CSCE 3602/DSCI 3415– Fundamentals of Machine Learning

Project

(Due on: May 21, 2025 at mid-night)
(Submit on Canvas as one .zip file)

In this project, each team needs to pick one of the following ideas to work on.

Project 1: Assessing Neurological States from Physiological Signals

Neurological states including different types of stress (physical, emotional and cognitive) could be recognized using different types of physiological signals. In this project, the aim is to identify neurological states using non-brain physiological signals such as electrodermal activity, accelerometer data, heart rate and many others. This could simplify the process of detecting neurological states.

Dataset Source: https://physionet.org/content/noneeg/1.0.0/

Project 2: Sleep Stage Prediction from Heart Rate and Motion Data

Insomnia represents one of the most irritating challenges people might have given that humans ability to function depends primarily on the quality of their sleep. In this project, data mining methods will be developed to recognize different sleep stages using heart rate and motion data recorded using wearable sensors.

Dataset Source: https://physionet.org/content/sleep-accel/1.0.0/

Project 3: Classification of Neurodegenerative Diseases using Gait Analysis

Diagnosis of different neurodegenerative diseases represents one of the most challenging tasks in medicine. While brain scans and blood tests have been shown to help in such problem, other methods that involve analysis of moving patterns of patients have been proposed. In this project, data mining methods will be developed to analyze and classify gait information of three different neurodegenerative diseases; namely Parkinson's, Huntington's and Amyotrophic Lateral Sclerosis.

Dataset Source: https://physionet.org/content/gaitndd/1.0.0/

Project 4: Predicting Intensive Care Unit (ICU) Mortality

This project aims at predicting which patients of the ICU are at high risk of mortality versus patients who could be cured at the ICU. This can help hospitals manage resources in a better way to help patients at high risk.

Dataset Source: https://physionet.org/content/challenge-2012/1.0.0/

THE AMERICAN
UNIVERSITY IN CAIRO

CSCE 3602/DSCI 3415– Fundamentals of Machine Learning

Project

(Due on: May 21, 2025 at mid-night)
(Submit on Canvas as one .zip file)

Project 5: Predicting Football Games Result using In-game Events

This project aims at predicting the results of the football games based on in-game events. Different feature extraction and classification techniques will be examined in order to predict during the game how the game will end. This can help coaches change their in-game tactics to avoid losing a game.

Dataset Source: https://figshare.com/collections/Soccer_match_event_dataset/4415000/3

Project 6: Bank Marketing Data

The dataset represents direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to assess if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

Dataset Source (Requires Login): https://data.world/data-society/bank-marketing-data

Project 7: MetroPT-3 Dataset

From a metro train in an operational context, readings from pressure, temperature, motor current, and air intake valves were collected from a compressor's Air Production Unit (APU). This dataset reveals real predictive maintenance challenges encountered in the industry. It can be used for failure predictions, anomaly explanations, and other tasks.

Dataset Source: https://archive.ics.uci.edu/dataset/791/metropt+3+dataset

Project 8: E-commerce - Users of a C2C fashion store

The data was scraped from a successful online C2C fashion store with over 9M registered users. The store was first launched in Europe around 2009 then expanded worldwide. Visitors do not appear in this dataset. Only registered users are included. "Visitors" cannot purchase an article but can view the catalog.

Dataset source (Requires Login): https://data.world/jfreex/e-commerce-users-of-a-french-c2c-fashion-store

CSCE 3602/DSCI 3415– Fundamentals of Machine Learning

Project

(Due on: May 21, 2025 at mid-night)
(Submit on Canvas as one .zip file)

Deliverables:

- Your Python code (.py or Jupyter notebook) to be submitted on Canvas on **May 21 at mid-night**

- A report (to be submitted on Canvas on **May 21 at mid-night**) that includes the following:
  - Introduction: A description of the problem the selected project addresses
  - Methods: A description of the approach followed to analyze the dataset
  - Results: A description of the achieved results
  - Conclusions

- Submission of the above items should be done as one .zip file by the deadline

- A presentation (to be presented in class on **May 22**). The presentation should follow the same flow as the submitted report.

Guidelines:

- This is a group project. A maximum of 3 students per group is allowed.

- Each team must send an e-mail by **Tuesday, April 29 at mid-night** specifying the members of the team and the name of the selected project.

- **Changing teams or changing the dataset of the project will not be allowed.**

- Project grading will be as follows (out of 20):
  - 5 points on the code or project submitted
  - 5 points on the submitted report
  - 5 points on the presentation
  - 5 points on the depth of the analysis and the obtained insights