JIMEA | Jurnal Ilmiah MEA (Manajemen, Ekonomi, dan Akuntansi)
Vol. 8 No. 2, 2024

UNDERSTANDING SERVICE QUALITY OF MOBILE VIDEO EDITING : MAPPING THE NEGATIVE IMPRESSION BY TEXT MINING APPROACH

Maya Ariyanti[1]; Yumna Tazkia[2]
School of Economics and Business, Telkom University, Bandung, Indonesia[1,2]
Email : ariyanti@telkomuniversity.ac.id[1]; yumnatazkia@student.telkomuniversity.ac.id[2]

ABSTRACT

KineMaster is a video editing application that supports the content creator industry; however, compared to its competitors, that app falls short in release year, download numbers, and ratings. This research aims to determine the service quality of the Android-based KineMaster application based on sentiment analysis and the classification of mobile app service quality (MASQ) dimensions. The data used is secondary data from 5,000 reviews of Google Play Store using Google Colab and processed using RapidMiner Studi version 10.2. Naïve Bayes and k-Nearest Neighbors (KNN) algorithms are applied to determine the best one. Negative sentiment data resulting from the worst MASQ dimension classification will be carried out by WordCloud using Google Colab to determine complaint priorities. The research results show that positive sentiment dominates at 62.24% using the KNN algorithm as the best algorithm in this research. Nevertheless, the 37.76% negative sentiment is not ignored. Based on the number of negative sentiments in each dimension, technical reliability is the worst dimension, valence is the second worst dimension, and performance is the third worst. Prioritized complaints are update reliability, watermarks, app, feature downloads, inability to open apps, export capabilities, high price, and processing speed.

Keywords : Complaint Prioritization; Dimension Classification; Sentiment Analysis; Service Quality; WordCloud

ABSTRAK

KineMaster merupakan aplikasi edit video yang mendukung industri pembuat konten; namun, jika dibandingkan dengan pesaingnya, aplikasi tersebut kurang dalam hal tahun rilis, jumlah unduhan, dan peringkat. Penelitian ini bertujuan untuk mengetahui kualitas layanan aplikasi KineMaster berbasis Android berdasarkan analisis sentimen dan klasifikasi dimensi mobile app service quality (MASQ). Data yang digunakan merupakan data sekunder dari 5.000 review Google Play Store menggunakan Google Colab dan diolah menggunakan RapidMiner Studi versi 10.2. Algoritma Naïve Bayes dan k-Nearest Neighbors (KNN) dipilih untuk menentukan algoritma yang terbaik. Data sentimen negatif hasil klasifikasi dimensi MASQ terburuk akan dilakukan oleh WordCloud menggunakan Google Colab untuk menentukan prioritas keluhan. Hasil penelitian menunjukkan bahwa sentimen positif mendominasi sebesar 62,24% dengan menggunakan algoritma KNN sebagai algoritma terbaik dalam penelitian ini. Meski demikian, sentimen negatif sebesar 37,76% tidak diabaikan. Berdasarkan jumlah sentimen negatif pada masing-masing dimensi, keandalan teknis menjadi dimensi terburuk, valensi menjadi dimensi terburuk kedua, dan kinerja menjadi dimensi terburuk ketiga. Keluhan yang diprioritaskan adalah keandalan pembaruan, watermark, aplikasi, pengunduhan fitur, ketidakmampuan membuka aplikasi, kemampuan ekspor, harga yang mahal, dan kecepatan pemrosesan.

Kata kunci : Analisis Sentimen; Klasifikasi Dimensi; Kualitas Layanan; Prioritas Keluhan; WordCloud

INTRODUCTION

Internet use in Indonesia has increased by 5.44% compared to last year (wearesocial.com, 2023). The increasing internet use in Indonesia has indirectly raised awareness among Indonesians in the age of digital and technological advancements. A new job trend is emerging, commonly known as content creator. Moreover, the market value of the Indonesian content creator industry is estimated to reach IDR 4 – IDR 7 trillion, and this value will increase fivefold by 2027 (katadata.co.id, 2022). The potential in this industry and creator ecosystem can significantly boost the Indonesian economy. This will assist in achieving the 8th Sustainable Development Goal (SDG), namely decent work and economic growth.

Many companies also support content creators' jobs by releasing special software products (applications) to help the work of content creators, one of which is Android-based video editing. According to cnnindonesia.com (2022), dailysocial.id (2022) and idntimes.com (2023), several video editing apps are recommended for Android devices, namely CapCut, Inshot, VN, or Kinemaster. Unfortunately, the KineMaster, despite being the first to launch, is not as advanced as its competitors on the Google Play Store.

The table 1 of comparison video editing apps from Google Play Store in 2023 clearly shows that the number of downloaders and KineMaster ratings are still lagging those of its competitors. In fact, KineMaster was released first in the Google Play Store. This means that the level of development or popularity of the Android-based KineMaster application remains low on the Google Play Store. Moreover, complaints will still appear until 2023 in Google Play Store reviews. Users are expressing their dissatisfaction through these complaints about the KineMaster application. Therefore, it is necessary to analyze service quality from reviews of the Android-based KineMaster application using the Mobile Apps Service Quality (MASQ) dimension approach to assess shortcomings and make necessary improvements.

