Team 1

1. Project Name: Recommender System for Movies

2. Team members:

   Changrong Chen

   Haobo Zhao

3. Summary

   For this project, our team plan to implement a recommender system for movies using Scala and Apache Spark.

   A recommender system is a program that seeks to predict user preferences and recommend items that are most possibly favored by the end user. It is used everywhere in our daily life, examples including YouTube, Amazon, etc.

   For the data source, we plan to use the dataset from Kaggle which contains credible dataset.

4. Use Cases

   1) Based on the preferences and searching histories of users, categories, rating, etc., provide potential movies that might be most appealing to the clients.

   2) Calibrate the services to the preferences of users through continuous usage of the recommender system by users.

5. Acceptance criteria

Our project will use several ways to evaluate the quality of this movie recommender system:

1) Recall: what proportion of movies that a user likes were actually recommended

2) Precision: Out of all the recommended movies, how many did the user actually like.

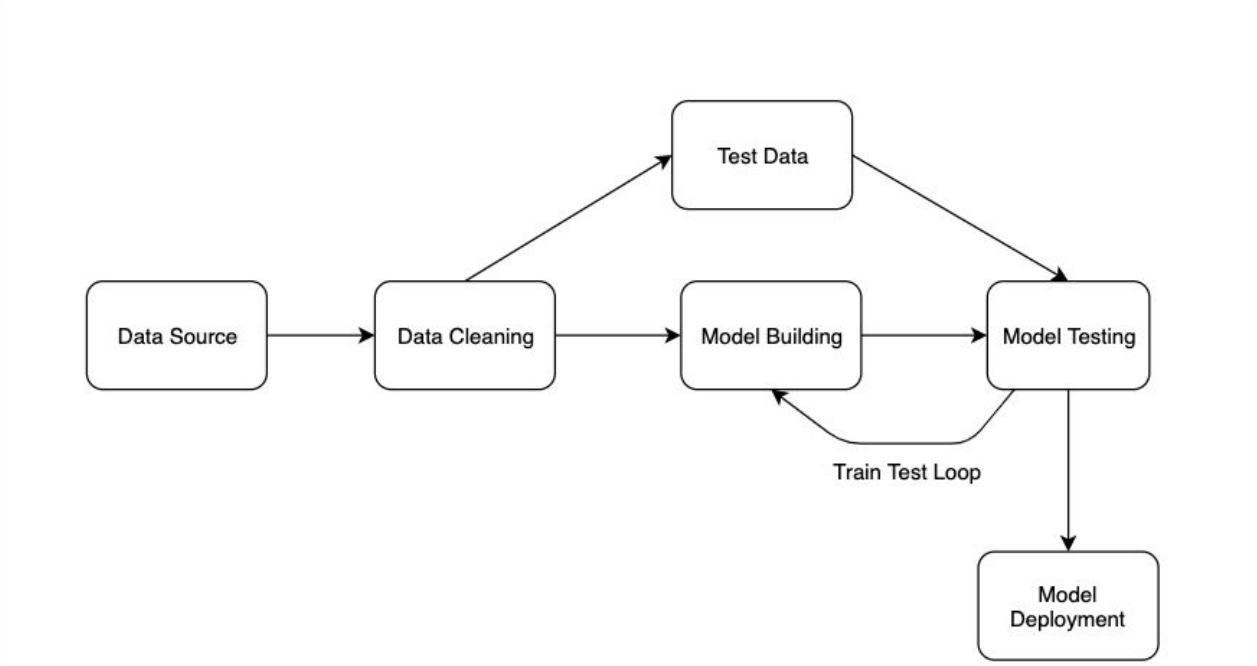
3) Reciprocal Rank: If the system recommends 3 movies, A, B and C, to a user, the user only likes the movie C. then in this case, the rank of movie C is 3, thus the reciprocal rank is 1/ 3. Larger the mean reciprocal rank is, better recommendations will be.

4) MAP at K (Mean Average Precision at cutoff K): the precision calculated by considering only the subset of the recommendations from rank 1 through K

6. Goals

- Practice Scala and Apache Spark, and enhance general programming skills
- Build a decent recommender system specific for movies
- Build teamwork spirit through collaborations
- Achieve the accredited criteria of course CSYE 7200

7. Project Structure



8. What we will do

    - Setting up development environments for Scala and Spark
    - Cleaning data and Ingesting
    - Building and training model
    - Model Testing
    - Making prediction
    - User Interface

9. Milestones

    - Week 1 (November 8 2019 - November 14 2019):

Setup the development of Scala and Spark for the project and run some small dataset for prototyping

- Week 2 (November 15 2019 - November 21 2019):

  Early-state implementation (data cleaning, ingestion, etc.)

- Week 3(November 22 2019 - November 28 2019):

  Model building, training and testing

- Week 4(November 29 2019 - December 6 2019):

  More testing, deployment and wrapping up the project