Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

Implementation of the Naïve Bayes Algorithm in the SMS Spam Filtering System

Diah Ayu Anggraini[1)*], Muhammad Ikhsan[2)], Suhardi[3)]
[1)2)3)] Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
[1)]diahanggrni14@gmail.com, [2)]mhd.ikhsan@uinsu.ac.id, [3)]suhardi@uinsu.ac.id

ABSTRACT

In the context of the escalating global spam activity, supported by data from CNN Indonesia in 2021, this research aimed to investigate the root causes and characteristics of this phenomenon. The approach employed in this study involved a series of exploration and classification stages of text messages with the clear objective: to determine whether each message fell into the spam category or not, utilizing the Naïve Bayes method. Additionally, the research aimed to identify the factors influencing the status of text messages, whether they were considered as spam or not. The Naïve Bayes classification method was chosen to facilitate the process of identifying spam-related messages. The dataset used in this research had an 80:20 ratio and was obtained from the Department of Communication and Informatics of Asahan Regency. This data was used to train and test the developed classification model. Data labeling processes were conducted to uncover the factors influencing the status of text messages as spam or non-spam. The research findings indicated that issues related to spam and non-spam messages remained a serious concern. The high accuracy rate, reaching 92%, achieved by the Naïve Bayes method in classifying messages, demonstrated the effectiveness of the method in detecting spam messages.

Keywords: Message Classification System, Naïve Bayes, Spam, SMS Messages, TF-IDF

INTRODUCTION

SMS served as a text-based communication tool operating on mobile devices (Fachri & Sembiring, 2020). The introduction of SMS began in Europe in 1992 with the integration of GSM, later evolving into CDMA and TDMA technologies (Sari et al., 2022). With the advancement of technology, SMS was not only used to exchange messages with known contacts but also with unfamiliar ones . Messages were categorized into two types: spam and non-spam/ham. Spam referred to irrelevant or inappropriate messages sent to a large number of recipients (Reviantika et al., 2021). On the other hand, non-spam/ham messages were genuine or important messages. Sending commercial messages to wireless devices without permission was considered a violation of the law according to the FCC (Panggabean et al., 2023). Spam could include requests for payments, money transfers, notifications of winnings, or fraudulent attempts (Adila et al., 2023).

According to a report from CNN Indonesia in 2021, Truecaller released a report on global spam call activities in the Truecaller Global Spam Report 2021. Indonesia ranked 6th out of 20 countries most affected by spam. Truecaller successfully assisted over 300 million users worldwide in blocking and identifying 37.8 billion spam calls. The report also outlined spam and scam trends over the past year, presenting some critical data, the current situation, and predictions for 2022. In January, the total number of reported spam calls reached nearly 12.6 million, which increased to 25.8 million in October 2021. On average, each Indonesian received 14 spam calls per month. Data on spam calls were collected from incoming and outgoing call logs and messages during the period from January 1, 2021, to October 31, 2021 (Ikhsan, 2021).

The Department of Communication and Informatics frequently received complaints from the public about various issues, including spam problems. However, the system they used still classified messages manually, which was deemed ineffective and time-consuming (Arisona et al., 2023). To address this issue, an automated message classification system was designed. This system utilized the Naïve Bayes Algorithm to filter messages, employing automatic classification and labeling methods.

The aim of this research was to implement an automated message classification system, particularly in the context of SMS, to assist the Department of Communication and Informatics in handling complaints from the public regarding spam and unwanted messages. It was hoped that with this system, the process of analyzing and classifying messages could be done more quickly and efficiently. This system would utilize the Naïve Bayes Algorithm to classify messages into two main categories: ham (clean messages without spam elements) and fraud.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

LITERATURE REVIEW

In the previous study titled "Design of Mobile Application for Detecting SMS Spam Messages in Indonesia," several algorithms were investigated for text classification, including Support Vector Machines (SVM), Decision Tree, K-Nearest Neighbor (KNN), Naïve Bayes, Neural Network, Association rule-based, and Boosting. The team conducted an accuracy analysis of each algorithm to determine the best one. From the analysis conducted, the SVM algorithm achieved an accuracy of 94.7%, while the Naïve Bayes algorithm reached 84%. However, among these algorithms, the Naïve Bayes algorithm emerged as the simplest, most efficient, and most commonly used algorithm. Based on the analysis, the Naïve Bayes algorithm was selected for the design of the SMS Spam detection application (Lumbantobing et al., 2021).

In the subsequent study titled "Comparison of Naïve Bayes, SVM, and Decision Tree Algorithms for SMS Spam Classification," the Naïve Bayes algorithm yielded the highest accuracy among other algorithms with a value of 0.94. It was determined from precision, recall, fl-score, and accuracy metrics that the Naïve Bayes algorithm outperformed SVM and Decision Tree algorithms in classifying Indonesian SMS spam (Fitriana et al., 2020).

The difference between your research and the previous studies lies in the research objective. Your research aimed to classify text messages to determine whether they are spam or not using the Naïve Bayes method. Meanwhile, previous studies focused on comparing various classification algorithms such as Naïve Bayes, SVM, and Decision Tree. In the research method, your study employed the Naïve Bayes classification method with a dataset obtained from the Communication and Informatics Department of Asahan Regency. However, previous studies utilized different classification methods and datasets. Your research achieved an accuracy of 92% in classifying messages, while previous studies obtained varied results depending on the algorithm used. Some previous studies achieved an accuracy of 84% for Naïve Bayes, while others achieved lower accuracies.

