Informatics and Software Engineering
p-ISSN/e-ISSN: 2988-2818/2988-2222
Homepage: https://sanscientific.com/journal/index.php/ise
1(1) 14-23 (2023)
DOI: https://doi.org/10.58777/ise.v1i1.55

Research Paper

Sentiment Analysis of Rising Fuel Prices on Social Media Twitter using the Naïve AlgorithmBayes Classifiers and AdaBoost

Imam Hendriyadi[1], Angela Febrianti Putri[2], Rahmawati[3], Nursofia Ningsih[4], Putri Rizki Widianti[5],Dedi Bi Saputra[6*]
[1,2,3,4,5,6] Information Systems, Nusa Mandiri University, Jakarta

Received: June 10, 2023; Accepted: June 25, 2023

Abstract
The government issued a policy of increasing the price of Indonesian fuel oil (BBM) in September 2022. This policy resulted from the war in Europe between Russia and Ukraine, which caused a surge in world oil prices because many respondents complained about the increase in fuel. This condition has caused much controversy or opinion among the public on social media, especially Twitter. With this phenomenon, sentiment analysis uses the Naïve Bayes classifiers algorithms to see how the public responds to government policies. The classification used in this sentiment analysisis Complaint or Not Complaint. Sentiment analysis of fuel rise on Twitter using Naïve Bayes classifier algorithm and AdaBoost Naïve Bayes classifier algorithm is applied to get the best classification by using hashtag tweets the increase in the price of fuel oil (BBM) which was taken 1000 tweets to be Used US a dataset. Data preprocessing consists of Text, Status, removal annotations, Remove hashtags, Remove urls, regexp, Indonesian stemming, and Indonesian stopword removal. The analysis results obtained an accuracy value of 70.69%, precision of 70.49%, recall of 71.45%, and AUC of 0.729 (good classification).

Keywords: Sentiments Analysis; Fuel Price hikes; Naïve Bayes Classifiers; Classification

Abstrak
Pemerintah mengeluarkan kebijakan menaikkan harga bahan bakar minyak (BBM) Indonesia pada September 2022. Kebijakan ini adalah hasil dari perang di Eropa antara Rusia dan Ukraina hal ini menyebabkan lonjakan harga minyak dunia karena banyak responden masyarakat yang mengeluhkan atas kenaikan BBM. Hal ini banyak menimbulkan kontroversi ataupun opini pada kalangan masyarakat di social media khususnya twitter. Dengan adanya fenomena tersebut, untuk melihat bagaimana tanggapan masyarakat terhadap kebijakan pemerintah maka dilakukan analisis sentimen menggunakan algoritma naïve bayes classifier. Klasifikasi yang digunakan pada analisis sentimen ini adalah Complaint atau Not complaint. Analisis sentimen kenaikan bahan bakar pada twitter menggunakan algoritma naive bayes classifier dan adaboost, algoritma naïve bayes classifier ini diterapkan untuk mendapatkan klasifikasi terbaik. Dengan menggunakan hashtag tweets Kenaikan harga bahan bakar minyak (BBM) yang diambil 1000 tweet untuk di jadikan dataset. Preprocessing data terdiri dari Text, Status, Remove annotations, Remove hastag, Remove url, Regexp, Indonesian stemming, Indonesian stopword removal. Hasil analisis tersebut didapatkan nilai accuracy 70,69%, precision 70,49%, recall 71,45%, dan AUC yang didapat sebesar 0,729 (good classification).

Kata Kunci: Analisis Sentimen, Kenaikan Harga BBM, Naïve Bayes Classifier

How to cite: Hendriyadi, I., Putri, A. F., Rahmawati, Ningsih, N., Widianti, P. R., Saputra, D. B., (2023). Sentiment Analysis of Rising Fuel Prices on Social Media Twitter using the Naïve Algorithm Bayes Classifiers and AdaBoost, *Informatics and Software Engineering (ISE) 1(1)*, 14-23

[*] Corresponding author: Dedi Dwi Saputra (dedi.eis@nusamandiri.ac.id)

1. Introduction

BBM (material burn oil): isa type of material burn (fuel) generated from the refinery (refining) crude oil (crude oil). Oil raw from the earth is processed in refining (refinery), especially Formerly to produce products oil (oil products), which are included in it are BBM. In addition to producing fuel, refining Crude oil produces a variety of products, including gas, such as naphtha, light sulfur wax residue (LSWR), and asphalt (Juliani et al., 2022).

