Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 09 Issue 01
2025

Research Article
https://doi.org/10.56979/901/2025

Enhancing Mobile App Quality: A Data-Driven Approach to User Feedback

Muhammad Bilal Azhar[1*], Naeem Aslam[1], Ahmad Naeem[1], Muhammad Fuzail[1], and Ahmad Murad[1]

[1]Department of Computer Science, NFC-IET, Multan, Pakistan.
[*]Corresponding Author: Muhammad Bilal Azhar. Email: Bilal.azhar9639@gmail.com

Received: April 01, 2025 **Accepted:** May 20, 2025

Abstract: In the very competitive mobile application arena, customer reviews are essential for success. Timely responses to these reviews can substantially enhance an app's rating and prominence. The proliferation of user-generated content has rendered the extraction of useful insights increasingly challenging for developers. To enhance user experience and enable prompt changes, it is essential to swiftly and properly identify the primary issues encountered by users. This research presents a dual-phase hybrid framework. Phase 1 computes the mean sentiment rating for user reviews of Zoom Cloud Meeting, Microsoft Teams, and Google Meet. According to these averages, Phase 2 offers pragmatic recommendations to developers. A text data augmentation strategy utilizing the advanced language comprehension capabilities of big language models such as ChatGPT was implemented. The framework uses a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) to calculate ratings from review datasets and BERT-base for labeling and analyzing feelings. The suggested hybrid model exhibited robust performance on datasets containing a substantial number of reviews, suggesting a tendency for positive ratings to surpass negative ones. The study utilized a benchmark dataset comprising 44,767 app reviews to compare its findings with actual ratings, yielding significant insights for enhancing app development. According to average evaluations, Microsoft Teams (4.52) and Zoom Cloud Meeting (4.16) outperformed Google Meet (3.12) and all other applications (3.40), indicating that users of Teams and Zoom will derive the most advantages from the latest versions. We anticipate that the study's recommendations will aid app makers in enhancing their offerings.

Keywords: Google Meet; User Feedback; Sentiment Analysis; Zoom Cloud Meeting; Microsoft Teams

1. Introduction

The importance of mobile applications has increased exponentially in the age of digital technology due to the widespread usage of mobile devices, especially smartphones and tablets [1]. These apps are becoming a part of our everyday lives and are essential tools for business, productivity, communication, and pleasure [2]. The various categories found in app stores facilitate user browsing and selection of the ever-changing landscape of mobile applications, which range from premium services to free utility [3]. When it comes to the number of apps available, the Google Play Store is the biggest app store. It is impossible to overestimate the influence of user reviews and ratings while choosing apps from different categories [4]. Users' decisions are shaped by this interaction between feedback and measured satisfaction, which has a big impact on them. As an example, comparatively few people download apps with a rating of three stars, and many users do not install apps with a rating of two stars [5]. Since COVID-19, many educational institutions have made online learning mandatory. Unfortunately, the lack of infrastructure and time to create and execute online education made the shift to online learning challenging [6, 7]. Social networking sites, educational platforms, and other apps are in high demand in developing nations. The need for online learning and teaching platforms is rising in these countries, and it's critical to comprehend how students use these resources [8]. Teachers must give students lectures, materials, and assessment tools as part of the learning process. Students learn online by completing assignments and transferring knowledge through technology [9]. Students who communicate online gain from this online learning technique [10]. Online learning has several advantages, such as convenience and flexibility. To support learning, many platforms can disseminate resources, and assets, or gather homework [11].

The rapid emergence and dissemination of COVID-19 has brought attention to the critical need for remedies regarding online instruction and meeting conduct [12]. Mobile app repositories offer useful access to often updated, substantial datasets about software [13]. Search engines, repositories, app shops, and other diverse, multipurpose platforms are among these repositories [3]. Publication app evaluations, in which users share a variety of aspects including personal thoughts or experiences, bug reports, questions, or requests, is one of the most well-liked contributions on these platforms [14, 15]. Using NLP [16], it is possible to extract meaningful insights from this fastexpanding data and filter out noisy words. Numerous NLP researchers have advanced in recent years to determine text qualities such as subjectivity, polarity, or emotion recognition, as well as document or context classification [17]. This need for researchers to categorize a setting as good, negative, or neutral has been satisfied by sentiment analysis (SA).

Opinion extraction, or SA, is defined as gathering data from publicly available content to produce people's opinions, attitudes, and expressions about news, customer goods, themes, or forum discussions [18]. There are various kinds of data augmentation techniques in NLP [19–21]. More recent techniques, such as word vector interpolation in the latent space [22] and back translation [23], use language models to produce trustworthy samples for more successful data augmentation. Inspired by the recent successes of large language models, particularly the development of ChatGPT, which demonstrated enhanced language comprehension abilities [24, 25], we employ a ChatGPT-based text data augmentation technique in this study. People may meet physically thanks to the commonly used video conference platforms Zoom Cloud Meeting [26], Google Meet [27], and Microsoft Teams [28] apps. Global pandemic conditions were brought on by COVID-19 in 2019 [29]. Everyone started searching for a platform that would assist with virtual meetings, virtual classes, and several other services [30]. Among the most widely used online communication and education systems are Zoom Cloud Meeting, Microsoft Teams, and Google Meet. The rating is becoming more significant for developers to update the versions of the Google Meet Apps and Zoom Cloud Meeting [31]. Many students provide negative evaluations because they dislike going to class, and the majority of users do not rate things correctly. Thus, the rating system might not accurately reflect the true impact of the revised version [32, 33]. We present a two-phase hybrid framework in this study to get around this restriction. The hybrid approach calculates the mean sentiment rating of user reviews for Zoom Cloud Meeting, Microsoft Teams, and Google Meet in the first phase. In the second step, it provides developers with recommendations based on the average ratings that were established in the previous phase. We also performed sentiment studies on a few other popular apps, such as Hangouts and Skype. The suggested hybrid model uses a BERT base for sentiment analysis and a combination of RNN-LSTM to determine the average scores on the review dataset.

1.1. Problem Statement

Since the pandemic has led to severe limitations on travel and in-person meetings, it was nearly inevitable that online teaching methods would be created during the pandemic through digitally mediated communication tools. The utilization of online learning environments is a current concern, and in-depth research in this area is lacking. The ability to use the newest learning technology is required in the new world. An app's ability to work and have an attractive user experience is crucial to its success. The application's developers must regularly release new program versions to provide the features required in this online help scenario. Thus, to improve the overall user experience and spur product development, designing a solution that can efficiently analyze user sentiment and behavior during video conferencing is imperative.

1.2. Contribution

The following are the primary contributions of the suggested hybrid model:

- We provide a unique hybrid method for identifying app deficiencies that computes average ratings on review datasets for Zoom Cloud Meeting, Google Meet, and Microsoft Teams apps. When analyzing user reviews, the suggested method can accurately identify positive and negative sentiments and produce review scores.

- We perform data augmentation to expand the size of the review datasets that were not readily available in significant length.
- We leverage the capabilities of RNN-LSTM to prevail over the weaknesses of each model, and we use BERT Base as a sentiment analyzer.
- We thoroughly analyzed the Hybrid Model's performance using the augmented reviews dataset for Microsoft Teams, Zoom Cloud Meeting, and Google Meet.
- This article represents the first attempt to offer businesses (Zoom Video Communications, Microsoft, and Google) suggestions for improving their apps' sections based on user feedback knowledge.

1.3. Organization

The following outlines how this research is arranged: A Related Work is presented in first 2. Detailed work and discussion of the hybrid technique are explained next 3. The next results of a hybrid model are presented 4. This study is concluded 5 with discussions of future work.

