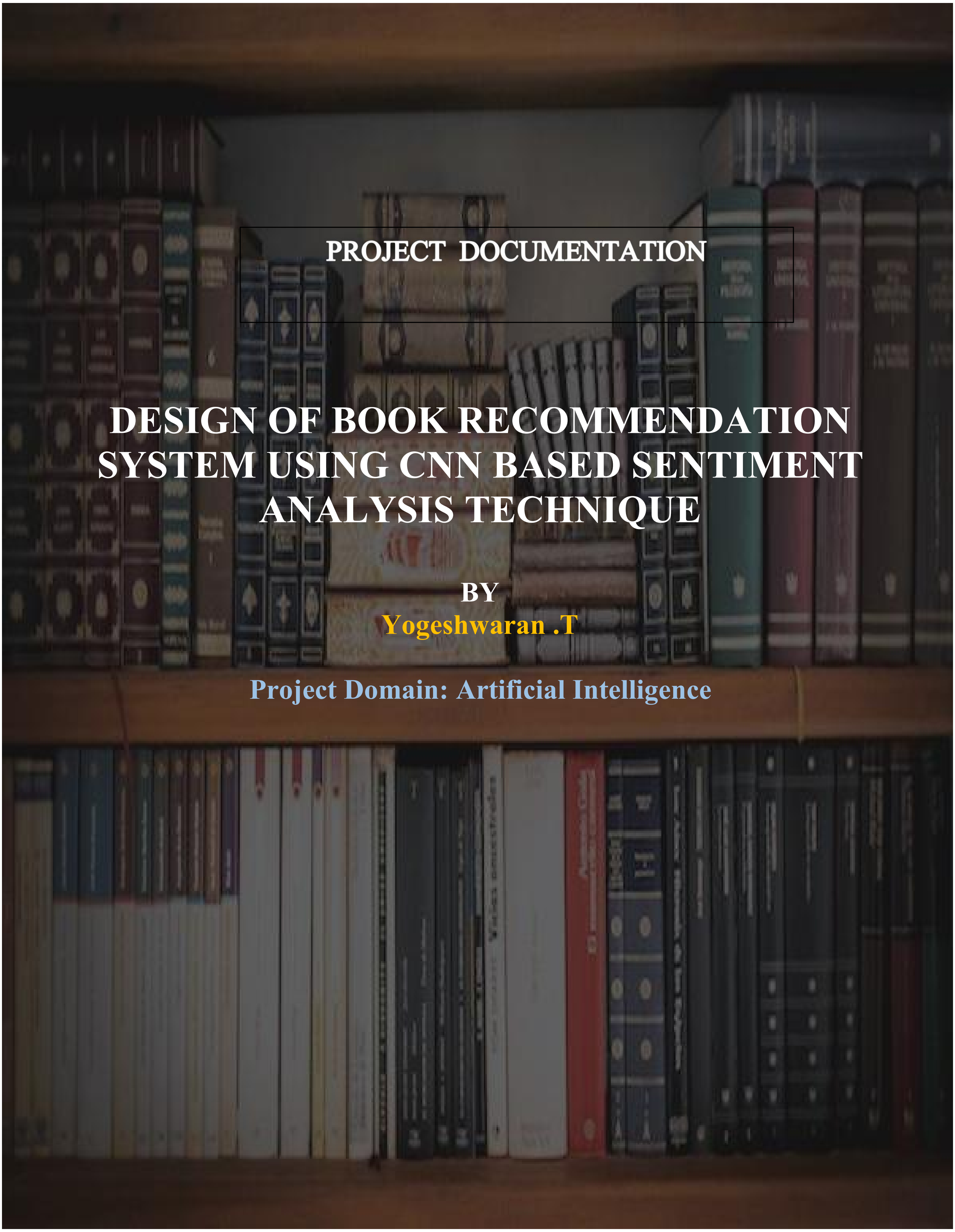


PROJECT DOCUMENTATION

DESIGN OF BOOK RECOMMENDATION SYSTEM USING CNN BASED SENTIMENT ANALYSIS TECHNIQUE

BY

Yogeshwaran .T

Project Domain: Artificial Intelligence

ABSTRACT

An online book review is a form of literary criticism in which a book is merely described (summary review) or analyzed based on content, style, and merit. A book review may be a primary source, opinion piece, summary review or scholarly review. A book review's length may vary from a single paragraph to a substantial essay. Online Book reviews are considered as one of the most essential sources of client opinion.

In present situation, consumers can learn about the books using online review resources to make decisions. The proposed methodology is to perform sentiment analysis for classification of book reviews into positive and negative. An extraction of book feature sentiment is to be done to get best accuracy within less time elapsing. Then, the reviewers are grouped based on the polarity of their review.

Over the past few years, recommended systems have been more and more effective in helping people identify their preferred resources from a large number of candidate objects. The Recommendation System assesses the preferences of users who have previously shown an interest in certain items, in order to recommend items which better match the target user's interests. It helps people save a considerable amount of shopping time.

KEYWORDS – Sentiment Analysis, Feature Extraction, Polarity of review, Recommender System, User Interest.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CNN	Convolutional Neural Network
NLP	Natural Language Processing
ReLU	Rectified Linear Unit
POS	Part-Of-Speech
LCS	Least Common Subsumer
DBOW	Distributed Bag-Of-Words
DMM	Distributed Memory Mean
SVD	Single Value Decomposition
CF	Collaborative Filtering

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

Sentiment analysis is the area which deals with judgments, responses as well as feelings, which is generated from texts, being extensively used in fields like data mining, web mining, and social media analytics because sentiments are the most essential characteristics to judge the human behaviour. Sentiments can be positive, negative, or neutral or it can contain an arithmetical score articulating the effectiveness of the sentiment.

Sentiments can be expressed by calculating the judgment of people on a certain topic, approach, and sensation toward a unit, where a unit can be an occurrence, a theme, or even a character.

Sentiment analysis and opinion mining are used interchangeably in several cases though there are occurrences where they hold minute dissimilarities among themselves. Sentiment analysis works on discovering opinions, classify the attitude they convey, and ultimately categorize them division-wise. The reviews are first collected in the process, their sentiment recognized, features selected, sentiments classified, and finally sentiment polarization determined or calculated. Finding the appropriate dataset is a very important concern while dealing with sentiment analysis.

Sentiment analysis can be functional for reviewing products for business, to ascertain the high and lows of stock markets, to understand the mentality of people reading news, and also views expressed by people in political debates. Sentiment analysis is done basically because not every review that is received gives a direct “good” or a “bad” notion. Though sentiment analysis is very much helpful, the enhancement of the analysis depends on the amount of training data that has been fed into the machine.

Deep learning is considered to be a major part of Machine Learning which is mainly supported on methods and actions formalized to gain knowledge about multiple levels of feature depiction. Inclusion of Deep Learning procedures and availability of large datasets has made possible a great substantial evolution in the field of sentiment analysis. Deep learning has a huge advantage as it carries out involuntary trait selection hence saving time and manual labour as feature engineering is not required. Different deep learning architectures like convolution neural networks and recursive neural networks, deep convolution neural networks, long short-term memory neural networks and other types of networks have yielded good results and outperformed numerous characteristic manufacturing techniques.

CNNs are basically several layers of convolutions with nonlinear activation functions like ReLU or tanh applied to the results. In a traditional feedforward neural network we connect each input neuron to each output neuron in the next layer. That's also called a fully connected layer, or affine layer. Instead, in CNN we use convolutions over the input layer to compute the output. This results in local connections, where each region of the input is connected to a neuron in the output. Each layer applies different filters, typically hundreds or thousands like the ones showed above, and combines their results. During the training phase, a CNN automatically learns the values of its filters based on the task you want to perform.

The input to most NLP tasks are sentences or documents represented as a matrix. Each row of the matrix corresponds to one token, typically a word, but it could be a character. That is, each row is vector that represents a word. Typically, these vectors are word embeddings (low-dimensional representations) like word2vec.

The goal of a recommender system is to generate meaningful recommendations to a collection of users for items or products that might interest them. Suggestions for books on Amazon, or movies on Netflix, are real-world examples of the operation of industry-strength recommender systems. The design of such recommendation engines depends on the domain and the particular characteristics of the data available. Recommender systems differ in the way they analyze these data sources to develop notions of affinity between users and items, which can be used to identify well-matched pairs.

1.2 PROBLEM DEFINITION

Maximum amount of existing research on text and information processing is focused on mining and getting the factual information from the text or information. When an organization needs to find opinions of the general public about its products and services, it conducts surveys on focused groups.

The sentiment analysis is carried out for only few datasets which is at the sentence level, feature level and document level with less accuracy. Static reviews are only taken into account.

These reviewers' sentiments serve as vital side information for improving the performance of recommender systems, most existing approaches ignore to fully exploit them in modelling the fine-grained user-item interaction for improving recommender system performance. Recommender systems have been in existence everywhere with most of them using single ratings in prediction. However, multi-criteria predictions have been proved to be more accurate

1.3 ORGANIZATION OF THE PROJECT

The presentation of the report is organized as follows:

- In the current chapter Overview, Problem Definition and finally the organization of the thesis are presented.
- In Chapter 2, an overview of the existing system and their merits and demerits are discussed.
- In Chapter 3, the implementation of the proposed system is explained with necessary figures.
- In Chapter 4, the various Hardware and Software Requirements of the Project are discussed.
- In Chapter 5, the results of the project work are discussed, and the output is presented.
- In Chapter 6, the project work is concluded along with the possible directions of the future workout lined.

CHAPTER 2
LITERATURE SURVEY

Ricardo B. Scheicher et al. in their paper- **Sentiment classification improvement using CNN with semantically enriched information** propose a method to improve sentiment classification performance with the use of semantically enriched information derived from domain expressions. An experimental evaluation was conducted by applying different classification algorithms to three datasets composed by reviews of different products and services. The results indicated that the proposed method enables the improvement of classification accuracy when dealing with reviews of a narrow domain. The experimental evaluation indicated that the method is suitable when the reviews relate to entities of the same nature [1].

In **Hybrid Architecture for Sentiment Analysis using Deep Learning**, Arnav Chakravathy et al. perform Sentimental Analysis using a deep learning model. The word embedding technique is used following which the input is fed into a deep neural network architecture. The proposed methodology uses a hybrid architecture, which consists of CNNs (Convolutional Neural Networks) and RNNs (Recurrent Neural Networks), to implement the deep learning model on the SAR14 and Stanford Sentiment Treebank data sets. The model performs sentence level sentiment analysis; however, its performance can be improved in the future by implementing an architecture that uses boosted CNNs [2].

Another paper-**Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis** by Zhiping Hou et al. identifies themes and compares differences in online travel reviews. A semantic association analysis was applied to extract thematic words and construct a semantic association network from reviews obtained from three major online travel agencies (OTAs) in China. The core idea of semantic association analysis is that semantics are defined by the co-occurrence of two words in a sentence with high-frequency words as a node, taking into consideration the frequency of high-frequency phrase co-occurrence as a link between nodes. Statistical analysis of thematic words, semantic association analysis and visualization are performed on the data [3].

In the paper- **A Novel Deep Multi-Criteria Collaborative Filtering Model For Recommendation System** by Nour Nassar et al. propose a novel multi-criteria collaborative filtering model based on deep learning. This model consists of two parts: in the first part, the model obtains the users and items features and uses them as an input to the criteria ratings deep neural network, which predicts the criteria ratings. Those criteria ratings constitute the input to the second part, which is the overall rating deep neural network and is used to predict the overall rating. The performance of this model was evaluated on the TripAdvisor1 dataset, a multi-criteria rating dataset, which contains users' ratings for hotels. The experimental results demonstrates the model's good performance by outperforming the other methods for all the evaluation metrics. In the future, its performance can be improved either by studying different deep learning methods and Autoencoder or by using more complex deep networks or other feature representation methods [4].

Table 2.1 Table for Comparison of Various Techniques

Sl. No.	Author Name	Title of the Paper	Journal Name / Year	Dataset Used	Parameters	Merits
1	RicardoB. Scheicher et al.	Sentiment classification improvement using CNN with semantically enriched information	Information Processing and Management, February, 2019 © Elsevier	The experimental evaluation was conducted using three datasets HuLiu2004, SemEval2014, SemEval2015	gBoED Frequency Distribution, gBoED Distance	It improves the classification of documents with low predictive confidence.
2	Arnav Chakra vathy et al.	Hybrid Architecture for Sentiment Analysis using Deep Learning	International Journal of Advanced Research in Computer Science, February 2018.	Stanford Sentiment Treebank dataset (11855 reviews) SAR14 movie review dataset (233600 movie reviews)	feature map, pooling operation, weight matrix and backward propagation through time	Using both CNN and RNN to increase the efficiency in analysis of data.

3	Zhiping Hou et al	Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis	International journal of Tourism Management, March 2019 © Elsevier	three major OTAs platforms (Ctrip, Tuniu and Tongcheng) in China	Average length of reviews, number of words	Expands the understanding of the methodological challenges and offers novel insights for mining the opinions for the benefit of tourists, hotels and tourism enterprises and OTAs.
4	Nour Nassar et al.	A novel deep multi-criteria collaborative filtering model for recommendation system	International Journal of Knowledge-Base Systems, June 2019 © Elsevier	TripAdvisor1 dataset, multi-criteria rating dataset	Mean Absolute Error (MAE), F-measure (Fβ), Mean Average Precision (MAP) , Mean Reciprocal Rank (MRR)	It predicts the criteria ratings and predicts the overall rating

CHAPTER 3
DESIGN OF PROPOSED WORK

In this proposal, 50 document level reviews are first trained using the semantic network and the tagged words (adverbs, adjectives, verbs and negations) are collected. In the **semantic network** the collectively identified words of each documents with similar meaning is added to the list for every iterations and finally appended as positive list and negative list.

80% of the reviews and the tagged words from the Semantic Network Module are collectively taken as training dataset. This data set is trained using **CNN and n - grams** method using distributed Bag Of Words. The sigmoid function is used to set the range from 1 to 0.

Then the testing reviews are tested with the training dataset to classify the reviews as positive and negative based on their polarity. The reviewers are **grouped** depending upon the polarity of their review in the database. From the updated database of books and reviewers the table of userid, bookid and rating are extracted based on unique id.

Then it is passed to **popularity recommendation model** to recommend top popular items meanwhile the same extracted data is considered and decomposed using single value decomposition and predicted data frame is created. Based on the similarity value we predict the user interest with regarding to item and recommend the top-N items. A **collaborative filtering algorithm** is to be used to recommend books, it will effectively improve data scarcity and real-time of the problem.

