CSYE7200
Book Recommendation System
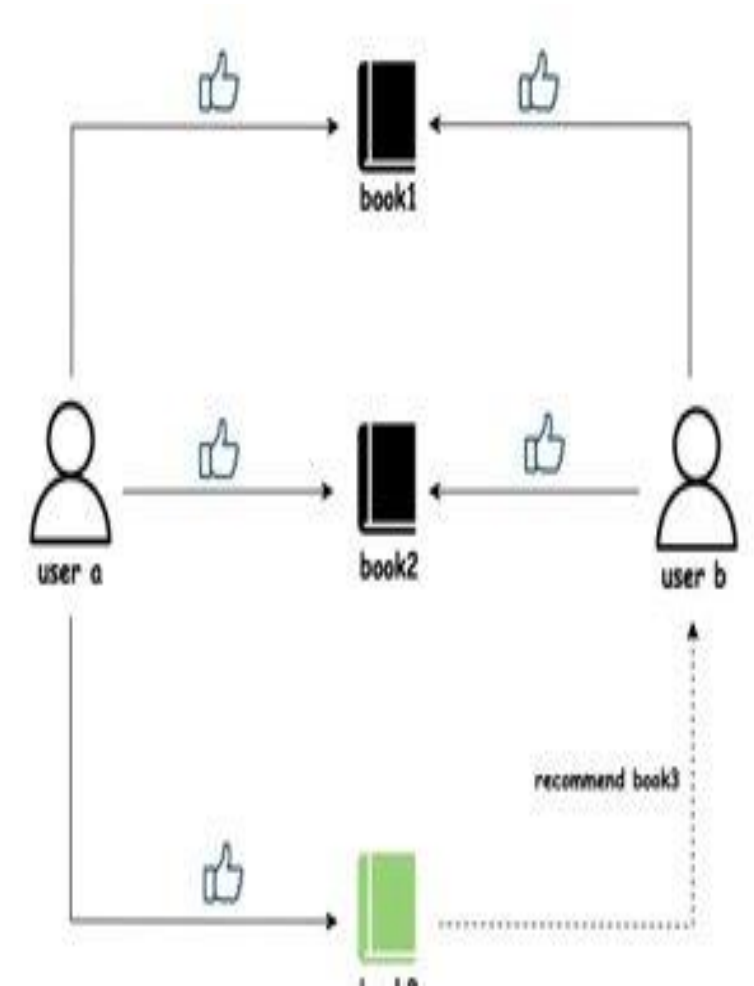# With Apache Spark and Scala

By

Team 3

Maria Gloria (002667315)

Sumanth Anumula (002809265)

Introduction

- A recommendation system is one of the top applications of data science. It uses Big Data to suggest or recommend additional products to consumers.

- A book recommendation system is a type of recommendation system where we have to recommend similar books to the reader based on his interest.

- In this project, we use Alternating Least Squares(ALS) Collaborative Filtering algorithm to build the recommendation model

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

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, Rating
Users.csv	278859	User-ID, Location, Age

BUILD MODEL

```scala
// Split the data into training and test sets (for example, 80% training and 20% test)
val Array(trainingData, testData) = jdf.randomSplit(Array(0.8, 0.2))

// build recommendation model with als algorithm
val als = new ALS()
  .setMaxIter(15)
  .setRegParam(0.01)
  .setUserCol("userId")
  .setItemCol("isbn")
  .setRatingCol("rating")
val alsModel = als.fit(trainingData)

// evaluate the als model
// compute root mean square error(rmse) with test data for evaluation
// set cold start strategy to 'drop' to ensure we don't get NaN evaluation metrics
alsModel.setColdStartStrategy("drop")
val predictions = alsModel.transform(testData)

val evaluator = new RegressionEvaluator()
  .setMetricName("rmse")
  .setLabelCol("rating")
  .setPredictionCol("prediction")
val rmse = evaluator.evaluate(predictions)
println(s"root mean square error $rmse")
```

Workflow

Tech Stack

Scala Programming Language 2.12.18

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	Train Recommendation Algorithms(Clustering, Matrix Factorization) and Tune the models.
Develop	Develop API to feed input to the model and get recommended Books.
Improve	Improve the user experience by providing personalized and relevant book recommendations to users.

