Twitter Influencer RecSys

Capstone Project II – Final Report

Group 7

Gavriel Owens Vincentio	BF21DSY004
Gargi Asthana	BS20DMU002
Viren Bachani	BS20DMU014
Radhika Barve	BS20DMU008

Table of Contents

1. User Requirement and Project Description

User Requirement

To optimise the outcome of our model, user requirements must be considered. IKEA, our client who intends to advertise their newly released furniture, needs to provide us with their desired target audience or niche for the advertising campaign to ensure the effective functioning of our model.

1. Keywords – IKEA needs to select a specific category from a pre-defined list of keywords or options we offer. For example, IKEA could select furniture or Decor as these categories are more relevant to a product like a keyboard.

2. Reach - Our clients need to specify the level of engagement or the number of followers they desire for the celebrity they wish to collaborate with. The popularity of the celebrity will depend on the chosen reach, and it's crucial to keep in mind that more popular celebrities will require a higher investment. Our goal is to assist our clients in making well-informed decisions by providing them with this information, allowing them to select a suitable celebrity within their advertising budget.

3. Marketing Budget - To provide practical recommendations to our clients, we consider not only the desired reach but also the marketing budget. This helps us to fine-tune our results and provide our clients with better-defined expectations based on their budgets.

By running the requirements through a pre-trained model, we can identify the most appropriate options according to the client's input. These options accurately capture the niche and engagement expectations when selecting an influencer or celebrity from Twitter.

Project Description

By running the requirements through a pre-trained model, we can identify the most appropriate options according to the client's input. These options accurately capture the niche and engagement expectations when selecting an influencer or celebrity from Twitter.

To execute our project, we need financial support to cover various expenses such as team salaries, food, and other perks. Additionally, we require resources to train our algorithms, which may include cloud services. We also need to budget for licenses, consultation, model development, and product testing. Our stakeholders comprise the project team, client, influencers, and other individuals or groups who are involved with the system in some way.

Our system's development consists of three primary stages: scraping, recommending, and predicting. The success of our system depends on technical resources such as internet

connectivity and platforms for data scraping, training, and testing. However, we are aware of the risks associated with our recommendation system, such as data privacy issues and the potential for influencers to fail in meeting client expectations after being recommended. Our limitations include using only one data access platform and the possibility of disagreements between the client and influencers. Our success criteria are based on the algorithm's ability to accurately recommend an influencer, with a KPI or accuracy score determined as the benchmark for success.

Our aim is to create a recommender system that recommends the most appropriate influencer for product marketing, based on the client's requirements. To automate the process of finding influencers, we are developing a collaborative filter model that eliminates biases that were previously present in the manual process. The model clusters influencers according to shared interests and ranks them based on their suitability for the product.

We are collecting data by scraping information from Twitter using snscrape, a social networking service (SNS) scraper that can collect millions of tweets, user profiles, hashtags, threads, and list posts. Our collaborative filtering recommendation system clusters Twitter users based on their preferences by filtering users based on similar reactions from other users.

Our approach involves searching for tweets using Twitter's query search and manually inputting ten niches, including makeup, technology, gaming, sports, illustrations, comedy, entertainment, music, photography, and culinary. We plan to save those tweets, ranging from 700K to 10M raw data, and integrate K-mode clustering algorithm, commonly used in applications like Amazon's product recommendations, Spotify, and iTunes' music recommendations, into our recommender system for IKEA.

2. Architecture of the Design Project

The architecture of the design project involves several subcategories including the work breakdown structure, project schedule, risk register and stakeholder register. The work breakdown structure involves how the team has segregated the division of the project, which is our case involves 3 categories i.e., the design stage, the development stage, and the deployment stage. The design stage comprises of

1. The performance requirement, which basically equates to the ability to which the software system accomplishes certain outputs under a given set of conditions.
2. The structure of the project which gives a detailed outline as to how the recommender system is built along with its stages from design to deployment.
3. The development stage consists of the code development which in our case would be using all the libraries used to scrape, clean and process data. The
4. UI/UX components of the recommender system that we build using Twitter as our platform.
5. The deployment stage involves (1) the marketing of the recommender system, followed by (2) the regular updates we need to provide to the system to maintain the efficiency and the accuracy of its outputs. Moving along to the project schedule, it consists of the timelines we have calculated taking various factors into consideration.

Risk register are a vital aspect of a project as they help identify the potential consequences associated with the project. This can range anywhere from the planning phase to the deployment of the software in question. Risk registers also inculcate the likelihood of a consequence occurring. In our case, data privacy is one of the problems that may arise resulting from data processing. Privacy, in today's world, directs strongly towards ethics and its effects have a high consequential probability, and hence can be counted as a high-risk issue. An example could be data breach of Personally Identifiable Information (PII). Another potential risk concerns around the recommendation itself i.e., there may be a possibility that the recommended influencer fails to meet the client's expectations. Even though, this may be a very rare scenario, a possibility still exists. A third risk that we can register would be a compatibility disagreement between client and system recommender influencer. In today's business world, this tends to be a commonality. Regardless of the limited number of risks we have outlined as part of the risk register, yet there may exists a lot more than the ones underlined.

