Peer Review File

Article Information: https://dx.doi.org/10.21037/jmai-23-58

Reviewer A

General Comments

This manuscript provides a comprehensive study of machine learning techniques, such as tokenization, decision trees, support vector machines (SVM), random forest, stochastic gradient boosting (SGB), and gated recurrent units (GRU). It explains how these techniques can be applied to process and analyze user reviews in the Google App Store. Additionally, the paper covers system design aspects, detailing the process of creating logical and physical layouts. It discusses input design, with an emphasis on minimizing user input and errors while enhancing security. Despite its robust content, the manuscript does not explicitly relate these techniques to applications in the medical field, making it less relevant for a medical audience. From the JMAI website: "The articles published in Journal of Medical Artificial Intelligence (JMAI) need to discuss the incorporation of artificial intelligence in the medical field. And it includes but not limited to, AI in bio- and clinical medicine, machine learning based decision support, robotic surgery, data analytics and mining, laboratory information systems and AI in medical education." It would be challenging to justify this manuscript based on any of the criteria.

 # Response to the Specific Comments

The introduction of the manuscript, while comprehensive, could use a little streamlining and much more concise writing.

- The Introduction (along with the rest of the paper) frequently repeats certain phrases and ideas, which increases its length unnecessarily. For example, there are repetitive mentions of the increasing importance of artificial intelligence and machine learning. Try to condense similar concepts and avoid unnecessary repetition to maintain the reader's attention and clarify your message.

Response: As per the reviewer's suggestions, the needful is done.

- The paragraph about app development seems somewhat out of place in the broader context of AI and machine learning. If this information is essential to the paper, it should be tied more closely to the main topic or be better integrated into the flow of the argument.

Response: As per the reviewer's suggestions, the needful is done and a paragraph about app development is removed.

- While you have provided a lot of background information, the problem statement could be better defined. Consider starting with the problem (discrepancy between star ratings and reviews) and then delving into its implications and why it matters. This would make it clearer to the reader why the study was undertaken.

Response: As per the reviewer's suggestions, a paragraph is added on page no. 4, paragraph no. 68-81. “Previous research has focused on using reviews and ratings to classify them into positive, negative, and neutral categories, but no work has been done to predict biased and unbiased ratings using deep learning models.

The objective of this manuscript is to address these issues in the following ways :

Develop an intelligent system to assist users based on their needs and interests. Investigate the prediction of an automated review rating system that can accurately predict unbiased ratings.

Focus on predicting biased and unbiased ratings using machine learning models, as previous research has primarily focused on sentiment classification.

Leverage the capabilities of machine learning models to extract complex patterns and representations from review data.

Enhance the transparency and fairness of online review platforms by providing users with objective and reliable information through unbiased ratings.

This article uses five different types of existing techniques to show a comprehensive comparison between the models to deliver the results as per the user’s needs and interests”.

- The transition from the background to the specific details of the Google Play Store and its applications seems abrupt. You might need a smoother transition or more clarification on why you're focusing on this specific topic.

Response: As advised, please find the paragraph no.44-53 on page no. 3. “Mobile phones have become an integral part of people's lives and the mobile application industry is growing rapidly, resulting in increased revenue for the global sector. However, with the growing number of mobile application designers, it is important for them to sustain their income in the market and move in the right direction. Google Play Store [2] [5] is one of the most popular app stores, with over 0.675 million Android applications available. With so many options available, online application surveys have created a significant impact compared to paid applications. It is difficult for potential clients/users to sort out all the reviews and ratings, and for developers to improve application performance based only on evaluations. Along with this, various challenges are faced on the Google play store due to an increase in the number of applications with no specific features and usage that misleads the users with app quality, security, privacy, app monetization, policy compliance, competition, and user engagement”.

- The way the paper is structured (section by section) is mentioned in the introduction. Usually, this information is placed at the end of the introduction, providing the reader with a roadmap for the rest of the paper.

Response: Roadmap is mentioned in Page no. 4, paragraph no. 96-98, “The paper is structured into different sections: Section 2 focuses on related work, Section 3 discusses the material and methods used, Section 4 presents the experiment, Section 5 goes into detail on the results and discussion, and Section 6 concludes the work”.

- The part of the introduction describing the survey conducted by Aralikatte et al. might be better placed in the literature review or "Previous Work" section of your paper, as it seems to break the flow of the problem statement.

Response: As reviewer's suggestions, Aralikatte et al. statement is moved to review work section Page no.5, paragraph no. 110.