In early September 2022, the government decided to issue a policy to increase the price of fuel oil (BBM) in Indonesia. This policy resulted from the war in Europe between Russia and Ukraine, which caused a surge in world oil prices. The economic downturn has had a major impact on oil prices, including in Indonesia. The Russian oil trade embargo triggered a supply shock that made prices higher in global markets (Kurniasih & Suseno, 2022). Fuel is currently a very important commodity for society, especially for the economic industry. Fluctuations in fuel prices affect the operational costs of people's daily activities. The policy of increasing fuel prices sparked controversy on social media, including on social media Twitter (Kurniasih & Suseno, 2022).

Therefore, the author tries to do a sentiment analysis of the increase in fuel oil (BBM) on social media Twitter. This study aims to look at public opinion/response, especially to Twitter users in Indonesia, towards policies that the government has decided after the increase in fuel prices. In this study, the authors used the Naïve Bayes and Adaboost classification algorithms to analyze sentiment because this algorithm aims as a classification method for the classification of complaints and not complaints. Data collection in this study was carried out by obtaining a dataset. Data is Tweets from the Twitter user community grouped according to positive, negative, or neutral labels using the Naive Bayes approach. (Saputra, Rahmad, 2022).

Previous research that has been carried out relates to problems faced by other authors with the title "Application of the Naïve Bayes Algorithm in the Analysis of Increases in Fuel Oil on Twitter" by Rahmad Aldi Saputra and Sejati Waluyo in 2022. The results of comparative research between training data and test data are mutually exclusive. Related to the comparison, 80-20 gets a True Accuracy value of 81.00%, 70-30 gets a True Accuracy value of 83.00%, and 60-40 gets a True Accuracy value of 77.50%. (Saputra, RA, & Waluyo, S., 2022).

Another research is "Analysis of Sentiments in Online Ojek Services Using the Naïve Bayes Method" (Nugroho et al., 2016). The test results show that the system can classify sentiments using Naïve Bayes with an accuracy of 80% based on 800 tweet data consisting of 300 training data and 500 test data, in the title "Twitter Sentiment Analysis Using Text Mining With the Naïve Bayes Classifier Algorithm," (Sudiantoro, AV, & Zuliarso, E., 2018). The results of testing using Naïve Bayes get an accuracy of 84%, which means that the performance of the Naïve Bayes classifier algorithm can classify text data very well. (Sudiantoro, AV, & Zuliarso, E., 2018).

2. Methods

The research method used is Naïve Bayes. Naïve Bayes is a data mining/machine learning method that aims to group or classify data into several classes by exploiting the probabilities or opportunities of the data. (Hidayah, NF, & Budiman, SN, 2022).

The framework in this study can be described in Figure 1 with the following explanation:

1. Data Collection from Twitter

The data collected is taken from Twitter, tweets with hashtags according to the topic raised, namely about rising fuel prices. The data is obtained by connecting the Twitter API through the developer's Twitter account, and the search is carried out using the rapid miner.

2. Processing Gata Framework

Gata Framework is a website using the PHP programming language to process text data (Setiawan et al., 2020). It can be concluded that the Gata framework is a framework for Indonesian text mining preprocessing that provides Indonesian stopword removal. Indonesian stemming, regular expression (regexp).