In simple words, stakeholders are an entity or group of entities that are involved as part of a product (or majorly a business), directly or indirectly. In our project, the stakeholder register aims to focus on the stakeholders that are involved in the entire process right from planning the project structure to software deployment. The stakeholders to the system of our product will primarily consist of the client (and their subset of family or friends). Our stakeholder may also include other clients where our recommender system will be used. The primary stakeholders in connection to the client may include the leading positions of the company. Furthermore, since we focus on a media platform, the marketing department of our client may end up being the primary stakeholders as well. Lastly, members of other departments

like Finance will also play a key part in the usage of our product and hence can be included as another set of primary stakeholders.

Our investors are categorized as important stakeholders simply because they have a vital part in the initial, development, and end stages of the influencer recommender system. Apart from investors or venture capitalists, banks and other financial entities who choose to invest in the recommender system built can also be considered stakeholders. The developing team who plans, develop, produce, and install the Twitter influencer recommender system make up for our secondary system of stakeholders. Despite having been coined as 'secondary', these entities exist right at the core of the circle of stakeholders. More secondary entities include the project manager. Moreover, any entity as part of the development, evaluation, or deployment of the goods or services are all included in the circle of secondary stakeholders as well.

Competitors are another set of entities that can be part of the tertiary group of system stakeholders. Though being on the other end of things, it is wise to make a note of these individuals as part of our feedback or improvement, which in turn, makes them part of our stakeholders. Lastly, the sole purpose of using the terminologies 'Primary', 'Secondary', & 'Tertiary' is segregating them based on location or any other common clusters and is not based on hierarchies between these clusters in any way.

3. Project Management Plan

Atlassian Agile

For effective workflow, we have allocated tasks based on expertise, interests, skills and availability. Having a well-planned work division has led to more productive use of time and skills which has led to the successful completion of our project. A software that we used is Atlassian Agile, a project management tool designed for collaboration and effective project management. Promoting a continuous delivery encourages the team to work in short spans of time and requires active presence along with frequent feedback. While we used a one-on-one method to communicate effectively, Agile has helped the team align important deadlines and aided in keeping track of division of work.

Some Key benefits of using Agile are:

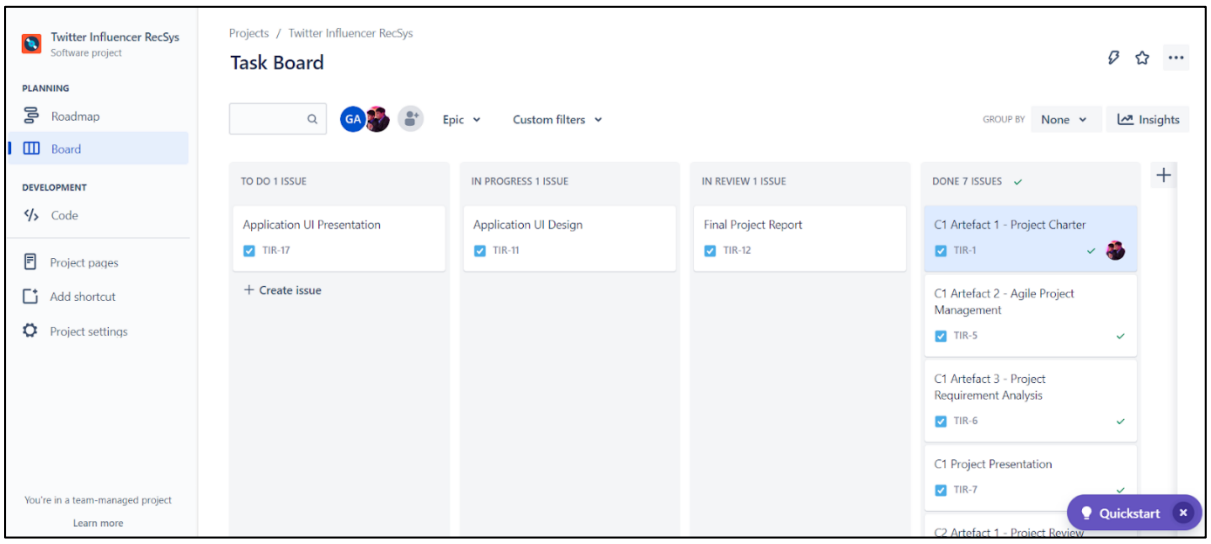
1. Increased Productivity: By breaking down tasks into smaller and attainable goals, we have been able to focus on the most valuable task at a given time.
2. Improved collaboration: Communication and collaboration of team members is highly encouraged aiming to improve the quality of inter-team relationships and trust as well as maintaining a high quality of work as everyone is on par with the current task on hand.
3. Flexibility and Adaptability: Being an easy-to-use tool, tasks can be easily updated as and when required.
4. Transparency and visibility: Being accessible to all team members, it encourages the team to be responsible and accountable for the work at hand. This also allows team members to be vary and attentive of each other's works.

