Project Proposal

of

User Behavior and Recommendation Analysis

Submitted by:

Name: Divyanshi Maurya

Reg. no.: 12219959

Section: K22UG

Roll no. : RK22UGA33

Submitted to:

Mr. Ved Prakash Chaubey (UpGrad, Instructor)

63892

Subject: CSE 353: EDA PROJECT

In partial fulfillment for the requirements of the award of the degree of

“Bachelors of Technology in Computer Science and Engineering”

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

TABLE OF CONTENTS

Project Title:

User Behavior and Book Recommendation Analysis Using Apriori Algorithm on Goodreads Data

1. Objective:

This project aims to analyze user shopping behavior and preferences by applying the Apriori algorithm to a retail dataset. The objective is to identify frequent itemsets that are commonly purchased together and generate meaningful product recommendations based on association rules. The analysis will provide insights into shopping trends and patterns, particularly in today's digital era, where personalized content curation is increasingly valuable. Additionally, this project will explore the influence of social media networks on consumer preferences and examine the broader impact of algorithmic recommendations in shaping personalized shopping habits.

2. Example of Today's Scenario:

IIn today's digital landscape, consumers are bombarded with an overwhelming amount of product choices, making it challenging to discover items that truly align with their preferences. Platforms like Amazon, Google, and Flipkart serve as social hubs where users shop and engage in e-commerce, providing a wealth of data on user behavior. By leveraging this data, personalized recommendation systems can enhance the shopping experience, promoting greater engagement and helping users discover products that resonate with their needs and interests..

3. Impact of the Project:

- Enhance user experience by offering tailored product recommendations based on their wishlist and viewing patterns.
- Assist users in discovering lesser-known products that align with their interests, promoting diverse recommendations.
- Demonstrate the effectiveness and limitations of using the Apriori algorithm for recommendation systems.
- Able to provide much more data-driven insights using data-driven datasets

4. Dataset:

Source: Retailrocket recommender system dataset

Description: The dataset consists of three files: a file with behaviour data (events.csv), a file with item properties (item_properties.csv) and a file, which describes category tree (category_tree.csv). The data has been collected from a real-world ecommerce website. It is raw data, i.e. without any content transformations, however, all values are hashed due to confidential issues. The purpose of publishing is to motivate researches in the field of recommender systems with implicit feedback.

5. Project Timeline:

Week 1-2: Data Collection and Preprocessing

- Objective: Collect data using the kaggle, clean and preprocess it for analysis.
- Then I will fetch the data related to user reviews, ratings and reading list.
- For this project, I will take around 15,000+ number of users to ensure the diversity and comprehensive data.
- Then we would perform data cleaning : Removing duplicates, handling missing and incorrect data, and normalizing text (removing uppercase, lowercase and special characters).

Week 3-4: Exploratory Data Analysis (EDA)

- I will perform EDA to understand the distribution of data, identify patterns, and uncover initial insights.
- After I clean the data, I will convert the cleaned data into pandas data frame, for easier manipulation.
- Then, we would perform basic statistics, rolling statistics, hypothesis testing, outlier analysis and check the distribution of the data
- Visualization of reading patterns across genres, identification of top 10 most-read books, how many books are typically read by users on average.
- It will help in refining the scope of the analysis by focusing on prominent reading trends.

Week 5-6: Implementation of Apriori Algorithm

- Now we would need to organize the data, so that each user’s reading history is a transaction.
- Then we will apply the Apriori algorithm using ‘apyori’ library to identify frequent itemsets and generate association rules.
- There are some parameters:

1. Support: Minimum threshold for how often a combination of books appears together.

2. Confidence: Minimum threshold for how often the rule has been found to be true.
3. Lift: A measure of the strength of association between books.

- Then we would extract and format the results into readable rules, to make use of recommendation
- Statistics: Expectation of generating over 1,000 significant association rules with a minimum support threshold.
- Core component of the project, leading to the creation of a functional recommendation engine.

Week 7-8: Analysis and Optimization

- Analyze the results, we will optimize parameters like support and confidence, and refine the recommendation engine.
- Comparison of rule performance with high lift ad confidence values,as these indicate strong associations between books, focusing on precision and recall metrics.
- Ensures that the recommendation engine is both accurate and relevant by parameter tuning and testing to user preferences.

Week 9: Visualization and Interpretation

- Visualize the results using graphs and charts, and interpret the findings.
- Creation of visual reports highlighting key association rules and their implications.
- Provides a clear and interpretable presentation of the analysis, making it accessible to non-technical stakeholders.

Week 10: Final Report and Presentation

- Compile the results, document the project, and present the findings.
- Generation of a comprehensive report with actionable insights and future recommendations.
- Disseminates the project’s outcomes, offering valuable learnings for future research and applications.

6. Results:

The project will yield a functional product recommendation system based on user behavior analysis. The findings will illustrate key trends in shopping habits and demonstrate the effectiveness of the Apriori algorithm in generating meaningful recommendations. The final system could be extended or integrated into existing e-commerce platforms to enhance user engagement and drive more personalized shopping experiences.

7. Future Goal:

This project will impact both individual consumers and the retail industry by promoting personalized product curation and providing data-driven insights into shopping trends. It will also contribute to the broader field by showcasing the applicability of association rule learning, particularly the Apriori algorithm, in developing effective recommendation systems.

8. Learnings:

- Gained experience in data collection and preprocessing using APIs.
- Applied the Apriori algorithm in a real-world context, learning about its strengths and limitations.
- Developed skills in data analysis, visualization, and interpretation.

9. Technologies Used:

- Python: For data collection, preprocessing, and implementation of the Apriori algorithm.
- Pandas, NumPy: For data manipulation and analysis.
- Matplotlib, Seaborn: For data visualization.
- Requests: For API interaction.
- Google Colab: For cloud-based development and collaboration.

10. Implementation Plan:

The project will be executed in a Google Colab environment, with Python serving as the primary programming language. The workflow will begin with data collection from the retail dataset, followed by data cleaning and exploration. The Apriori algorithm will then be applied to identify patterns within the data, which will be analyzed and visualized to draw meaningful conclusions. The entire process will be meticulously documented and presented in a comprehensive final report.

This comprehensive proposal outlines the steps necessary to achieve the project's objectives, ensuring a structured approach to exploring user behavior and delivering a functional recommendation system.

11. Conclusion

This project showcases how data-driven methods can help us understand user behavior and improve product recommendations. By collecting and analyzing data from a retail dataset using the Apriori algorithm, we identified patterns in shopping habits and generated meaningful product recommendations. The results showed that personalized recommendations can greatly enhance user engagement by suggesting products that match individual preferences.

The project also provided insights into the impact of algorithmic recommendations, emphasizing the need for diverse data sources and careful parameter tuning to achieve the best results. The lessons learned from this project, including data collection, preprocessing, and applying the Apriori algorithm, contribute to the growing field of recommendation systems and can be applied to other areas beyond retail.

In conclusion, the project successfully met its goals by creating a functional recommendation system and deepening our understanding of user behavior in online shopping. The methods and tools used in this project can be adapted for future work, making it a strong foundation for ongoing research and applications in personalized content curation.