**ENCS5342– Second Sem 2022-2023 – Course Project**

**Optimal Review Ranking for Improving Shopper's Decision Making**

Sara Ammar, Hala Ziq, Ayham Maree

1191052@student.birzeit.edu 1191637@student.birzeit.edu 1191408@student.birzeit.edu

***Abstract***

In the digital age, online shopping has gained significant popularity, providing convenience and a wide range of product options. However, navigating through a vast volume of user-generated reviews poses a challenge for online shoppers, hindering their ability to make informed purchasing decisions. Traditional ranking techniques fall short in delivering personalized recommendations for users. This project aims to reinvent the review ranking process by considering various factors, including previous user reviews for other products. The goal is to develop an Arabic reviews rating system that leverages algorithms and machine learning techniques to offer tailored review recommendations based on individual user preferences and contextual aspects. By providing users with personalized and relevant reviews, the proposed system aims to enhance decision-making capabilities and improve user satisfaction, thereby revolutionizing the online shopping experience.

1. **INTRODUCTION**

In the digital era, online shopping has become increasingly popular, offering convenience and puts a huge selection of products at our fingertips. However, one challenge faced by online shoppers is navigating through the immense volume of user-generated reviews to make informed purchasing decisions. With popular items amassing countless reviews, it can be arduous for individuals to find relevant and trustworthy information that aligns with their unique preferences and needs.

Simple ranking techniques, such as sorting reviews by time, rate, or helpfulness votes, are frequently used on traditional online buying platforms. While these techniques might offer some basic insights, they fall short of delivering a customized experience for every user. Is it possible to reinvent the review ranking process by taking into account numerous elements, such as the previous user reviews for other products? This project seeks to answer this question.

Our project's goal is to create an Arabic reviews rating system that empowers customers to make wiser decisions by offering tailored review recommendations. We want to develop a system that takes into account both individual user preferences and contextual aspects in addition to filtering through the massive volume of user-generated material by leveraging the power of algorithms and machine learning techniques.

We will use the user's reviews on the products to learn more about their cares to tailor the review recommendations based on the user's interests and areas of concentration. We will also be able to better understand the relevance of reviews to certain customers.

Our proposed system will significantly improve the online shopping experience by eliminating the frustration of sifting through numerous reviews. Instead, users will be presented with a personalized and curated set of reviews that are most relevant to their unique requirements. By offering tailored recommendations based on a holistic understanding of the user, we aim to enhance decision-making capabilities and boost user satisfaction.

the Optimal Review Ranking project seeks to revolutionize the way shoppers’ access and assess user-generated reviews. By harnessing the power of personalized recommendations, our system will empower users to make more informed decisions, saving time and ensuring a seamless and satisfactory online shopping experience.

1. **DATA DESCRIPTION & ANALYSIS**

**2.1.1 BRAD: Books Reviews in Arabic Dataset**

This dataset contains 510,600 book reviews in Arabic language. The reviews were collected from GoodReads.com [1] website during June/July 2016.

We download the bal-clean-reviews.tsv file from

BRAD: Books Reviews in Arabic Dataset [2] GitHub repository.

Format of the data set:

rating: 2.0 review\_id: “1657686339”

user\_id: “22103652” book\_id: “13637412”

review: “لطيفه :). كأنك بتتفرج ع مسلسل بس نوعا ما لطيف”

Some of the terms are explained below:

Rating: rating of the book, it takes 1,2, for bad book and 4,5 for good book.

Review\_id: Unique review identifier.

User\_id: Unique reviewer(user) identifier.

Book\_id: Unique book identifier.

Review: Text of the review.

**2.1.2 BRAD Analysis**

in Book Review in Arabic Dataset, there are totally 231392 unique review, 27530 unique user, and 10620 unique books, in the summary table presented below, we can see that the median number of reviews by a user is 6 and the median number of reviews associated with each book in the dataset is 9. Furthermore, the maximum number of reviews for a product is 837.

|  |  |  |  |
| --- | --- | --- | --- |
|  | #of user reviews | #of book reviews | review length |
| Mean | 8.405 | 21.788 | 306.849 |
| Min | 3 | 1 | 0 |
| Max | 264 | 1880 | 23157 |
| std | 6.26 | 62.696 | 544.553 |
| 25% | 5.0 | 6.0 | 47.0 |
| 50% | 6.0 | 9.0 | 139.0 |
| 75% | 9.0 | 18.0 | 337.0 |

Table 1 Summary Table BRAD Analysis

**2.2.1 Filtered BRA Dataset**

As the BRAD analysis show that the dataset is huge, that mean it maybe has unhelpful and unnecessary data to use on the model, so we filtered the data to increase the accurse of the model, and choose the reviews that implement the following conditions:

1. Review length is between 20 and 50 words.
2. User with reviews count between 20 and 40 reviews.

Choose 10 random users from the filtered dataset after applying the above conditions.

