**E-BOOKS RECOMMENDATION SYSTEM USING REVIEWER RATINGS AND FEEDBACK**

**PROJECT REPORT**

***Submitted by***

# K.SASI KIRAN

**(2019202049)**

***submitted to the faculty of***

**INFORMATION AND COMMUNICATION ENGINEERING**

***In partial fulfillment of the requirement for the degree***

***of***

**MASTER OF COMPUTER APPLICATIONS**



**DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY COLLEGE OF ENGINEERING, GUINDY**

**ANNA UNIVERSITY**

**CHENNAI 600025**

**May 2022**

# ANNA UNIVERSITY, CHENNAI

## BONAFIDE CERTIFICATE

Certified that this Report titled “**EBooks recommendation system using reviewer ratings and feedback**” is the bonafide work of **K.Sasi Kiran (2019202049)** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**PLACE**:CHENNAI KANIMOZHI.S

**DATE**: 02/05/2022 TEACHING FELLOW

PROJECT GUIDE

DEPARTMENT OF IST,CEG

ANNA UNIVERSITY

CHENNAI 600025

**COUNTERSIGNED**

**Dr. S. SRIDHAR**

**HEAD OF THE DEPARTMENT**

**DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY**

**COLLEGE OF ENGINEERING, GUINDY**

**ANNA UNIVERSITY**

**CHENNAI 600025**

**ACKNOWLEDGEMENT**

Its my privilege to express my sincere thanks to my project guide **Ms.S.KANIMOZHI**, TEACHING FELLOW, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for her keen interest, inspiring guidance, constant encouragement and support with my work during all the stages, to bring this thesis into fruition.

I deeply express my sincere thanks to **Dr. S. SRIDHAR**, Professor and Head of the Department, Department of Information Science and Technology, College of Engineering, Guindy, Anna University, Chennai for extending support.

I would like to express my sincere thanks to the project committee members, PROFESSOR Dr.SASWATI MUKHERJEE, Dr.M.VIJAYALAKSHMI, ASSOCIATE PROFESSOR, Dr.E.UMA, ASSISTANT PROFESSOR, MS.P.S.APIRAJITHA, TEACHING FELLOW, MS.C.M.SOWMYA, TEACHING FELLOW, Department of Information Science and Technology, Anna University, Chennai for giving their valuable suggestions, encouragement and constant motivation throughout the duration of my project.

K.SASI KIRAN

## TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.**  1 | **TITLE**  **ABSTRACT**  **INTRODUCTION** | **PAGE NO.**  7 |
|  | 1.1 General | 8 |
|  | 1.2 Problem Statement | 9 |
|  | 1.3 Objective | 9 |
| 2 | **LITERATURE REVIEW** |  |
|  | 2.1 Data Preprocessing | 10 |
|  | 2.2 Word Mapping  2.3 Filtering user feedback  2.4 Recommendation System | 10  11  11 |
|  |
| 3 | **OVERALL ARCHITECTURE**  3.1 Architecture design | 12 |
|  | 3.2 Architecture Explanation | 13 |
|  | 3.3 List of Modules | 16 |
|  | 3.4 Modules Explanation  3.4.1 Data Preprocess Module  3.4.2 Sentiment Analysis Module | 17  17  18 |
| 4 | **RESULTS** | 19 |
| 5  6 | **CONCLUSION AND FUTURE WORK**  **REFERENCES** | 23  24 |

# LIST OF FIGURES

[3.1 **E-Book Recommendation System Architecture**](#_bookmark17) 11

3.4.1 **Data Preprocess Module** 17

3.4.2 **Sentiment Analysis Module** 18

# LIST OF ABBREVIATIONS

SO Subjectivity – Objectivity Polarity

PN Positivity – Negativity Polarity

# ABSTRACT

The e-books application is built for recommending different books, which the user is interested in. The dataset is collected and it is preprocessed so as all the null values and duplicated values also any unwanted fields are removed. The books are separated into categories. The dataset is studied and using manual analysis, few categories of books are selected. Genre count is identified and it is needed for the analysis. The books are recommended majorly in two ways, by taking the ratings into consideration and by taking the feedback into consideration. If the user gives direct books name as the input, then the user will be getting the recommendations based on the ratings of the recommended books (high to low). If the user gives set of words as the input, those words are extracted and a keyword is found which matches with the genre. If the keyword is present in the genre, then another set of books are recommended and they are shown based on the sentiment analysis, positivity to negativity. This type of recommendation is to be proposed in this project to bring out the user’s feedback into consideration in the time of search for the book itself.