In order to have a well-planned and smoothly run project, here are some of the steps that were involved in the work division and project planning:

1. Defining the project scope: This included defining the project goals (including long term and short-term goals) and some key deliverables. This was majorly executed during the first phase of the project i.e., 'the project planning phase'.
2. Identifying tasks: We then identified tasks required to complete the project. We defined user requirements and KPIs that we intended to achieve. For software development we identified that some of the key tasks that will be executed are data collection, cleaning, engineering, model development and deployment.
3. Determining team member skills: Understanding that each team member has expertise in different areas, two team members were assigned the project planning and report writing tasks while the other two focussed more on the coding and software development.
4. Defining deadlines and monitoring progress: We used Agile to set up well defined deadlines and kept a close eye on details such as quality of work and ensuring that the team members were able to meet project deadlines.

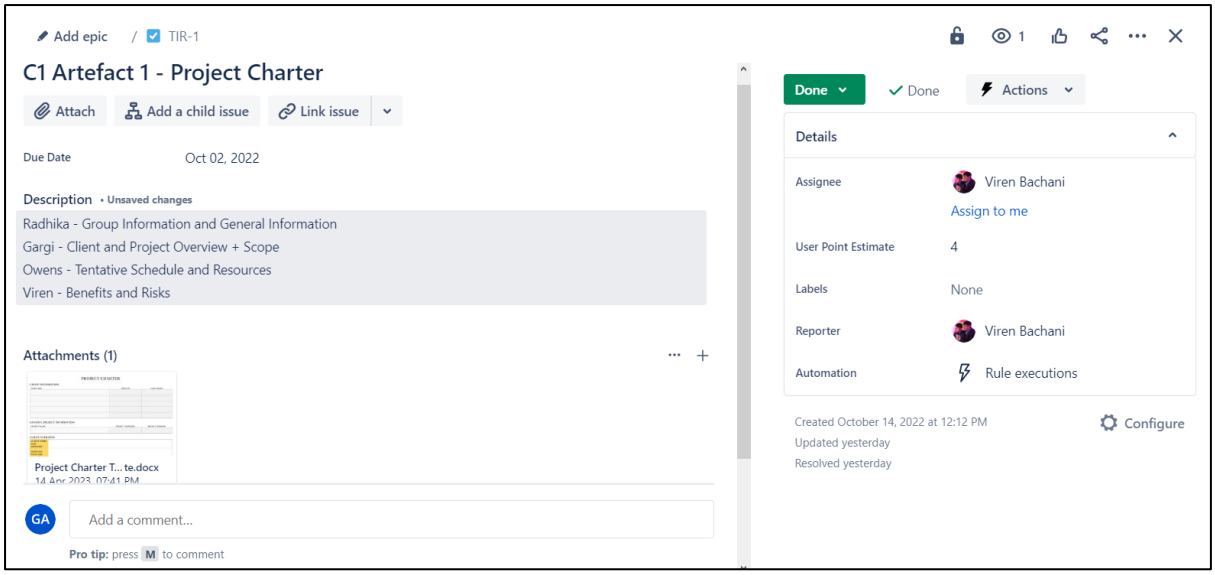
Walkthrough of Project Management and Planning Tools

We set up a project task board where tasks would be updated frequently as they came up. Here is an overview of how the task board looks when we have pending tasks. All tasks were initially added to the “To Do” section. When in progress or in review they were added to their respective sections. Only when the task was thoroughly reviewed, completed, and submitted, it was finally marked as “Done”. This provided us with clarity on the expectation of a task in hand.



View of Project Board

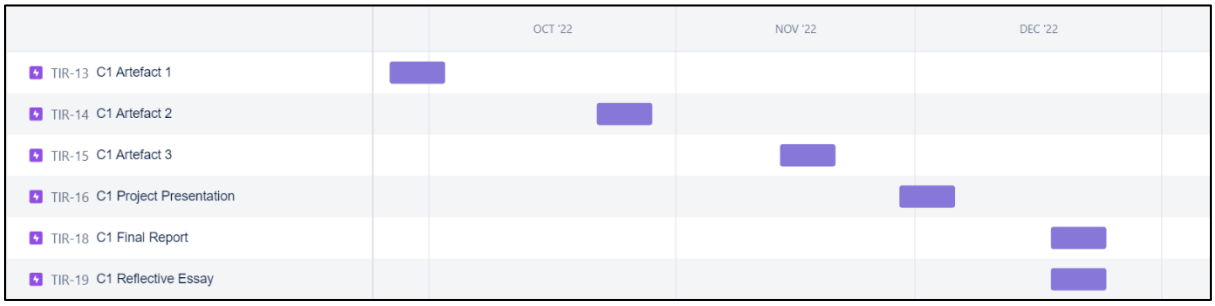
Each task that was issued had all the required information within it so that all members had easily accessible resources. Meeting minutes were also included in the description when required to keep track of details of the tasks. An example of the same is shown below to illustrate the execution of Capstone 1 Artefact 1 – Project Charter where the due date, work division, link to the Google Document file with everyone’s complied work and the reference dummy project charter was included.



