Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13729
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Analyzing Public Sentiment Towards BSI Service Disruptions Through X: Naïve Bayes Algorithm

Yudhistira[1)*], Aini Suri Talita[2)]
[1)2)]Universitas Gunadarma, Indonesia
[1)]yudhistira.elwinhaner@gmail.com, [2)]ainisuri@staff.gunadarma.ac.id

Submitted:Jun 5, 2024 | **Accepted**: July 2, 2024 | **Published**: July 10, 2024

Abstract: Disruptions to banking services can negatively affect customer trust and happiness, thus affecting the bank's reputation in the eyes of the public. Analysis of sentiment expressed on social media is very important because it can provide a direct picture of individual perceptions and responses in real time. This research aims to analyze public sentiment towards disruptions in Bank Syariah Indonesia (BSI) services through social media using the Naive Bayes algorithm. Through this analysis, the research seeks to understand the pattern of public responses and perceptions of BSI disruptions and evaluate the performance of the Naive Bayes algorithm in classifying sentiment on related tweet data. The data used came from specific social media platforms, where sentiment analysis was conducted by categorizing the data into positive, negative, and neutral categories. The research findings show that the sentiment analysis of the community towards BSI service disruptions through X social media platform shows a diverse pattern of responses and perceptions. This finding recorded 525 data points with negative sentiment, 325 data points with neutral sentiment, and 141 data points with positive sentiment. The research also compared the performance of the Naive Bayes algorithm with the Google Cloud Natural Language API, which showed an accuracy rate of 81.03%. This research provides valuable insights for Bank Syariah Indonesia in understanding public perception of BSI services on social media.

Keywords: Sentiment Analysis, Bank Syariah Indonesia, Social Media, Naïve Bayes Algorithm, Google Cloud Natural Language API

INTRODUCTION

Digital banking services have become an integral part of the daily life of modern society. People rely on digital banking services to conduct financial transactions, transfer funds, pay bills, and access other banking information. However, sometimes disruptions to digital banking systems can occur, which can negatively impact customer satisfaction and the company's image.

On May 8–11, 2023, Bank Syariah Indonesia (BSI) experienced a disruption that resulted in customers being unable to conduct financial transactions at branch offices, ATMs, or through the BSI Mobile application Zulfikar Hardiansyah (2023). During this period, people were able to express their sentiments, opinions, and experiences through social media platforms such as X.

In this context, analyzing people's sentiments towards the BSI disruption through social media is important to understand people's perceptions and responses to the incident. By analyzing people's sentiments, valuable insights can be gained into how customers responded to the disruption, how the company responded, and the impact on BSI's image.

One method used for sentiment analysis is the Naïve Bayes algorithm. The Naïve Bayes algorithm is a classification method based on Bayes' theorem with the assumption that the features used are mutually independent. This method has been widely used in sentiment analysis to classify text or opinions into positive, negative, or neutral categories.

In this context, the use of the Naïve Bayes algorithm in analyzing community sentiment related to BSI disruptions through social media can provide a deeper understanding of the sentiment that emerges, the pattern of community responses, and its impact on the image of BSI. As such, this research has significant relevance in the context of understanding and improving digital banking services and can provide insights for BSI in addressing disruptions and improving customer satisfaction.

*name of corresponding author

A number of previous studies have been conducted to analyze public sentiment on various relevant topics. The first research by Pratmanto et al. (2020) focused on sentiment analysis of e-commerce applications on the Google Play Store. This study aims to classify user reviews into positive or negative categories using the Naïve Bayes algorithm. By collecting data from user reviews, this study managed to achieve a high accuracy rate of 96.667. The second research by Pristiyono et al. (2021) is a sentiment analysis of the COVID-19 vaccine in Indonesia through data from Twitter. In this study, using the Naïve Bayes method, we successfully analyzed the sentiment of the Indonesian people towards the COVID-19 vaccine. The results showed that about 56% of the tweets analyzed were negative. The third study by Novendri et al. (2020) explores sentiment analysis of comments on movie trailers on YouTube, with a focus on television drama series trailers. Although the series had a high level of demand from viewers, there were also some negative comments. Through analysis using the Naïve Bayes algorithm, this study successfully classified the comments with an accuracy rate of 81% and achieved precision and recall of 74.83% and 75.22%, respectively.

Also, studies that perform sentiment analysis with different algorithms, such as the fourth study by Furqan et al. (2022), focus on sentiment analysis of the new normal policy in Indonesia using the K-Nearest Neighbor method. In this study, it was found that around 56% of the tweets analyzed showed negative sentiments. The last research by (Afdhal et al., 2022) leads to sentiment analysis of comments on YouTube related to islamophobia using the Random Forest algorithm. In this study, we successfully identified comments that were negative or contained islamophobia with an accuracy of 79%.