2. Related Work

Several studies in machine learning and deep learning are available for sentiment analysis of app evaluations and data from websites or blogs. Nevertheless, sentiment analysis works in the context of app development are now scarce. There aren't many publications in the literature on sentiment analysis based on app reviews.

2.1. User Reviews and Importance of Feedback

Since end-user reviews frequently provide important insights into the real-world performance and user experience of applications, sentiment analysis of these reviews has grown in importance within the app development industry. To gather data on app evolution, including new features, non-functional requirements, and problems, app researchers have recently put forth several research methodologies incorporating the analysis of end-user feedback [34]. The authors [35] draw attention to the depth of study conducted in this field and stress the value of using app reviews to gain an understanding of software progress. Moreover, the examination of machine learning (ML)-based cloud-based bug-tracking software defects research by [36, 37] highlights the practical uses of these approaches in lowering the time and expense needed for app development. App ratings and reviews can be important factors in a user's decision-making process when choosing an app within a particular app category [4]. App ratings offer a quantifiable assessment of user happiness, whereas reviews offer insightful commentary from people who have used the app directly [34]. In this study, we use review datasets from Google Meet, Microsoft Teams, and Zoom Cloud Meeting to identify sentiments.

2.2. Apps Development based on Feedback Aspects

Through app reviews, users can report both good and negative elements, problems, and bugs as well as share their experiences, opinions, and suggestions about the app. Developers may optimize their app and increase the user experience by looking through app reviews, identifying recurrent themes and patterns, and using this feedback to inform data-driven decisions for updates, bug fixes, feature enhancements, and general app optimization [38]. On the other hand, users assign the app a star rating between 1 and 5, usually depending on how they feel about it all. One star indicates the lowest degree of satisfaction, whereas five stars indicate the maximum level of satisfaction [39]. When a user feels that an app is of great quality and they have had an excellent experience with it, they could rate it with five stars. Three stars typically indicate that consumers' experiences with the app are mediocre or neutral. Aspects of utilizing any product or design are known as user experience factors [40]. Finding the user experience factors is essential for app developers in the digital world because it allows them to focus on creating usable apps that satisfy user needs while also offering them a competitive edge by providing a positive user experience [41]. To enable developers to enhance the user experience appropriately, it is essential to comprehend the key elements that contribute to both positive and negative user experiences. For instance, unfavorable reviews are linked to factors like compatibility, resource usage, requests for improvements, bugs, crashes, and [41]. Based on the average rating computation that was applied to user reviews in review datasets, we offer recommendations to the enterprises.

2.3. Need of BERT Model and Data Augmentation

In 2018, Devlin et al. presented BERT [42], a bidirectional transformer training-based system for language modeling. It uses the attention mechanism to extract the true context of the text and is composed of two main parts, which are used for pre-training and fine-tuning: an encoder for the input text and a decoder for the task prediction or output [43]. There are two primary ways to train the BERT model: Discreet language modeling involves randomly hiding 15% of the words in a sentence. The model then uses the context of the other, non-masked words in the sentence to try to anticipate the concealed word. Using pairs of sentences as input, the model predicts whether the second sentence in a pair is the sentence that comes after in the original text [44]. Text classification model training is frequently enhanced by data augmentation, which is the synthetic creation of fresh text through modifications. The data augmentation techniques now in use in NLP operate at many granularity levels, including characters, words, phrases, and documents [45]. The process of randomly adding, removing, swapping, or altering characters in a text document is known as data augmentation at the character level [46]. This technique increases the NLP model's resilience to noise. Word-level data augmentation is also effective. Two words in the text are randomly swapped in random swap augmentation, and some words are randomly deleted in random deletion augmentation [47]. We applied data augmentation techniques to the larger review dataset and ran sentiment analysis on it using the BERT BASE pre-trained model.

2.4. Sequence Models Implications

An effective model for sentiment analysis is the recurrent neural network (RNN) [48]. Recent works employing RNN for sentiment analysis include [49–51]. The vanishing gradient problem is caused by its improved temporal complexity and ability to learn more recent words than earlier words [52]. Recurrent neural networks that can learn order dependence in sequence prediction tasks are called long short-term memory networks (LSTM networks) [53] [54]. Because LSTM is ideal for classifying, processing, and predicting time series with time lags of uncertain duration, it is widely employed in NLP [55]. It provides the greatest degree of control and produces superior outcomes. Reviews and analyses have benefited greatly from the use of LSTM networks [56]. However LSTMs are prone to overfitting, and using the dropout strategy to prevent this problem is challenging. We suggest a hybrid model that combines BERT's promising performance with RNN and LSTM's drawbacks. It will use both RNN and LSTM in straightforward configurations. The LSTM model will receive its recent word characteristics from a basic RNN model. These traits will be stored using LSTM, which will also determine the long- and short-term relationships between them. The SoftMax activation function will be used to identify sentiment using the output of an additional basic RNN that receives data from the LSTM. On a scale of 1 to 5, the hybrid model that has been suggested will give the average ratings for the review datasets.

3. Materials and Methods

Three degrees of analysis are possible for SA: aspect, sentence, and document levels [57]. Sentence-level data are the focus of the current investigation. Sentiment analysis in app development has always been a difficult undertaking. In this regard, the current study suggests a hybrid model that uses three widely used video conferencing review datasets. First, data augmentation is selected as a pre-processing technique to help in training to fine-tune the BERT BASE for effective results. To suggest more robust and user-satisfying app development, we provide a novel hybrid model that carries out sentiment analysis on the three most popular review datasets [58]. After calculating the average rating of user reviews, this model suggests features that can have an impact on the market for video conferencing apps.

3.1. App Selection Process.

We used an analysis of Android apps from the Google Play app store to carry out our research. These apps were gathered by employing a specific search query to do a systematic scrape of the Google Play website. The purpose of the search query is to identify the Android app's category. To include Microsoft Teams, Google Meet, and other relevant apps in the review dataset, the search query also looks for them. Play-scrapper6 and google-play-scrapper7, two Python-based Google Play app store scraper modules, were employed that can be found at Play Scrapper1 and Google Play Scrapper2 respectively. The most widely used video conferencing apps are provided with a few sample reviews in Table 1.

Table 1. Most Popular Video Conferencing Sample Review Dataset

App	Review	Score

Google Meet	Hi please as soon as possible put raise hand option...	2
	THIS is a very good app but there is one problem...	4
	Just the most amazing app I have seen in my life....	5
Zoom	Awful. I used to be able to use Zoom...	1
	Loved using it. Can survive with...	5
	All experience is good but please make...	3
Skype	Sometimes it works, sometimes it doesn’t...	1
	I love this app since it was one of...	5
	It looks like Skype doesn’t support HEVC video...	3

3.2. Review Datasets

To create datasets with a range of sizes and characteristics, we gather evaluations from different sources. After undergoing preprocessing, the data gathered from various sources is included in our hybrid model. According to Table 2, the reviews gathered for the Zoom Cloud Meeting App is 4018, followed by Google Meet 7822, Microsoft Teams 2685, Hangouts 1739, and Skype 10,000 and can be found at Meeting App Reviews3.

Table 2. Most Popular Video Conferencing Review Dataset

App	No of Sources	Collected Reviews
Google Meet	31	7822
Microsoft Teams	24	2685
Zoom Cloud Meeting	53	4018
Hangout	15	1739
Others	55	4543
Skype	21	10,000
Grand Total	199	30,807

3.3. Text Augmentation

This method creates fresh data with different data orientations. Researchers may now rest easy knowing that data augmentation reduces overfitting and maximizes data generation from sparse data [59, 60]. Table 3 contains the augmented review datasets. Despite being simple to use and not requiring any outside resources, these strategies significantly increase performance [61]. However, the text augmentation methods that we employ to improve the dataset are covered in the list items. To increase the accuracy of the algorithms, we have compiled the preprocessing methods, looked into data augmentation methods, and conducted experiments to see if it is possible to automatically generate training data applied for small review datasets. Because of the review dataset's limitations, this is a vital phase that enriches the datasets and reduces the time and expense of creating a pre-labeled dataset for a particular area.

