Saood Usmani
Initial Proposal

Introduction and topic and/or problem statement – A short introduction and summary of the goals of the project

- Goal: Multi label classification of finance transactions with natural language processing
- I volunteer at a non profit, and one of the things that I help with is classifying miscellaneous finance transactions. This can be quite tedious, as there can be hundreds of transactions per month. I aim to build a machine learning model that can automatically classify these transactions. I believe the model will perform fairly well as the way we categorize these transactions is very standard, and something the model should be able to pick up on.

Data sources that will be used – A reference to any datasets utilized in the project

- I will use all of the data for the past miscellaneous transactions that are categorized so far. This is around 3000+ data points, which I believe will be plenty for my model as the patterns are pretty basic. Each data point also has around ten different fields to train on, such as amount, date, donor name, memo(important), etc. One consideration though is that this data sheet is very sensitive as it contains lots of sensitive donor information, so I couldn't share it publicly online, but I can privately train on it.

List of high-level methods, techniques and/or technologies that you are considering using.

- I am considering using some sort of transformer architecture, as the primary field that determines where the miscellaneous transaction goes is the memo field, where the user that is donating the money can write a short description of what it is for. I would need to use natural language processing for this field, hence a transformer.

Products to be delivered – what are the primary deliverables for the project? This is what we will be grading

- **Juypter notebook with test accuracy reported**
- I would love to get a website up where I and other members of the finance team could actively upload new unclassified transactions and run the ML model to classify them; however, I am not sure how feasible this is as I don’t know how to upload and run a ML model to a website.
- I was also thinking of maybe comparing my model to another model(GPT 3 maybe?) and seeing the performance differences