracking progress often involves conducting a Functional Threshold Power (FTP) test. This crucial metric represents the maximum power a cyclist can sustain for one hour. Traditionally, this test demands an all-out effort for a full hour, but a more feasible alternative is the 20-minute test. In this version, cyclists exert maximum effort for 20 minutes following a structured warmup that includes three one-minute efforts and a five-minute all-out effort [1]. The resulting 20-minute max power is then multiplied by 0.95 to estimate the 1-hour power (FTP). To optimize performance, cyclists strategically train in specific zones tailored to achieve distinct outcomes. Training in zone 2, for instance, proves beneficial for enhancing endurance and establishing a solid foundation for subsequent higher-intensity training phases [1] [3]. This holistic approach to cycling training, coupled with the insights provided by platforms like Strava, underscores the comprehensive nature of performance improvement within the cycling community.

1.2 Research Objectives

RO1: To describe the trends within the time spent in training zones over the duration of training periods.
RO2: To describe the trends within the performance changes over the duration of training periods.
RO3: To predict the value of functional threshold power (FTP) using the Strava data of the training block.
RO4: To defend the model for performing the prediction of FTP in RO3.
RO5: To evaluate the causal relationships implied by the RO3 model.

2 RELATED WORK

In their research, Ronald Andrew Stockwell and Andrea Corradini introduced a model to predict a cyclist's Functional Threshold Power (FTP) using Multiple Linear Regression with data from Golden Cheetah [4]. While our shared objective involves predicting FTP through machine learning, key distinctions exist. I utilize Strava data, whereas they employed a preexisting dataset from Golden Cheetah. Unlike their exclusive focus on power fields, my approach explores models that combine power fields with non-power fields like heart rate and moving time. A notable difference lies in their inclusion of 30 min max power, requiring a maximal effort test, while my goal is to predict FTP without necessitating such tests. In essence, while both studies share the overarching theme of predicting FTP through machine learning, the disparities in dataset, feature selection, and avoidance of maximal effort tests distinguish my approach.

In a study by Giorgos Demosthenous, Marios Kyriakou, and Vassilis Vassiliades [5], they focused on predicting the average speed of a cycling ride using data from various sensors. While our research shares a commonality in predicting continuous metrics from cycling activity data, it diverges in its aim, as it does not target the prediction of Functional Threshold Power (FTP) or any power-related data for cyclists. Moreover, their dataset comprises only 14 rides, later reduced to 12, differing significantly from the extensive dataset utilized in my research. Additionally, their inclusion of unconventional sensors like back sensors and inclination percentage detection sensors sets their study apart, introducing a distinctive approach to data collection compared to the more conventional sensor usage in cycling activities like power meter, heart rate belt, etc.

In their study, Iztok Fister, Duan Fister, and Simon Fong utilized a collection of GPX files from sports activities to construct a personalized trainer, employing various data mining techniques such as association, classification, clustering, and prediction [6]. Although their focus doesn't specifically involve predicting Functional Threshold Power (FTP), their research stands as an illustrative example of the diverse applications achievable through the exploration of cycling activity data. The study showcases the potential for leveraging data mining techniques to develop a personalized trainer, shedding light on the broader possibilities within the realm of utilizing cycling data for insightful and practical outcomes.

Simon A. Jobson, Louis Passfield, Greg Atkinson, Gabor Barton, and Philip Scarf conducted a comprehensive review of methodologies for quantifying training load in their research [7]. They explored various metrics such as training volume, session rating of perceived exertion, and heart rate-derived training impulse (TRIMP). While not centered on predicting Functional Threshold Power (FTP), their paper delves into the challenges of analyzing cycling training data, particularly addressing the inherent variability in power output. This review provides valuable insights into methodologies employed by other researchers in the analysis of cycling data, offering a broad perspective that aligns with my work. Additionally, it concludes with suggestions on utilizing cycling training data to enhance overall performance, adding a practical dimension to the discussion.

Agrin Hilmkil, Oscar Ivarsson, Moa Johansson, Dan Kuylenstierna, and Teun van Erp developed a predictive model that anticipates heart rate for the entire duration of a ride, utilizing factors such as power, speed, and time. Although their focus is on predicting heart rate rather than Functional Threshold Power (FTP), their methodology aligns closely with my research as they employ similar metrics [8]. Moreover, their model also incorporates the heart rate from 30 seconds prior to the current time, revealing an innovative approach to modeling heart rate dynamics throughout a ride.