This research contributed by implementing an automated message classification system using the Naïve Bayes method. Additionally, it noted that previous studies predominantly utilized manual approaches in the classification process. Therefore, this research provided innovation by reducing manual involvement in the classification process, thereby enhancing the efficiency and speed of message analysis. Consequently, this research can be regarded as a significant contribution to the field of spam message classification, particularly in the context of SMS, with a focus on automating the process using the Naïve Bayes method.

METHOD

The methodology employed in completing this research involved utilizing the Naïve Bayes algorithm as a guideline in the applied research method. There were several stages in the framework of this research process, where the Naïve Bayes algorithm played a role. Naïve Bayes was utilized to develop new ideas, refine products, and serve as a learning tool and work aid that supported efficiency and productivity.

Here are the stages of the research framework (Aulia et al., 2023):

Planning

The planning phase in this research framework was conducted through several stages. The first stage was determining the research topic, where the researcher conducted observations to identify the issues to be addressed (Batubara & Nasution, 2023). The chosen topic for this final project was the Implementation of the Naïve Bayes Algorithm in SMS Spam Message Filtering Systems. Then, the second stage was determining the research object, where SMS Spam Messages were selected as the research object. This was followed by the problem formulation stage, where the researcher identified the problems to be studied along with their scope or limitations. The next stage was determining the research title, where based on observations on the research object, the researcher determined a title that aligned with the studied issue. Finally, the objective determination was carried out to clarify the goals of this research.

Data Collection Techniques

The data collection techniques in this research included observation, literature review, and interviews. Observation was conducted through direct observation at the Asahan District Communication and Information Office, where the researcher observed the work processes of the relevant employees and conducted an initial survey to obtain data. Literature review was carried out by collecting data from various relevant sources, including journals, e-books, and references from the official website of the Asahan District Information and Communication Office. Finally, interviews were conducted with Mrs. Riri in the IT Department to obtain information about the system to be developed and the issues faced by the community.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

Needs Analysis
The needs analysis stage was a process to obtain information about the requirements needed to identify problems in the research (Syahranitazli & Samsudin, 2023). This analysis involved specific methods to evaluate problems that arose during the research journey, starting from identifying the SMS spam message filtering system, then through model analysis to the testing stage in implementing the Naïve Bayes algorithm in the system.

Design
Designing is a process of depiction, planning, and detailed sketching or arrangement of several separate elements into a unified whole that functions (Shenita & Suendri, 2023).

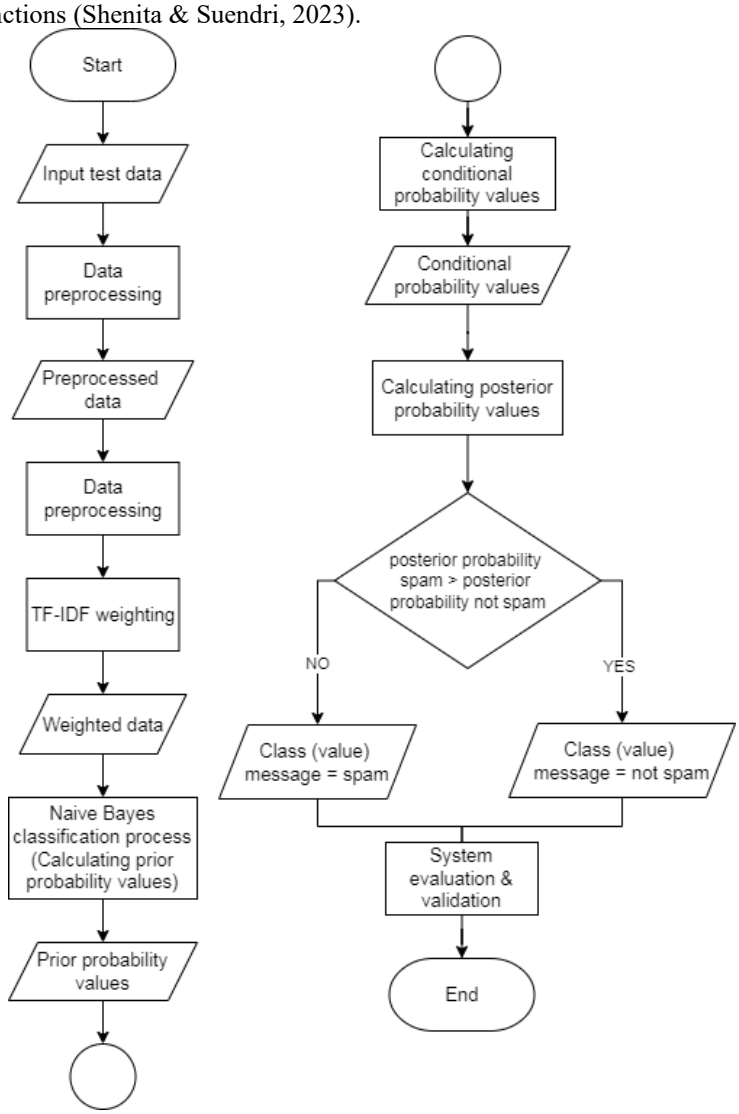


Fig. 1 Flowchart of the Classification System.

Testing
This testing process was conducted using a dataset with an 80% training data and 20% testing data, where the testing data consisted of manually labeled SMS spam messages. The dataset was then classified using the Naïve Bayes method. Through the classification process, it was determined whether the inputted SMS spam messages belonged to spam or non-spam messages and could be separated based on their respective classes.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

Application or Use
The application/use of the system in implementing the Naïve Bayes method can assist in calculating results for testers and help minimize errors in calculations within the system.