3.1 HIGH LEVEL ARCHITECTURE DIAGRAM

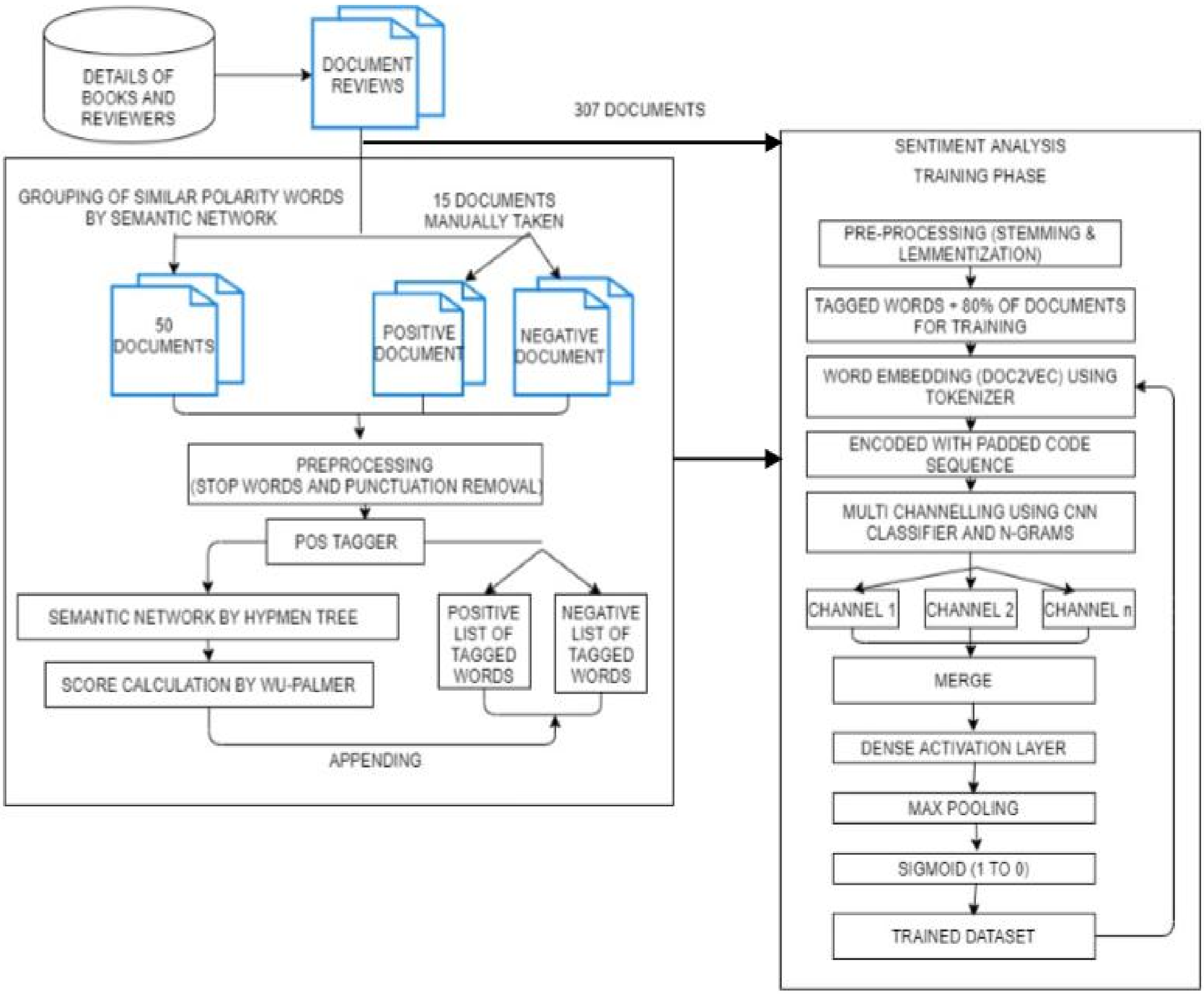


Fig 3.1 Training Phase

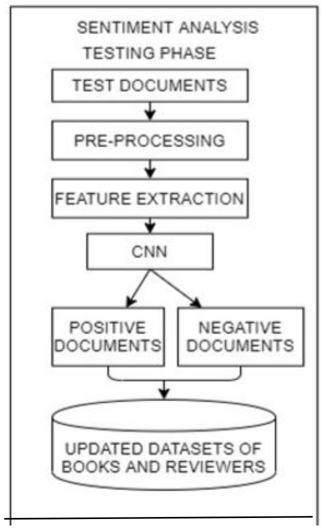


Fig 3.2 Testing Phase

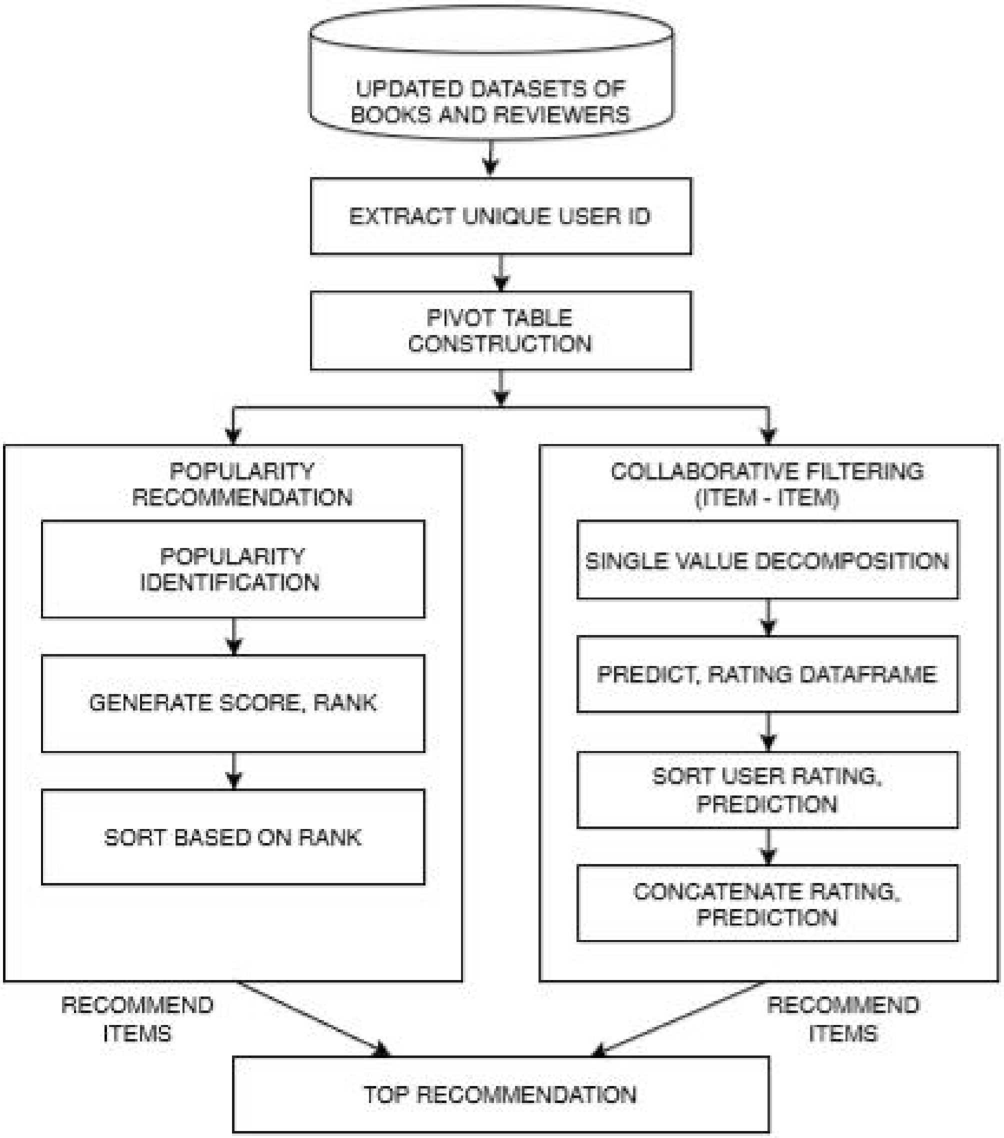


Fig 3.3 Overall Architecture of Recommendation System

3.2 MODULE DESCRIPTION

The proposed system has the following modules:

Module 1: Grouping Similar Tagged Words By Semantic Network

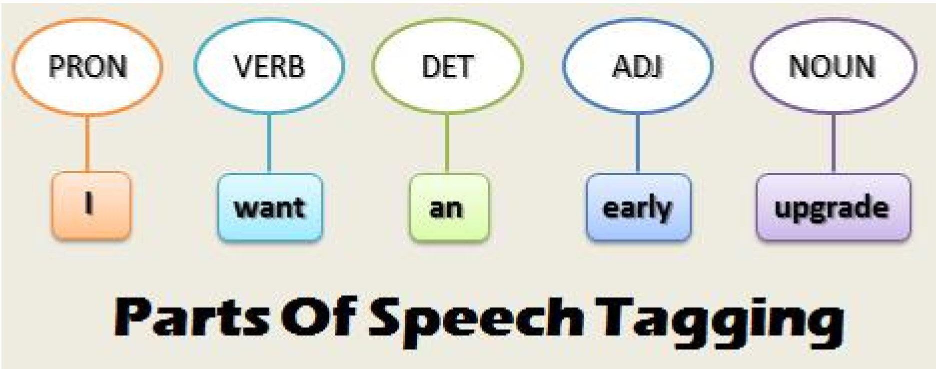
Module 2: Sentiment Analysis On Book Reviews

Module 3: Identification Of Reviewers (based on age/locality/gender)

Module 4: Recommendation System

3.2.1 Module 1: Grouping Similar Tagged Words By Semantic Network

- In this module, few sentences (15) which are manually identified as positive and negative sentences are collected from the datasets of book reviews and these documents are passed for the pre-processing.
- The pre-processed documents of positive and negative are fed into the POS tagger to obtain tagged words such as verbs, adjectives, adverbs and negated words. These tagged words from each category are collected as a different list by avoiding repeated words.
- The list of POS tagged words are fed into the semantic network. The new reviews are given into the semantic network where once again the POS tagger is used to identify the words as adverbs, adjectives, verbs and negation then the tagged words are fed into WordNet where each of the words are connected and by synsets, their various synonyms are also considered and compared with the list of tagged words and return a score for each word. The score for each word is calculated by WuPlamer formula.



- The WuPlamer is used to calculate the relatedness by taking into account the depths of two synsets in the WordNet taxonomies, including the depth of Least Common Subsumer (LCS).

$$\mathbf{Wu - Palmer} = 2 * \frac{\mathbf{depth\ (lcs(s1, s2))}}{\mathbf{(depth\ (s1)\ +\ depth\ (s2))}}$$

------------->(3.1)

- The score generated is 0<=score<=1 which can be represented in decimal. In the equation, the depth represents the similarity of each word compared with the words in the list. The score can be zero when there is no way of similar matching. This equation (3.1) calculates the similarity based on how similar the word senses are and where the synsets occur relative to each other in the hypmen tree.

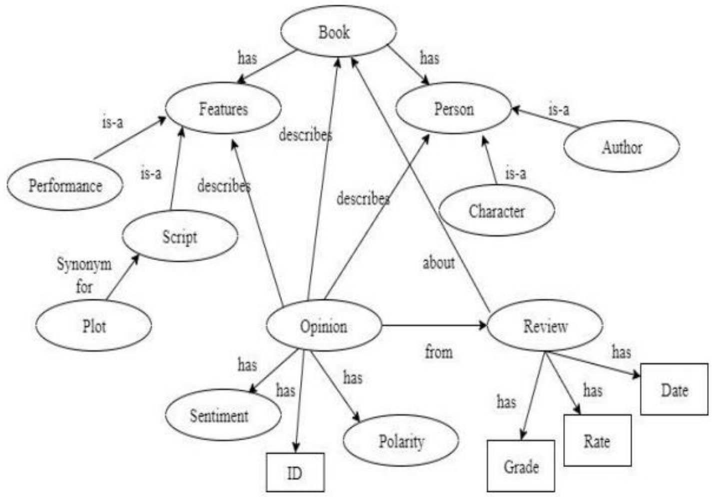


Fig 3.2.1 a) Typical structure of hypmen tree

- Based on the generated three scores by comparing to three different lists, the best score is chosen and the input sentence is classified as positive and negative. After classification, each word from the classified review is appended to the category of the list which is identified as output. Hence this improves the accuracy as the new words are included in the given list.

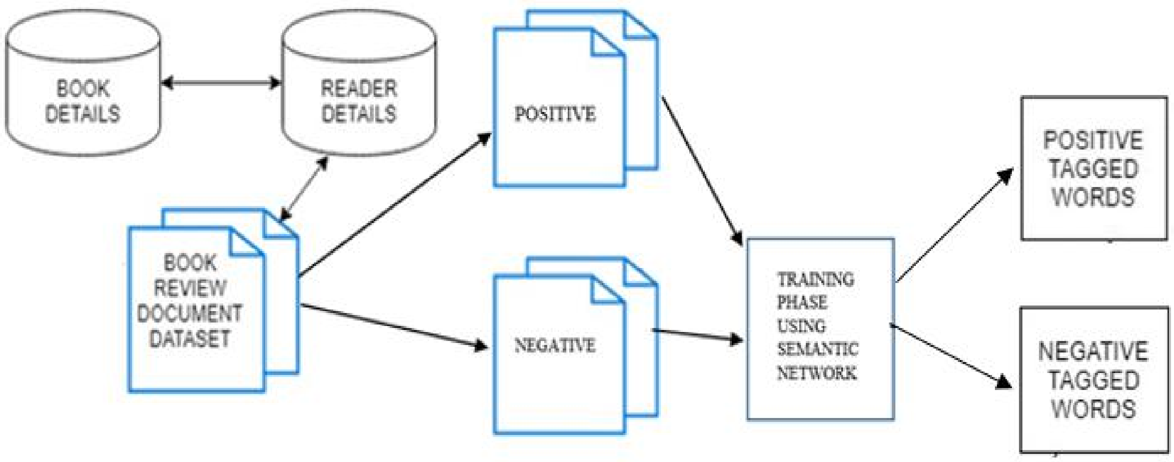
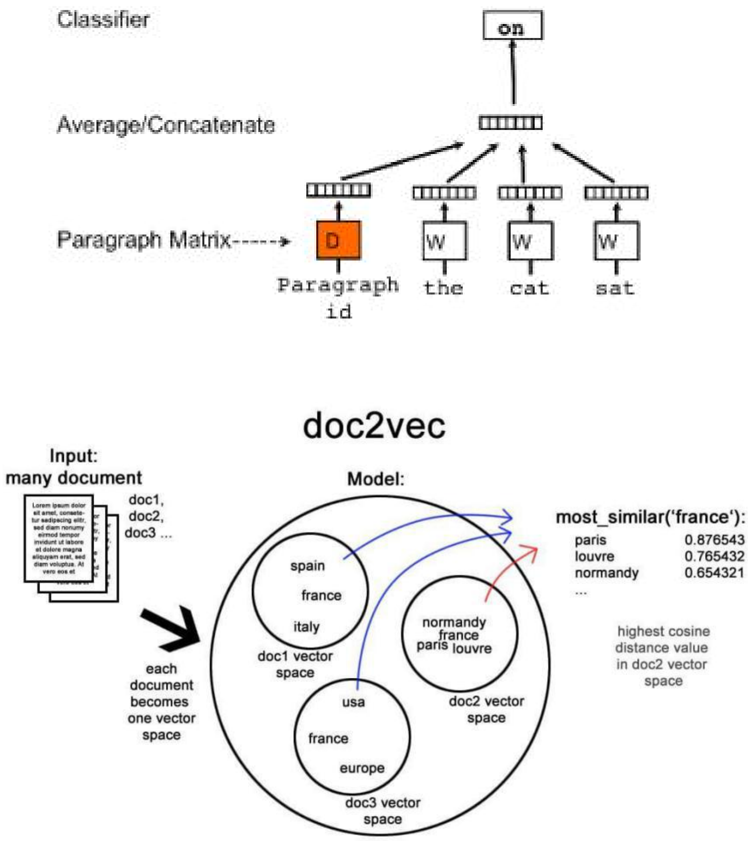


Fig 3.2.1 Grouping Similar Tagged Words By Semantic Network

3.2.2 Module 2: Sentiment Analysis On Book Reviews

- In this module, the reviews are pre-processed to avoid the tags, stop words and punctuation then stemming and lemmatization are done to identify the exact root meaning of the word.
- Then the word embedding is done to convert the pre-processed data into the Doc2vec (document to vector model) where the similar type of words are represented by the same vector, the tagged words from the previous module are also taken in this word embedding layer to collectively form the training dataset.



Structure of Doc2Vec model

- Doc2vec is an unsupervised algorithm [16] to generate vectors for sentence/paragraphs/documents. The algorithm is an adaptation of word2vec which can generate vectors for words. Doc2Vec computes a feature vector for every document in the corpus. The underlying intuition of Doc2Vec is that the document representation should be good enough to predict the words in the document.
- The vectors generated by doc2vec can be used for tasks like finding similarity between sentences/paragraphs/documents. Doc2Vec's learning strategy exploits the

idea that the prediction of neighbouring words for a given word strongly relies on the document also.

- A sigmoid function is a mathematical function having a characteristic "S"-shaped curve or sigmoid curve. A standard choice for a sigmoid function is the logistic function shown in the first figure and defined by the formula:

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} \tag{3.2}$$

- And this is fed into the Convolution neural network [14] where the sentiment of the words are identified by N-grams along with Distributed Bag Of Words (DBOW) method. The sigmoid function is used to set the range from 1 to 0.
- The DBOW model "ignores the context words in the input, but force the model to predict words randomly sampled from the paragraph in the output."
- After classification the corresponding data is stored in the database by referring to the reviewers.

