Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

Sentiment Analysis Against IndiHome and First Media Internet Providers Using Ensemble Stacking Method

Arya Rafif Muhammad Fikri[1], Jondri[2,*], Widi Astuti[3]

School of Computing, Informatics Study Program, Telkom University, Bandung, Indonesia
Email: [1]aryarafif@student.telkomuniveristy.ac.id, [2,*]jondri@telkomuniveristy.ac.id, [3]widiwdu@telkomuniversity.ac.id
Email Penulis Korespondensi: jondri@telkomuniversity.ac.id
Submitted:**25/07/2022**; Accepted:**18/08/2022**; Published: **30/09/2022**

Abstract−Customer satisfaction is one of the factors that can be used to measure the success of service in a company. In the era of the 2000s until now, internet service providers have continued to grow throughout the world, including in Indonesia. IndiHome and First Media are companies that provide internet services that make it easy for the public to communicate and obtain information. With many uses of IndiHome and First Media internet services, there are often several obstacles that cause various responses from users. Users usually channel these responses to IndiHome or First Media customer care on Twitter. Sentiment analysis on Twitter needs to be done to see how IndiHome and First Media users respond to the internet services that have been provided. The first process in sentiment analysis is taking data from Twitter, then data cleaning and feature extraction are carried out, after that the data is classified using the ensemble stacking method. The dataset for this study was obtained from Twitter using the Twitter API and the Tweepy library. The dataset that has been collected is 6.962 tweets for the IndiHome dataset and 8,089 tweets for the First Media dataset. This study conducts sentiment analysis using the Ensemble Stacking with three base classifiers and a meta classifier. The base classifier used is Naïve Bayes, K-Nearest Neighbor, and Decision Tree, while the meta classifier used is Logistic Regression. This study uses the term frequency-inverse document frequency (TF-IDF) to determine the frequency value of a word in a document. This study uses two test scenarios: testing without oversampling and testing with oversampling on the dataset. The results show that Ensemble Stacking with term frequency-inverse document frequency feature extraction produces the highest accuracy, with an accuracy value of 88.27% on the IndiHome dataset and 92.56% on the First Media dataset by oversampling on both datasets.

Keywords: IndiHome; First Media;Oversampling; Ensemble Stacking; TF-IDF

1. INTRODUCTION

In today's technological developments, the internet has become one of the media that is often used to obtain and search for information from various worlds. In 2018 internet users in Indonesia reached 171.7 million people, this figure has increased by 10.12% after being seen and compared to the previous year, which was only 143.26 million people, this was issued by APJII [1]. In Indonesia, one of the internet service providers is IndiHome and First Media, each of which currently has many customers. With a large number of uses of IndiHome and First Media internet services, it will give various opinions, both positive and negative.

There are often obstacles in using internet services, including a slow, unstable internet connection, even until the internet connection suddenly disconnects. IndiHome and First Media have customer care on several social media, one of the social media used is Twitter. IndiHome has customer care on Twitter with the username @IndiHomeCare, and FirstMedia has customer care with the username @FirstMediaCares. @IndiHomeCare and @FirstMediaCares is one of the forums provided by IndiHome and First Media for its users to issue opinions on IndiHome and First Media internet services. The admins of @IndiHomeCare or @FirstMediaCares will respond to the results of opinions that have been submitted. Opinions expressed by users are subjective expressions that describe the user's feelings. Therefore, the opinions that users on Twitter have expressed can be used as data to conduct sentiment analysis. This study classifies the responses of IndiHome and First Media internet service users into positive, neutral, and negative sentiments.

In a previous study conducted in 2019, Shakina Rizkia, Erwin Budi Setiawan, and Diyas Puspandari with the topic of customer satisfaction sentiment analysis on the internet provider IndiHome on Twitter using the Decision Tree and TF-IDF weighting aimed to find out how much Indihome customer satisfaction is at social media Twitter and to determine the accuracy generated by the Decision Tree which is implemented in sentiment analysis. In this study, it can be concluded that based on the results of the tests that have been carried out by Shakina, the best accuracy results are 80.1% in the bigram scenario with TF-IDF weighting [2].

In 2019, research conducted by Ainun Nisa on the topic of sentiment analysis using the Naive Bayes selection feature Chi-Square for telecommunication service providers shows that the highest performance value is found in the Naive Bayes using the Chi-Square with the result of 85.5% accuracy, recall 86%, precision 83%, and f1-score 84% [3].

K-Nearest Neighbor was chosen to be one of the base classifiers in this study because in previous research conducted by Akhmad Deviyanto and M. Didik R. Wahyudi with the topic of applying sentiment analysis to Twitter users using the K-Nearest Neighbor with TF-IDF word weighting and Cosine Similarity function produces the greatest accuracy at k=5 with an accuracy value of 67.2% [4]. The results of the accuracy of these studies are still relatively low.

