Applied Information System and Management (AISM)
Volume 6, (1) 2023, p. 13-20
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v6i1.26045

The Comparison of Sentiment Analysis of Moon Knight Movie Reviews between Multinomial Naive Bayes and Support Vector Machine

Abdul Azzam Ajhari

***Abstract*— Online movie streaming platforms have changed the current pattern of watching movies. Besides providing convenience in watching anywhere and anytime, this service is provided at a lower cost to moviegoers. The increase in moviegoers on online streaming platforms has resulted in easy-to-find reviews. This review can determine whether they decide to watch the film or not. The moviegoers' reviews can be easily and quickly found for analysis using sentiment analysis to find a film's worthiness. This study used sentiment analysis in classifying Twitter data predictions using the Multinomial Naive Bayes (MNB) and Support Vector Machine (SVM). In the sentiment analysis of labeling with positive and negative categories, a distilled version of BERT (DistilBERT) was used in this study. With a little human assistance in preprocessing, the model worked objectively with an overall accuracy performance on the confusion matrix of 64.50% for the Multinomial Naive Bayes model and 64.12% for the Support Vector Machine model. Performance evaluation was also carried out by calculating the cross-validation accuracy, which resulted in an accuracy of 72.38% for the MNB. Meanwhile, the SVM model obtained an accuracy of 70.19%.**

***Index Terms*—Sentiment analysis, movie reviews, multinomial naïve bayes, support vector machine, distilBERT.**

I. Introduction

The rapid development of technology in the film industry has changed people's watching habits, from physically watching movies in theaters to online streaming platforms. According to a Harris survey in the Wall Street Journal in 2019, the average American uses three to six streaming services [1]. One of the popular online movie streaming services or platforms is Disney+ (Plus) Hotstar. Based on statista.com data,

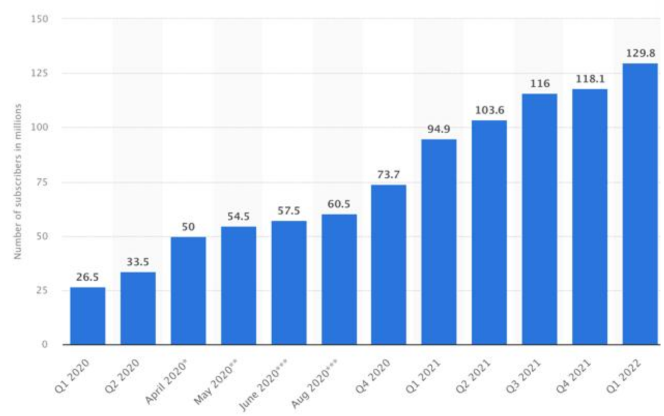

Fig. 1. Disney+ Subscriber Growth Statistics 1st quarter 2020 to 1st quarter 2022 (in millions) [2]

there was a significant increase in subscribers from the first quarter of 2020 to the first quarter of 2022, as shown in Fig. 1.

With the rapid increase in subscribers who watch movies on this platform, it is necessary to balance the quality of the films provided by film production houses. This can be exemplified by the latest series, "Moon Knight," which has just been released by the Marvel Cinematic Universe production house on March 30, 2022, on the Disney+ (plus) Hotstar streaming service. The large number of Marvel fans who watch and review this series can affect the feasibility of the Moon Knight series. By analyzing moviegoers' opinions, this film's eligibility can be determined. It then can impact the budget savings incurred and the increase in revenue received to determine whether this film can be continued or not.

One of the most widely used methods to get the opinion of moviegoers is sentiment analysis. It is defined as a process of analyzing opinions on datasets to determine whether they are positive or negative [3]. Sentiment analysis can be used on the electability of political figures [4], such as presidential candidates [5], customer satisfaction with products [6-7], and film reviews [8-12]. This research used two types of machine learning options, called Multinomial Naïve Bayes (MNB) and Support Vector Machine (SVM), to determine the feasibility of Moon Knight episodes 1-4. The data were collected from

Received: 15 May 2022; Revised: 15 June 2022; Accepted: 5 July 2022

A. A. Ajhari, Pusat Pengkajian dan Pengembangan Teknologi Keamanan Siber dan Sandi, Badan Siber dan Sandi Negara, Jakarta, Indonesia (e-mail: abdul.azzam@bssn.go.id).

people's posts on Twitter because they have been widely used in sentiment modeling in several studies.

