INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Book Recommendation System with Sentiment Analysis

A Project Proposal by

Mr. S.C Fernando

Supervised by

Ms. Vayangi Ganepola

This Project Proposal is submitted in partial fulfilment of the requirements for the BSc (Hons) Software engineering degree at the University of Westminster

November 2022

I. Table Of Content

II. List Of Tables

III. List Of Figures

1. Introduction

In the age where the internet is the leading source of information recommendation systems are very common to be seen because of the abundance of data that is available. In this Research the author researches the currently available book recommendation systems and tries to find a improvement that can be done to significantly improve the performance of the system.

This document will discuss the identified problem, the significance of the research idea, the steps taken to approach the research, a review of existing work and the milestones and date of deliveries of the project.

2. Problem domain

2.1 NLP and Sentiment analysis

Natural language processing (NLP) is one of the core branches of machine learning that uses models to understand and analyze human language in depth. This technology is widely used in text to-speech, google assistant, text analysis to name a few. Sentiment analysis is one of the sub-branches of NLP where use cases for such a technology can be seen in the Amazon or eBay product recommendation system where it processes our likes and dislikes of certain products and recommends items, we are more likely to be interested in.

This technology is further used in analyzing online reviews on certain products and predicting the customers emotion towards that certain product.

2.2 Recommendation Systems with Sentiment Analysis

Recommendation Systems are very common in the current age of the internet, recommendation systems at the base level use multiple filters to filter data based on many factors such as the likes and dislikes of a user and give a recommendation of something the user may like (Huang, 2022). Systems such as these can be seen done on a large scale such as google ads which show ads of things the user searched previously on social media (Edelman Michael Ostrovsky Michael Schwarz et al., 2005), further streaming services such as Spotify recommends songs or playlists according to the users likes or dislikes uses a combination of machine learning and sentiment analysis (Anderson et al., 2020).

3. Problem Definition

With the abundance of data users produce on the internet daily gives an opportunity for machine learning models to analyze the data to reveal patterns that couldn't be seen before. Reading books in general, can be a very effective way to spend time or to learn new things , books sometimes have stories or information that is not readily available on the internet, books can also be a kind of therapy to brighten up the mood, but choosing a book can often be a daunting task for beginners because of the abundance of genres that are available for books, from the past surveys that have been done to understand the attitude of people towards reading 60% of the people replied positively (Fujimoto and Murakami, 2022) .

The Current research done for book recommendation systems goes over the use of sentiment analysis to identify the user's mood using their past social media activity and use that information to find a related book (Fujimoto and Murakami, 2022), in the current system there is room for improvement with the introduction of complex emotions and a machine learning model which could yield more accurate results than the current system.

3.1 Problem Statement

The existing book recommendation system (Fujimoto and Murakami, 2022)could benefit in increasing accuracy by implementing complex emotions and a machine learning-powered recommendation system.

4. Research Motivation

In the modern era where technology and social media are a part of the day to day life the author believes that taking a break from social media to read a book could be the perfect way to relax for a lot of people, due to the time it takes to find a book someone likes people drift away from reading, an improved book recommendation system could be the perfect solution for people trying to find a book they like.

5. Existing work

Citation	Brief Description	Limitations	Contribution
(Fujimoto and Murakami, 2022)	A sentiment analysis powered book recommendation system.	The analysis is done on a limited number of emotions.	An effective sentiment analysis power book recommendation system.
(Huang, 2022)	A Deep Learning powered recommendation system.	Low accuracy that could be improved.	Improved results compared to traditional algorithms.
(Kurmashov, Latuta and Nussipbekov, 2016)	A book recommendation system that gives fast qualitative recommendations.	Low accuracy of results.	Fast book recommendations.
(Sariki and Kumar, 2018)	A book recommendation system that recommends book according to the named entities present in the books.	The Efficiency could increase in the ranking method.	Ability to recommend books according to the named entity, ability to generate character lists of a book.
(Pera and Ng, 2011)	A book recommendation system that recommends books based on the user's social media friend's interest in books.	Relies solely on social media friend list data to give accurate results.	Outperforms the recommendation systems of Amazon and Library Things.