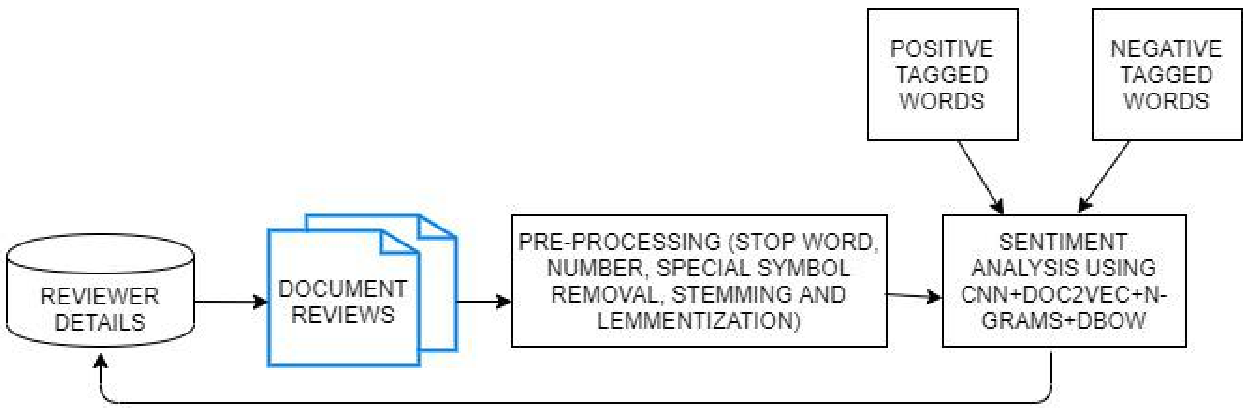


Fig 3.2.2 Sentiment Analysis On Book Reviews

3.2.3 Module 3: Identification Of Reviewers (Based On Age/Locality/Gender)

- The reviewers are grouped based on the type of review they are given at first and then based on the reviewers' age, gender and locality they are grouped using the MySql. The queries and the groups may be of more than two groups.

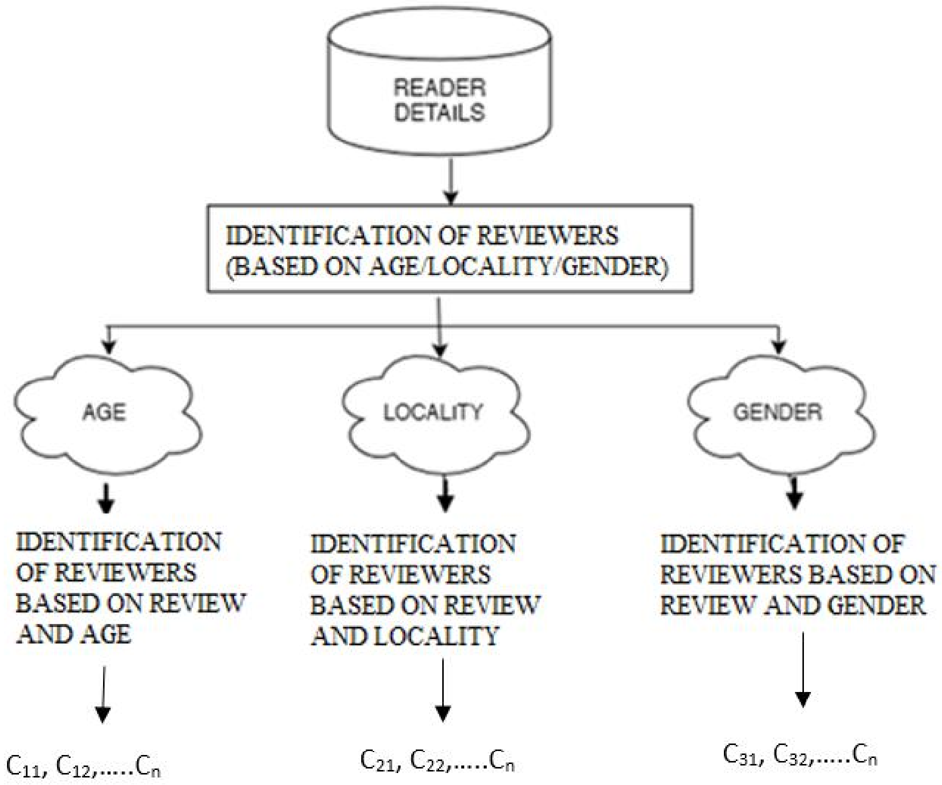


Fig 3.2.3 Identification of Reviewers (Based on age/locality/gender)

3.2.4 Module 4: Recommendation System

Recommendation based on popularity and collaborative filtering algorithm:

- The evolution of the internet has provided the available access to numerous items. Over the past few years, recommender systems have been more and more effective in helping people identify their preferred resources from a large number of candidate objects.
- The Recommendation System [9] assesses the preferences of users who have previously shown an interest in certain items, in order to recommend items which

better match the target user’s interests. It helps people save a considerable amount of shopping time.

- The main advantage of collaborative [18] approaches is that they require no information about users or items and, so, they can be used in many situations. Moreover, the more users interact with items the more new recommendations become accurate: for a fixed set of users and items, new interactions recorded over time bring new information and make the system more and more effective.
- Recommendation System of a book selects and suggests the contents to meet reviewers’ preference automatically using data of previous readers stored in database.
- The rating, bookid and userid dataset from the database table is extracted from the database and from that the unique userid are taken into account because there can be repeated userid which have given many reviews with same id.
- Then a pivot table is constructed for the ratings.This pivot table will be used in both popularity recommendation and collaborative filtering techniques.
- In popularity model, identification is done by generating score and the rank based on the top rank [15] the books are recommended for the users.
- In collaborative filtering model the users activities of browsing is taken into account and based on their usage of the website the books are recommended.
- In this collaborative filtering, single value decomposition tool is incorporated.
- SVD is one major tool in the development of implicit based similarity recommender systems. Instead of having access and using the explicit contents’ and users’ features, SVD uses the rating of all uses on the items they have rated, to implicitly discover the item and user features.
- Though, these discovered features are not the original explicit features. The implicit features discovered are – in general – a non-linear combination of the original features and may not represent any physical or explicit interpretation.
- The implicit similarity based recommender systems are often called collaborative filtering (CF). CF methods do need domain specific knowledge of features of contents nor users. In simple words “customers who have bought this items also bought these items”

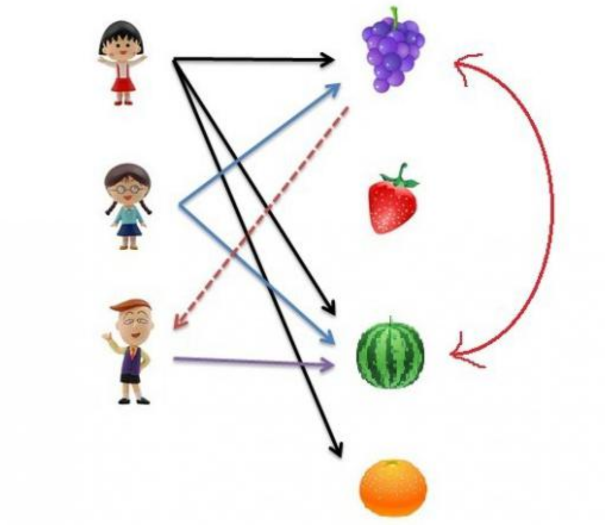


Fig 3.2.4 a) item-item Collaborative filtering - depicts that the user1 bought grapes,watermelon and orange , the user3 bought watermelon alone but the user three will get a recommendation that users who have bought watermelon also bought grapes.

- They could use the feedback of users on a contents to compute the unknown interactions. To make results of CF to be accurate, large amount of data is needed.
- One simple model t compute the unknown entries in the matrix X is to use the baseline model. The basic idea is that an unknown rating of user I on the content j, x_{ij}, could be computed based on the average rating the user I gives to all contents rated by the user, the average rating the content j has received and the average of all rating.

- Baseline estimate for x_{ij} is;

$$b_{ij} = v + t_i + t_j \tag{3.3}$$

Where,

v = mean rating of all users over all items
t_i = rating bias of user i = mean of all ratings by user i – v
t_j = rating bias of item j = mean of all ratings on item j – v

- Then collectively from both the modules the books are recommended for the readers.

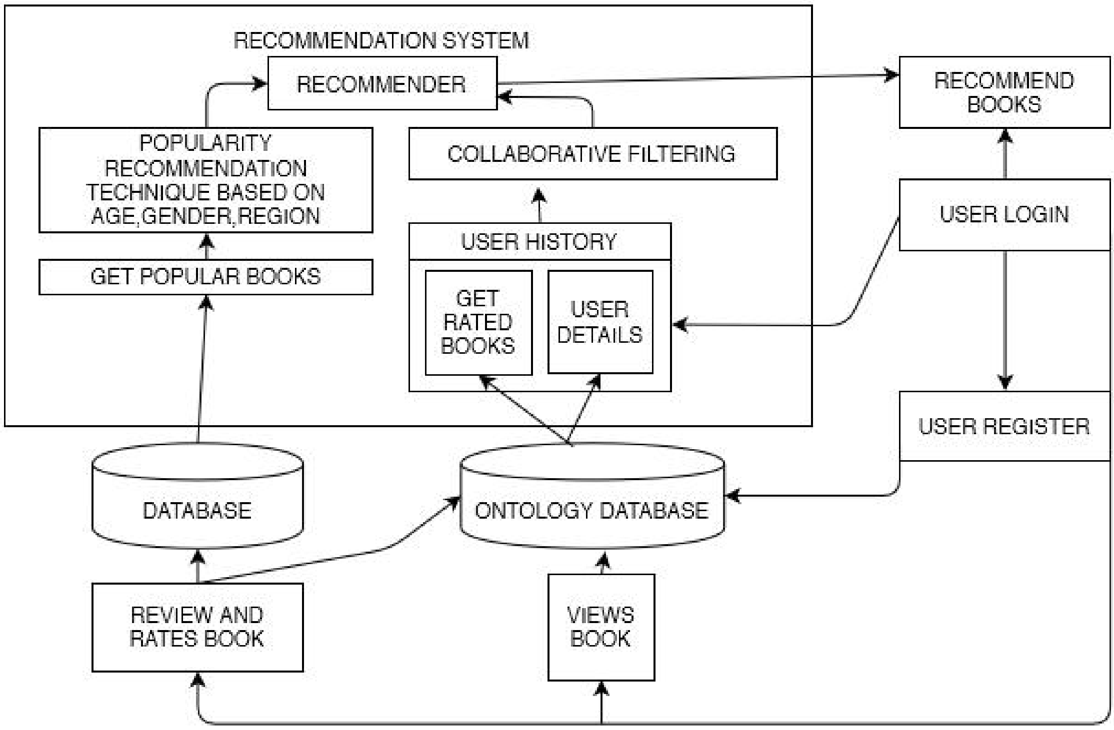


Fig 3.2.4 Recommendation System

Our proposed work used the semantic tagged words from the reviews to understand the reviews better. The reviews are represented in document to vector form so that the document sized reviews are well handled for the sentiment analysis phase to get better results compared to other existing methods of about 90.32%.

The recommendation model which is built efficiently as the latest item to item collaborative filtering along with the popularity algorithm are used for recommending books. The Books which are popular and based on the user interest are recommended more relatively to the users.

CHAPTER 4
IMPLEMENTATION ENVIRONMENT

4.1 HARDWARE REQUIREMENTS

Processor	: 3.0 gigahertz (GHz) frequency or above
RAM	: 4 GB
Hard disk	: 20 GB
Internet connection	: Broadband/4mps speed
Browsers	: Chrome*36+, Edge* 20+Mozilla Firefox 31+, IE 11+.

4.2 SOFTWARE REQUIREMENTS

Spyder

Spyder is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. It is released under the MIT license. It features a unique combination of the advanced editing, analysis, debugging, and profiling functionality of a comprehensive development tool with the data exploration, interactive execution, deep inspection, and beautiful visualization capabilities of a scientific package.

The Spyder IDE is used in the module1 to run semantic network

Jupyter Notebook

Jupyter Notebook (formerly IPython Notebooks) is a web-based interactive computational environment for creating Jupyter notebook documents. The "notebook" term can colloquially make reference to many different entities, mainly the Jupyter web application, Jupyter Python web server, or Jupyter document format depending on context. A Jupyter Notebook document is a JSON document, following a versioned schema, and containing an ordered list of input/output cells which can contain code, text (using Markdown), mathematics, plots and rich media, usually ending with the ".ipynb" extension.

The CNN and recommendation modules(2 and 4) uses jupyter notebook.

Language - Python 3.6.5v

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming.

Python is used to code the complete project in algorithm involvement parts.

Backend – MySql

MySQL is a fast, easy-to-use RDBMS being used for many small and big businesses. MySQL is developed, marketed and supported by MySQL AB, which is a Swedish company. MySQL is becoming so popular because of many good reasons −

- MySQL is released under an open-source license. So you have nothing to pay to use it.
- MySQL is a very powerful program in its own right. It handles a large subset of the functionality of the most expensive and powerful database packages.
- MySQL uses a standard form of the well-known SQL data language.
- MySQL works on many operating systems and with many languages including PHP, PERL, C, C++, JAVA, etc.
- MySQL works very quickly and works well even with large data sets.

Framework - Django 2.2.1

Django is a Python-based free and open-source web framework, which follows the model-template-view (MTV) architectural pattern. Django's primary goal is to ease the creation of complex, database-driven websites. The framework emphasizes reusability and "pluggability" of components, less code, low coupling, rapid development, and the principle of don't repeat yourself. Python is used throughout, even for settings files and data models.

The web development side of the project is carried out in django framework.

CHAPTER 5

RESULT DISCUSSION

5.1 INPUT – OUTPUT SPECIFICATION

Input Specification

a) Input data - Review document, user details, book details
b) Format - .txt and .csv

The text format reviews are converted to doc2vec format and fed into the CNN. This gives a classified output and updated into the database.

Output Specification

Based on the identified polarity the books are mapped and Reviewers are identified based on their age, locality and gender. Then the books are recommended to the users by collaborative filtering and popularity algorithm.

5.2 DATASET DESCRIPTION

The Book details were collected from the kaggle website. Details about nearly 1098 books were collected and stored in the database.

The reader details were collected from the fakenamegenerator, in order to avoid the repetition of the names; the first name, middle name and last name is generated for each reader. The readers are generated with different countries. All these details about the reader are then stored into another database.

The reviews collected are in the document level. These reviews are collected from the amazon website. The reviews constitute a 5GB dataset.

Total number of Reviews = 307

The following table consists of the total number of review count and the count of reviews used for training and testing.

Table 5.1 Total Number Of Reviews

Sentiment	Training	Testing	Total Count
Positive	140	38	178
Negative	105	24	129

The following table consists of categories and count in which the books are present in the dataset.

Table 5.2 Category Of Books

Category	Count	Category	Count	Category	Count
Action	87	Crime	51	Musical	45
Adventure	25	Document	160	Mystery	79
Animation	44	Drama	94	Romance	65
Children	71	Fantasy	91	Sci-Fi	56
Comedy	57	Horror	59	Thriller	68
War	51	Modern	70	Literature	109

The table 5.2 depicts eighteen different categories in which the books are considered for the recommendation.