The problem identification in this study begins with an analysis of the disruption that occurred at Bank Syariah Indonesia (BSI), but is focused on emphasizing the analysis of public sentiment towards the disruption. The BSI disruption that occurred from May 8 to 11, 2023, included the inability of customers to conduct financial transactions through branch offices, ATMs, and the BSI Mobile application. This issue could cause dissatisfaction, inconvenience, and financial loss to customers, which is the main focus of the sentiment analysis. In addition, indications of cyberattacks and ransomware associated with the disruption raise concerns about system security and customer data protection. BSI's lack of communication response and transparency to the issue is also a concern, given the uncertainty that may be felt by customers and the general public. The negative impact on BSI's image and trust is also an important part of the sentiment analysis, as service disruptions and cyber-attacks can affect people's perceptions of this banking institution. Service quality and data protection are highlighted, where weaknesses in BSI's systems and infrastructure can result in the vulnerability of customer data. Thus, this study aims to analyze public sentiment towards Bank Syariah Indonesia (BSI) service disruptions through social media using the Naive Bayes algorithm. The goal of this study is to find out how people generally react and think about BSI disruptions and how well the Naive Bayes algorithm works at figuring out how people feel about related tweet data.

LITERATURE REVIEW

This literature review will examine prior works on sentiment analysis, specifically focusing on the implementation of the Naive Bayes algorithm and the diverse uses of sentiment analysis in different contexts. The purpose is to compare and connect these studies in order to obtain a full understanding of the present state of knowledge and identify any gaps that exist. The Naive Bayes technique is extensively employed in sentiment analysis across diverse domains. Alsaeedi & Khan (2019)conducted a study to investigate several methods of sentiment analysis applied to Twitter data in a broad sense. This study presents a comprehensive analysis of the merits and drawbacks of different techniques, such as Naive Bayes, for handling textual information extracted from social media platforms.

(Husen et al., 2023) public opinion on social media, particularly Twitter, has become a valuable source of information for companies like BSI Bank. Sentiment analysis can help BSI Bank understand public perceptions of their services. This study developed and applied machine learning algorithms (SVM, naïve Bayes, and logistic regression) to analyze sentiments towards BSI Bank on Twitter. Using tweet data from Kaggle, the study found that SVM achieved 0.88% accuracy, logistic regression 0.86%, and naïve Bayes 0.76%. The results indicated that SVM outperformed the other algorithms. The analysis also revealed a higher percentage of negative sentiments towards BSI Bank, highlighting significant public concern and dissatisfaction with the company's services.

In addition, Khomsah (2020) conducted research aimed at enhancing the performance of the Naive Bayes algorithm for analyzing the sentiment of hotel reviews. The findings demonstrated that Naive Bayes may be enhanced to enhance the accuracy of sentiment categorization, particularly in specialized domains like hotel reviews. (Turmudi Zy et al., 2022) conducted a study where they utilized the Naive Bayes method to examine the sentiment surrounding data breaches on the X platform. This study showcases the efficacy of the Naive Bayes method in classifying prevalent unfavorable attitudes related to data security concerns.

The study conducted by Manguri et al. (2020) utilized sentiment analysis to examine the worldwide dissemination of COVID-19 using data from Twitter. This research primarily examines global health challenges and explores how public sentiment on crisis situations might be assessed and analyzed using social media. (Firmansyah & Puspitasari, 2021) conducted a study to investigate the public's attitude towards COVID-19

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13729 p-ISSN : 2541-044X

immunization. This study employs the Naive Bayes algorithm to demonstrate that attitudes about vaccination are varied, with a significant number of negative viewpoints arising in response to material disseminated on social media.

(Apriani & Gustian, 2019) and Fitri et al. (2020) conducted research that demonstrated the use of Naive Bayes in analyzing the sentiment of popular applications like Tokopedia and Ruangguru. Both studies demonstrate the application of the Naive Bayes algorithm to analyze app user feedback with the purpose of identifying issues and enhancing services.

Although numerous research have utilized the Naive Bayes algorithm for sentiment analysis across different domains, there are still certain areas that require further investigation. The majority of studies concentrate on particular areas such as healthcare, hotel reviews, and digital apps, however there is a scarcity of research that specifically focuses on the Islamic banking sector. The issue of service disruption at Bank Syariah Indonesia (BSI) has not received extensive attention in the context of sentiment analysis using social media.

The objective of this research is to address the existing void by specifically examining individuals' sentiment concerning BSI service outages. This research aims to analyze the pattern of community responses and evaluate the performance of the approaches utilized by employing the Naive Bayes algorithm and comparing it with the Google Cloud Natural Language API. This research aims to offer novel perspectives to Bank Syariah Indonesia by examining public perceptions of their services on social media. Furthermore, the findings of this study might serve as a significant point of reference for other banks in effectively handling service crises and enhancing customer satisfaction by gaining a deeper comprehension of community attitude.This literature evaluation validates the importance and immediacy of this research within the existing literature, while emphasizing the anticipated advancements in the areas of sentiment analysis and Islamic banking service management.