Leonid Kholkine, Tom De Schepper, Tim Verdonck, and Steven Latré presented a machine learning approach in their research, aiming to predict road cycling race performance [9]. It's crucial to note that their focus was on forecasting race results rather than Functional Threshold Power (FTP). The dataset they utilized comprised statistics from the previous two years' race results, diverging from the cycling activity data central to my research. Consequently, while both studies employ machine learning for predictive purposes in the cycling domain, the dissimilarity lies in the specific performance metric targeted and the nature of the data employed.

In the broader landscape of data analysis and machine learning applications on sports activities data, particularly in the context of cycling, a notable gap exists in research specifically targeting the prediction of Functional Threshold Power (FTP). While various studies explore diverse aspects of cycling data, such as race performance, heart rate prediction, and training load quantification, none have specifically delved into predicting FTP using Strava data with metrics like average heart rate, time in zones, and moving time. This observation suggests that within the domain of cycling activity data analysis and machine learning, there is a discernible gap in the existing literature, paving the way for the current research to contribute valuable insights.

3 EXPLORATORY DATA ANALYSIS

The dataset for this project was generated by fetching my own Strava activities data using the Strava APIs [10]. The APIs require an access token that can be obtained by following their authentication steps. This access token is used for all subsequent API calls. The APIs were used to fetch activities, zones per activity and the power stream per activity.

Various preprocessing techniques were applied to the data to serve distinct objectives such as generating summary statistics, visualizations, and models using different data subsets. Notably, some data underwent aggregation, transforming time series data into independent, unrelated data points. This facilitated smoother randomization and data augmentation processes. The example of preprocessed data provided in Table 1 illustrates the synthesis of ride metrics, zones, FTP, power, and heart rate, offering a glimpse into the comprehensive dataset and its associated meanings.

During the data exploration phase, I delved into various relationships within the dataset, employing scatter matrix plots for key comparisons such as distance vs average speed, distance vs total elevation gain, and total elevation gain vs average speed. Additionally, histograms were generated for location country and moving time, providing insights into ride distribution. Noteworthy patterns emerged, revealing a positive correlation between total elevation gain and ride length, while higher average speeds correlated with shorter rides and lower elevation gains. The histograms underscored a concentration of rides in India due to an extended stay, and moving time exhibited a broad distribution with minimal variance.

TABLE 1
INTERPRETABLE DATA

Column	Data	Meaning
distance	63539.8	Distance traveled in meters
moving_time	8189	Duration of the ride cyclist was moving
start_date_local	2023-10-17T10:00:06Z	Start time and date of the ride
kilojoules	1276.9	Amount of work done in the ride in kj
average_heartrate	152.1	Average heartrate for the entire ride in beats per minute
HR Zone 1	4257.0	Time spent in heart rate zone 1 in seconds
...		Covers HR Z2-4
HR Zone 5	0.0	Time spent in heart rate zone 5 in seconds
Power Zone 1	618.0	Time spent in power zone 1 in seconds
...		Covers power Z2-10
Power Zone 11	20.0	Time spent in power zone 11 in seconds
ftp	123.1160417	Estimated ftp by calculating 0.95 * 20min max power in watts