Sentiment analysis on Twitter needs to be done to see how IndiHome and First Media users respond to the internet services that have been provided, as well as to provide feedback to IndiHome and First Media to maintain or

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

improve their internet services. The data used to conduct sentiment analysis are tweets containing information related to opinions and complaints on IndiHome and First Media internet services. This study performs sentiment analysis using the Ensemble Stacking with three base classifiers and a meta classifier. The base classifier used is Naïve Bayes, K-Nearest Neighbor, and Decision Tree, while the meta classifier used is Logistic Regression. This study uses the term frequency-inverse document frequency (TF-IDF) to determine the frequency value of a word in a document. The Ensemble Stacking was chosen because it can improve the model's accuracy in predicting the sentiment of IndiHome and First Media users by combining the prediction results from several base classifier models, which will be predicted again by the meta classifier to get the best accuracy in both scenarios. This study uses two test scenarios: testing without oversampling the dataset and testing by oversampling the dataset to see the best final accuracy results.

2. RESEARCH METHODOLOGY

2.1 System Design

The system design that was built in this study was used to analyze the sentiment of each *tweet* mentioned by Twitter users to the @IndiHomeCare and @FirstMediaCares accounts and to obtain the accuracy value of each model used. The first stage is to fetch data from Twitter to get the dataset, then label each tweet manually. Data that has been labeled will be cleaned at the preprocessing stage so that the data can be used optimally. The feature extraction used in this research is TF-IDF, and then the data is divided into data tests and data trains. The classification method used is Ensemble Stacking with the base classifier of *Naïve Bayes*, *K-Nearest Neighbor*, *Decision Tree*, and *Logistic Regression* as meta classification. The flow of system design can be seen in Figure 1.