METHOD

This study adopted a quantitative approach to analyze public sentiment towards BSI interference. The research will involve collecting data from tweets on social media platform X relating to BSI and the disruptions that occurred within a specific time span. Data will be collected through crawling techniques using the Tweet Crawler. Data collection will utilize NodeJS libraries, such as Tweet Harvest, to access and collect tweets relating to BSI and disruption. The data collected will include the tweet text, date of posting, user who sent the tweet, and additional relevant information. This process resulted in 991 tweets being collected by the Tweet Crawler.
The data collected will be analyzed using sentiment analysis methods to understand people's perceptions and responses to BSI disruptions.

Below, these steps will be explained in detail, along with a flowchart image of the data processing.

Fig. 1 Data Processing Flowchart

In the data processing stage, the incoming text goes through the following steps before it can be analyzed further:
1. Input: The process starts with the input of the text to be analyzed. This text can be a single sentence, a paragraph, or even a larger text document.

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13729 p-ISSN : 2541-044X

2. Tokenizing: After receiving the input, the text is divided into small units called tokens. These tokens can be words, phrases, or specific symbols. This process allows the computer to understand the text better.
3. Stopword Removal: After tokenization, the next step is to remove words that do not contribute significantly to sentiment analysis. Words like "and," "or," "at," and the like often do not have significant sentimental meaning. The removal of these words is referred to as stopword removal.
4. Score Calculation: Once tokenization and stopword removal are done, the next step is to calculate the sentiment score of the text. This calculation process involves counting the occurrences of words contained in the dictionary of positive, neutral, and negative words and multiplying the occurrences of these words by their weights to determine the final sentiment score.
5. Classification: Once the sentiment score is calculated, the text is classified into the appropriate sentiment category based on the score that has been obtained. These sentiment categories can be positive, neutral, or negative.

The Naive Bayes algorithm is a probabilistic classifier based on Bayes' Theorem with the assumption of independence between predictors. It is particularly effective for large datasets and is often used in text classification tasks such as sentiment analysis. The core principle involves calculating the posterior probability of a class given a set of features and then selecting the class with the highest probability.

Steps in Applying Naive Bayes in This Research:

1. Data Collection: Gather a dataset containing text data and corresponding sentiment labels (positive, negative, neutral).
2. Preprocessing: Clean the text data by removing noise such as stop words, and punctuation, and performing stemming or lemmatization.
3. Feature Extraction: Convert the text data into numerical features using techniques like TF-IDF (Term Frequency-Inverse Document Frequency).
4. Training the Model: Use the training dataset to calculate the prior probabilities of each class and the likelihood of each feature given a class.
5. Classification: Apply the Naive Bayes classifier to new, unseen data by computing the posterior probabilities and predicting the class with the highest probability.
6. Evaluation: Assess the performance of the classifier using metrics such as accuracy, precision, recall, and F1-score.

Here's a simple flowchart to visualize the Naive Bayes process:

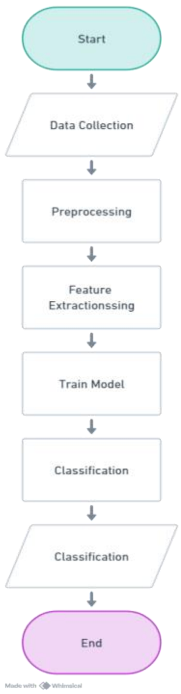

Fig 2. Naive Bayes Flowchart

RESULT

Results of the Data Collection Process

Data was collected through a crawling process using a NodeJS library called Tweet Harvest. This library allows researchers to access and collect data from X based on certain keywords and periods. The data retrieved includes

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13729 p-ISSN : 2541-044X

tweets containing user opinions, complaints, or experiences related to BSI interference during the specified period. Below is an image of the tweet crawling process using Tweet Harvest.

```
MINGW64:/c/Users/Yudhistira/Desktop/node-tweet

Yudhistira@MIS-YUDHISTIRA-NB MINGW64 ~/Desktop/node-tweet
$ npx tweet-harvest@0.0.35 -f 12-05-2023 -t 19-05-2023 -s "bank bsi" -l 1000

Welcome to the Twitter Crawler

This script uses Chromium Browser to crawl data from Twitter with *your* Twitter a
uth token.
Please enter your Twitter auth token when prompted.

Note: Keep your access token secret! Don't share it with anyone else.
Note: This script only runs on your local device.

? What's your Twitter auth token? [redacted]
√ What's your Twitter auth token? ****************************************
npm WARN optional SKIPPING OPTIONAL DEPENDENCY: fsevents@2.3.2 (node_modules\@play
wright\test\node_modules\fsevents):
npm WARN notsup SKIPPING OPTIONAL DEPENDENCY: Unsupported platform for fsevents@2.
3.2: wanted {"os":"darwin","arch":"any"} (current: {"os":"win32","arch":"x64"})
npm WARN Yudhistira No description
npm WARN Yudhistira No repository field.
npm WARN Yudhistira No license field.

+ @playwright/test@1.34.3
updated 1 package and audited 441 packages in 3.189s

3 packages are looking for funding
  run `npm fund` for details

found 42 vulnerabilities (6 moderate, 26 high, 10 critical)
  run `npm audit fix` to fix them, or `npm audit` for details

Opening twitter search page...


Filling in keywords: bank bsi since:2023-05-12 until:2023-05-19


Got some tweets, saving to file...
Your tweets saved to: C:\Users\Yudhistira\Desktop\node-tweet\tweets-data\bank_bsi_
29-05-2023_13-27-42.csv
Total tweets saved: 25

Got some tweets, saving to file...
Your tweets saved to: C:\Users\Yudhistira\Desktop\node-tweet\tweets-data\bank_bsi_
29-05-2023_13-27-42.csv
Total tweets saved: 48
```

