BOOKHUB

PROJECT SYNOPSIS
OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY

Computer Science Engineering

Submitted By

KONARK LOHAT (1905801)

GAUTAM JAIN (1905785)

JASKAMAL SINGH (1905006)

September 2022

GURU NANAK DEV ENGINEERING COLLEGE, LUDHIANA

TABLE OF CONTENTS

1. INTRODUCTION

A platform where users can register themselves as a creator and a reader. The goal of this idea is to entice the excitement for books in the modern generation with the help of AI and the recommendation engine for books (minor project). A full-stack application comprising a stand-alone front-end, back-end, and database server(s) to ensure the scalability of the platform.

Users can create books with the help of an easy UI/UX, a full authentication and authorization system, and users will be recommended based on other similar users (collaborative filtering) and similar books (content-based filtering). Books can be purchased with the help of a credit system.

Technology Stack -

1. Frontend - ReactJS

   ReactJS is a front-end JavaScript library for creating user interfaces that is open-source. Facebook and a group of independent developers and businesses manage ReactJS. For online and smartphone applications, it is used to manage the display layer. We can build reusable UI components with ReactJS.

2. Backend - NodeJS/Django
   - Node.js is an open source server environment
   - Node.js is free
   - Node.js runs on various platforms (Windows, Linux, Unix, Mac OS X, etc.)
   - Node.js uses JavaScript on the server

3. Database - MySQL

   MySQL is a fast, easy-to-use RDBMS being used for many small and big businesses. MySQL is developed, marketed and supported by MySQL AB, which is a Swedish company. MySQL is becoming so popular because of many good reasons − MySQL is released under an open-source licence. So you have nothing to pay to use it.

4. Version control - GIT/GitHub

   Git is a **version-control system** for tracking changes in computer files and coordinating work on those files among multiple people. Git is a Distributed Version Control System. So Git does not necessarily rely on a central server to store all the versions of a project's files.

2. RATIONALE

Book market of India is the 6th largest market in the world and 2nd largest for the English language book market. Online book users are fairly less as compared to other countries. We believe that being online could be the new way to connect to like-minded people based on the similar interest for various types of content.

A platform to share, read and explore the content writing of other users based on their interests. This platform can act as an open collection of content to explore new options. The proposed platform would also be able to carve a revenue model for content writers/readers.

3. OBJECTIVES

1. Design a simple UI/UX interface which enables a user to create/read content.
2. Develop two stand-alone servers to handle authentication and recommendations.
3. Generating dataset based on the activities of active users, for recommendation purposes.
4. Linking of four stand-alone servers for scalability.

4. LITERATURE

1. Google Developers Community [1]: Recommendation System by Google Developers Platforms

- Proposed Learnings:
  - Describe the purpose of recommendation systems.
  - Understand the components of a recommendation system including candidate generation, scoring, and re-ranking.
  - Develop a deeper technical understanding of common techniques used in candidate generation.
  - Content-based filtering uses item features to recommend other items similar to what the user likes, based on their previous actions or explicit feedback.
  - Collaborative filtering uses similarities between users and items simultaneously to provide recommendations

2. Joeran Beel et al. [2]: Research-paper recommender systems: a literature survey

- This paper has introduced recommender systems to new research. This paper has also identified key problems which need research in recommender systems.
- This paper can help Ph.D. and Master's students in choosing their area of research. The research gap is already presented in this paper to form different problems of recommender systems.
- The recommendation system finds its utility in major areas of web Applications. As these problems get solved more and more useful recommendation systems will become.

3. Nitin Mishra et al. [3]: Research Problems in Recommender systems

- It is a system that helps users to choose items that they may need. A Good Recommender System saves user's time and keeps the user engaged in the system resulting in higher revenue.
  - Collaborative Filtering
    - KNN Model - KNN is a Supervised Learning algorithm that uses labelled input data sets to predict the output of the data points.
    - Pearson Correlation - Pearson correlation makes use of Pearson's Correlation Coefficient to make predictions by means of linear correlation between two data points.
  - Content-based Filtering
    - TF-IDF Model - This Model makes use of tf-idf metric is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in searches of information retrieval, text mining, and user modelling.