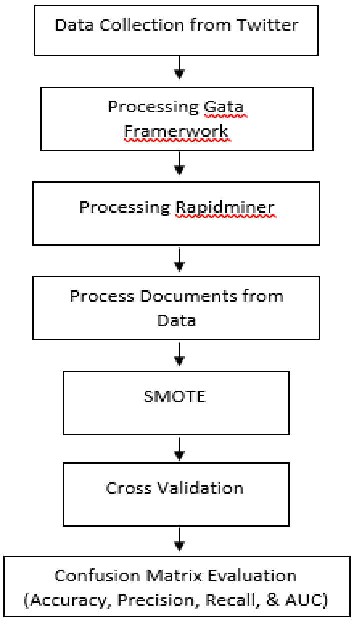

Figure 1. Framework sentiment analysis

Next, a separate sentence withthe words Which No needed with:

a. Annotations removal process For removing text Which owns annotation @ & # (Putra et al., 2022).
b. Transformation – Remove URLs: Fordelete URLs Which There is on say.
c. Regexp: Regular Expressions are processed to remove sign read and number so that results appearonly in words.
d. Indonesian Stemming: Words withthe same root word will be grouped.
e. Stopwords removal: processomit common wordsusually seen in amount Lots And considered No relevantlike "uwwwuu," "wkwk," "re," etc. will be deleted.

3. Processing Rapidminer

RapidMiner is an open-source based data mining application. It contains stand-alone applications for data analysis and mining machines, such as data loading, transformation, modeling, and visualization. (Nofitri, R., & Irawati, N., 2019).
Rapid Miner is also the software used for science and developed by a company with the same name. The main function of this application is to run a business or commercial and is often used as a means of research, education, and training (Muhammad et al., 2018).

5. Process Documents from Data

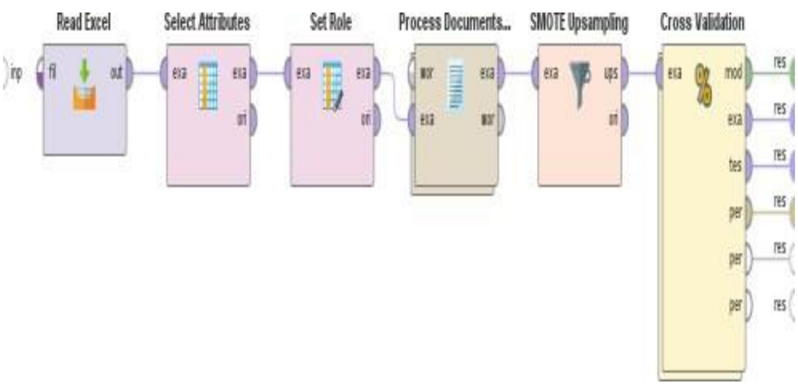

Figure 2. Processing Rapid Miner

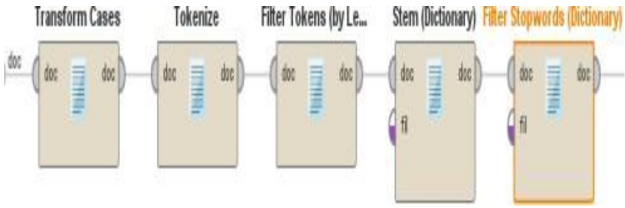

Figure 3. Process Documents From Data

In this process, several stages are applied to the dataset that has been imported, namely:

a. Transform Cases: The stage of converting all capital letters to lowercase (Hardi et al., 2021).
b. Tokenize: The stage of separating words in a sentence, such as words, phrases, symbols, or other meaningful elements (Ristyani Slamet et al., 2022).
c. Filter Tokens (by Length): The process of taking important words from the token results; in this process, words with a certain length will be deleted.
d. Stemming: The process of changing affixed words into root words.
e. Stopword: The process of eliminating words that are not important or irrelevant to the object of research.

5. SMOTE

SMOTE is a popular method for dealing with class imbalance. This technique synthesizes new samples from the minority class to balance the dataset by re-sampling the minority class samples (Sulistiyowati & Jajuli, 2020).

6. Cross Validation

Cross-validation (CV) is a methodstatistics that can be used to evaluate performance models or algorithmsWhere data is separated to become two subsets. That is the data learning process and data validation/evaluation. *Cross-validation* is sharing a dataset into two parts one part made data *training* And part Whichother made data testing. (Rilvani, Trisnawan & Santoso, 2019).