Fig. 3 Scraping process using Tweet Harvest

The process yields 991 tweet results for the crawler.

Dataset Description

The collected dataset contains a set of tweets relevant to BSI disruptions in the period from May 8 to 11, 2023. Each tweet in the dataset comes with attributes such as tweet text, date and time of posting, and user information. The dataset covers a wide range of sentiments, including positive, negative, and neutral, which will form the basis for sentiment analysis using the Naïve Bayes algorithm and other relevant algorithms. Here is the dataset that was successfully collected from X.

id_str	full_text
1638104542914180000	No time to get ready for your meeting? Use xpression camera to be dressed for all occasions.
1628347462003090000	Buka Tabungan di BCA mobile sekarang, bisa dapet voucher Rp100 Ribu pilihan sendiri 😎
1659341247553900000	@bankbsi_id min tolong BSI Mobile nya kapan bisa, ini mahasiswa kesusahan mau makai duit
1659191934441910000	You'd think this would be the first thing the bank tell their customers after any suspicion of brea
1659345047396180000	@catwomaaannn @bankbsi_id Coba clear cache. Di gw bisa sih
1661908300592730000	📢 Exciting Announcement! 🎉\n\n🌍 We are thrilled to unveil ZTE 2022 Sustainability Report F
1659331153114050000	Gaes yang punya rekening BSI pindahin uangnya ke rekening bank lain yang menurut kalian puny
1659322930487260000	@bankbsi_id (klo kyk gini trus) User setia dr jaman BSM hruskah consider ke tetangga sebelah?
1659347053091360000	Anggota Komisi XI DPR RI Puteri Anetta Komarudin mendesak Bank Syariah Indonesia (BSI) seger
1659345770171210000	@saxifralip Password, Kata sandi dan Kode OTP kepada pihak manapun termasuk pihak BSI, sert
1659338115629330000	@flip_id ka..saya mau tarik koin ke rek bsi saya ko masih blm bisa aja..knp ya..kira² brp lama pros
1659336669919630000	@Tullllllllll @Fikri_syi @bankbsi_id skrg gmn kaaa? apa udh balik saldonya?
1659337561834420000	@lekyunghoo @Basuhsuh @bankbsi_id akhirnya kamu dtg ke CS langsung?
1659329586436660000	@sssilian @bankbsi_id Ke bank
1659345812038780000	@saxifralip Terima kasih telah menggunakan layanan kami, semoga berkah dan sehat selalu. Wa
1659345499009470000	@bankbsi_id Buat apa jaga pin dan password kalo data nasabah juga bocor dan dijual di dark wel
1659196539112430000	Bismillah\nYuu istirahat\nTong hilap BerWudhu sebelum tidur\n3 ayat terakhir al baqoroh\n3 x
1659326163632930000	@bankbsi_id halo, saya mau tutup rekening bsi gmn caranya?? saya sudah tanya di whatsapp dai

Fig. 4 Example sentiment dataset

By using the dataset collected from X, it is hoped that this research can provide a comprehensive understanding of the public's perception of BSI's disruption and how that sentiment can affect the company's image. This description of the data and dataset forms the basis for the next stage in the sentiment analysis process to be carried

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13729
e-ISSN : 2541-2019
p-ISSN : 2541-044X

out in this research.

Dictionary of Words

In sentiment analysis, the dictionary of words plays an important role in determining the polarity or sentiment of the text being analyzed. In this context, the dictionary of words is divided into three main categories: positive, negative, and neutral. Each category contains a list of words that represent a particular sentiment, which will be used as a reference when performing sentiment classification on the analyzed text.

The reference for this dictionary of words is obtained from the GitHub repository masdevid/ID-OpinionWords, which is a list of opinion words (positive and negative) in Indonesian. This dictionary was initiated by Liu (2022) with modifications and translations for Bahasa Indonesia.

Positive Words Dictionary

A positive word dictionary is a collection of words that have positive connotations or describe expressions of satisfaction, joy, or approval. Examples of words in this dictionary include "happy," "successful," "praiseworthy," and the like. The following is table 1, namely examples of these words, which will be a reference for recognizing positive sentiments in the text being analyzed.