***Keywords:*** Preprocessing data, reviewer ratings, feedback (positive/negative), Sentiment analysis, Recommendation system

# CHAPTER 1

# INTRODUCTION

**1.1 General**

* Recommender systems aim to predict users' interests and recommend product items that quite likely are interesting for them.
* They are among the most powerful machine learning systems that online retailers implement in order to drive sales.
* The impact is very huge which helps in marketing the products worldwide.
* Recommender systems are beneficial to both service providers and users.

The dataset is pre-processed and stored, also the sentiment analysis is made for getting the feedback of the user into consideration. Sentiment analysis is a technique in natural language processing that aims to retrieve the "sentiment" of a piece of text—positive, negative, or neutral. This is an easy way of summarizing the contents of a piece of text, and one that is easily understood.

Note, however, that sentiment analysis is a difficult problem. Humans agree on the sentiment of sentences only 80% of the time, and the best classifiers can get around that level of accuracy, but we're going to just use a built-in analyser in the NLTK (Natural Language Toolkit) python library.

***Technologies****:*

Programming Language : Java, Python

Front-end Technologies: HTML, CSS and JS

Algorithms: Sentiment Analysis

DataBase : MySQL

IDE : NetBeans

Framework: Python jupyter notebook

* 1. **Problem Statement**

The shortcomings of content based and collaborative based recommendation system are degrading the system. The user satisfaction is very mere when these algorithms are being used. More number of products are not been recommended, also they’re not efficient and faster. For a good recommendation to happen the ancient techniques is not much supporting, so in order to consider the user feedback also, this recommendation system helps. Sentiment analysis also will be of great help in order to categorize the review as positive, neutral or negative review and get the probability values for each columns and make analysis based on those values.

* 1. **Objective**

The objective of recommender systems is to provide recommendations based on recorded information on the users’ reviews and ratings. This system is proposed for two cases. The motive to take the two cases is to differentiate the way of getting the recommended books (for ratings and for feedback). The case 1 has user input as the direct book name itself. In this case the recommender system will process and brings out the recommending books with ratings as the categorization. The case 2 has user input as set of words (sentence), here the keyword is extracted and matched, if the keyword is found in the genre array, then the recommender system will recommend books based on genre and feedback as the categorization. This is the overall objective of this system.

**CHAPTER 2**

**LITERATURE REVIEW**

This Chapter explains about the literature survey made on the existing system, analyzing the problem statements and issues with the existing system and proposed objectives for the new system.

**2.1 Data Preprocessing**

Namrata Chaudhary proposed the general Sources of Errors and Data Cleaning Strategies in which python jupyter is used in order to perform data cleaning using pandas library. Data should be preprocessed before storing them onto a database to reduce complexity and noise of raw data. After the dataset collection, it is cleaned and processed for analysis. Also, it is processed in such a way that the dataset contains only those rows which are needed and eliminating the null values etc.

**2.2 Word Mapping**

Keywords are words that come to you naturally, or that may be part of a specific discipline vocabulary (e.g. terms used only by midwives), or that you brainstorm when planning your search. If the 'map term to subject heading' box is ticked the database will attempt to map your keyword or phrase to a subject heading in the database’s thesaurus. If the ‘map term to subject heading’ box is not ticked, the word or phrase will be treated as a keyword. Searching by keyword finds only those results where your keyword appears as an exact match in several fields including the title or abstract. This works particularly well if you are looking for a specific spelling, product, term, or phrase. It is important to recognise that different spelling and terminology may exist for the same search topics. Keyword searching will not differentiate between spellings.

**2.3 Filtering User Feedback**

This approach is nothing but to get the positive, neutral and negative of the user feedback and to filter them based on their matter of behaviours. The creators of the comments share their opinions on different topics, discuss current issues even spot accidents or any flu outbreaks. These are the valuable source of opinions and sentiments as huge amount of posts are posted by the users according to their used products and services, or express their different views on different perspectives. Researchers are using these posts to measure the public sentiment and to do sentiment analysis. They are trying to determine the “PN-polarity” of subjective terms i.e, identifies whether a term expresses the opinion which could have positive or negative connotation.