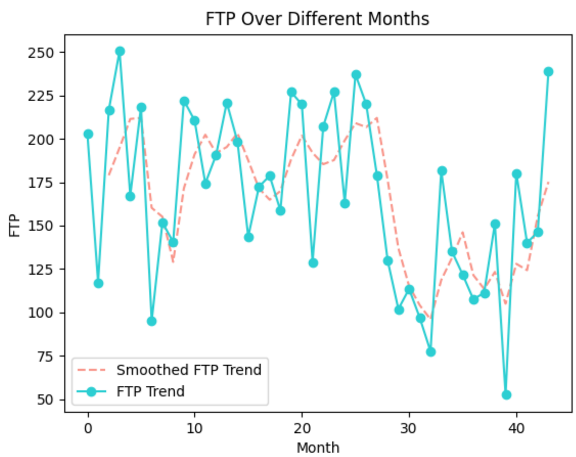

Fig. 1. Trends in FTP over Different Months

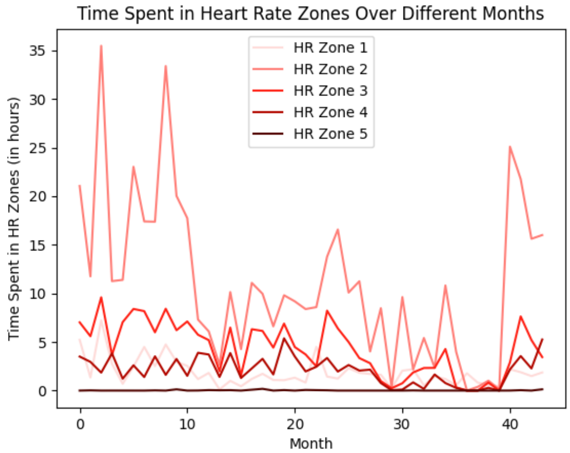

Fig. 2. Trends in time spent in HR Zones over Different Months

To delve deeper into performance trends, two additional plots were constructed. The first visualized the fluctuation in FTP over training months, showcasing periodic declines during off periods followed by recovery during intense training. However, a plateau was evident between 225 and 250. The second plot unveiled the evolution of time spent in heart rate zones over months, indicating consistent time in Zone 4 and variable time in Zone 2, with higher values in the initial and final training phases. These analyses contribute to a nuanced understanding of performance dynamics throughout the training period and satisfy the objectives RO1 and RO2.

4 METHODOLOGY

To predict Functional Threshold Power (FTP), I leveraged aggregated data from the pre-processing stage, focusing on the sum of all parameters within specific training blocks. Multiple datasets, including zones_ftp_power_agg, zones_ftp_hr_agg, and zones_ftp_power_hr_agg, were employed, each comprising a blend of zones, FTP, power, and heart rate data. The key distinctions are that power_agg only has power zones, hr_agg only has hr zones and power_hr_agg has both power and hr zones. To enhance the dataset size, augmentation techniques were applied, resulting in six datasets—three standard aggregates and three augmented variants. The training phase involved experimenting with three regression models: Linear Regression, Ridge Regression, and Lasso Regression, leading to a total of 18 models.

The datasets were split into training and testing sets using an 80-20 ratio, reserving 20% for testing. To validate prediction accuracy, two FTPs were predicted separately, evaluated using the metrics Mean Squared Error and R squared error. Additionally, experiments were conducted to assess the impact of incremental changes in two potentially crucial features, average_heartrate and Power Zone 7 on FTP predictions. Specifically, the Lasso Regression model was applied to the zones_ftp_power_hr_agg dataset for both experiments. This involved systematically increasing and decreasing the values of these features to observe their effects on FTP predictions, contributing valuable insights into feature sensitivity and model behavior.

5 Results and Discussion

Among the 18 models employed for prediction and evaluation in the preceding phase, the majority exhibited a notable error exceeding 500 Mean Squared Error (MSE), equivalent to approximately 22 watts. Given a 250w FTP, this translates to an error of 8.8%, a substantial deviation that could require a cyclist months to achieve. Despite this, there was a standout model achieving an MSE of 382.7, approximating to ~19.5w, although still not ideal. Table 2 presents the performance of the three best models, revealing that the top performers were all trained on a dataset variant incorporating both power and heart rate zones. Lasso outperformed the others, closely followed by Ridge. While data augmentation did have an impact, it did not surpass the performance achieved with actual data, emphasizing the importance of the dataset quality in model training. This satisfies the objective RO3.

To validate the reliability of the model and fulfill objective RO4, I conducted a sanity check by predicting FTP on aggregate data for two distinct rows. The predicted FTP values closely aligned with the actual FTP, registering at 100.9 vs 111.3 and 184.25 vs 180.3. The proximity of these values to the ground truth values instilled confidence in the model's accuracy and its ability to effectively predict FTP. This process served as a crucial step in ensuring the model's robustness and its alignment with real-world FTP outcomes.

