Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

Aspect-based Sentiment Analysis on Social Media Using Convolutional Neural Network (CNN) Method

Ananta Ihza Ramadhan[1], Erwin Budi Setiawan [2,*]

School of Computing, Informatics, Telkom University, Bandung, Indonesia
Email: [1]anantaihzar@student.telkomuniversity.ac.id, [2]erwinbudisetiawan@telkomuniversity.ac.id
Correspondence Author Email: erwinbudisetiawan@telkomuniversity.ac.id
Submitted: **06/02/2023**; Accepted: **27/03/2023**; Published**: 31/03/2023**

Abstract−Social media are a platform for people to express their opinions on various topics, one of which is Twitter. Movie reviews are a frequently found topic on Twitter that contains a person's opinion of a movie that has been watched. But since opinions are subjective, it is difficult to determine an accurate assessment of a movie. In addition, the diverse aspects of a movie make it difficult to judge whether a review is positive or negative. Referring to that problem, a method is needed to perform sentiment analysis of the problem to be used as an analysis in increasing audience satisfaction with films in the future. In this study, sentiment analysis of movie reviews was carried out based on aspects of plot, acting, and director. This research also performs classification using a CNN model and combines several techniques, that is TF-IDF feature extraction, FastText feature expansion, and SMOTE to calculate the accuracy value and F1-Score. The final results obtained in this study are in the aspect of the plot getting an accuracy of 73.81% (+12,22%) and F1-score 73.72% (+15,93%), the acting aspect obtaining an accuracy value of 89.30% (+0,54%) and F1-score 89.26% (+50,80%), and in the aspect of the director having an accuracy of 87.37% (+0,28%) and F1-score 87.35% (+84,39%). Based on these results, each application of techniques such as TF-IDF, FastText, and SMOTE can increase the accuracy value and F1-Score of the model built.

Keywords: Sentiment Analysis; Social Media; Convolutional Neural Network; Aspect; FastText; SMOTE

1. INTRODUCTION

The internet is commonly used by people to access social media. Social media is a platform that builds networks between users by sharing information such as text, videos, photos, audio, or sentiment information [1][2]. One of the benefits of social media can be used in improving the quality of products or movies through reviews [2]. Social media is used as a means to express their perspective on various topics discussed such as on the Twitter application. The Twitter application allows users to show their interest by sharing opinions about the knowledge they have, such as giving reviews of movies [3]. Movie reviews are someone's opinion in giving an assessment of the movie that has been watched. With movie reviews, filmmakers can use these reviews as evaluation material in making better movies [4]. However, the large number of reviews resulted in discrepancies in assessments of the movie [5]. In addition, a movie review can include several aspects of a movie such as plot, acting, and director aspects [5][6]. This creates difficulties in determining whether a review tends to be a positive or negative opinion. Therefore, a study is needed to help classify opinions on this problem. The classification results in this study can be used as analytical material in making films so as to increase audience satisfaction.

Sentiment analysis is a technique for classifying and evaluating individual opinions based on the emotions contained in a person's subjective information [7][8]. Sentiment analysis detects opinion polarity by classifying positive, negative, or neutral opinions depending on the words used. Sentiment analysis can be applied at several levels, one of which is the aspect level by identifying sentiment on certain aspects [7]. In this research, it focuses on aspect-based sentiment analysis by identifying sentiments on aspects of plot, acting, and director in movies. Sentiment analysis can be done using deep learning techniques, which are part of machine learning. One of the algorithms used in deep learning is Convolutional Neural Network (CNN) which has independent computational analysis capabilities [9].

Some previous research has been done on sentiment analysis, especially using CNN classification and word embedding methods. Research [6] used Naïve Bayes on a movie review dataset to analyze sentiment based on aspects. The aspects used in this research are screenplay, music, acting, plot, movie, and direction. The Naïve Bayes method produces the highest accuracy value of 79.372%. Another study in 2020 [10], compares various methods such as CNN, Naïve Bayes, SVM, and ANN in analyzing sentiment based on aspects. The dataset used is movie reviews. The results stated that the CNN method obtained the best accuracy value of 72.17%. Research in 2020 [11] also uses the CNN method with Word2Vec feature extraction in analyzing sentiment based on aspects. The dataset used is a review of bukalapak.com online stores with the classification of aspects, that is accuracy, quality, service, price, packaging, and delivery. This research resulted in an accuracy of 85.54%. While in research [12] comparing the performance of word embedding, that is Word2Vec, Glove, and Fasttext in classifying text using the CNN method, it was concluded that FastText produced better performance. In research [13] using the CNN method in sentiment analysis. This research uses word embedding, i.e., Word2Vec, FastText and GloVe. The dataset used comes from Twitter social media in a product review. It was concluded that the CNN method using FastText has a higher F-measure value, i.e., in the skip-gram scheme of 91.6% and the CBOW scheme of 86.3%.

