**Mulualem Asmare:** MSDS 696 Project proposal

1. **Title of the project:** Book Recommender System Using Neural Network
2. **High level description of the project: what question or problem are you addressing?**

Currently, we live in an era where different aspects of our lives are dominated by artificial intelligence (AI). Businesses that implement AI in their decision-making processes experience improved results. One type of artificial intelligence application is recommendation systems.

The recommendation system provides personalized recommendations for users by analyzing patterns in their behavior as well as patterns in similar users. Its implication in business results in improved customer retention and engagement, which increases product sales. For instance, in 2011-2012, when Amazon first deployed its collaborative filtering-based recommender system, it recorded a 29% increase in sales by the second fiscal quarter of 2012. Today, recommendation engines are used in various sectors such as e-commerce, movie streaming services, social media services, health, and fitness, and more. The recommendation system market is predicted to increase from $2.12 billion in 2020 to $15.3 billion US dollars by 2026.Recognizing the benefits of recommendation systems, this project aims to develop a collaborative book recommendation system to provide personalized book suggestions based on users' interaction history, as well as users with similar interests. This keeps users engaged, maximizes retention and user satisfaction, and increases revenue.

**References:**

Ahramovich, A. (n.d.). Recommendation Systems and Machine Learning. Recommendation Systems and Machine Learning. <https://www.itransition.com/machine-learning/recommendation-systems>

1. **What types of data science task is it?**

* Exploratory Data Analysis
* Data visualization
* Feature engineering
* Data transformation
* Recommendations Model development

1. **Data: Brief description of data. How big do you expect the data will be? Is amount of your data too big or too small? If you're web-scraping or collecting data, how long do you expect to collect the data?**

I plan to use datasets obtained from Kaggle and GitHub repositories. The Kaggle dataset will be formed by merging three data frames using their common key. The dataset from GitHub will be formed by merging two data frames using a common key. Once each dataset is merged, I will combine the two datasets to create one large dataset. The merged data from Kaggle contains 3.04 GB of data and includes information about 2 million book reviews for 212,404 unique books. The other merged dataset from GitHub contains 6 million ratings for 10,000 of the most popular books.

1. **How will you analyze the data? What machine learning methods do you plan to use, and/or what business intelligence aspect do you plan on incorporating?**

To analyze the data, I will use Jupyter Notebook with Python 3 for interacting with and analyzing the data. Depending on the computational load, I may utilize Google Colab. I will employ data understanding to comprehend each dataset and to combine datasets. Exploratory Data Analysis (EDA) will be performed to gain insights about the datasets. Data preprocessing will be applied for data cleaning, handling missing values, feature engineering, and normalization of numerical rating features (if the data that will be combined have different rating scales). Additionally, data encoding will be used to transform categorical variables into numerical representations for machine learning. I plan to implement neural network machine learning to develop the model. I will utilize cosine similarities to test the model and obtain similar books for specified users. Additionally, I will employ appropriate evaluation metrics based on the data, considering factors such as sparsity, rating scale, and outliers, which may affect the choice of metrics.

1. **Describe any anticipated difficulties and problems. Discuss how you may overcome the problems**.

I may encounter several anticipated difficulties and problems during the project, such as system and library compatibility issues, error codes, code issues, and various unexpected challenges. I plan to resolve these issues by utilizing resources such as Assignment I did in my past class, Google, reading articles, watching related videos, and visiting GitHub repositories, Stack Overflow, Kaggle, and other relevant platforms.

1. **Suggest a timeline for the project.  This should be a weekly breakdown of what you plan on doing each week.**

**Week1:** Project proposal

**Week2:** Project proposal and Data collection

**Week3:** Data Cleaning and Exploration

**Week4:** Data preparation

**Week5:** Data preparation

**Week6:** Model development

**Week7:** Model evaluation and Testing

**Week8:** Finish up project, conclude the observation and findings.