II. Related Work

Sentiment analysis studies are divided into three categories: sentiment analysis of film reviews, sentiment analysis of film comments, and sentiment analysis of film ratings.

A. Sentiment Analysis of Film Reviews

Film review text preprocessing, construction of domain sentiment dictionary, negative word dictionary, and degree adverb dictionary were four sentiment dictionaries proposed in this study [13] to improve sentiment classification accuracy in reviews.

Meanwhile, the optimization of the CNN and Bi-LSTM models was proposed in another study [14] to analyze the sentiment of film reviews with the addition of an attention mechanism to become an effective model.

Three different calculations, namely, Naïve Bayes (NB) and Support Vector Machine (SVM) using an n-gram approach such as unigram, trigram, bigram, bigram+trigram, unigram+bigram, and unigram+bigram+trigram were proposed in a study [15] to analyze sentiment on film reviews on the Rotten Tomatoes platform.

B. Sentiment Analysis of Film Comments

Several studies have been carried out to analyze film comments. For instance, research [16] suggested that comments in the form of text on films can be analyzed by sentiment analysis using the Bi-LSTM model to attain the best accuracy performance.

In addition, other research [17] analyzed comments and opinions on the Money Heist season 4 trailer on Youtube using sentiment analysis with the Multinomial Naïve Bayes model. The best accuracy reached 81%. Finally, research [18] conducted sentiment analysis on Myanmar film comments using the Naïve Bayes classifier with the best accuracy of 83.60%.

C. Sentiment Analysis of Film Ratings

The success of the box office films on a movie streaming device can be predicted [19] using text mining and sentiment analysis with the SVM model, producing a classification accuracy of 81.69%. Additionally, other research [20] outlined the method of disclosing user attitudes on certain dimensions of the film to obtain a rating review expressed by moviegoers.

Based on previous research on the sentiment analysis of film reviews, comments, and ratings, it can be concluded that the most widely used machine learning models are Multinomial Naïve Bayes (MNB) and Support Vector Machines (SVM). Therefore, the present research decided to use both models. In addition, little information could be found from previous studies about how to process the labels assigned to each dataset. Thus, this research also used the distilled version of the BERT (DistilBERT) model to label datasets based on sentiment polarity.

D. Multinomial Naïve Bayes

Multinomial Naïve Bayes classifier is one of the most popular algorithms used for text mining due to its convenience [21]. Besides, it has a fast processing time, an easy implementation with a reasonably simple structure, and an excessive level of effectiveness [22]. MNB calculates the probability of a class primarily based on its attributes and determines the class that has the highest possibility. It classifies classes primarily based on simple possibilities by assuming that each attribute in the data is mutually unique. Within the opportunity model, each k class and the quantity of attributes can be written as within (1).

$$P = (y_1|x_1, x_2, ..., x_n) \tag{1}$$

The appearance of feature X_a data in the $Y_kP(x_a|y_k)$ class category is the probability from the MNB calculation, which is multiplied by the probability of the $P(y_k)$ class category. The distribution of feature data $P(x_a)$ will occur from the results of previous calculations. Hence, the new calculation is determined in (2).

$$P(y_k|x_a) = \frac{P(y_k)P(x_a|y_k)}{P(x_a)} \tag{2}$$

Then, the highest probability value is chosen from each opportunity class to determine the optimal class. In (3) is the formula for choosing the highest value.

$$y(x_i) = \arg\max P(y) \prod_{i=1}^{a} P(x_1|y) \tag{3}$$

The classification accuracy not only depends on probability but also can use weights on each class [23]. In this way, the attribute can increase the predictive effect.

E. Support Vector Machine

According to [24], the support vector machine (SVM) has a solution that the maximum margin classifier or hyper-plane concept can overcome the prediction classification problems that occur in other linear classifiers. In classifying the correct prediction of the SVM method, the following mathematical calculations are known.

Solving the enlargement and subtraction of two distances can be done by maximizing the margin (4) and minimizing (5) equations to obtain the sum of the distances from the separating hyper-plane to the closest point.

$$Margin = \frac{2}{\overrightarrow{||w||}^2} \tag{4}$$

$$L(w) = \frac{\overrightarrow{||w||}^2}{2} \tag{5}$$

The emergence of finite optimization problems can be done by using an approach using numerical calculations in (6) below.