RESULT

Analysis
In the data analysis for classifying SMS spam messages using the Naïve Bayes method, data was collected, and information on the data collection needs was sought. The data consisted of the content of SMS messages relevant to the research topic. Data was obtained through manual input from the research site, then stored and processed in .xlsx format. Each data was labeled (classified) for training during the classification process. The labeling process used calculations to categorize the data into two classes, namely spam and non-spam. Before further processing, the data went through a data preparation stage, including data preprocessing, word weighting, dividing between training and test data, and finally, classification of the dataset.
The classification method used was Naïve Bayes with probability techniques (Rizki et al., 2021). This method was used to evaluate the system's performance in classifying messages in the test data based on training on the training data. The system classified and generated information about the classification and prediction on the test data, presented in matrix form. Confusion matrix was used to evaluate the accuracy level of the Naïve Bayes method in classifying the SMS message dataset.

Data Representation
This section will elucidate the processes undertaken in this research, starting from data collection to the classification process using the Naïve Bayes method.

Dataset Collection
The dataset in this research consists of SMS message contents containing both spam and non-spam sentences. The total dataset collected amounted to 124 messages containing both spam and non-spam content. The dataset comprises message contents received by users on the SMS messaging application. The table below shows an example of raw data in the messaging application on a mobile phone.

Table 1. Dataset Display

No	Phone Number	Message Content
1	+62 858-2411-XXXX	ass…kmi dri Permata Bank M-nawarkan p1njaman d4na t4npa agunan/survey??? 5jta s/d 500jta U/Info Chat WhatsApp:085824615675
2	3Topup	Pengisian Rp.5000 sukses. Masa aktif s/d 06/08/2023. Max masa aktif 365hr. PASTI MURAH kuota 24jam, temukan di https://bit.ly/3Jid1Ak atau *111*1#
…	…	….
124	+62 856-5683-XXXX	Selamat anda mendapatkan RP. 10JT dri TikTOK ID 55TH77K info chat admin WA.087759899091 Trimakasih.

Labeling
After the research dataset was collected, the next step was to label each message to determine whether it was spam or not. The labeling process was automated by the system based on indicators of spam and non-spam messages obtained from reference sources. In the labeling process, several things needed to be considered.

Table 2. Labeling Indicators
Sumber : (Sutra Dewi, 2022)

Spam	Not Spam
1. Contains something we didn't do. 2. Contains requests to fill in personal and confidential data. 3. Offers pharmaceutical or health products.	1. Contains desired information for the user. 2. The sender's number is from one of the user's contacts. 3. The message sender is from a government institution. 4. The message is not harmful and does not disclose personal data.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

4. Messages informing the number owner of winning a lottery.	

The dataset provided earlier will undergo the labeling process. The classification classes categorized as spam and non-spam will be automatically processed using keywords considered as spam such as 'winner', 'loan', 'lottery', '@', 'million', 'prize', and then utilizing phone numbers considered as spam.

```
def labelling(isi_pesan):
  keyword_spam = ['pemenang', 'pinjaman', 'undian', 'd@n@', 'jt', 'hadiah', 'selamat', 'minat', 'terpilih', 'bantuan',
                  'mendapatkan', 'http', 'bit.ly', 'pnjaman', 'hutang', 'online', 'klik', 'keuntungan', 'penerima', 'uang',
                  'meraih', 'butuh', 'modal', 'bunga', 'kode', 'menang', 'agen', 'hubungi', 'mengganggu', 'tunai', 'www', 'segera',
                  'penipuan', 'cek', 'menawarkan', 'spin', 'Tips', 'berhak', 'jaminan', 'sms', 'ass', 'resmi',
                  'mendpatkan', 'whatsapp', 'kami', 'Hub', 'pin', 'chat', 'giveaway', 'bonnus', '@nd@', 'utk']
  if not any(kata in isi_pesan.lower() for kata in keyword_spam):
    return True
  return False

data = pd.read_csv('dataset_sms_nolabel.csv', sep = ';', encoding = 'ISO-8859-1')
data['label'] = data.apply(lambda row: "Bukan Spam" if labelling(row['isi_pesan']) else "Spam", axis=1)
data.head(125)
```

Fig. 2 The data labeling process in Google Colaboratory

After labeling or assigning labels to the data, the data will be structured in a data structure called a dataframe. It's like organizing data in the form of a table, where each row represents one entity or data sample, and each column represents attributes or features associated with that entity. A DataFrame is a useful way to organize and store data for further analysis.