Another study conducted in 2022 [14], Sukma and Erwin discussed sentiment based on aspects by applying TF-IDF as feature extraction and FastText as feature expansion to improve performance on the NBSVM classification

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

model. The research resulted in an F1-score of 91.24% on the signal aspect and 88.75% on the service aspect by applying SMOTE in overcoming unbalanced data labeling. And in other aspect-based sentiment research [15] Alhakiem and Erwin in 2022, applied TF-IDF feature extraction and FastText feature expansion to logistic regression, then overcame unbalanced data by using SMOTE. It was concluded that by using SMOTE, F1-Score increased by 3.33% in the signal aspect and 12.91% in the service aspect. In research [16] using the SVM method in aspect-based sentiment. The dataset used is tweet data related to products from Telkomsel with signal and service aspects. This study uses TF-IDF feature extraction, as well as FastText as feature expansion using a corpus of tweets, Indonews, and a combination of tweet corpus and Indonews. In this study, the problem of unbalanced datasets was handled using SMOTE. The results obtained are this research produces an F1-score of 95.93% in the signal aspect and 94.53% in the service aspect. Meanwhile research in 2022 [17], Rimdani and Erwin implemented the Gradient Boosting Decision Tree classification technique with SMOTE and Random Undersampling on signal and service aspects. The results of this study show that the use of SMOTE can produce the best performance value compared to Random Undersampling. The final performance value obtained is the F1 score for the signal aspect of 96.035%, an increase of 25.260%, and the F1 score for the service aspect of 90.256%, an increase of 27.918%. Also in research [18], Aditya and Erwin carried out an aspect-based sentiment analysis on Twitter by applying the Glove feature expansion to the Random Forest classification. The accuracy value obtained on the signal aspect is 80.37% and on the service aspect is 80.12%. The usage of the GloVe feature expansion is also able to increase the accuracy value by 13.15% in the signal aspect and 5.37% in the service aspect.

Based on previous research, these studies have not implemented a sentiment analysis based on movie aspects using CNN with added the FastText feature expansion. Therefore, in this research conducted a sentiment analysis based on the movie aspect using the CNN classification model with TF-IDF feature extraction and FastText feature expansion. In addition, unbalanced dataset labeling is handled using SMOTE to maximize the performance of the model.

This research aims to analyze sentiment based on aspects such as plot, acting, and director using the CNN method and FastText feature expansion by overcoming data imbalance using SMOTE. The results of this study provide accuracy and F1-Score values for each aspect as well as an analysis of the effect of FastText feature expansion on the resulting accuracy. The problem limitation in this research is that the dataset used is Indonesian language movie review tweet data, and the method used is FastText feature expansion and CNN classification algorithm. The aspects used in this study are based on the movie review tweet data that has been collected i.e., plot, acting, and director.

2. RESEARCH METHODOLOGY

2.1 Research Stages

This research was conducted based on the stages shown in the flowchart in Figure 1.

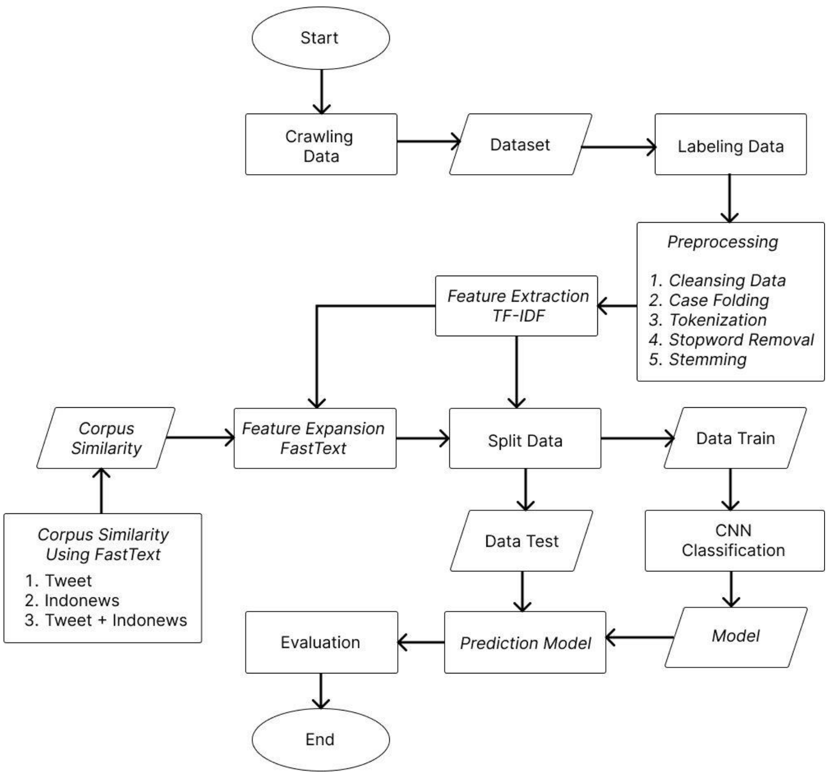

Figure 1. Flowchart of sentiment analysis based on aspects using CNN

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

Figure 1 is a flowchart of conducting sentiment analysis based on aspects using the convolutional neural network (CNN) method with a case study of movie reviews originating from Twitter.

2.2 Data Collection

The data collection stage is carried out using the Python module, snscrape. The dataset used comes from Twitter social media in the form of tweets using keywords as sentiment labels. Table 1 shows keywords based on movie title reviews based on aspects of plot, acting, and director. The dataset is then stored in Pickle Dataframe (.pkl) format. The number of datasets collected is 17,247 data.

Table 1. Keywords and Aspects of Movie Reviews

Category Aspect	Keywords
Plot	plot, cerita, alur, jalan cerita, dialog, ending, skrip
Acting	acting, aktris, actor, pemeran, pemain, karakter, performansi
Director	sinematografi, director, pembuatan film, penyutradaraan, sinematik, director

Table 1 shows the keywords used in crawling data. Keywords that are used more than one and contain information related to each aspect.