Table 1. Examples of Positive Words

(^-^)	hakistimewa	palingdikenal
(^.^)	halal	rahmat
(^_^)	halus	rajin
acunganjempol	ideal	ramah
adaptif	idola	sabar
adil	imut	sah
bagus	jaminanmutu	salut
bahagia	janji	taat
Baik	jempol	tahanlama
cakap	kagum	uangkembali
cantik	karismatik	ulung
damai	karya	unggul
dapatdiandalkan	lancer	unik
dapatdicapai	layak	variasi
adukatif	lebihbaik	visioner
efektif	mahir	wah
efetivitas	maju	whooa
efisien	makmur	whooooa
fantastis	nikmat	wow
fasih	nyaman	yakin
favorit	optimal	yay
gagah	optimis	yeah
gembira	optimism	yihaa
giat	pahlawan	
hadiah	palingberuntung	

Negative Word Dictionary

In contrast, a negative word dictionary consists of a list of words that describe expressions of dissatisfaction, disappointment, or discomfort. Examples in Table 4.2 are words in this dictionary, including "sad," "frustrated," "disappointed," and the like. The use of this negative word dictionary helps in identifying negative sentiments in the text.

Table 2. Examples of Negative Words

:(	gakaruan	pahit
:*(	gapeka	palsu
:,(	habis	racun
abnormal	halangan	radikal
absurd	iblis	ragu
acak	jahat	ragu
badai	kabur	ramai
bahaya	kabut	sabar

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13729
e-ISSN : 2541-2019
p-ISSN : 2541-044X

balasdendam	kacau	sakit
cacat	lalai	salah
cacian	lamban	tabu
calo	maaf	tajam
deadline	mabuk	tandus
defensif	mahal	ulang
edan	naif	ultimatum
egois	najis	virus
ejek	nakal	vulgar
fanatisme	obsesif	wabah
frustasi	ocehan	yasilahkansaja
gajelas	omelan	yahudi

Neutral Word Dictionary

The neutral word dictionary includes words that do not have a strong sentiment polarity or words that are generally considered neutral in the context of sentiment analysis. The examples in Table 4.3 include common descriptive words such as "new," "latest," "information," and others. Neutral word dictionaries are used to deal with texts that do not imply a clear positive or negative sentiment.

Table 3. Examples of Neutral Words

;)	emotion	half
;o)	emphasize	stop
waspada	include	heavy duty
sepanjangwaktu	alluring	hefty
mengalegorisasikan	emosi	raksasa
Persekutuan	menekankan	memungur
aliansi	mengikutsertakan	sangar
kiasan	memikat	pertumbuhan
sindiran	seluruh	setengah
menilai	terasa	berhenti
penilaian	tukangsuap	tugasberat
anggapan	lantai	besardankuat
astronomis	meramalkan	tinggi
sikap	benteng	bertenagatinggi
rata-rata	terusterang	hm
sadar	sering	hmm
kesadaran	penuh	menghipnotis
bersertifikat	skalapenuh	ide
nyanyian	sepenuhnya	menyalakan
klaim	mendasar	berfilsafat
curhat	fundamental	praktis
efektif	didanai	berdoa
rumit	menggembleng	tekanan
perwujudan	gerakan	berdiri
menggerakan	kekuatan	kecenderungan

Data Preprocessing

Before carrying out sentiment analysis on the dataset collected from X, a data preprocessing stage is carried out to clean and prepare the data so that it is ready to be processed. Data preprocessing is an important step in text analysis, especially when dealing with data from social media platforms like X, which tend to be unstructured and contain a lot of noise. The following is an example of the sentiment that will be processed.

@bankbsi_id siap kak.. saaya percaya sama BSI dari dulu

Fig. 5 Examples of sentiments to be analyzed

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i3.13729 p-ISSN : 2541-044X

Tokenizing

The tokenizing stage aims to remove irrelevant or disturbing information, special characters, punctuation marks, and certain symbols, or replace capital letters with lowercase letters so that the data is more consistent. The following is an example of the resulting sentiment after the tokenizing process.

Stopword Removal

Stopwords are common words that often appear in language that do not have a specific meaning and do not make a significant contribution to text analysis. Examples of stop words in Indonesian are "and," "or," "yang," "of," and so on. Stopword removal is carried out to reduce data dimensions and increase analysis efficiency, so that only words that have important meaning are included in sentiment analysis. The following is an example of the resulting sentiment after the stopword removal process.

Fig. 6 Examples of sentiments that have gone through the stopword removal stage

After going through the data preprocessing stage, the clean and structured data is ready to be used to carry out sentiment analysis using the Lexicon method and the Naïve Bayes algorithm, gaining insight into sentiment and public responses to BSI digital banking disruptions via the X social media platform.

DISCUSSIONS

Utilizing Lexicon and Naive Bayes in Sentiment Analysis

The Naïve Bayes algorithm states the relationship between posterior probability and prior probability as well as the likelihood of an event. Mathematically, Bayes' theorem can be written as follows:

$$P(y|x) = \frac{P(x|y) * P(y)}{P(x)} \tag{1}$$

From equation 2 in the context of sentiment analysis, y is the sentiment category (positive, negative, or neutral), and x is the text to be classified. The goal of the Naïve Bayes algorithm is to calculate the posterior probability P(y|x) for each sentiment category y based on the features (words) contained in text x.