**2.4 Recommendation System**

[Michael J. Pazzani](https://link.springer.com/chapter/10.1007/978-3-540-72079-9_10#auth-Michael_J_-Pazzani) discusses content-based recommendation systems, i.e., systems that recommend an item to a user based upon a description of the item and a profile of the user’s interests. Content-based recommendation systems may be used in a variety of domains ranging from recommending web pages, news articles, restaurants, television programs, and items for sale. Although the details of various systems differ, content-based recommendation systems share in common a means for describing the items that may be recommended, a means for creating a profile of the user that describes the types of items the user likes, and a means of comparing items to the user profile to determine what to re commend. The profile is often created and updated automatically in response to feedback on the desirability of items that have been presented to the user.

**CHAPTER 3**

# 3.1 EBook Recommendation System Architecture

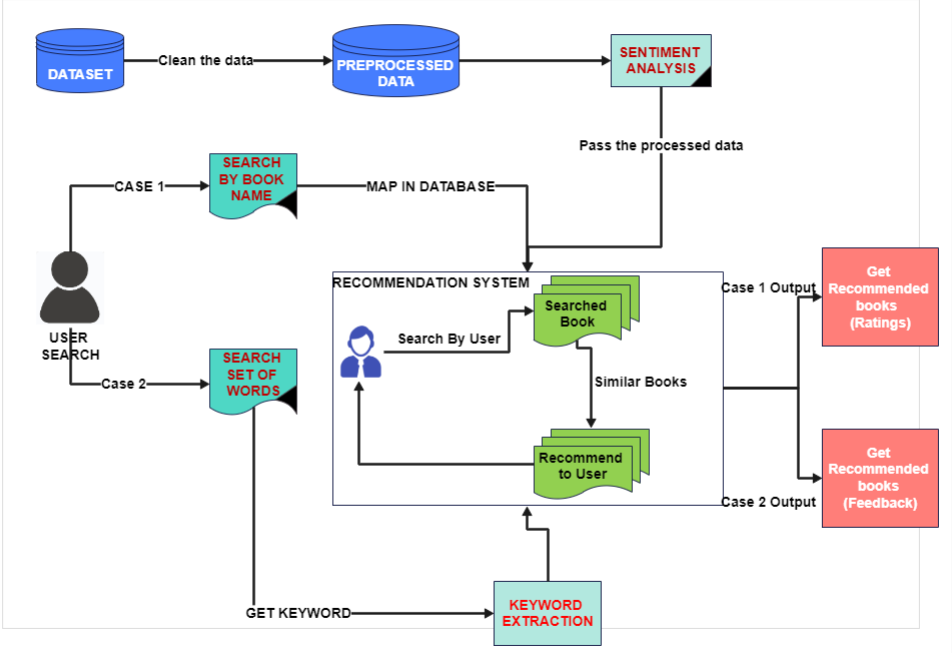


Fig1 : E-Book Recommendation system Architecture diagram

# 3.2 ARCHITECTURE EXPLANATION

# DATASET:

# The dataset for books should be collected from kaggle, it contains books.csv, users.csv and ratings.csv files. The size of dataset is large enough to make our implementations. Each csv file consists of more than 2 lakh records which consists of all the information regarding the books, users and their ratings.

# PROCESSED DATA:

# After the dataset collection, it is cleaned and processed for the analysis. Also it is processed in such a way that the dataset contains only those rows which are needed and eliminating the null values etc. After the analysis and filtering out the unwanted data, we are left with only forty thousand rows of information. In Fig1, it clearly explains that the combined with the users and ratings regarding to the book “isbn” and the “user-id”.

# SENTIMENT ANALYSIS:

# The purpose of sentiment analysis is based on the two sectors:

# Classifying Documents

# Classifying documents or any passages according to sentiment orientation such as positive vs. negative.

# Gathering Information

# Extracting information of opinions which contains information of particular aspects of interest and the corresponding sentiment orientation in a structured form from a set of unstructured data. The tasks of classifying documents of the sentiment analysis can be divided into three sub-tasks:

# Identifying SO polarity:

# Whether the comment or post is referring a situation or event without disclosing the subjectivity (positive or negative opinion) on it or expressing opinion on its subject matter. Briefly, it means that identify the subjective or objective polarity of a post or comment.