Table 1-Existing Work (Self – Composed)

6. Research Gap

After reviewing the existing literature, a handful of implementations of book recommendation systems were discovered. Implementations of a book recommendation system can be seen that centers around providing fast recommendations with minimum input data but yields slightly inaccurate results (Kurmashov, Latuta and Nussipbekov, 2016) , a problem that can be seen in this system is the priority of speed over accuracy in a book recommendation system, where low accurate results could affect in users disliking the book recommendation. Books in general, take time to read and people and most people spend only around 3.5 hours a week on average(Fujimoto and Murakami, 2022). In this Book recommendation system research (Fujimoto and Murakami, 2022) the paper talks about the use of sentiment analysis in finding the user's emotions and recommending books according to the emotion, an improvement in the existing can be done by adding a machine learning model for the book recommendation system and introducing complex emotions apart from the basic happy and sad emotions (Li, Li and Jin, 2020), this could be taken as a gap in research that could be further explored.

7. Contribution to the Body of Knowledge

Here we will be looking at the contributions this project will be doing to the existing work that has been done in the domain of recommendation systems.

7.1 Technological contribution

The current book recommendation system (Fujimoto and Murakami, 2022)depends on basic human emotions to give recommendations such as happy and sad, the result accuracy can be further improved using a variety of more complex emotions such as laughter, panic, joy, etc. (Li, Li and Jin, 2020). The recommendation system the current system uses a basic sentiment analysis system, but this could further improve if a machine learning model is used in combination with the standalone sentiment analysis model. The new model the author is proposing will yield higher accurate results compared to the currently existing models the new model will be having the capacity to update with any new book genres.

7.2 Contribution to the research domain

This research will directly contribute to the domain of computer science. Due to the increased accuracy and ease of use the new book recommendation system proposes this could attract non-readers and even readers to read more books which could inspire recommendation systems to be used in other fields of research, more research into recommendation systems could lead into the creation of less resource intensive and accurate models which could greatly benefit the domain of computer science and machine learning.

8. Research Challenge

This research project will be actively using machine learning recommendation systems and sentiment analysis. Recommendation systems have widespread use in many parts of the online advertisement industry such as google ads, YouTube ads, Spotify ads, amazon recommendations, etc.(Anderson et al., 2020). Sentiment analysis has a good track record of being used as a good marketing tool to find complex patterns in what the user likes and dislikes. A few decades prior machine learning involved calculations were done on very powerful large-scale supercomputer but will the passing years it has been possible to execute and train machine learning models in an average laptop computer, even will a huge arsenal of tools available today working and producing effective results with a machine learning model still holds complexity hence some of the possible challenges the author has discovered has been listed.

- Finding or creating a viable dataset that contains updated and enough data of books to satisfy the machine learning model to produce satisfactory results.
- Designing a machine learning recommendation model to work in conjunction with the sentiment analysis model.
- Finding appropriate language and libraries for the machine learning model and GUI.

9. Research question/s

RQ1: What are the problems faced by the existing book recommendation system?

RQ2: In what ways does the new system increase the speed and efficiency of the existing system?

RQ3: What kind of technologies/algorithms are used in the existing model?

RQ4: What are the design improvement that can be made in the new system?

RQ5: What testing methods will be used when testing the system?

10. Research Aim

This Research aims to design, develop, test and evaluate a system that recommends books by taking the users past activity on social media using sentiment analysis to find the users mood and giving recommendations of books powered by a machine learning model.

This research aims to design a system to recommend books to users with an up-to-date sentiment analysis model to find the users mood this information will be extracted from the users past social media activity. The recommendation system will be done using a machine learning recommendation algorithm which would recommend books from a premade library of books.

The Proof of concept for this research will be done by involving real users extracting their mood from their past activity and manually evaluating if the book recommendation they got is of their liking, this will help us evaluate if the system if performing up to expectations.

11. Research Objective

Research Objectives	Explanation	Learning Outcome	Research Questions
Problem Identification	Identifying a problem that haven't been addressed and needs to be solved. I. Research the current book recommendation system domain and identify any viable problem. II. Research and find how big the problem is and the scale of the problem is. III. Research what technological improvements that can be made in the new system.	LO1, LO4	RQ1

Literature Review	Reviewing the existing literature and finding any gaps. I. Analyzing and reviewing the existing model for any effective improvements that can be made. II. Understanding the size and effectiveness of the gap. III. Hypothesizing the effectiveness of the proposed system compared to the old system using existing literature.	LO1, LO4, LO3, LO6, LO8	RQ2
Data Gathering and Analysis	Gathering the required data needed for running the model. I. Analyzing the proposed model and collecting the needed datasets. II. Reading and understanding relevant sources of information if the datasets are being created. III. Reading existing papers related to book recommendation systems to find the procedures they took to find data or create. IV. Collecting the needed frameworks and libraries.	LO2, LO3, LO6	RQ3, RQ1
Research Design	Making the blueprint of the proposed system. I. Designing the blueprint of the system to collect the user's activity. II. Designing the blueprint of the recommendation model. III. Designing the blueprint of the user emotion detection model. IV. Designing the blueprint of the GUI.	LO3, LO2, LO4	RQ4
Implementation	I. Coding and implementing the system from the blueprints. II. Creating the system for collecting user activity from social media.	LO2, LO3, LO5, LO7	

