COS 472&572

Group Project
(35% of the total score)

Instructor: Yuqi Song

Fall 2024

University of Southern Maine

Group Project

- Each group includes 1-3 person, more people on the team, more effort required

- The project includes 2 main parts:
  - Project proposal presentation in class (after midterm, 5-10 minutes per group) →10%

  - Project final presentation (25%, last week, 20-25 minutes), submit project report (4+ pages) and project source code→ 25%

Topics for your group project

1. Analysis and predict turnover intention for restaurant employees
2. Predict material property
3. Kaggle active competition: Child Mind Institute — Problematic Internet Use

Project 1: Turnover intention prediction

- The primary goal of this data mining task is to analyze historical employee data to identify patterns and factors that contribute to turnover intention among restaurant employees.

- By understanding these factors, management can implement strategies to improve employee retention, enhance job satisfaction, and optimize operational efficiency.

Project 1: Turnover intention prediction

- Main tasks:
  - Data preprocessing, data analysis and data visualization
  - Extracting and selecting features that are most relevant to the properties being predicted
  - Design the machine learning - based framework for predicting employee retention
  - Evaluate and improve the model’s performance

Project 2: Material property prediction

- In the field of materials science and material informatics, predicting the properties of materials before they are synthesized or subjected to various processing conditions can significantly accelerate innovation and improve the efficiency of material design.

- This project aims to leverage advanced data analytics and machine learning techniques to predict key material properties, thus reducing the need for costly and time-consuming experiments.

Project 2: Material property prediction

- Main tasks:

  - **Data Collection and Preparation**: Gather comprehensive datasets from existing materials databases.
  - **Feature Engineering**: Develop a framework for extracting and selecting features that are most relevant to the properties being predicted.
  - **Model Development**: Utilize and compare several machine learning models to find the best predictor for material properties.
  - **Validation and Optimization**: Implement rigorous model validation techniques to ensure accuracy and reliability.

Project 3: Child Mind Institute

- Child Mind Institute — Problematic Internet Use

- Predict the level of problematic internet usage exhibited by children and adolescents, based on their physical activity.

