**Recommendation System Documentation**

**Introduction**

This documentation provides an overview of a recommendation system implemented using the ALS (Alternating Least Squares) algorithm in PySpark. The system uses collaborative filtering to recommend movies to users based on their past ratings and the ratings of other users.

**Data Sources**

- ratings\_df: Contains user ratings for movies.

- movies\_df: Contains information about movies (e.g., title, genres).

- links\_df: Contains links to external movie databases.

**Implementation Steps**

1. Setting Up Spark Session: A Spark session is created to initialize the Spark context and SQL context.

2. Loading Data: The ratings, movies, and links data are loaded into Spark DataFrames from CSV files stored in an S3 bucket.

3. Data Preparation:

- The ratings DataFrame is split into training and validation sets.

- ALS parameters such as `maxIter`, `regParam`, and `rank` are set.

4. Model Training and Evaluation:

- An ALS model is trained on the training set.

- Predictions are made on the validation set.

- Root Mean Square Error (RMSE) is calculated to evaluate the model's performance.

5. Hyperparameter Tuning:

- The model is trained and evaluated for different values of the `rank` parameter.

- RMSE is calculated for each value of `rank` to find the optimal value.

6. Generating Recommendations:

- Recommendations are generated for all users and all movies.

- Top 5 recommendations are selected for each user and each movie.

7. Displaying Recommendations:

- Recommendations are joined with movie information to display the movie title, genres, and predicted rating.

8. Custom Recommendations:

- Recommendations are generated for a specific user (user ID 12) and a specific set of movies.

- Recommendations are displayed along with movie information and external links.

9. Subset Recommendations:

- Recommendations are generated for a subset of users and movies.

- Top 10 recommendations are displayed for each user and each movie.

10. New User Recommendations:

- Recommendations are generated for a new user by providing a list of movie IDs and user IDs.

- Predicted ratings for the new user's ratings are displayed.

**Conclusion**

This documentation provides an overview of a recommendation system implemented using the ALS algorithm in PySpark. The system demonstrates how collaborative filtering can be used to recommend movies to users based on their past ratings and the ratings of other users. The system also showcases how hyperparameter tuning can improve the model's performance and how recommendations can be generated for new users and subsets of users and movies.