Cross Validation using the *Naïve Bayes* And *AdaBoost* algorithm can increase mark *accuracy, precision, recall, and AUCs.*

7. Confusion Matrix

This research results from the testing phasewill in evaluated using a table *Confusion Matrix* that is, *accuracy, precision, recall, and AUCs.*

Table 1. Confusion Matrix

Correct t Classification	Classified as +	-
+	true not complaint	False complaint
-	Falsecomplaint	true, not a complaint

The confusion matrix is a matrix that displays the results of a binary classification in a dataset. Several general formulas can be used in this matrix to calculate classification performance (Andika et al., 2019).

a. accuracy

Accuracy is the degree of closeness of the predicted value to the actual value (Prabowo & Fauzan, 2021).

b. Precision

Precision is the level of accuracy between the requested data and the predicted results provided by the model (Prabowo & Fauzan, 2021).

c. recall
   A recall is the model's success in redefining information (Prabowo & Fauzan, 2021).

8. Adaboost

Adaboos t used For classifydata on class respectively - respectively. *Adaboost*look for category class based on the value of the weight owned by the class. This process is repeated so the class has an updated mark. On *adaboost*, the weight value will continue to increase on each iteration from the weight mark Which Wrong on every iteration. *Adaboost* is a typical ensemble learning algorithm, and the results have high accuracy. (Novianti et al., 2022).

3 . Results and Discussion

Based on the results of the stages and research methods carried out, the results and discussion of the research are

3.1. Data Collection From Twitter

At this stage, the collection is carried outdata from social *media Twitter* based on, say, keys that become the background behind the problem taken. Process collection data: *Rapid Miner* is used as the tool by using *the search operator Twitter*. Data taken only speak Indonesia with amount record as much 1,000 data *tweets* Which use say keyincrease fuel price, on process This added *remove duplicate operator* for delete sentences or tweets that have a similarity. Here is the process of *crawling data on Twitter.*

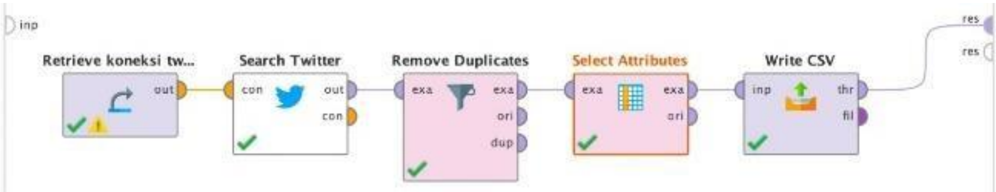

Figure 4. Process *Crawling Twitter Data*

The data generated in this process is saved in .csv format. On those files, the duplicate removal process will be carried out back to using the *remove duplicate feature* from *Microsoft Excel*, Which produces as many as 632 data *tweets*.

3.2. Data Labeling

After doing stage *crawl dataTwitter*, the next stage is *labeling* or labeling data in a manner manual by the writer. Labeling data on 632 data *tweets*, produce data with two classifications viz *complaint* and *not a complaint.* Following are the results from labeling data that has been done.

NO	Text	Kelas Pembanding
1	@sundingjalila Semoga sering2 di gelar ya pasar murah ini, karena sangat membantu warga dalam memenuhi kebutuhan pokoknya apalagi sekarang semuanya pada serba mahal seiring dengan kenaikan harga BBM	Not Complaint
2	@ZeboLady Bubarkan Partai Politik yang setuju dan mendukung kenaikan harga BBM..!!	Complaint
3	RT @TheReal_Rizkhy: @detikcom Pas ada demo kenaikan BBM, dia malah ngerayain ultah.	Not Complaint
4	RT @HeriMaspur: @niluhdjelantik @NasDem Mana ada pejuang rakyat termarjinalkan setuju dengan UU cilaka dan kenaikan BBM... Pejuang palsu	Complaint
5	Alokasikan Rp 10,41 Miliar untuk Subsidi Masyarakat Terdampak Kenaikan%BBM https://t.co/Sn3zCByIZx	Not Complaint
6	Dampak Kenaikan Harga BBM, Ojol dan Angkot Dapat Keringanan Pajak Nol Persen, Syaratnya Cukup%Mudah https://t.co/BZqs0sz0M3	Not Complaint
7	Apakah sudah beradaptasi dg kenaikan BBM dan inflasi yg mengiringi ???. Kok rasa2nya bisa diatasi oleh Kepolisian yak ? ??? lanjutkan Ndan....	Complaint
8	@niluhdjelantik @NasDem Memeluk rakyat termarjinalkan tapi gak suka ada orang demo kenaikan BBM, itu sih ANJING https://t.co/heVGevfaOR	Complaint
9	Pemkot Cimahi Lakukan Pendataan Pelaku UMKM Terdampak Inflasi atau Kenaikan Harga BBM https://t.co/OrZcdLMOFU	Not Complaint