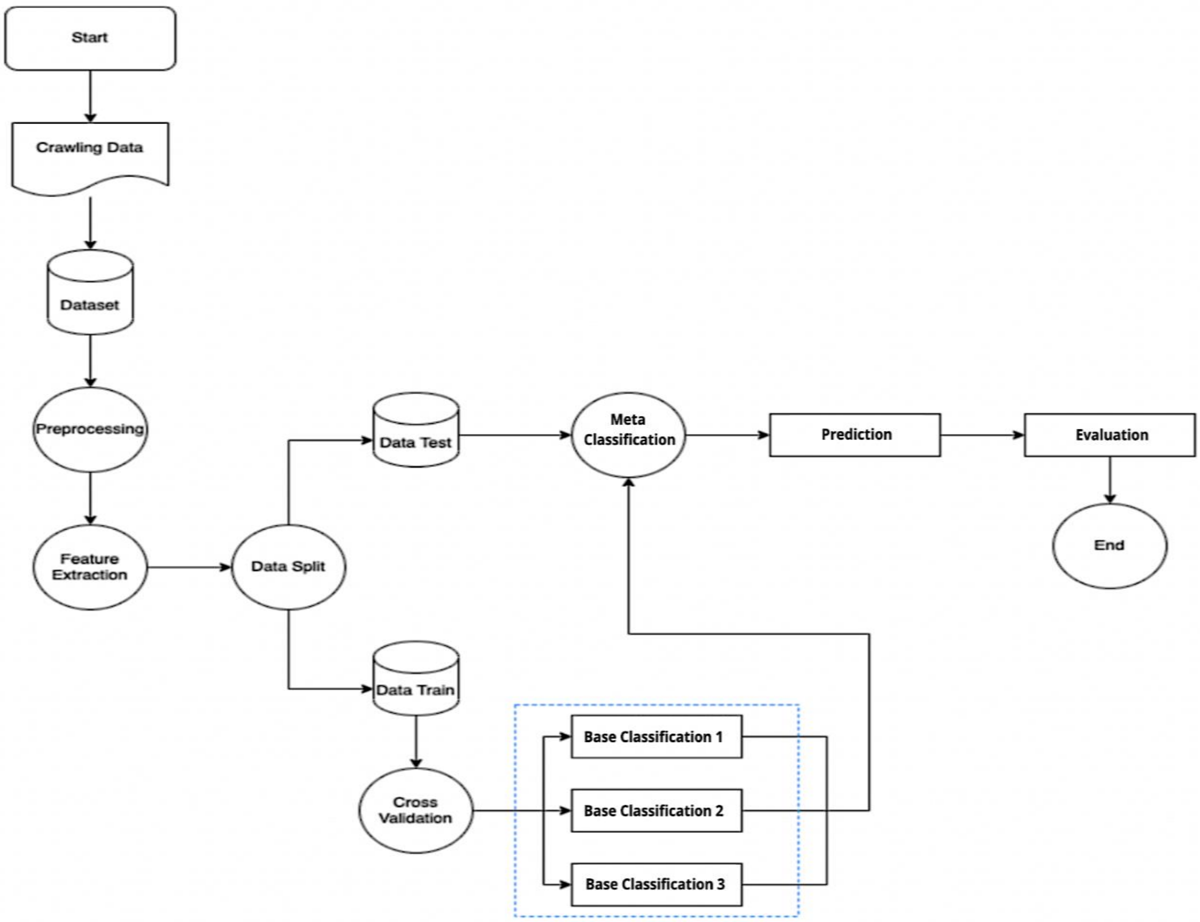

Figure 1. System Design

2.2 Dataset

Dataset for this study was taken from Twitter using the Twitter API *(Application Programming Interface)* with the *tweepy library*. Tweets taken are tweets that are mentioned by Twitter users to the @IndiHomeCare and @FirstMediaCares accounts. Tweet data retrieval was carried out in March 2022, with a total of 6.962 tweets collected for the Indihome dataset and 8.089 tweets for the First Media dataset. The next stage is to manually label each tweet with the provision of three sentiments, namely positive, neutral, or negative. The final result of the sentiment on each tweet is the decision of three people. Table 1 shows sample tweets from both datasets along with sentiment results.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

Table 1. Example of a Dataset

Provider	*Tweet*	Sentiment
IndiHome	@IndiHomeCare abangnya langsung dateng, bintang 5	Positive
	@IndiHomeCare Saya berlangganan dan membayar untuk Internet Speed 30 Mbps tapi saya cek kok cuma 21.9 Mbps , tolong dong dibenerin speednya	Negative
	@IndiHomeCare min cek dm	Neutral
First Media	@FirstMediaCares Sudah normal kembali. Terima kasih atas atensi nya @FirstMediaCares	Positive
	@FirstMediaCares ini ada kendala apalagi yaa~? Pulang2 internet mati. Tapi pas dicek di cek.firstmedia tdk ada gangguan. Mohon perhatian dan penjelasannya dong.. idpel:11480976	Negative
	@FirstMediaCares sudah saya dm detailnya	Neutral

2.3 *Preprocessing*

Preprocessing is done to clean raw data into data that is ready to be used in the next process, namely feature extraction using TF-IDF [5]. The *preprocessing* has several stages: Case *Folding*, *Remove Punctuation*, *Tokenization*, *Stopword*, and *Stemming*.

Case Folding is the process of changing capital letters to lowercase [6], *Remove Punctuation* to clean documents from characters that cannot be read by the system [7], *Tokenization* to separate sentences into words [8], *Stopword Removal* is a process removing words that have no important meaning [8], and *Stemming* is the process of eliminating affixes for each word [9]. Examples of the results of each pre-processing stage can be seen in tables 2 and 3.

Table 2. Example of tweet *Preprocessing* @IndiHomeCare

Tweet	*Preprocessing* Stage	results
@TelkomIndonesia @IndiHomeCare ini mau putus wifi tapi kok tlpon rumah ikutan harus diputus ? \n BENERAN INI SEPERTI INI PROSEDURNYA ????????'	*Folding Case*	@telkomindonesia @indihomecare ini mau putus wifi tapi kok tlpon rumah ikutan harus diputus ? \n beneran ini seperti ini prosedurnya ????????'
	Remove Punctuation	telkomindonesia indihomecare ini mau putus wifi tapi kok tlpon rumah ikutan harus diputus beneran ini seperti ini prosedurnya
	Tokenization	['telkomindonesia', 'indihomecare', 'ini', 'mau', 'putus', 'wifi', 'tapi', 'kok', 'tlpon', 'rumah', 'ikutan', 'harus', 'diputus', 'beneran', 'ini', 'seperti', 'ini', 'prosedurnya']
	Stopword Removal	['mau', 'putus', 'wifi', 'kok', 'tlpon', 'rumah', 'ikutan', 'diputus', 'beneran', 'prosedurnya']
	Stemming	['mau', 'putus', 'wifi', 'kok', 'tlpon', 'rumah', 'ikut', 'putus', 'beneran', 'prosedur']

