DiscoverEase: Tailored App Recommendations at Your Fingertips

GROUP 13 TEAM MEMBERS

- Swarn Gaba (sgaba@iu.edu)
- Gandhar Ravindra Pansare (gpansar@iu.edu)
- Chaithra Lal Nair (cnair@iu.edu)

Together, our team aims to leverage our combined skills and expertise to create **DiscoverEase: Tailored App Recommendations at Your Fingertips**, a comprehensive solution designed to enhance the app discovery experience for users by offering personalized, data-driven app recommendations. Each team member's contribution is vital to our project's success, ensuring that DiscoverEase addresses the practical needs of users and stands out in the digital landscape.

PROJECT SUMMARY

The sheer volume of apps available in app store can be overwhelming for the users. A system that helps navigate these choices by providing personalized recommendations can significantly enhance user experience by making it easier to discover apps that align with their interests and needs. Our goal is to develop an innovative app recommendation system, supported by a vast dataset of over 1.2 million records from the Apple App Store, acquired through Kaggle.

By prioritizing data-centric techniques and analyzing key dataset characteristics such as 'App_Name,' Genre,' and 'Content_Rating,' our system will provide personalized suggestions. It will allow users to refine their preferences based on genre, content rating, and other app features, ensuring recommendations are tailored to individual needs.

This approach will help us to create a platform that enables users' engagement to explore new and upcoming apps, find undiscovered treasures, and keep up with the latest app industry trends, all customized to their individual preferences.

PROJECT DESCRIPTION

Objectives:

Our main goals are to make finding apps easier and to provide a personalized experience for users. We want to simplify the app discovery process so users don't have to go through millions of apps to find ones that genuinely interest them. Additionally, we aim to personalize app recommendations, allowing users to customize their results based on their preferences and specific parameters like genre, user rating etc. This way, users can easily discover apps that suit their tastes and needs without the hassle, making the entire process efficient and enjoyable.

Our goals include:

- **Customized User Experiences**

  The aim here is to tailor app recommendations specifically to each user's preferences. This means the system will consider individual likes, dislikes, and past interactions with apps to suggest new apps.

- **Maximizing Dataset Potential**

  This objective focuses on making the most of the extensive data available from the Apple App Store. The dataset includes various app attributes such as genre, content rating, and user ratings. By analyzing this data, the system can identify patterns and preferences, leading to more accurate and useful app suggestions.

- **Initial Experience Enhancement**

  The 'cold start' problem is a common challenge in recommendation systems, referring to the difficulty of providing personalized recommendations to new users who have not yet provided enough data about their preferences. This objective addresses this challenge by finding innovative ways to offer meaningful recommendations to new users, possibly by utilizing generalized data or trends among similar users, until the system has gathered enough individual data to personalize suggestions.

- **Broadening the Recommendation Spectrum**

  In addition to personalization, the project seeks to broaden users' horizons by providing a variety of app suggestions. This means the system will not only focus on users' established preferences but also introduce them to apps they might not have discovered on their own.

Usefulness:

Our app recommendation platform strives to serve a wide range of users by utilizing the power of a comprehensive database, becoming an invaluable tool for anyone seeking personalized app suggestions. The database, complemented by an interactive interface, promises significant utility by enhancing app visibility, particularly for smaller developers and niche applications struggling in a saturated market. Moreover, the database offers deep insights into market trends, user behaviors, and preferences, which are invaluable for app developers and industry stakeholders to make informed decisions on app development and marketing strategies.

Our database stands out by incorporating a wealth of data from the Apple App Store, comprising around a million app entries. While our platform shares some inspiration with the Google Play Store dataset, our focus on the Apple App Store data sets us apart, offering a specialized approach to app recommendations for iOS users. This differentiation makes our database particularly relevant for

individuals seeking tailored app suggestions within the Apple ecosystem, enriching the app discovery experience.

Dataset:

The dataset is obtained from Kaggle:

https://www.kaggle.com/datasets/gauthamp10/apple-appstore-apps?select=appleAppData.csv

The dataset we obtained is provided in CSV format and features an extensive compilation of data from the Apple App Store. This dataset consists of approximately 1.2 million records, distributed across 42 distinct columns, capturing a wide array of information including but not limited to "App_Name," "Primary_Genre," "Content_Rating," "User_Rating," and "Description." Originally, the dataset was curated for the purposes of cataloging apps and facilitating in-depth analysis. As such, it stands as a significant resource for our recommendation system, offering a rich source of data regarding the variety of applications available on the Apple App Store.

With its vast collection of approximately 1.2 million entries and 42 different attributes detailing various aspects of the apps, this dataset provides a comprehensive foundation for our system. However, to align more closely with our project's scope, we will be utilizing a smaller subset of this dataset. This approach allows us to normalize the data effectively and organize it into separate tables as required for our analysis.

The dataset was meticulously compiled by Gautham Prakash in October 2021, drawing inspiration from a similar dataset that was designed for Google Play Store applications. The primary objective behind the creation of this dataset was to establish an extensive and detailed database of applications found in the Apple Store, thereby supporting a wide range of data-driven projects and analyses.

Technology Stack:

In the development of DiscoverEase, our tailored app recommendation system, we have carefully selected a technology stack that optimizes performance, facilitates efficient data management, and ensures a seamless user experience. Our chosen technologies are **Streamlit, Python, and SQLite databases.**

- **Streamlit:** We will utilize Streamlit to transform our Python scripts into interactive, user-friendly web applications, as it will significantly accelerate the development of our project's front-end interface.

- **Python**: It will serve as the backbone of our project, as it is renowned for its versatility and powerful libraries that are particularly beneficial in data analysis and web development.

- **SQLite:** It will provide us a lightweight, efficient database solution for our project, offering seamless data storage and quick retrieval capabilities crucial for managing the extensive dataset and user preferences.

Together, Python, Streamlit, and SQLite will form a cohesive and powerful technology stack that underpins the DiscoverEase project. This combination will allow us to present our insights through an intuitive web interface, and manage a vast dataset effectively.

Sharing:

To ensure seamless collaboration and efficient management of our project deliverables, we have established a comprehensive plan for sharing and tracking our work. Central to our strategy is the utilization of a GitHub repository, which serves as the primary platform for version control and collaboration. The repository can be accessed via the following link:

https://github.iu.edu/sgaba/DiscoverEase

Our communication strategy encompasses regular meetings conducted through Google Meet, supplemented by in-person gatherings at Herman B Wells Library. These sessions are essential for maintaining open lines of communication, addressing any challenges promptly, and aligning our project goals and timelines.

The GitHub repository mentioned above will host all project-related materials, including but not limited to, source code, datasets, documentation, and progress reports. This platform will enable us to efficiently manage contributions from all team members, track changes, and review code. It will also allow us for the integration of issue tracking and feature requests, making it easier to prioritize tasks and manage project milestones.

This approach will ensure that all team members are aligned and can contribute effectively to the success of the DiscoverEase project.

Group Contribution During Data Description and Planning Phase:

Name	Tasks	Average Time Spent (per milestone)
Swarn Gaba	Leading project conceptualization, defining objectives, contributing to dataset selection, tech stack decisions, contributing technical insights and report creation.	12 hours
Gandhar Ravindra Pansare	Preliminary data analysis, key attribute identification, contributing to dataset selection, tech stack decisions, and report creation.	12 hours
Chaithra Lal Nair	Planning technology stack, initial planning for future database design, contributing to dataset selection, and report creation.	12 hours