

DATA 643 Project 5 | Implementing a Recommender System on Spark

The goal of this project is give you practice beginning to work with a distributed recommender system. It is sufficient for this assignment to build out your application on a single node.

Adapt one of your recommendation systems to work with Apache Spark and compare the performance with your previous iteration. Consider the efficiency of the system and the added complexity of using Spark. You may complete the assignment using PySpark (Python), SparkR (R), sparklyr (R), or Scala.

Please include in your conclusion: For your given recommender system's data, algorithm(s), and (envisioned) implementation, at what point would you see moving to a distributed platform such as Spark becoming necessary?

You may work on any platform of your choosing, including Databricks Community Edition or in local mode. You are encouraged but not required to work in a small group on this project.

Please submit a link to your GitHub repository for your Jupyter notebook or RMarkdown file (and rpubs.com link).