- InsertCharAugmentation [62]: By introducing noise into the data, this technique enhances the model's capacity for generalization by randomly inserting characters at various points across the text.
- WordNetSynonymAug [62]: It substitutes terms in the WordNet thesaurus with their synonym.
- SwapWordAug [47]: It swaps words in the text at random. This technique is a variation of Wei et al.'s Easy Data Augmentation (EDA) technique.
- ContextualWordAugUsingBert(Substitute): In this approach, which replaces randomly chosen words in the text with < mask > tokens based on context, BERT is used [63, 64]. BERT is then allowed to anticipate the token at that point.
- Spelling Augmentation [65]: It purposefully misspells some words to produce fresh content. The Oxford Dictionary provides a list of English terms that are frequently misspelled. For example, the term "because" is sometimes misspelled as "becouse."

Table 3. Augmented Review Datasets

App	Enhanced Reviews	Total Reviews
Google Meet	3850	11,672
Microsoft Teams	3500	6185
Zoom Cloud Meeting	3550	7568
Hangout	1530	3269
Others	1530	6073
Skype	0	10,000
Grand Total	13960	44,767

3.4. Preprocessing

To obtain well-structured data, we employ a few preparation techniques [66] as shown in Figure 1. Text pre-processing chores are essential for implementing the rule-based method with supported lexicons. Data cleaning is the process of processing incomplete data; for example, certain rows are eliminated due to missing information [67]. The same data is presented in the dataset in a variety of ways. We have resolved this issue such that the data cannot contradict one another.

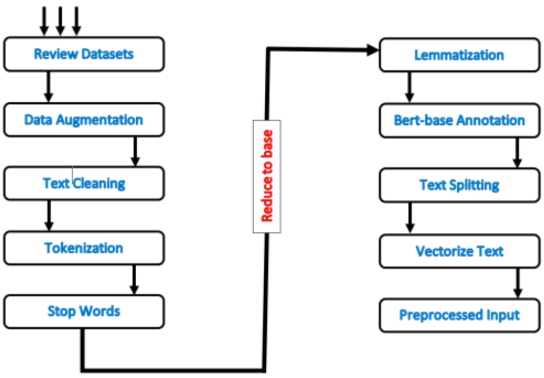

Figure 1. Preprocessing Steps in Hybrid Model

Data implementation involves the use of data transformation. Tokenization involves dividing the data into pieces, each of which is referred to as a token. This token is easily definable by a machine. Tokenization is one of the most crucial preprocessing steps, which is why we utilized it [68]. Two of the models in our proposed study make use of preprocessing techniques, and one of the models BERT BASE is pre-trained. Given that a substantial portion of the reviews are delivered from difficult-to-write mobile devices, the data must be cleaned before being analyzed. Our preprocessing steps include tokenization, stemming, parts of speech (pos) tagging, and the removal of stop words and superfluous characters. Typically, stop words (such as the, as, was, with, in, etc.) only have syntactic meanings [69]; that is, they lack sentiment. Word forms that differ syntactically but are semantically equal can be grouped using stemming and lemmatization. As an illustration, the terms "saw" and "sees" are combined to form the term "see" [70]. We use Keras and TensorFlow neural network packages concurrently Keras4 to preprocess our data. Tokenizer, text to sequences Tokernizer5, and pad sequences P adding6 are Keras tools that we employ.

3.5. Tokenization

By evaluating each token's significance, the tokenization stage, as shown in Table 4, aids in analyzing the review process. Consequently, 769 user reviews that contained only numbers or non-English letters were disqualified.

Table 4. Sample Tokenized Reviews

Tokenized Reviews
{'Mostly':1,'spyware':2, 'data':3, 'gathering':4, 'app':5,'designed':6,..} {'Good':1,'use':2, 'daily':3, 'meetings':4, 'work':5,'great':6,..} {'like':1,'but':2, 'network':3, 'problem':4,..} {'very':1,'useful':2, 'video':3, 'audio':4, 'clearly':5,'love':6,..}

3.6. Review Dataset Factors

Thematic analysis [71] was utilized to curate these seven aspect labels through an all-encompassing investigation of the sampled reviews that were utilized to train the automatic classification models. Table 5 lists the sentiment and aspect labels that were utilized. To annotate the tokenized review datasets, we employed the BERT-based language model.

Table 5. Different Aspects of Review Datasets

Specifications	Values
Technical	Stating hardware or software issues
Positive	Strong sense of positivity, tone of satisfaction
Negative	Strong sense of negativity, expressing dislike
Neutral	Expresses mixed feelings
Utility	How easily the user can navigate
Feature Required	Addition of a new feature
Learning	Learning new ideas or concepts

3.7. BERT-base Sentiment Annotation

In this work, we investigate dataset annotation using the multi-head self-attention mechanism of the BERT. The BERT attention mechanism's Query (Q), Key (K), and Value (V) components begin a linear transformation to" dynamically" generate weights for different connections. These weights are then put into the scaling dot product [72], which is provided in below Equation.

$$Att[Q,K,V] = SoftMax[\frac{QK^T}{\sqrt{d_k}}]V$$

N self-attention is computed in parallel by the multi-head attention by using Equation.

$$H_j = Att[QW_j^Q, KW_j^K, VW_j^V]$$

Below Equation gives the total attention received by all independent participles in the layers and heads, which is the attention for each token.

$$AttW_{token} = \sum layers \sum headsAtt_j$$

Below Equation illustrates how we add up the attention of a word's tokens to determine the attention of each word [72].

$$AttW_{token} = \sum_{token \in word} AttW_{token}$$

Usually, the model is trained and tested on an annotation sample before being applied to the data set. Positive, negative, or neutral user reviews can be manually annotated by researchers or delegated to others. Furthermore, rather than employing manual annotation, several researchers tended to construct corpora annotated according to user rates [73, 74]. For instance, each user can rate a product on Google Play from 1 to 5 stars to indicate how satisfied they are [75]. Reviews that have been tokenized are sent into the BERT model, which uses the BERT-base model's encoding method to encode the reviews and produce sentiment scores. Next, we contrast BERT scores with the actual ratings from the top four apps: Microsoft Teams, Zoom Cloud Meeting, Skype, and Google Meet. We decided to classify the ratings with one and two stars as negative sentiment and the remaining ratings as favorable. We used BERT-base on the remaining ratings, ignoring the first or second star rating. We compare the two settings to see if the sentiments are comparable, and if so, we incorporate the data into our review dataset. Using Algorithm 1, we start with all reviews in Rv and tokenize and pad the sequence to get Ri . Next, examine the dataset RD that was produced by using the data cleansing qualities that were described in the section on preprocessing. Reviews that had similar ratings positive or negative in the real ratings, X, from the review dataset and the BERT-base generated ratings, X0, were the only ones that were included.

3.8. Recurrent Neural Network

An RNN is a feed-forward neural network whose output depends on the previous state for sequence modeling and data. To capture the association between the current and past time steps, it feeds back fresh state information to the preceding layer while iterating over the elements in the sequence [51, 52]..