	III. Creating the book recommendation model with the machine learning model. IV. Creating the model to give the emotion of the user from the retrieved user activity. V. Training the machine learning model with the dataset. VI. Creating the GUI.		
Testing and Evaluation	Testing and benchmarking the final prototype. I. Creating testcases for every functionality present in the system. II. Testing the system on mock data and in real-world data with users. III. Evaluating if the model is performing up to expectations and tuning it accordingly.	LO4, LO2, LO8	RQ5

Table 2-Research Objectives (Self-Composed)

12. Project Scope

12.1 In-scope

- A system that recommends books using the users past online activity.
- Adding a machine learning model to the old model that only uses sentiment analysis.
- Factoring in complex emotions to the already existing simple emotions such as happy, sad and neutral.
- Creating a system that collects the activity of the users in one type of social media.
- Having a large but limited library of books to use in the process of the machine learning model.
- GUI interface for the user to interact with to get recommendations.

12.2 Out-scope

- Additional advanced complex emotions for the sentiment analysis.

- Having all the books that have been published in the library.
- Machine learning recommendation algorithm that autotunes according to the users likes and dislikes of the previous recommendations the user got.
- Creating a less resource-intensive model and application.

12.3 Diagram showing prototype feature

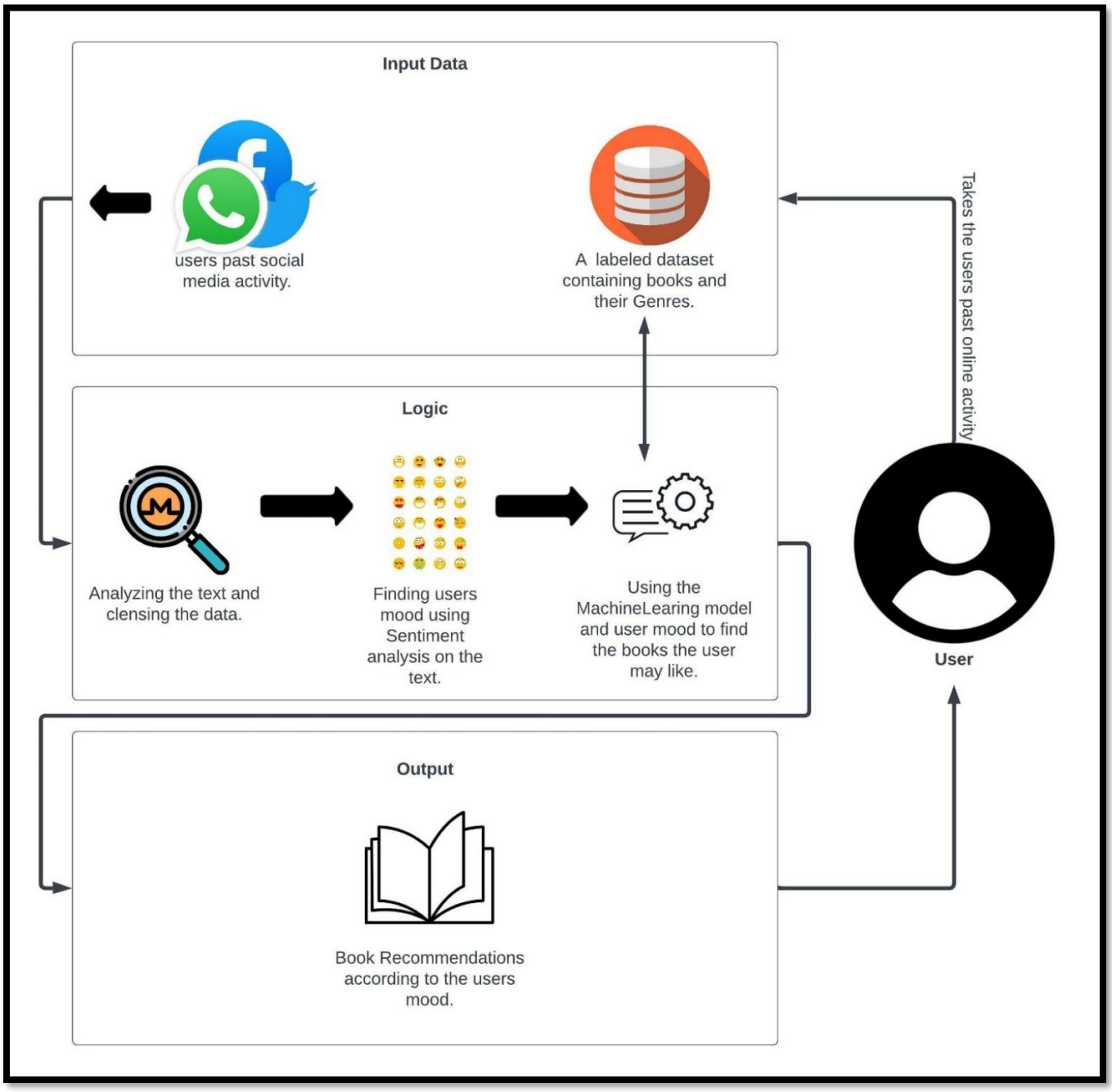

Figure 1-Prototype feature Diagram (Self-Composed)

13. Methodology

13.1 Research methodology

Research Philosophy	Among the 3 main research paradigms (Interpretivism, Positivism & Pragmatism) this project is better suited into the category of **pragmatism**(Goldkuhl, 2012)**.** This research involves both quantitative and qualitative data but weights mostly towards the use of **qualitative** data in the project itself with the collection of data from the survey's some of the data falls under the category of **quantitative** data.
Research Approach	A **deductive** research approach was chosen for this project, as this project aims to improve a pre-existing theory by increasing the accuracy, even while the recommendation system is altered by adding a machine learning model the basic theory of this research stays the same.
Research Strategy	The **survey** method is chosen as a way to collect the data for the requirement gathering, a mix of **quantitative** and **qualitative** will be collected in the survey. In conjunction with the survey, questionnaires will be performed which will strictly collect only quantitative data.