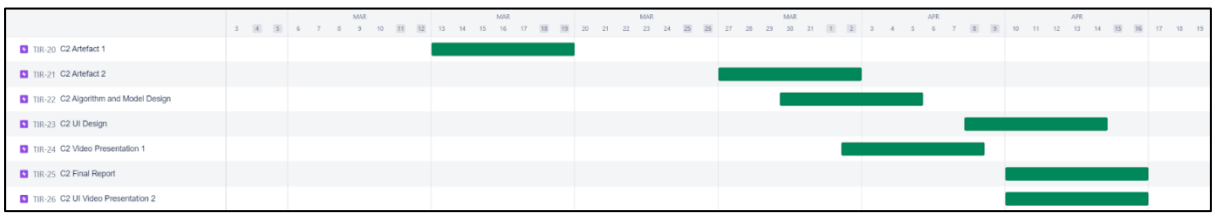
Example of Tasks in Project Board

Due dates for tasks were mentioned along with the work division to maintain transparency. Adequate time was given to team members and one on one communication was encouraged when assistance was required.

As the project was segregated into two parts – the “Project Planning and Management Phase” (C1) and the “Project Deployment Phase” (C2), we constructed a roadmap for better workflow and visual representation of deadlines. The following are the constructed roadmaps for C1 and C2 respectively.



C1 Project Planning and Management Phase



C2 Project Deployment Phase

Overall Project Roadmap

The overall Project Roadmap gives us a better idea about how project deadlines were spaced out. Eventually all deadlines were successfully met due to a robust and organized project management system.

For any further communication, we preferred direct or one on one conversations or issued a team meeting where participation of all members was highly encouraged.

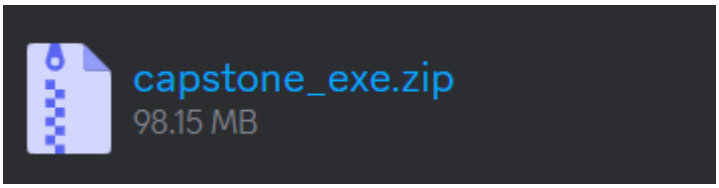
4. System Requirements for Installation

Our software is easy to run and requires a system that is installed with Windows 10 or above to run the front end. Please keep in mind that the file does not run in iOS. No other installations or requirements are necessary.

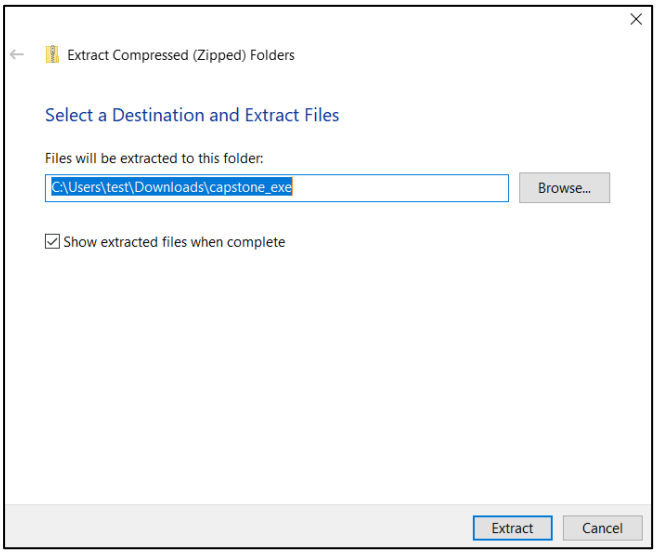
5. Step by Step Installation Process

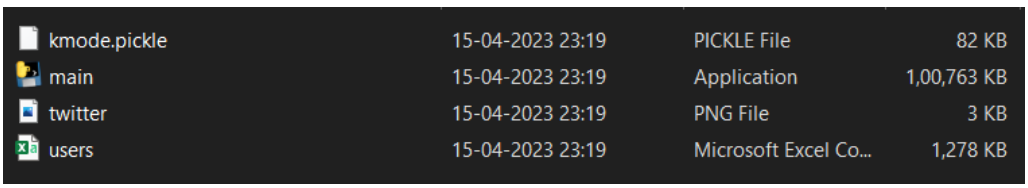
While our software is easy to install and run, we recommend that you closely follow the following steps such that the application runs smoothly.