Figure 5. Results *Labelling*

Furthermore, the results of the data that has been labeled will be next to stage pre- *Processing with Gata Framework.*

3.3. Processing Gata Framework

Then results labeling, which has been collected, will *proceed with Gata Framework*. These files shared become files containing 25 to 50datasets. Share this file that has existing limitations to servers on the *website Gata framework.* Following is the view from *processing* using the website from *Gata Frameworks.*

Figure 6. Processing with Gata Frtamework

Several stages are done on *the website Gata framework* that is:

a) *Annotations removal* on website Gata Frameworks, on stage This omitted the *mentioned mark* or @ on sentence so that later sentence in *tweets* No own sign @.
b) *# (Hashtags) removal*, process This done to remove hashtags Which there is on calm tweets
c) *transformation Remove URLs* on process This, URLs removed on sentence Which there are tweets inweb Gata Framework.
d) *Tokenization Regular expressions (regexp)* On stage This, done disappearance symbol - symbol Which there is in *tweets* so that Whichappear only just the sentence.
e) *Indonesian Stemming* On process This sentence Which contains affix removed so that say Which contains affixe s thebe the appropriate base wordstructure Language Indonesia say Whichaffix beginning like *mem,mem, in* and so forth on the website Gata Frameworks.
f) *Indonesian Stop Word Removal* stage This is the process of final *processing with Gata* Framework, i.e., deletion of words Which No relevant to analysis sentiment, likethe words *but, for, with, which,* on, and, say continue other on *the website Gata Frameworks.* Followingis the result of *Processing with Gata Framework* :

Figure 7. Processing with Gata Framework

3.4. Processing with Rapid Miner

At this stage, testing is carried out using the *tools Rapidminer.* Stages This is done with some process tests tried on the dataset to produce more accurate data for implementation *machine learn n ing.*

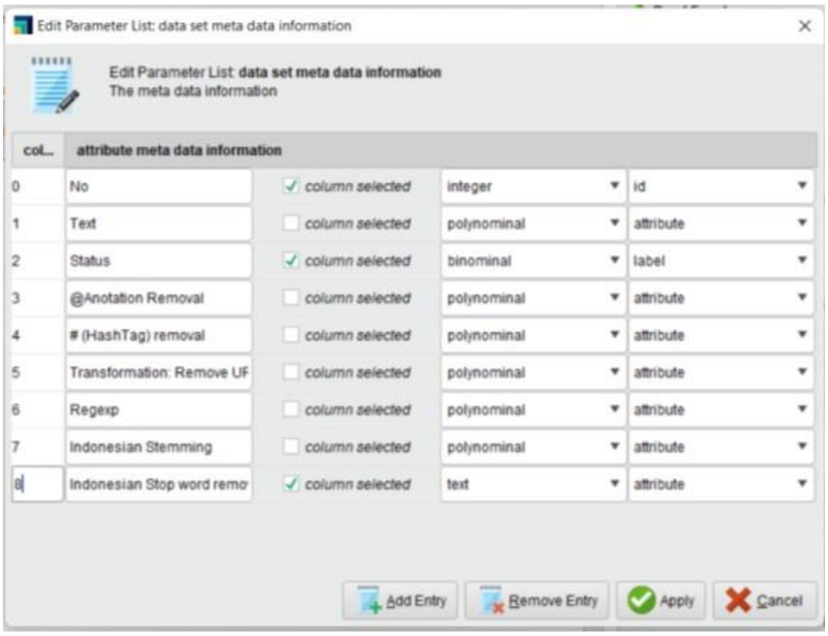

Figure 8. Imported Data

In stages, this is done by *importing* data already in *processing with the Gata Framework website application* to the application *RapidMiner* using operator *read* excel.