2.3 Labeling

Labeling is done manually on each tweet in the dataset that has been collected for use in the classification process. Data labeling is done based on aspects of movie reviews i.e., plot, acting, and director. Table 2 shows the number of sentiment labels on each aspect.

Table 2. Number of Sentiment Labels for Each Aspect

Category Aspect	Positive	Neutral	Negative
Plot	6999	6566	3682
Acting	2907	13770	570
Director	1907	14753	587

Based on Table 2, Sentiment labeling is done by giving label 1 if the discussed aspect is positive, label -1 if the discussed aspect has a negative sentiment, and label 0 if the discussion of an aspect is considered neutral.

2.4 Preprocessing

The data that has been taken from tweets is still raw data that needs to be cleaned to be used in the next stage. At this stage, the following preprocessing is carried out:

a. **Cleansing Data**: Data cleaning is carried out at this stage to remove emojis, numbers, punctuation marks, symbols, hashtags, URLs, and mentions because the classification stage is not needed.
b. **Case Folding**: The case folding stage aims to homogenize the text by converting uppercase letters into lowercase letters.
c. **Tokenization**: The tokenization process is carried out to convert sentences in tweets into a collection of tokens or a collection of word fragments that are entered into an array.
d. **Stopword Removal**: This stage aims to select and retrieve important words. Generally, words with a large number of occurrences will be considered unimportant words.
e. **Stemming**: Stemming is the process of converting words that contain affixes into their base words.

2.5 Feature Extraction TF-IDF

TF-IDF is used as a feature extraction that aims to convert words that have gone through the preprocessing stage into a number vector form. TF-IDF (Term Frequency Inverse Document Frequency) functions to extract word features based on the frequency of occurrence of words and provide word weights in documents [19]. Calculating the frequency of words that often appear in documents can be calculated using (1) TF (Term Frequency) [20].

$$TF(t,d) = \frac{count(t)}{count(d)} \tag{1}$$

The greater the TF value, the more often the word appears in the document. As for (2) IDF (Inverse Document Frequency) functions to calculate the importance of a word, the higher the IDF value indicates that the word is not common or rarely used in documents. But the lower the IDF value explains that the word is commonly used.

$$IDF(t,D) = log\frac{number\ of\ documents}{number\ of\ documents\ containing\ t} \tag{2}$$

Based on this formula, (3) TF-IDF can be formulated as follows.

$$TF - IDF\ (t,d,D) = TF(t,d) \times IDF(t,D) \tag{3}$$

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

t is the number of occurrences of each word in all documents, d is the number of documents containing that word, and D is the total number of documents.

2.6 SMOTE

Based on table 2, it can be seen that the data used is not balanced between classes. Therefore, the Synthetic Minority Over-sampling Technique (SMOTE) method is used, which is a data resampling technique to overcome unbalanced data. SMOTE works by creating synthesized data based on minority classes in the cluster [21][22]. The class with the least number will be resampled by SMOTE so that the class is close to or equal to the number of the largest class.

2.7 Feature Expansion FastText

FastText is one of the word embedding methods that create word vectors or performs text classification, as well as overcoming time problems in model training [23][24]. FastText enables the identification of undefined word vectors by taking into account subword information in word embedding and constructing word vectors as n-grams of characters [25][26]. In simple terms, when performing model training and the input word is not identified, the word is grouped into n-grams to identify its vector insertion [23].

In this research, the FastText model is used to perform the feature expansion process using the gensim module. The FastText model used includes a corpus or collection of words taken from movie and news review datasets. The corpus used in this research is a tweet corpus, Indonews corpus, and a combination of Indonews and tweet corpus. The corpus can be used to determine the similarity of a word. An example of similiarity can be seen in Table 3.

Table 3. Example similarity of the word "sinematografi".

Rank	Features
1	cinematografi
2	sinematografer
3	biografi
4	visual
5	cinematograpy
6	koreografi
7	cinematography
8	bank
9	sinematik
10	visinema

Based on Table 3, it shows the ranking of the similiarity of the word "sinematografi" from 1 to 10. The highest rating indicates the word is very similar to the word "cinematography". After the corpus and similarity are obtained, the next step is to expand the features by replacing the 0-vector in the feature extraction with the similarity vector of the searched word [27].

2.8 Split Data

At this stage, data is separated into train data and test data. Train data is data that will be used to train and find the best CNN model, while test data is used for CNN classification testing. Split data is done with more train data than test data. The comparison of train data and test data in this study used a ratio of 90:10, 80:20, and 70:30.

2.9 Classification with Convolutional Neural Network

A convolutional neural network (CNN) is a deep learning type of artificial neural network. Generally, CNN is utilized for computer vision purposes to perform image classification, but CNN has good performance in text classification [28]. If in image classification the input is an image, then in text classification on CNN the input is a word vector generated by word embedding [29]. In performing classification on text, the CNN model has four parts namely the embedding layer, convolutional layer, pooling layer, and fully connected layer. The structure of the CNN model can be shown in Figure 2.