Table 3. Example of tweet *Preprocessing* @FirstMediaCares

Tweet	*Preprocessing* Stage	Result
Jaringan internet mati total di daerah otista 1a jaktim. Mohon bantuannya untuk segera di cek. Terima kasih @FirstMediaCares'	*Case Folding*	jaringan internet mati total di daerah otista 1a jaktim. mohon bantuannya untuk segera di cek. terima kasih @firstmediacares'
	Remove Punctuation	jaringan internet mati total di daerah otista a jaktim mohon bantuannya untuk segera di cek terima kasih firstmediacares
	Tokenization	['jaringan', 'internet', 'mati', 'total', 'di', 'daerah', 'otista', 'a', 'jaktim', 'mohon', 'bantuannya', 'untuk', 'segera', 'di', 'cek', 'terima', 'kasih', 'firstmediacares']
	Stopword Removal	['jaringan', 'internet', 'mati', 'total', 'daerah', 'otista', 'jaktim', 'mohon', 'bantuannya', 'segera', 'cek', 'terima', 'kasih']

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

Stemming	['jaringan', 'internet', 'mati', 'total', 'daerah', 'otista', 'jaktim', 'mohon', 'bantuan', 'segera', 'cek', 'terima', 'kasih']

Figure 3 and 4 shows the number *tweets* from each sentiment in the IndiHome and First Media datasets after passing the *preprocessing*. The IndiHome dataset, which previously contained 6,962 *tweets*, became 5,179 *tweets* and the First Media dataset, which previously contained 8,089 *tweets*, became 7,940 *tweets*. This study uses two test scenarios: testing without *oversampling* and testing with *oversampling* on the dataset. Oversampling was carried out using the *Synthetic Minority Oversampling Technique* (SMOTE) to balance the number of classes in the two datasets [10]. The number of tweets in the IndiHome dataset after oversampling became 7,335 tweets and the number of tweets in the First Media dataset after oversampling became 15,327 tweets.

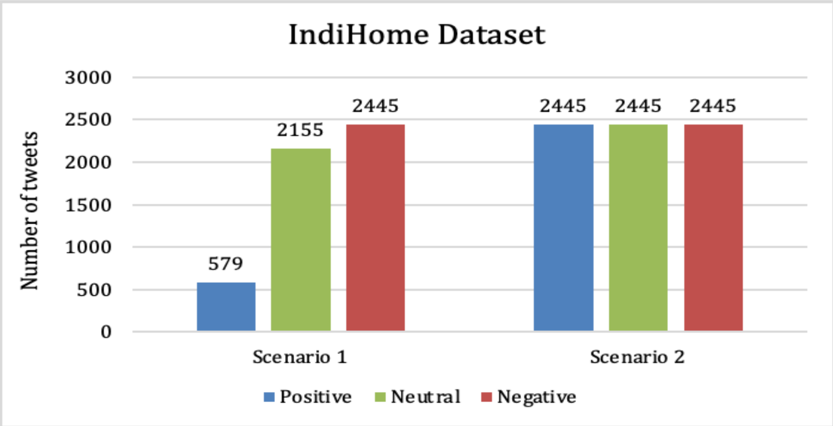

Figure 3. Distribution IndiHome dataset class before and after *oversampling*

Figure 3 shows the data distribution of each sentiment class on the Indihome dataset. The first scenario is a scenario without oversampling, and scenario 2 is a scenario with oversampling.

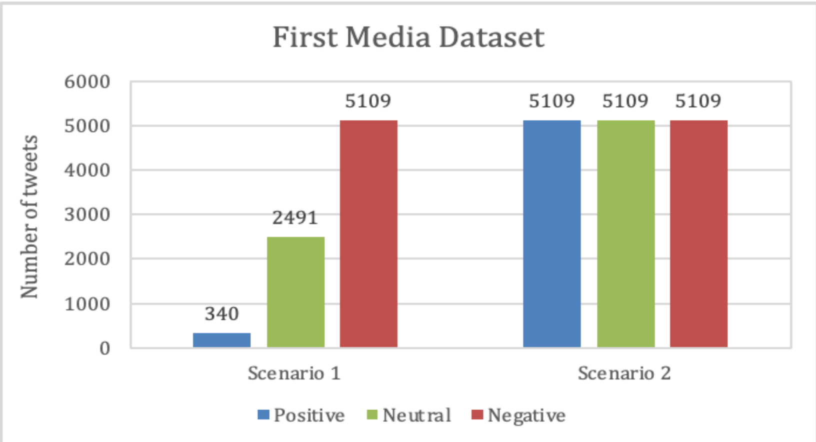

Figure 4. Distribution of First Media dataset classes before and after *oversampling*

Figure 4 shows the data distribution for each sentiment class in the First Media dataset. The first scenario is a scenario without oversampling, and scenario 2 is a scenario with oversampling.