Acceptance criteria

- RMSE Value (Root Mean Square Error): Despite an RMSE value of 5.5, the system demonstrates a decent level of accuracy in predicting user preferences and generating book recommendations.

- Our system's current recommendation response time of around 8 seconds deviates from the original target in 500 milliseconds. However, while not meeting the initial goal, the system ensures a satisfactory user experience and adheres to functional requirements.

```
root mean square error 5.511924543561396
Server online at http://localhost:8080/
Press RETURN to stop...
```

Results: Top 10 Book Recommendations for a User

Top 10 Book Recommendations for User 11944

User ID	Recommended Book	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M
11944	340666269	Dog Who Loved Too Much Tales Treatments	Dr Nicholas Dodman	0	Trafalgar Square	http://images.amazon.com/images/P/0340666269.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0340666269.01.THUMB
11944	553282182	Alicia: My Story	Alicia Appleman-Jurman	1990	Bantam Books	http://images.amazon.com/images/P/0553282182.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0553282182.01.THUMB
11944	590445774	Encyclopedia Brown Sets the Pace (Encyclopedia Brown (Paperback))	Donald J. Sobol	1991	Scholastic	http://images.amazon.com/images/P/0590445774.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0590445774.01.THUMB
11944	671500139	MARK KISTLER'S IMAGINATION STATION : LEARN HOW TO DRAW IN 3-D WITH PUBLIC TELEVISION'S FAVORITE DRAWING TEACHER	Mark Kistler	1994	Fireside	http://images.amazon.com/images/P/0671500139.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0671500139.01.THUMB
11944	671656945	Mark Kistler'S Draw Squad	Mark Kistler	1988	Fireside	http://images.amazon.com/images/P/0671656945.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0671656945.01.THUMB
11944	932379982	The Women Who Hate Me: Poetry 1980-1990	Dorothy Allison	1991	Firebrand Books	http://images.amazon.com/images/P/0932379982.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0932379982.01.THUMB
11944	671670050	The BRIDE	Julie Garwood	1989	Pocket	http://images.amazon.com/images/P/0671670050.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0671670050.01.THUMB

Top 10 Book Recommendations for User 276746

User ID	Recommended Book	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	
276746	886777844	Summon the Keeper	Tanya Huff	1998	Daw Books	http://images.amazon.com/images/P/0886777844.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886777844.01.THUMBZZZ.jpg	http:/
276746	886778344	Fearful Symmetries: The Return of Nohar Rajasthan	S. Andrew Swann	1999	Daw Books	http://images.amazon.com/images/P/0886778344.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886778344.01.THUMBZZZ.jpg	http:/
276746	886777704	Prince of Dogs (Crown of Stars/Kate Elliott, Vol 2)	Kate Elliott	1998	New Amer Library	http://images.amazon.com/images/P/0886777704.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886777704.01.THUMBZZZ.jpg	http:/
276746	886778034	Owlsight (Valdemar: Darian's Tale, Book 2)	Mercedes Lackey	1999	Daw Books	http://images.amazon.com/images/P/0886778034.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886778034.01.THUMBZZZ.jpg	http:/
276746	886778204	Battle Magic	Martin Harry Greenberg	1998	Daw Books	http://images.amazon.com/images/P/0886778204.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886778204.01.THUMBZZZ.jpg	http:/
276746	886777984	This Alien Shore (Daw Books Collectors, No. 1096)	C. S. Friedman	1998	Daw Books	http://images.amazon.com/images/P/0886777984.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886777984.01.THUMBZZZ.jpg	http:/
276746	886778174	The Reel Stuff	Martin H. Greenberg	1998	Daw Books	http://images.amazon.com/images/P/0886778174.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886778174.01.THUMBZZZ.jpg	http:/
276746	886777054	Exile's Song (Daw Book Collectors,	Marion Zimmer Bradley	1996	Daw Books	http://images.amazon.com/images/P/0886777054.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0886777054.01.THUMBZZZ.jpg	http:/