# Identifying PN-polarity:

# Whether a subjective post or comment is expressing positive or negative.

# Identifying the Degree of PN-polarity:

# This step gives the impression of the degree of positivity or negativity on that opinion. Positivity could be weakly positive, mildly positive or strongly positive and same could be for the negative opinion.

# "I think this is going to be one of the most important datasets of this era, because we are looking at what people are talking about in real time at the scale of an entire society," says Mislove, an assistant professor of computer science at Northeastern University. So, sentiment analysis is doing the right thing as he told

# ARCHITECTURE EXPLANATION:

The User will be able to search the books by two ways:

1. *Search the books direct by giving the book name* : In this case, the user will be allowed to enter any book name of his own, it is then checked if present in the database. If the book is present in the database, then that particular book name is parsed into the recommendation system, where the name of the book is taken and related books are recommended for the user.
2. *Search the book indirect by giving set of words*: In this case, the user will be entering a set of words or like a sentence, that sentence is parsed into the keyword extraction algorithm and the keyword is extracted which is present in the genre array. The genre is considered as array with unique values. This keyword, after extraction, if matches in the array, it then gets that particular genre and passes into the recommendation system. As mentioned earlier, the book genre is taken and related genre books are recommended for the user.

In Fig1, for case 1, the user will get books displayed in hierarchical order of ratings (high to low). And also in Fig1, for case 2, the user will be getting the books displayed in the order of positive, neutral and negative feedback which is parsed and taken in the sentiment analysis.

# 3.3 LIST OF MODULES

* Dataset Preprocessing.
* Sentiment Analysis.
* Search by name.
* Search by set of words.
* Recommendation System.

# LIST OF COMPLETED MODULES

**Data Preprocessing :**

***Input*** *:* All three (Books, Users, Ratings) CSV files.

Process the data, manipulate it, clean it using Pandas, Numpy in jupyter platform.

***Output*** *:* Cleaned and processed Data which is loaded into another CSV file

**Sentiment Analysis :**

***Input :*** The preprocessed csv file.

Import nltk sentiment vader library and loop the column for which the polarity needs to be given to. The two functions, one for stop words to detect the language and the other function for splitting up of the words. The polarity is found for positivity, neutral and negativity for the feedback column and combined with the csv file with respect to the User-ID.

***Output :***  The polarities for each column is found and they’re included in the dataset, i.e. compound, positive, neutral and negative columns

# 3.4 MODULES EXPLANATION

**3.4.1 DESCRIPTION OF DATA PREPROCESS MODULE:**

1. **Data Cleaning**

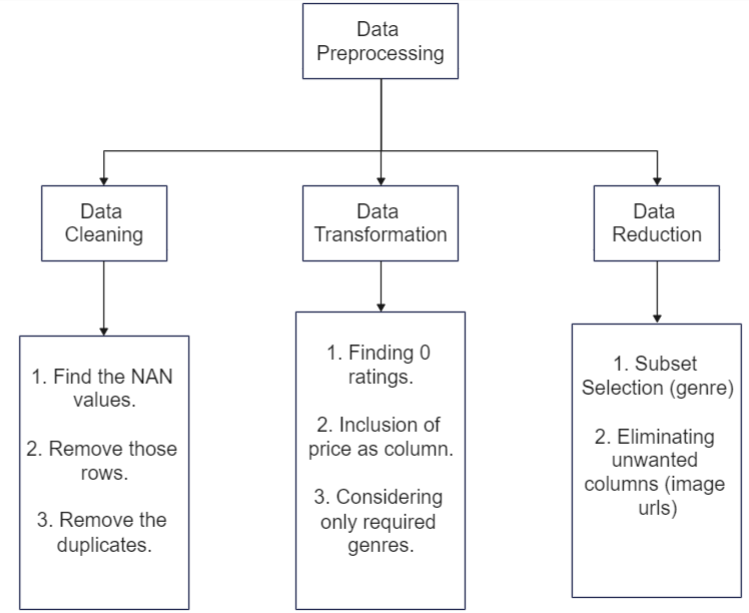
The CSV files will have more null values, so they must be eliminated. Remove the duplicates (rows and columns), this is what explained in Fig2.