2.3 TF-IDF *(Term frequency-inverse document frequency)*

Term frequency-inverse document frequency is a weighting stage by calculating the value of the *term frequency* and the occurrence of a word in the entire document [11]. *Term frequency* is a word in a particular document [12], where the more often the word appears, the greater the value of *term frequency* [13]. The TF-IDF method counts how many words are in a text document [14].

$$W_{ij} = tf_{ij} \, x \, log \left(\frac{N}{df_i}\right) \quad (1)$$

Equation (1) W_{ij} is the calculation weight of i which is a word from a document j , tf_{ij} is the frequency i on document j , df_i is the number of documents containing the word i and N is the number of documents. After the TF-IDF process, the next step is to split the data, where the IndiHome and First Media datasets are divided into two parts, namely data train and data test. The data train used is 80%, and the data test is 20%.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

2.4 *Ensemble Stacking*

Ensemble Stacking has two stages in the learning process, namely in the first stage each base classifier is trained using a dataset that has been weighted by TF-IDF to generate its respective predictions and the second stage is meta-classification taking the prediction results from the three base classifiers and returning Final predictions depend on the values of the three base classifiers [15].

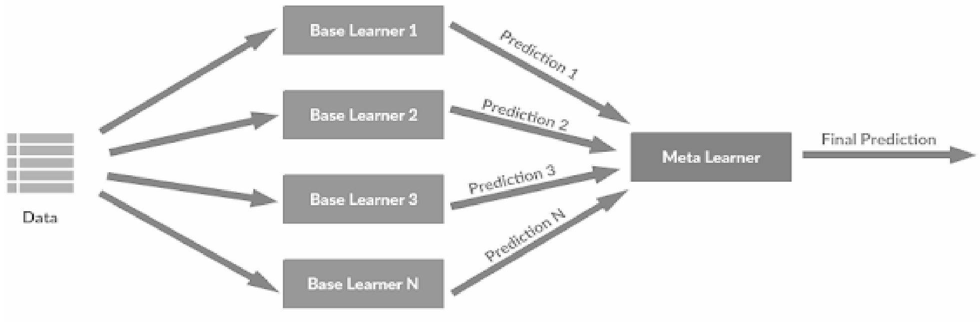

Figure 5. Ensemble Stacking

In this process, three base classifiers are used: Naive *Bayes*, *K-Nearest Neighbor*, and *Decision Tree*. In contrast, the meta classification used in the second stage of *Ensemble Stacking* is *Logistic Regression*.

3. RESULTS AND DISCUSSION

The first stage in this research is to take data from Twitter to get a dataset, then label each tweet manually by three people. Data that has been labeled will be cleaned at the preprocessing stage so that the data can be used optimally. After the data has passed the preprocessing stage, feature extraction is performed using TF-IDF, then the data is divided into test data and training data. After feature extraction, the two datasets were classified with two scenarios to get the best accuracy results using the Ensemble Stacking Method with the base classifier Naïve Bayes, K-Nearest Neighbor, Decision Tree, and Logistic Regression as meta classification.

3.1 First Scenario Testing

The first test scenario is a test without oversampling the dataset. The test method used is Ensemble Stacking with three base classifiers, namely Naive Bayes, K-Nearest Neighbor, Decision Tree and the meta classifier used is Logistic Regression. Tests were carried out without hyperparameter tuning and with hyperparameter tuning.

3.1.1 Cross Validation Accuracy

Cross Validation Accuracy displays the predicted accuracy of all base classifiers. Table 4 shows the cross validation accuracy of the IndiHome dataset without oversampling, K-Nearest Neighbor provides the lowest prediction accuracy at 66.2%, while Naive Bayes provides the highest accuracy at 73.7%.

Table 4. Cross Validation Accuracy on the IndiHome dataset without oversampling

Naive Bayes	K-Nearest Neighbor	Decision Tree
0.737	0.662	0.727

Table 5 shows the Cross Validation Accuracy of the First Media dataset without oversampling, K-Nearest Neighbor provides the lowest prediction accuracy at 69.1% while Naive Bayes provides the highest accuracy at 75.9%.

Table 5. Cross Validation Accuracy on First Media dataset without oversampling

Naive Bayes	K-Nearest Neighbor	Decision Tree
0.759	0.691	0.755

3.1.2 Evaluation Model Without Hyperparameter Tuning

The classification process uses the Naive Bayes classification model, K-Nearest Neighbor, Decision Tree, and Ensemble Stacking. Tables 6 and 7 are the results of the performance of each classification in the IndiHome and First Media datasets.