index	isi_pesan	label
0	asskmi dri Permata Bank M-nawarkan p1njaman d4na t4npa agunan/survey??? 5jta s/d 500jta U/Info Chat WhatsApp:085824615675	Spam
1	Mohon maaf, persediaan voucher sudah habis. Nantikan program Telkomsel Poin berikutnya.	Bukan Spam
2	Nomor 085207057991, Per Besok 24 Okt, +2000 COIN PULSA Anda akan Hangus! SEGERA Ambil di *500*75# dan Tukar Jadi PULSA Hari Ini!* Hub *500*75# dan Pilih 1 YA!	Spam
3	Kabari aku ya kalo ada info di grup angkatan, aku hari ini off wa dulu	Bukan Spam
4	@ss bpk/ibu YTH kred1t t4np4 1n99un4n den94n pl4fon dri 5>500jt minat whts@@s WA:085341719177	Spam
5	Assalamualaikum Ada msAlah keuangn Atau butu modl usha & lain2 kmi menawarkan pijaman minimal 5jt/500jt WA.082394979022	Spam
6	Rekening SeaBank kamu sudah berhasil dibuka! Klik seabank.co.id/c/home untuk mulai menabung. Bunga tinggi s.d. 6% cair setiap hari!	Spam
7	3LOKER: IT/COMPUTER SOFTWARE, PT. Akar Inti Teknologi, Jakarta, Syarat: Pengalaman 3 thn http://triloker.com/q/277297	Spam
8	Nomor 082213290961, Per Besok 02 Aug, +2000 COIN PULSA Anda akan Hangus! SEGERA Ambil di *500*75# dan Tukar Jadi PULSA Hari Ini!* Hub *500*75# dan Pilih 1 YA!	Spam
9	Belajar dari pandemic, masuk dalam agenda Presidensi G20 Indonesia. Hal ketersediaan vaksin dan sistem kesehatan global jadi topik bahasan utama. s.id/G20pedia	Spam
10	Kamu Terpilih Mndaptkn Rp100jt dri TIKT0K Ket_Selanjutnya Chat Wa -81250856345	Spam
11	(#PESAN RESMI#) PT.WhatsApp INDONESIA Selamat Anda Terpilih Sebagai Penerima DANA BANTUAN &) JUTA Dari PT/WhatsApp HUB.WA:083186078999 Bpk:RAHMANSYAH	Spam
12	PROMO Spesial Beli: -Rp55Rb/17GB+25min Tsel 30hr dg balas AH1	Bukan Spam
13	Terimakasih Anda sudah menukar 1 POIN untuk kesempatan menang Samsung Galaxy Z Flip 4. Tunggu pengumuman Pemenang tgl 2UG23 di MyTelkomsel. SKB	Spam
14	Pelanggan 082213290961, Selamat! Nomor Kamu di Hape Ini Baru Saja Dapat +2000 COIN PULSA di *500*75 SEGERA Hub *500*75 utk Ambil. Bisa Kamu Tukar Jadi PULSA!	Spam
15	Pelanggan 081268560347, Kamu Dapat Pesan. Untuk Baca, Hubungi *101*11# Sekarang, Chatting Tanpa Kuota Internet & Raih bonus Pulsa	Spam
16	k@m1 d@r1 te@m B@P@U m3nguc@pk@n @nd@ mdp@tk@n Rp.50.000.000 info vi@ wh@ts@p: 082236546775	Spam
17	Selamat, Paket Kebangetan Data 30K 14GB 30D /30 hari Rp 30000 telah aktif, berlaku s/d tgl 11/08/2023 pkl. 23:59 WIB. Cek status/berhenti berlangganan melalui My Telkomsel Apps atau hub *363#. Info : 188.	Spam
18	Pelanggan 081268560347, Kamu Dapat Pesan. Untuk Baca, Hubungi *101*11# Sekarang, Chatting Tanpa Kuota Internet & Raih bonus Pulsa	Spam

Fig. 3 Labeled Message Data

The labeled dataset was then saved in ".csv" format to facilitate the message classification process that was to be conducted. An example of the labeled dataset can be seen in the following table.

Text Preprocessing

Text Pre-Processing was an essential stage in cleaning the dataset before conducting the classification process using predefined methods. This process was highly beneficial in facilitating classification. The stages included data cleansing by removing irrelevant characters and reducing noise, standardizing lowercase letters throughout the dataset, separating words in sentences, removing unimportant words like conjunctions, and stripping affixes from words. After completing all pre-processing stages, the data was saved in .csv format for use in the next step.

TF-IDF Weighting

In this stage, the system calculates the weight for each word in the document to determine how important those words are in their context. This process is called TF-IDF, which stands for "Term Frequency-Inverse Document Frequency". This process can be seen in a flowchart illustrating the steps taken to determine the weight of words. In other words, the system measures how often a word appears in the document (TF), but also takes into account how common the word is across the entire collection of documents (IDF), so that words that rarely appear but are important will have a higher weight.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

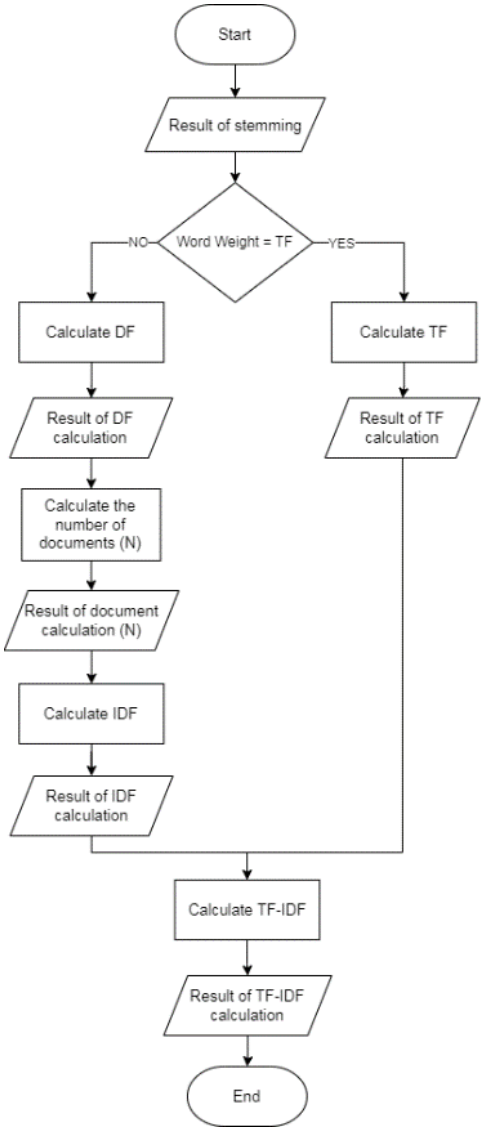


Fig. 4 TF-IDF Process Flowchart

The mathematical equation to calculate TF (Term Frequency) is:

