Team name: BEAMING
Team members: Dilay Ozkan, Nghia Nguyen, Benjamin Torres, Malavika Mampally, Egemen Elver
Date: 02/23/2024

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".**

Each member tried to research topics that interested them and in the end, we unanimously decided to work on some sort of image dataset, since it was something that none of us had worked with. The first goal is to decide on a concrete dataset that everyone is fine with. The next step will be to clearly state the aim of our project, such that it adds value to whichever industry dataset we choose.

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

The first week was spent individually gathering information on potential datasets that we could work on and then proposing that topic to everyone during the team meeting.

First Meeting (Date: 02/09/2024) 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.

Benjamin came up with an age-predicting model using image regression. We tried to assess the complexity of the technique and the industries that might benefit from it. Since age is an easily available information, we moved on to another dataset.

We next looked at an image-classification model that could predict whether a person is a smoker or a non-smoker using already labeled images of people which Malavika discovered. Egemen pointed out that although health insurance companies could benefit from this model, the accuracy will have to be optimum to prevent repercussions on incorrect classification.

Dilay had two more ideas on image regression.

1. *Prediction of book prices using Book Cover Pictures:*

Although the concept of predicting book prices based on cover design is intriguing, in the era of online bookstores, customers have ready access to book prices, and factors like authorship, genre, and reviews may carry more weight in purchasing decisions. Hence this idea was also turned down.

2. *Housing Prices Based on House Pictures:*

This idea also did not meet the expectations we had due to the broader context of house hunting, recognizing that the exterior appearance holds limited influence compared to factors like neighborhood, accessibility, and interior features, reflects a keen understanding of end-users priorities.

Everyone was equally fascinated with studying images, hence we decided to give it another week before going ahead and finalizing a dataset.

Nghia designed and created the team website and name. Finally, Malavika filled out the team registration form due that night.

Second Meeting (Date: 02/23/2024) All members were present

Agenda: Check on the progress made by everyone with respect to finding interesting datasets that could be effectively analyzed using machine learning techniques and provide a useful outcome.

Dilay ideated the weekly meetings so that we keep everyone in the loop with our work and would also help in creating a collaborative working environment where everyone’s input is heard and valued.

Nghia found a lung X-ray dataset that could be used to classify whether the person has pneumonia, COVID-19, or tuberculosis. We discussed the methodologies and the final conclusions of the study. Although it was limited to the healthcare industry, it would have a significant role in detecting critical diseases at earlier stages.

Benjamin found a new angle to study the image regression dataset that predicts the age of a person. Instead of trying to get a perfect model, we could try to come up with ways to make the model fail such that we can understand situations where this model will definitely not work.

Egemen and Malavika had a few ideas, too which we decided to go through later on, at our own convenient time since we had other commitments right after the hour-long meeting. We briefly reviewed points to add to the biweekly report and adjourned the meeting.

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

- Some team members were hesitant to use complex techniques like CNN since no one had experience working with it. Still, we ended up agreeing to learn it due to the existence of similar modeling techniques on the internet.
- Time issues: Our group consists mostly of PhD students who are either instructors or TA/RA, which made it difficult to schedule meetings when every member was available.
- Health issue: A few members were not keeping well, so we had to switch to online meetings

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. The weekly meeting is scheduled to be on Thursdays at 7 PM.
2. Our goal is to try to finish the project proposal by the end of two weeks.
3. Benjamin and Malavika are responsible for finalizing the dataset, checking the specifications of the dataset, whether the computational power of our PCs is enough, and so on.
4. Egemen will validate the dataset and identify metrics or techniques that can be used.
5. Nghia and Dilay will be responsible for writing up the proposal, defining the objective of our study, and the steps that we will follow to analyze the data. Finally, Nghia will update the same on our website.