Understanding service quality refers to a company's ability to accurately assess and evaluate its services' quality (Aditya *et al*., 2023). By understanding service quality, complaints can be identified. Complaint is an action that arises as a reaction to dissatisfaction in the provision of services by service providers with various choices of demands or settlements with and/or external parties (Bafadhal, 2022). Thus, complaints about service quality can cause user dissatisfaction, where user dissatisfaction can reduce the level of user loyalty to the service (Zuliestiana & Setiawan, 2022) . In this way, service quality becomes an important element in customer expectations and a dominant element in customer evaluation (Zeithaml *et al*., 2018). Therefore, negative impression can improve the service (Ariyanti & Rifaldi, 2019).

The study by Wang et al. (2022) introduced the application of text data mining technology in analyzing customer service complaints, indicating the practical utilization of text mining in assessing service quality (Wang et al., 2022). Additionally, Ordenes & Zhang (2019) highlighted the use of text mining in service research to measure consumer sentiment, experiences, and service quality, further emphasizing its relevance in this context (Ordenes & Zhang, 2019). These references collectively support the notion that text mining is being actively employed to gauge service quality by analyzing consumer sentiments and experiences. The study by Mao (2020) utilized text mining software to identify critical retail quality dimensions associated with sporting goods stores, demonstrating the practical application of text mining in evaluating the quality of retail services. Similarly, Mahr et al. (2019) emphasized the rich opportunities for sensorial research through text mining insights, particularly in studies on servicescape, further highlighting the diverse applications of text mining in understanding customer service experiences.

Therefore, text mining approach will be used in this research. In this way, this research aims to determine the service quality of the Android-based KineMaster application based on sentiment analysis (binary classification) and Mobile Apps Service Quality (MASQ) dimension classification (multiclass classification) using the best algorithm. The negative sentiment results in the worst MASQ dimension, which will be formed using WordCloud to determine the priority of the problem or complaint. That way, the most complaints about services can be resolved quickly.

LITERATURE REVIEW

Mobile Apps Service Quality (MASQ)

In general, service quality is the value felt by customers, which is indicated by the better the service, the greater the profits for the company (Hafidz & Muslimah, 2023). This is because service quality is one of the factors that form customers who are loyal to the company (Agiesta et al., 2021). Customer loyalty is a form of advantage for the company because loyal customers will use the company's products or services more often. This is also reinforced by the statement Abdurochman and Tantra (2023) that loss of customer loyalty can result in loss of income stream. Therefore, it is necessary to measure service quality.

The conceptual model to measure service quality is continuously evolving based on necessity. In accordance with statement Abdurochman and Tantra (2023) that as knowledge develops, SERVQUAL can be explored further to more precisely see the quality of service in a particular industry. Thus, the conceptual model of Mobile Apps Service Quality (MASQ) is a concept development of service quality focusing more on mobile application services running on mobile devices (Wulfert, 2019). The MASQ scale measurement consists of three main dimensions, seven of which form the second part. This means that the scale for measuring MASQ applies two levels of dimensional hierarchy, but the measurement still uses more detailed sub-dimensions (Wulfert, 2019). Based on Figure 1, there are three main dimensions: interaction quality, environment quality, and outcome quality. Interaction quality consists of responsiveness, information, security, and privacy. Environment quality consists of design and performance. Outcome quality consists of technical reliability and valence. To find out more, table 2 shows an explanation of the dimensions and sub-dimensions of Mobile Apps Service Quality.

Text Mining

Analysis of service quality can use text mining. Text mining is concerned with the extraction of patterns and knowledge from textual data (Senave et al., 2023). Jo (2019) stated the same thing which text mining is the process of extracting knowledge to make important decisions from textual data. Sentiment analysis is a part of text mining, especially text classification. Sentiment analysis is a study to analyze opinions, sentiments, or emotions applied in every business and social domain (Yahya *et al*., 2023) which can be divided into three sentiments: positive, negative, and neutral. However, according to Valdivia *et al*. (2018), neutral sentiment provides less information because it is ambiguous. Because of that, the sentiment can be classified into positive and negative. It is known as binary classification. Meanwhile, dimension classification is known as multiclass classification because, according to Alamsyah and Rachmadiansyah (2018), multi-class classification is a classification method with more than two target classes. Previous research states that implementing multiclass classification has many advantages in identifying customer relationship management problems because it can provide real-time complaints (Arusada *et al*., 2017). Multiclass classification, part of text mining, answers the challenge of the unstructured nature of online data. This is because it can identify complaints based on appropriate dimensions.

When carrying out binary and multiclass classification, algorithms such as Naïve Bayes, Decision Tree, Random Forest, KNN, or SVM can be used. Naïve Bayes algorithm is widely used because the resulting accuracy is high and supports processing extensive data appropriately (Sari *et al*., 2018; Pratmanto *et al*., 2020; Harahap *et al*., 2022). In the research of Sari *et al*. (2018), the Naïve Bayes accuracy value was 90%, and in the research of Pratmanto *et al*. (2020) was 96.67%. On the other hand, KNN is an easy and simple algorithm because it only pays attention to its nearest neighbors. Research by Harahap *et al*. (2022), which compared Naïve Bayes and KNN, showed that Naïve Bayes is the best algorithm. The accuracy of Naïve Bayes in sentiment analysis was 98.95% and 71.75% for multiclass classification. Otherwise, the research by Limbong *et al*. (2022) compared Naïve Bayes and k-Nearest Neighbors (KNN) to show that the KNN was superior at 92.9% to Naïve Bayes at 91.4% accuracy. There is also research by Damarta *et al*. (2021) that obtained an accuracy value of 87.41% on k-Nearest Neighbors (KNN), so the KNN algorithm can also be used to measure service quality. Thus, Naive Bayes and KNN are compared to find out which is better for binary and multiclass classification.