1. **Data Transformation**

Find the zero ratings data and include a separate column as price. Genre column has multi label genres, so they must be eliminated which comes under data reduction part in Fig2.

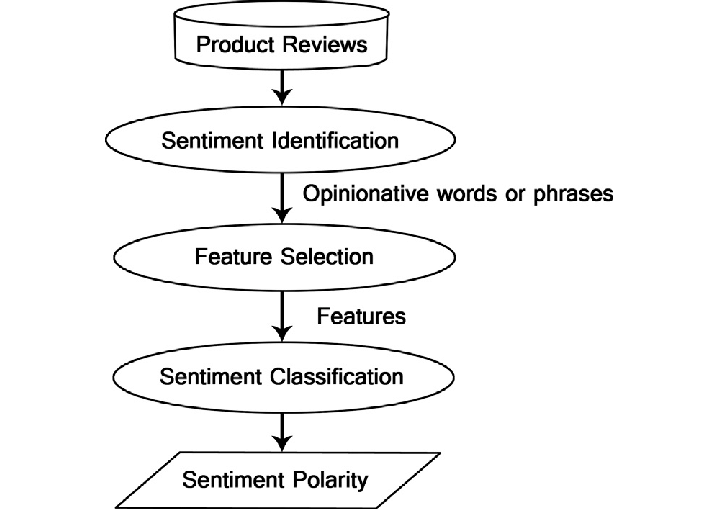
1. **Data Reduction**

For any data study to make, the unwanted rows/columns must be deleted. Also the subset values must be eliminated.





**3.4.2 DESCRIPTION OF SENTIMENT ANALYSIS MODULE:**



Sentiment Analysis is the computational study of people’s opinions, attitudes and emotions toward an entity. In Fig3, Sentiment Analysis identifies the sentiment expressed in a text then analyzes it. Therefore, the target of SA is to find opinions, identify the sentiments they express, and then classify their polarity. The data sets used in SA are an important issue in this field. The main sources of data are from the product reviews. These reviews are important to the business holders as they can take business decisions according to the analysis results of users’ opinions about their products. The reviews sources are mainly review sites. SA is not only applied on product reviews but can also be applied on stock markets, news articles,  political debates. They are also used as data sources in the SA process.

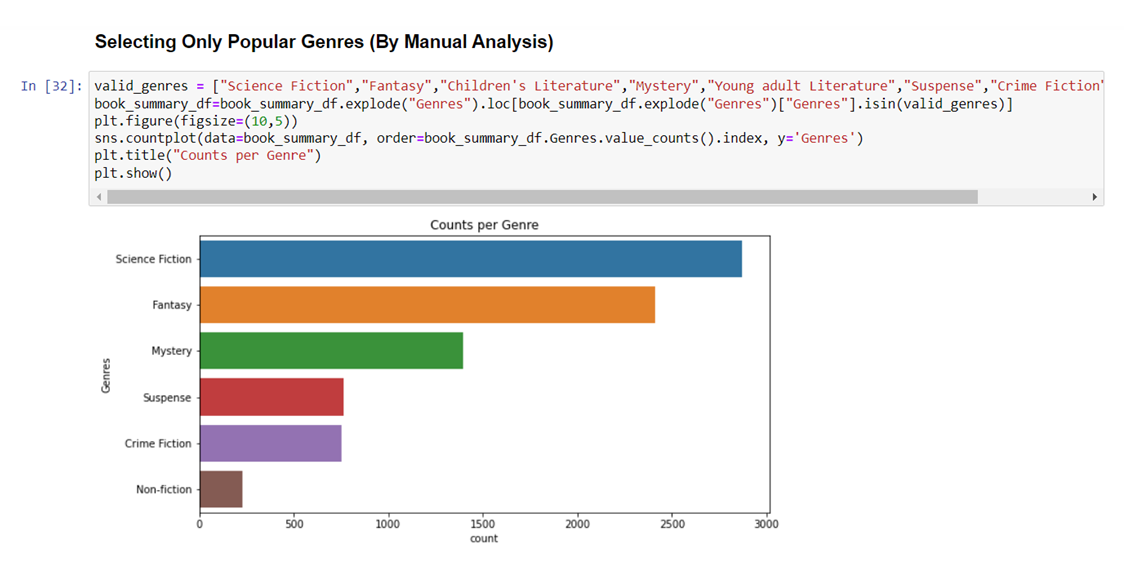
In the preprocessed data, we need to identify the reviews of each book, whether they’re positive or neutral or negative. This is done using the sentiment analysis, it makes our task easy by just applying the algorithm to the dataset and getting the required results.