Algorithm 1 Review Dataset Preparation Procedure

Require: Annotated Review Dataset
Ensure: R_v = User Reviews

$R_i \Leftarrow$ words from R_v
$X \Leftarrow$ Actual ratings
$R_D \Leftarrow \{\}$
while $\bar{X}_{ik} \Leftarrow Score_k$ **do**
    $Score_k \Leftarrow$ Sentiments of given review
    if $\bar{X}_{ik}, \bar{X}_k \geq 3$ **then**
        $R_D \Leftarrow R_D \cup \{Score_k, \bar{X}_k\}$
    end if
    Match for each $Score_k$
end while
Return R_D

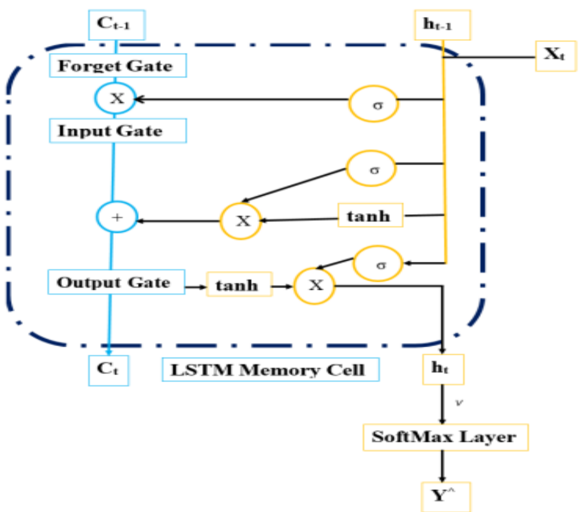

Figure 2. LSTM Architecture

Our experiment makes use of a straightforward 32-layer RNN. At a given timestamp, the previous input is pulled from the previous hidden layer and fed back into the input of the current hidden layer. An inserted Spatial Dropout with 0.25 is placed on the RNN layer before it. After the RNN layer, we add a BatchNormalization, Dropout with 0.25, and GlobalMaxPool1D in that order. Finally, a three-connected layer with a sigmoid activation function is added to the Dense layer.

3.9. Long short-term Memory

The model constructed with the LSTM network [76] will function better in sentiment analysis training and review average rating computation. LSTM is gradually applied to different large language models to address gradient disappearance caused by recurrent neural networks. In the recurrent neural network (RNN), which is based on the traditional cyclic neural network, the LSTM unit assumes the role of the neurons. Figure 2 also displays the fundamental components of an LSTM architecture [76].

3.9.1. Forget Gate

A value of 0 denotes complete forgetfulness of the information in this bit, whereas a value of 1 denotes complete retention. The computation formula [76] is shown in equation 5.

$$G_{ti} = \sigma[\omega_{gi} * (h_{t-1}, x_{ti}) + b_{Gi}] \tag{5}$$

3.9.2. Input Gate

The input-gate augments the data needed by the new cell state to the maximum extent possible. The input gate's output is the current cell state multiplied by the sigmoid function, which has a value range of 0 − 1. The formula for calculation in Equations 6 and 7 is as follows.

$$I_n = \sigma[\omega_{in} * (h_{t-1}, x_{ti}) + b_{Ii}] \tag{6}$$

$$\bar{C}_t = \sigma[\tanh_{ci} *(h_{t-1}, x_{ti}) + b_{Ci}] \tag{7}$$

Data from both the old and new states can then be combined to form the final new cell state. in accordance with Equation 8.

$$C_{ti} = [G_{ti} * C_{t-1}] + [I_{ti} * \bar{C}_t] \tag{8}$$

3.9.3. Output Gate

The output gate generates a sigmoid function with a value between 0 and 1. The computation of the activation function is demonstrated by Equations 9 and 10 [76].

$$OP_{ti} = \sigma[\omega_{OT} * (h_{t-1}, x_{ti}) + b_{OT}]h_t = OP_t * \tanh C_{ti} \tag{9}$$

$$hp_t = OP_{ti} * \tanh C_{ti} \tag{10}$$

3.10. Hybrid Model Training

To train our hybrid model, we separate the review dataset into training, validation, and test datasets as given in Table 6.To reduce the impact of RNN's limits and take advantage of LSTM's strengths, we combine RNN and LSTM in our hybrid model to calculate the average ratings of reviews. We use the previously generated BERT annotation output to prepare the model's review dataset. We use the review datasets from Table 3, including the most widely used video conferencing apps, to train our model.

Algorithm 2 Average Rating Calculation Procedure

Require: Average Actual Ratings
Ensure: $R_D \Leftarrow$ BERT-base Annotated Reviews
$R_{total} \Leftarrow \{\}$
$R_{count} \Leftarrow \{\}$ $Avg_{rating} \Leftarrow \{\}$
if $R_D = \{\}$ **then**
  No Ratings found for the app
return
end if
while ratings in R_D **do**
  $R_{total} = R_{total} +$ ratings
  $R_{count} = R_{count} + 1$
  $Avg_{rating} = \frac{R_{total}}{R_{count}}$
end while
Return Avg_{rating}

We seeded the algorithm with preprocessed review datasets for every app. Next, we use the review datasets provided in Table 3 to determine the actual ratings of the Skype, Microsoft Teams, Zoom Cloud Meetings, and Google Meet apps. The textual algorithm for rating calculation is displayed in Algorithm 2. We used LSTM layers after RNN layers to accomplish a difficult rating computation assignment. In the suggested hybrid model, we employ the softmax activation function at the top layers as given in Equation 11.

$$CR_{Entropy} = -\sum_{i}^{5} X_j log f(Sc_j) \tag{11}$$

The scores are computed using Equation 12, and the probability of ratings is found using the SoftMax activation function.

$$f(Sc_j) = \frac{e^s j}{\sum_k^5 e^s j} \tag{12}$$

Table 6. Training Testing Dataset

Source	Training	Validataion	Test	Total Reviews
Google Meet	9338	1167	1167	11672
Microsoft Teams	4948	619	618	6185
Zoom Cloud Meeting	6054	757	757	7568
Hangout	2615	327	327	3269
Others	4858	607	608	6073
Skype	8000	1000	1000	10,000
Grand Total	35,813	4,477	4,477	44,767

4. Results & Discussion

To forecast the average sentiment scores for the user reviews, we employed a hybrid model. Then, to find the commonalities in the reasons given in the reviews for the developer to assist in creating apps that may satisfy users.

4.1. Evaluation Metrices

The accuracy formula, defined as the ratio of properly classified samples to the total number of samples [76], may be found in Equation 13.

$$Accu = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{13}$$

Equation 14 provides the percentage of correctly categorized positive and negative attitudes, which is calculated using the Precision formula.

$$PR = \frac{TP}{(TP + FP)} \quad (14)$$

It matters more to have a high Recall rate. The proper recognition ratio (Recall) is the sum of the positive and negative attitude counts. Equation 15 presents the procedure for computing the Recall, a critical indicator in sentiment analysis [76].

$$R = \frac{TRP}{TRP + FAN} \quad (15)$$

It provides a reasonable evaluation of accuracy and recall. Equation 16 provides the formula for determining F1, the harmonic mean of Precision and Recall.

$$F1 = \frac{2 * (PR * R)}{(PR + R)} \quad (16)$$

4.2. Experimental Setup

An experimental setup for a hybrid model with different parameters and their values is shown in Table 7.