The basic assumption of Naïve Bayes is that each feature (word) in text x are independent of each other, although in reality some features may be related. Nevertheless, this assumption simplifies probability calculations and makes this algorithm efficient in text classification.

Implementation of Naïve Bayes in sentiment analysis involves the following steps:

1. Defining Likelihood: Next, the likelihood P(y|x) of each feature (words) in text x for each sentiment category y is calculated based on the distribution of words in each sentiment category in the dictionary of positive, negative, and neutral words.
2. Calculating Posterior Probability: The posterior probability P(y|x) of each sentiment category y is calculated based on Bayes' theorem by multiplying the prior probability P(y) by the likelihood P(y|x) of each feature in text x.
3. Classification: After getting the posterior probability for each sentiment category, text x is classified into the sentiment category with the highest probability.

In this research, the Naïve Bayes model will be implemented using the PHP programming language, namely the Laravel framework, by utilizing libraries or modules that support text processing and sentiment analysis. The data that has been previously processed and represented will be used as input for the Naïve Bayes model so that it can classify sentiment in public texts related to BSI interference in X.

bankbsiid siap kak saaya percaya sama bsi dari dulu

The Sentiment Analysis Program is being implemented.

The sentiment analysis program implemented in this research uses the Laravel framework as a development base. This framework was chosen because of its flexibility and ability to simplify web application development. In addition, this program uses a SQL Server database to store and manage tweet data to be analyzed.This program has the main function of carrying out sentiment analysis on tweet data taken from social media. Some of the main

bankbsiid siap kak saaya percaya bsi dulu

features of this program include data collection, data preprocessing, sentiment analysis, and visualization of results. Tweet data is collected using the Tweet Harvest library and stored in a SQL Server database. Next, the data goes through a preprocessing process to clean and prepare it for sentiment analysis using algorithms from the phpInsight

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13729
e-ISSN : 2541-2019
p-ISSN : 2541-044X

library. The results of sentiment analysis are visualized in the form of a pie chart to show the proportion of positive, negative, and neutral sentiment in the dataset.

Program Description

The sentiment analysis program implemented in this research uses the Laravel framework as a development base. This framework was chosen because of its flexibility and ability to simplify web application development. In addition, this program uses a SQL Server database to store and manage tweet data to be analyzed.

Program Functions and Features

This program has the main function of conducting sentiment analysis on tweet data taken from social media. This program includes several key features.

- Data Collection: The program can collect tweet data using the Tweet Harvest library, then save it into a SQL Server database for further processing.
- Data Preprocessing: Tweet data will go through a preprocessing process, including data cleaning, text normalization, stopword removal, and stemming using the yasirutomo/php-sentianalysis-id and JWHennessey/phpInsight libraries.
- Sentiment Analysis: After preprocessing, the program uses the sentiment analysis algorithm from the phpInsight library to classify each tweet into positive, negative, or neutral categories.
- Visualization of Results: The results of the sentiment analysis will be visualized using a pie chart to show the proportion of positive, negative, and neutral sentiment in the dataset.

Program Workflow

This program's workflow begins with collecting tweet data from X using the Tweet Harvest library. The data is then stored in a SQL Server database. After that, the data will go through a preprocessing process to be cleaned and prepared for sentiment analysis. The sentiment analysis process is carried out using an algorithm from the phpInsight library, and the results are visualized in a pie chart.

Program Output

The following is a screenshot of the display results from the sentiment analysis program that has been implemented:

#	Tweet	Hasil	Naive Bayes
1	#1 - @flip_id ka..saya mau tarik koin ke rek bsi saya ko masih belum bisa aja..knp ya..kira^2 brp lama prosesnya..apakah masih terkait error nya bsi atau saya tarik koin bukan di jam official bank nya ?	Negatif	0.61538461538405, 0.30769230769295, 0.076923076923006,
2	#2 - #Indonesia #News #MediaBerita Tautan : https://t.co/NnnnIJglRp Reporter : Muhammad Nasir BLANGPIDIE - Diduga karena jaringan Bank Syariah Indonesia (BSI) masih error, mengakibatkan keterlambatan pengiriman bahan bakar minyak (BBM) ke Stasiun Pengisian Bahan Bakar Umum (SPBU...	Negatif	0.49999999999962, 0.25000000000056, 0.24999999999981,
3	#3 - Ketahui apa itu saldo diblokir BSI, apakah saldo mengendap Rp 50 ribu bisa diambil? Yuk simak penjelasan dari Bank BSI. #bsierror #bsimobile https://t.co/OEzdoe8izr	Negatif	0.49999999999962, 0.25000000000056, 0.24999999999981,
4	#4 - Gila sih ini bank maunya apaaan dah error lama betul woiii, mau bayar ini itu gk bisa!!!! Gw butuh duit sekarang!!!!!!!!! ?? Yap siapa lagi kalau bukan BSI!!! ?? Ini stafnya pada tidur kah!??? Setidaknya konfirmasi kek kapan bagusnya ??	Negatif	0.72727272727253, 0.18181818181813, 0.090909090909339,
5	#5 - BSI ini bank apaan ya. Error kok lama bgt. Gk punya duit buat benerin apa ya #bsierror #bsidown #bsingawur @bankbsi_id	Negatif	0.66666666666633, 0.16666666666708, 0.16666666666658,