1. Install the file capstine_exe.zip onto your local system.



2. Unzip the file. This will yield an executable file that already has all the required dependencies, and the user does not have to install anything separately. The user can simply use and run this independent file.
   This will yield four sub files consisting of the .exe file, a pickle file of the model, an image, and the data.

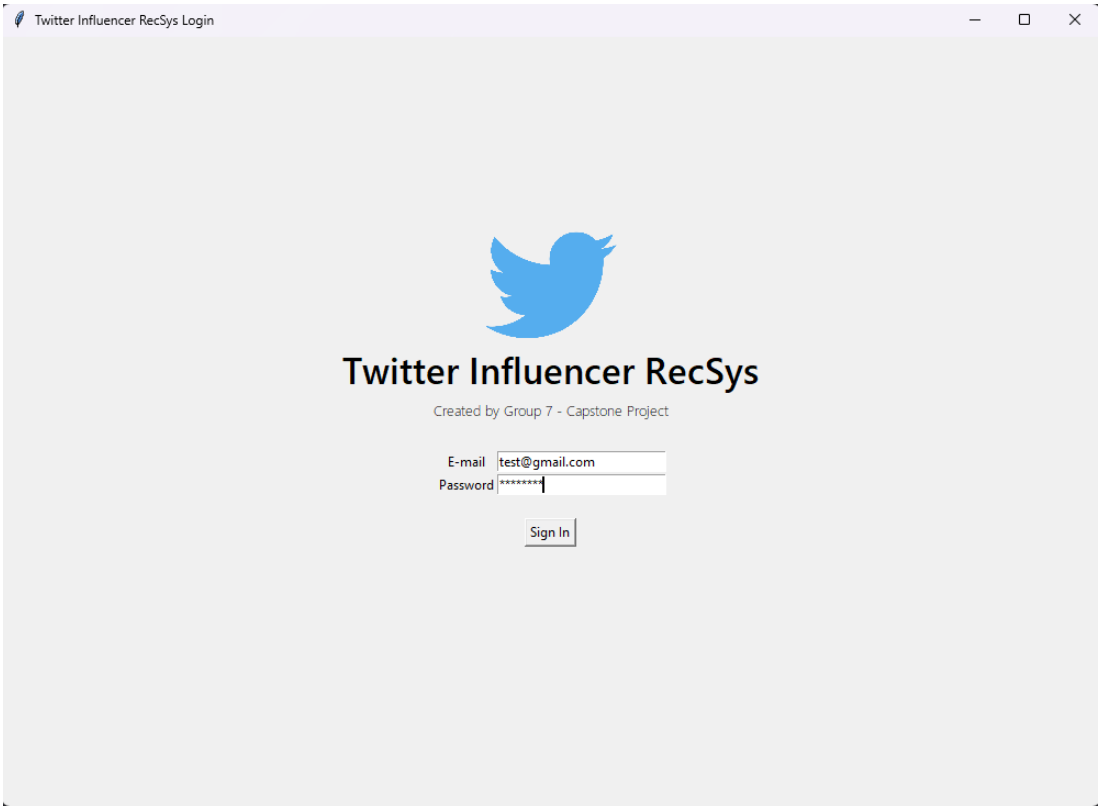




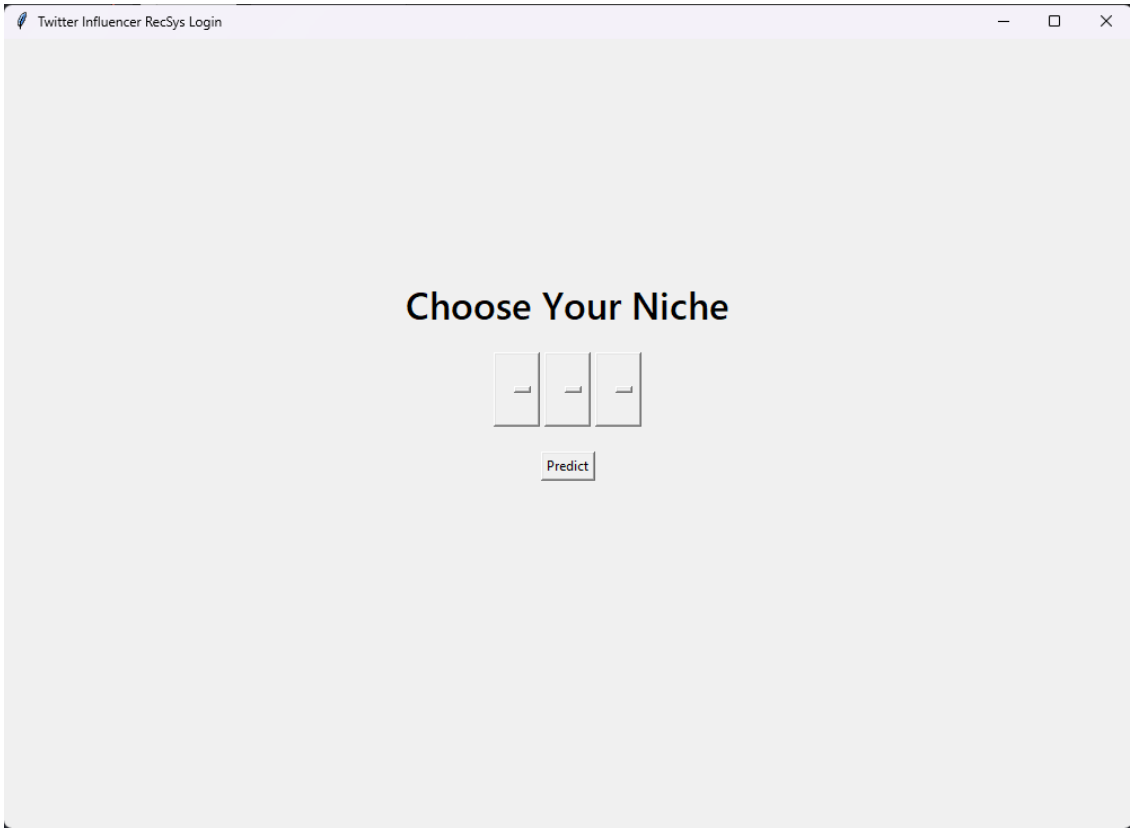
6. Manual of Use

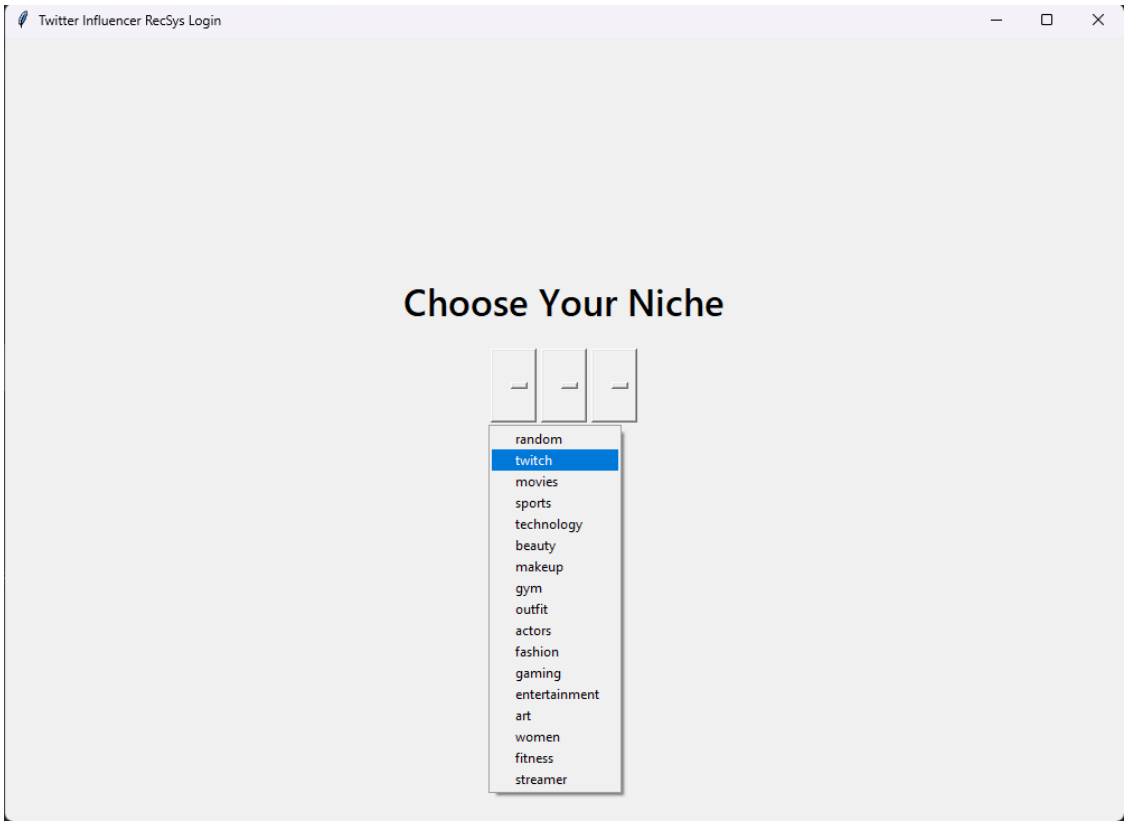
1. Run the main.exe file as it is. The command prompt will open, and this might take a few moments. This will launch the user interface.