$$TF = \frac{f_{t,d}}{\sum t' \in d f_{t,d}} \quad (1)$$

description :

TF	: *term frequency*
d	: documen
t	: *term*
ft,d	: number of terms (tokens/words) in each document
$\sum t' \in d\ ft,d$	: number of documents containing t

The mathematical equation for calculating *IDF* (idf_i)

$$idf_t = \ln\left(\frac{1+N}{1+dft}\right) + 1 \quad (2)$$

The equation for calculating the weight of each document

$$W_{dt} = TF_{dt} \times idf_t \quad (3)$$

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

description:
idf_t : the *idf* value of the term (token) t
N : The number of available documents
dft : the frequency of the word appearing in the document
tf_{td} : the frequency of the word t appearing in document d

Then normalize the obtained weights using the following equation:

$$Norm\ W_{dt} = \frac{W_{dt}}{\sqrt{\sum_{i=1}^{n} x_i^2}} \quad (4)$$

Splitting the dataset
Dataset splitting is one of the steps in dividing the dataset into two parts consisting of training data and test data, where the composition of the split will allocate a larger proportion to the training data compared to the test data. In this research, a dataset split with an 8:2 ratio will be applied, where 80% of the total dataset will be used as training data, while the remaining 20% will be used as test data, as referenced in the previous research.

Naïve Bayes Classifier Classification
After the word weighting step is completed, the classification of the test data will be conducted using the Naïve Bayes Classifier method. The Naive Bayes method was utilized for predicting probabilities. It constituted a form of statistical classification known as the Naïve Bayesian Classifier. This method is grounded on the simple assumption that attribute values are conditionally independent of each other when the output value is known. Its advantage lies in requiring minimal training data and often proving more effective than anticipated in complex real-world situations. The Bayes' theorem was a method used to transform uncertain data into certain data by comparing between the yes and no data, particularly in the context of this research to classify data as spam or non-spam.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (5)$$

Description:
X : Data with an unknown class.
H : The hypothesis that the data belongs to a specific class.
P(H\X) : The probability of hypothesis *H* given condition *X* (posterior probability)
P(X\H) : The probability of *X* given condition *H*
P (H) : The probability of hypothesis *H* (prior probability)
P (X) : The probability of *X*

Message Filtering System Implementation
All the stages conducted during the data analysis will then be implemented into a system aimed at filtering messages. In the message filtering system process, the Python programming language is used, and the text editor tool used is Google Colaboratory.

Data Preprocessing
All the steps performed in the data preprocessing process will result in a new dataset in the form of new documents, which will then be used in the classification process.

Cleaning
Removing noise means cleaning text from unnecessary or distracting characters. For example, removing unnecessary punctuation, excessive spaces, or irrelevant symbols. The purpose of this step is to make the text more organized and easier to read. Below is an example of the results of this cleaning step performed using Google Colaboratory.

index	isi_pesan	label
0	ass kmi dri Permata Bank M nawarkan p njaman d na t npa agunan survey jta s d jta U Info Chat WhatsApp	Spam
1	Mohon maaf persediaan voucher sudah habis Nantikan program Telkomsel Poin berikutnya	Bukan Spam
2	Nomor Per Besok Okt COIN PULSA Anda akan Hangus! SEGERA Ambil di dan Tukar Jadi PULSA Hari Ini! Hub dan Pilih YA!	Spam
3	Kabari aku ya kalo ada info di grup angkatan aku hari ini off wa dulu	Bukan Spam
4	ss bpk ibu YTH kred t t np n un n den n pl fon dri jt minat whts s WA	Spam

Fig. 5 The Result of the Cleaning Process

Case Folding
Converting all letters in the dataset to lowercase means changing all uppercase letters to lowercase. This means that

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

if there are capital letters in the text, they will be converted to lowercase. For example, "Hello" would become "hello". The purpose of this step is to make the data consistent in terms of letter case, which facilitates further processing and analysis. Below is an example of the results of this stage performed using Google Colaboratory.

index	isi_pesan	label
0	ass kmi dri permata bank m nawarkan p njaman d na t npa agunan survey jta s d jta u info chat whatsapp	Spam
1	mohon maaf persediaan voucher sudah habis nantikan program telkomsel poin berikutnya	Bukan Spam
2	nomor per besok okt coin pulsa anda akan hangus! segera ambil di dan tukar jadi pulsa hari ini! hub dan pilih ya!	Spam
3	kabari aku ya kalo ada info di grup angkatan aku hari ini off wa dulu	Bukan Spam
4	ss bpk ibu yth kred t t np n un n den n pl fon dri jt minat whts s wa	Spam

