ANALISIS SENTIMEN MENGENAI PELAYANAN INTERNET SERVICE PROVIDER DI INDONESIA PADA MEDIA SOSIAL DENGAN MEMBANDINGKAN HASIL KINERJA ALGORITMA KLASIFIKASI NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

TUGAS AKHIR

Mohamad Afrizal
41517010058

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021

https://lib.mercubuana.ac.id/

ANALISIS SENTIMEN MENGENAI PELAYANAN INTERNET SERVICE PROVIDER DI INDONESIA PADA MEDIA SOSIAL DENGAN MEMBANDINGKAN HASIL KINERJA ALGORITMA KLASIFIKASI NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Mohamad Afrizal
41517010058

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021

https://lib.mercubuana.ac.id/

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:
NIM : 41517010058
Nama : Mohamad Afrizal
Judul Tugas Akhir : Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan di dalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 17 September 2021

SEPULUH RIBU RUPIAH 10000 METERAI TEMPEL 636AJX787036554

Mohamad Afrizal

UNIVERSITAS MERCU BUANA

https://lib.mercubuana.ac.id/

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Mohamad Afrizal
NIM	:	41517010058
Judul Tugas Akhir	:	Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

UNIVERSITAS
MERCU BUANA

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 17 September 2021

METERAI
TEMPEL
B65AJX787036553

Mohamad Afrizal

https://lib.mercubuana.ac.id/

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Mohamad Afrizal
NIM	:	41517010058
Judul Tugas Akhir	:	Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	v
		Jurnal Nasional Terakreditasi	v		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
	Disubmit/dipublikasikan di :	Nama Jurnal	: Seminar Nasional Ilmu Komputer dan Sistem Informasi (SNIKSI 2021)		
		ISSN	:		
		Link Jurnal	: https://conference.binus.ac.id/ocs/index.php/SNIKSI/		
		Link File Jurnal Jika Sudah di Publish	:		

UNIVERSITAS
MERCU BUANA

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Mengetahui
Dosen Pembimbing TA

Dr. Mujiono Sadikin, MT. CISA, CGEIT

Jakarta, 17 September 2021

SEPULUH RIBU RUPIAH
METERAI TEMPEL
ED9AJX787036555

Mohamad Afrizal

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41517010058
Nama	:	Mohamad Afrizal
Judul Tugas Akhir	:	Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 03 Agustus 2021

Menyetujui,

(Devi Fitrianah, Dr., MTI)
Dosen Penguji 1

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41517010058
Nama	:	Mohamad Afrizal
Judul Tugas Akhir	:	Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 03 Agustus 2021

Menyetujui,

(Desi Ramayanti, S.Kom., MT)
Dosen Penguji 2

UNIVERSITAS
MERCU BUANA

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41517010058
Nama	:	Mohamad Afrizal
Judul Tugas Akhir	:	Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 03 Agustus 2021

Menyetujui,

(Vina Ayumi, S.Kom., M.Kom)
Dosen Penguji 3

UNIVERSITAS MERCU BUANA

LEMBAR PENGESAHAN

NIM	:	41517010058
Nama	:	Mohamad Afrizal
Judul Tugas Akhir	:	Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 03 Agustus 2021

Menyetujui,

(Dr. Mujiono Sadikin, MT. CISA, CGEIT)
Dosen Pembimbing

UNIVERSITAS
MERCU BUANA

Mengetahui,

(Wawan Gunawan, S.Kom., MT)
Koord. Tugas Akhir Teknik Informatika

(Hery Derajad Wijaya, S.Kom., MM)
Ka. Prodi Teknik Informatika

https://lib.mercubuana.ac.id/

KATA PENGANTAR

Terimakasih dan rasa syukur saya panjatkan kepada Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, dikarenakan Tugas Akhir yang berjudul "Sentiment Analysis of Internet Service Provider in Indonesia on Twitters" dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Penulis menyadari bahwa pembuatan Tugas Akhir ini tidak terlepas dari bantuan dan bimbingan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua. Ayah dan Ibu, yang tak pernah lelah mendukung dan selalu percaya bahwa saya bisa menyelesaikan kuliah dengan baik, juga tak pernah luput mendoakan yang terbaik untuk proses meraih gelar sarjana bagi saya.
2. Ibu Dr. Ida Nurhaida, S.T., M.T selaku Dosen Pembimbing Akademik
3. Bapak Dr. Mujiono Sadikin, MT. CISA, CGEIT selaku Dosen Pembimbing Tugas Akhir yang telah memberikan masukan saat bimbingan dan meluangkan waktu sebagian besarnya untuk melakukan bimbingan dalam penyusunan tugas akhir ini hingga selesai.
4. Seluruh Dosen Program Studi Teknik Informatika yang sudah memberikan ilmu yang bermanfaat selama kuliah berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
5. Sahabat dan kerabat, yang telah percaya bahwa saya bisa melewati dan menyelesaikan Tugas Akhir ini dan mendapatkan gelar sarjana dengan baik, juga tidak pernah bosan memberi dukungan dan doa.
6. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan, terima kasih atas semua pelayanan dan arahannya.
7. Senior atas berbagai saran yang diberikan dan junior yang dengan semangat mendukung dikala bertemu.