Framework

Based on previous literature, it can be briefly that Mobile Apps Service Quality (MASQ) is a concept development of service quality that focuses more on mobile application services that run on mobile devices (Wulfert, 2019). This development can be seen from the MASQ scale measurements, namely the three main variables and seven dimensions which form the second part. What this means is that the scale for measuring MASQ applies two levels of dimensional hierarchy, but the measurement still uses more detailed sub-dimensions (Wulfert, 2019).

In accordance with research from Alamsyah & Rachmadiansyah (2018), service quality analysis will be carried out by text mining, especially with sentiment analysis (binary classification) and dimensional classification (multi-class classification). The binary classification is based on positive and negative sentiment. Positive sentiment is represented by reviews that tend to be satisfied or happy, while negative sentiment represents reviews that tend to feel dissatisfied or disappointed. Meanwhile, multi-class classification will be carried out based on the selected dimensions. In Alamsyah & Rachmadiansyah's (2018) research, the dimension used was transportation service quality. However, in this research the dimensions used are the Mobile Apps Service Quality dimensions from Wulfert (2019) to measure the quality of application services specifically for smartphones using its sub-dimensions. The best algorithm for classification will be known from the accuracy between Naïve Bayes and KNN.

According to research from Alamsyah & Rachmadiansyah (2018), service quality results based on binary classification (sentiment) and multiclass classification (dimensions) will be analyzed. Then, prioritizing problems or complaint priorities from the worst dimension will be carried out using WordCloud. That is illustrated in the framework in Figure 2.

RESEARCH METHOD

This research is a mixed method. The mixed method research aims to answer the problem formulation, not only using a qualitative or quantitative approach (Bougie & Sekaran, 2020). This research uses the Naïve Bayes and k-Nearest Neighbors algorithms, which involve statistical or quantification processes. The results are in the form of numbers that must be interpreted. On the other hand, WordCloud visualizations in the form of text are also interpreted to understand the meaning of user complaints qualitatively.

This research collected data by scraping KineMaster application review data on the Google Play Store using Google Colab. As many as 5,000 data were collected from February 1st, 2023, to November 30th, 2023. So, this research uses data sources from User Generated Content (UGC), which includes secondary data. Therefore, the population consisted of users of the Android-based KineMaster application, which was reviewed in the Google Play Store application. The samples in this research were users of the Android-based KineMaster application who submitted their reviews on the

Google Play Store over approximately ten months, specifically from February 1st, 2023, to November 30th, 2023.

The data obtained from scraping with Google Colab must be prepared, starting with data cleaning and labeling, text pre-processing, TF-IDF weighting, use of classification algorithms, and evaluation or comparison of algorithm performance. That is illustrated in figure 3.

Data obtained from scraping with Google Colab will be cleaned by deleting meaningless, ambiguous, or irrelevant data. According to Putri *et al*. (2020), cleaning is a process for deleting irrelevant or meaningless data. As a result of cleaning, the data amounted to 2,990 data. Then, the data will be divided into training and test data, with a ratio of 70 to 30. 70% will be training data, and 30% will be test data because, according to Masrury *et al*. (2019), the division of training and test data is effective for the maximum result is 70: 30. Thus, 2,093 data will be labeled binary and multiclass whereas 897 data will be non-labeled.

The data is labeled into two categories, namely binary classification, divided into positive and negative sentiment, as in the example of binary classification label data in Table 3. Meanwhile, Table 4 shows an example of labeling multiclass classification data, which is divided into seven categories according to Wulfert (2019) of Mobile Application Service Quality dimensions.

After the training and test data are collected, the text pre-processing process is carried out to prepare existing data by eliminating data that does not match the desired format. The pre-processing process carried out in this research is tokenizing, filtering, and stemming using RapidMiner Studio version 10.2. Sari *et al*. (2018) state that tokenizing divides words, phrases, or symbols. The filtering process is a process that functions to remove common words that do not have important or necessary meanings, such as *'dan', 'dari', 'untuk'*. Finally, the stemming process is carried out by changing the words into their basic words, such as *'membantu'* becomes *'bantu'*. An example of pre-processing is in table 5.

After that, make sure RapidMiner version 12.0 uses TF-IDF for weighting. TF-IDF is a weighting scheme that helps determine how far a term is associated with a document by assigning weight to each word (Damarta *et al*., 2021). Moreover, use the

Naïve Bayes and KNN (k=7) algorithm to see the performance. A confusion matrix can evaluate performance (Widyawati *et al*., 2021).

Based on the values of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) in figure 4, the performance evaluation also can be seen from accuracy, precision, and recall (Widyawati *et al*., 2021).

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Note:
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

Accuracy is the ratio of the amount of review data that is predicted correctly to the total data. Precision measures how precise the classifier is in predicting text data, and recall measures how precisely the classifier remembers groupings of text data. We can see the accuracy of comparing two algorithms because (Siddik, 2023) accuracy measures how well the classification model can classify data correctly as a whole.

Based on the evaluation and comparison results, the best algorithm or model between Naïve Bayes and KNN will be selected in this research so that the results of sentiment analysis (binary classification) and dimension classification (multi-class classification) are based on the best algorithm. Next, based on the number of negative sentiments, the negative sentiment data from the three worst MASQ dimensions will be made into a WordCloud visualization to determine the main complaints.

Based on figure 5, WordCloud will be used in this research after obtaining negative review data in the worst Mobile Apps Service Quality (MASQ) dimension to look for the words they complained about the most, which could be seen from the WordCloud visualization. The word that has the biggest size means that it has a large number of frequencies or is a word that appears frequently. Thus, from the WordCloud visualization, we can interpret the priority of user problems or complaints in the worst Mobile Apps Service Quality (MASQ) dimension.