Applied Information System and Management (AISM)
Volume 6, (1) 2023, p. 13-20
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v6i1.26045

$$f(\vec{x}_i) = \begin{cases} 1 & \text{if } \vec{w} \bullet \vec{x}_i + b \geq 1 \\ -1 & \text{if } \vec{w} \bullet \vec{x}_i + b \leq -1 \end{cases} \quad (6)$$

III. Research Method

A. Data Collection

This study uses a tool to crawl data from Twitter in collecting data.

Table 1 describes the data crawled using Twint with the keyword "moon knight". The language filter is "en" or English on Twitter since the launch of the moon knight series on March 30, 2022.

Table 1.
Dataset Crawling Parameters

No	Twint Param. Name	Data Type	Description
1.	Search	String	Search terms
2.	Lang	String	Language used on Twitter
3.	Since	String	Filtering tweets posted data base on since date

Table 2.
Dataset Explanation

No	Twint Feature Name	Data Type	Description
1.	date	string	Tweets posted date
2.	username	string	The name of the account that posted the tweet
3.	tweet	string	Tweet opinion post
4.	language	string	Language used
5.	nretweets	integer	Number of retweets on tweet posts
6.	nlikes	integer	Number of likes on tweet posts

As presented in Fig. 2, the total data obtained from Twitter on April 18, 2020, until April 20, 2022, was 40,008 data. The data contained six attributes or features, namely date, username, tweet, language, nretweets, and nlikes, as explained in Table 2.

Unfortunately, even though the language has been filtered to English, the dataset still had data in a different language (not English). Then the data was saved into a comma-separated values (csv) format document.

B. Research Method

This section describes the research stages which are divided into 3 parts, namely preprocessing, data separation, and model validation.

	date	username	tweet	language	nretweets	nlikes
0	2022-04-20 16:18:25	UncoolTimi	So I'm watching #Moonknight…did I miss somethi...	en	0	0
1	2022-04-20 16:18:24	UselessPure	Spoilers for moon Knight. https://t.co/82xcbs...	en	0	0
2	2022-04-20 16:18:24	DempseyPillot	Who is Taweret? Marvel's Hippopotamus Goddess ...	en	0	0
3	2022-04-20 16:18:22	fxlmmm24	#MoonKnight E4 รีวิวด้วย Emoji แบบใบ้ๆ - 👩❤️...	th	0	0
4	2022-04-20 16:18:21	lizzieshrooms	#MoonKnight spoilers!! - - - - - - THE KISS W...	en	0	0
...	...	...	...	...	...	...
40003	2022-04-18 06:02:00	JebacRohana	cant wait na me and the boys (ja i marika) wsp...	en	0	0
40004	2022-04-18 06:01:57	scifichick25	Wonder if Moon Knight is indulging in matzo br...	en	1	5
40005	2022-04-18 06:01:52	MizoAsc	@rihabd ا بتفكريني بصفحة لبنانية متعصبة لمسلسل...	ar	0	0
40006	2022-04-18 06:01:46	niadacosta	e agora vou começar moonknight https://t.co/0...	pt	0	8
40007	2022-04-18 06:01:35	JackLaridian	@FastTalkKing @driiftyfilm What like comic boo...	en	0	1

40008 rows × 6 columns

Fig. 2. Total Data Tweets Moon Knight on Twitter

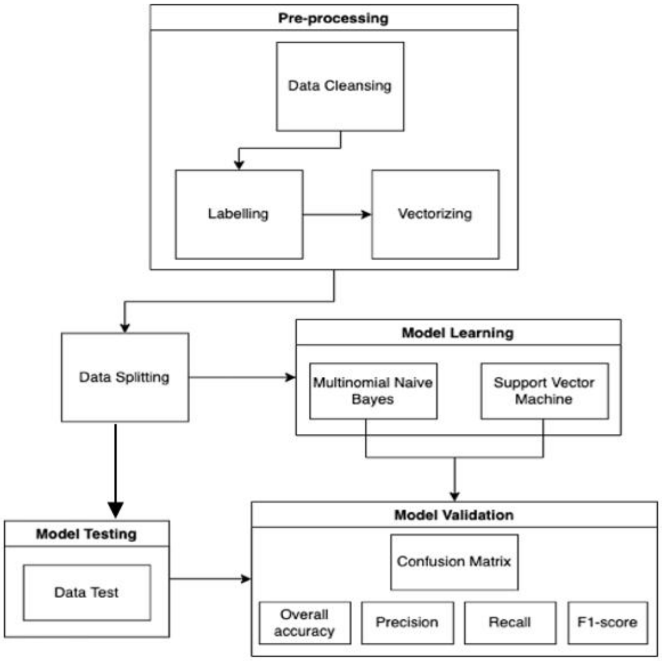

Fig. 3. Framework Sentiment Analysis

Sentiment analysis framework [25] is used as a research methodology which can be seen in Fig. 3.