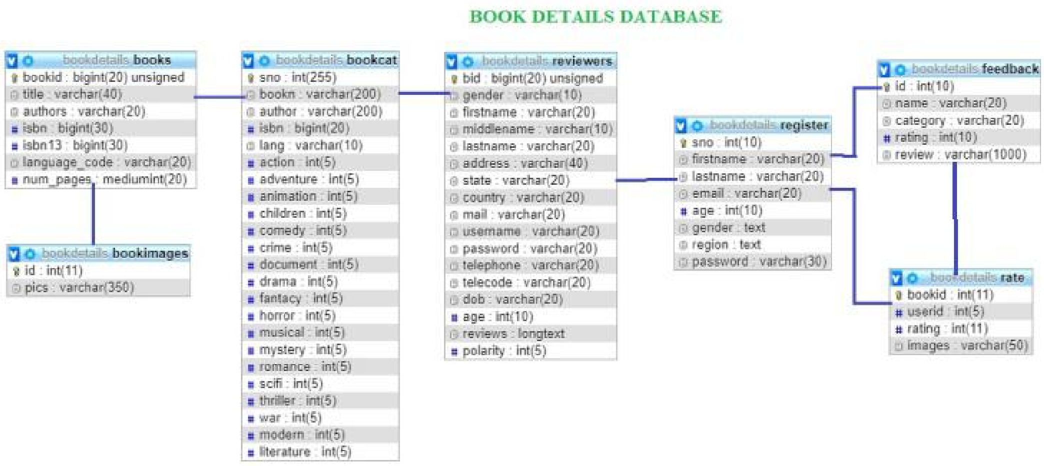


Fig 3.2.5 Database Scheme diagram

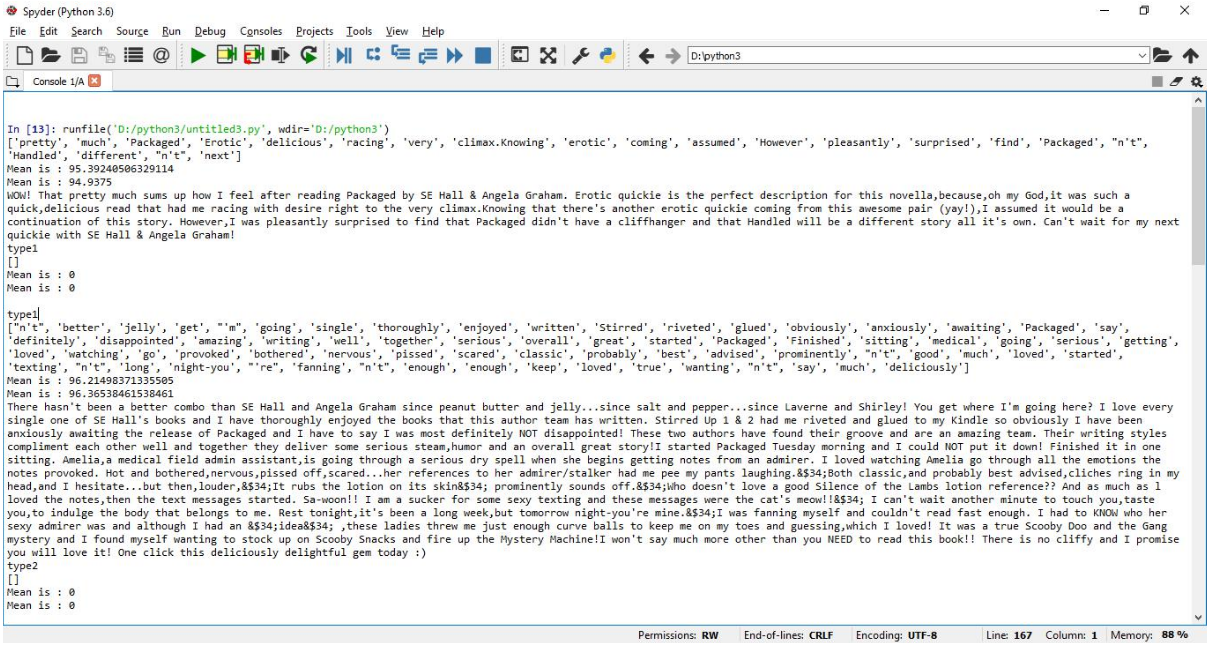
Table 5.2.1 Database description

S.no	Name of the data table	Use/relations
1.	books	This contains necessary information of the books
2.	bookimages	Each book has a front cover image which is stored in this table and related to books table
3.	bookcat	This table consists of categories of the books . a book can belong to one or more categories depending upon its journe so the 0 and 1 is used to represent their category. For example: the harry potter book comes under category of novel,adventure,fantasy and these columns will carry 1 for that book.
4.	reviewers	After registration the people who review any book will be stored in this table with their personal information along with analysed polarity of their review.
5.	register	It contains the registration details of users
6.	feedback	This table is used to store the users review temporary and then it is moved to the reviewers database with personal details. And moved to cnn for identify their polarity.
7.	rate	This table contains the just the ratings of the user for the bookid so image for that book is easy to retrieve during recommendation.

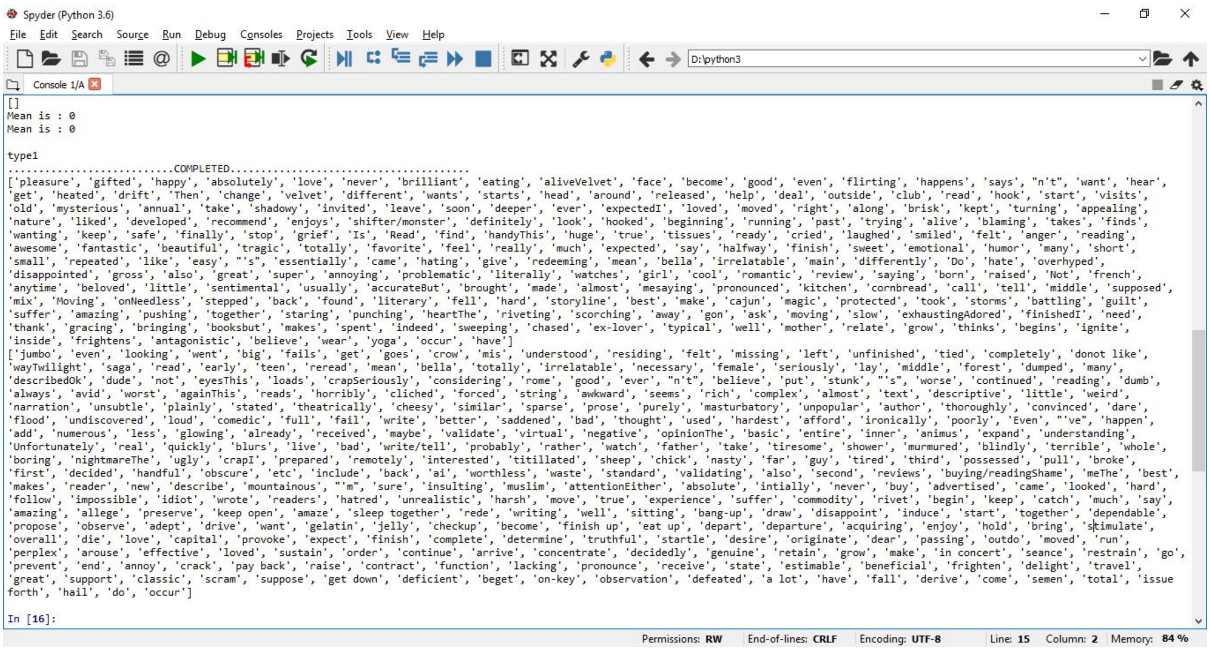
5.3 INPUT/OUTPUT SCREENSHOTS

Module 1

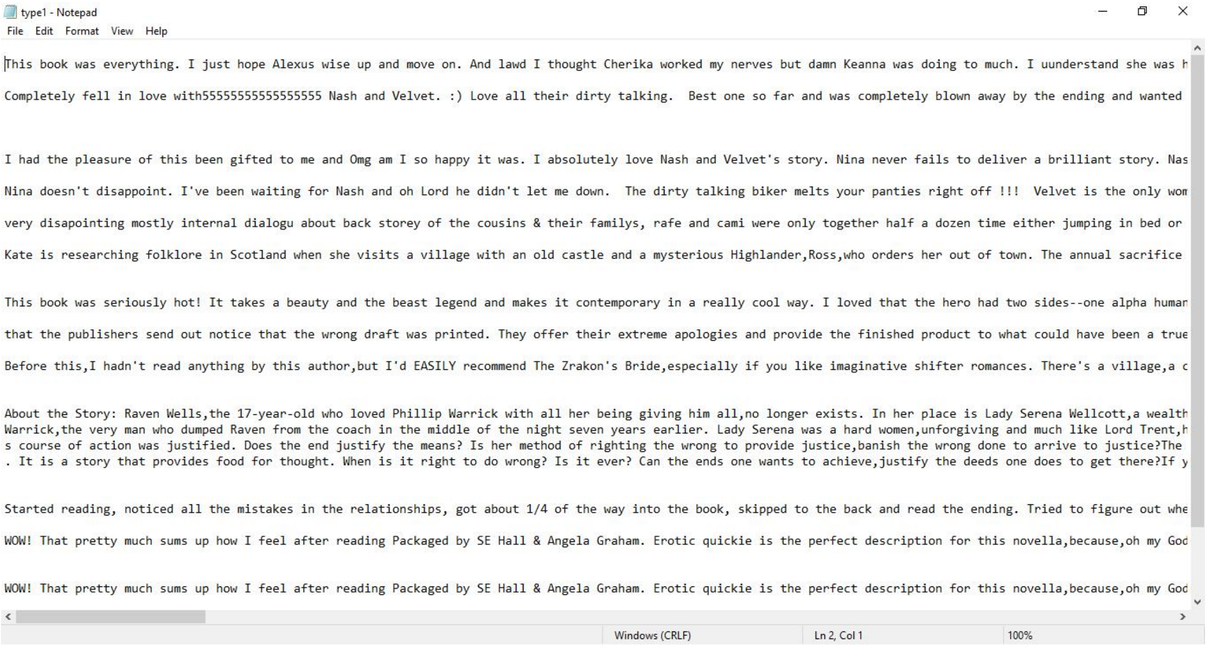
1. The semantic network identifies the types of documents based on similarity



2. The set of corpus (verbs, adverbs and adjectives) collected as two lists which will be used in CNN to predict the sentiments.

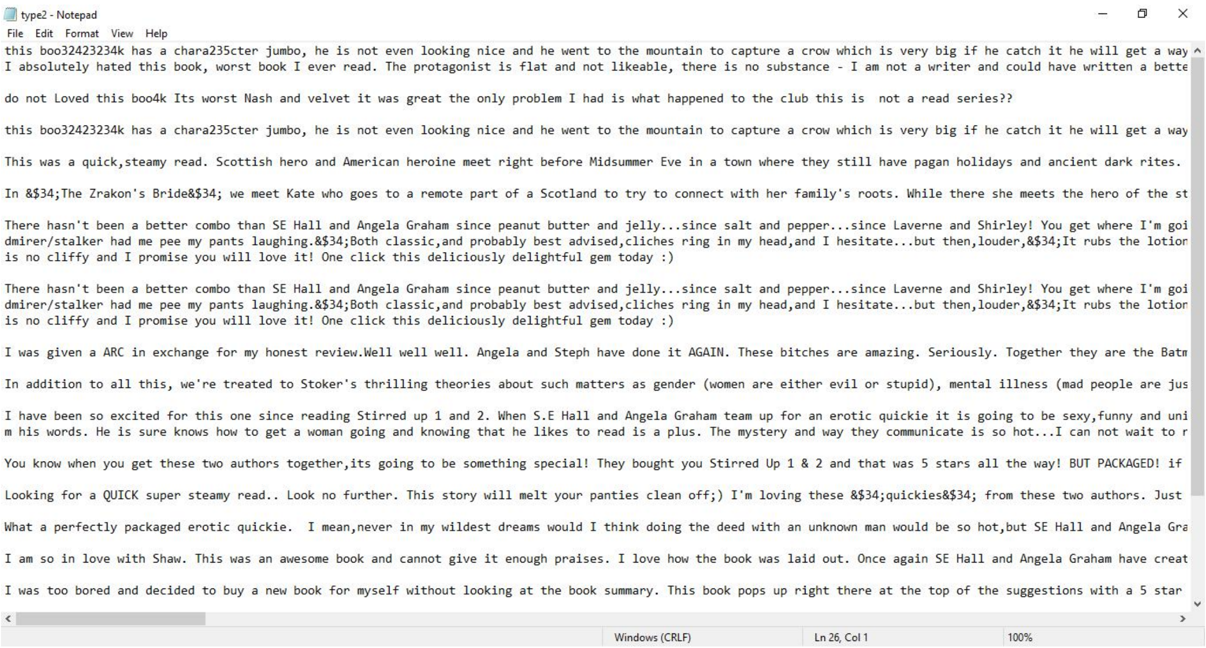


3. Type1 document level reviews are store in a separate 'type1.txt' file.



```
type1 - Notepad
File Edit Format View Help
This book was everything. I just hope Alexus wise up and move on. And lawd I thought Cherika worked my nerves but damn Keanna was doing to much. I uunderstand she was h
Completely fell in love with55555555555555555 Nash and Velvet. :) Love all their dirty talking.  Best one so far and was completely blown away by the ending and wanted

I had the pleasure of this been gifted to me and Omg am I so happy it was. I absolutely love Nash and Velvet's story. Nina never fails to deliver a brilliant story. Nas
Nina doesn't disappoint. I've been waiting for Nash and oh Lord he didn't let me down.  The dirty talking biker melts your panties right off !!!  Velvet is the only wom
very disapointing mostly internal dialogu about back storey of the cousins & their familys, rafe and cami were only together half a dozen time either jumping in bed or
Kate is researching folklore in Scotland when she visits a village with an old castle and a mysterious Highlander,Ross,who orders her out of town. The annual sacrifice

This book was seriously hot! It takes a beauty and the beast legend and makes it contemporary in a really cool way. I loved that the hero had two sides--one alpha human
that the publishers send out notice that the wrong draft was printed. They offer their extreme apologies and provide the finished product to what could have been a true
Before this,I hadn't read anything by this author,but I'd EASILY recommend The Zrakon's Bride,especially if you like imaginative shifter romances. There's a village,a c

About the Story: Raven Wells,the 17-year-old who loved Phillip Warrick with all her being giving him all,no longer exists. In her place is Lady Serena Wellcott,a wealth
Warrick,the very man who dumped Raven from the coach in the middle of the night seven years earlier. Lady Serena was a hard women,unforgiving and much like Lord Trent,h
s course of action was justified. Does the end justify the means? Is her method of righting the wrong to provide justice,banish the wrong done to arrive to justice?The
. It is a story that provides food for thought. When is it right to do wrong? Is it ever? Can the ends one wants to achieve,justify the deeds one does to get there?If y

Started reading, noticed all the mistakes in the relationships, got about 1/4 of the way into the book, skipped to the back and read the ending. Tried to figure out whe
WOW! That pretty much sums up how I feel after reading Packaged by SE Hall & Angela Graham. Erotic quickie is the perfect description for this novella,because,oh my God

WOW! That pretty much sums up how I feel after reading Packaged by SE Hall & Angela Graham. Erotic quickie is the perfect description for this novella,because,oh my God
Windows (CRLF)    Ln 2, Col 1    100%
```

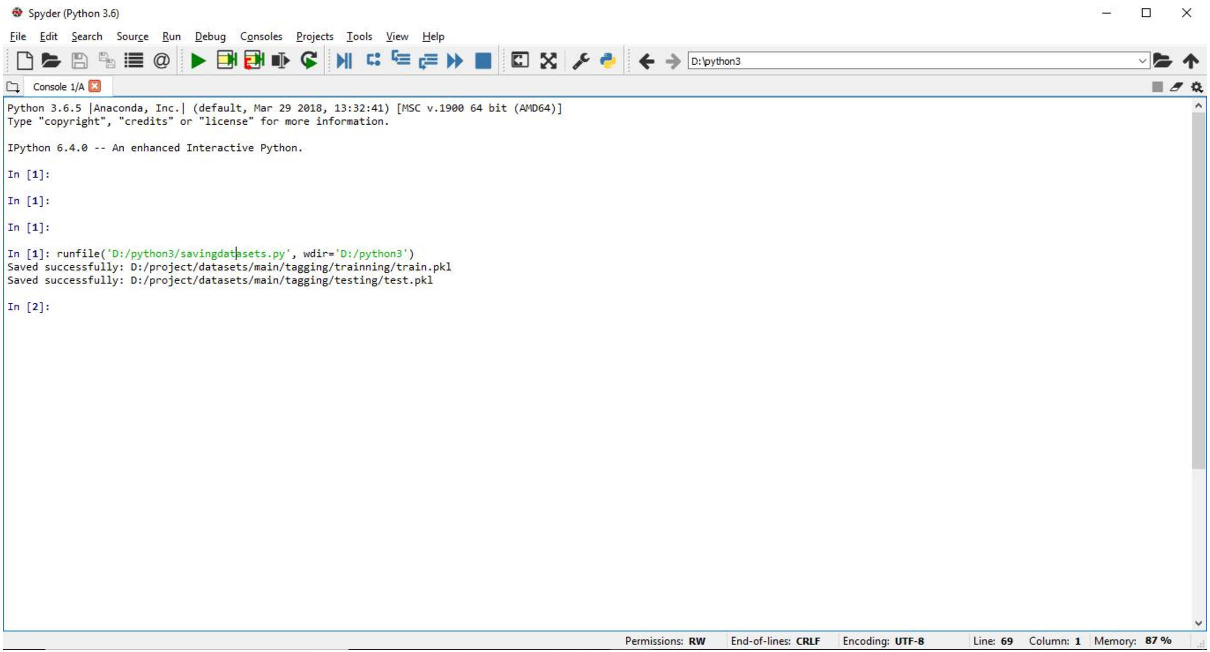
4. Type2 document level reviews are store in a separate 'type2.txt' file.