- Your usage of specific machine learning terminologies, such as "supervised learning" and "Support Vector Machine (SVM)", could be explained briefly for the benefit of readers who might not be familiar with them.

Response: As per the reviewers, “Supervised learning symbolically relates where a model is trained using labeled examples (input data with corresponding target outputs) in order to predict unknown data. Although SVMs were first introduced in the 1960s, they gained popularity only in the 1990s due to their ability to produce remarkable results. Sentiment analysis, classification, and prediction are different tasks performed with a robust and reliable approach to extracting valuable insights” is added in page no.9 paragraph no.240.

- The authors provided a comprehensive overview of tokenization and its significance in user review analysis, detailing Word Embeddings tokenization with Word2Vec and various classification techniques (Decision Tree, SVM, RF, SGB, and GRU).

- The algorithmic representation of tokenization, word embedding, and accessing word embeddings is commendable, highlighting the crucial role of tokenization in feature extraction, sentiment analysis, and text classification. However, the explanations of classification techniques could benefit from additional context explaining when and why each would be advantageous over others in specific use cases.

Response: As advised the needful has been added and highlighted.

- More information about the advantages of GRUs compared to traditional RNNs and LSTM, especially concerning computational resource efficiency with large datasets, would enhance the understanding.

Response : As stated, the above advised is mentioned in Page no. 11, “The Gated Recurrent Unit (GRU) is an advanced version of recurrent neural networks (RNNs) that was introduced by [12]. It is commonly utilized to tackle classification problems and it solves the "vanishing gradient" issue that is typically associated with standard RNNs [27]. The GRU employs two gates - the update gate and the reset gate - to handle the vanishing gradient issue that standard RNNs encounter [12, 27]. The gates can learn to keep relevant information from the past while discarding irrelevant information, and pass the relevant information to future events to improve prediction accuracy. The update gate controls the amount of information from previous time steps that should be passed on to future steps [27][30]. In contrast, the reset gate decides how much of the previous information (history) should be disregarded by the network. Using a GRU can enhance the RNN's memory capacity and make it easier to train the model. The GRU architecture utilizes an update gate to regulate the flow of information from the past to the future, and a reset gate to decide how much previous information should be discarded. This design enhances the network's capacity to retain significant data and make better predictions. Adopting GRUs instead of standard RNNs can improve the network's memory capacity and ease the training process. This is especially beneficial for applications such as automatically identifying and categorizing requirements from app reviews, where accuracy and efficiency are critical for achieving success. The GRU architecture is very

adaptable and simple as compared to Long short-term memory (LSTM) which helps in faster computation speed and memory efficiency [27]. There are various advantages of GRU as compared to RNN and LSTM based on specific databases and tasks".

- The clear definitions of the logical and physical design aspects of a system are beneficial; however, an explanation of their interaction would be more insightful.

Response: As per suggestions, statements are added in Section 5.1 and Section 5.2 in respective pages no. 12-14 Explaining the logical and physical design aspects of a system.

- The authors effectively emphasized user-friendly and secure input design; adding real-world examples or case studies would clarify these considerations.

Response: As advised by the reviewers, the Google App store has many real-world using data applications as mentioned in the manuscript.

- The authors' writing style can become convoluted in technical sections; refining this to be more clear and more concise would enhance readability and comprehension.

Response: As advised the needful has been added and highlighted.

- The manuscript primarily focuses on tokenization and text analysis in user reviews, which is a specific area of machine learning and NLP, potentially making it less relevant to a medical audience.

Response: As advised the needful has been added and highlighted. The Google App Store there are various health care applications related to the topics. So, by using these techniques and comparison the patients can conclude to use a specific app for their usage and management.

- The detailed exploration of techniques like Word2Vec, SVM, RF, SGB, and GRU is more aligned with computer science or data science than medical research or practice.

- The lack of a clear connection or applicability to medical topics, such as patient care, medical research, or health policy, might make the manuscript less interesting for the medical audience.

Response: As advised by the reviewers, in the Google App Store there are various health care applications related to the topics. So, by using these techniques and comparison the patients can conclude to use a specific app for their usage and management.

- The manuscript doesn't delve into healthcare-specific use cases, such as medical text analysis, EHR data processing, or medical literature reviews, limiting its usefulness for healthcare professionals.

Response: As advised by the reviewers, in the Google App store there are various health care applications related to the topics. So, by using these techniques and comparison the patients can conclude to use a specific app for their usage and management.