RESULT AND DISCUSSION

Evaluating The Algorithm

Finally, the performance evaluation can be seen from accuracy, precision, and recall based on TP, TN, FP, or FN calculations. The performance results of Naive Bayes and KNN can be compared in Tables 6 and 7. As the ratio of the amount of review data that is predicted correctly to the total data, accuracy is important. This is also confirmed by Siddik (2023), who states that accuracy measures how well the classification model can classify data correctly as a whole. In terms of the accuracy of the algorithm, KNN is superior to Naïve Bayes in both binary classification and multi-class classification. The accuracy of KNN in binary classification is 87.39%, and the accuracy in multiclass classification is 70.33%, whereas the accuracy of Naïve Bayes is 82.46% in binary classification and 42.48% in multiclass classification. So, KNN is used in this research.

Sentiment Analysis (Binary Classification)

According to previous results, the KNN algorithm is used. The sentiment results are shown in Figure 6. The results show that the sentiment dominated by user reviews of the Android-based KineMaster application is 62.24% positive sentiment, or 1861 positive reviews, compared to negative sentiment, only 37.76% or 1129 negative reviews. This means that most Android-based KineMaster application users are pretty satisfied with the services provided by the company. This can be seen from the percentage of positive sentiment from reviews above 50% and negative sentiment below 50%.

The dominance of positive sentiment does not mean companies can ignore negative sentiment or existing complaints. Evaluating complaints (negative sentiment) must be done to improve the performance of applications, even companies. This means that in the future, the percentage of negative sentiment can be suppressed (reduced) while the percentage of positive sentiment can be increased. In this way, further analysis was carried out regarding the quality of the KineMaster application service based on the Mobile Apps Service Quality (MASQ) dimension to help the developer make decisions regarding increasing mobile app dimensions.

Submitted : 13/04/2024 |Accepted : 12/05/2024 |Published : 17/07/2024
P-ISSN; 2541-5255 E-ISSN: 2621-5306 |

Dimension Classification (Multi-Class Classification)

Based on figure 7 of the multiclass classification result, the valence dimension outperforms positive sentiment with 1429 data or 76.79% of the positive sentiment data. Even though the valence dimension is the highest in positive sentiment, this dimension is also still the second worst category based on negative sentiment. If we compare all the data, the number of positive valence dimensions is 1429, and the number of negative ones is only 177. So, it can still be said that most users still give an excellent final impression like *"puas"* or *"bagus"*. However, some feel unsatisfied with the KineMaster application, such as "jelek" or "mahal," which is indicated by the second worst negative sentiment results compared to other dimensions of negative sentiment. Thus, the valence dimension holds the key to overall satisfaction or dissatisfaction regarding service quality, which is also important in improving service quality. According to Wulfert (2019), the valence dimension is the final impression of satisfaction from the user after completing the service.

The technical reliability dimension is the second most frequently discussed after the valence dimension. The technical reliability dimension concerns the consistency and accuracy of mobile application operations, like mobile app and feature reliability, availability of m-services, and continuous operation after updating (Wulfert, 2019). Unlike the valence dimension, the technical reliability dimension is dominated by negative sentiment, which means that many users complain about the technical reliability of the Android-based KineMaster application, such as *"update", "fitur",* or *"layanan tidak tersedia".* The number of negative sentiments is 741 negative reviews, while the positive reviews amounted to 171 data from all reviews. Seeing that the number of negative sentiments is greater, further management is needed to identify deficiencies and ensure reliability in editing updated videos, which becomes an advantage.

Like the technical reliability dimension, the performance dimension is also dominated by negative sentiment. This means that many people still complain about application performance, either regarding processing speed or related to the device. The keywords are *"lama", "ngelag, "macet", "error", "cepat", "penggunan kuota", "penggunaan data",* and *"versi Android"*. There are 145 negative sentiments in the performance dimension, while only 28 gave positive reviews. Therefore, this dimension must be considered to make it easier for users and save costs and time. That way, the user will feel comfortable.

In contrast to the technical dimensions of reliability and performance, the design dimension is dominated by positive reviews. This means that the application is attractive or easy to use. Generally, the keywords can use *"ukuran teks bagus"*, *"ukuran logo pas"*, *"mudah digunakan"* or *"mudah dipahami"*. The number of positive reviews collected 225 data, and negative reviews amounted to 60. The attractive appearance has succeeded in achieving user satisfaction, so it has become an advantage in Kinemaster. The advantages of a product or service are important in marketing management so that KineMaster becomes more famous and can create value for customers.

The main dimensions of interaction quality, which consist of responsiveness, information, security, and privacy, are the dimensions that are rarely discussed. In the overall data, only two reviews of the responsiveness dimension are positive. This means that the application service provider's ability to move quickly and politely in resolving user problems is good. So, the keyword is *"respon cepat"*. Fast response from management is also key to the quality of the service. If customer complaints are not responded to quickly and appropriately, then dissatisfaction will arise from the customer's side. Luckily, there are two positive sentiments and no negative sentiments regarding the responsiveness of KineMaster's management.

Likewise, the information dimension is dominated by positive sentiment, totaling four reviews, and negative sentiment, only one review. This means that the company or application developer has provided sufficient and useful information for users, such as *"informasi lengkap"*. However, the security and privacy dimension is still dominated by negative sentiment, which totals five reviews, and positive sentiment, which totals two reviews. That way, system protection, and personal data still need to be improved so that users feel safe using the KineMaster app in the short and long term.

