MOVIE RECOMMENDATION SYSTEM

Project Description

What is a recommender system?

A sort of information filtering system is a recommender system. The system's algorithm can discover precise customer preferences by analysing large data sets. You can recommend new, relevant material to your users after you know what they like. Everything from movies and music to love partners is affected by this.

- Recommender systems are used by Netflix, YouTube, Tinder, and Amazon, to name a few. Users are enticed with relevant suggestions depending on their decisions by the systems.

Project Description

Recommender systems can also improve the following experiences:

- Websites for News
- Knowledge Bases for Computer Games
- Platforms for Social Media
- Support Systems for Stock Trading

Problem Statement

Develop a movie recommendation system based on user feedback.

System of general recommendations:

- In terms of recommendations, we always propose movies with the highest average rating and more than a specified amount of ratings using this method.

Recommendation system based on user input:

- Similarities between user ratings could be used to forecast the user's recommendations.

Implementation

- Exploratory Data analysis on the Netflix data set can be done by loading the Netflix data set into Spark.
- Train, validate, and split your tests.
- Finding the appropriate hyperparameter for a machine learning model.
- Predicting user ratings for movies they haven't seen yet.
- Visualization in action.

Implementation

SPARKMLLIB :

Spark's scalable machine learning library, MLlib, is a collection of common learning methods and utilities.

- Classification, regression, clustering, collaborative filtering, dimensionality reduction, and fundamental optimization primitives are all included.
- Spark ML is also compatible with model training and production, allowing trained models to be quickly deployed to production.

Data-Set

Instead of streaming data, I'd want to use Kaggle's

Netflix movie recommendation dataset.

This is the data set that I would want to use for the project: https://www.kaggle.com/netflix-inc/netflix-prize-data

To check the dataset for duplicate entries and superfluous columns using exploratory data analysis.

Thank you!