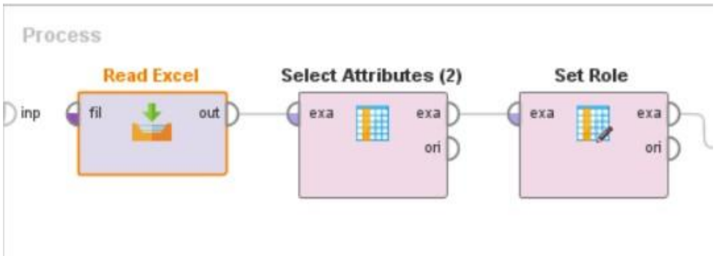

Figure 9. Operator attributes and Set Roles

This stage is done to throw away attributes Which No needed. Then do change on parameter roles become a state. This will make it possible. We choose sub gathering column for saved data.

3.5. Process Documents From Data

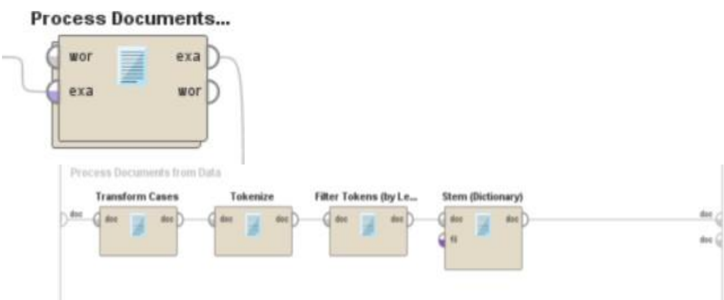

Figure 10. Processing Documents From Data

Stage furthermore Which covers stages Process Documents from Data asfollows:
a) Tokenize, the tokenizing process is the process of cutting the input string based on each word of its composition. This stage is carried out to separate word for word from a sentence text.
b) Transform Cases; this stage makes tweet data into all lowercase letters, from capital letters to ordinary lowercase letters.
c) Filter Token by Length (4.25), this stage aims to get words that are
d) between 4 and 25 characters long.
e) Steaming Additions, this stage aims to compare distributor data generated by removing data affixes such as "di," "me," "meng," "ter," "ber," "kan," "nya," and others.
f) Stopword the process of eliminating words that are not important or irrelevant to the object of research.

3.6. Cross Validation (CV)

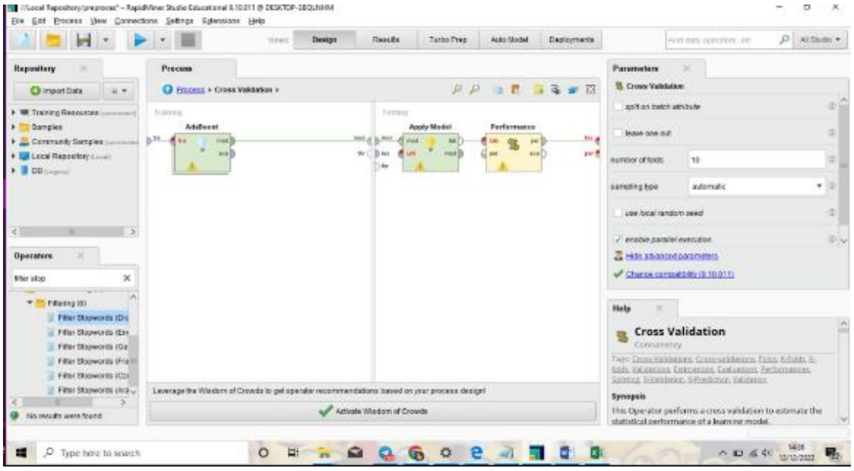

Figure 11. NB+AdaBoost cross-validation

The cross-validation operator is added. This operator functions as a testing tool with *Adaboost* and *Naïve Bayes operators*. Then the applied model and performance operators are added to the process to determine the level of *accuracy, precisions, recall,* and *AUC (Area Under Curve)* in the dataset.