Fig. 6 Example of displaying program implementation results

Additionally, here is a pie chart showing the distribution of sentiment in the dataset:

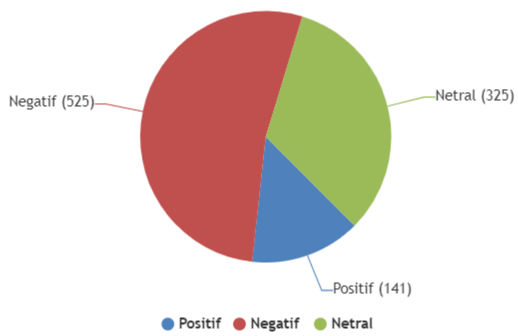

Fig. 7 Pie chart results from Naïve Bayes sentiment analysis

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13729
e-ISSN : 2541-2019
p-ISSN : 2541-044X

After the sentiment analysis program is run, a screenshot of the program shows the sentiment classification results for each processed dataset. Users can clearly see how each tweet or text is classified into positive, negative, or neutral categories. Additionally, the pie chart displays the distribution of the number of sentiments in the dataset visually. This diagram provides a more intuitive picture of how large the proportion of positive, negative, and neutral sentiment is in the analyzed dataset. This way, users can quickly understand sentiment patterns and trends in their data.

After displaying the results of the analysis using the Naive Bayes algorithm, this research also conducted sentiment analysis using the Google Cloud Natural Language API to compare the results. The analysis results from the two methods are then compared to evaluate the consistency and accuracy of sentiment classification. As an attachment, the next picture shows a comparison of the analysis results from the two methods. This will help you get a better idea of how well the Naive Bayes algorithm works for sentiment analysis on social media about BSI service disruptions.

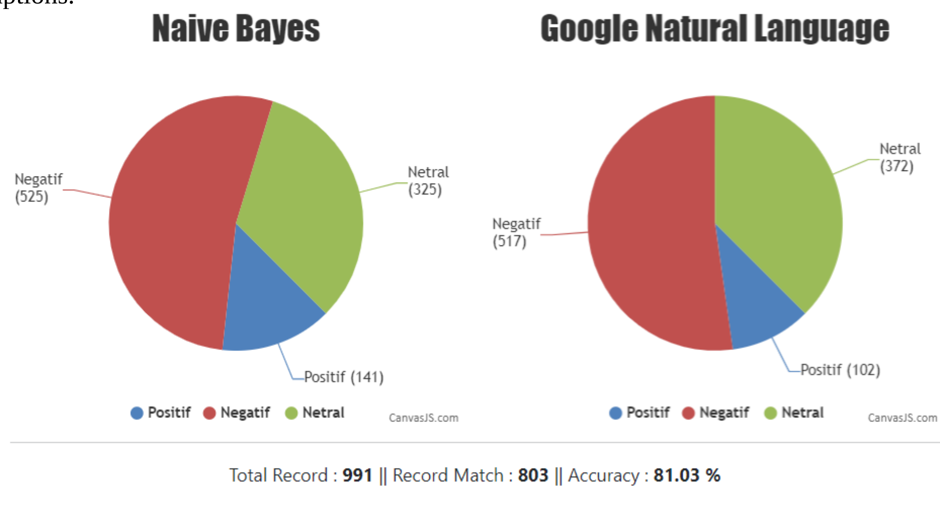

Fig. 8 Pie chart image of Naïve Bayes analysis results and Google Cloud Natural Language API

The image above displays many sentiments from each positive, neutral, and negative class. The similarity in accuracy between the two methods is 81.03% for 803 data points out of a total of 991 data points.

In this section, we present a comparative analysis of different methods used in our research for sentiment analysis of BSI service disruptions. The table below provides a detailed comparison of each method, including the amount of data utilized and the resulting accuracy. This comparison aims to highlight the effectiveness of the Naive Bayes algorithm in relation to other techniques and to underscore the robustness of our chosen method.

