Your First Step in Machine Learning: Exploring Public Data through the Lens of Flow

Bachelor/Master project "Information Systems"

Table of Contents

Prof. Dr. Mario Nadj
University of Duisburg-Essen
Faculty of Computer Science

Chair and Website
Information Systems and AI in Marketing
ai.ris.uni-due.de

Address Campus Essen
Universitätsstraße 9
D-45141 Essen

www.uni-due.de

1: Project Overview

Background: Machine learning enables the classification of flow states using physiological data from wearable devices. However, research progress is limited by data scarcity and the lack of accessible, publicly available datasets.

Project Goal: The goal of this project is to identify relevant datasets for flow state classification and assess their suitability for machine learning models. This evaluation aims to support future research on automated flow state detection.

Highlights: Gaining practical insights into applied machine learning problems; close collaboration with the supervising chair, providing guidance and support throughout the project to achieve its objectives; and the development of a complete machine learning pipeline suitable for inclusion in a GitHub portfolio.

2: Project Background

In recent years, machine learning (ML) has been increasingly applied across a wide range of domains. This rapid adoption is also reflected in market projections. The ML market is expected to grow at an annual growth rate of 31.72% between 2026 and 2031, reaching a projected volume of USD 445.27 billion by 2031 (Statista, 2026). These advances have opened up new possibilities in applications that rely on complex pattern detection, such as the recognition of the flow state through physiological data. However, a major challenge for data scientists when developing ML models, in particular, more advanced deep learning models, remains data scarcity (Alzubaidi, 2023).

Researchers focusing on flow state classification are likewise affected by this challenge. Flow, defined as the feeling of being completely absorbed into an activity, is regarded as a desirable state due to its association with improved performance and well-being (Peifer, 2020). Consequently, recent research approaches focus on the development of tools capable of detecting the flow state with wearable devices, such as ECG chest belts, to support individuals in entering or maintaining this state. Researchers have already accomplished important initial steps toward implementing such tools (Irshad et al., 2023; Rissler et al, 2020). For example, Rissler et al. (2020) developed a classification model using a random forest algorithm that was able to distinguish between high and low flow states with an accuracy of 70.6%, based on data from a field experiment in which software developers wore an ECG chest belt during their workday.

Despite these advances, the limited number of publicly available datasets hinders further progress in flow state classification. In addition, many potentially relevant datasets are not openly published due to data privacy regulations and are therefore only accessible upon request. Moreover, there is no standardized platform for publishing such datasets, which makes existing data difficult to find and access. To address these limitations, the objective of

this student project is to identify relevant datasets and to investigate their potential for developing ML models capable of classifying flow states.

3: Project Objective

The objective of this specific Bachelor/Master project is to classify flow states based on physiological data. To achieve this, students will identify publicly available datasets (e.g., BIRAFEE-dataset (Kutt et al., 2022) or PhySF (Irshad et al., 2023)), which are then used to build a flow classification model. To evaluate the effectiveness of the developed model, its performance will be compared with a rule-based classification approach (e.g., high heart rate indicating a flow state) in a small pilot study (n = 5).

The **objectives of this specific project** are the following:

- Review of the relevant literature on flow physiology
- Identification of publicly available flow datasets
- Preprocessing of data, e.g.:
    - Filtering and processing of physiological signals
    - Feature extraction
- Training and testing of different ML algorithms, e.g.:
    - Implementation of selected algorithms
    - Evaluation of algorithms via performance metrics
- Evaluation of model performance against a rule-based approach in a small pilot study
    - Conception and execution of the pilot study
    - Conducting the study
    - Analysis of the results

Depending on group size, adjustments to the project scope can be made. Adjustments will be announced at the kickoff and/or in subsequent milestone meetings.

Each finished artifact has to be properly documented to facilitate understanding and further development by students in follow-up projects.

