Meeting 1 Agenda

Introduction

Andy Wang

andywang0321@gmail.com - Fourth year - Data Theory, minor Data science engineering - I come from the math/stats side of data science, though I am quite comfortable with ML programming - Languages: Python, SQL, R

Tiffany Li

kehan1230@gmail.com - Fourth year - Financial Actuarial Math major and data science engineering minor - My skills in programming are using Python to write the report by analyzing and visualizing.

Matthew Liu

- Fourth year
- Math of Comp, minor Data science engineering
- working at generative AI startup in SF
- generative AI lab
- deep learning classes
- AWS, S3, Spark, Redis

Anvesha Dutta

dutta.anvesha06@gmail.com - Fourth year - Data Theory, minor Data science engineering - Python, Streamlit, Tableau, scikitlearn, Tensorflow

Sam Hopkins

samthehopkin@gmail.com - Fourth year - CS Major - Python, ml packages like PyTorch, etc.

Adithi Ramesh

adithi.ramesh02@gmail.com - Fourth year - CS major, taken most of the ML and Data Science electives offered - basic understanding of the diff models and have worked with - quite confident with python, pandas, matplotlib, etc.

Project goal

- get A on assignment?

  General group preference

- get onto leaderboard?

Project choice

Incorrect assignment detection (IND)

Given the paper assignments of each author and paper metadata, the goal is to detect paper assignment errors for each author.

- simplicity
- **general group preference**

Academic question answering (AQA)

Given professional questions and a pool of candidate papers, the objective is to retrieve the most relevant papers to answer these questions.

- seems more involved

Paper source tracing (PST)

Given the full texts of each paper, the goal is to automatically trace the most significant references that have inspired a given paper.

- challenging
- requires very powerful transformer model capable of processing scientific papers
- "inspiration" is very vague

Dummy submission (19th of April)

- All teams need to submit a dummy submission to the contest portal for their chosen task
- Deadline: April 19, 11:59PM

Steps:

- Run the baseline code provided for your chosen task
- Prepare the dummy submission file according to the specified format
- Submit the dummy file to the contest portal
- Verify that the submission was successful and meets the requirements

GitHub / Google Drive setup

- Matthew will create group shared repo