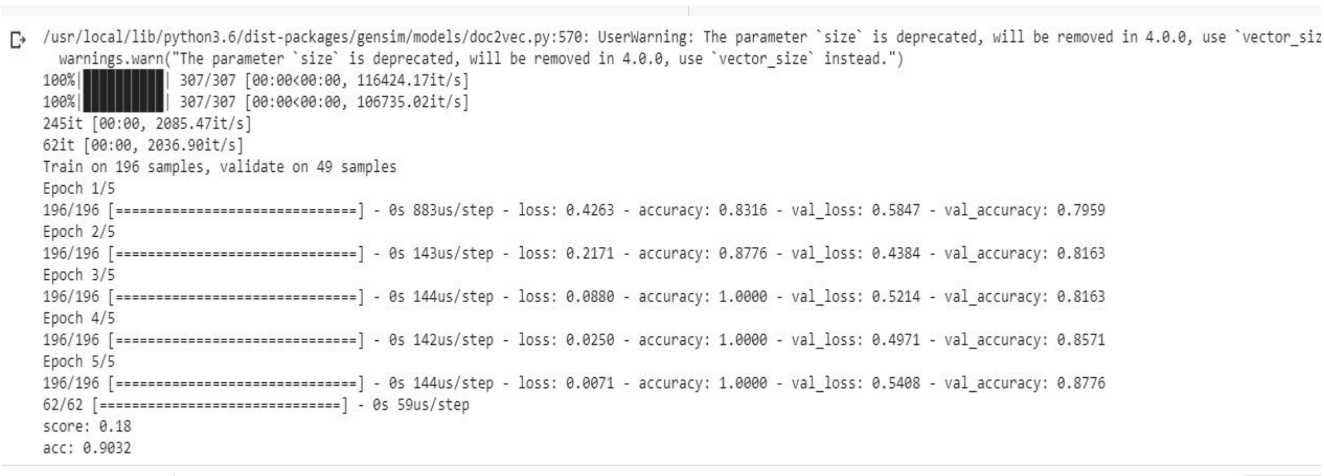
```
type2 - Notepad
File Edit Format View Help
this boo32423234k has a chara235cter jumbo, he is not even looking nice and he went to the mountain to capture a crow which is very big if he catch it he will get a way
I absolutely hated this book, worst book I ever read. The protagonist is flat and not likeable, there is no substance - I am not a writer and could have written a bette

do not Loved this boo4k Its worst Nash and velvet it was great the only problem I had is what happened to the club this is  not a read series??

this boo32423234k has a chara235cter jumbo, he is not even looking nice and he went to the mountain to capture a crow which is very big if he catch it he will get a way

This was a quick,steamy read. Scottish hero and American heroine meet right before Midsummer Eve in a town where they still have pagan holidays and ancient dark rites.

In &$34;The Zrakon's Bride&$34; we meet Kate who goes to a remote part of a Scotland to try to connect with her family's roots. While there she meets the hero of the st

There hasn't been a better combo than SE Hall and Angela Graham since peanut butter and jelly...since salt and pepper...since Laverne and Shirley! You get where I'm goi
dmirer/stalker had me pee my pants laughing.&$34;Both classic,and probably best advised,cliches ring in my head,and I hesitate...but then,louder,&$34;It rubs the lotion
is no cliffy and I promise you will love it! One click this deliciously delightful gem today :)

There hasn't been a better combo than SE Hall and Angela Graham since peanut butter and jelly...since salt and pepper...since Laverne and Shirley! You get where I'm goi
dmirer/stalker had me pee my pants laughing.&$34;Both classic,and probably best advised,cliches ring in my head,and I hesitate...but then,louder,&$34;It rubs the lotion
is no cliffy and I promise you will love it! One click this deliciously delightful gem today :)

I was given a ARC in exchange for my honest review.Well well well. Angela and Steph have done it AGAIN. These bitches are amazing. Seriously. Together they are the Batm

In addition to all this, we're treated to Stoker's thrilling theories about such matters as gender (women are either evil or stupid), mental illness (mad people are jus

I have been so excited for this one since reading Stirred up 1 and 2. When S.E Hall and Angela Graham team up for an erotic quickie it is going to be sexy,funny and uni
m his words. He is sure knows how to get a woman going and knowing that he likes to read is a plus. The mystery and way they communicate is so hot...I can not wait to r

You know when you get these two authors together,its going to be something special! They bought you Stirred Up 1 & 2 and that was 5 stars all the way! BUT PACKAGED! if

Looking for a QUICK super steamy read.. Look no further. This story will melt your panties clean off;) I'm loving these &$34;quickies&$34; from these two authors. Just

What a perfectly packaged erotic quickie.  I mean,never in my wildest dreams would I think doing the deed with an unknown man would be so hot,but SE Hall and Angela Gra

I am so in love with Shaw. This was an awesome book and cannot give it enough praises. I love how the book was laid out. Once again SE Hall and Angela Graham have creat

I was too bored and decided to buy a new book for myself without looking at the book summary. This book pops up right there at the top of the suggestions with a 5 star
Windows (CRLF)    Ln 26, Col 1    100%
```

Module 2

1. The trained data and test data are cleaned by removing stopwords, punctuation and alphabets and converted into pkl format.



2. The execution of proposed work in five epoch this picture contains loss, value loss, value accuracy and accuracy of the module



Module 4

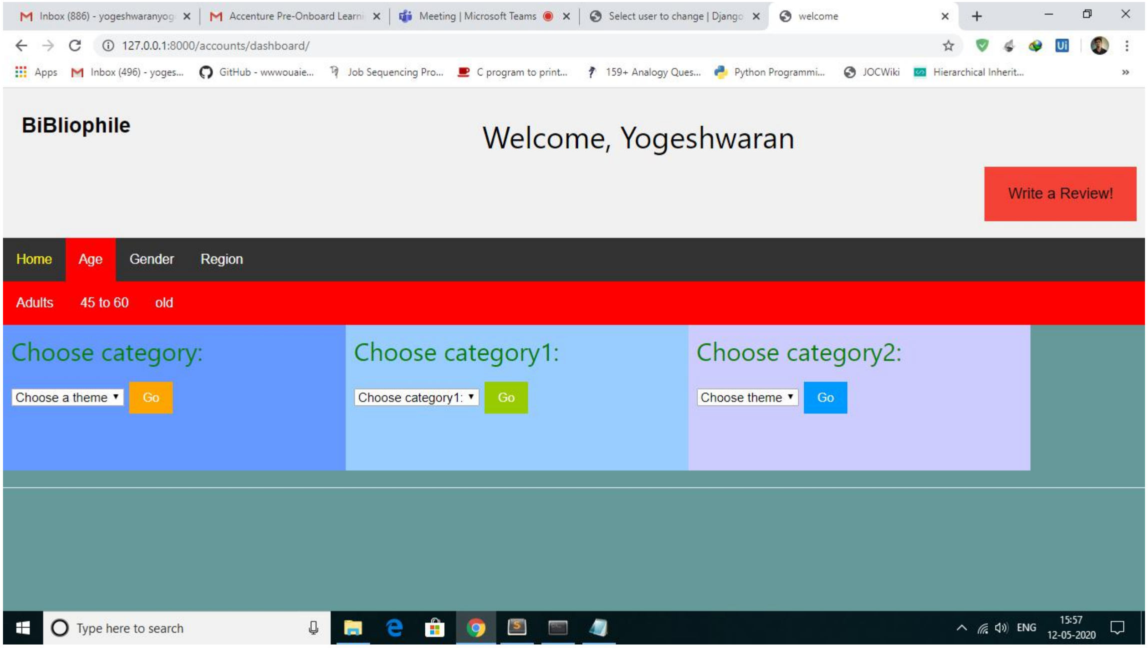
1. This is the screenshot of the recommendation which is given by the popularity model where the results are same for two users since it recommend the most popular items to the users.

```
The list of recommendations for the userId: 10

    rank  userId  score
9    1.0      10      7
12   2.0      10      7
18   3.0      10      7
40   4.0      10      7
2    5.0      10      6

The list of recommendations for the userId: 100

    rank  userId  score
9    1.0     100      7
12   2.0     100      7
18   3.0     100      7
40   4.0     100      7
2    5.0     100      6
```

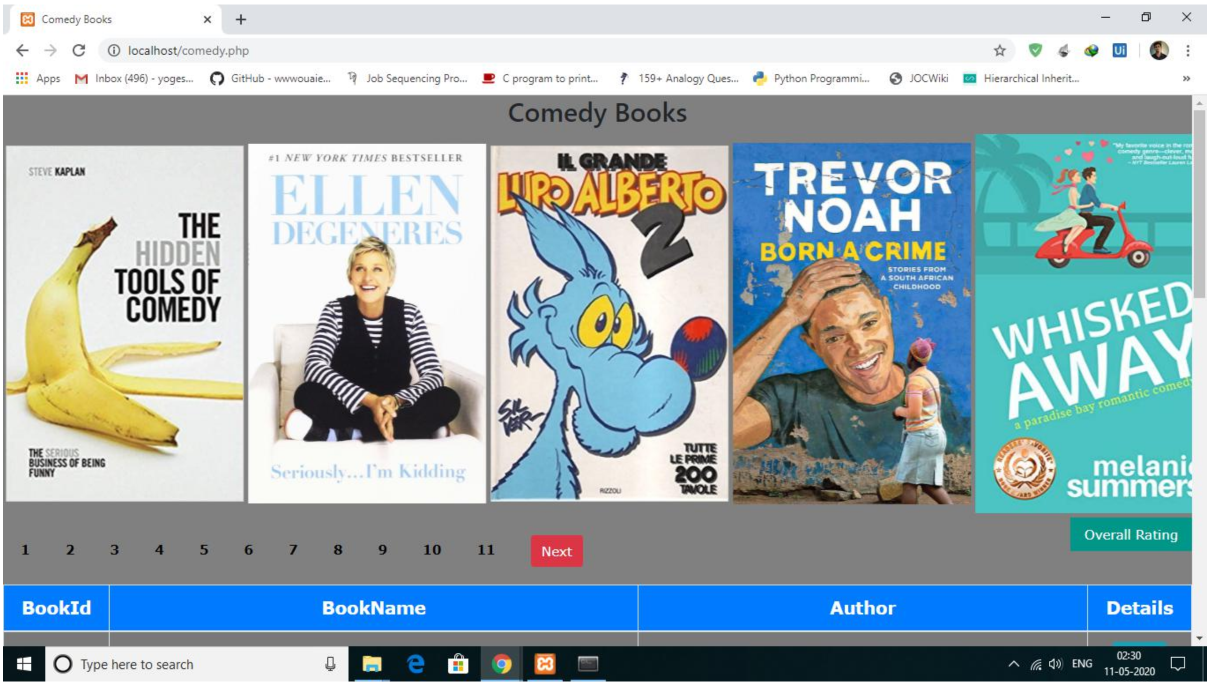
2. This screenshot shows the recommendation of books to the users based on collaborative filtering and we can note that the books recommended to each user id is different and it is based on the collaborative filtering.

```
[ ]
Below are the recommended items for user(user_id = 4):

                   user_ratings  user_predictions
Recommended Items
253                         0.0      6.373814e-16
73                          0.0      3.463204e-16
241                         0.0      2.289596e-16
188                         0.0      2.104843e-16
235                         0.0      1.739557e-16

[ ] userID = 59
    num_recommendations = 5
    recommend_items(userID, pivot_df, preds_df, num_recommendations)

Below are the recommended items for user(user_id = 59):

                   user_ratings  user_predictions
Recommended Items
140                         0.0      2.746916e-16
18                          0.0      8.823644e-17
144                         0.0      3.651063e-17
188                         0.0      3.050763e-17
245                         0.0      7.592663e-31
```

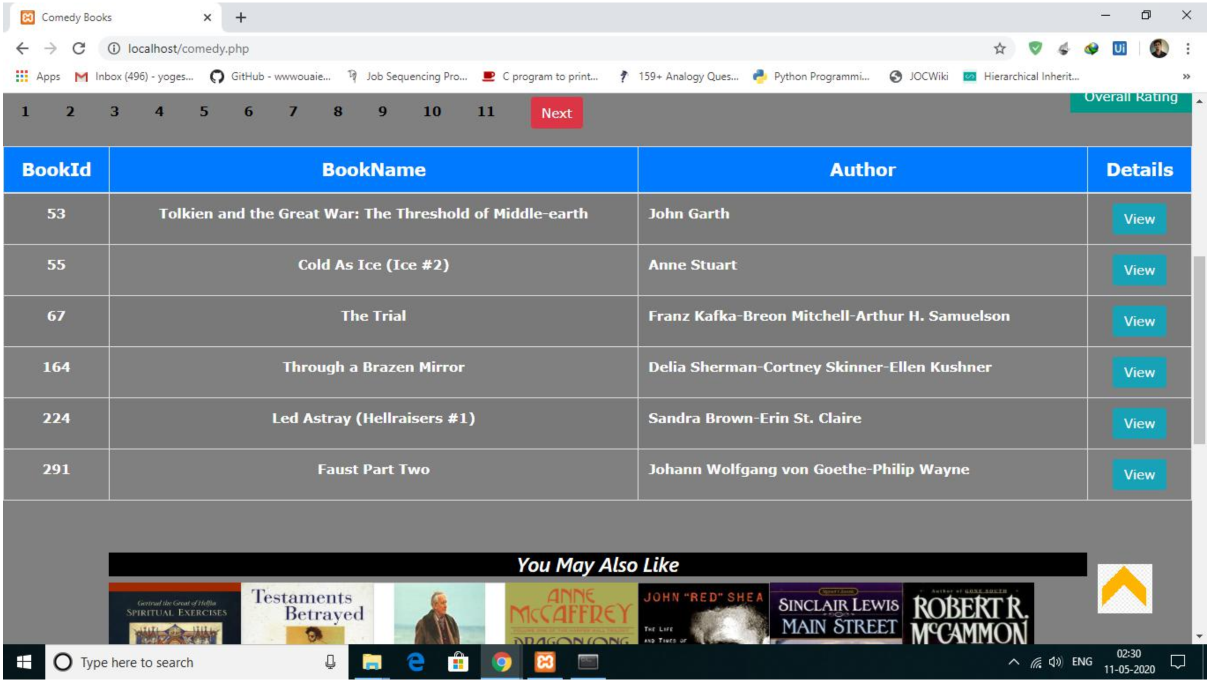
3. The home page for the recommendation user interface developed where the book are divided into three categories. Which has option to review a book and various navigation based on age, gender and region.