In short, based on the number of negative sentiments, the technical reliability dimension is the worst because it most frequently complained about 741 reviews or 65.63% of the negative sentiment data. This is followed by the valence dimension with 177 reviews or 15.68%, then the performance dimension with 145 reviews or 12.84% of the negative sentiment data. The three worst dimensions are analyzed further using

WordCloud so that the management or developer can find out the most complaints because negative user impressions can help improve services.

Complaint Priorities

Based on the negative sentiment, the technical reliability dimension is the worst. Figure 8 is a word cloud visualization of the technical reliability. The results showed that the most complained about was the application's inability to update the application or the application's inability to operate after updating. It can be seen from the keyword “update”. This is in accordance with Bustami and Noviaristanti (2022), which states that it is difficult to operate the application after the update. This means that “update” is the keyword for most complaints in both this research and previous research conducted by Bustami and Noviaristanti (2022). Not only is there an update problem, but it is also not uncommon for people to complain about the watermark. The keyword is “watermark”. Technically, many users ask to remove the feature of the watermark because users cannot remove it themselves unless they subscribe. Unfortunately, users also have problems with subscribing. The existence of the watermark feature is also in accordance with the theory of Wulfert (2019) regarding feature reliability, so this appears to be complained about. Apart from that, users also cannot download applications or features. Some even complain that they cannot open the applications or features that have been prepared, such as *"gak bisa buka layar kanan"*. Another thing that users have complained about that has caused disappointment is the application's inability to export the results of their video edits. This aligns with the theory of Wulfert (2019) that the availability of m-service is also a point in the technical reliability dimension.

The three main aspects related to "update", "feature reliability", and "availability of m-services" are key for improving KineMaster management so that user perception tends more towards a positive impression. Users feel satisfied because there are no technical interruptions when using the application, and the Android-based KineMaster application is still operating consistently.

The valence dimension is the final impression of the user. Even though the valence dimension is the highest dimension of positive sentiment, there are still many complaints about 177 negative reviews, so valence is the second worst dimension after technical reliability. More about the words that appear most often can be seen in Figure

9, which is the result of WordCloud visualization of the valence dimension of negative sentiment.

Based on the word cloud visualization of the valence dimension in Figure 9, it can be seen that at least the words "aplikasi," "kinemaster," "jelek," "mahal," "premium," or *"sampah"* appear. Thus, there are still users who are not quite satisfied with the KineMaster application because they feel that the KineMaster application is still bad and has a lot of rubbish. The valence dimension describes customer feelings after experiencing m-services (Wulfert, 2019), both good and bad. This shows emotional benefits. That way, *"jelek"* is a bad word that shows the bad feelings of users after using the KineMaster application. Moreover, the words *"mahal"* and "premium" still appear, which are also points in the visualization. This means that users feel that the premium price of the KineMaster application is very expensive.

This final impression is also important to pay attention to because the valence dimension shows dissatisfaction. In this way, the overall Kinemaster service and its prices are being reconsidered by the Kinemaster company to increase the number of users without harming the company. It means that both the user and the company benefit, namely satisfaction felt by the user and consumer loyalty also gained by the company.

The third worst dimension is the performance dimension based on the number of negative sentiments, namely 145 negative data. The negative sentiment data of the performance dimension can be visualized using WordCloud, as shown in Figure 10. Based on Figure 10 regarding WordCloud performance dimensions of negative data, the words "aplikasi," "kinemaster," "lama," "lemon," "lag," "must," and *"error"* appear. In that way, the complaints given are that they often experience problems with the speed of the process because it "takes a long time," "lags," "slow," and even the process "errors."

Those results align with the findings of Bustami and Noviaristanti (2022) that the mobile app still has frequent errors or long processes in the performance dimension. Because of that, processing speed in operating mobile apps, especially in video editing in KineMaster, must be managed again by the developer so that users save more time using the KineMaster application. Likewise with mobile games, according to Arifin *et al*. (2023) study, a mobile game requires effectiveness and efficiency in use. Surely, a mobile application with a fast process will be more effective and efficient in its use. Therefore, it is easier and more convenient to use a mobile app.

CONCLUSIONS AND RECOMMENDATIONS

Based on the accuracy results, the k-Nearest Neighbors (KNN) algorithm is superior to Naïve Bayes in both binary classification and multi-class classification, so the sentiment results show that positive sentiment is 62.24% superior compared to negative sentiment, which is only 37.76%. However, that doesn't mean ignoring the negative sentiment. Complaints related to negative sentiment must still be prioritized to resolve the problem and be able to compete with other competitors.

The three worst dimensions are the dimensions of technical reliability, valence, and performance. In this way, the priority problems or complaints must be resolved immediately, namely updates, watermarks, downloading applications and features, inability to open applications, and export capabilities in the technical reliability dimension. Moreover, KineMaster is too expensive. The price is a complaint of valence dimension. Lastly, the complaint of processing speed in the performance dimension. Therefore, the complaints prioritized that must be resolved immediately by the company or developer are updates, watermarks, downloading applications and features, inability to open the applications, export capabilities, high price, and processing speed.

Improving services by paying attention to negative user impressions is one way for KineMaster developers to reduce the number of complaints regarding its application services. Eventually, positive impressions can be increased, considering that positive impressions influence purchasing decisions for a product or service and improve the brand image of the KineMaster application. Thus, the results of this research could be an opportunity for mobile app developers to increase or improve the quality of services to increase the number of users, ratings, and even loyal consumers.