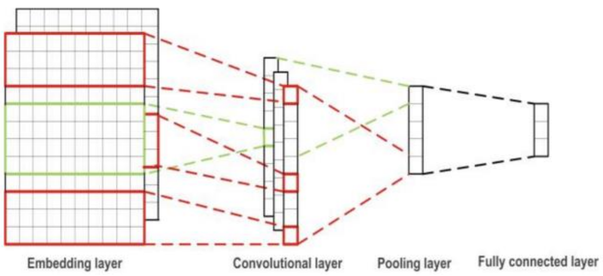

Figure 2. CNN model structure

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

Figure 2 is the structure of the CNN model, and the following is the definition of parts in the CNN model [30]:

a. Embedding layer: The word vector matrix is sorted from top to bottom based on the words in the sentence. The matrix is defined as m × n if a sentence has m words and the dimension of the word vector is n.
b. Convolutional layer: The previous layer obtains multiple feature maps using a convolution operation. Each convolutional window is defined as k × n, k being the number of longitudinal words and n being the dimension of the word vector. Then this layer generates a group of feature maps with one column.
c. Pooling layer: This layer is also called the sub-sampling layer. Its function is to reduce the size of the input data by max-pooling by taking the largest value of the input data.
d. Fully connected layer: This layer is the last layer that is generally connected to the other fully connected layers and its output is the probability distribution of the labels.

After labeling, preprocessing, feature extraction, feature expansion, and split data, the classification will be carried out using Convolutional Neural Network (CNN). The results issued at this stage are sentiment models that classify opinions into negative, neutral, and positive.

2.10 Performance Evaluation

The performance evaluation process is performed on each trial and calculated using the confusion matrix. The confusion matrix describes the amount of test data that is correctly classified and misclassified by the model [31]. The confusion matrix is generally made in the form of a table, as in Table 4 [32][33].

Table 4. Confusion Matrix

Actual Class	Prediction Class	
	Positive Prediction	Negative Prediction
Actual Positive	TP	FN
Actual Negative	FP	TN

Based on Table 4, True Positive (TP) is a positive class that is correctly predicted as a positive class, False Positive (FP) is a negative class that is incorrectly predicted as a positive class, False Negative (FN) is a positive class that is incorrectly predicted as a negative class, and True Negative (TN) is a negative class that is correctly predicted as a negative class. By using a confusion matrix, we can calculate performance measurements such as (4) accuracy, (5) precision, (6) recall, and (7) F1 Measure. The following is the formula for calculating system performance:

a. Accuracy is a calculation of the ratio of correctly predicted data (positive or negative) and compared to all data.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

b. Precision is the ratio of data predicted to be true positive and compared to the overall result of data predicted to be positive.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

c. Recall is the ratio of predicted true-positive data to all true-positive data.

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

d. F1 Measure is a weighting that considers precision and recall values.

$$F1\ Measure = \frac{2 \times (Recall \times Precision)}{Recall + Precision} \quad (7)$$

3. RESULT AND DISCUSSION

In this study, testing was carried out which was divided into four scenarios using a classification model built by CNN, that is:

a. The first scenario tests using the baseline.
b. The second scenario performs baseline testing by adding TF-IDF feature extraction.
c. The third scenario adds TF-IDF feature extraction and FastText feature expansion.
d. The fourth scenario applies SMOTE to address unbalanced data.

Each scenario uses the CNN classification model and executes the program 5 times. The test results of each scenario are in the form of average accuracy and F1-Score values.

3.1 Evaluation Results

The first scenario was tested to find a baseline model using CNN classification. The dataset used in the test amounted to 17247 movie review data with a train size and test size ratio of 90:10, 80:20, and 70:30. The results of this test obtained the average accuracy value for each aspect as shown in Table 5.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

Table 5. CNN baseline test results

Category Aspect	Test Size					
	90:10		80:20		70:30	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
Plot	**65,77**	**63,59**	64,63	61,78	64,62	62,16
Acting	**88,82**	**59,19**	88,32	56,30	88,15	55,34
Director	**87,12**	**47,37**	86,78	45,36	87,05	44,78

Based on the test results in Table 5, on the aspect of the plot, the highest accuracy value is 65.77% and F1-Score is 63.59% with a test size ratio of 90:10. In the acting aspect, the highest accuracy value is 88.82% and F1-Score is 59.19% with a test size ratio of 90:10. While the director aspect obtained the highest accuracy value of 87.12% and F1-Score of 47.37% with a ratio of 90:10. Based on these results, the 90:10 ratio produces the highest accuracy value for each aspect. The test results in the first scenario will then be used as the basis for the next scenario.

In the second scenario, the baseline model in the previous scenario is then added with feature extraction using TF-IDF. Testing is done by comparing the accuracy results for the max number of features in TF-IDF, that is 1000, 5000, and 10000. The test results in this scenario can be seen in Table 6.

Table 6. Baseline test results with TF-IDF

Category Aspect	Max Feature					
	1000		5000		10000	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
Plot	66,64 (+132)	66,28 (+4,23)	**67,51 (+2,64)**	**67,35 (+5,91)**	65,76(-0,02)	65,70 (+3,32)
Acting	**89,98 (+1,30)**	**68,40 (+15,56)**	89,43 (+0,69)	65,34 (+10,39)	89,62 (+0,90)	66,33 (+12,06)
Director	**88,01 (+1,02)**	**48,95 (+3,33)**	87,94 (+0,94)	51,93 (+9,63)	87,84 (+0,83)	49,86 (+5,26)

Table 6 shows the highest accuracy value in the plot aspect of 67.51% and F1-Score of 67.35% with a max feature count of 5000. The increase in accuracy against the baseline in the plot aspect is 2.64% and F1-Score is 5.91%. In the acting aspect, the highest accuracy value of 89.98% and F1-Score of 68.40% was obtained with a max feature count of 1000. The accuracy and F1-Score values increased by 1.30% and 15.56% respectively. In the director aspect, the highest accuracy value of 88.01% and F1-Score of 48.95% was obtained with a max feature count of 1000. The increase in accuracy in the plot aspect is 1.02% and F1-Score is 3.33%.