5. FEASIBILITY STUDY

1. **Technical Feasibility -** The project is evaluated to be technically feasible as the technology to be used is easily available and open source. The technology and related dependencies are up to date with the current technical requirements and provide the best possible way for implementation of the project.

2. **Market Feasibility -** The project is beneficial for the software engineers/software designers/system architecture, writing blogs as the recommendation system provides a platform/opportunity for them to connect to their desired target audience. The project being open source provides a feasible and easy way for them to customise the system according to their needs and use it to boost their target space.

3. **Economic Feasibility -** The project is economically feasible for the technology-related content creators to implement a powerful recommendation system on their product with virtually no additional added cost (excluding the hosting and its related costs). As the recommendation system grows further with contributions, the cost-benefit factors related to various components flattens. The technologies used in the project are easily available for free.

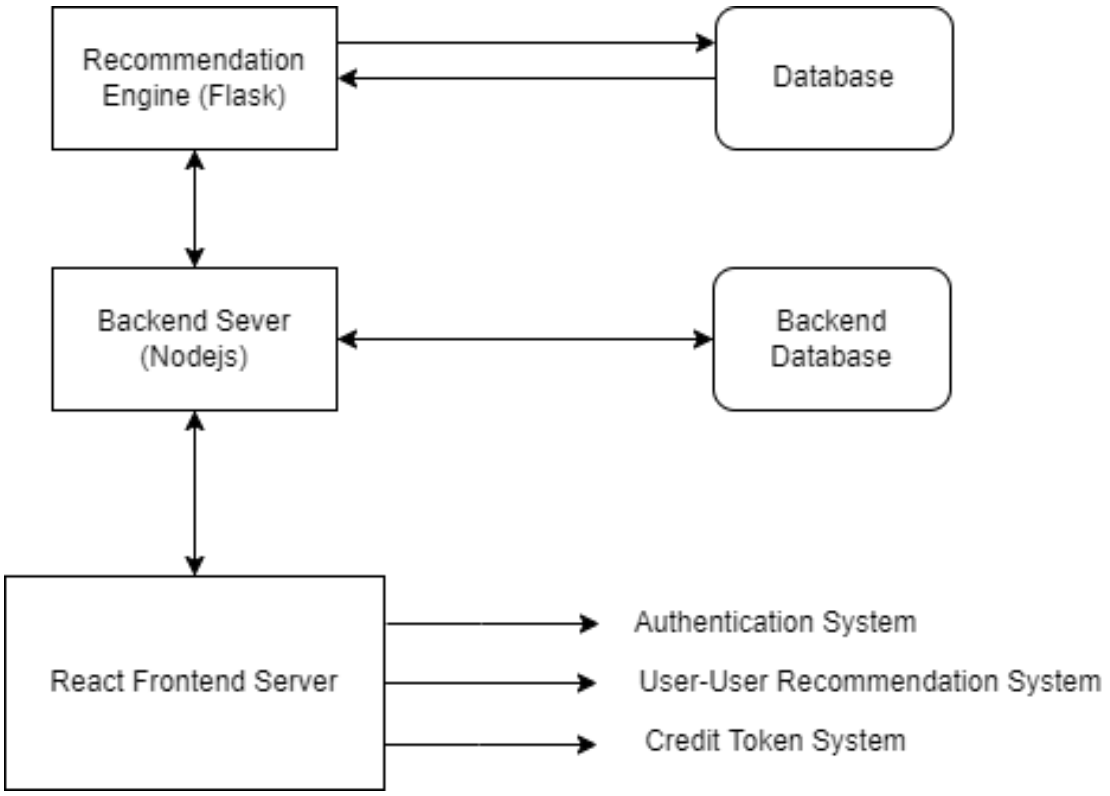
4. **Operational Feasibility -** The project is operationally Feasible as the recommendation system is easily customizable to meet the requirements of a particular product. The project being open source provides horizons to continuous improvements which are beneficial for the product to get the best accurate/optimised recommendations.