This research has limitations. A limitation of this research is that there are only two algorithm comparisons. Reviews are provided only through the Google Play Store's user-generated content (UGC). User-generated content is not only available on the Google Play Store. That way, for further research, it is hoped that we will be able to compare the Naïve Bayes, KNN, Decision Tree, Random Forest, and Support Vector Machine algorithms or even deep learning. Apart from that, data sources such as

Instagram comments or other forums are also added so that the service quality evaluation results are more representative.

REFERENCES

Abdurochman, A. F., & Tantra, T. (2023). Pengaruh Airlines Service Quality dan Brand Image terhadap Customer Loyalty Penumpang Maskapai Lcc. *Jurnal Ilmiah MEA (Manajemen, Ekonomi, Dan Akuntansi)*, *7*(2).

Aditya, I. A., Haryadi, F. N., Haryani, I., Rachmawati, I., Ramadhani, D. P., Tantra, T., & Alamsyah, A. (2023). Understanding service quality concerns from public discourse in Indonesia state electric company. *Heliyon*, *9*(8). https://doi.org/10.1016/j.heliyon.2023.e18768

Agiesta, W., Sajidin, A., & Perwito. (2021). Pengaruh Kualitas Pelayanan dan Kepuasan Pelanggan terhadap Loyalitas Pelanggan KA Lokal Bandung Raya. *Jurnal Ilmiah MEA (Manajemen, Ekonomi, Dan Akuntansi)*, *5*(2).

Alamsyah, A., & Rachmadiansyah, I. (2018). Mapping online transportation service quality and multiclass classification problem-solving priorities. *Journal of Physics: Conference Series*, *971*(1). https://doi.org/10.1088/1742-6596/971/1/012021

Arifin, M. S., Ariyanti, M., & Nurhazizah, E. (2023). Economics and Digital Business Review Analisis Kualitas Mobile Games Berdasarkan Ulasan Platform Google Play Di Indonesia Menggunakan Metode Text Mining. *Economics and Digital Business Review*, *4*(1), 357–368.

Ariyanti, M., & Rifaldi, R. (2019). Monitoring social media with Social Network Analysis Method and Text Network Analysis as Business Intelligence. *Test Engineering & Management*, *81*, 2780–2786.

Arusada, M. D. N., Putri, N. A. S., & Alamsyah, A. (2017, October 18). Training data optimization strategy for multiclass text classification. *2017 5th International Conference on Information and Communication Technology, ICoIC7 2017*. https://doi.org/10.1109/ICoICT.2017.8074652

Bafadhal, A. S. (2022). *Manajemen Komplain Dan Kualitas Layanan Pariwisata*. Yogyakarta: Deepublish Publisher.

Bougie, R., & Sekaran, U. (2020). *Research Methods for Business: A Skill-Building Approach*. Hoboken: John Wiley & Sons, Inc.

Bustami, D. K., & Noviaristanti, S. (2022). Service Quality Analysis of Tokopedia Application Using Text Mining Method. *International Journal of Management, Finance and Accounting*, *3*(1), 1–21. https://doi.org/10.33093/ijomfa.2022.3.1.1

cnnindonesia.com. (2022, January 12). 7 Rekomendasi Aplikasi Edit Video di HP Android Terbaik. *Cnnindonesia.Com*. https://www.cnnindonesia.com/teknologi/20220110144140-190-744744/7-rekomendasi-aplikasi-edit-video-di-hp-android-terbaik

dailysocial.id. (2022). *7 Rekomendasi Aplikasi Edit Video Terbaik untuk Android*. https://dailysocial.id/post/7-rekomendasi-aplikasi-edit-video-terbaik-untuk-android

Damarta, R., Hidayat, A., & Abdullah, A. S. (2021). The application of k-nearest neighbors classifier for sentiment analysis of PT PLN (Persero) Twitter account service quality. *Journal of Physics: Conference Series*, *1722*(1). https://doi.org/10.1088/1742-6596/1722/1/012002

Hafidz, G. P., & Muslimah, R. U. (2023). Pengaruh Kualitas Layanan, Citra Merek, Kepercayaan Pelanggan dan Kepuasan Pelanggan terhadap Loyalitas Pelanggan

Produk Herbalife. *Jurnal Ilmiah MEA (Manajemen, Ekonomi, Dan Akuntansi)*, *7*(No.1).

Harahap, R. M., Kusumahadi, K., & Nurhazizah, E. (2022). Analysing Service Quality of Mobile Health Platforms Using Text Analytics: A Case Study of Halodoc and Alodokter. *Asian Journal of Research in Business and Management*, 168–182. https://doi.org/10.55057/ajrbm.2022.4.1.14

idntimes.com. (2023). *12 Aplikasi Edit Video Terbaik di Android Terbaru 2023*. https://www.idntimes.com/tech/gadget/amp/patricia-firscha/aplikasi-edit-video-terbaik-android-gratis?page=all#page-2

katadata.co.id. (2022). *Potensi Industri Konten Kreator Indonesia*. https://katadata.co.id/amp/desysetyowati/digital/626a3444da848/potensi-industri-konten-kreator-indonesia-ditaksir-senilai-rp7-triliun

Limbong, J. J. A., Sembiring, I., & Hartomo, K. D. (2022). Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes Dan K-Nearest Neighbor Analysis of Review Sentiment Classification On E-Commerce Shopee Word Cloud Based With Naïve Bayes And K-Nearest Neighbor Methods. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, *9*(2), 347–356. https://doi.org/10.25126/jtiik.202294960