- Preprocessing

There are three proposed processes, namely data cleansing, labeling, and vectorizing in the preprocessing stage.

1) Data cleansing

- The data were filtered based on the nretweets and nlikes features with a parameter limit of more than zero. Thus, it can be said that the tweets are in accordance with what other users feel. The function then generated the final 4,592 data that were compatible with most users.
- During this process, data redundancy might occur, that is, the repetition of the same data or data set in a dataset. To reduce the level of data redundancy, it is necessary to remove duplicate data in the tweet feature. The final data from the function was 4,584 unique data. These data did not have data duplication.
- Not all the data is clean, there are still symbols, use of emoticons, and non-standard words. First, cleaning the regex data to remove text that is mixed with symbols and emoticons. Then, the text is converted to lowercase. After that, the text in the form of sentences is divided into several words using tokenizes. Then each word is filtered with English stopwords, so that data that is still mixed is removed from the dataset. Finally, the text data for each appropriate word is searched for the basic word using the stemming function.
- Furthermore, not all the data is clean. The data still contained symbols, emoticons, and non-standard words. First, these data were processed by cleaning the regex data to remove texts that were mixed with symbols and emoticons. Then, the texts were converted to lowercase. Since the texts were still in sentence form, they should be divided into several words using tokenizes. Then, each word was filtered with English stopwords to get the mixed data removed from the dataset. Finally, the text data for each appropriate word was searched for the basic word using the stemming function.
- The tokenize and stemming function technique is a feature extraction which modifies the tweets feature into a new feature called text_clean and text_preprocessing.
- In the text_preprocessed feature, the data were still in the form of snippets of basic words. The data were combined into one whole sentence and replaced the text_clean feature. Then, the text_preprocessed feature was removed.
- Total data from the data cleansing process obtained 3,088 clean data that does not have a label with six features: date, username, language, nretweets, nlikes, and text_clean.

2) Labeling

In this study, data that did not have a label was determined based on the tweet's sentiment using the distilled version of BERT (DistilBERT) method [26] . DistilBERT is a model trained with the same data as BERT, namely BookCorpus. The model is taught from 11,038 collections of un-published books and the English Wikipedia (excluding lists, tables, and headers).

DistilBERT is a model for text classification with the same architecture used in BERT. The token and pooler type embeds were removed while the number of layers was reduced by a factor of 2. Most of the operations used in the Transformer architecture (linear layer and layer normalization) were highly

	date	username	language	nretweets	nlikes	text_clean	label	score
0	2022-04-20 16:17:18	LENNIEC0N	en	4	10	end ep	NEGATIVE	0.980303
4	2022-04-20 16:17:13	Kraven_sD	en	1	1	mr knight moon knight one favourite shows mome...	POSITIVE	0.998832
5	2022-04-20 16:17:12	getFANDOM	en	3	26	episode number us	POSITIVE	0.995056
6	2022-04-20 16:17:09	TeresaFortesLuz	en	1	1	spoilers hug everything ill thinking wholesome	NEGATIVE	0.632042
7	2022-04-20 16:16:28	hexxtechs	en	1	2	take moment appreciate may el calamawys acting...	POSITIVE	0.979506
...	...	...	...	...	...	...	...	...
3529	2022-04-18 06:05:41	ryn5hnd	en	1	4	phase purple	NEGATIVE	0.989150
3530	2022-04-18 06:04:07	TheReelTalkLive	en	5	8	going live mins talk dc overhaul announcement ...	NEGATIVE	0.988737
3531	2022-04-18 06:03:55	MoFromStreamr	en	1	18	feel sick	NEGATIVE	0.999687
3532	2022-04-18 06:03:10	mrjafri	en	7	16	people say superheroes need watch moon knight	POSITIVE	0.627477
3533	2022-04-18 06:01:57	scifichick25	en	1	5	wonder moon knight indulging matzo brei faith ...	POSITIVE	0.978787

Fig. 4. Dataset that has been labeled by DistilBERT

optimized in the modern linear algebra framework. Our investigations then revealed that variations in the last dimension of the tensor (size dimension) hidden) had less impact on computational efficiency (for fixed-parameter budgets) than variations on other factors such as the number of layers. DistilBERT has fewer layers than the BERT architecture.