Table 6 shows the results of the three base classifiers and Ensemble Stacking on the IndiHome dataset without oversampling. Naive Bayes obtained the highest accuracy value among other base classifiers, which was 79.05%, while K-Nearest Neighbor obtained the lowest accuracy value among other base classifiers, which was 71.81%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 84.94%.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

Table 6. Classification performance on the IndiHome dataset without oversampling

	Naive Bayes	K-Nearest Neighbor	Decision Tree	Ensemble Stacking
Accuracy	0.790541	0.718147	0.780888	**0.849421**
Precision	0.780655	0.757320	0.781060	0.849338
Recall	0.790541	0.718147	0.780888	0.849421
F1 Score	0.773466	0.702547	0.780576	0.849351

Table 7 shows the classification results of the three base classifier and Ensemble Stacking on the First Media dataset without oversampling. Naive Bayes obtains the highest accuracy value among other base classifiers, which is 77.64% while K-Nearest Neighbor obtains the lowest accuracy value among other base classifiers, which is 71.50%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 81.98%.

Table 7. Classification performance on the First Media dataset without oversampling

	Naive Bayes	K-Nearest Neighbor	Decision Tree	Ensemble Stacking
Accuracy	0.776448	0.715058	0.765113	**0.819899**
Precision	0.745951	0.748023	0.772523	0.818034
Recall	0.776448	0.692695	0.76511881	0.7651188
Score	0.692695	0.76511881	0.870.77881	0.692695

Table 8 shows the classification results of the three base classifiers and Ensemble Stacking on the IndiHome dataset with hyperparameter tuning. Decision Tree obtains the highest accuracy value among other base classifiers, which is 78.86% and K-Nearest Neighbor obtains the lowest accuracy value among other base classifiers, which is 74.13%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 85.52%, but in the Ensemble Stacking hyperparameter tuning, the hyperparameter of Naive Bayes is not used because the final accuracy value of Ensemble Stacking has decreased.

Table 8. Hyperparameter tuning performance on IndiHome dataset without oversampling

Classifier	Hyperparameter	Value	Accuracy	Precision	Recall	F1 Score
Naive Bayes HT	Alpha	0.1	0.746139	0.779923	0.746139	0.753313
	binarize	None				
K-Nearest Neighbor HT	n_neighbors	7	0.741313	0.752766	0.741313	0.735554
Decision Tree HT	Criterion	gini	0.788610	0.788516	0.788610	0.788254
	min_samples_leaf	1				
Ensemble Stacking HT			**0.855212**	0.855212	0.855212	0.855212

Table 9 displays the classification results of the three base classifiers and Ensemble Stacking on the First Media dataset with hyperparameter tuning. Naive Bayes obtains the highest accuracy value among other base classifiers, which is 80.22% and K-Nearest Neighbor obtains the lowest accuracy value among other base classifiers, which is 72.04%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 82.49%.

Table 9. Hyperparameter tuning performance on First Media dataset without oversampling

Classifier	Hyperparameter	Value	Accuracy	Precision	Recall	F1 Score
Naive Bayes HT	alpha	0.1	0.802267	0.800292	0.802267	0.793083
	binarize	none				
K-Nearest Neighbor HT	n_neighbors	7	0.720403	0.767088	0.720403	0.730189
Decision Tree HT	criterion	gini	0.772040	0.770058	0.772040	0.770777
	min_samples_leaf	5				
Ensemble Stacking HT			**0.824937**	0.826224	0.824937	0.825481

3.2 Second Scenario Testing

The second test scenario is a test with oversampling the dataset. The test method used is Ensemble Stacking with three base classifiers, namely Naive Bayes, K-Nearest Neighbor, Decision Tree and the meta classifier used is Logistic Regression. Tests were carried out without hyperparameter tuning and with hyperparameter tuning.

3.2.1 Cross Validation Accuracy

Cross Validation Accuracy displays the final prediction accuracy of all base classifiers. Table 10 shows the Cross Validation Accuracy of the IndiHome dataset that has been oversampled, K-Nearest Neighbor provides the lowest accuracy at 65.4% while Naive Bayes provides the highest accuracy at 86.9%.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

Table 10. Cross Validation Accuracy on the IndiHome dataset with oversampling

Naive Bayes	K-Nearest Neighbor	Decision Tree
0.869	0.654	0.818

In table 11 shows the Cross Validation Accuracy of the First Media dataset after oversampling, K-Nearest Neighbor gives the lowest accuracy at 68% while Naive Bayes and Decision Tree provide the same highest accuracy at 87.3%.

Table 11. Cross Validation Accuracy on First Media dataset with oversampling

Naive Bayes	K-Nearest Neighbor	Decision Tree
0.873	0.680	0.873

3.2.2 Evaluation Model Without Hyperparameter Tuning

The classification process uses Naive Bayes classification, K-Nearest Neighbor, Decision Tree, and Ensemble Stacking. Tables 12 and 13 are the results of the performance of each classification in the IndiHome and First Media datasets.

Table 12 shows the results of the three base classifiers and Ensemble Stacking on the IndiHome dataset after oversampling. Naive Bayes obtained the highest accuracy value among other base classifiers, which was 86.77%, while K-Nearest Neighbor obtained the lowest final accuracy value among other base classifiers, which was 65.43%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 87.73%.