In addition to model evaluation, I conducted targeted experiments to gauge the influence of two specific features on FTP prediction. The selected features, denoted as A (average heartrate) and B (Power Zone 7), were incrementally adjusted to observe their effects on FTP. Systematically increasing A, B, and both concurrently consistently resulted in FTP increments, indicating a positive correlation between these features and FTP. Notably, the simultaneous increase of both features yielded the highest FTP predictions. Intriguingly, when A was decreased while B increased, the initial iterations saw an FTP rise, followed by a subsequent decline. This observation suggested that A (average heartrate) held greater influence in the model compared to Power Zone 7, emphasizing its higher weight in shaping FTP predictions. These findings contribute valuable insights into feature importance and their nuanced impact on FTP predictions. This also satisfies the objective RO5.

TABLE 2
Top 3 Performing Models

Dataset	Model	MSE	R2
zones_ftp_power_hr_agg	Lasso	382.7173371675	0.744259593
zones_ftp_power_hr_agg	Ridge	503.7133687483	0.6634072997
zones_ftp_power_hr_agg_augmented	Ridge	512.0358563414	0.7544735501

6 Conclusion

In conclusion, this research advances our understanding of predicting FTP in cycling through machine learning applied to Strava data. Exploring various models, including Linear Regression, Ridge, and Lasso Regression, revealed challenges in achieving precise predictions, with many models exhibiting substantial errors. Notably, the Lasso Regression model on zones_ftp_power_hr_agg dataset showed promise, emphasizing the role of data quality. Sanity checks aligned predicted FTP closely with actual values, validating the model's reliability. Feature experiments underscored the significance of average heartrate and Power Zone 7 in FTP predictions, shedding light on nuanced relationships. This study not only contributes to the evolving landscape of cycling data analysis but also highlights avenues for future research and model refinement.

Future research endeavors should aim to enhance the model's precision, targeting an error reduction to less than 1%. Experimentation with diverse models will be crucial in understanding their efficacy in predicting Functional Threshold Power (FTP). Currently, the model relies solely on individual data, introducing challenges related to data inconsistency and infrequent FTP tests. To overcome this limitation, future studies could incorporate data from various users, enabling the construction of distinct or amalgamated models. Addressing issues of non-consistent data and sporadic FTP tests requires a more controlled approach to data collection, ensuring a richer dataset. Additionally, dealing effectively with incomplete data, where certain rides lack heart rate or power information, is imperative for robustness. Striving for a model that accommodates variations in data richness will enhance its applicability and utility for a broader audience.

References

[1] H. Allen, A. Coggan and S. McGregor, Training + Racing with a Power Meter, Boulder, CO, USA: Velopress, 2019.

[2] J. Maria Peterson, E. Kemps, L. Lewis and I. Prichard, "Associations Between Commercial App Use and Physical Activity: Cross-Sectional Study," *Journal of Medical internet Research,* vol. 22, no. 6, 2020.

[3] E. Faria, D. Parker and I. Faria, "The Science of Cycling," *Sports Medicine,* vol. 35, pp. 285-312, 2005.

[4] R. Andrew Stockwell and A. Corradini, "A Machine Learning Approach to Predict Cyclists' Functional Threshold Power," in *Intelligent Data Engineering and Automated Learning – IDEAL 2023* , 2023.

[5] G. Demosthenous, M. Kyriakou and V. Vassiliades, "Deep reinforcement learning for improving competitive cycling performance," *Expert Systems with Applications,* vol. 203, 2022.

[6] I. Fister, D. Fister and S. Fong, "Data Mining in Sporting Activities Created by Sports Trackers," in *2013 International Symposium on Computational and Business Intelligence*, New Delhi, India, 2013.

[7] S. Jobson, L. Passfield, G. Atkinson, G. Barton and P. Scarf , "The Analysis and Utilization of Cycling Training Data," *Sports Medicine,* vol. 39, p. 833–844, 2009.

[8] A. Hilmkil, O. Ivarsson, M. Johansson, D. Kuylenstierna and T. van Erp, "Towards Machine Learning on data from Professional Cyclists," in *12th World Congress on Performance Analysis of Sports*, Opatija, Croatia, 2018.

[9] L. Kholkine, T. De Schepper, T. Verdonck and S. Latré , "A Machine Learning Approach for Road Cycling Race Performance Prediction," in *Machine Learning and Data Mining for Sports Analytics* , 2020.

[10] Strava, "Strava API Documentation," Strava, [Online]. Available: https://developers.strava.com/docs/reference/. [Accessed November 2023].