The text_clean feature was used as input data. It was then scored and labeled with the names POSITIVE and NEGATIVE by DistilBERT, as illustrated in Fig. 4. The labeling process resulted in 1,496 data, in which the positive data sentiment consisted of 653 data or 43.65%. Meanwhile, 843 data (56.35%) represented the negative data sentiment. Because the data was not balanced, undersampling was carried out on the imbalanced sentiment dataset of negative data. It was reduced to as much as 190 data randomly. Therefore, the total data resulted in 1,306 data, with the proportions of each amounting to 653 (balanced).

3) Vectorizing

The text_clean feature data is typically studied for vocabulary and converted from the previous text into matrix data by analyzing each word from the whole sentence.

- Data Splitting

The total of 1,306 data was divided into 80% train data and 20% test data using the train_test_split function. The input data used was 80 percent of the training data from the text_clean feature with its class feature in the form of a label trained with an MNB model compared to an SVM.

- Model Validation

The remaining 20% of data (261 data) were used as test data by utilizing the cross-validation accuracy score to calculate the maximum value for accuracy validation. In addition, the confusion matrix was used to calculate the performance of the test evaluation in the form of overall accuracy, precision, recall, and f1-score values.

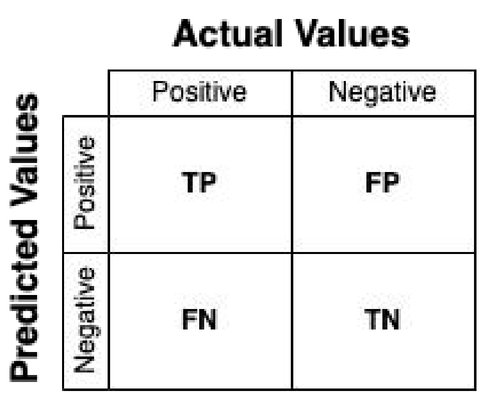

Fig. 5. Confusion Matrix

Figure 5 indicates there are abbreviations of TP, TN, FP, and FN. True positive (TP) is usually used to predict the positive data according to the positive class. True negative (TN) is usually used to predict the negative data according to the negative class. False positive (FP) is a type one error where data that should be predicted to be negative were mistakenly misclassified into a positive class. False negative (FN) is a type two error, where data should be predicted to be positive, but they were misclassified into a negative class.

To measure the performance results, the confusion matrix classification results have four terms, namely overall accuracy, precision, recall, and f1-score, as explained below.

Performance measurement was predicted using the overall accuracy as the following equation (7). Model accuracy score reflected the model's capacity to predict both accurately positive and negative of all results.

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (7)$$

Then, to measure the level of positive observation ratio that was correctly predicted, precision equation (8) was used.

$$precision = \frac{TP}{TP+FP} \quad (8)$$

The ratio of positive observations was correctly predicted to all statements in the actual class to measure the recall (8) value. The precision score was model accuracy. The score reflected the model's ability to predict accurately all the positive predictions it generates [27].

$$recall = \frac{TP}{TP+FN} \quad (9)$$

Finally, to see the balance of performance, the f1-score calculation (10) was carried out from the average precision and recall value. The recall evaluated the effectiveness of the proposed model in identifying positive samples.

$$f1 - score = 2x\frac{precision\ x\ recall}{precision + recall} \quad (10)$$

IV. Result and Discussion

The proportion of data in Fig. 6 explains that the 1,306

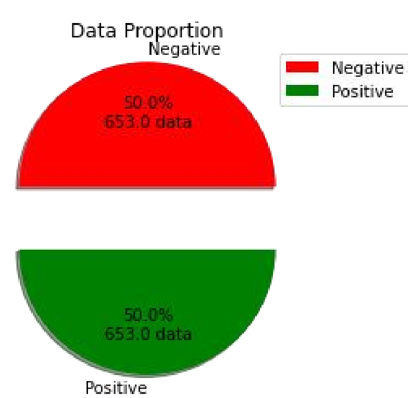

Fig. 6. Proportion Data

datasets were divided into two classes: negative and positive.

The total of 1,306 data was divided into 80% train data and 20% test data using the train_test_split function. Before forming the model, hyperparameter tuning was done by adding a random seed function to help the model choose data points randomly and activate the stratify function to take stratified samples. The input data was the text_clean feature. The class feature is a label that is trained and tested with a Multinomial Naïve Bayes which is compared with a Support Vector Machine. 80% of the data (1,045 data) were trained into the model.