3.7. Confusion Matrix
From the overall classification results obtained, this process is carried out for test results classification using the *confusion matrix method* witha number of data that have been tested. At this stage, the author looks for *accuracy, precision, recall,And AUC.*

accuracy: 70.69% +/- 5.57% (micro average: 70.70%)

	true Not Complaint	true Complaint	class precision
pred. Not Complaint	240	98	71.01%
pred. Complaint	103	245	70.40%
class recall	69.97%	71.43%	

Figure 12. Results accuracy Algorithm NB

Figure 12 shows that mark *accuracy* is big, 70.69%, with an error tolerance of 5.57%,with a *true complaint value of* 245 *records* and*a true not complaint* of 240 *records*.

precision: 70.49% +/- 4.61% (micro average: 70.40%) (positive class: Complaint)

	true Not Complaint	true Complaint	class precision
pred. Not Complaint	240	98	71.01%
pred. Complaint	103	245	70.40%
class recall	69.97%	71.43%	

Figure 13. Precision NB Algorithm Result

Figure 13 shows that mark precision is big 70,49 % with an error tolerance of 4.61%, with*true complaint* value 245 *records* and *true notcomplaint* 240 *records*.

recall: 71.45% +/- 9.47% (micro average: 71.43%) (positive class: Complaint)

	true Not Complaint	true Complaint	class precision
pred. Not Complaint	240	98	71.01%
pred. Complaint	103	245	70.40%
class recall	69.97%	71.43%	

Figure 14. Results recall Algorithm NB

Figure 14 shows that mark *recall* as big 71.45% with a toleranceerror of 9.47%, with a *true valueof complaint* 245 *records* And *true not complaint* 240 *records*.

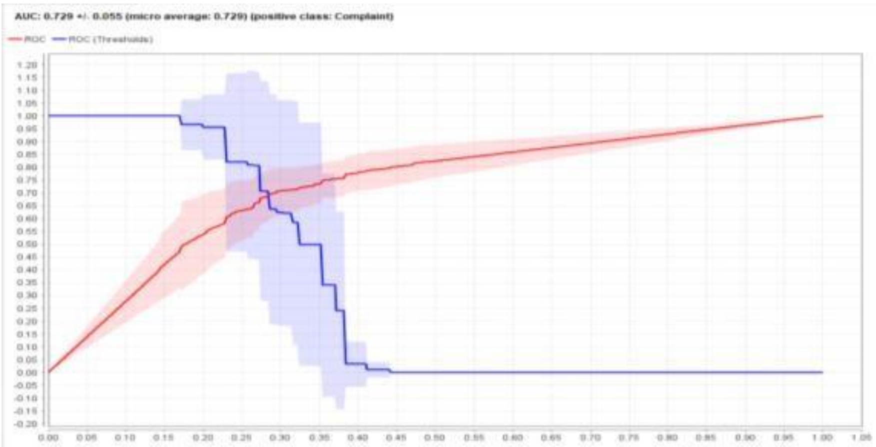

Figure 15. Chart ROC Algorithm Naïve Bayes

Based on the results, testing *performance* produces *curve* ROC like in Figure 15 And *mark* AUC, which obtained as big0.729 (*good classification*).

4. Conclusion

Testing the data crawl resultsfrom social media Twitter with queries #Rise in fuel prices with algorithmNaïve Bayes has succeeded. The approach uses the method text proven mining and Naïve Bayes algorithm effectively to classify perspectives tweet *Not Complaint* and *Complaint;* this thing supported with it generates markaccuracy of 70.69%, precision of 70.49%, recall of 71.45%, And AUC Which got a big 0.729 which was evaluated with a confusion matrix.

References

Andika, L. A., Amalia, P., & Azizah, N. (2019). *Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier*. *2*(1), 34–41.

Hardi, N., Alkahfi, Y., Handayani, P., Gata, W., Firdaus, M. R., Studi, P., Ilmu, M., Studi, P., Informasi, S., Pusat, K. J., & Khusus, D. (2021). *Analisis Sentimen Physical Distancing pada Twitter Menggunakan Text Mining dengan Algoritma Naive Bayes Classifier*. *10*, 131–138.