Fig. 6 The Result of the Case Folding Process

Tokenizing

Tokenization is like breaking a sentence puzzle into pieces, or "tokens", which are individual words. For example, the sentence "I like eating fried rice" would be separated into tokens like "I", "like", "eating", "fried", "rice". Each token can be considered as a separate piece that is easier to process, analyze, or manipulate further. This tokenization process helps computers understand and work with text more efficiently. Below is an example of the results of this stage performed using Google Colaboratory

index	isi_pesan	label
0	ass,kmi,dri,permata,bank,m,nawarkan,p,njaman,d,na,t,npa,agunan,survey,jta,s,d,jta,u,info,chat,whatsapp	Spam
1	mohon,maaf,persediaan,voucher,sudah,habis,nantikan,program,telkomsel,poin,berikutnya	Bukan Spam
2	nomor,per,besok,okt,coin,pulsa,anda,akan,hangus!,segera,ambil,di,dan,tukar,jadi,pulsa,hari,ini!,hub,dan,pilih,ya!	Spam
3	kabari,aku,ya,kalo,ada,info,di,grup,angkatan,aku,hari,ini,off,wa,dulu	Bukan Spam
4	ss,bpk,ibu,yth,kred,t,t,np,n,un,n,den,n,pl,fon,dri,jt,minat,whts,s,wa	Spam

Fig. 7 The result of the Tokenizing Process

Stopword Removal

This process is a step to remove words that do not provide significant meaning in the data. For example, words like "and", "or", "I", which often appear in text but do not convey useful information for analysis. Removing these words helps simplify the text and enhance focus on more relevant words. Here, we use the nltk (Natural Language Toolkit) library to assist in this removal process. Below is a display of the results of this stage shown in the Google Colaboratory environment.

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

1 to 5 of 5 entries Filter

index	isi_pesan	label
0	ass,kmi,dri,permata,bank,m,nawarkan,p,njaman,d,na,t,npa,agunan,survey,jta,s,d,jta,u,info,chat,whatsapp	Spam
1	mohon,maaf,persediaan,voucher,habis,nantikan,program,telkomsel,poin	Bukan Spam
2	nomor,besok,okt,coin,pulsa,hangus!,ambil,tukar,pulsa,ini!,hub,pilih,ya!	Spam
3	kabari,ya,kalo,info,grup,angkatan,off,wa	Bukan Spam
4	ss,bpk,yth,kred,t,t,np,n,un,n,den,n,pl,fon,dri,jt,minat,whts,s,wa	Spam

Fig. 8 Result of Stopword Removal Process

Stemming

This process aims to remove affixes attached to words in the text messages. Affixes are parts of words that add meaning or change their function. For example, in the word "playing," the prefix "play-" indicates that it is a verb in infinitive form. Removing affixes helps simplify words so they are easier to understand and analyze. To perform this step, we use the Sastrawi library, which is a tool for natural language processing in the Indonesian language. The results of this stage are displayed in the Google Colaboratory environment.

```
ass kmi dri permata bank m nawarkan p njaman d na t npa agun survey jta s d jta u info chat whatsapp
mohon maaf sedia voucher habis nanti program telkomsel poin
nomor besok okt coin pulsa hangus ambil tukar pulsa ini hub pilih ya
kabar ya kalo info grup angkat off wa
ss bpk yth kred t t np n un n den n pl fon dri jt minat whts s wa
```

Fig. 9 The result of the stemming process

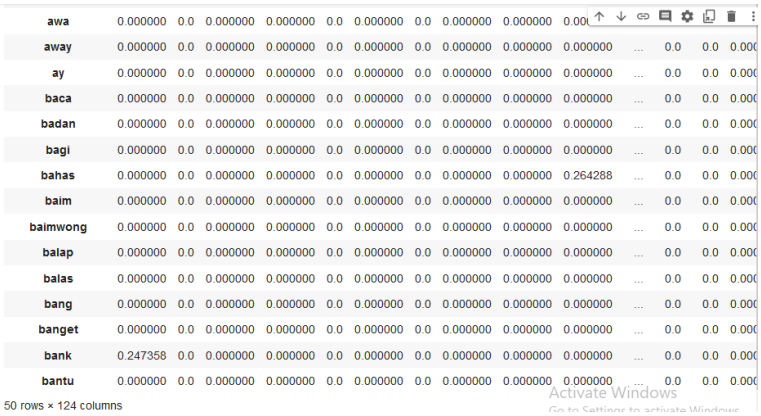
TF-IDF

This stage involves calculating how important each word is in the document based on how often the word appears and how unique the word is in the dataset. The method used to measure this is called TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF assigns a higher weight to words that appear more frequently in a specific document but rarely appear in other documents, as these words are considered more important in describing the content of that document specifically. In this study, the result is 124 words that have TF-IDF weights.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024



awa	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.00				
away	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
ay	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
baca	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
badan	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
bagi	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
bahas	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.264288	...	0.0	0.0	0.000
baim	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
baimwong	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
balap	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
balas	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
bang	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
banget	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
bank	0.247358	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000
bantu	0.000000	0.0	0.000000	0.000000	0.0	0.000000	0.0	0.000000	0.000000	0.000000	...	0.0	0.0	0.000

50 rows × 124 columns

Fig. 10 The TF-IDF process generates the Document-Term Matrix

Split Dataset

This stage is to divide the dataset into two parts: training data (used to train the model) and testing data (used to evaluate the model). In this study, an 80:20 ratio is used, which means 80% of the dataset is used for training data and the remaining 20% is used for testing data. This is done so that the model can learn from the majority of the data but also be evaluated on data it has never seen before to assess its overall performance.

```
#splitting data
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(x_tfidf, dk_clean['Label'], test_size = 0.2, random_state = 1)
```