**2.2.2 Filtered BRAD Analysis**

The filtered dataset contains 10 random users with their all reviews on different books, where there was 242 unique reviews for all 10 users, the data which shown on the table below determent that each user comment at max 2 review on the same books, this mean the selected data was good to use it to train the ranking model.

|  |  |  |  |
| --- | --- | --- | --- |
|  | #of user reviews | #of book reviews | review length |
| Mean | 24.2 | 1.038 | 852 |
| Min | 20 | 1 | 660 |
| Max | 39 | 2 | 1337 |
| std | 5.493 | 0.193 | 198.86 |
| 25% | 21.25 | 1 | 719.5 |
| 50% | 22.5 | 1 | 803 |
| 75% | 24 | 1 | 889.5 |
|  | Sentiment score | Rating |
| Mean | 0.0021 | 2.501 |
| Min | -0.3661 | 3.27 |
| Max | 0.3461 | 1.75 |
| std | 0.2406 | 0.502 |
| 25% | -0.162 | 2.15 |
| 50% | -0.0042 | 2.49 |
| 75% | 0.196 | 2.94 |

Table 2 Summary Table Filtered BRAD Analysis for each user

**2.3.1 Filtered BRAD Data Preprocessing**

To achieve the user satisfaction on the optimal reviews

ranking model and improving the shoppes’ to make a

perfect decision, we apply some preprocesses on the

filtered dataset to train the model and gives more

accurse results, so we follow these steps:

1. **Data Clean:** remove punctuation, numbers, links, underscores, and parentheses, etc, from the review text.

Original review text:

“لطيفه :). كأنك بتتفرج ع مسلسل بس نوعا ما لطيف ”

Clean review:

“لطيفه كأنك بتتفرج ع مسلسل بس نوعا ما لطيف”

1. **Stemming:** we stem the review text to reduce the vocabulary size and improving computational efficiency in text analysis tasks.

Clean review:

“لطيفه كأنك بتتفرج ع مسلسل بس نوعا ما لطيف”

Stemmed Review:

“لطيف كأن تفرج ع مسلسل بس نوع ما لطيف”

1. **Extract Keywords:** after Stemming the review, we extract the noun and the adjectives from the stemmed review text to find the most users cares.

Stemmed Review:

“لطيف كأن تفرج ع مسلسل بس نوع ما لطيف”

Extracted keywords Review:

“[لطيف, نوع, بس, مسلسل, كأن]”

1. **Stop Word Removal:** after Extracting keywords step, we remove the Arabic stop words and other words without meaning from the stemmed noun and adjectives which extracted from the review text.

Extracted keywords Review:

“[لطيف, نوع, بس, مسلسل, كأن]”

Removed stop word Review:

“مسلسل نوع لطيف”

**3.0 Work Methodology**

to improve the decision making and more satisfying user experience, the model should rank the reviews and the recommendation of the book on personalized taking into consideration users cares and their interests, to achieve this model of personalized review ranking we used the following work Methodology Steps:

**3.1 Sentiment Analysis**

Sentiment analysis is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, study affective states and subjective information, and used to determine the sentiment or emotional tone expressed in a piece of text. [3]

Natural language processing (NLP) techniques are frequently used by sentiment analysis algorithms to extract relevant features from the text and award sentiment scores depending on specified sentiment.

Sentiment score range takes values between [-1,1] where this score is graded as following:

* A score of [-1,0[ represents the negative sentiment for the review text.
* A score equal 0 represent the natural sentiment for the review text.
* A score of ]0,1] represents the positive sentiment for the review text.

We apply the sentiment analysis for the review text, because there was a conflict on the dataset between the book rating and the review text, like when the rating is less than 3 and the sentiment score was positive.



Figure 1 Sample of Dataset Conflict

**3.2 Keywords Extraction**

the automated process of extracting the most relevant words and expressions from review text. we use this process to break down the human language and let the model able to analyze the review text, so to do this process we follow these steps: [4]

1. Part-of-Speech Tagging.
2. Stemming
3. Apply frequency-based method (TF-IDF)

**3.2.1 Part-of-Speech Tagging (Pos Tagging)**

POS tagging is an important natural language processing application used in machine learning, the main issue that must be addressed in part of speech tagging is that of ambiguity.

We do this step to assign a pos tags for each word in the review text, (eg., verb, noun, adjective,…,etc) then we filter the pos tags list from verb and other unnecessary tags to create user profile by collect the nouns and adjectives from review text. [5]

**3.2.2 Stemming**

we use this technique to extract the base for each word after apply pos tagging step to has more accurse result when apply frequency-based method to find the TF-IDF for each word on the review text.

**3.2.3 Apply Frequency-base Method (TF-IDF)**

We use this method to assign a weight for each term on the review text by find the term frequency on the review and find the inverse document frequency.