Table 4
Comparison of Sentiment Analysis Methods

	Naïve Bayes	Google Natural Language
Total Data	991	
Total Positive	141	102
Total Neutral	325	372
Total Negative	525	517
Accuracy	81,03%	

CONCLUSION

Based on the results of the research that has been carried out, it is concluded that analysis of public sentiment regarding BSI service disruptions via the X social media platform reveals diverse patterns of responses and perceptions. Of the total 991 tweets analyzed using the Naive Bayes algorithm, 525 were recorded with negative sentiment, 325 with neutral sentiment, and 141 with positive sentiment. Meanwhile, the Google Cloud Natural Language API method shows 517 negative sentiment data, 102 positive sentiment data, and 372 neutral sentiment data. The accuracy level of the Naive Bayes algorithm for the Google Cloud Natural Language API is 81.03%. This analysis provides a deeper understanding of how communities responded to BSI service disruptions, highlighting differences between the results obtained from the two methods. Nevertheless, the results obtained show that the Naive Bayes algorithm is generally quite good at classifying sentiment in tweet data related to BSI interference, although the differences in results between the Naive Bayes method and the Google Cloud Natural Language API need to be considered in interpreting the results.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 8, Number 3, Juli 2024
DOI : https://doi.org/10.33395/sinkron.v8i3.13729
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Based on the results of this research, several suggestions can be given for further development. First, expand data sources by collecting information from various social media platforms, such as Twitter, Facebook, and others, to provide a more comprehensive picture of public sentiment regarding BSI service disruptions. Second, increasing accuracy can be achieved by researching and developing more sophisticated and accurate sentiment analysis methods, such as deep learning or ensemble learning techniques. Third, a deeper analysis of the pattern of public responses and perceptions of BSI service disruptions is needed, including understanding the factors that influence changes in sentiment over time. Fourth, the integration of sentiment analysis methods, such as combining results from Lexicon-Based, Naive Bayes, and Natural Language Processing approaches, will provide a more holistic understanding of public sentiment. Finally, further research is needed to understand the impact of public sentiment on a company's reputation and performance, as well as its strategic implications for managing crises and improving customer service.

REFERENCES

Afdhal, I., Kurniawan, R., Iskandar, I., Salambue, R., Budianita, E., & Syafria, F. (2022). Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia. *Penerapan Algoritma Random Forest Untuk Analisis Sentimen Komentar Di YouTube Tentang Islamofobia*, *5*(1), 122–130.

Alsaeedi, A., & Khan, M. Z. (2019). A study on sentiment analysis techniques of Twitter data. *International Journal of Advanced Computer Science and Applications*, *10*(2), 361–374.

Apriani, R., & Gustian, D. (2019). Analisis Sentimen Dengan Naïve Bayes Terhadap Komentar Aplikasi Tokopedia. *Jurnal Rekayasa Teknologi Nusa Putra*, *6*(1), 54–62. https://doi.org/10.52005/rekayasa.v6i1.86

Firmansyah, Z., & Puspitasari, N. F. (2021). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Berdasarkan Opini Pada Twitter Menggunakan Algoritma Naive Bayes. *Jurnal Teknik Informatika*, *14*(2), 171–178.

Fitri, S. G., Selsi, R., Rustam, Z., & Pandelaki, J. (2020). Naïve bayes classifier models for cerebral infarction classification. *Journal of Physics: Conference Series*, *1490*(1), 012019.

Furqan, M., Sriani, S., & Sari, S. M. (2022). Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia. *Techno.Com*, *21*(1), 51–60. https://doi.org/10.33633/tc.v21i1.5446

Khomsah, S. (2020). Naive Bayes Classifier Optimization on Sentiment Analysis of Hotel Reviews. *Jurnal Penelitian Pos Dan Informatika*, *10*(2), 157. https://doi.org/10.17933/jppi.2020.100206

Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.

Manguri, K. H., Ramadhan, R. N., & Amin, P. R. M. (2020). Twitter sentiment analysis on worldwide COVID-19 outbreaks. *Kurdistan Journal of Applied Research*, 54–65.

Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes. *Bulletin of Computer Science and Electrical Engineering*, *1*(1), 26–32. https://doi.org/10.25008/bcsee.v1i1.5

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

Pristiyono, Ritonga, M., Ihsan, M. A. Al, Anjar, A., & Rambe, F. H. (2021). Sentiment analysis of COVID-19 vaccine in Indonesia using Naïve Bayes Algorithm. *IOP Conference Series: Materials Science and Engineering*, *1088*(1), 012045. https://doi.org/10.1088/1757-899X/1088/1/012045

Turmudi Zy, A., Nugroho, A., Rivaldi, A., & Afriantoro, I. (2022). Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes. *Jurnal Teknologi Informatika Dan Komputer*, *8*(2), 202–213. https://doi.org/10.37012/jtik.v8i2.1240

Husen, R. A., Astuti, R., Marlia, L., Rahmaddeni, R., & Efrizoni, L. (2023). Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(2), 211–218. https://doi.org/10.57152/malcom.v3i2.901

Zulfikar Hardiansyah. (2023). *Kronologi Layanan BSI Eror, Down Berhari-hari dan "Dipalak" Hacker Ransomware Ratusan Miliar*. Kompas.Com. https://tekno.kompas.com/read/2023/05/17/09010077/kronologi-layanan-bsi-eror-down-berhari-hari-dan-dipalak-hacker-ransomware?page=all

*name of corresponding author