Research Choice	As discussed in the **research philosophy** out of the 3 paradigms **pragmatism** is chosen as this method is a mix of both **qualitative** and **quantitative** methods. The surveys and questionaries boast both quantitative and qualitative data thus making the pragmatism method more suited (Goldkuhl, 2012).
Time zone	Between **cross-sectional** or **longitudinal** time horizons, the **cross-sectional** method was chosen as data that is collected in a single point in time and this proves more effective due to the project requiring more up-dated data. Longitudinal methodology involves collecting

	data from an extended time frame which fits poorly with the approach of this project.

Table 3-Research Methodology (Self-Composed)

13.2 Development methodology

13.2.1 System development life cycle model (SDLC)

The one-person **Agile** SDLC model will be used as enables upgrades in the UI and models based on test, this model proves flexible in adding new and updating old features on the go. Plenty of changes will take place in the proposed project and the flexibility of the model will prove very beneficial.

13.2.2 Design methodology

Between the available design methodologies OODA (object-oriented analysis and design) and **SSADM** (Structured Systems Analysis & Design Method). **OODA** proved to be better due to its simple structure of the act, observe, orient and design which could easily adapt to rapid changes the project might go through, while SSADM is mainly focused on projects that have large groups of people working on it.

13.2.3 Requirement Elicitation Methodology

Questionnaires and surveys will act as the main way to collect the requirements that are needed for the prototype.

13.2.4 Testing methodology

Testing the ML models and the GUI will be important in recognizing the need for any tuning of the model and if the expected results are achieved, the testing for the prototype and model will be evaluated separately.

I. Prototype testing

The prototype or the GUI of the application will be hosted as a webpage online or as a standalone desktop application the quality of the app will be tested by:

- **Latency** – the network responsiveness webpage or application.
- **Load speed** – how fast the resources will load when the application or webpage opens.

II. Model testing

The sentiment analysis model and machine learning recommendation model will be tested on their output results:

- **Accuracy** – How accurate the output results are?
- **Precision** – How consistently can the model give accurate outputs?
- **Recall** – The number of correct predictions vs the number of predictions that should have been predicted correctly.