Table 7. Training Testing Dataset

Specifications	Values
Total Review	44,767
No. of Datasets	5
Models	BERT followed by RNN + LSTM
Number of Epochs	10-70
Platform	Jupiter Notebook
Languages	Python
Hardware	RAM 16GB + SSD
GPU	RTX3070, and Intel Xeon W1370

Table 8. Sentiment based Reviews Distribution

Sentiment	Review	Percentage (%)
Positive	29,098	65
Negative	13,431	30
Neutral	2,238	5
Total	44,767	100

4.3. Results and Evaluation

Following the model's application to the preprocessed reviews, the reviews were divided into two stages: thematic analysis and sentiment analysis. The distribution of reviews by sentiment (negative, neutral, and positive) is displayed in Table 8. According to the data, 65% of the reviews are favorable, indicating that most users of these programs were happy and found them to be beneficial. 30% of the reviews were negative, and the neutral reviews accounted for no more than 5%. Out of all the negative feedback, 4 frequent issues accounted for 60% of the unfavorable ratings.

Table 9. Sentiment based Reviews Distribution

Sentiment	Review	Percentage (%)

Positive	29,098	65
Negative	13,431	30
Neutral	2,238	5
Total	44,767	100

With 26% of unfavorable reviews, issues with frequent crashes were the most common. Other issues with technical reasons came in second with 14%, learning challenges with 15%, and reporting updates with 5%.

Table 10. Google Meet Platform

F1 Score	Recall	Precision	Ratings
0.57	0.62	0.59	1
0.34	0.41	0.45	2
0.31	0.37	0.51	3
0.29	0.24	0.46	4
0.83	0.79	0.65	5

The model is trained on five distinct datasets using an RNN and LSTM combination. This smoothness demonstrates how well-trained and broadly applicable the model is to validate the data. Additionally, it demonstrates that neither overfitting nor underfitting exists, which is encouraging for the suggested hybrid model.

Table 11. Zoom Cloud Meeting Platform

F1 Score	Recall	Precision	Ratings
0.59	0.72	0.61	1
0.58	0.39	0.50	2
0.33	0.31	0.54	3
0.32	0.26	0.44	4
0.86	0.81	0.66	5

Table 12. Microsoft Teams Platform

F1 Score	Recall	Precision	Ratings
0.61	0.72	0.68	1
0.42	0.49	0.57	2
0.41	0.51	0.44	3
0.32	0.46	0.54	4
0.89	0.91	0.76	5

Because this model performs better overall than SOTA models, it will give developers the features they need to know what users think when new versions of the apps are released. Table 9 provides an overview of the model’s performance for each of the five rating classes. We found that models trained on Zoom Cloud Meeting, Google Meet, and Microsoft Teams perform better overall based on each model we used. In 2017, Google created Google Meet, the initial iteration of Hangouts Meet. It was enhanced for new needs after being employed in the business service packages of the company [77]. Google Meet allowed collaboration on documents, holding audio and video chats, and using other tools and Google disc resources while working on documents. With the COVID-19 epidemic spreading, Google decided to make this product available without payment in March 2020. This move increased the service’s global appeal.

Table 13. BERT-base Model

F1 Score	Recall	Precision	Ratings

0.62	0.71	0.61	1
0.32	0.59	0.52	2
0.43	0.41	0.54	3
0.42	0.52	0.44	4
0.81	0.81	0.66	5

It was not as good as its rivals, though, in terms of instruction. Table 10 presents the outcomes of the Google Meet review dataset used in this work. Our accuracy rate for analyzing Google Meet review ratings was 0.67. Teachers can use the provided system's capabilities while engaging in textual and oral communication on Zoom's online platform. The option to record conferences is one of this resource's main benefits. This will allow the teacher to review earlier lessons that the students missed at a later time. The Zoom platform directly aids in the development of the aforementioned abilities by encouraging the utilization of many creative opportunities [78]. The results of the Zoom Cloud Meeting review dataset utilized in this experiment are shown in Table 11. We analyzed review ratings for Zoom Cloud Meetings, and our accuracy rate was 0.77. In 2017, Microsoft Teams was introduced. The project's creator intended for Office 365 users to be able to interact with one another via video conference. Microsoft Teams is included in Office 365, can be integrated with other programs, and is a component of the Microsoft Office Suite subscription. Users could only access an alternate version of this program in 2018, which operated without Office 365 membership and had several limitations (such as a cap on the number of people who could join a video conference). In European schools, Microsoft Teams has become more and more popular.

Table 14. RNN Model

F1 Score	Recall	Precision	Ratings
0.71	0.62	0.58	1
0.32	0.59	0.67	2
0.51	0.41	0.54	3
0.42	0.56	0.34	4
0.79	0.85	0.66	5

The U.S. university system also makes use of the platform [77]. Based on its four primary features—chat, a teamwork center, customization possibilities, and strong security—Microsoft Teams is an adaptable platform for collaboration. Lectures and seminars can also be given by instructors using Microsoft Teams. Table 11 displays the outcomes of the Microsoft Teams review dataset used in this investigation. Our accuracy rate was 0.70 while analyzing Microsoft Teams review ratings. The single performance of the cutting-edge pre-trained models we employed in our work was compiled in Tables 13,14, and 15. We discovered that models trained on Zoom Cloud Meeting, Microsoft Teams, and Google Meet outperform models such as BERT-base, and RNN-LSTM in terms of accuracy, based on the overall performance of all the models. We verify that other review datasets similarly show improvements in precision, recalls, and f1-scores. We verify that other review datasets similarly show improvements in precision, recalls, and f1-scores.

Table 15. LSTM Model

F1 Score	Recall	Precision	Ratings
0.60	0.52	0.58	1
0.52	0.49	0.67	2
0.61	0.61	0.54	3
0.42	0.66	0.64	4
0.90	0.94	0.79	5

4.4. Discussion

Table 16 summarises the real ratings for the Microsoft Teams, Zoom Cloud Meeting, and Google Team app evaluations along with the calculated ratings for each of the five models that were created through training on five distinct datasets.

Table 16. Average Ratings of Review Datasets

Avg Ratings	Microsoft Teams	Zoom Cloud Meeting	Google Meet	All other Apps
Calculated	4.05	3.68	3.42	3.15
Rating on Datasets	4.52	4.16	3.12	3.40

The result shows that the average rating obtained from our surveys is greater than the average ratings found in the dataset, suggesting that the reviews' true assessments of the product are higher than the assessments they gave when reviewing it. We can provide developers with feedback and state that our work is operating well based on the actual ratings of upgraded versions. Because the model was trained on small datasets like Hangout, it can only yield findings that are as discerning as the true average rating. The model's output, which was trained on extensive datasets such as Zoom, Meet, and Teams, is more discerning than the average rating itself. This will provide the software developers with the input they require and validate the positive perception that users have of the most current iterations of the apps that were the focus of this investigation. The comparison of the app features shows that Zoom and Google Meet are not less feature-rich than Microsoft Teams as shown in Table 17.

Table 17. Features Comparison of Review Apps

Features	Zoom Cloud Meeting	Microsoft Teams	Google Meet
Wave hand	Y	Y	Y
Talk feature	Y	Y [100 Participant]	Y
Material distribution	Organizer	All	N
Transferring Files	N	Y	Y
Document Function	Y	Y	NA
Track of Visitors	Local/Cloud	Cloud	Y
Remote Management	300 People	250 People	Y
Calendar integration	Probable	Accessible	NA
Waiting room	Y	Y	-
Context blurry feature	Y	Y	NA
Sharing Videos	All	Conference Presenter	Y

Electronic Board	Y	Y	N

The ”Raise Hand” feature is the only item lacking. It is asserted, nevertheless, that if every student in the class has their cameras turned on, then this won’t be an issue. Zoom does not have the file transfer functionality at this time. Though we believe Microsoft Teams provides excellent functionality overall, we also notice that Google Meet has added more capabilities in recent years. Because of the numerous elements contributing to the rapid development of remote learning, it is anticipated that this platform will be used more extensively and more actively in higher education institutions. In conclusion, higher education institutions undervalue the potential of the Google Meet platform. Naturally, this resource’s tools enable a thorough examination of students’ training, the timely and high-quality transfer of essential information, the comfortable conduct of classes, etc. The advancement of higher education can benefit from the adoption of Microsoft Teams. Our hybrid model results, together with the features these applications offer, suggest that developers update their apps to give users improved video conferencing facilities and stay current with their features.