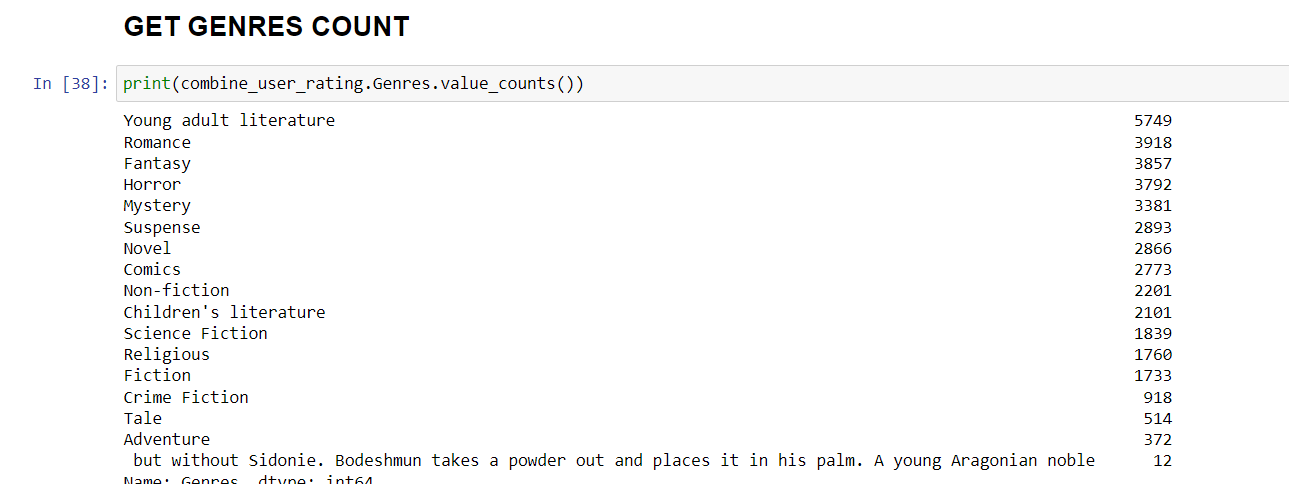
**CHAPTER 4**

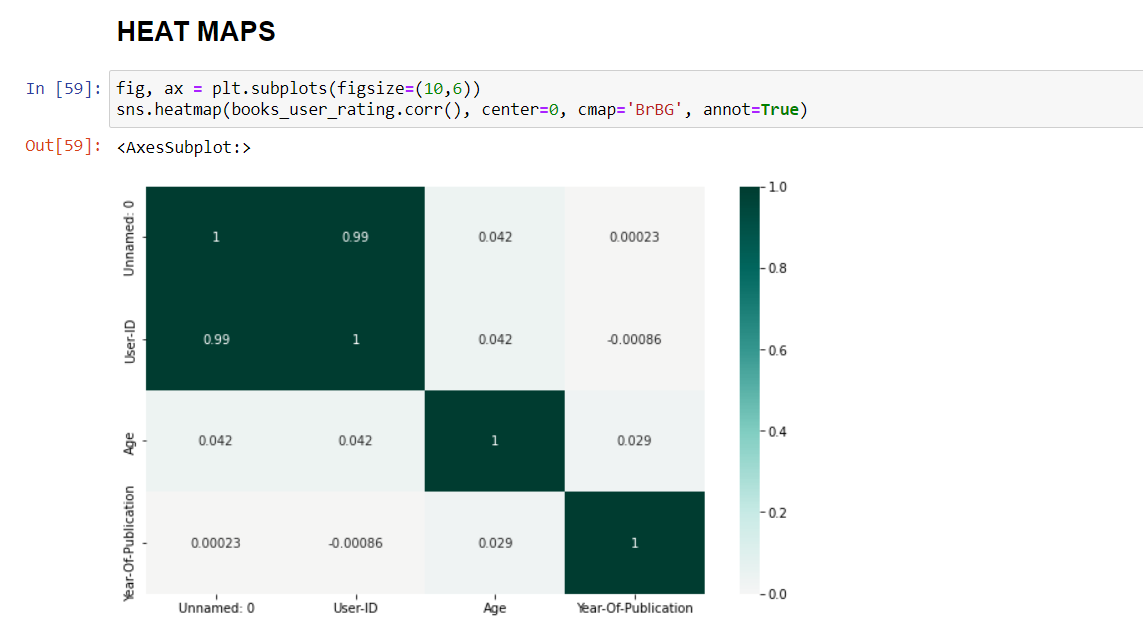
**RESULTS**

***Data Preprocessing***

****

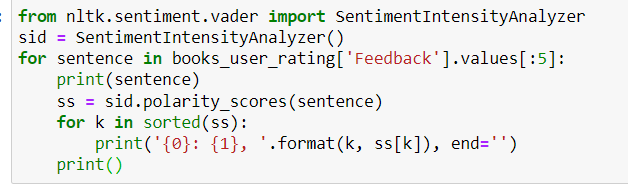