In the third scenario, testing is done by adding FastText feature expansion to the model from the previous scenario. Testing using a similarity corpus consisting of a Tweet corpus, Indonews corpus, and a combination corpus of Tweet and Indonews. Each corpus will be calculated from the top 1, top 5, and top 10. The test results on the plot aspect can be seen in Table 7.

Table 7. Test result of baseline plot aspect with TF-IDF and FastText

Top	Plot Aspect					
	Tweet Corpus		Indonews Corpus		Tweet + Indonews Corpus	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
1	68,00 (+3,39)	67,79 (+6,60)	67,73 (+2,98)	67,56 (+6,24)	**68,50 (+4,15)**	**66,77 (+5.00)**
5	65,14 (-0,96)	66,62 (+4,76)	67,59 (+2,77)	67,42 (+6,02)	68,12 (+3,57)	66,31 (+4,28)
10	65,10 (-1,02)	63,45 (-0,22)	67,17 (+2,13)	66,98 (+5,33)	64,13 (-2,49)	66,08 (+3,92)
20	63,13 (-4,01)	60,42 (-4,99)	67,68 (+2,90)	66,20 (+4,10)	67,25 (+2,25)	65,43 (+2,89)

Based on Table 7, the highest accuracy value in the plot aspect of 68.50% and F1-Score of 66.77% was produced by top 1 using a combination of Tweet and Indonews corpus. The increase in accuracy and F1-Score values in the plot aspect is 4.15% and 5.00%, respectively.

Table 8. Test result of baseline acting aspect with TF-IDF and FastText

Top	Acting Aspect					
	Tweet Corpus		Indonews Corpus		Tweet + Indonews Corpus	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
1	**89,95 (+1,27)**	**64,76 (+9,41)**	89,61 (+0,89)	66,56 (+12,45)	89,69 (+0,98)	66,90 (+13,03)
5	89,43 (+0,69)	63,54 (+7,35)	89,79 (+1,09)	65,34 (+10,39)	89,29 (+0,53)	66,11 (+11,69)
10	89,30 (+0,54)	64,35 (+8,72)	89,77 (+1,07)	64,76 (+9,41)	89,54 (+0,81)	64,41 (+8,82)
20	88,73 (-0,10)	61,54 (+3,97)	89,58 (+0,86)	64,49 (+8,95)	89,74 (+1,04)	63,02 (+6,47)

In Table 8, the highest accuracy value on the acting aspect of 89.95% and F1-Score of 64.76% was produced by the top 1 using the Tweet corpus. The increase in accuracy and F1-Score on the acting aspect is 1.27% and 9.41% respectively.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

Table 9. Test result of baseline director aspect with TF-IDF and FastText

Top	Director Aspect					
	Tweet Corpus		Indonews Corpus		Tweet + Indonews Corpus	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
1	88,01 (+1,02)	49,78 (+5,09)	87,23 (+0,13)	51,50 (+8,72)	87,43 (+0,36)	52,17 (+10,13)
5	**88,02 (+1,03)**	**49,97 (+5,48)**	87,33 (+0,24)	52,70 (+11,25)	87,30 (+0,21)	53,33 (+12,58)
10	87,91 (+0,91)	48,33 (+2,03)	87,56 (+0,51)	52,40 (+10,62)	87,52 (+0,46)	52,82 (+11,51)
20	87,71 (+0,68)	49,40 (+4,29)	87,72 (+0,69)	52,41 (+10,64)	87,88 (+0,87)	53,82 (+13,62)

In Table 9, the highest accuracy value in the director aspect of 88.02% and F1-Score of 49.97% was produced by the top 5 using the Tweet corpus. The increase in accuracy and F1-Score values in the director aspect are 1.03% and 5.48%, respectively.

In the fourth scenario, the application of the SMOTE algorithm was added to the results of the previous scenario to overcome unbalanced data. The test results using the SMOTE algorithm can be seen in Table 10.

Table 10. Test results with the addition of SMOTE

Category Aspect	Without SMOTE		With SMOTE	
	Accuracy (%)	F1-Score (%)	Accuracy (%)	F1-Score (%)
Plot	68,50 (+4,15)	66,77 (+5,00)	**73,81 (+12,22)**	**73,72 (+15,93)**
Acting	89,95 (+1,27)	64,76 (+9,41)	**89,30 (+0,54)**	**89,26 (+50,80)**
Director	88,02 (+1,03)	49,97 (+5,49)	**87,37 (+0,28)**	**87,35 (+84,39)**

Based on Table 10, each aspect experienced a higher increase in accuracy after using SMOTE compared to without SMOTE. The increase in accuracy value on the plot aspect with SMOTE is 12.22% and F1-Score is 15.93%. In the acting aspect, the increase in accuracy value is 0.54% and F1-Score is 50.80% after the application of SMOTE. As for the director aspect, the increase in accuracy value is 0.28% and F1-Score is 84.39% after using SMOTE.