- Use http://localhost:8080/recommendations/user/userId
- Replace "userId" with any User ID number in the dataset when testing and the results will be generated

Results: Top 10 Highly Rated Books for a User

Top 10 Highly Rated Books for User 11944

User ID	Recommended Book	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	
11944	151002177	Animal Farm: A Fairy Story	George Orwell	1996	Harcourt	http://images.amazon.com/images/P/0151002177.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0151002177.01.THUMBZZZ.jpg	http://ima
11944	618002243	The Return of the King (The Lord of The Rings, Part 3)	J. R. R. Tolkien	1999	Houghton Mifflin Company	http://images.amazon.com/images/P/0618002243.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0618002243.01.THUMBZZZ.jpg	http://ima
11944	618002219	The Hobbit: or There and Back Again	J.R.R. Tolkien	1999	Houghton Mifflin Company	http://images.amazon.com/images/P/0618002219.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0618002219.01.THUMBZZZ.jpg	http://ima
11944	618002235	The Two Towers (The Lord of the Rings, Part 2)	J. R. R. Tolkien	1999	Houghton Mifflin Company	http://images.amazon.com/images/P/0618002235.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0618002235.01.THUMBZZZ.jpg	http://ima
11944	618002227	The Fellowship of the Ring (The Lord of the Rings, Part 1)	J. R. R. Tolkien	1999	Houghton Mifflin Company	http://images.amazon.com/images/P/0618002227.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0618002227.01.THUMBZZZ.jpg	http://ima
11944	1047973	Brave New World	Aldous Huxley	1999	Trafalgar Square Publishing	http://images.amazon.com/images/P/0001047973.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0001047973.01.THUMBZZZ.jpg	http://ima

Top 10 Highly Rated Books for User 276984

User ID	Recommended Book	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	
276984	590426702	Karen's Haircut (Baby-Sitters Little Sister (Paperback))	Ann M. Martin	1990	Scholastic	http://images.amazon.com/images/P/0590426702.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0590426702.01.THUMBZZZ.jpg	http://i
276984	590442996	Karen's Worst Day (Baby-Sitters Little Sister, 3)	Ann M. Martin	1990	Scholastic	http://images.amazon.com/images/P/0590442996.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0590442996.01.THUMBZZZ.jpg	http://i
276984	590442589	Karen's School Picture (Baby-Sitters Little Sister, 5)	Ann M. Martin	1993	Scholastic	http://images.amazon.com/images/P/0590442589.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0590442589.01.THUMBZZZ.jpg	http://i
276984	590448234	Karen's Carnival (Baby-Sitters Little Sister (Paperback))	Ann M. Martin	1995	Scholastic	http://images.amazon.com/images/P/0590448234.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0590448234.01.THUMBZZZ.jpg	http://i
276984	590259970	Karen's Lemonade Stand (Baby-Sitters Little Sister, 64)	Ann M. Martin	1995	Scholastic	http://images.amazon.com/images/P/0590259970.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0590259970.01.THUMBZZZ.jpg	http://i
276984	590456504	Karen's Snow Day (Baby-	Ann M. Martin	1993	Scholastic	http://images.amazon.com/images/P/0590456504.01.THUMBZZZ.jpg	http://images.amazon.com/images/P/0590456504.01.THUMBZZZ.jpg	http://i

- Use path http://localhost:8080/recommendations/userchoice/userId
- Replace "userId" with any User ID number in the dataset when testing and the results will be generated

Conclusion

We successfully designed and implemented a book recommendation system using collaborative filtering with Alternating Least Squares (ALS) in Apache Spark. Leveraging user ratings and book information datasets, we created a collaborative filtering model that predicted book recommendations for users based on their preferences.

The system provided personalized recommendations for individual users, enhancing their browsing experience and potentially boosting user engagement and satisfaction.

By employing ALS in Spark's machine learning library, we efficiently handled the recommendation task, utilizing user-item interactions to generate accurate predictions. The system effectively identified patterns in user preferences and made reliable book recommendations, contributing to a more tailored and enriched user experience.

Thank You