MERCU BUANA

https://lib.mercubuana.ac.id/

8. Semua pihak dan personal yang tidak dapat disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Akhir kata, hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam eksperimen, cara penjelasan maupun kekeliruan penulisan. Untuk itu, kritik dan saran pembaca sangat dihargai dan diharapkan. Semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 3 Agustus 2021
Mohamad Afrizal

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/

DAFTAR ISI

UNIVERSITAS
MERCU BUANA

NASKAH JURNAL

SENTIMENT ANALYSIS OF INTERNET SERVICE PROVIDER IN INDONESIA ON TWITTERS

Mohamad Afrizal*, Mujiono Sadikin

Department of Informatics, Faculty of Computer Science, Universitas Mercu Buana
Jl. Raya Meruya Selatan, Kembangan, Jakarta, Indonesia 11650

*Corresponding author, e-mail: 41517010058@student.mercubuana.ac.id, mujiono.sadikin@mercubuana.ac.id

Abstract – *In suppressing the spread of COVID-19, the Indonesian government has established a PSBB policy or Large-Scale Social Restrictions. Many aspects and areas affected by the policy include education and offices. APJII or the Association of Indonesian Internet Service Providers explained that there was an increase in the number of internet users in Indonesia from last year's penetration of 64% to 73.7%. One of the reasons for this increase was the COVID-19 pandemic. On the Twitter platform, they often find various kinds of public responses that they give about the services of the Internet Service Provider, both negative and positive. In this study, sentiment analysis was conducted to determine public opinion on the performance of Internet Service Providers. The method used is the Naïve Bayes classification algorithm and Support Vector Machine assisted by RapidMiner and Python tools. The experimental results show that the Support Vector Machine algorithm provides the highest accuracy values of 93% and 92% for the two data tested, both Indihome and Firstmedia.*

Keywords: Internet, Naïve Bayes, Support Vector Machine, Algorithm, Sentiment

I. INTRODUCTION

In 2019, a new virus was discovered from Wuhan, China, namely the Corona virus (COVID-19). The wider the spread of this virus, including in Indonesia, has made the Indonesian government seek various ways to prevent the spread of this virus in Indonesia. One of the efforts made by the Indonesian government is to establish PSBB or Large-Scale Social Restrictions.

With the enactment of the PSBB, there are many aspects and areas that are affected by the government's policy, including education and office activities. All schools including universities in Indonesia are closed, requiring students and students to be encouraged to study online from home and also many offices must implement the WFH (Work From Home) policy or work from home online. Since the enactment of the PSBB policy, making internet needs very important for workers, teachers, lecturers, students and students to support teaching and learning activities and office activities.

APJII or the Association of Indonesian Internet Service Providers revealed through a survey that has been conducted, explaining that there is an increase in the number of internet users in Indonesia today from last year's penetration of 64.8% to 73.7%. One of the reasons for this increase was the COVID-19 pandemic. APJII also noted that the majority of

https://lib.mercubuana.ac.id/

internet users in Indonesia use Indihome and Firstmedia internet providers. (APJII, 2020).

The problem is, they often find various kinds of public responses that they put on Twitter regarding the services of the Internet Service Provider, both negative and positive. According to them, the services provided by the ISP are still considered unsatisfactory to their customers. Among them from internet speed, frequent technical problems, and so on. From the survey conducted by APJII and these problems, the authors found a solution by conducting research on sentiment analysis or public opinion on the services of the two providers, namely Indihome and Firstmedia.