3.2 Discussion

Based on the tests that have been carried out using four scenarios, each scenario can increase the accuracy and F1-score values of the CNN classification model. Figure 3 shows the increase in accuracy and F1-Score values for each scenario in the plot aspect.

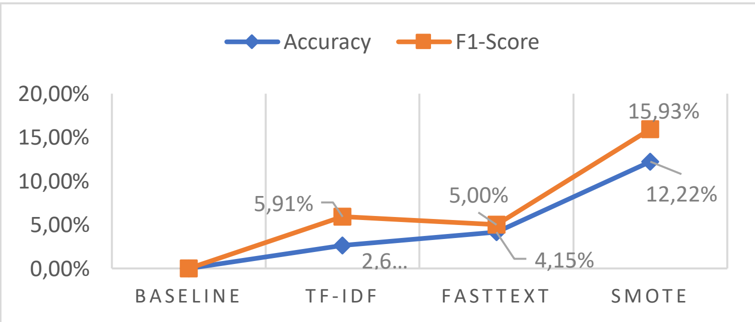

Figure 3. Increase in Accuracy and F1-Score on plot aspects

In the plot aspect, the baseline model with the best accuracy and F1-score results was obtained with a test size ratio of 90:10. In the second scenario, there is an increase in accuracy and F1-Score values of 2.64% and 5.91% respectively by using the number of max features 5000. In the scenario third, an increase in accuracy value of 4.15% and F1-Score of 5.91% was obtained with top 1 in the Twitter and news combination corpus. In the fourth scenario, there was a significant improvement after the application of the SMOTE algorithm. The increase in accuracy and F1-Score values in the fourth scenario was 12.22% and 15.93%, respectively.

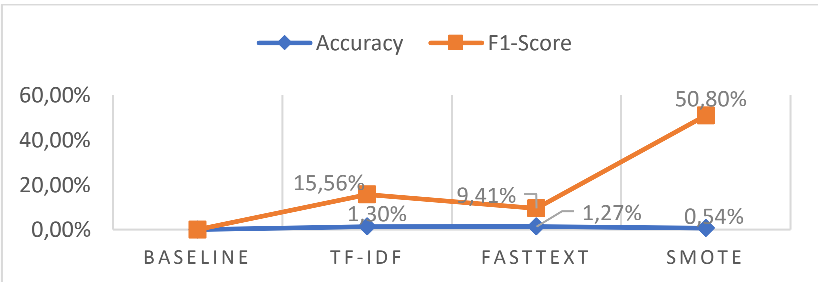

Figure 4. Increase in Accuracy and F1-Score on acting aspects

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

Figure 4 shows the rate of increase in accuracy and F1-Score values for each scenario in the acting aspect. In the acting aspect, the baseline model with the best accuracy and F1-score results is obtained with a test size ratio of 90:10. In second scenario, there was an increase in accuracy and F1-Score values of 1.30% and 15.56% respectively by using a max feature count of 1000. In third scenario, an increase in accuracy value of 1.27% and F1-Score of 9.41% was obtained with top 1 in the Twitter corpus. In fourth scenario, there was an increase after the application of the SMOTE algorithm. The increase in accuracy and F1-Score values in the fourth scenario was 0.54% and 50.80%, respectively.

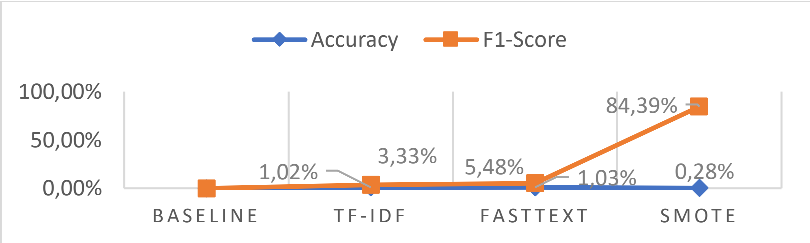

Figure 5. Increase in Accuracy and F1-Score on director aspects

Figure 5 shows the rate of increase in accuracy and F1-Score values for each scenario in the director aspect. In the director aspect, the baseline model with the best accuracy and F1-score results is obtained with a test size ratio of 90:10. In the second scenario, there was an increase in accuracy and F1-Score values of 1.02% and 3.33% respectively by using a max feature count of 1000. In the third scenario, an increase in accuracy value of 1.03% and F1-Score of 5.48% was obtained with the top 5 in the Twitter corpus. In the fourth scenario, there was an improvement after the application of the SMOTE algorithm. The increase in accuracy and F1-Score values in the fourth scenario was 0.28% and 84.39%, respectively.

Based on the test results, all scenarios can affect the increase in accuracy and F1-Score values for each aspect of the CNN classification model. The largest increase in accuracy and F1-Score values is obtained in the fourth scenario compared to other scenarios. This happens because in the fourth scenario, the SMOTE algorithm can balance the amount of data for each class so that the performance of the model used becomes better.