13.3 Project management methodology

13.3.1 Schedule

13.3.1.1 Deliverables

Deliverable	Date
Submission of Project Proposal (Final PP)	3rd November 2022
Submission of SRS	24th November 2022
Project Specifications Design and Prototype (PSDP)	2nd February 2023
Test and Evaluation Report	23rd March 2023
Final Project Report	27th April 2023
Final Viva Presentation	4th to 15th May 2023

Table 4-Deliverables (Self-Composed)

13.3.1.2 Gantt chart

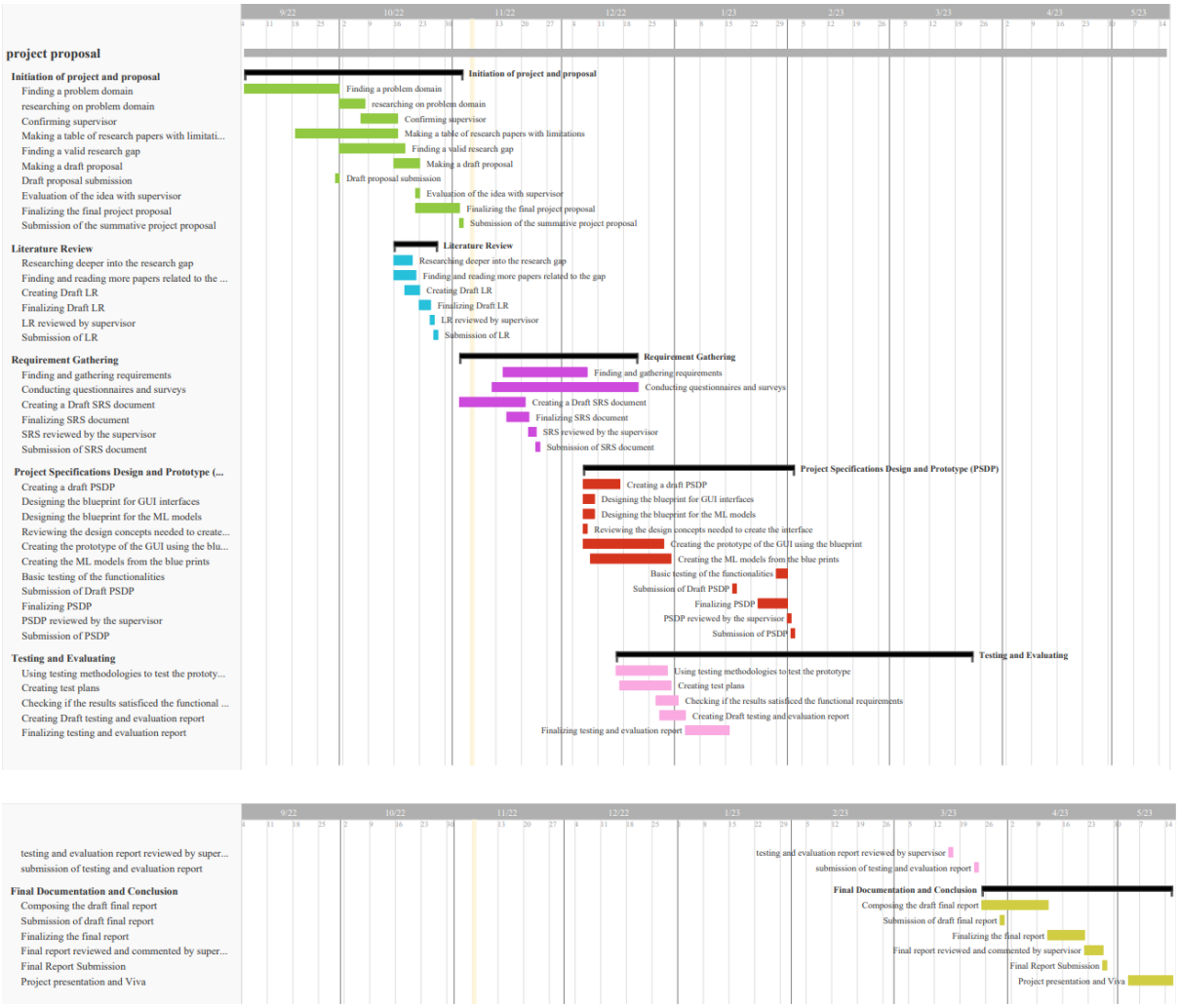

Figure 2 Gantt chart (Self-Composed)

13.3.2 Resource Requirements

With the size and scope of the project the resource requirements the identified requirements are listed, and the unidentified requirements are hypothesized as software, hardware, data and skill requirements.