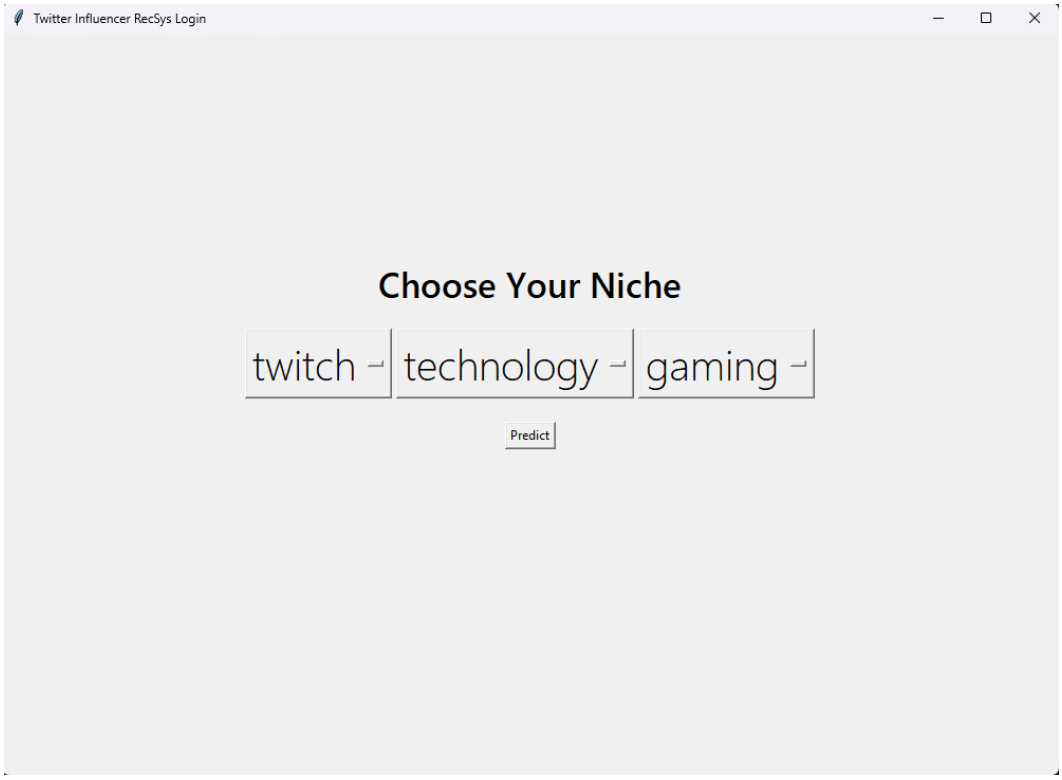
2. Enter the username and password as given below:
   E-mail: test@gmail.com
   Password: password
   (While this is a dummy model of the user interface and does not require a username password, we have included this section to make it look and feel more authentic and dynamic.)



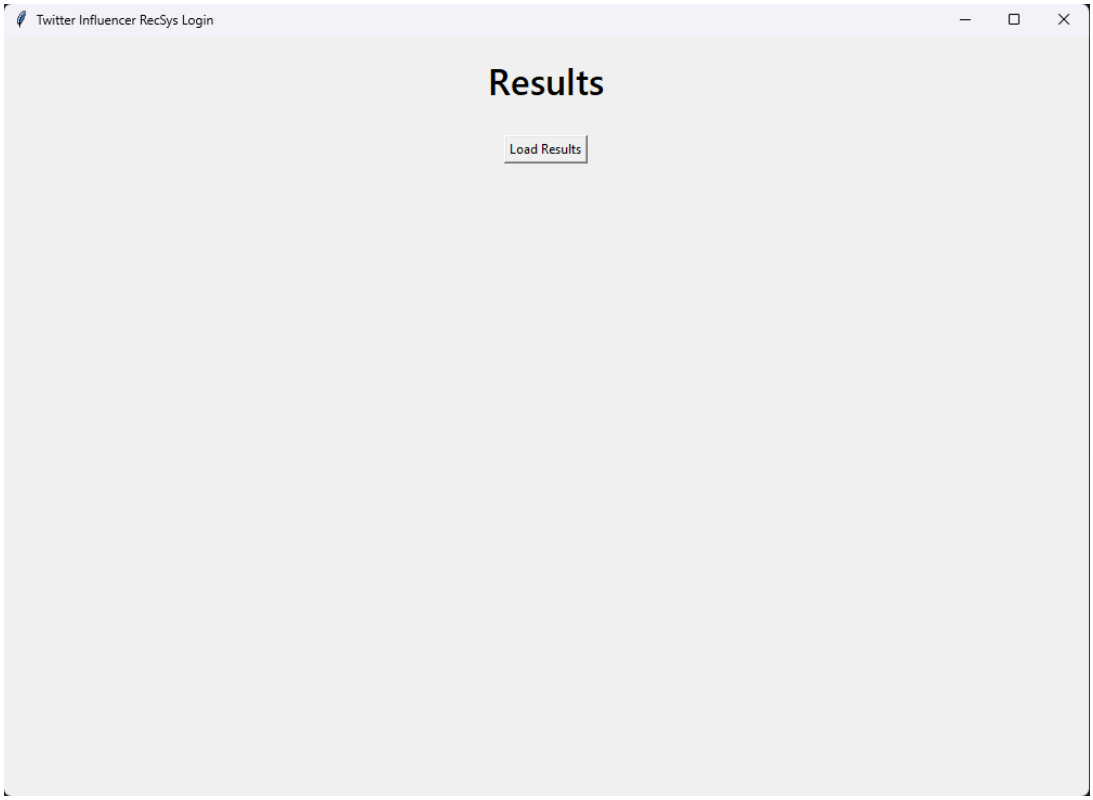
3. Using the dropdown menu select your desired niches. For example, we have selected “twitch”, “technology” and “gaming”. Then click “predict”.