We use the TfidfVectorizer library which tokenize the review text then gives weight for each term according this equation: [6]

Where:

After applying these three steps we can extract the keywords by filtering the most importance word on the review text which has high weight or score.

**3.3 User Profiling**

To construct the users profiles from the reviews that they commented it on different books, after applying the TF-IDF technique some words has a more than one weights because it was exist in more than one review text, so we assign the weight of these words on the same user corpus by **the sum of these weights**, we preferred the sum on choosing the Maximum weight of this word or Mean value of the weights, because there was words has high weight, but it frequent on one review text, that means the user is less interest with these words than the most frequent words.

we add more vectors(eg, rating, sentiment score) to update the weights, so we apply different equation for each case:

**Case I: Reviews Text with conflict**

In this case we check the value of sentiment score if it less than 0 and the rating is more than 3 and

visa versa, the weight of the term remains the same.

Where:

−IDF Technique

**Case II: Review Text without Conflict**

In this case we check the value of sentiment score if it less than zero and the rating less than 3, else if the sentiment score more than zero and the rating more than 3, to let the sentiment score and the rating improve the weights we apply the following equation:

Where:

−IDF Technique.

: semantic score for review j

rating: user rating for review j.

After updated the weight for each term we choose the to 30 terms that have the maximum weights to construct the user profile.



Figure 2 user profile for user id 4047738

The above figure shows the top 30 terms that user interest with and use it frequently.

After we construct the user profile, we use it to be the query that used to find the similarity between it and the documents(reviews)

**3.4 Personalized Ranking of Reviews**

To rank the book reviews according to the user profile, we take randomly 10 unique books from BRA Dataset, then we apply stemming process only on the review text.

Now, we have the documents(reviews), and the query (user profile), so we will pass it for BM25 model in elastic search engine to ranking the top 5 reviews of specific book according the user profile.

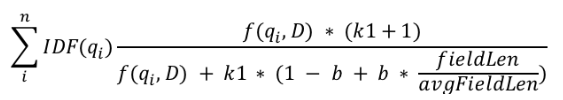
* **Elastic Search Engine:**

Is an engine designed to handle huge data, then apply fast search algorithm for different type of data.

Elastic search takes the data (documents) then indexing the documents, then let the user to apply complex query the data provided, it makes a tokenization for each word on the document and store it on inverted index list, and identifies all of the documents each term occurs on it, to make the review researchable near real time, it use BM25 as a default model for ranking. [7]

* **BM25:**

Is a ranking method used on elastic search engine to find the similarity and the relevance between the documents and the query that passed to elastic search, BM25 apply the following formula to rank the documents:



Where:

.

f(: the term i frequency of the query on the document D.

firldlen: length of the document

avgFieldLen: average documents length on the collection.

K1,b: tuning parameter to control the term frequency and the documents length.

BM25 is an improvement over TF-IDF that takes into account the length of the document and the average length of documents in the corpus. It also uses a probabilistic approach to ranking documents based on their relevance to a query, so it's considered it was more effective than TF-IDF Ranking. [8]

After pass the query and the documents to BM25 algorithm the top-ranking reviews for

book id **“13575970”** and the user **“4047738”** profile was shown below.

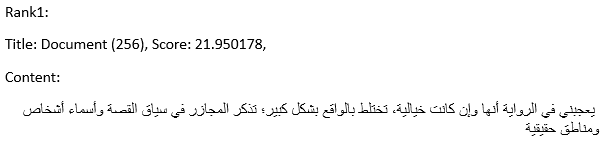
****

Figure 3 Rank 1 Review for book 13575970 for user 4047738 profile

**Note: the other reviews ranking is attached on the**

**Zip files.**

**4.0 NLP Tools & APIs Used**

1. **AraBERT:** it was used to put the sentiment score on the review text, and it is a modern model that analyzes the sentiments on the Arabic text and gives it a polarity score towards the positive, negative and neutral categories. It was trained on a large corpus of Arabic language data.
2. **Farasa:** It was used for segmentation, stemming and for part-of-speech tagging for reviews. It is an advanced tool that provides a range of text analysis capabilities. Farasa takes advantage of deep learning techniques to accurately analyze the Arabic text. It simplifies and allows for a deeper understanding of the structures and meanings in the Arabic language.
3. **NLTK's**: The NLTK's model was used to remove the stop words of the Arabic language, which are common and unimportant words such as prepositions, for example, and removing them can reduce the size of the data and the noise in it.
4. **TfidfVectorizer:** it was used for extracting features efficiently from the review text.
5. **Elastic Search:** it is a powerful search and analytics engine, it was used here for analyzing book reviews with the BM25 ranking function, determine the relevance of documents to a given query with the aid of this rating function, which is perfectly suited for information retrieval work.