**** ****



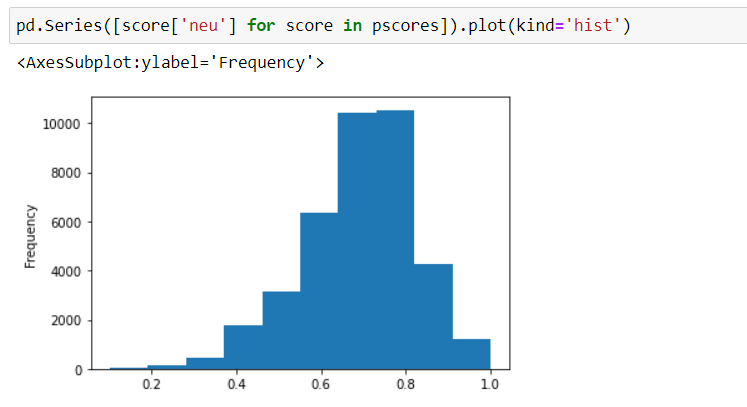


***Sentiment Analysis :***

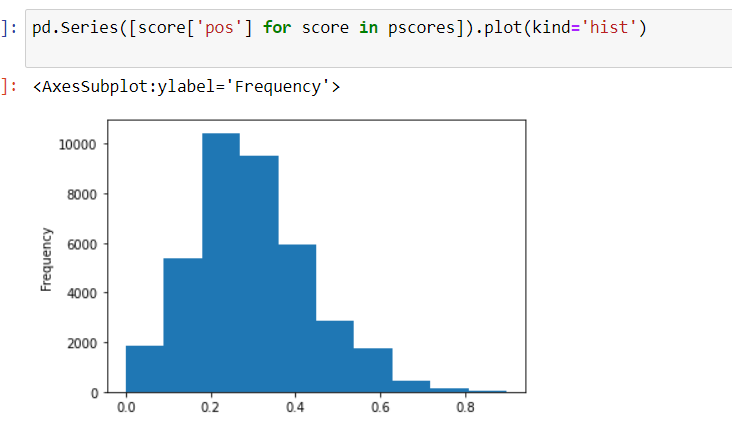
*Importing the nltk and required sentiment libraries for the analysis*