Twitter Influencer RecSys Login
Choose Your Niche
Predict


Twitter Influencer RecSys Login
Choose Your Niche
random
twitch
movies
sports
technology
beauty
makeup
gym
outfit
actors
fashion
gaming
entertainment
art
women
fitness
streamer

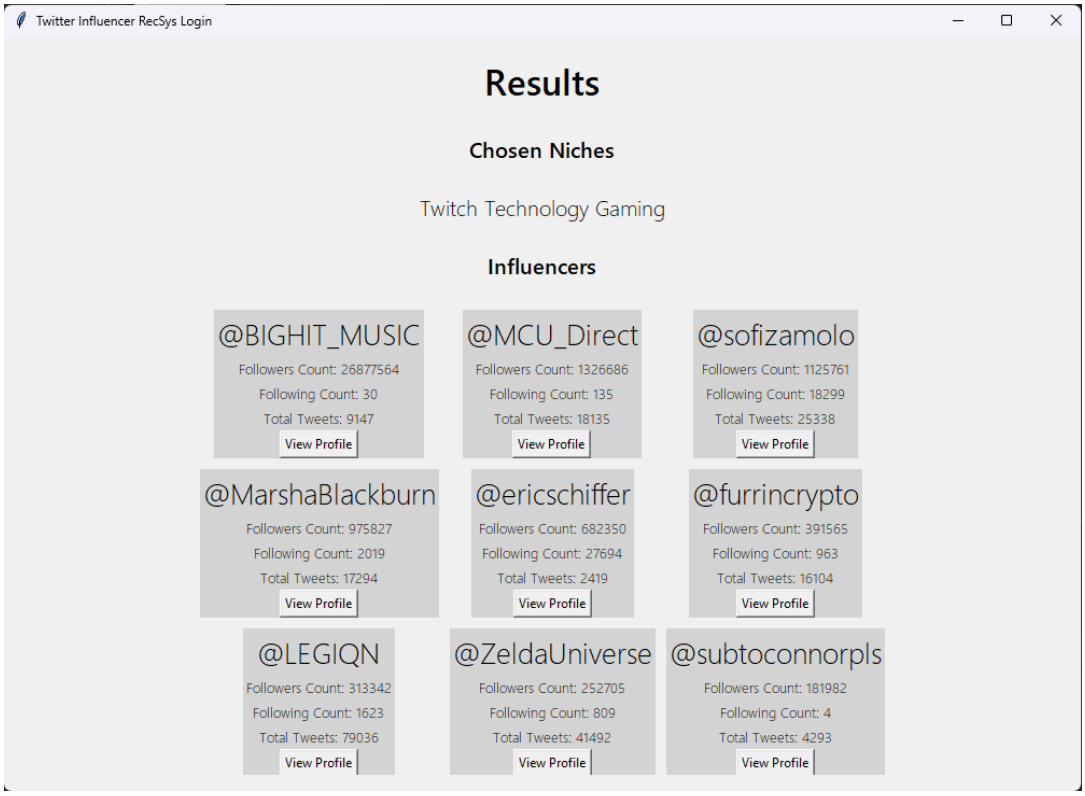


4. Next, click “Load Results”. This yields a page with the predicted niches and the users associated with the same.



5. You can now click on the influencers which will navigate you to their twitter profiles where other specifics and information about them such as followers count, following

count, tweets, etc. This gives our client a better idea and more data to make an informed decision from the list of recommended users.



As mentioned in the walkthrough video, the results are not always accurate, and many times do not represent a niche. However, being a recommendation system, we encourage our clients/users to make an informed decision based on an independent research.