4: Project Registration and further Information

All Bachelor- and Master students interested in this IS-project may apply via our online application form. The link to the application will be **available here** (Link hinterlegt) as of the start of the application period on **Wednesday, 25.03.2026.** The deadline for the application is **Wednesday, 08.04.2026, 23:59.**

For successful completion of the project, Master students receive 12 ECTS (corresponding to 360 working hours) and Bachelor students receive 6 ECTS (corresponding to 180 working hours)

- Master students will invest 18 hours per week into the project.
- Bachelor students will invest 9 hours per week into the project.

In addition to project-specific deliverables during the process, we also require some fixed deliverables from each group. On the 05.06.2026, the project team needs to submit a one-page management summary on the current state of the project. A template will be provided by the Chair. As final deliverables, the project team has to submit three items until 24.07.2026. These are (1) a 5-minute video presenting the results (we will provide an example video for orientation), (2) a scientific poster summarizing the results (whether the poster has to be presented will be determined through a centralized vote among the Information Systems Chairs), and (3) a project report along with the project code. The project report will contain the final results on 10-20 pages. We will provide a template for the final project report. The code needs to be accessible as Open Source via GitHub with appropriate documentation. We want to put up your posters (2) at the Chair's premises to highlight your achievements. The final video (1) and a link to your code (3) will be shown on the Chair page.

5: Project Schedule (Dates are to be announced)

Kickoff	Kickoff with accepted participants Clarify project scope and responsibilities
Milestone 1	Identification of publicly available flow datasets
Milestone 2	Development of a flow classification model
Milestone 3	Conception and execution of the pilot study
Milestone 4	Evaluation of the pilot study
Milestone 5	Submission of one-page management summary, project poster, report, code & 5-minute video illustrating the project

6: Grading

The final grade will be calculated as following based on the sum of the milestones:

- Milestone 1: 15%
- Milestone 2: 15%
- Milestone 3: 15%
- Milestone 4: 15%
- Milestone 5: 40%

7: Sources

- Alzubaidi, L., Bai, J., Al-Sabaawi, A. *et al.* A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *J Big Data* 10, 46 (2023). https://doi.org/10.1186/s40537-023-00727-2
- Irshad, M. T., Li, F., Nisar, M. A., Huang, X., Buss, M., Kloep, L., Peifer, C., Kotzsnik, B., Pollak, A., Pyszka, A., Flak O. & Grzegorzek, M. (2023). Wearable-based human flow experience recognition enhanced by transfer learning methods using emotion data. *Computers in Biology and Medicine*, *166*, 107489.
- Kutt, K., Drążyk, D., Żuchowska, L., Szelążek, M., Bobek, S., & Nalepa, G. J. (2022). BIRAFFE2, a multimodal dataset for emotion-based personalization in rich affective game environments. *Scientific Data*, *9*(1), 274.
- Peifer, C., Syrek, C., Ostwald, V., Schuh, E., & Antoni, C. H. (2020). Thieves of flow: how unfinished tasks at work are related to flow experience and wellbeing. *Journal of Happiness Studies*, *21*(5).
- Rissler, R., Nadj, M., Li, M. X., Loewe, N., Knierim, M. T., & Maedche, A. (2020). To be or not to be in flow at work: physiological classification of flow using machine learning. *IEEE transactions on affective computing*, *14*(1), 463-474.
- Statista. (2024). *Machine learning market size worldwide from 2017 to 2031*. Statista. https://www.statista.com/outlook/tmo/artificial-intelligence/machine-learning/worldwide?srsltid=AfmBOooAAsiihUXLuEqEjzji0XIevLk2YE267ZMvhUtAsk5tp3qx7fym

Recommendation:

- Knierim, M. T., Rissler, R., Dorner, V., Maedche, A., & Weinhardt, C. (2018). The psychophysiology of flow: A systematic review of peripheral nervous system features. *Information Systems and Neuroscience: Gmunden Retreat on NeuroIS 2017*, 109-120.