# CSYE7200
# Book Recommendation System With Apache Spark and Scala

By

**Team 3**

Maria Gloria (002667315)

Sumanth Anumula (002809265)

# Use Case

- Administrator loads user's historical data with which model is trained.
- User should be able to invoke API by providing User Id and get Recommended Books based on User's historical behavior(from the trained model).

# Data Source

https://www.kaggle.com/datasets/saurabhbagchi/books-dataset/

| File Name | Row Count | Columns |
|-----------|-----------|---------|
| Books.csv | 271379 | ISBN,Book-Title,Book-Author,Year-Of-Publication,Publisher,Image-URL-S,Image-URL-M,Image-URL-L |
| Ratings.csv | 1149780 | User-ID, ISBN, Book-Rating |
| Users.csv | 278859 | User-ID, Location, Age |

# Workflow

| Data-Ingestion | → | Data Pre-processing | → | Train Model Using Data | → | Model Evaluation and Tuning | → | API Development |

# Tech Stack

- Scala Programming Language 2.13.7

-  Apache Spark , Core, SQL, MLIB 3.5.0 for Data Ingestion, Data Analysis, Data Cleaning, Training ML Model.

- Akka Http toolkit 10.6.0 for API Development

- Git Repo: https://github.com/Maria-Obono/Book-Recommendation-System

# Goal

- Train Recommendation Algorithms(Clustering, Matrix Factorization) and Tune the models.

- Develop API to feed input to the model and get recommended Books.

- Improve the user experience by providing personalized and relevant book recommendations to users.

# Acceptance criteria

- RMSE values to be around 0.7 (Lower the better)

- Ensure that the system provides recommendations within a specified time frame (e.g., under 500 milliseconds).

# Milestones

- 11/9 - 11/12  Exploratory Data Analysis and feature Extraction
- 11/13 - 11/18 Evaluate Recommendation Algorithms.
- 11/19 - 11/22 Error Evaluation
- 12/23 - 11/29 API Development, Unit Tests
- 11/30 - 12/07 Unit Tests, Readme and final presentation.

# Thank You