**5.0 Limitation and Future works**

In this research, we forced several limitations, most notably the inability to access the personal information of users on shopping sites, such as their shopping activities, their browsing list, or their subscriber groups. We just have users’ reviews on the products in addition to their rating of that product, which was sometimes inconsistent with the sentiment analysis that we did for the text. In addition, the sources of data are few in the Arabic language, and the colloquial language that exists within the Arabic text, the tools that we have used, are unable to deal with or understand it.

In order to work on developing the effectiveness of rating systems for optimal review to improve the shopper's decision-making process, it is possible to develop the way users' personal files work by using the history of shopping, browsing and searching activities owned by shopping sites. However, priority must be given to the privacy and security of user data and ensuring the processing of personal information and browsing history. Developing such a system also requires large-scale data collection and analysis. Furthermore, it is important to design the system to deal with biases and unfair influence; Because it may lead to a skewed classification of the review, which may harm certain users or products, the system must also be adaptable and scalable to accommodate the dynamic nature of online shopping platforms, especially as it is increasing over time as products, categories, and specific preferences by users of certain products increase. Accordingly, review rating algorithms must be rapidly adaptable and provide accurate and relevant ratings as the number of products and reviews on them increases, and so on. Finally, user feedback must be sought and actively shared and integrated into the development and improvement processes, by involving users in the design and decision-making, so that their needs and concerns can be better addressed, as this leads to a more user-centered review rating system, and also in this topic can be used Regular feedback loops, usability studies and surveys to continuously assess user satisfaction and identify areas for improvement and enhancement.

**6.0 Conclusion**

The Optimal Review Ranking project endeavors to transform the way users access and evaluate user-generated reviews. By utilizing personalized recommendations, the system empowers users to make more informed decisions, saving time and ensuring a seamless and satisfactory online shopping experience. The project focused on developing an Arabic reviews rating system that incorporates individual user preferences, contextual aspects, and machine learning algorithms. The work methodology involved sentiment analysis, keyword extraction, user profiling, and personalized ranking of reviews using the BM25 model in Elastic Search engine. Several NLP tools and APIs, such as AraBERT, Farasa, NLTK's, and TfidfVectorizer, were utilized to preprocess and analyze the data. Although the project faced limitations, such as limited access to personal user information and difficulties in handling colloquial Arabic language, it provides a foundation for future works in enhancing the accuracy and scope of the review ranking system. By continuously improving and expanding the system, it is possible to revolutionize the online shopping experience by providing users with tailored and reliable review recommendations.

**7.0 References**

[1] “Meet Your next Favorite Book.” Goodreads, www.goodreads.com/. Accessed 7 July 2023.

[2] Elnagara. “Elnagara/Brad-Arabic-Dataset: Brad: Books Reviews in Arabic Dataset.” GitHub, github.com/elnagara/BRAD-Arabic-Dataset. Accessed 7 July 2023.

[3] “Sentiment Analysis.” Wikipedia, 4 June 2023, en.wikipedia.org/wiki/Sentiment\_analysis. Accessed 7 July 2023.

[4] “Keyword Extraction: A Guide to Finding Keywords in Text.” MonkeyLearn, monkeylearn.com/keyword-extraction/. Accessed 7 July 2023.

[5] Chiche, Alebachew, and Betselot Yitagesu. “Part of Speech Tagging: A Systematic Review of Deep Learning and Machine Learning Approaches - Journal of Big Data.” SpringerOpen, 24 Jan. 2022, journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00561-y. Accessed 7 July 2023.

[6] Madan, Rohit. “TF-IDF/Term Frequency Technique: Easiest Explanation for Text Classification in NLP with Python.” Medium, 27 Nov. 2019, medium.com/analytics-vidhya/tf-idf-term-frequency-technique-easiest-explanation-for-text-classification-in-nlp-with-code-8ca3912e58c3. Accessed 7 July 2023.

[7] “What Is Elasticsearch?” Elastic, www.elastic.co/what-is/elasticsearch?fbclid=IwAR1\_fNF0CIuYMsTh29qhU4el23wcNFdsKzkUtUw9wu3fs6oXT-s2CNzbmIs. Accessed 7 July 2023.

[8] Connelly, Shane. “Practical BM25 - Part 2: The BM25 Algorithm and Its Variables.” Elastic Blog, 1 Apr. 2019, www.elastic.co/blog/practical-bm25-part-2-the-bm25-algorithm-and-its-variables?fbclid=IwAR1cHsvPHD4SB6d3wk-eRfsxxm8opzmUjRshtCOfUJFXGjyJTl41BKICFgU. Accessed 7 July 2023.

**8.0 Appendix**

[Google Colab CODE link](https://colab.research.google.com/drive/1RJKUGVBpMDYpV64jTN6coTZhNYE29h20?usp=sharing)