4.5. Swot Analysis of Apps

We performed a SWOT study to compare the pros and cons of Google Meet, Zoom Cloud Meeting, and Microsoft Teams to provide a comprehensive review of our hybrid model. The results are compiled in Table 18. Swot analysis provides a quick summary of the study’s apps, indicating that they are straightforward to use and have intuitive interfaces for training management. Every platform offers a wide range of tools to raise the standard of instruction.

Table 18. SWOT Analysis of Review Apps

App	Advantages	Disadvantages
Zoom Cloud Meeting	Simple design Setting up training sessions Personal conversation	Time limit 100 participant Messages after joining
Microsoft Teams	Part of the Office365 Max 300 Persons Common Mode Discussions	No Firefox support Complex interface Meetings on closed System
Google Meet	Enterprise package Digital whiteboard Friendly interface Protected by Google security policy	Time limit 1hr Poor sound quality No recording in virtual room No individual chat

Furthermore, a conspicuous drawback of the Zoom and Google Meet platforms is the upper limit of conference attendees (one hundred). As a result, it is difficult to envision the planning of an extensive virtual conference with lots of attendees. Nevertheless, there is no such restriction on the platforms’ premium versions. Working online demands extra caution because of the modern hybrid risks and cyber problems. Although Microsoft Teams, Zoom, and Google Meet are well-established businesses, several issues exist with their operations. Less personal information should be shared with apps, according to experts. Learning how to use privacy and security settings and becoming comfortable with them is also crucial [79]. To begin with, it is preferable to register and create a video conference using a different email address or account that is not linked to any bank accounts, social media profiles, etc. Using a secure password with a variety of letters and digits is also a smart idea. Remember that video conferences are recorded, so if at all possible, switch off the camera and microphone. Because Microsoft Teams uses encrypted communication and Microsoft security features, we think it’s the most secure platform. Zoom is a less secure platform. Zoom bombing bullying has been done with a straightforward numerical meeting ID. It involves inviting strangers to the conference to demean every one of them and interfering with the proceedings. Additionally, the proprietors of Zoom are accused of disclosing the personal information of its clients [80]. Consequently, several governmental organizations declined to collaborate with this platform. As a result, every online learning platform under study has unique risks and opportunities.

4.6. Recommendations for App Developers

Table 19. Recommendations based on Average Ratings

App	Recommendations
Zoom Cloud Meeting	AI Powered Meeting Summaries Presenter Tracking Security Enhancements Customized Waiting Room Improved File Sharing Better Mobile Experience Advanced Background Features
Microsoft Teams	Persistent Chat across Meetings Improved Meeting Transcriptions Dynamic Layouts Optimized Resource Usage Better External Collaboration
Google Meet	Enhanced Noise Cancellation Integrated Whiteboard Improved Breakout Rooms Meeting Analytics Offline Recording Access Integration with Third-Party Apps Customized Meeting Layouts

We propose many areas of app improvement, as shown in Table 19, by training a hybrid model with BERT-base and RNN-LSTM on various datasets and basing our recommendations on the difference between the average review ratings computed and real ratings. The suggestions will assist app developers in enhancing the functionality of their video conferencing applications to better meet the needs of users, advance business objectives, and play a supporting role.

5. Conclusion and Future work

These days, governmental and scholarly materials published online are widely used worldwide. During this period, the majority of educational institutions and organizations used Microsoft Teams, Google Meet, or Zoom Cloud Meeting. For this reason, to enhance the development of apps, the developers must release some essential features. Users' feedback is necessary for app developers to address these and other technical problems. The issue, though, is that some individuals failed to offer accurate ratings in the reviews. To prevent overfitting or underfitting, we used data augmentation techniques to improve the review dataset. To better understand user perspectives and improve functionality for Google Meet, Microsoft Teams, Zoom Cloud Meeting, and all other apps, this study gave recommendations utilizing a BERT base, RNN, and LSTM technique. A benchmark dataset comprising 44,767 reviews from various apps was made available. We compared our results with actual ratings using BERT-base for annotation and RNN-LSTM for average rating computation. This allowed us to give developers input on how to improve the app development process. Based on average ratios, the following apps have been rated: Zoom Cloud Meeting (4.16), Google Meet (3.12), Microsoft Teams (4.52), and all other apps (3.40). It suggests that users of the Teams and Zoom apps will benefit from the most recent version. Having a huge review dataset will improve the model's performance. However, there aren't enough reviews for us to do so. Because of this, we will continue working on this project as long as there are enough reviews to eliminate any uncertainty. We concluded that, although our hybrid technique produced fewer negative reviews and more positive ones, it had a greater influence on developers' suggestions regarding how to improve app features than the actual review scores. We think that the suggestions made in this study will assist app developers in making their apps better. Furthermore, the dataset and analysis that are offered will serve as a foundation for further study of this subject.