Sentiment analysis is a branch of science from text mining, natural language programs and artificial intelligence. Sentiment analysis itself or also commonly referred to as opinion mining is one part of text mining. The process of grouping the text contained in a word, sentence or document and then determining whether the opinion expressed in the sentence or document is positive or negative is called sentiment analysis (Putra Nuansa, 2017). This field conducts the study of people's opinions, sentiments, evaluations, behavior and emotions towards an entity such as products, services, organizations, individuals, problems, topics, events and their attributes (Prager, nd).

It is hoped that later this research can help internet providers in responding to these problems and assist providers in evaluating the products and services they provide to their customers.

II. METHODS

This study uses a quantitative method where the data that has been collected is processed through the Twitter API using Rapid Miner tools. In this study, python and r tools were assisted to perform preprocessing and labeling on Google Colab. In this study, experiments were also conducted by comparing the performance of the Naïve Bayes classification algorithm and the Support Vector Machine to find out which algorithm has the best level of accuracy. The research flow is illustrated in Figure 1. Research Methodology.

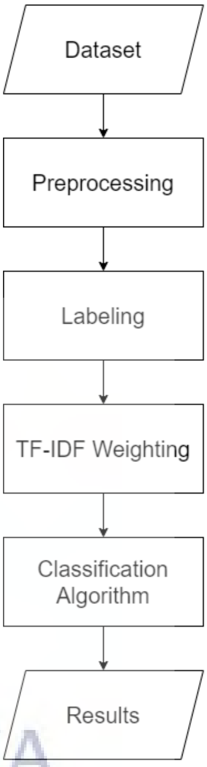


Figure 1. Research Methodology

The initial stage of this research is to prepare the Indihome and Firstmedia datasets. Dataset collection is done using Rapidminer tools taken through the Twitter API by entering relevant keywords according to the topic. The next stage is the preprocessing stage to clean the dataset that has been collected. In this study, the preprocessing stages carried out are cleansing, case folding, tokenization, stopword removal, stemming and remove duplicates.

After the preprocessing stage, the next step is the labeling stage to label the cleaned dataset. In this study, the labeling process was assisted by using the python

https://lib.mercubuana.ac.id/

and r tools on Google Colab. Labeling the dataset is useful to simplify the classification process.

The next stage is the TF-IDF weighting. TF-IDF weighting is the process of assigning weights to each word contained in a document. At the TF-IDF weighting stage, it is carried out to convert data in the form of text into numbers so that it can be processed by a computer at the classification stage (Oyebode & Orji, 2019).

The last stage is the classification process using the Naïve Bayes algorithm and the Support Vector Machine. Before the classification stage is carried out, the data is first divided into training data and test data. The training data is used for the learning process on the Naïve Bayes algorithm and Support Vector Machine. The next step is to use test data for the process of testing the Naïve Bayes classification model and Support Vector Machine.

In this study, the Support Vector Machine algorithm is used because it is suitable for text classification and how the algorithm works that can overcome outlier data. While the Naïve Bayes algorithm is suitable for text data because Nave Bayes uses the probability method, word opportunities appear.

UNIVERSITAS MERCU BUANA

A. Dataset

At this stage, data collection is carried out through the Twitter API using the Rapid Miner tools. The process carried out is to enter keywords that are relevant to the topic discussed. The author uses several keywords to collect related data, including “internet Indihome”, “koneksi Indihome”, “jaringan Indihome”, “pelayanan Indihome” and “provider Indihome” while for Firstmedia tweets using the keywords “internet Firstmediacares”, “koneksi Firstmediacares”, “jaringan Firstmediacares”, “pelayanan Firstmediacares” and “provider Firstmediacares”. From the tweets that have been collected, the data that has been collected is contained in table 1. Twitter data.

Table 1. Twitter data

Provider	Tweets
Indihome	14097
Firstmedia	13798

B. Preprocessing Data

Data preprocessing aims to clean datasets from raw data into ready-to-use data in order to facilitate the classification process into positive and negative. Data pre-processing is a data mining technique that involves transforming raw data into an easy-to-understand format. The data pre-precossing step is needed to solve several types of problems including noisy data, data redundancy, missing data values, etc. (Syadid, 2019). The stages or steps carried out in this study consist of 5 steps, including cleansing, case folding, tokenization, stopword removal and stemming. The explanation of the steps taken:

- Cleansing: Cleansing in this study aims to remove RT, username, hashtag, number and URL on tweets.

  Example of a tweet: Malem teh pengen santai maen game abis kuliah seharian. Ehhhh udah 2 jam internet mati. Gimana nih @IndiHome

  Cleansing result: Malem teh pengen santai maen game abis kuliah seharian. Ehhhh udah jam internet mati. Gimans nih?