6. Methodology / Planning of work

An AI powered community for book creators / readers. This platform enables a user to create or read books free of cost or with the use of credit tokens based on the content owner's preference.

The platform also acts as a revenue generation platform for amature, rookie content creators to generate books, journals, novels, etc.

Four Standalone servers ensure the scalability of the application, with the frontend, backend, recommendation engine and database server.



6.1 Flow Chart of the application servers.

Major phases in our project include -

1. Recommenders Procedure:

1.1. Candidate Generations: It will be responsible for generating smaller subsets of candidates to recommend to a user, given a huge pool of thousands of items.
1.2. Scoring Systems: Candidate Generations can be done by different generators, so, we need to standardise everything and try to assign a score to each of the items in the subsets.
1.3. Re-Ranking Systems: After the scoring is done, the system takes into account other additional constraints to produce the final rankings.
1.4. Based on these scores, recommendations will be made and displayed to the users based on common interests.
1.5 Models used for the recommendation engine

1.5.1 **TF-IDF Model** - TF-IDF (Term Frequency - Inverse Document Frequency) is a handy algorithm that uses the frequency of words to determine how relevant those words are to a given document. It's a relatively simple but intuitive approach to weighting words, allowing it to act as a great jumping off point for a variety of tasks.

1.5.2 **Pearson correlation Model** - The Pearson correlation measures the strength of the linear relationship between two variables. It has a value between -1 to 1, with a value of -1 meaning a total negative linear correlation, 0 being no correlation, and + 1 meaning a total positive correlation.

1.5.3 **KNN Model (Cosine Similarity) -** Cosine similarity is used as a metric in different machine learning algorithms like the KNN for determining the distance between the neighbours, in recommendation systems, it is used to recommend movies with the same similarities and for textual data, it is used to find the similarity of texts in the document.

2. Dataset Generation

2.1. Dataset will be generated based on the user activities in real time.
2.2. This dataset will be fed to the recommendation engine to generate real-time recommendations.
2.3. Dataset will be generated at frontend and processed at backend and then fed to the recommendation engine, while correspondingly getting saved to the database in process.

3. Interconnection between standalone servers

3.1. Frontend React server with simple UI/UX to generate dataset.
3.2. Backend Node/flask/Django server to process the data.
3.3. Recommendation engine which recommends content based on the user-user connection.
3.4. My Sql server which stores all the data.

7. FACILITIES REQUIRED

1. **Software Requirement**
   - 1.1. Programming Language - Python v3.10+
   - 1.2. Collaboration - GitHub
   - 1.3. Browser - Chromium-based
   - 1.4. Database - SQL database

2. **Hardware Requirement**
   - 2.1. Recommended OS - Windows 11
   - 2.2. Recommended RAM - 8 GB
   - 2.3. Recommended Processor - Intel i3 or above

8. EXPECTED OUTCOMES

1. The project aims at making predictions based on the user's interest.
2. A platform for reading and creating books/content.
3. A credit system to ensure the viability of the project according to market standards.

9. REFERENCES

[1] Google Developers Community (2020). Recommendation System by Google developers platforms. Link - https://developers.google.com/machine-learning/recommendation

[2] Joeran Beel, Bela Gipp, Stefan Langar, Corinna, Breitinger et al., 2016. Research-paper recommender systems: a literature survey. Journal on Digital Libraries. Link - https://d-nb.info/1147681678/3442-6596/1717/1/012002/pdf

[3] Nitin Mishra, Saumya Chaturvedi, Aanchal Vij, Sunita Tripathi et al., 2020. Research Problems in Recommender systems - Journal of Physics: Conference Series. Link - https://iopscience.iop.org/article/10.1088/1742-6596/1717/1/012002/pdf

[4] Towards Data Science, Medium Blog Channel. Recommendation System Series. Link - https://towardsdatascience.com/

[5] 5 Awesome Online Communities for Book Lovers Link - https://bookclubbites.com/5-best-online-communities-for-book-lovers/