1 The Project

My project aims to use machine learning to evaluate rowing training to predict training and performance outcomes. This will require the collection of data from users, basic analysis of this data to find basic trends and insights, and then the use of Machine Learning techniques to produce predictions of training outcomes or expected performances. The data collection is already in progress using a purpose-built website to onboard participants and manage the collection of data from various data providers. Data analysis and insights generated serve two purposes. Firstly, it allows me to become familiar with the data collected and build necessary data pipelines and processes which may be used in the machine learning step. Additionally, it allows me to provide some feedback or value to potential users to keep them interested in continuing to provide data. Finally, the machine learning step will aim to build a model to predict performances or suggest training sessions and intensities to hit a given target. The exact outcome of this model will be clarified once data is collected and explored. My primary motivation is to provide better feedback to rowers, something I have found lacking when using various training platforms like Whoop and Strava. I will aim to:

1. Obtain research ethics approval and recruit a minimum of 10 participants

2. Establish a data pipeline to automatically collect data from commonly used training tracking services and APIs (Strava, Concept2 Logbook)

3. Determine weekly mileage, time in training zones, trends in training for each week, and each athlete through data analysis and insight generation.

4. Provide feedback to athletes using the above data analysis.

5. Using machine learning, predict an athlete's performance given a set of training data. This might include:

    (a) suggesting sessions (mileage and intensity) given previous trends for an athlete
    (b) predicting performances on ergo tests (2k, 500m, 6k, 30min r20)

Much of the machine learning part of the project will be guided by the PhD Thesis of Tania Churchill [1] which developed a hybrid artificial neural network ensemble model (HANNEM) to model the relationship between training and performance for elite cyclists.

2 The Literature

Rowing is classified as a strength-endurance sport, this means training focuses on building aerobic, anaerobic, and physical power while developing rowing technique [2]. Much of the time spent training is on building endurance. There are multiple training approaches to develop endurance, with two basic training patterns being threshold training and polarized training [3]. These patterns dictate how much time an athlete spends at different intensities, or "training zones". For most amateur athletes, 5 aerobic zones are used, labelled zone 1 - zone 5, where each zone corresponds with a percentage of a calculated maximum Heart Rate ($\mathrm{HR_{max}}$). Athletes with access to specialised equipment can build a more scientifically accurate set of training zones using a lactate test. For rowing this is typically a "step test", where rowers will go through a series of timed intervals, working at prescribed power levels. After each interval the blood lactate is measured (in mmol/L) this is used, in conjunction with the wattage output, to determine training zones. These training zones are normally defined as (T1) basic oxygen utilization training (UT2) [lactate = 0-2 mmol/L], (T2) oxygen utilization training (UT1) [lactate = 2-3.5 mmol/L], (3) anaerobic threshold training (AT) [lactate = 3.5-4.5 mmol/L], (4) oxygen transport training (TR) [lactate = 4.5-6 mmol/L], and (5) anaerobic capacity training (AN) [lactate $\geq$ 6 mmol/L] [4]. Depending on how rigorous the testing protocol was, Heart Rate zones may be calculated for each zone, these may vary from the aerobic zones calculated from $\mathrm{HR_{max}}$. Calculating time spent in these zones can be used to approximate an athlete's fatigue and help manage loads.

There have been many approaches used to model human performance. Starting with the Bannister model in 1975 where performance was calculated as, Performance = Fitness - Fatigue [5]. This method of modelling performance is quite limiting as biological processes are inherently non-linear – many variables affect performance beyond just fitness and fatigue. Neural Networks have been used to predict performance as early as 2002 where Edelmann-nusser *et al.* [6] predicted the Olympic finishing times for swimmers. Machine learning, particularly neural networks, generally requires a large volume of data in order to be effective. Churchill (2014) [1] presented a HANNEM approach which allowed for only three athletes' data to be used when developing a neural network