4. CONCLUSION

This research conducts sentiment analysis based on three aspects, i.e., plot, acting, and director aspects. The model used is CNN by applying feature extraction using TF-IDF, feature expansion using FastText, and overcoming the problem of unbalanced label datasets using SMOTE. The dataset used is in the form of movie review tweets totaling 17,247 data. This study was conducted to analyze the accuracy results of the four scenarios. The results show that by adding feature expansion, the three aspects experience an increase in accuracy and F1-Score against the baseline. In the plot aspect, the accuracy was 68.50% and F1-Score was 66.77%, using the Tweet+Indonews corpus generated by Top-1. In the acting aspect, the resulting accuracy is 89.95% with an F1-Score of 64.76%, these results use the Tweet corpus on Top-1. The director aspect produced an accuracy of 88.02% and F1-Score of 49.97% using the Tweet corpus with Top-5. By overcoming the unbalanced labeling of the dataset, applying SMOTE resulted in a significant improvement. In the plot aspect, the accuracy increased to 73.81% with an increase of 12.22% and F1-Score 73.72% with an increase of 15.93% from the baseline. Accuracy in the acting and director aspects decreased when compared to the feature expansion model without SMOTE, but F1-Score experienced a very high increase in both aspects. The acting aspect obtained an accuracy of 89.30% with an increase of 0.54% and F1-Score of 89.26% resulting in an increase of 50% from the baseline, then the director aspect resulted in an increase of 0.28% with an accuracy of 87.37% and F1-Score of 87.35% and an increase of 84.39% from the baseline. Suggestions for further research can be done with a larger number of datasets and a more balanced number of labeling.

REFERENCES

[1] F. Li, J. Larimo, and L. C. Leonidou, “Social media marketing strategy: definition, conceptualization, taxonomy, validation, and future agenda,” J. Acad. Mark. Sci., vol. 49, no. 1, pp. 51–70, 2021, doi: 10.1007/s11747-020-00733-3.

[2] Z. Drus and H. Khalid, “Sentiment analysis in social media and its application: Systematic literature review,” Procedia Comput. Sci., vol. 161, pp. 707–714, 2019, doi: 10.1016/j.procs.2019.11.174.

[3] N. A. AlSomaikhi and Z. A. Alzamil, “Twitter Users’ Classification Based on Interest,” Int. J. Inf. Retr. Res., vol. 10, no. 1, pp. 1–12, 2020, doi: 10.4018/ijirr.2020010101.

[4] C. Nanda, M. Dua, and G. Nanda, “Sentiment Analysis of Movie Reviews in Hindi Language Using Machine Learning,” Proc. 2018 IEEE Int. Conf. Commun. Signal Process. ICCSP 2018, pp. 1069–1072, 2018, doi: 10.1109/ICCSP.2018.8524223.

[5] N. A. Shafirra and I. Irhamah, “Klasifikasi Sentimen Ulasan Film Indonesia dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolutional Neural Network (CNN),” J. Sains dan Seni ITS, vol. 9, no. 1, 2020, doi: 10.12962/j23373520.v9i1.51825.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 4, Maret 2023 Page: 1828−1836
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i4.3103