Table 12. Classification performance on the IndiHome dataset with oversampling

	Naive Bayes	**K-Nearest Neighbor**	**Decision Tree**	**Ensemble Stacking**
Accuracy	0.867757	0.654397	0.826176	**0.877301**
Precision	0.866940	0.778407	0.825599	0.876547
Recall	0.867757	0.654397	0.826176	0.877301
F1 Score	0.866648	0.564067	0.825619	0.876884

Table 13 shows the results of the three base classifiers and Ensemble Stacking on First Media dataset after oversampling. Naive Bayes obtains the highest accuracy value among other base classifiers, which is 86.88%, while K-Nearest Neighbor obtains the lowest accuracy value among other base classifiers, which is 67.54%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 90.90%.

Table 13. Classification performance on the First Media dataset with oversampling

	Naive Bayes	K-Nearest Neighbor	Decision Tree	Ensemble Stacking
Accuracy	0.868885	0.675473	0.861709	**0.909002**
Precision	0.870401	0.786872	0.861125	0.908683
Recall	0.868885	0.675473	0.861709	0.909002
F1 Score	0.867162	0.590156	0.861032	0.908580

Table 14 shows the results of the three base classifiers and Ensemble Stacking on the IndiHome dataset that has been oversampled. This test uses hyperparameter tuning to improve the accuracy of the classification model. Naive Bayes obtained the highest accuracy value among other base classifiers, which was 87.59% and K-Nearest Neighbor obtained the lowest accuracy value among other base classifiers, which was 68.91%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 88.27%.

Table 14. Hyperparameter tuning performance on IndiHome dataset with oversampling

Classifier	Hyperparameter	Value	Accuracy	Precision	Recall	F1 Score
Naive Bayes HT	alpha	0.1	0.875937	0.874194	0.875937	0.874810
K-Nearest Neighbor HT	n_neighbors	1	0.689162	0.795204	0.689162	0.622391
Decision Tree HT	Criterion min_samples_leaf	Gini 1	0.816633	0.816109	0.816633	0.816040
Ensemble Stacking HT			**0.882754**	0.881722	0.882754	0.882168

Table 15 shows the results of the three base classifiers and Ensemble Stacking on the First Media dataset that has been oversampled. This test uses hyperparameter tuning to improve the accuracy of the classification model. Naive Bayes obtained the highest accuracy value among other base classifiers, which was 89.49% and K-Nearest Neighbor obtained the lowest accuracy value among other base classifiers, which was 69.99%. Ensemble Stacking using Logistic Regression as a meta classification produces the highest accuracy value among all base classifiers, which is 92.56%.

Building of Informatics, Technology and Science (BITS)
Volume 4, No 2, September 2022 Page: 924−931
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v4i2.1969

Table 15. Hyperparameter tuning performance on First Media dataset with oversampling

Classifier	Hyperparameter	Value	Accuracy	Precision	Recall	F1 Score
Naive Bayes HT	alpha	0.01	0.894977	0.895307	0.894977	0.893979
K-Nearest Neighbor HT	n_neighbors	1	0.699935	0.798670	0.699935	0.636519
Decision Tree HT	Criterion min_samples_leaf	Gini 1	0.862035	0.861415	0.862035	0.861423
Ensemble Stacking HT			**0.925636**	0.925720	0.925636	0.925536

4. CONCLUSION

This study conducted a sentiment analysis on Twitter social media towards internet providers IndiHome and First Media using the TF-IDF feature extraction and the Ensemble Stacking. The base classifier used is Naive Bayes, K-Nearest Neighbor, and Decision Tree, while the meta classification used is Logistic Regression. There are two scenarios carried out in this study, the first scenario is testing without oversampling and the second scenario is testing with oversampling on the dataset. The results of this study indicate that the final accuracy value of Ensemble Stacking is higher than Naive Bayes, K-Nearest Neighbor, and Decision Tree. In the first scenario, Ensemble Stacking using hyperparameter tuning produces an accuracy value of 85.52% for the IndiHome dataset and 82.49% for the First Media dataset. There is an increase in the absolute accuracy results in the second scenario, Ensemble Stacking using hyperparameters tuning obtained an accuracy value of 88.27% for the IndiHome dataset and 92.56% for the First Media dataset. The second scenario produces higher accuracy because oversampling has been carried out so that the number of classes in the data**.** From the results of the two scenarios, it can be concluded that Ensemble Stacking gives better results when compared to Naive Bayes, K-Nearest Neighbor, and Decision Tree.