Hidayah, N. F., & Budiman, S. N. (2022). Penerapan Metode Naive Bayes Dalam Analisis Sentimen Aplikasi Sentuh Tanahku Pada Google Play. *Jati (Jurnal MahasiswaTeknik Informatika)*, *6*(2), 679-683.

Kurniasih, U., & Suseno, A. T. (2022). Analisis Sentimen TerhadapBantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan BakarMinyak (BBM). *Jurnal MediaInformatika Budidarma*, *6*(4), 2335–2340. https://doi.org/10.30865/mib.v6i4.49.2058

Muhammad, Z., Rahmadhani, R., Rizqifaluthi, H., & Yaqin, M. A. (2018). *Process Mining Akademik Sekolah Menggunakan RapidMiner*. *10*(2), 47–51.

Nugroho, D.G., Chrisnanto, Y. H., & Wahana, A. (2016). Analisis Sentimen Pada Jasa Ojek Online Menggunakan Metode NaiveBayes. Prosiding SNST Fakultas Teknik, 1(1).

Nofitri, R., & Irawati, N. (2019). IntegrasiMetode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, *6*(1), 35-42.

Novianti, N., Zarlis, M., & Sihombing, P. (2022). Penerapan Algoritma Adaboost Untuk Peningkatan Kinerja Klasifikasi Data Mining Pada Imbalance Dataset Diabetes. Jurnal Media Informatika Budidarma, 6(2), 1200-1206.

Prabowo, E., & Fauzan, A. C. (2021). *Implementasi Algoritma Naive Bayes Classifier Untuk Penentuan Kelayakan Penerima Beasiswa Peningkatan Prestasi Akademik (PPA) Bagi Mahasiswa*. *3*(2), 165–177.

Putra, A. P., Pratama, Y., & Krisnadi, E. K. (2022). *Text Mining untuk Sentimen Analisis dengan Metode Naïve Bayes , SMOTE , N-Gram dan*. *6* (September), 961–973.

Rilvani, E., Trisnawan, A. B., & Santoso, P. (2019). Penentuan KelulusanSiswa Yayasan Cerdas Bakti Pertiwidengan menggunakan Alghoritma Naive Bayes dan Cross Validation. Pelita Teknologi, 14(2), 145-153.

Ristyani Slamet, Gata, W., Novtariany, A., Hilyati, K., & Jariyah, F. A. (2022). *Analisis Sentimen Twitter Terhadap Penggunaan Artis Korea Selatan Sebagai Brand Ambassador Produk Kecantikan Lokal*. *5*.

Saputra, R. A., & Waluyo, S. (2022).Penerapan Algoritma Naive Bayes Dalam Analisis Kenaikan Bahan Bakar Minyak Pada Twitter. In *Seminar Nasional MahasiswaFakultas Teknologi Informasi (SENAFTI) Universitas Budi Luhur* (Vol. 1, No. 1, pp. 156-163).

Saputra, Rahmad, S. (2022). Penerapan Algoritma Naive Bayes DalamAnalisis Kenaikan Bahan Bakar Minyak Pada Twitter. *SeminarNasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Jakarta-Indonesia*, *September*,156–163. https://senafti.budiluhur.ac.id/index.php

Setiawan, K., Rahmatullah, B., Burhanuddin, Paryanti, A. B., & Fauzi, F. (2020). *Komparasi Metode Naive Bayes Dan Support Vector Machine Menggunakan Particle Swarm Optimization Untuk Analisis Sentimen Mobil Esemka*. *4* (3), 102–111.

Sudiantoro, A. V., & Zuliarso, E. (2018). Analisis sentimen twittermenggunakan text mining dengan algoritma Naïve Bayes Classifier.

Sulistiyowati, N., & Jajuli, M. (2020). Integrasi Naive Bayes Dengan Teknik Sampling SMOTE Untuk Menangani Data Tidak Seimbang. Nuansa Informatika, 14 (1), 34-37.

Yuliani, D., Saryono, Apriani, Maghfiroh, & Ro. (2022). Dampak Kenaikan Harga Bahan Bakar Minyak (BBM) Terhadap Sembilan Bahan Pokok(Sembako) Di Kecamatan Tambun Selatan Dalam Masa Pandemi. *Jurnal Citizenship Virtues, 2* (2), 320–326.