References

1. Ramzy, M., Ibrahim, B.: User satisfaction with arabic covid-19 apps: Senti- ment analysis of users' reviews using machine learning techniques. Information Processing & Management 61(3), 103644 (2024)
2. Abo, M.E.M., Raj, R.G., Qazi, A.: A review on arabic sentiment analysis: state-of- the-art, taxonomy and open research challenges. IEEE Access 7, 162008–162024 (2019)
3. Motger, Q., Miaschi, A., Dell'Orletta, F., Franch, X., Marco, J.: T-frex: A transformer-based feature extraction method from mobile app reviews. In: 2024 IEEE International Conference on Software Analysis, Evolution and Reengineer- ing (SANER), pp. 227–238 (2024). IEEE
4. S¨allberg, H., Wang, S., Numminen, E.: The combinatory role of online ratings and reviews in mobile app downloads: an empirical investigation of gaming and pro- ductivity apps from their initial app store launch. Journal of Marketing Analytics 11(3), 426–442 (2023)
5. Raman, R., Kumar, V., Pillai, B.G., Rabadiya, D., Bhoyar, P.K., Kumbhojkar, N.: Analyzing sentiments in motorku x app reviews using the naive bayes clas- sifier approach. In: 2024 Second International Conference on Data Science and Information System (ICDSIS), pp. 1–5 (2024). IEEE
6. Kumar, A., Bala, P.K., Chakraborty, S., Behera, R.K.: Exploring antecedents impacting user satisfaction with voice assistant app: A text mining-based analysis on alexa services. Journal of Retailing and Consumer Services 76, 103586 (2024)
7. Okoro, E.P.: Application of zoom e-learning management system in business edu- cation in universities nigeria. International Journal of Religion 5(5), 317–326 (2024)
8. A'isy, Q., Mufaridah, F., Nurkamilah, N.: The challenges of teaching speaking to high school students through zoom video conference application. Pubmedia Jurnal Pendidikan Bahasa Inggris 1(2), 8–8 (2024)
9. Rajeshkumar, C., Rajakumari, K.: E-learning technologies with interactive fac- tors and elements on the perspective of knowledge management. In: Human Machine Interaction in the Digital Era, pp. 35–38. CRC Press (2024)
10. Al-Maroof, R.S., Salloum, S.A., Hassanien, A.E., Shaalan, K.: Fear from covid-19 and technology adoption: the impact of google meet during coronavirus pandemic. Interactive Learning Environments 31(3), 1293–1308 (2023)
11. Maleki, A.: "it is not only about us!": Investigating efl learners' perspectives towards reasons of online exam cheating. Journal of Academic Ethics, 1–16 (2024)
12. Shevchenko, V., Malysh, N., Tkachuk-Miroshnychenko, O.: Distance learning in ukraine in covid-19 emergency. Open Learning: The Journal of Open, Distance and e-Learning 39(1), 4–19 (2024)
13. Zhao, Y., Hou, X., Wang, S., Wang, H.: Llm app store analysis: A vision and roadmap. arXiv preprint arXiv:2404.12737 (2024)
14. Sarma, A., Chen, N.: Effective teaching through code reviews: Patterns and anti- patterns. Proceedings of the ACM on Software Engineering 1(FSE), 1262–1283 2024
15. Malhotra, R., Meena, S.: A systematic review of transfer learning in software engineering. Multimedia Tools and Applications, 1–62 (2024)
16. Lyu, Q., Apidianaki, M., Callison-Burch, C.: Towards faithful model explanation in nlp: A survey. Computational Linguistics, 1–67 (2024)
17. Jim, J.R., Talukder, M.A.R., Malakar, P., Kabir, M.M., Nur, K., Mridha, M.: Recent advancements and challenges of nlp-based sentiment analysis: A state-of- the-art review. Natural Language Processing Journal, 100059 (2024)
18. Shen, Z., Pan, Y., Tan, K., Ji, H., Chu, S.: Educational innovation in the digital age: The role and impact of nlp technology. OLD AND NEW TECHNOLOGIES OF LEARNING DEVELOPMENT IN MODERN CONDITIONS 281 (2024)
19. Liu, Z., Chen, C., Wang, J., Huang, Y., Hu, J., Wang, Q.: Guided bug crush: Assist manual gui testing of android apps via hint moves. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2022)
20. RIEZLER, S.H., Hagmann, M.: Validity, Reliability, and Significance: Empirical Methods for NLP and Data Science. Springer, (2024)
21. Naveed, H., Anwar, S., Hayat, M., Javed, K., Mian, A.: Survey: Image mixing and deleting for data augmentation. Engineering Applications of Artificial Intelligence 131, 107791 (2024)
22. Kangaslahti, S., Alvarez-Melis, D.: Continuous language model interpolation for dynamic and controllable text generation. arXiv preprint arXiv:2404.07117 (2024)
23. Lu, Z., Zhou, A., Ren, H., Wang, K., Shi, W., Pan, J., Zhan, M., Li, H.: Math- genie: Generating synthetic data with question back-translation for enhancing mathematical reasoning of llms. arXiv preprint arXiv:2402.16352 (2024)

24. Nam, D., Macvean, A., Hellendoorn, V., Vasilescu, B., Myers, B.: Using an llm to help with code understanding. In: Proceedings of the IEEE/ACM 46th International Conference on Software Engineering, pp. 1–13 (2024)
25. Zhang, M., Jiang, G., Liu, S., Chen, J., Zhang, M.: Llm–assisted data augmen- tation for chinese dialogue–level dependency parsing. Computational Linguistics, 1–24 (2024)
26. Nuryana, Z., Pangarso, A., Zain, F.M., et al.: Factor of zoom cloud meet- ings: Technology adoption in the pandemic of covid-19. International Journal of Evaluation and Research in Education 10(3), 816–825 (2021)
27. Kansal, A.K., Gautam, J., Chintalapudi, N., Jain, S., Battineni, G.: Google trend analysis and paradigm shift of online education platforms during the covid-19 pandemic. Infectious Disease Reports 13(2), 418–428 (2021)
28. Ilag, B.N., Sabale, A.M.: Microsoft teams overview. In: Troubleshooting Microsoft Teams: Enlisting the Right Approach and Tools in Teams for Mapping and Troubleshooting Issues, pp. 17–74. Springer (2022)
29. Daniel, S.J.: Education and the covid-19 pandemic. Prospects 49(1), 91–96 (2020)
30. Tarkar, P.: Impact of covid-19 pandemic on education system. International Journal of Advanced Science and Technology 29(9), 3812–3814 (2020)
31. Liptrot, E., Pearson, H.A., Montazami, A., Dub´e, A.K.: Why this app? how user ratings and app store rankings impact educators' selection of educational apps. Computers & Education 218, 105080 (2024)
32. Sun, R., Xue, M., Tyson, G., Wang, S., Camtepe, S., Nepal, S.: Not seen, not heard in the digital world! measuring privacy practices in children's apps. In: Proceedings of the ACM Web Conference 2023, pp. 2166–2177 (2023)
33. Ekambaranathan, A., Zhao, J., Chalhoub, G.: Navigating the data avalanche: Towards supporting developers in developing privacy-friendly children's apps. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 7(2), 1–24 (2023)
34. Dkabrowski, J., Letier, E., Perini, A., Susi, A.: Analysing app reviews for software engineering: a systematic literature review. Empirical Software Engineering 27(2), 43 (2022)
35. Malgaonkar, S., Licorish, S.A., Savarimuthu, B.T.R.: Prioritizing user concerns in app reviews–a study of requests for new features, enhancements and bug fixes. Information and Software Technology 144, 106798 (2022)
36. Batra, H., Punn, N.S., Sonbhadra, S.K., Agarwal, S.: Bert-based sentiment analysis: A software engineering perspective. In: Database and Expert Sys- tems Applications: 32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I 32, pp. 138–148 (2021). Springer
37. Hai, T., Zhou, J., Li, N., Jain, S.K., Agrawal, S., Dhaou, I.B.: Cloud-based bug tracking software defects analysis using deep learning. Journal of Cloud Computing 11(1), 32 (2022)
38. Noei, E., Zhang, F., Zou, Y.: Too many user-reviews! what should app developers look at first? IEEE Transactions on Software Engineering 47(2), 367–378 (2019)
39. Li, S., Guo, J., Fan, M., Lou, J.-G., Zheng, Q., Liu, T.: Automated bug reproduc- tion from user reviews for android applications. In: Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering in Practice, pp. 51–60 (2020)
40. Tushev, M., Ebrahimi, F., Mahmoud, A.: Domain-specific analysis of mobile app reviews using keyword-assisted topic models. In: Proceedings of the 44th International Conference on Software Engineering, pp. 762–773 (2022)
41. Nakamura, W.T., Oliveira, E.C., Oliveira, E.H., Redmiles, D., Conte, T.: What factors affect the ux in mobile apps? a systematic mapping study on the analysis of app store reviews. Journal of Systems and Software 193, 111462 (2022)
42. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
43. Malik, K. R., Sajid, M., Almogren, A., Malik, T. S., Khan, A. H., Altameem, A., ... & Hussen, S. (2025). A hybrid steganography framework using DCT and GAN for secure data communication in the big data era. Scientific Reports, 15(1), 19630.
44. Mangnoesing, G.V., Tru¸scˇa, M.M., Frasincar, F.: Pattern learning for detecting defect reports and improvement requests in app reviews. In: Natural Language Processing and Information Systems: 25th International Conference on Applica- tions of Natural Language to Information Systems, NLDB 2020, Saarbru¨cken, Germany, June 24–26, 2020, Proceedings 25, pp. 126–134 (2020). Springer
45. Belinkov, Y., Bisk, Y.: Synthetic and natural noise both break neural machine translation. arXiv preprint arXiv:1711.02173 (2017)
46. Liesting, T., Frasincar, F., Tru¸scˇa, M.M.: Data augmentation in a hybrid approach for aspect-based sentiment analysis. In: Proceedings of the 36th Annual ACM Symposium on Applied Computing, pp. 828–835 (2021)
47. Sajid, M., Aslam, N., Abid, M. K., & Fuzail, M. (2022). RDED: recommendation of diet and exercise for diabetes patients using restricted boltzmann machine. VFAST Transactions on Software Engineering, 10(4), 37-55.