Mahr, Dominik, Susan Stead and Gaby Odekerken-Schröder, (2019), Making sense of customer service experiences: a text mining review. Journal of Services Marketing. https://doi.org/10.1108/jsm-10-2018-0295

Mao, Luke Lunhua (2020). Understanding retail quality of sporting goods stores: a text mining approach. *International Journal of Sports Marketing and Sponsorship* https://doi.org/10.1108/ijsms-03-2020-0029

Masrury, R. A., Fannisa, & Alamsyah, A. (2019, July 1). Analyzing tourism mobile applications' perceived quality using sentiment analysis and topic modeling. *2019 7th International Conference on Information and Communication Technology, ICoICT 2019*. https://doi.org/10.1109/ICoICT.2019.8835255

Ordenes, Francisco Villarroel and Shunyuan Zhang (2019). From words to pixels: text and image mining methods for service research. Journal of Service Management. https://doi.org/10.1108/josm-08-2019-0254

Pratmanto, D., Rousyati, R., Wati, F. F., Widodo, A. E., Suleman, S., & Wijianto, R. (2020). App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm. *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012043

Putri, N. N. S., Alamsyah, A., & Widiyanesti, S. (2020, June 1). Fulfillment and Responsiveness on Online Travel Agencies Using Multiclass Classification. *2020 8th International Conference on Information and Communication Technology, ICoICT 2020*. https://doi.org/10.1109/ICoICT49345.2020.9166457

Sari, P. K., Alamsyah, A., & Wibowo, S. (2018). Measuring e-Commerce service quality from online customer reviews using sentiment analysis. *Journal of Physics: Conference Series*, *971*(1). https://doi.org/10.1088/1742-6596/971/1/012053

Siddik, A. Muh. A. (2023). Comparison of Transfer Learning Algorithm Performance in Hand Sign Language Digits Image Classification. *Jurnal Matematika, Statistika Dan Komputasi*, *20*(1), 75–89. https://doi.org/10.20956/j.v20i1.26503

Valdivia, A., Luzón, M. V., Cambria, E., & Herrera, F. (2018). Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Information Fusion*, *44*, 126–135. https://doi.org/10.1016/j.inffus.2018.03.00

Wearesocial.com. (2023). *Digital 2023 Indonesia*.

Widyawati, R., Irawan, H., & Ghina, A. (2021, May 19). Content Analysis of Tourist Opinion based on Tourism Quality (TOURQUAL) by Text Mining Online Reviews: The Case of Borobudur. *International Conference on Sustainable Management and Innovation*. https://doi.org/10.4108/eai.14-9-2020.2304462

Wulfert, T. (2019). Mobile App Service Quality Dimensions and Requirements for Mobile Shopping Companion Apps Junior Management Science. *Junior Management Science*, *4*(3), 339–391. https://doi.org/10.5282/jums/v4i3pp339-391

Xin Wang, Jiming Tian and Fei L. (2022). Text data mining of power based on natural language processing technology. Journal of Physics: Conference Series, Volume 2221, 2022 2nd International Conference on Electronics, Circuits and Information Engineering (ECIE-2022) 07/01/2022 - 09/01/2022 Online. https://doi.org/10.1088/1742-6596/2221/1/012050

Yahya, R. H., Maharani, W., & Wijaya, R. (2023). Disaster Management Sentiment Analysis Using the Bilstm Method. *Jurnal Media Informatika Budidarma*, *7*(No.1), 501–508. https://doi.org/10.30865/mib.v7i1.5573

Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2018). *Services Marketing: Integrating Customer Focus Across the Firm* (7th ed.). New York: McGraw-Hill Education

Zuliestiana, D. A., & Setiawan, A. N. (2022). Pengaruh E-Service Quality terhadap E-Customer Satisfaction dan Dampaknya terhadap E-Customer Loyalty Pada Pengguna Aplikasi BCA Mobile. *Jurnal Ilmiah MEA (Manajemen, Ekonomi, Dan Akuntansi)*, *6*(2).

FIGURES, GRAPHICS, AND TABLES

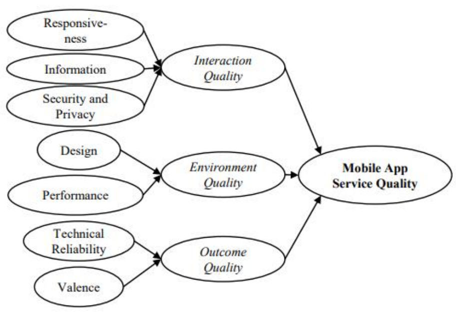

Figure 1. Mobile Apps Service Quality Dimensions
Source: Wulfert (2019)

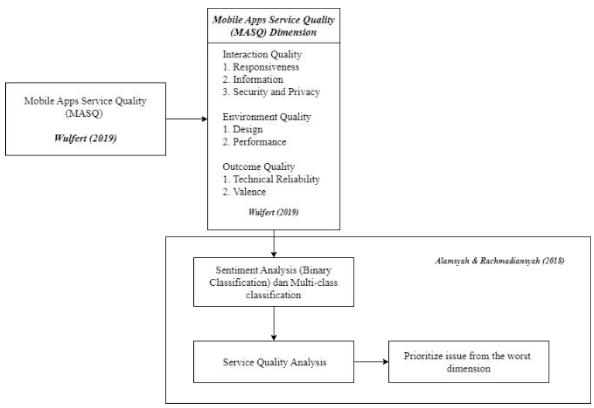

Figure 2. Framework
Source: Wulfert (2019), Alamsyah & Rachmadiansyah (2018)