13.3.2.1 Software Resources

- **Microsoft Word** – all the documents are created and exported from word.
- **Mendeley** – Citation management software to keep track of research papers, used in MS Word as a plugin.
- **IntelliJ IDEA/ VS code** – Will be used as the 2main IDE for coding the implementation.

- **Python** – Will be used as the main tool to clean, train and run the machine learning models with many available python data science frameworks.
- **React JS**- Will be used as the tool to design the frontend GUI application.
- **Java** – Will be used with spring boot framework to run REST API.
- **Gradient/Google Colab** – will be used as a cloud platform to train the ML models in case of a resource limitation.
- **Operating System (OS) Windows 10**- All the software and final prototype will be made and tested on the windows 10 system with a OS build of : 19045.2130.

13.3.2.2 Hardware Resources

- **Core i7 4th gen or higher** – To run and train large data sets with machine learning models.
- **8GB ddr4 dual channel memory** – To manage resource-intensive workload of the ML models.
- **20GB or more disk space** – A relatively large amount of space will have to be dedicated to datasets, code, resource material, models and training data including code for the front-end GUI application.

13.3.2.3 Data Requirements

- **Twitter data** -Data from Twitter API will have to be extracted.
- **ML model data**- Kaggle datasets, datahub.io or google open-source datasets related to books will be needed.

13.3.2.4 Skill Requirements

- Prior knowledge working with python and its machine learning frameworks.
- Basic knowledge of front-end application development in react.
- Knowledge in sentiment analysis and machine learning recommendation systems.

13.3.3 Risk and Mitigation

Risk Item	Severity	Frequency	Mitigation Plan
Loss or deletion of the code/work- Power cuts or disk corruption could severely impact the progress of a project in a negative way.	**5**	**1**	Saving multiple backups to the cloud and having a separate physical copy of the work.
Hardware limitations- Training very large datasets could be very resource intensive on average computers which would slow down the progress of the project.	**3**	**2**	With the help on google cloud services it's possible to run resource intensive machine learning models online.
Outdated software or frameworks- With new plugins and frameworks releasing everyday old plugins and frameworks could go out of service resulting inability to complete some functionalities.	**4**	**2**	Having multiple frameworks or plugins that does the same task in case of outdated software or incompatibility.

Table 5-Risk Management (Self-Composed)

References

Anderson, A. et al. (2020). Algorithmic Effects on the Diversity of Consumption on Spotify. *The Web Conference 2020 - Proceedings of the World Wide Web Conference, WWW 2020*. 20 April 2020. Association for Computing Machinery, Inc, 2155–2165. Available from https://doi.org/10.1145/3366423.3380281.

Edelman Michael Ostrovsky Michael Schwarz, B. et al. (2005). *NBER WORKING PAPER SERIES INTERNET ADVERTISING AND THE GENERALIZED SECOND PRICE AUCTION: SELLING BILLIONS OF DOLLARS WORTH OF KEYWORDS Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords*. Available from http://www.nber.org/papers/w11765.

Fujimoto, T. and Murakami, H. (2022). A Book Recommendation System Considering Contents and Emotions of User Interests. 23 September 2022. Institute of Electrical and Electronics Engineers (IEEE), 154–157. Available from https://doi.org/10.1109/iiaiaai55812.2022.00039.

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, 21 (2). Available from https://doi.org/10.1057/ejis.2011.54.

Huang, G. (2022). E-Commerce Intelligent Recommendation System Based on Deep Learning. *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2022*. 2022. Institute of Electrical and Electronics Engineers Inc., 1154–1157. Available from https://doi.org/10.1109/IPEC54454.2022.9777500.

Kurmashov, N., Latuta, K. and Nussipbekov, A. (2016). Online book recommendation system. *Proceedings of the 2015 12th International Conference on Electronics Computer and Computation, ICECCO 2015*. 23 February 2016. Institute of Electrical and Electronics Engineers Inc. Available from https://doi.org/10.1109/ICECCO.2015.7416895.

Li, Z., Li, R. and Jin, G. (2020). Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary. *IEEE Access*, 8, 75073–75084. Available from https://doi.org/10.1109/ACCESS.2020.2986582.

Pera, M.S. and Ng, Y.K. (2011). With a little help from my friends: Generating personalized book recommendations using data extracted from a social website. *Proceedings - 2011 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2011*. 2011. 96–99. Available from https://doi.org/10.1109/WI-IAT.2011.9.

Sariki, T.P. and Kumar, B. (2018). *A book recommendation system based on named entities*.