

Afrive Technology Justification Document

A Collaborative Project by Group 2

Faith Obodo - Data Scientist

Stutern,

June 2022



CONTENTS

Introduction

Data Sourcing And Web Scrapping

Rationale Behind Python Libraries Used

Methodology And Model

Deployment And Api Generation

INTRODUCTION

The revolution of E-commerce globally is changing in recent times and it has given more users around the globe, accessibility to their needs. E-commerce has given a wider visibility to small business owners, bringing them closer to their target customers, engaging in business to customers and business to business transactions.

It also gives users the ease to be able to select carefully what exactly what they want with clear stipulated specifications. It has also provided job opportunities.

Afrive project is an E-commerce website, that has only African books in its repository, available to users for purchase.

This website has several features apart from users being able to purchase books, they will also be able to rate and post reviews of whatever book they purchase.

Whatever books that users click on or add to their cart, the website will also recommend related books based on the genre with good ratings to the user.

This recommendation on the website will allow users easily get the kinds of books that they will like to have.

The language used in this project is python and html

Data Sourcing and Web Scrapping

The data for the project was sourced from this [website](#), a list of African books with other detail information such as author name, book title, country, and publisher. The website is scrapped using beautiful soup library, to read the html language, over a hundred of books was gotten from the website. The genre and rating of these books were gotten manually. The dataset was then tweaked to suit the features needed in an excel sheet, to be used in the code editor.

Rationale Behind Python Libraries Used

A recommendation system for new website is usually best to cold start, receiving feedback from users as to the kind of books that they like, and this is used to run the recommendation system.

A popularity-based feature is used to start the system which then subsequently runs with collaborative filtering.

The language used in the code editor is python and the libraries that are used includes:

NumPy: This is an open-source library, that is used for working with arrays. It also has functions for working in domain of algebra, matrices, and Fourier transform.

Pandas: This is an open-source library, built on top of the NumPy library. This is python library for and data analysis, it offers data structures for manipulating data. It is popularly for its ease in importing dataset into a code editor.

Matplotlib: This is a 2-dimensional plotting library for data visualizations, and graphical plotting it supports various types of graphical representations such as bar graphs, histogram, scatter plot, stem plots, line graphs etc.

Seaborn: Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Seaborn works easily with data frames and the Pandas library. The graphs created can also be customized easily.

Truncated SVD from Scikit-learn: Dimensionality reduction using truncated SVD. This transformer performs linear dimensionality reduction by means of truncated singular value decomposition (SVD), this estimator does not center the data before computing the singular value decomposition. This means it can work with sparse matrices efficiently.

Pickle: is primarily used in serializing and deserializing a Python object structure. In other words, it's the process of converting a Python object into a byte stream to store it in a file/database, maintain program state across sessions, or transport data over the network

Flask: is a web framework, it's a Python module that lets you develop web applications easily. It has a small and easy-to-extend core, with features like URL routing, template engine.

Methodology And Model

The method that is being used in the recommendation system is the collaborative filtering. What this does exactly is to pick the items that have been liked or rated by users, this then collects the items with the highest number of high ratings, to return similar books to users.

After cleaning, exploring and visualizing the dataset, it is then converted into a pivot table where the index is the genre of books, the values are the ratings of the books and the columns represent the title of the books, this is then transposed. Only this transposed dataset can be passed into the algorithm, so it is saved as a csv file and downloaded as a `df_t.csv` file.

The transposed dataset is then passed to a truncated SVD algorithm for decomposition, and this gives us the model matrix, which is then saved as a pickle file and downloaded as `correlation.pkl` file.

The result gotten from the algorithm then returns as index numbers which are read as values to return the names of books.

Deployment And Api Generation

An `app.py` file is created to deploy into the flask pipeline, which returns a json file and an embedded message in a url link.

In the `app.py` file, the `[GET, POST]` method is used to get data from the application, based on the post done by the user as a request on the website, and returns results for the user's benefits.

A `requirements.txt` file is created to be read by Heroku.

A procfile with the web: `gunicorn app:app`

`Runtime.txt` file to save what language was used and in what version, in this case is python 3.9.11

To generate the API, an app is created on Heroku with the name `afriverecommendation` and connects to the git-hub folder to enable app deployment, it runs, and the web app goes online.

The API is then generated for subsequent development.