[6] V. Parkhe and B. Biswas, “Aspect based sentiment analysis of movie reviews: Finding the polarity directing aspects,” Proc. - 2014 Int. Conf. Soft Comput. Mach. Intell. ISCMI 2014, pp. 28–32, 2014, doi: 10.1109/ISCMI.2014.16.
[7] A. Alsaeedi and M. Z. Khan, “A study on sentiment analysis techniques of Twitter data,” Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 2, pp. 361–374, 2019, doi: 10.14569/ijacsa.2019.0100248.
[8] M. Cindo and D. P. Rini, “Literatur Review: Metode Klasifikasi Pada Sentimen Analisis,” Semin. Nas. Teknol. Komput. Sains, pp. 66–70, 2019, [Online]. Available: https://seminar-id.com/semnas-sainteks2019.html
[9] P. A. Nugroho, I. Fenriana, and R. Arijanto, “Implementasi Deep Learning Menggunakan Convolutional Neural Network Pada Ekspresi Manusia.” Jurnal ALGOR, 2020.
[10] R. Kumar and S. Garg, Aspect-Based Sentiment Analysis Using Deep Learning Convolutional Neural Network, vol. 933. Springer Singapore, 2020. doi: 10.1007/978-981-13-7166-0_5.
[11] M. T. Ari Bangsa, S. Priyanta, and Y. Suyanto, “Aspect-Based Sentiment Analysis of Online Marketplace Reviews Using Convolutional Neural Network,” IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 14, no. 2, p. 123, 2020, doi: 10.22146/ijccs.51646.
[12] A. Nurdin, B. A. S. Aji, A. Bustamin, and Z. Abidin, “Perbandingan Kinerja Word Embedding Word2Vec , Glove ,” J. TEKNOKOMPAK, vol. 14, no. 2, pp. 74–79, 2020.
[13] Aytuğ Onan, Deep Learning Based Sentiment Analysis on Product Reviews on Twitter, vol. 1054. 2019. doi: 10.1007/978-3-030-27355-2_8.
[14] S. D. Lestari and E. B. Setiawan, “Sentiment Analysis Based on Aspects of Telkomsel Users on Twitter Using FastText Feature Expansion and NBSVM Classification,” J. Comput. Syst. Informatics, vol. 3, no. 4, pp. 469–477, 2022, doi: 10.47065/josyc.v3i4.2202.
[15] H. R. Alhakiem and E. B. Setiawan, “Aspect-Based Sentiment Analysis on Twitter Using Logistic Regression with FastText Feature Expansion,” J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 6, no. 5, pp. 840–846, 2022, doi: 10.29207/resti.v6i5.4429.
[16] Muhammad Afif Raihan and Erwin Budi Setiawan, “Aspect Based Sentiment Analysis with FastText Feature Expansion and Support Vector Machine Method on Twitter,” J. RESTI (Rekayasa Sist. dan Teknol. Informasi), vol. 6, no. 4, pp. 591–598, 2022, doi: 10.29207/resti.v6i4.4187.
[17] R. A. Annisa and E. B. Setiawan, “Aspect Based Sentiment Analysis on Twitter Using Word2Vec Feature Expansion Method and Gradient Boosting Decision Tree Classification Method,” IEEE, 2022.
[18] A. M. Zakaria and E. B. Setiawan, “Aspect-Based Analysis of Telkomsel User Sentiment on Twitter Using the Random Forest Classification Method and Glove Feature Expansion,” J. Teknol. dan Sist. …, 2022, [Online]. Available: https://jtsiskom.undip.ac.id/article/view/14558
[19] D. A. Widodo, N. Iksan, and B. Sunarko, “Sentiment Analysis of Twitter Media for Public Reaction Identification on Covid-19 Monitoring System using Hybrid Feature Extraction Method.” Internasional Journal of Intelligent System and Applications In Engineering, 2022.
[20] N. Ritha, N. Hayaty, T. Matulatan, A. Uperiati, M. Rathomi, and M. Bettiza, “Sentiment Analysis of Health Protocol Policy Using K- Nearest Neighbor and Cosine Similarity,” 2023, doi: 10.4108/eai.11-10-2022.2326274.
[21] W. Chaipanha and P. Kaewwichian, “Smote Vs. Random Undersampling for Imbalanced Data-Car Ownership Demand Model,” Commun. - Sci. Lett. Univ. Žilina, vol. 24, no. 3, pp. D105–D115, 2022, doi: 10.26552/com.C.2022.3.D105-D115.
[22] A. Harris and E. Yanti, “Optimizing Attack Detection for High Dimensionality and Imbalanced Data with SMOTE , Chi-Square and Random Forest Classifier,” Int. J. Informatics Comput. Sci., vol. 6, no. 1, pp. 1–14, 2022, doi: 10.30865/ijics.v6i1.3890.
[23] E. M. Dharma, F. L. Gaol, H. L. H. S. Warnars, and B. Soewito, “the Accuracy Comparison Among Word2Vec, Glove, and Fasttext Towards Convolution Neural Network (Cnn) Text Classification,” J. Theor. Appl. Inf. Technol., vol. 100, no. 2, pp. 349–359, 2022.
[24] T. Yao, Z. Zhai, and B. Gao, “Text Classification Model Based on fastText,” Proc. 2020 IEEE Int. Conf. Artif. Intell. Inf. Syst. ICAIIS 2020, pp. 154–157, 2020, doi: 10.1109/ICAIIS49377.2020.9194939.
[25] B. Kuyumcu, C. Aksakalli, and S. Delil, “An automated new approach in fast text classification (fastText): A case study for Turkish text classification without pre-processing,” ACM Int. Conf. Proceeding Ser., pp. 1–4, 2019, doi: 10.1145/3342827.3342828.
[26] J. Choi and S. W. Lee, “Improving FastText with inverse document frequency of subwords,” Pattern Recognit. Lett., vol. 133, pp. 165–172, 2020, doi: 10.1016/j.patrec.2020.03.003.
[27] R. A. Yahya and E. B. Setiawan, “Feature Expansion with FastText on Topic Classification Using the Gradient Boosted Decision Tree on Twitter,” pp. 322–327, 2022, doi: 10.1109/icoict55009.2022.9914896.
[28] A. Jacovi, O. S. Shalom, and Y. Goldberg, “Understanding Convolutional Neural Networks for Text Classification,” EMNLP 2018 - 2018 EMNLP Work. BlackboxNLP Anal. Interpret. Neural Networks NLP, Proc. 1st Work., pp. 56–65, 2018, doi: 10.18653/v1/w18-5408.
[29] I. Santos, N. Nedjah, and L. De Macedo Mourelle, “Sentiment analysis using convolutional neural network with fasttext embeddings,” 2017 IEEE Lat. Am. Conf. Comput. Intell. LA-CCI 2017 - Proc., vol. 2018-Janua, pp. 2–6, 2017, doi: 10.1109/LA-CCI.2017.8285683.
[30] Z. Wang and Z. Qu, “Research on Web text classification algorithm based on improved CNN and SVM,” Int. Conf. Commun. Technol. Proceedings, ICCT, vol. 2017-Octob, pp. 1958–1961, 2018, doi: 10.1109/ICCT.2017.8359971.
[31] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” J-SAKTI (Jurnal Sains Komput. dan Inform., vol. 5, no. 2, pp. 697–711, 2021.
[32] F. Novitasari and M. D. Purbolaksono, “Analysis Sentiment Aspect Level on Beauty Product Reviews Using Chi-Square and Naïve Bayes,” J. Data Sci. Its Appl., vol. 4, no. 1, pp. 18–30, 2021, [Online]. Available: https://commdis.telkomuniversity.ac.id/jdsa/index.php/jdsa/article/view/72
[33] A. Indriani, “Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier,” Semin. Nas. Apl. Teknol. Inf. Yogyakarta, vol. 1, no. 1, pp. 21–2014, 2014, [Online]. Available: www.bluefame.com,