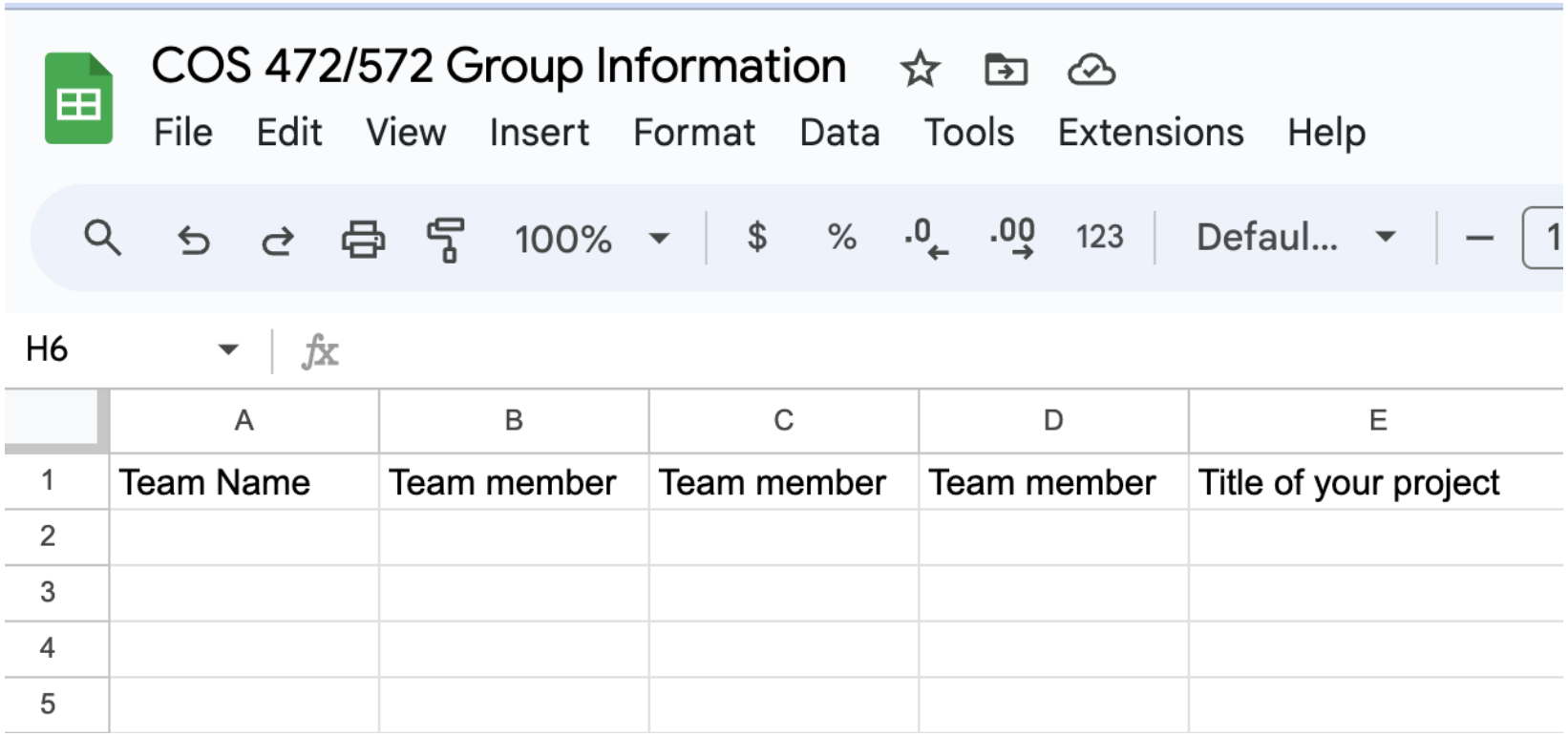
- The goal of this competition is to develop a predictive model that analyzes children's physical activity and fitness data to identify early signs of problematic internet use. Identifying these patterns can help trigger interventions to encourage healthier digital habits.

Project 3: Child Mind Institute (dataset)

- The Healthy Brain Network (HBN) dataset is a clinical sample of about five-thousand 5-22 year-olds who have undergone both clinical and research screenings.
- The objective of the HBN study is to find biological markers that will improve the diagnosis and treatment of mental health and learning disorders from an objective biological perspective.
- Two elements of this study are being used for this competition: physical activity data (wrist-worn accelerometer data, fitness assessments and questionnaires) and internet usage behavior data.
- The goal of this competition is to predict from this data a participant's Severity Impairment Index (sii), a standard measure of problematic internet use.

Fill in your information on Google sheet

- COS 472/572 Group Information
- Reserve your proposal presentation date: Oct 22/Oct 24/Oct 29/Oct 31



	A	B	C	D	E
1	Team Name	Team member	Team member	Team member	Title of your project
2					
3					
4					
5					

Project proposal presentation

- Introduction
  - Briefly introduce the project topic, its significance, and why it's worth your work and effort.
  - Provide an overview of the problem or question the project aims to address.

- Data
  - Briefly describe your dataset and the data preprocessing involved

Project proposal presentation

- Methodology
  - Describe the approach, methods, and techniques you intend to use for conducting the project.
- Evaluation
  - How will you evaluate your algorithm’s performance
- Division of your labor
  - If your team has multiple members, how do you allocate tasks to each individual?

Teamwork Tools

- GitHub
  - A version control platform that allows collaborative coding, tracking changes, and managing projects effectively.

- Overleaf
  - It is a collaborative editing platform that allows multiple users to work on a document simultaneously, it also provides real-time previewing and easy sharing of LaTeX documents.
  - Proposal, paper, report, slides, resume, recommendation letter, etc.

Good luck, and enjoy the learning and collaboration experience!