Recommender system

Last updated by | Harish Nedunuri | 21 Mar 2023 at 22:09 GMT

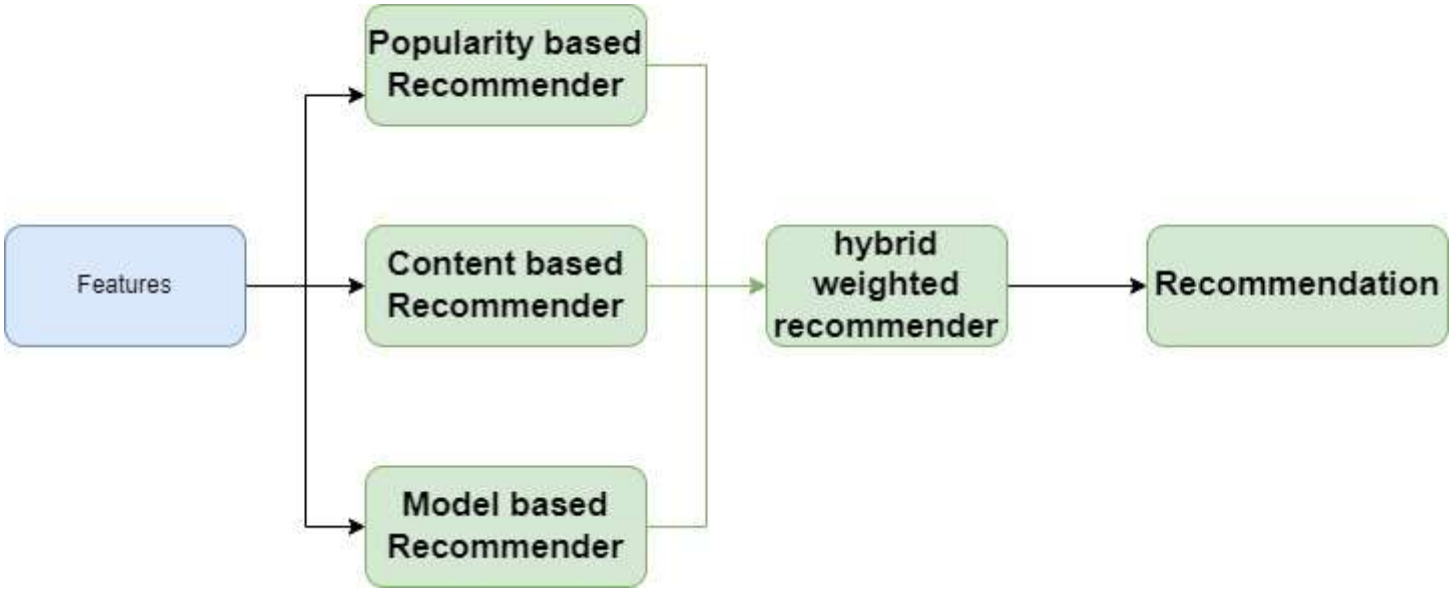
Project Table of Contents

Child Pages

Overview

This project is an attempt to create an effective recommendation system. A python based package suitable for cloud deployment is realised.
The python package will be cloud agnostic. A DevOps solution to make the code operate as a data bricks workflow job is outlined.
Recommendations will be extracted based on three methods and a consolidated recommendation will be provided to the user

1. popularity-based model,
2. the content-based model, and
3. the ML model-based model.



Dataset

An open-source dataset from the Kaggle data repository is used to develop this project.
reference: https://www.kaggle.com/datasets/surajjha101/bigbasket-entire-product-list-28k-datapoints

This dataset contains 10 fields :

- index - Index
- product - Title of the product (as they're listed)
- category - Category into which the product has been classified
- sub_category - Subcategory into which the product has been kept
- brand - Brand of the product
- sale_price - The price at which the product is being sold on the site
- market_price - The market price of the product
- type - Type into which product falls
- rating - Rating the product has got from its consumers
- description - Description of the dataset (in detail)

How do you handle drift in data or bias in data?

How your ML model can handle a biased dataset?