

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. Vishmi Ganepola

This Project Proposal is submitted in partial fulfilment of the requirements for the BSc

(Hons) Software engineering degree at the University of Westminster

**October 2022**

© The copyright for this project and all its associated products resides with Informatics Institute of Technology

Contents

[1. Introduction 4](#_Toc118222370)

[2. Problem domain 4](#_Toc118222371)

[2.1 NLP and Sentimental analysis 4](#_Toc118222372)

[2.2 Recommendation Systems with sentiment analysis 4](#_Toc118222373)

[3. Problem Definition 5](#_Toc118222374)

[3.1 Problem Statement 5](#_Toc118222375)

[4. Research Motivation 5](#_Toc118222376)

[5. Existing work 6](#_Toc118222377)

[6. Research Gap 6](#_Toc118222378)

[7. Contribution to the Body of Knowledge 7](#_Toc118222379)

[7.1 Technological contribution 7](#_Toc118222380)

[7.2 Contribution to the research domain 7](#_Toc118222381)

[8. Research Challenge 8](#_Toc118222382)

[9. Research question/s 8](#_Toc118222383)

[10. Research Aim 9](#_Toc118222384)

[11. Research Objective 9](#_Toc118222385)

[12. Project Scope 10](#_Toc118222386)

[12.1 In-scope 10](#_Toc118222387)

[12.2 Out-scope 10](#_Toc118222388)

[12.3 Diagram showing prototype feature 11](#_Toc118222389)

[References 14](#_Toc118222390)

# Introduction

In the age where 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 on the existing work and the milestones and date of deliveries of the author.

# 2. Problem domain

## 2.1 NLP and Sentimental analysis

Natural language processing also known as NLP are one of the core branch in 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 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 a base level uses multiple filters to filter data based on many factors such as likes and dislikes of a user and gives a recommendation of something the user may like (Huang, 2022). Systems such as these can be seen done in a large scale such as google ads which shows 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 it gives opportunity for machine learning models to analyze the data to revel 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 system goes over the use of sentiment analysis to identify the users 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 is 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, a 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 amount of emotions. | A 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. |
|  |  |  |  |
|  |  |  |  |

# 6. Research Gap

After reviewing the existing literature a handful of implementations of book recommendation systems were discovered. Implementations of 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 effect in users disliking the book recommendation. Books in general takes time to read and people and majority of 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 users emotion and recommendation books according to the emotion, a 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 have 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 which 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 update with any new book genres.

## 7.2 Contribution to the research domain

This research will be directly contributing in 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 been 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 a 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 have a good track record for being used as a good marketing tool to find complex patterns in what the user likes and dislikes. Few decades prior machine learning involved calculations were done on very power large scale super computer but will the passing years it has been possible to execute and train machine learning models in a 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 datasets that contains updated and enough data of books to satisfy the machine learning model to produce satisfiable results.
* Designing a machine learning recommendation model to work in conjunction with the a sentiment analysis model.
* Finding appropriate language , library 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 from the existing system?

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

# 10. Research Aim

*This Research aims to develop a system that recommends books by taking the users past activity in 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 a 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 the 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 |
| Problem Identification | Identifying a problem that haven’t been addressed and needs to be solved.   1. Research the current book recommendation system domain and identify any viable problem. 2. Research and find how big the problem is and the scale of the problem is. 3. Research what technological improvements that can be made in the new system. | LO1 |
| Literature Review | Reviewing the existing literature and finding any gaps.   1. Analyzing and reviewing the existing model for any effective improvements that can be made. 2. Understanding the size and effectiveness of the gap. 3. Hypothesizing the effectiveness of the proposed system compared to the old system using existing literature. | LO1 |
| Data Gathering and Analysis | Gathering the required data needed for running the model.   1. Analyzing the proposed model and collecting the needed datasets. 2. Reading and understanding relevant sources of information if the datasets are been created. 3. Reading existing papers related to book recommendation systems to find the procedures they took to find data or create. 4. Collecting the needed frameworks and libraries. | LO2, LO3 |
| Research Design | Making the blueprint of the proposed system.   1. Designing the blueprint of the system to collect the users activity. 2. Designing the blueprint of the recommendation model. 3. Designing the blueprint of the user emotion detection model. 4. Designing the blueprint of the GUI. |  |
| Implementation | 1. Coding and implementing the system from the blueprints. 2. Creating the system for collecting user activity from social media. 3. Creating the book recommendation model with the machine learning model. 4. Creating the model to give the emotion of the user from the retrieved user activity. 5. Training the machine learning model with the dataset. 6. Creating the GUI. |  |
| Testing and Evaluation | Testing and benchmarking the final prototype.   1. Creating testcases for every functionality present in the system. 2. Testing the system on mock data and in real-world data with users. 3. Evaluating if the model is performing up to expectations and tuning it accordingly. |  |

# 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 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 the 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

Diagram

Description automatically generated

# 13. Methodology

## 13.1 Research methodology

|  |  |
| --- | --- |
| Research Philosophy | The author of the research has selected the positivism as the research philosophy |
| Research Approach | Deductive or inductive why? |
| Research Strategy | Experiment, survey => questionnaire (can be quantitative or qualitative) or interview (can be quantitative or qualitative), |
| Research Choice | Mono method => only one method can quantitative (Positivist) or qualitative (interpretivist), Multi method (More than one method but all belong to same paradigm (positivist or interpretivist)) or Mixed method (only pragmatist can mix the method => mixing the method from positivism and interpretivism) |
| Time zone | Cross-sectional or longitudinal |
|  |  |
|  |  |
|  |  |

## 13.2 Development methodology

* 1. **What is the life cycle model and why?**
  2. **Design methodology => SSADM or OOAD or Anything else?**
  3. **Evaluation methodology => Evaluation metrics and/or benchmarking**

## 13.3 Project management methodology

* 1. **Schedule using the Gantt Chart after doing a WBS (Do not have to provide the WBS)**
  2. **Deliverables, milestones and dates of deliverables**
  3. **Resource requirements** 
     1. **Hardware requirements**
     2. **Software requirements**
     3. **Skills requirements**
     4. **Data Requirements**
  4. **Risk Management**

|  |  |  |  |
| --- | --- | --- | --- |
| **Risk Item** | **Severity** | **Frequency** | **Mitigation Plan** |
|  | **5** | **5** |  |
|  | **5** | **4** |  |
|  | **5** | **1** |  |
|  |  |  |  |

# 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.

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.