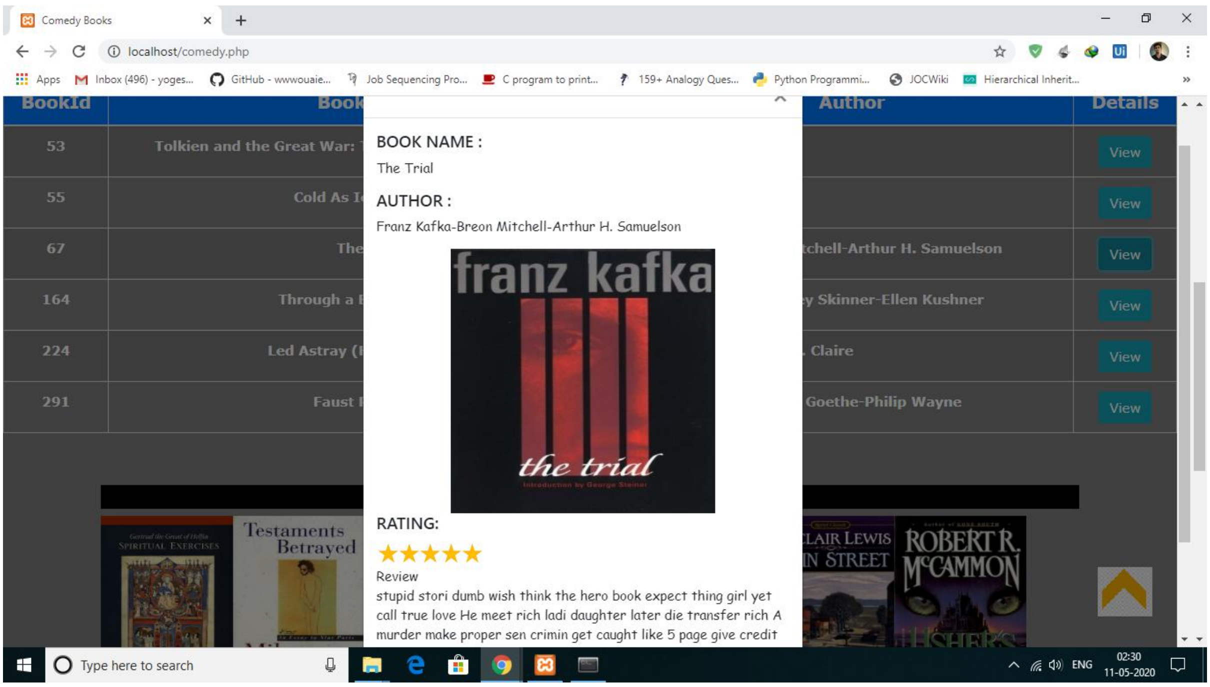
4. This is the comedy category selected in which the latest books are recommended on the top of the page



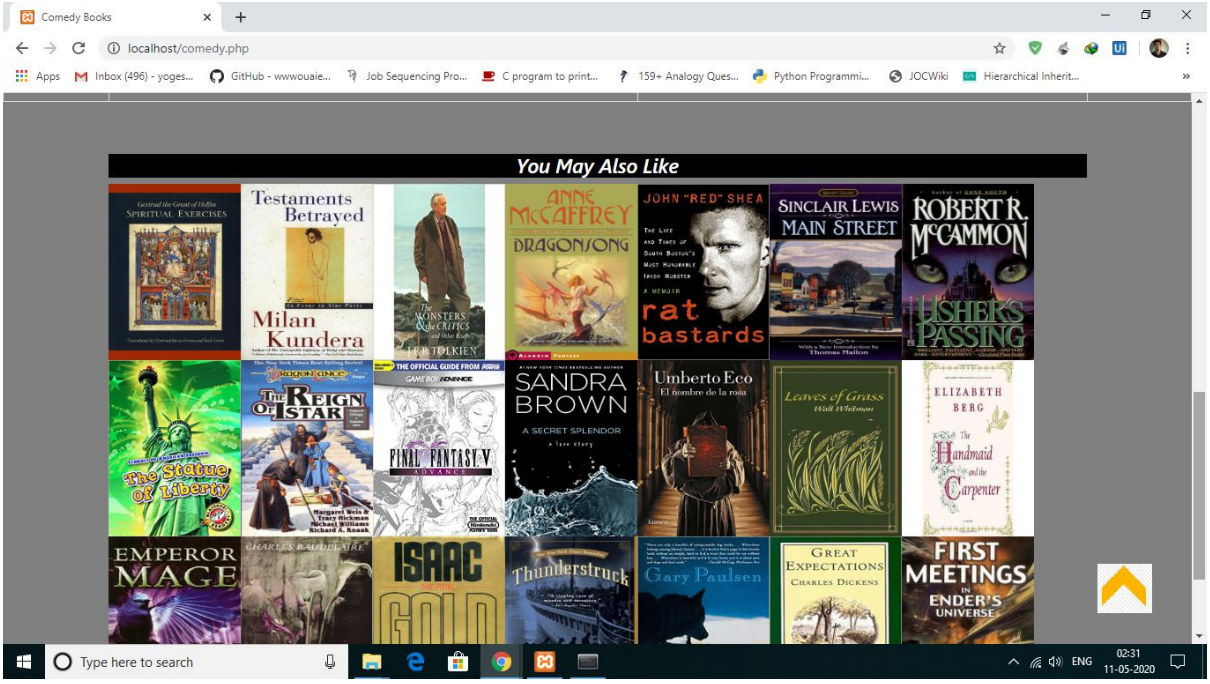
5. It represents the recommended books for the users where using view they can see the details of books and this table has books which has top ratings arranged on top page.



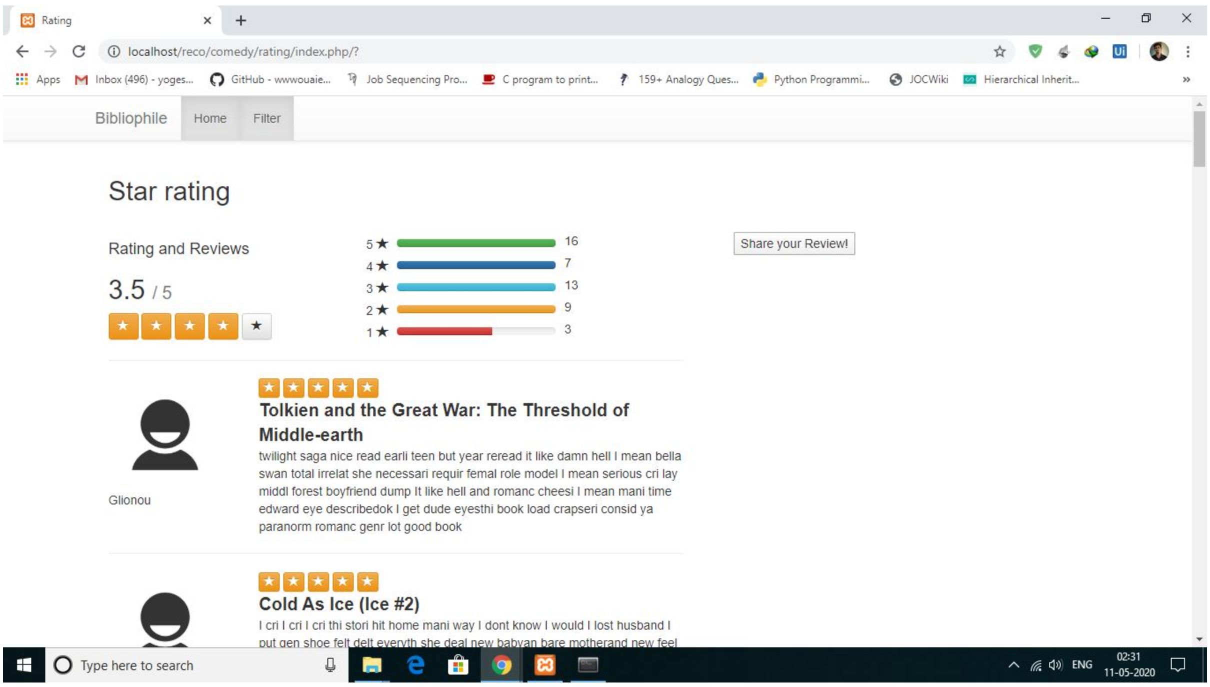
6. This is the view option which pop's up and in which the user can find the details of books like book name, author, rating and the review of the book given by the reviewers.



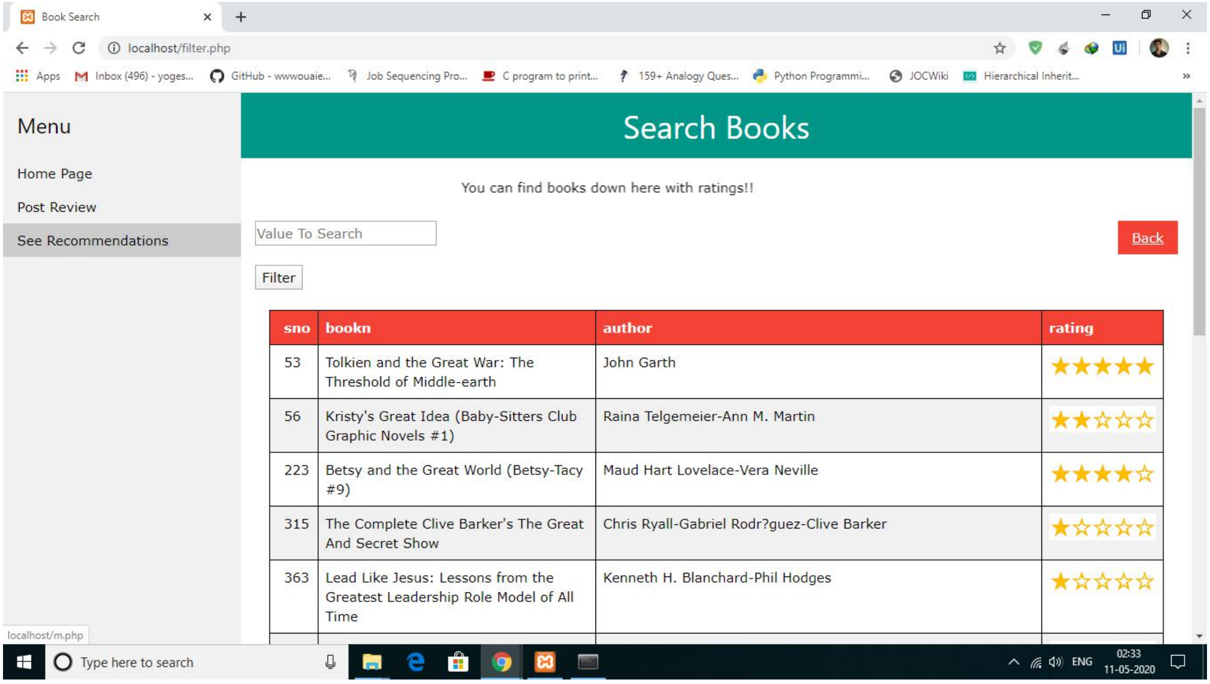
7. At the bottom of the book there are yet other recommendation for the users from the combination of children with other category.



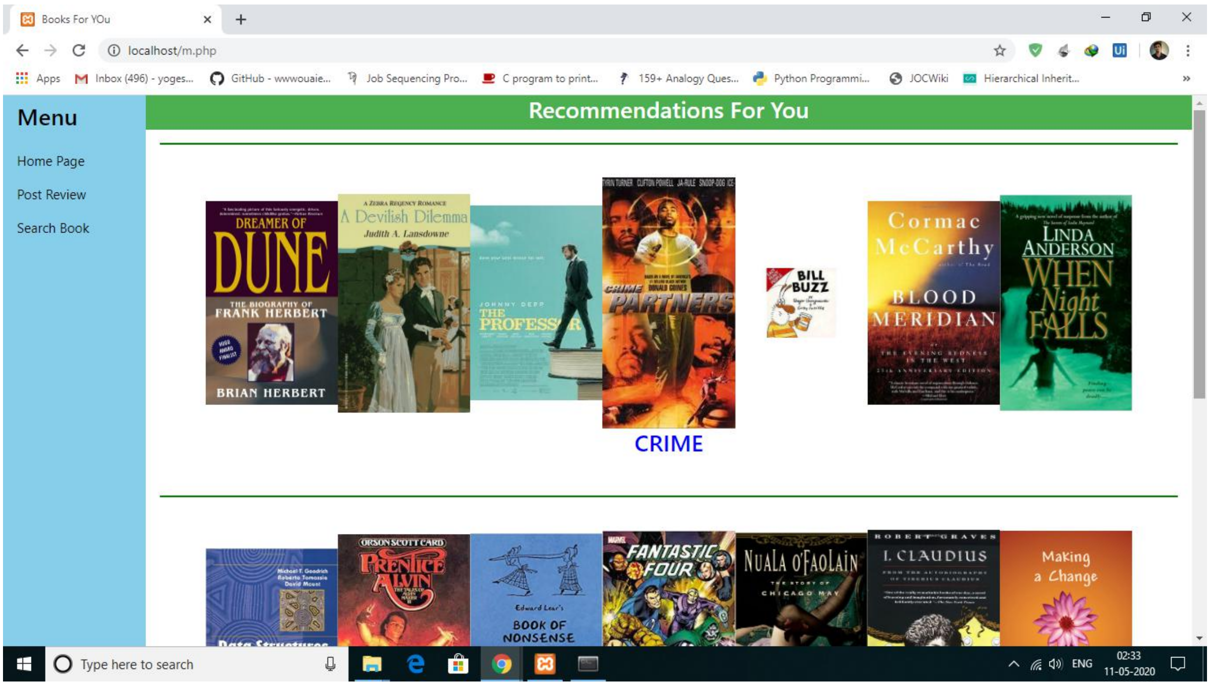
8. the overall rating table is created for each category where the total number of reviewers for each category is taken along with their rating and display the average rating for that category along with top reviews with book name and reviewer name for the books which are posted latest in that category.



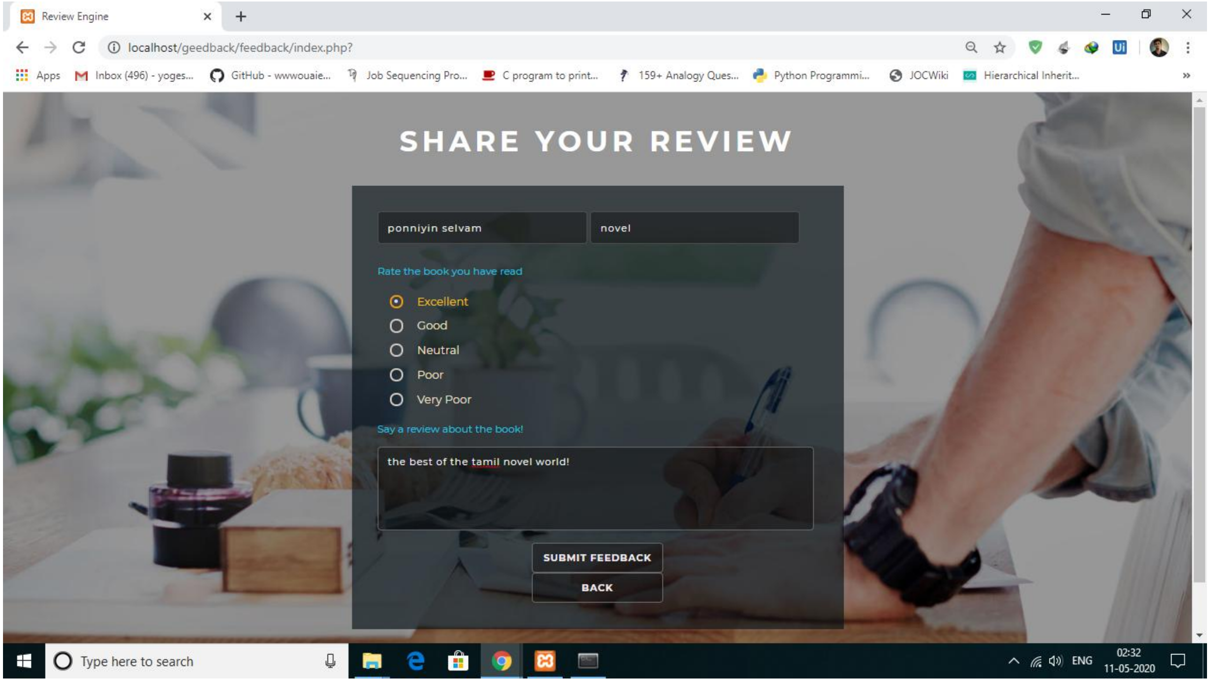
9. Search books page has the facility to search the books based on voice or by typing where the user can search a book based on just a key word from a book.



10. The recommendation of books for the users based on item to item collaborative filtering is seen in this page.



11. In this page the user can review a book by giving rating and posting a review.



5.4 SENTIMENT ANALYSIS – GRAPHS

Keras provides the capability to register callbacks when training a deep learning model. One of the default callbacks that is registered when training all deep learning models is the History callback. It records training metrics for each epoch. This includes the loss and the accuracy (for classification problems) as well as the loss and accuracy for the validation dataset, if one is set.