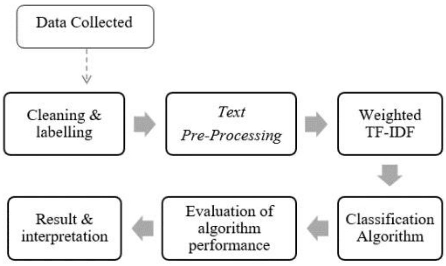

Figure 3. Binary and Multiclass Classification Data Analysis

		Predicted Label	
		Positive	Negative
Actual Label	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 4. Confusion Matrix
Source: Widyawati, *et al* (2021)

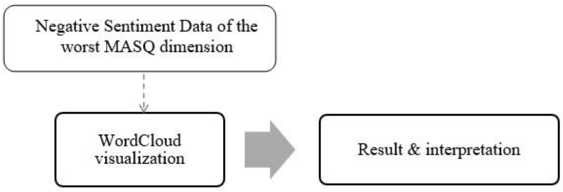

Figure 5. WordCloud

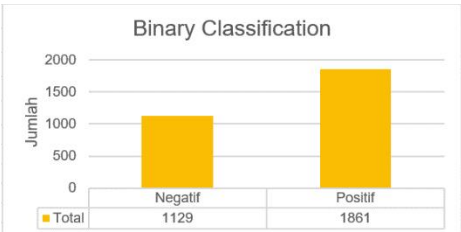

Figure 6. Result of Sentiment Analysis (Binary Classification)

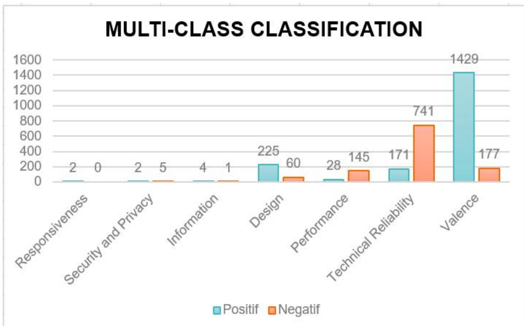

Figure 7. Result of Multiclass Classification

Figure 8. WordCloud Technical Reliability Dimension

Figure 9. WordCloud Valence Dimension

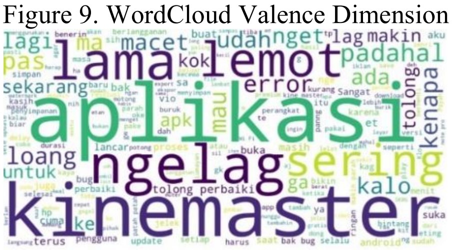

Figure 10. WordCloud Performance Dimension

Table 1. Comparison Video Editing Apps in Google Play Store (2023)

Application	Release	Number of Downloaders	Rating
CapCut	2020	500.000.000+	4.5
Inshot	2014	500.000.000+	4.9
VN	2018	100.000.000+	4.7
KineMaster	2013	100.000.000+	4.5

Table 2. Mobile Apps Service Quality Dimensions of Wulfert (2019)

Dimension	Description
Interaction Quality	It reflects on the quality of interactions between customers and mobile application service providers.
Responsiveness	The application service provider's ability to move quickly and courteously to resolve user problems.
Information	It is concerned with providing adequate and useful information for users.
Security and Privacy	It is concerned with protecting systems, user data, and network resources from attacks.
Environment quality	It reflects how the application provides its mobile application services to satisfy users.
Design	It is related to the application's visual aesthetics and ease of use.
Performance	Mobile application performance is related to processing speed, device storage usage, mobile network usage, and device or connection quality.
Outcome quality	It reflects the technical quality of application service delivery and user satisfaction with mobile services.
Technical Reliability	It is concerned with the consistency and accuracy of mobile application operations.
Valence	The final impression of the user after completion of service delivery.

Table 3. Sentiment Analysis Labeling Example (Binary Classification)

Text	Sentiment
Bagus dan cepat untuk mengedit	Positive
Semenjak update aplikasinya, buat ngedit video jadi patah patah Mulu. Mending yang versi lamanya lebih bagus dari yang sekarang. Kecewa berat	Negative

Table 4. Multiclass Classification Labeling Example

Text	Dimension
Interaction Quality	
Developer respon cepat.. Sangat membantu..	Responsiveness
Informatif	Information
Baru juga di install belom di buka aplikasinya udah di suruh update data aneh	Security and Privacy
Environment Quality	
Mudah dalam pemakaian	Design
Bagus dan cepat untuk mengedit	Performance
Outcome Quality	
Semenjak update aplikasinya, buat ngedit video jadi patah patah Mulu. Mending yang versi lamanya lebih bagus dari yang sekarang. Kecewa berat	Technical Reliability
Nelong mah sih app nya harusnya bintang 0 tp gk ada pilihan !!	Valence

Table 5. Pre-Processing Example

Raw Data	*aplikasi edit video ini sangat bermanfaat dan cepat di ingat caranya*
Tokenizing	*aplikasi, edit, video, ini, sangat, bermanfaat, dan, cepat, di, ingat, caranya*
Filtering	*aplikasi, edit, video, bermanfaat, cepat, caranya*

Table 6. Comparison Algorithm Performance
Binary Classification

Algorithm	Accuracy	Precision	Recall
Naïve Bayes	82.46%	84.63%	86.32%
KNN	87.39%	84.67%	96.32%

Table 7. Comparison Accuracy Performance
Multiclass Classification

Algorithm	Accuracy
Naïve Bayes	42.38%
KNN	70.33%