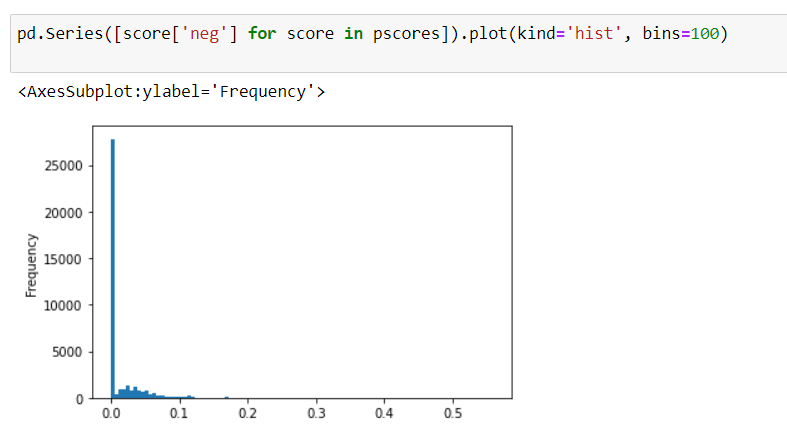
**Neutral Score:**



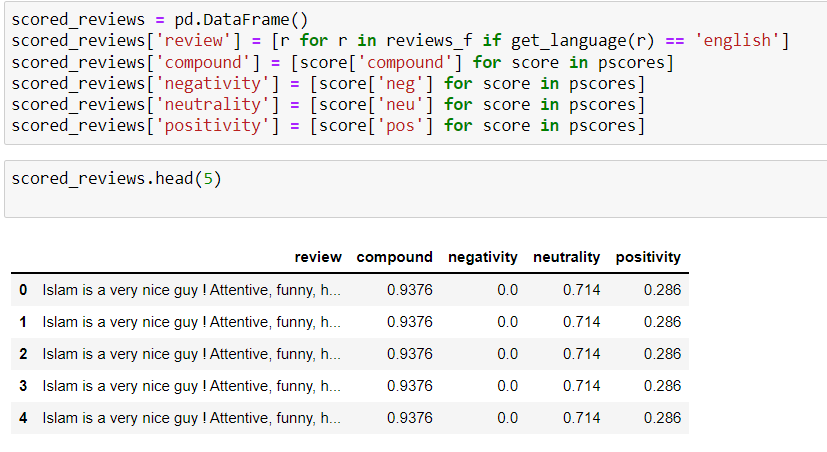
**Positive Score:**



**Negative Score:**



**Getting the polarity columns:**



**CHAPTER 5**

**CONCLUSION AND FUTURE WORK**

**Conclusion:**

The dataset after cleaning and getting the information which is necessary for the algorithm and recommendation is stored separately into a csv file and in future it is then used to get the most recommended system. The ratings count are also used for future analysis also. The genres count is taken for future work. The sentiment analysis is made and the data is stored separately in csv file and this data will be loaded into the database in future and the next set of work will be proceeded. This csv file will be stored for the database and to get the required queries.

**Future Work in Phase 2 :**

The preprocessed data is taken into separate csv file. Develop UI, where the user will be able to search for the book using the books name and get the recommended books based on the ratings. Next, the user will also be able to input set of words and get the recommended books based on the previously done sentiment analysis. The dataset should be stored in the mysql database and it needs to be connected to the front end. The data must be mapped according to the user request. The keyword extraction algorithm should also be written and it thus be able to get the specific word which match the genre. The genre needs to be converted to the array first and it then be able to match the extracted keyword from the user input. Thus, when the match is successful, it returns another set of recommended books, by passing this information to the recommendation system. This will thus be categorized as most positive feedback to the least one.

**CHAPTER 6**

# REFERENCES

* <https://data-science-blog.com/en/blog/2018/11/04/sentiment-analysis-using-python/>
* <https://arxiv.org/ftp/arxiv/papers/1403/1403.3185.pdf>
* <https://link.springer.com/chapter/10.1007/978-3-540-72079-9_10>
* <https://ieeexplore.ieee.org/abstract/document/6965073?casa_token=smrJx4qFy1wAAAAA:H4oCy03htAEqXuLqlV3lumBMTYGQjBjC4OHi5MVq44jo1ozKGklwmk63lQ9AAvLtuNxfWZvggg>
* <https://libguides.library.cqu.edu.au/c.php?g=760913&p=5456502>
* <https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset>
* <https://www.kaggle.com/code/sercanyesiloz/book-recommendation-system#Preprocessing>
* <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2>
* <https://www.mdpi.com/2076-3417/11/9/3986>
* <https://www.researchgate.net/profile/Haque-Nawaz-Lashari-2/publication/332731286_Student_Feedback_Sentiment_Analysis_Model_Using_Various_Machine_Learning_Schemes_A_Review/links/5d961d2c299bf1c363f5759a/Student-Feedback-Sentiment-Analysis-Model-Using-Various-Machine-Learning-Schemes-A-Review.pdf>