The history object is returned from calls to the fit() function used to train the model. Metrics are stored in a dictionary in the history member of the object returned. For a model trained on a classification problem with a validation dataset, this might produce the following listing:

['accuracy', 'loss', 'val_accuracy', 'val_loss']

Data Collected In The History Object To Create Plots

The plots can provide an indication of useful things about the training of the model, such as:

1. It's speed of convergence over epochs (slope).
2. Whether the model may have already converged (plateau of the line).
3. Whether the mode may be over-learning the training data (inflection for validation line).

Each deep learning algorithm has two graphs drawn,

Graph 1: Model Accuracy

A plot of accuracy on the training and validation datasets over training epochs.

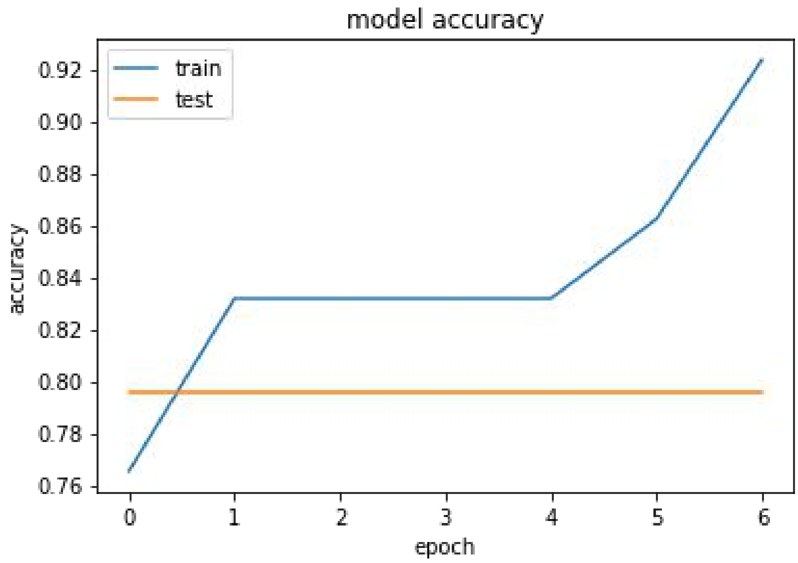
Graph 2: Model Loss

A plot of loss on the training and validation datasets over training epochs.

Note: The graphs are drawn using code in python 3.x from the model executed. And for validation 20% is for testing and 80% for training is given.

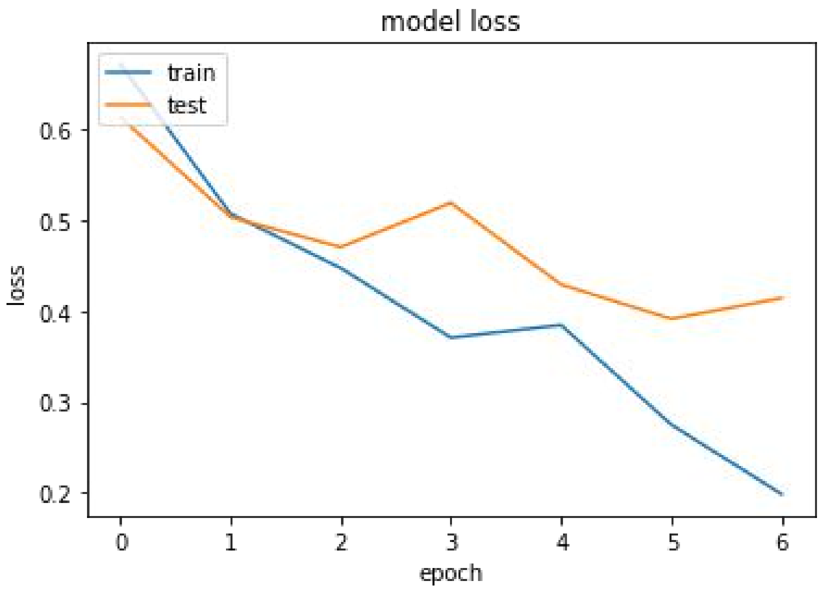
5.4.1 Convolutional Neural Network (CNN) - Accuracy: 75.81

Graph 1:



The graph is over the model accuracy for the validation set with the training data. The Y-axis shows the accuracy and X-axis shows the amount of epochs. It shows an accuracy of 75.81% over testing data and 92% over the training dataset.

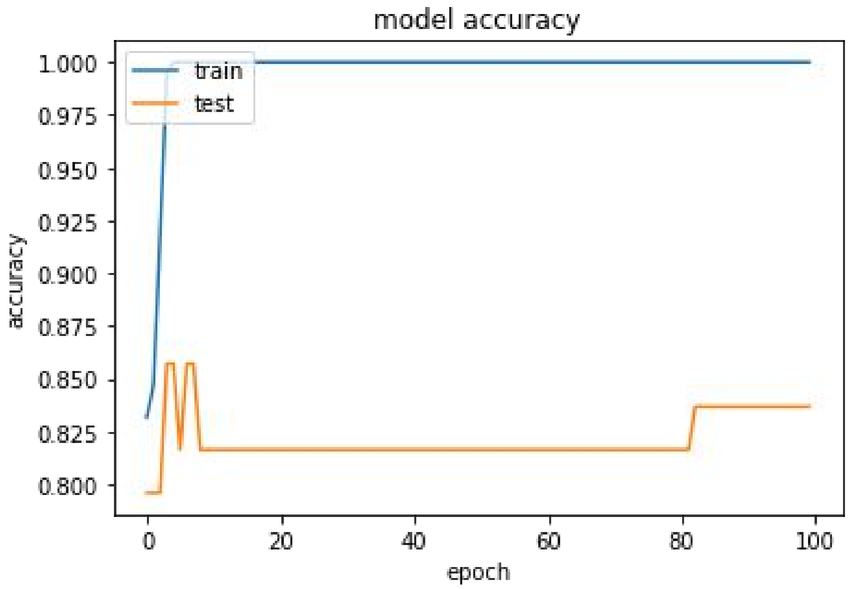
Graph 2:



The graph is over the model loss for the validation set with the training data. The Y-axis shows the amount of loss and X-axis shows the amount of epochs.

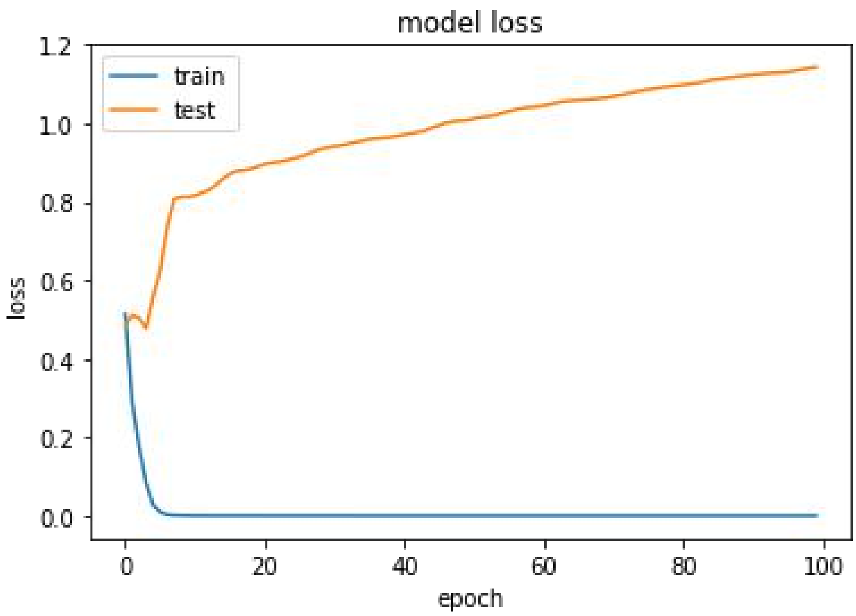
5.4.2 Word2Vec + CNN - Accuracy: 85.48

Graph 1:



The graph is over the model accuracy for the validation set with the training data. The Y-axis shows the accuracy and X-axis shows the amount of epochs. It shows an accuracy of 85.48% over testing data and 99.8% over the training dataset.

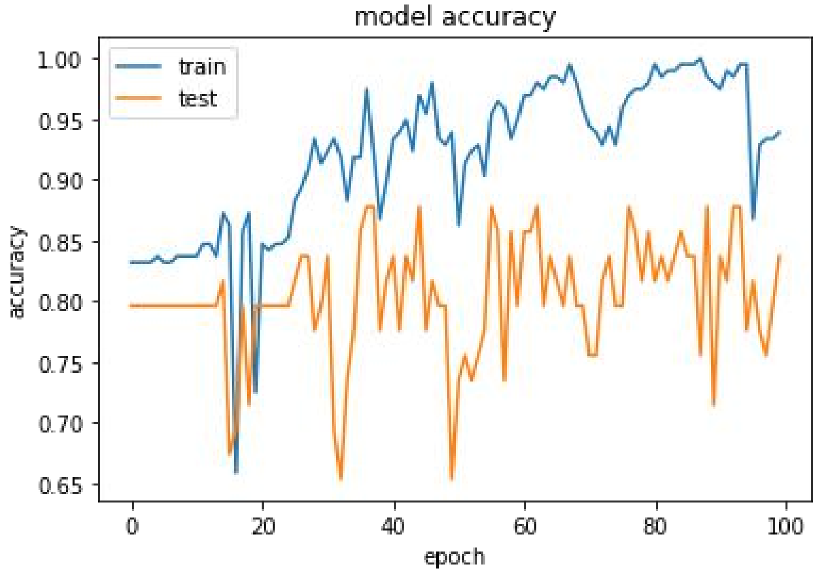
Graph 2:



The graph is over the model loss for the validation set with the training data. The Y-axis shows the amount of loss and X-axis shows the amount of epochs.

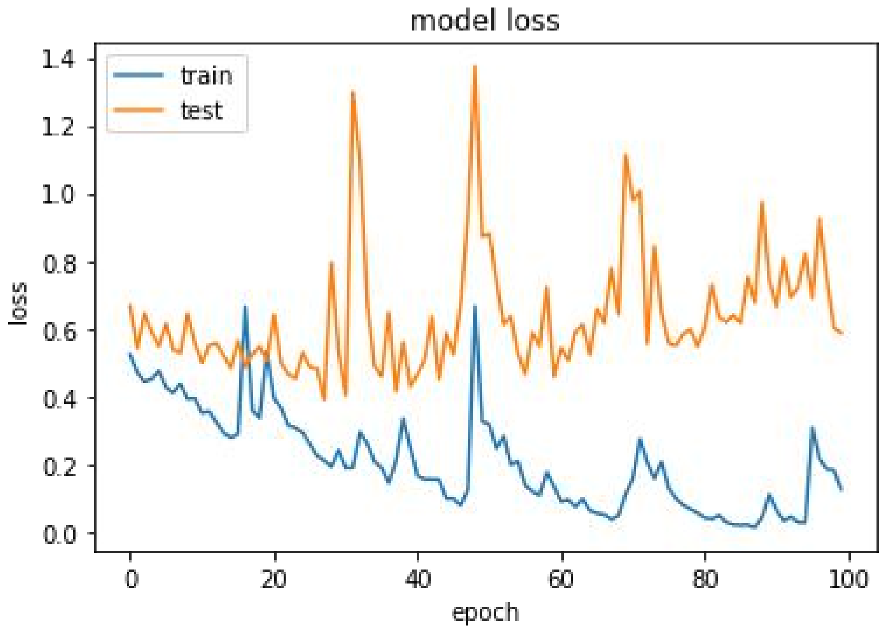
5.4.3 Doc2Vec With Word Embedding - Accuracy: 74.19

Graph 1:



The graph is over the model accuracy for the validation set with the training data. The Y-axis shows the accuracy and X-axis shows the amount of epochs. It shows an accuracy of 74.19% over testing data and 94% over the training dataset.

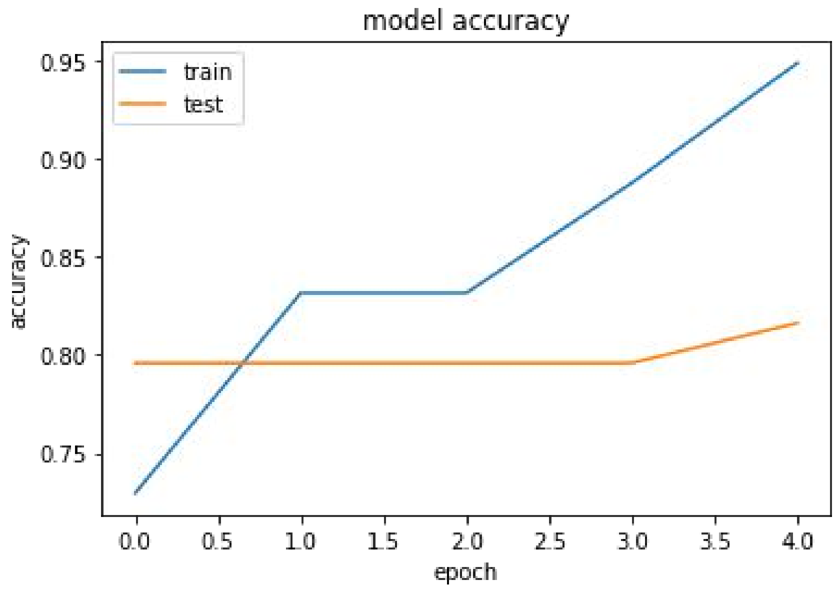
Graph 2:



The graph is over the model loss for the validation set with the training data. The Y-axis shows the amount of loss and X-axis shows the amount of epochs.

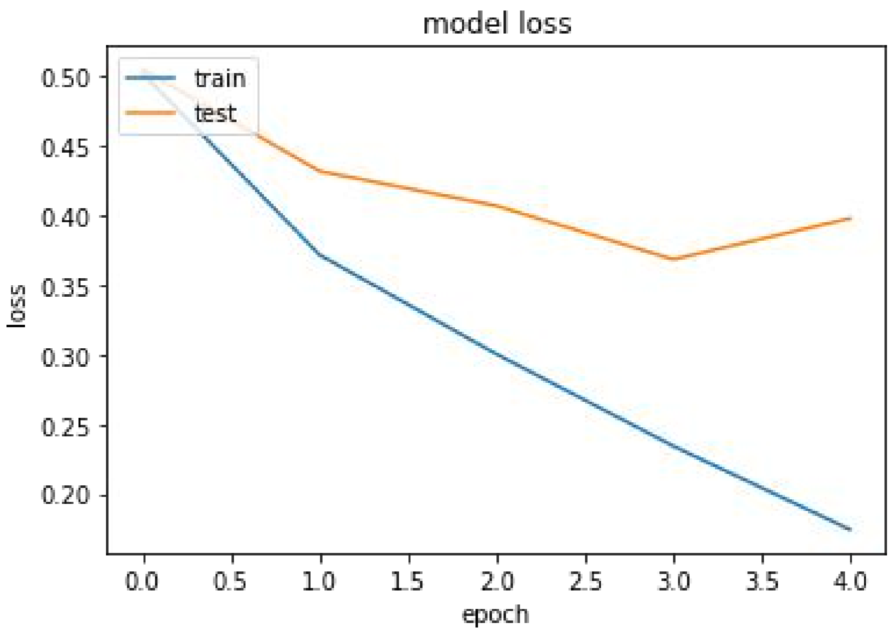
5.4.4 Doc2Vec Distributed Memory Concatenation - Accuracy: 81%

Graph 1:



The graph is over the model accuracy for the validation set with the training data. The Y-axis shows the accuracy and X-axis shows the amount of epochs. It shows an accuracy of 81% over testing data and 95% over the training dataset.