Further, the remaining 20% (261 data) were used as test data by utilizing the cross-validation accuracy score to calculate the maximum value for accuracy validation. In addition, the confusion matrix was used to calculate the performance of the test evaluation in the form of the overall accuracy, precision, recall, and f1-score values. The performance result evaluation model is presented in two results performance validations: cross-validation accuracy score and confusion matrix.

The conclusion is that a shorter from previous section serves two functions. The first is to summarize and bring together the main areas covered in writing, which might be called 'looking back', intended to answer the research problems or purposes. It helps the readers understand why your research should matter to them after they have finished reading the paper. The second is to give a final comment or judgment. The final comment may also include the research's limitation or constraint, making suggestions or recommendations for improvement, and speculating on future research/work.

Table 3.
Cross-Validation Results Accuracy

No	Model	Accuracy
1.	Multinomial Naïve Bayes (MNB)	72.38%
2.	Support Vector Machine (SVM)	70.19%

As shown in Table 3, Multinomial Naïve Bayes obtained the maximum performance at a cross-validation accuracy score of 72.38%. In comparison, the Support Vector Machine only had a maximum cross-validation accuracy score of 70.19%. Even so, it is necessary to re-confirm the accuracy of the cross-validation score using the confusion matrix evaluation performance in Table 4.

Table 4.
Confusion Matrix Results Accuracy

No	Model	Overall Accuracy	Precision	Recall	F1-score
1.	MNB	64.50%	65.83%	60.30%	62.94%
2.	SVM	64.12%	64.12%	64.12%	64.12%

The best evaluation performance of each model calculated using the confusion matrix is explained as follows. The Multinomial Naive Bayes obtained an overall accuracy of 64.50%, precision of 65.83%, recall of 60.30%, and F1 score of 62.94%. Then, the support vector machine model had an overall accuracy performance of 64.12%, precision of 64.12%, recall of 64.12%, and F1 score of 64.12%. This model is different from the research described in the related work, especially with research [18]. Because we used the DistilBERT model to label each data set with a limited dataset, it was not good enough to study the characteristics of audience review sentences.

In short, the moviegoers have negative sentiments towards the Moon Knight series. The level of dissatisfaction reached 56.35%, as shown by tweets on Twitter.

V. CONCLUSION

Using the confusion matrix calculation, the evaluation results indicate that the Support Vector Machine has the best performance. This model can be used to determine sentiment in the form of opinions from the habits of moviegoers, although it is necessary to increase the percentage of performance appraisals with a lot of tweet data by crawling data. However, it is necessary to note that this study has some limitations. The data obtained on Twitter since the new Moon Knight film release was limited. Hence, further research needs to add more data so the model can work better. In addition as a comparison,

	text_clean	score	label
1	people saying oscar isaac english accent moon knight bad literally u london lot sound like sorry	0.999798	NEGATIVE
2	steven line feels like scarcely move fronting makes lot sense	0.999783	NEGATIVE
3	worst thing say moon knight shockingly bad technical issues really shouldnt happening anything access resources disney	0.999776	NEGATIVE
4	went free style edits talking bad cgi lol	0.999739	NEGATIVE
5	noticed past couple weeks many drawing mcu moon knight fucking pale	0.999701	NEGATIVE
6	cant decipher moon knight posters im stupid	0.999691	NEGATIVE
7	way men acted joker came out moon knight episode joker much worse	0.999689	NEGATIVE
8	spoilers anyone else feel fucking awful seeing marc drugged like barely able register anything around like boy deserves better	0.999671	NEGATIVE
9	sorry inform random chick one important xmen characters famous moon knight especially released stop saying bs know anything	0.999665	NEGATIVE
10	im sorry person become every wednesday fault	0.999639	NEGATIVE

Fig. 7. Top 10 Negative Sentiment Predictions

	text_clean	score	label
1	love art style lily much gorgeous character designs amazing highly recommend family	0.999887	POSITIVE
2	ending amazing amazing episode moon knight especially watching along	0.999881	POSITIVE
3	moon knight good oscar isaac beautiful well done	0.999868	POSITIVE
4	loves thank much love awesome time episode best	0.999865	POSITIVE
5	episode hands favorite episode disney show yet absolutely fantastic	0.999860	POSITIVE
6	horror aspect moon knight incredible really going	0.999859	POSITIVE
7	set design show phenomenal definitely historians set wow	0.999855	POSITIVE
8	truly beautiful thing wake today thank kind making tiktok art wise words definitely make sure check	0.999854	POSITIVE
9	brand new poster proud seeing cairo done right first time thanks	0.999845	POSITIVE
10	amazing new full length moonknight couple killer tshirts well	0.999845	POSITIVE