- The heavy focus on technical details and algorithms might make the manuscript too complex for medical professionals who are not deeply familiar with machine learning or NLP.

Response: As advised by the reviewers, in the Google App store there are various health care applications related to the topics. So, by using these techniques and comparison the patients can conclude to use a specific app for their usage and management.

Reviewer B

The citations of Figures 2 and 8 are missing the main text.

Figure 2: The app name is Chinese. Please replace it with its English name.

3893	4 in a Row	GAME	3.8	4257	Varies with device	500,000+
1488	乐屋网: Buying a house, selling a house, renting ...	HOUSE_AND_HOME	3.7	2248	15M	100,000+

Response:

The figure 2 is modified and added.

Please consider changing Figure 2 to Tables and cite them in the main text.

(You have mentioned "table" in the following two paragraphs. Are they part of the legend of Figure 2?)

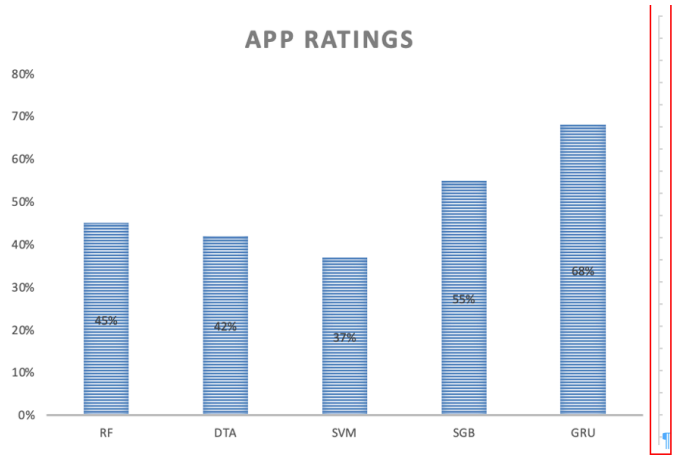
The table above shows the classification results based on various features of apps, including rating, reviews, size, installs, type, price, content rating, genres, and last update. The study focused on the art and design area of apps and categorized over ten apps based on their features to help users choose the best app for their needs. Similarly, different types of apps can be classified based on the given parameters.¶

The table above displays the projected final outcomes of the Gated Recurrent Unit (GRU) classifier, with an overall accuracy of 0.89. Out of all the classifiers, GRU classified almost 5616 apps, while the other classifiers only classified 3744 apps. When considering various predictions, the GRU classifier provides the most precise results for both users and developers.¶

Response:

As advised, the proper correction is done.

Please remove the extra bar in Figure 3.



Response:

Done.

Please check Figure 5. In the legend, SVM=21.5%. Should below 21% be 22%?

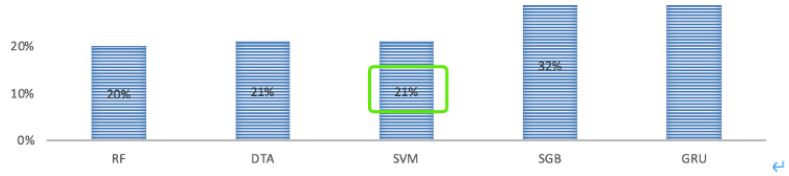


Figure 5. Demonstrates the predicted outcomes given by all the algorithms mentioned above based on the "rating score" feature.

The X-axis shows different algorithms, while the Y-axis indicates the predicted rating score. The output outcomes for different algorithms such as RF=20%, DT=21%, SVM=21.5%, SGB = 32.1%

Response:

Done.

Figures 3-8: Please check whether the below "DTA" should be "DT".

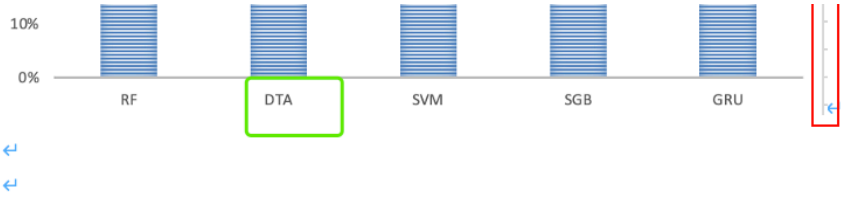


Figure 3. displays a bar chart that presents the anticipated outcomes of various algorithms based on the "reviews" feature.

The X-axis shows different algorithms, while the Y-axis shows the accuracy of the output based on reviews. The output results for different algorithms such as RF=45%, DT=42%, SVM=37%, SGB

Response:

Done.