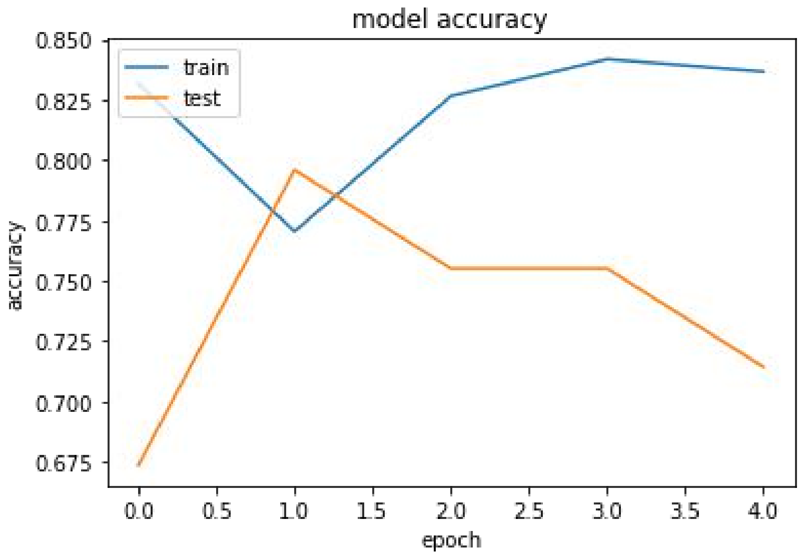
Graph 2:



The graph is over the model loss for the validation set with the training data. The Y-axis shows the amount of loss and X-axis shows the amount of epochs.

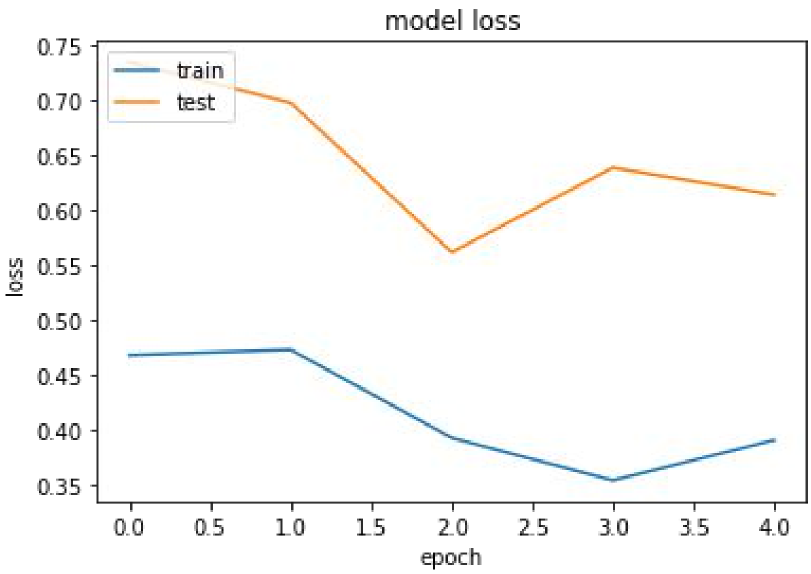
5.4.5 Doc2Vec Distributed Memory Mean - Accuracy: 75.81

Graph 1:



The graph is over the model accuracy for the validation set with the training data. The Y-axis shows the accuracy and X-axis shows the amount of epochs. It shows an accuracy of 75.81% over testing data and 83.50% over the training dataset.

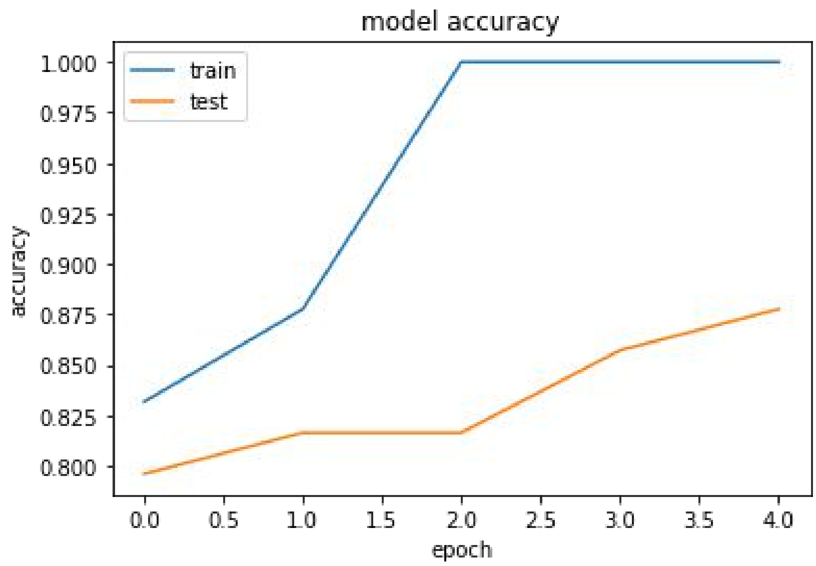
Graph 2:



The graph is over the model loss for the validation set with the training data. The Y-axis shows the amount of loss and X-axis shows the amount of epochs.

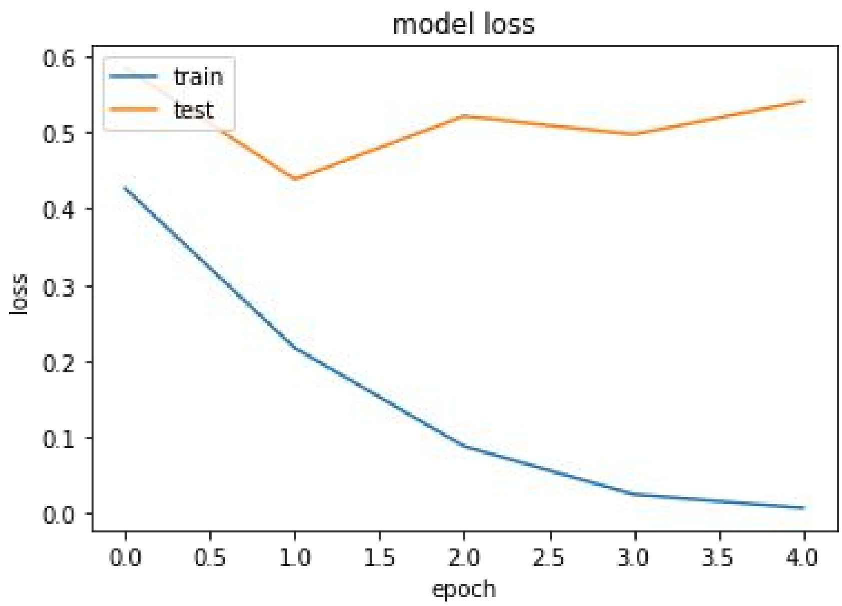
5.4.6 Doc2Vec Distributed Bag Of Words + Distributed Memory Concatenation - Accuracy: 90.32

Graph 1:



The graph is over the model accuracy for the validation set with the training data. The Y-axis shows the accuracy and X-axis shows the amount of epochs. It shows an accuracy of 90.32% over testing data and 100% over the training dataset.

Graph 2:



The graph is over the model loss for the validation set with the training data. The Y-axis shows the amount of loss and X-axis shows the amount of epochs.

5.5 RESULT

The implementation of the proposed algorithm is carried out on Polarity book review dataset. Kfold cross validation algorithm is implemented for testing.

Precision: It gives the exactness of the classifier. It is the ratio of number of correctly predicted positive reviews to the total number of reviews predicted as positive.

Recall: It measures the completeness of the classifier. It is the ratio of number of correctly predicted positive reviews to the actual number of positive reviews present in the corpus.

F-measure: It is the harmonic mean of precision and recall. F-measure can have best value as 1 and worst value as 0.

Accuracy: It is one of the most common performance evaluation parameter and it is calculated as the ratio of number of correctly predicted reviews to the number of total number of reviews present in the corpus.

Table 5.3 Evaluation Parameters

Evaluation Parameter	Formula
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F - measure	$\frac{2*Precision*Recall}{Precision+Recall}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$

Confusion matrix is generated to tabulate the performance of any classifier. This matrix shows the relation between correctly and wrongly predicted reviews. In the confusion matrix, TP (True Positive) represents the number of positive book reviews that are correctly predicted whereas FP (False positive) gives the value for number of positive book reviews that are predicted as negative. Similarly, TN (True Negative) is number of negative reviews correctly predicted and FN (False Negative) is number of negative reviews predicted as positive.

Table 5.4 Confusion Matrix for Proposed Work

Corrected Labels	Positive	Negative
Positive	130	46
Negative	31	98

The results obtained from various deep learning methods over our obtained dataset.

Table 5.5 Accuracy Calculation

Method	Polarity of reviews	Precision	Recall	F-Measure	Accuracy
CNN	Positive Negative	0.62 0.61	0.58 0.54	0.62 0.65	75.81%
CNN + Word2Vec	Positive Negative	0.75 0.69	0.71 0.72	0.76 0.78	85.48%
CNN + Doc2Vec	Positive Negative	0.79 0.70	0.75 0.67	0.73 0.80	74.19%
Doc2Vec + Distributed Memory Concatenation	Positive Negative	0.78 0.70	0.75 0.65	0.72 0.79	81%
Doc2Vec + Distributed Memory Mean	Positive Negative	0.72 0.69	0.69 0.63	0.66 0.70	75.81%
CNN + Doc2Vec (DMC) + DBOW (n-grams)	Positive Negative	0.85 0.80	0.77 0.83	0.82 0.84	**90.32%**

5.5.1 Result Of Recommendation System

Eighteen categories are considered for recommendation of books. There are three age groups namely adult, between 45 to 60 and old (above 60). Mainly the reviews which are analyzed are from the continents Australia, Europe and America. Presented below are the tables which show the people group and the category of books which they prefer the most.

Table 5.6 Categories Preferred By Various Groups

Sl. No.	Group	Most preferred category
	Based on age	
1.	Adult (below 45)	romance, sci-fi, drama
2.	Between 45 to 60	horror, fantasy, children
3.	old	document, literature, modern
	Based on gender	
1.	male	Document, sci-fi, literature
2.	female	drama, modern, comedy
	Based on region	
1.	Australia	animation, action
2.	America	adventure, horror, fantasy
3.	Europe	crime, thriller, comedy

Table 5.7 Categories Preferred By Various Countries

Sl. No.	Continent	Country	Age Group (Most preferred category)		
			Adult	Between 45 to 60	Old
1.	Australia	Melbourne	drama	Document	children
2.	America	United States	romance	action	thriller
		Canada	mystery	war	adventure
3.	Europe	Czech republic	Horror	Comedy	sci-fi
		Austria	war	Musical	drama
		Belgium	Mystery	Document	thriller
		Finland	Modern	Crime	document
		Greenland	Romance	Fantasy	horror
		Italy	Crime	Horror	Adventure
		Netherlands	Romance	Modern	literature
		New Zealand	Sci-Fi	Animation	literature

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

Sentiment analysis is important for any online retail company to understand customer's response. Reviews need to be analyzed for building recommendation algorithms to target customers based on their needs. In our work, we present the analyses of customer document level book reviews. This work present a method to semantically identify the positive and negative tagged words from the reviews and that tagged words are used in word embedding layers of convolution neural network to increase the performance of the sentiment analysis of the reviews with a proposed method.

The sentiment analysis results in the classification of reviews based on polarity as positive and negative. This also includes a comparative study of accuracy and loss of various classifiers. CNN with DMC and bow was able to give average accuracy of 90.32%.After the classification the books are recommended to the users based on item to item collaborative filtering and popularity recommendation algorithms which helps to recommend users the most appropriate books based on their reviews, age, gender and region.

The scope of this work can be extend by incorporating various recurrent neural network algorithms to analyze the reviews and then We can use hybrid recommender's like collaborative filtering with other Combination of recommending algorithms for Better recommendation algorithm.

REFERENCE PAPERS

[1] Asad Abdi, Aiti Mariyam Shamsuddin, Shafaatunnur Hasan and Jalil Piran, "***Deep Learning based Sentiment Classification of evaluative text based on multi feature fusion***" Information Processing and Management, Springer 2019.

[2] Rachana Bandana, "***Sentiment Analysis of Movie Reviews Using Heterogeneous Features***", 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) 2018.

[3] Ramon Zatarain Cabada et al., "***Mining of Educational Opinions with Deep Learning***" Journal of Universal Computer Science 2018.

[4] Arnav Chakravathy et al. "***Hybrid Architecture for Sentiment Analysis using Deep Learning***", International Journal of Advanced Research in Computer Science, February 2018.

[5] Aminu D'au et al. "***Sentiment-Aware Deep Recommender System With Neural Attention Networks***", Journal of Web Engineering, Vol. 11, No. 1 (April 2019) 051-063 © IEEE.

[6] Soufiene Jaffali, Salma Jamoussi, and Abdelmajid Ben Hamadou, "***Grouping Like-Minded Users Based on Text and Sentiment Analysis***", ICCCI 2014, LNAI 8733, pp. 83–93, 2014. c Springer International Publishing Switzerland 2014.

[7] Qi Ji et al, "***Research on Personalized Book Recommendation Model for New Readers***", 2018 3rd International Conference on Information Systems Engineering, 2160-1291/18/©2018 IEEE DOI 10.1109/ICISE.2018.00022.

[8] Jaspreet Kaur and Brahmaleen Kaur Sidhu, "***Sentiment Analysis Based on Deep Learning Approaches***", Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018).

[9] V. K. Kiran, R. E. Vinodhini, R. Archanaa and K. Vimalkumar, "***User Specific Product Recommendation and Rating System by Performing Sentiment Analysis on Product Reviews***", in International Conference on Advanced Computing and Communication Systems (ICACCS-2017), pp. 978-983, Coimbatore, 2017.

[10] Nour Nassar et al. "***A novel deep multi-criteria collaborative filtering model for recommendation system***" Journal of Knowledge Based Systems, June 2019 © 2019 Elsevier.

[11] Jinoh Oh, Sun Park et al, "***Novel Recommendation based on Personal Popularity Tendency***", 2011 11th IEEE International Conference on Data Mining, 1550-4786/11 © 2011 IEEE DOI 10.1109/ICDM.2011.110.

[12] Bolanle A. Ojokoh and Olumide Kayode, ***"A Feature–Opinion Extraction Approach To Opinion Mining"***, Journal of Web Engineering, Vol. 11, (2018)
© Rinton Press.

[13] Heng-RuZhang, FanMin, XuHe, and Yuan-YuanXu, "***A Hybrid Recommender System Based on User-Recommender Interaction***", Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2015, Article ID 145636, http://dx.doi.org/10.1155/2015/145636.

[14] Ricardo B. Scheicher et al, "***Sentiment classification improvement using CNN with semantically enriched information***", in ACM Symposium on Document Engineering 2019 (DocEng '19), September 23–26, 2019, Berlin, Germany. ACM, New York, NY, USA, https://doi.org/10.1145/3342558.3345410.

[15] Chaloemphon Sirikayon et al, "***A Collaborative Filtering Based Library Book Recommendation System***", 2018 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 978-1-5386-5254-1/18/$31.00 ©2018 IEEE.

[16] K. S. Srujan, S. S. Nikhil, H. Raghav Rao, K. Karthik, B. S. Harish and H. M. Keerthi Kumar, "***Classification of Book Reviews Based on Sentiment Analysis***", Information Systems Design and Intelligent Applications, Springer, vol. 672, pp. 401-411, 2018.

[17] Shuai Zhang, Lina Yao, Aixin Sun and Yi Tay, "***Deep Learning Based Recommender System: A Survey and New Perspectives***", Deep Learning Based Recommender System: A Survey and New Perspectives. ACM Computer Survey. 52, 1, Article 5 (February 2019), https://doi.org/10.1145/3285029.

[18] Tong Zhao, Julian McAuley, Irwin King, "***Leveraging Social Connections to Improve Personalized Ranking for Collaborative Filtering***", CIKM'14, November 3–7, 2014, Shanghai, China. Copyright 2014 ACM 978-1-4503-2598-1/14/11 http://dx.doi.org/10.1145/2661829.2661998.

LIST OF PUBLICATION

1. PRESENTATIONS IN NATIONAL CONFERENCES

1. **Yogeshwaran. T, Neena Jasmine. S, Kishore Kumar. B and Dr. S. Saraswathi,** "Sentiment Analysis of Book Reviews Using Semantic Network", International Innovative Research Journal of Engineering and Technology ISSN: 2456-1983 Vol: 5 Issue: 3 March 2020 *TEQIP – III National Conference on Big Data for Health Prediction (NABDAHP - 2020), Jan.09-10, 2020, PEC, Puducherry.*