Fig. 8. Top 10 Positive Sentiments Predictions

data labeling techniques other than sentiment polarity are needed from DistilBERT to improve prediction evaluation performance. This study also predicts tweets that are not included in the model to be predicted in real-time in Fig. 7 and Fig. 8. This study also predicts tweets that are not included in the model to be predicted in realtime in Fig. 7 and 8.

Figure 7 shows the top 10 negative sentiments. The Figure indicates that 56.35% of viewers dislike actor Oscar Isaac, who plays Stevent Grant's character as Moon Knight, for various reasons. In fact, if we look at the top 10 positive sentiments in Fig. 8, it is estimated that 43.65% of viewers liked the Moon Knight character and the history inserted into the film.

This study found that the use of sentiment analysis can predict the feasibility of a film, where the feasibility is determined by the opinion of moviegoers on social media. The opinion is predicted, and the proportion of positive and negative classes is determined. The result can help film producers make a decision that will impact on increasing the effectiveness of the budget issued in the future. These predictions can change from time to time to the opinions of film audiences through social media platforms. Therefore, a film producer needs a need analyst data to understand the wishes of the moviegoers.

References

[1] Y. Jia., “The Streaming Service Under Pandemic with the Example of Performance of Disney,” *Advances in Social Science, Education and Humanities Research. Proceedings of the 2021 International Conference on Social Development and Media Communication (SDMC 2021)*, vol. 631, 2021. [Online]. Available: https://www.atlantis-press.com/proceedings/sdmc-21/125968448.

[2] Statista, “Number of Disney Plus subscribers worldwide from 1st quarter to 1st quarter 2022” [Online]. Available: https://www.statista.com/statistics/1095372/disney-plus-number-of-subscribers-us/ (accessed February, 2022).

[3] E. Kontopoulos, C. Berberidis, T. Dergiades and N. Bassiliades, “Ontology-based sentiment analysis of twitter posts,” *Expert Systems with Applications*, vol. 40, pp. 4065-4074, doi: 10.1016/j.eswa.2013.01.001.

[4] K. Sigit, A. P. Dewi, G. Windu, Nurmalasari, T. Muhamad and N. Kadinar, “Comparison of Classification Methods on Sentiment Analysis of Political Figure Electability Based on Public Comments on Online News Media Sites,” *IOP Conf. Ser.: Mater. Sci. Eng.*, 2019, doi: 10.1088/1757-899X/662/4/042003.

[5] Syahriani, A. A. Yana and T. Santoso, “Sentiment analysis of facebook comments on indonesian presidential candidates using the naïve bayes method,” *J. Phys.: Conf. Ser.*, 2020, doi: 10.1088/1742-6596/1641/1/012012.

[6] S. W. Handani, D. I. S. Saputra, Hasirun, R. M. Arino and G. F. A. Ramdhan, “Sentiment Analysis for Go-Jek on Google Play Store,” *J. Phys.: Conf. Ser.*, 2019, doi: 10.1088/1742-6596/1196/1/012032.

[7] H. Wisnu, M. Afif and Y. Ruldevyani, “Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes,” *J. Phys.: Conf. Ser.*, 2020, doi: 10.1088/1742-6596/1444/1/012034.

[8] F. Ratnawati and E. Winarko, “Sentiment Analysis of Movie Opinion in Twitter Using Dynamic Convolutional Neural Network Algorithm,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 12, no. 1, 2018, doi: 10.22146/ijccs.19237.

[9] J. Zheng, L. Zheng and L. Yang, “Research and Analysis in Fine-grained Sentiment of Film Reviews Based on Deep Learning,” *J. Phys.: Conf. Ser.*, 2019, doi: 10.1088/1742-6596/1237/2/022152.

[10] R. Maulana, P. A. Rahayuningsih, W. Irmayani, D. Saputra and W. E. Jayanti, “Improved Accuracy of Sentiment Analysis Movie Review Using Support Vector Machine Based Information Gain,” *J. Phys.: Conf. Ser.*, 2020, doi: 10.1088/1742-6596/1641/1/012060.