Fig. 11 Dataset Splitting

After undergoing preprocessing, the dataset has been divided into 124 labeled data. Out of this total, 80% is utilized as training data, meaning there are 99 data points designated for training the model. The remaining 20% is allocated as testing data, comprising 25 data points that will be used to evaluate how well the model performs on unseen data. With this division, the model can learn from a significant portion of the data while also being tested on new data to assess its overall performance.

```
Size of x_train: (99, 651)
Size of y_train: (99, )
Size of x_test: (25, 651)
Size of y_test: (25, )
```

Fig. 12 The ratio of Training Data to Test Data Size

Naïve Bayes Classifier Classification

In this stage, the training data, which already has classes (values), will be used to train the classification model system built to make predictions and classifications on the testing data. Consequently, during the testing process, the system can determine the class of each message content in the testing data. The following are the steps to perform classification prediction using the Naïve Bayes Classifier method.

```
[16] clf = MultinomialNB().fit(X_train, y_train)
     predicted = clf.predict(X_test)

     print(f'confusion matrix data testing:\n {confusion_matrix(y_test, predicted)}')
     #print(f'confusion matrix data training :\n {confusion_matrix(y_train, clf)}')
     print('=====================================================\n')
     print(classification_report(y_test, predicted, zero_division=0))
```

Fig. 13 Classification Process of the Dataset

Testing

After the design and implementation of the message classification system were completed, the next step was testing the system to evaluate its success in predicting and classifying messages as spam or not spam. This research utilized Google Collaborator as a tool for data analysis, with the dataset stored in .csv format. The classification method employed was the Naïve Bayes Classifier, which generated predictions for the test data. Information from the confusion matrix presented the number of correct and incorrect predictions for each class (spam and not spam),

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

providing an overview of the system's accuracy. From this evaluation, information about the accuracy, precision, recall, and f1-score of the classification system could be obtained. Below is the display of the confusion matrix from the classification results.

```
confusion matrix data testing:
 [[ 0  2]
 [ 0 23]]
==================================
```

Fig. 14 The results of the system testing

In this study, the evaluation of the system's results is presented in the form of a confusion matrix. The confusion matrix provides information in the form of numbers arranged in a 2×2 matrix. From the confusion matrix, values such as accuracy, precision, recall, and f1-score can be calculated as the outcome of the classification system using the Naïve Bayes method. The classification results of the confusion matrix are presented as follows.

Tabel 3

Confusion Matrix of Classification Results

ij		**Predicted Class (j)**	
		Not Spam	**Spam**
Actual Class (*i*)	**Not Spam**	0	0
	Spam	2	23

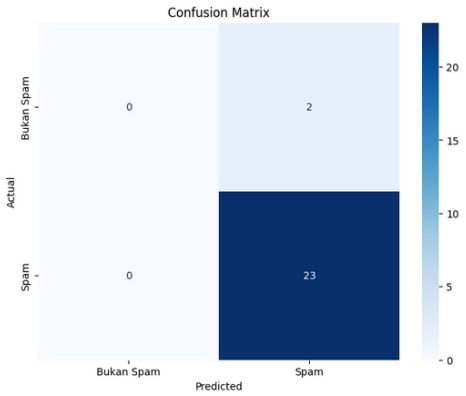


Fig. 15 The Confusion Matrix graph of the Classification Results

From Table 3, the values of accuracy, precision, recall, and f1-score can be computed:

$$Accuracy = \frac{23+0}{23+0+2+0} = 0{,}92 \times 100\% = 92\%$$

$$Precision = \frac{23}{23+2} = 0{,}92 \times 100\% = 92\%$$

$$Recall = \frac{23}{23+0} = 1 \times 100\% = 100\%$$

$$f1\text{-}score = \frac{2\times100\times96}{100+96} = 0{,}96 \times 100\% = 96\%$$

The overall values above could be presented in a classification report. Herein, the classification report of the confusion matrix from the testing of the validation data using the Naïve Bayes Classifier method is presented.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

```
confusion matrix data testing:
 [[ 0  2]
 [ 0 23]]
=====================================================

              precision    recall  f1-score   support

  Bukan Spam       0.00      0.00      0.00         2
        Spam       0.92      1.00      0.96        23

    accuracy                           0.92        25
   macro avg       0.46      0.50      0.48        25
weighted avg       0.85      0.92      0.88        25
```

Fig. 16 Classification Report

From the calculations above, it can be seen that the number of test data was 25, with an accuracy value of 92%, precision of 92%, recall of 100%, and an f1-score of 96%.

Implementation

The system developed in this research was applied to classify SMS messages as either spam or non-spam. Naïve Bayes Classifier was employed as the classification method, with accuracy testing conducted. The system aimed to facilitate and assist students, the general public, and even the government in classifying messages amidst the proliferation of scams presented through research journal articles and mass media.

DISCUSSIONS

By implementing the Naïve Bayes Algorithm in the SMS Spam Filtering System using a classification model, this research proposes several aspects of development for future studies. One of them is the ability to classify data containing messages that can be categorized as spam or non-spam, as well as identifying important features that can help distinguish between the two types of messages. These features may include keywords, excessive exclamation marks, the use of capital letters, and so on. Furthermore, the research will attempt to use methods such as K-Fold Cross Validation or applying other word weighting methods to evaluate and compare the performance of various classification methods. This will help determine which method is most effective in classifying messages. Additionally, in the system's development, a more user-friendly user interface will be built, based on web, mobile, or other platforms. With a more user-friendly interface, it is hoped that it will facilitate access for other users to use this system.