48. Cho, K., Van Merri¨enboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder- decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
49. Balakrishnan, V., Shi, Z., Law, C.L., Lim, R., Teh, L.L., Fan, Y.: A deep learning approach in predicting products' sentiment ratings: a comparative analysis. The Journal of Supercomputing 78(5), 7206–7226 (2022)
50. Tran, D.D., Nguyen, T.T.S., Dao, T.H.C.: Sentiment analysis of movie reviews using machine learning techniques. In: Proceedings of Sixth International Congress on Information and Communication Technology: ICICT 2021, London, Volume 1, pp. 361–369 (2022). Springer
51. Joy, D.T., Thada, V., Srivastava, U.: Sentiment analysis on global warming tweets using na¨ıve bayes and rnn. In: Data Engineering for Smart Systems: Proceedings of SSIC 2021, pp. 225–234 (2022). Springer
52. Sany, M.M.H., Keya, M., Khushbu, S.A., Rabby, A.S.A., Masum, A.K.M.: An opinion mining of text in covid-19 issues along with comparative study in ml, bert & rnn. In: International Conference on Deep Learning, Artificial Intelligence and Robotics, pp. 1–16 (2021). Springer
53. Javed, N., Muralidhara, B.: Emotions during covid-19: Lstm models for emotion detection in tweets. In: Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2021, pp. 133–148 (2022). Springer
54. Huang, Z., Xu, W., Yu, K.: Bidirectional lstm-crf models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
55. Rani, S., Bashir, A.K., Alhudhaif, A., Koundal, D., Gunduz, E.S., et al.: An efficient cnn-lstm model for sentiment detection in# blacklivesmatter. Expert Systems with Applications 193, 116256 (2022)
56. Liu, Y., Wang, L., Shi, T., Li, J.: Detection of spam reviews through a hierarchical attention architecture with n-gram cnn and bi-lstm. Information Systems 103, 101865 (2022)
57. Sajid, M., Khan, A. H., Malik, T. S., Bilal, A., Ahmad, Z., & Sarwar, R. (2025). Enhancing melanoma diagnostic: harnessing the synergy of AI and CNNs for groundbreaking advances in early melanoma detection and treatment strategies. International Journal of Imaging Systems and Technology, 35(1), e70016.
58. Islam, M.J., Datta, R., Iqbal, A.: Actual rating calculation of the zoom cloud meetings app using user reviews on google play store with sentiment annotation of bert and hybridization of rnn and lstm. Expert Systems with Applications 223, 119919 (2023)
59. Sajid, M., Sanaullah, M., Fuzail, M., Malik, T. S., & Shuhidan, S. M. (2025). Comparative analysis of text-based plagiarism detection techniques. PloS one, 20(4), e0319551.
60. Gorad, B., Kotrappa, S.: Novel dataset generation for indian brinjal plant using image data augmentation. In: IOP Conference Series: Materials Science and Engineering, vol. 1065, p. 012041 (2021). IOP Publishing
61. Duong, H.-T., Nguyen-Thi, T.-A.: A review: preprocessing techniques and data augmentation for sentiment analysis. Computational Social Networks 8(1), 1 (2021)
62. Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., Zhao, L., Xu, S., Liu, W., Liu, N., et al.: Auggpt: Leveraging chatgpt for text data augmentation. arXiv preprint arXiv:2302.13007 (2023)
63. Sajid, M., Razzaq Malik, K., Ur Rehman, A., Safdar Malik, T., Alajmi, M., Haider Khan, A., ... & Hussen, S. (2025). Leveraging two-dimensional pre-trained vision transformers for three-dimensional model generation via masked autoencoders. Scientific Reports, 15(1), 3164.
64. Kumar, V., Choudhary, A., Cho, E.: Data augmentation using pre-trained transformer models. arXiv preprint arXiv:2003.02245 (2020)
65. Coulombe, C.: Text data augmentation made simple by leveraging nlp cloud apis. arXiv preprint arXiv:1812.04718 (2018)
66. Papageorgiou, G., Economou, P., Bersimis, S.: A method for optimizing text preprocessing and text classification using multiple cycles of learning with an application on shipbrokers emails. Journal of Applied Statistics, 1–35 (2024)
67. Siino, M., Tinnirello, I., La Cascia, M.: Is text preprocessing still worth the time? a comparative survey on the influence of popular preprocessing methods on transformers and traditional classifiers. Information Systems 121, 102342 (2024)
68. Rodrigues, A.P., Vijaya, P., Fernandes, R., et al.: Tulu language text recognition and translation. IEEE Access (2024)
69. Abo, M.E.M., Idris, N., Mahmud, R., Qazi, A., Hashem, I.A.T., Maitama, J.Z., Naseem, U., Khan, S.K., Yang, S.: A multi-criteria approach for arabic dialect sen- timent analysis for online reviews: Exploiting optimal machine learning algorithm selection. Sustainability 13(18), 10018 (2021)
70. Malik, H., Shakshuki, E.M., Yoo, W.-S.: Comparing mobile apps by identifying 'hot'features. Future Generation Computer Systems 107, 659–669 (2020)

71. Sajid, M., Razzaq Malik, K., Ur Rehman, A., Safdar Malik, T., Alajmi, M., Haider Khan, A., ... & Hussen, S. (2025). Leveraging two-dimensional pre-trained vision transformers for three-dimensional model generation via masked autoencoders. Scientific Reports, 15(1), 3164.
72. Luo, X., Ding, H., Tang, M., Gandhi, P., Zhang, Z., He, Z.: Attention mecha- nism with bert for content annotation and categorization of pregnancy-related questions on a community q&a site. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1077–1081 (2020). IEEE
73. Sajid, M., Khan, A. H., Malik, K. R., Khan, J. A., & Alwadain, A. (2025). A new approach of anomaly detection in shopping center surveillance videos for theft prevention based on RLCNN model. PeerJ Computer Science, 11, e2944.
74. Nemes, L., Kiss, A.: Social media sentiment analysis based on covid-19. Journal of Information and Telecommunication 5(1), 1–15 (2021)
75. Mourad, A., Elbassuoni, S.: A large-scale analysis of covid-19 tweets in the arab region. Social Network Analysis and Mining 12(1), 71 (2022)
76. Sajid, M., Malik, K.R., Almogren, A., Malik, T.S., Khan, A.H., Tanveer, J., Rehman, A.U.: Enhancing intrusion detection: a hybrid machine and deep learning approach. Journal of Cloud Computing 13(1), 123 (2024)
77. Kem, D.: Personalised and adaptive learning: Emerging learning platforms in the era of digital and smart learning. International Journal of Social Science and Human Research 5(2), 385–391 (2022)
78. Alfadda, H.A., Mahdi, H.S.: Measuring students' use of zoom application in language course based on the technology acceptance model (tam). Journal of Psycholinguistic Research 50(4), 883–900 (2021)
79. Yovenko, L., Novakivska, L., Sanivskyi, O., Sherman, M., Vysochan, L., Hnedko, N.: Pedagogical analysis of the phenomenon of digital competence. International journal of computer science and network security: IJCSNS 21(6), 7–10 (2021)
80. Wedari, L.K., Fatihah, A.N., Rusmanto, T.: Zoom application acceptance in online learning: An analysis with the technology acceptance model. International Journal of Information and Education Technology 12(9), 821–830 (2022)