REFERENCES

[1] Marsya Nabila, “APJII Survey: Internet Users In Indonesia Reached 171.17 Million Throughout 2018,” Dailysocial.Id, 2019. Https://Dailysocial.Id/Post/Pengguna-Internet-Indonesia-2018 (Accessed May 28, 2022).
[2] S. Rizkia, E. Budi, S. S. Si, D. Puspandari, And M. Pd, “Analisis Sentimen Kepuasan Pelanggan Terhadap Internet Provider Indihome Di Twitter Menggunakan Metode Decision Tree Dan Pembobotan TF-IDF,” E-Proceeding Of Engineering, Vol. 6, Pp. 9683–9693, 2019.
[3] A. Nisa, E. Darwiyanto, And I. Asror, “Analisis Sentimen Menggunakan Naive Bayes Classifier Dengan Chi-Square Feature Selection Terhadap Penyedia Layanan Telekomunikasi,” E-Proceeding Of Engineering , Vol. 6, Pp. 8650–8658, 2019.
[4] A. Deviyanto, M. R. Didik Wahyudi, And T. Informatika UIN Sunan Kalijaga Yogyakarta Jl Marsda Adi Sucipto No, “PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR,” Jurnal Informatika Sunan Kalijaga), Vol. 3, No. 1, Pp. 1–13, 2018, [Online]. Available: Https://Twitter.Com/Search?L=Id&Q=AHY%20since%3A2017-01-01%20until%3A2017-01-
[5] J. A. Septian, T. M. Fahrudin, And A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF Dan K-Nearest Neighbor,” JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION, Pp. 43–49, 2019, [Online]. Available: Https://T.Co/9wloawpfd5
[6] F. Selva Jumeilah, “Penerapan Support Vector Machine (SVM) Untuk Pengkategorian Penelitian,” Rekayasa Sistem Dan Teknologi Informasi, Vol. 1, No. 1, Pp. 2580–0760, 2017, [Online]. Available: Http://Jurnal.Iaii.Or.Id
[7] N. S. Wardani, A. Prahutama, And P. Kartikasari, “ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA DENGAN KLASIFIKASI NAÏVE BAYES UNTUK MODEL BERNOULLI DAN MULTINOMIAL,” JURNAL GAUSSIAN, Vol. 9, Pp. 237–246, 2020, [Online]. Available: Https://Ejournal3.Undip.Ac.Id/Index.Php/Gaussian/
[8] R. N. Chory, M. Nasrun, And C. Setianingsih, “SENTIMENTAL ANALYSIS ON USER SATISFACTION LEVEL OF MOBILE DATA SERVICES USING SUPPORT VECTOR MACHINE (SVM) ALGORITHM,” 2019, [Online]. Available: Https://Github.Com/Jefferson-Henrique/Getoldtweets-Python.
[9] A. Fikri Zulfikar, “PENGEMBANGAN ALGORITMA STEMMING BAHASA INDONESIA DENGAN PENDEKATAN DICTIONARY BASE STEMMING UNTUK MENENTUKAN KATA DASAR DARI KATA YANG BERIMBUHAN,” UNIVERSITAS PAMULANG, Vol. 143, No. 3, Pp. 143–146, 2017.
[10] A. Fernández, S. García, F. Herrera, And N. V Chawla, “SMOTE For Learning From Imbalanced Data: Progress And Challenges, Marking The 15-Year Anniversary,” Journal Of Artificial Intelligence Research, Vol. 61, Pp. 863–905, 2018.
[11] K. H. Manguri, R. N. Ramadhan, And P. R. Mohammed Amin, “Twitter Sentiment Analysis On Worldwide COVID-19 Outbreaks,” Kurdistan Journal Of Applied Research, Pp. 54–65, May 2020, Doi: 10.24017/Covid.8.
[12] M. Farhan Muzakki And R. Febrian Umbara, “ANALISIS SENTIMEN MAHASISWA TERHADAP FASILITAS UNIVERSITAS TELKOM MENGGUNAKAN METODE JARINGAN SARAF TIRUAN DAN TF-IDF,” 2019.
[13] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, And S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF Dan Support Vector Machine,” JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY, Vol. 2, No. 1, 2022.
[14] S. Fransiska And A. Irham Gufroni, “Sentiment Analysis Provider By.U On Google Play Store Reviews With TF-IDF And Support Vector Machine (SVM) Method,” Scientific Journal Of Informatics, Vol. 7, No. 2, Pp. 2407–7658, 2020, [Online]. Available: Http://Journal.Unnes.Ac.Id/Nju/Index.Php/Sji
[15] R. Ananda Fitriansyah And Saparudin, “Penerapan Ensemble Stacking Untuk Klasifikasi Multi Kelas,” Vol. 2, No. 1, 2016, [Online]. Available: Http://Ars.Ilkom.Unsri.Ac.Id