CONCLUSION

Based on the research findings regarding the filtering of spam and non-spam SMS messages using the Naïve Bayes Classifier classification method, it can be concluded that this study utilized a dataset comprising 124 SMS messages obtained automatically and categorized into two groups: spam and non-spam SMS messages, using Google Collaborator. The dataset was obtained through labeling techniques with a total of 124 data. Labeling was performed by matching keywords such as "Winner", "loan", "lottery", "d@n@", "million", "prize", and the number of mobile phone numbers. Labeling in this study employed keyword-based techniques, where the system automatically counted the number of data detected based on keywords that met the criteria. By applying this technique, message labeling was done automatically without manual intervention. The accuracy achieved using the Naïve Bayes Classifier method in the spam filtering system using the classification model can be considered good, with an accuracy of 92%, precision of 91%, recall of 100%, and an f1-score of 95%, based on the evaluation of a dataset comprising 124 data with a split of 8:2 for training and testing data and term weighting using the TF-IDF method.

REFERENCES

Adila, N., Khasanah, S., & Sutabri, T. (2023). STRATEGI PERANCANGAN SISTEM AMAVIS DAN SPAMASSASSIN PADA SPAM MAIL. *Jurnal Sain Dan Teknik*, *5*(2), 154–166.

Arisona, D. C., Wibowo, G. N. A., Siswanto, & Gunawan. (2023). Klasifikasi Pesan Biasa, Operator, Spam, dan Debt Collector Menggunakan K-Nearest Neighbor. *Insypro*, *8*(2), 1–6.

Aulia, Z. N., Jati, G. K., & Santoso, I. (2023). ANALISIS SENTIMEN TANGGAPANPUBLIC MENGENAI E-TILANG MELALUI MEDIA SOSIAL YOUTUBE MENGGUNAKAN ALGORITMA NAIVE BAYES.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 2, April 2024
https://doi.org/10.47709/cnahpc.v6i2.3875

Submitted : May 17, 2024
Accepted : May 28, 2024
Published : June 1, 2024

Jurnal IKRAITH-INFORMATIKA, *7*(2), 150–156.

Batubara, M. Z., & Nasution, M. I. P. (2023). Sistem Informasi Online Pengelolaan Dana Sosial Pada Rumah Yatim Sumatera Utara. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(3), 164–171.

Fachri, B., & Sembiring, R. M. (2020). Pengamanan Data Teks Menggunakan Algoritma DES Berbasis Android. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *4*(1), 110–116. https://doi.org/10.30865/mib.v4i1.1700

Fitriana, D. N., Setifani, N. A., & Yusuf, A. (2020). PERBANDINGAN ALGORITMA NAÏVE BAYES, SVM, DAN DECISION TREE UNTUK KLASIFIKASI SMS SPAM. *JUSIM (Jurnal Sistem Informasi Musirawas)*, *5*(2), 167–174.

Ikhsan, M. (2021). Spam di RI Naik Dua Kali Lipat 2021, Aksi Penipu Tepat Sasaran. Retrieved from CNN Indonesia website: https://www.cnnindonesia.com/teknologi/20211220131919-185-736203/spam-di-ri-naik-dua-kali-lipat-2021-aksi-penipu-tepat-sasaran

Lumbantobing, R. D. H., Manalu, E. M., Sitinjak, D. S. P., & Manurung, T. W. (2021). Rancangan Aplikasi Mobile Pendeteksi Spam SMS di Indonesia. *JURNALTIO*, *2*(1), 24–29.

Panggabean, E. S. … Iqbal, M. (2023). Memahami Spam Terhadap Digitalisasi Masyarakat Desa Perkebunan Sei Balai Kecamatan Sei Balai Kabupaten Batubara. *Jurnal Pengabdian Masyarakat (JAPAMAS)*, *2*(2), 270–278.

Reviantika, F., Azhar, Y., & Marthasari, G. I. (2021). Analisis Klasifikasi SMS Spam Menggunakan Logistic Regression. *Jurnal Sistem Cerdas*, *4*(3), 155–160.

Rizki, M., Arhami, M., & Huzeni. (2021). PERBAIKAN ALGORITMA NAIVE BAYES CLASSIFIER MENGGUNAKAN TEKNIK LAPLACIAN CORRECTION. *Jurnal Teknologi*, *21*(1), 39–45.

Sari, M., Purnomo, H. D., & Sembiring, I. (2022). Review : Algoritma Kriptografi Sistem Keamanan SMS di Android. *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, *2*(1), 11–15.

Shenita, E., & Suendri. (2023). Web-Based Village Fund Assistance Distribution Information System Using the Quota Based Method. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, *8*(2), 708–718.

Sutra Dewi, I. (2022). *Perlindungan Hukum Bagi Konsumen Atas Sms Spam Yang Dikirim Oleh Operator Seluler*.

Syahranitazli, & Samsudin. (2023). SISTEM INFORMASI GEOGRAFIS PERSEBARAN PONDOK PESANTREN KABUPATEN LANGKAT DAN BINJAI MENGGUNAKAN LEAFLET. *Jurnal Pendidikan Teknologi Informasi (JUKANTI)*, *6*(1), 2621–1467.

* Corresponding author