[11] U. U. Acikalin, B. Bardak and M. Kutlu, “Turkish Sentiment Analysis Using BERT,” *2020 28th Signal Processing and Communications Applications Conference (SIU)*, pp. 1-4, 2020, doi: 10.1109/SIU49456.2020.9302492.

[12] U. D. Gandhi, P. M. Kumar, G. C. Babu and G. Karthick, “Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM),” *Wireless Pers Commun*, 2021, doi: /10.1007/s11277-021-08580-3.

[13] K. Lu and J. Wu, “Sentiment analysis of film review texts based on sentiment dictionary and SVM,” *Association for Computing Machinery*, pp. 73-77, 2019, doi: 10.1145/3319921.3319966.

[14] L. Shang, L. Sui, S. Wang and D. Zhang, "Sentiment analysis of film reviews based on CNN-BLSTM-Attention," *J. Phys.: Conf. Ser.*, 2020, doi: 10.1088/1742-6596/1550/3/032056.

[15] P. Tiwari, B. K. Mishra, S. Kumar and V. Kumar, "Implementation of n-gram Methodology for Rotten Tomatoes Review Dataset Sentiment Analysis," *Cognitive Analytics: Concepts, Methodologies, Tools, and Applications*, pp. 689-701, 2020, doi: 10.4018/978-1-7998-2460-2.ch036.

[16] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *IEEE Access*, vol. 7, pp. 51522-51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[17] R. Novendri, A. S. Callista, D. N. Pratama and C. E. Puspita, "Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes," *Bulletin of Computer Science and Electrical Engineering*, vol. 1, no. 1, pp. 26-32, 2020, doi: 10.25008/bcsee.v1i1.5.

[18] W. W. Thant and E. E. Mon. (2019). "Analyzing Sentiment of Myanmar Movie Comments Using Naïve Bayes Classifier", in *Proceedings of the Myanmar Universities' Research Conference (MURC 2019)*. 2019. [Online]. Available: https://www.uit.edu.mm/storage/2020/08/16.-Analyzing-Sentiment-of-Myanmar-Movie-Comments-Using-Naive-Bayes-Classifier.pdf.

[19] Y. Kim, M. Kang and S. R. Jeong, "Text mining and sentiment analysis for predicting box office success," *KSII Transactions on Internet and Information Systems*, vol. 12, pp. 4090-4102, 2018, doi: 10.3837/tiis.2018.08.030.

[20] A. S. Rathore, S. Arjaria, S. Khandelwal, S. Thorat and V. Kulkarni, "Movie rating system using sentiment analysis," *Advances in Intelligent Systems and Computing*, vol. 742, pp. 85-98, 2018, doi: 10.1007/978-981-13-0589-4_9.

[21] M. Hall, "A Decision Tree-Based Attribute Weighting Filter for Naive Bayes," *International Conference on Innovative Techniques and Applications of Artificial Intelligence. Research and Development in Intelligent Systems XXIII, SGAI 2016*., pp. 59-70, doi: 10.1007/978-1-84628-663-6_5.

[22] S. Taheri and M. Mammadov, "Learning The Naive Bayes Classifier with Optimization Models," *Int. J. Appl. Math. Comput. Sci*., vol. 23, no. 4, pp. 787-795, 2013, doi: 10.2478/amcs-2013-0059.

[23] I. Rish., "An Empricial Study of the Naïve Bayes Classifier", 2001. [Online], Available: https://www.researchgate.net/publication/228845263_An_Empirical_Study_of_the_Naive_Bayes_Classifier (accessed Mar, 2022).

[24] V. Jakkula., "Tutorial on support vector machine". School of EECS, Washington State University, [Online], 2006, Available: https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf.

[25] E. Sutoyo and A. Almaarif, "Twitter sentiment analysis of the relocation of Indonesia's capital city," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1620-1630, 2020, doi: 10.11591/eei.v9i4.2352.

[26] V. Sanh, L. Debut, J. Chaumond and T. Wolf., "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter". arXiv, 2020. [Online]. Available: https://arxiv.org/abs/1910.01108 (accessed March, 2022).

[27] K. B. Soni, M. Chopade and R. Vaghela, "Credit Card Fraud Detection Using Machine Learning Approach," *Appl. Inf. Syst. Manag.*, vol. 4, no. 2, pp. 71-76, 2021, doi: 10.15408/aism.v4i2.20570.