Team name: BEAMING
Team members: Dilay Ozkan, Nghia Nguyen, Benjamin Torres, Malavika Mampally, Egemen Elver
Date: 03/07/2024

Team roles for this report (write down name):

Facilitator(s): Dilay Ozkan, Egemen Elver

Recorder(s): Malavika Mampally

Deliverer(s): Nghia Nguyen

Planner(s): Benjamin Torres

Team Contact: Dilay Ozkan (dilay@unc.edu)

- **Describe briefly what the main goal of your team is (so the peer reviewer has some context). E.g. we are working on image classification for blah de blah. Our goal is blah de blah etc. In the initial part of the semester before your proposal it is ok to put down "we are still coming up with ideas on team project".**

Initially, our plan involved working with an image dataset. However, we reconsidered and opted for a more manageable yet interesting approach that allows everyone to contribute equally to the project. Our primary objective was to put an end to the search and finalize the data.

I. **What was done during the report period regarding the project:**

Since we could not come up with a dataset last week, we gave one last shot to identify a dataset that everyone is satisfied with and is happy to work with, since we want to make sure that it is suitable for each member's portfolio.

First Meeting (Date: 03/05/2024) All members were present

Agenda: Discuss and look at each dataset that the team members came up with. Understand the purpose of the datasets and various methods that could be used to analyze them, identify the pros and cons of working with that dataset, its application in the real world and the motivation of our final project.

Each of us took turns showing datasets that we found over the weekend.

Some of them are listed below:

- *Credit card fraud detection (eliminated due to insufficient data)*
- *Web page phishing detection*
- *Heart disease classification*
- *Eye disease classification*
- *Diabetes classification*
- *Chest Xray dataset*
- *Cardiovascular disease classification*
- *Identification of contradiction in texts*
- *Traffic signs image classification*

Due to numerous options and no consensus on one, we decided to rank each dataset from best (1) to worst (8) and then choose the one with minimum sum.

Web page Phishing dataset turned out to have the least sum hence we decided to go ahead with it.

Next, we delegated the work to each member so that we could complete the proposal by Friday of that week.

Dilay and Nghia were assigned to write the goals and motivation part of the proposal, Malavika and Egemen worked on data description, and Benjamin used Python to perform exploratory data analysis on the data.

Malavika took down the minutes of the meeting for reference during the biweekly report. We did not meet again but communicated on a group chat regarding the progress of the proposal.

II. What were obstacles faced if any in working on the project?

- We went back and forth on the issue of whether image datasets could be complex to analyze or not. Ultimately, we dropped the idea.
- There was a bit of uncertainty in our meetings since we had not yet finalized a dataset.
- Midterms and other commitments led to postponement of the weekly meeting on Thursday but we made it up the following Tuesday.

III. What is the plan for the next reporting period including what each team member is planning to work on. Describe goals and potential timelines.

1. We came up with an initial few techniques of analysis that we think are appropriate for such a dataset. Each of us will spend the week testing out those techniques and reconvene to discuss our results.
2. To make the most of the course modules, we selected the recently taught modeling concepts, and each picked one to explore further, like logistic regression and decision trees (Dilay and Nghia), K-nearest neighbors (Malavika), and support vector machines (Egemen) to model the data and come up with an accurate classification solution. We also intend to explore the random forest (Benjamin) as an alternative.
3. To avoid piling up work during the preparation of the final project report, the team collectively decided to update the contents as and when they were done, such as writing down the EDA and literature review as well as the methodology selected. This way, the only thing left to do before presentations would be to write down the findings and conclusion.