to predict performance. This showed that highly personalised predictions can be made to predict performance based on the training behaviour and response of each athlete. Finally, there is a project called AIROW, based in the University of Vienna, which is exploring the use of machine learning and big data to optimise training load of top endurance athletes, beginning with rowers [7].
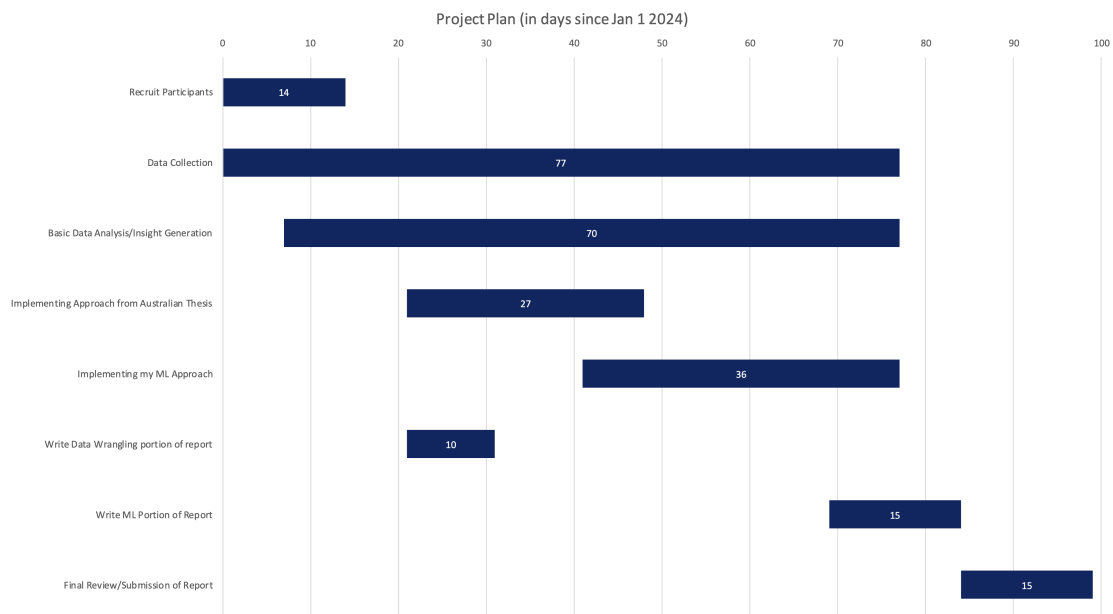
3   The Plan

Figure 1: The Project Plan Gantt Chart

Data collection and the basic analysis pieces will run throughout most of the project in order to continually provide new data to train models on, as well as to provide feedback to keep participants interested in continuing to provide data to the project. A majority of time is spent actually working on the project with roughly 30 days total dedicated to writing the overall report. Some writing will happen during the course of the development work to make sure I am able to remember what to write about.

4   The Ethics

An ethics approval was requested and granted for the collection and storage of user data. There is a wide range of data that could be collected to help provide context to a potential model for training and performance. Data collected includes session data (type, mileage, heart rate, timing) and where available, recovery data (heart rate variability, daily resting heart rate, sleep duration). Stored data will be psuedonomised with a translation key for each participant securely stored on my local machine. Given that the target particpant group would be of a reasonable size, it would take considerable effort for anyone with unauthorised access to figure out which generated key relates to which participant as many training programs are quite similar with similar sessions and timings. Participation for participants is entirely voluntary and they may ask at any time prior to the report due date for their data to be erased from any databases. There is no risk of personal data being revealed as a result of participating in this project. There is a potential for some GPS data to be collected when tracking certain on the water sessions. There is not a risk of personal information being revealed as a result of this GPS data as the locations where on the water rowing takes place are publically known.

References

[1]   T. Churchill, "Modelling athletic training and performance," Ph.D. dissertation, University of Canberra, Australian Capital Territory, Australia, 2014.

[2]   J. Mäestu, J. Jürimäe, and T. Jürimäe, "Monitoring of performance and training in rowing," *Sports Medicine*, vol. 35, no. 7, pp. 597–617, Jul. 2005. DOI: 10.2165/00007256-200535070-00005.

[3]   K. S. Seiler and G. Ø. Kjerland, "Quantifying training intensity distribution in elite endurance athletes: is there evidence for an "optimal" distribution?" *Scandinavian Journal of Medicine & Science in Sports*, vol. 16, no. 1, pp. 49–56, 2006, ISSN: 0905-7188. DOI: 10.1111/j.1600-0838.2004.00418.x.

[4]   A. Das, U. S. Kaniganti, S. J. Shenoy, P. Majumdar, and A. K. Syamal, "Monitoring training load, muscle damage, and body composition changes of elite indian rowers during a periodized training program," *Journal of Science in Sport and Exercise*, vol. 5, no. 4, pp. 348–359, Nov. 2022. DOI: 10.1007/s42978-022-00197-7.

[5]   E. W. Banister, T. W. Calvert, M. V. Savage, and T. Bach, "A systems model of the effects of training on physical performance," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-6, no. 2, pp. 94–102, Feb. 1976. DOI: 10.1109/tsmc.1976.5409179. [Online]. Available: https://doi.org/10.1109/tsmc.1976.5409179.

[6]   J. Edelmann-nusser, A. Hohmann, and B. Henneberg, "Modeling and prediction of competitive performance in swimming upon neural networks," *European Journal of Sport Science*, vol. 2, no. 2, pp. 1–10, Apr. 2002. DOI: 10.1080/17461390200072201. [Online]. Available: https://doi.org/10.1080/17461390200072201.

[7]   [Online]. Available: https://airow.univie.ac.at/.