- Case folding: The text in tweets tends to have various types of writing, one of which is upper and lower case writing. The solution to this problem is to change the text in lowercase (Nur Habibi & Sunjana, 2019). Case folding is used to make searching easier. Not all data are consistent in the use of capital letters (Gunawan et al., 2018).

Case folding results: malem teh pengen santai maen game abis kuliah seharian. ehhhh udah jam internet mati. gimana nih

- Tokenization: Tokenization in this study aims to separate words in tweets into individual words (de Oliveira et al., 2020). The result of the word split will be represented as a token.

Tokenization results: malem, teh, pengen, santai, maen, game, abis, kuliah, seharian, ehhhh, udah, jam, internet, mati, gimana, nih,

- Stopword removal: at this stage the conjunction is removed. (Tineges et al., 2020). If there is a word that is not in the stopword list, it will be removed. A stopword is defined as a term that is irrelevant to the main subject of the database even though the word is often present in the document. The following are examples of stopwords in Indonesian: yang, juga, dari, dia, kami, kamu, aku, saya, ini, itu, atau, dan, pada, dengan, adalah, yaitu, ke, tak, tidak, di, pada, jika, maka, ada, pun, lain, saja, hanya, namun, seperti, kemudian, etc. (Syadid, 2019).

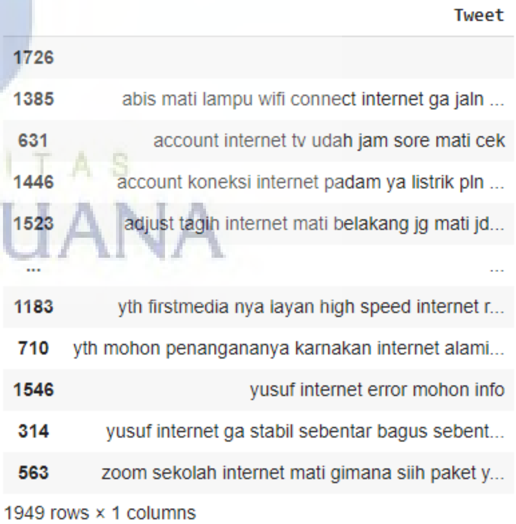
Stopword removal results: malem, teh, santai, maen, game, abis, kuliah, seharian, ehhhh, udah, jam, internet, mati, gimana, nih,

- Stemming: this stage is using stemming to form the basic words of the tokenization process. This study uses a library provided by Python, namely the Sastrawi library. Words that appear in documents often have many morphological variants. Therefore, every word that is not stopwords is reduced to a suitable stemmed word (term). The word is stemmed to get its root form by removing the prefix or suffix. In this way, we get groups of words that have similar meanings but differ in syntactic form from one another. The group can be represented by one particular word. For example, the word menyebutkan, tersebut, disebut, can be said to be similar or a group and can be represented by a common word called sebut (Syadid, 2019).

Stemming results: malem, teh, santai, maen, game, abis, kuliah, hari, ehhhh, udah, jam, internet, mati, gimana, nih,

- Remove Duplicate: this step is used to remove data that has duplicates.

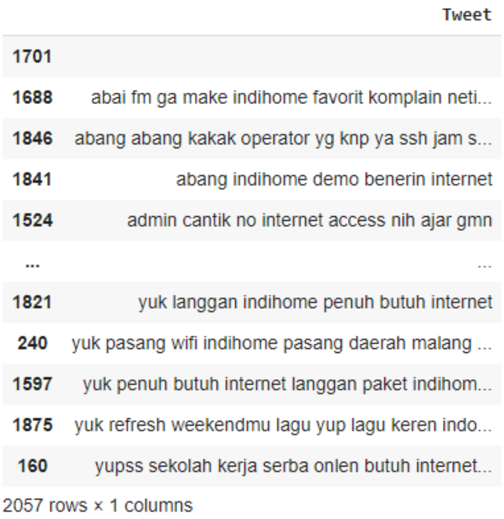
UNIVERSITAS MERCU BUANA



	Tweet
1726	
1385	abis mati lampu wifi connect internet ga jaln ...
631	account internet tv udah jam sore mati cek
1446	account koneksi internet padam ya listrik pln ...
1523	adjust tagih internet mati belakang jg mati jd...
...	...
1183	yth firstmedia nya layan high speed internet r...
710	yth mohon penangananya karnakan internet alami...
1546	yusuf internet error mohon info
314	yusuf internet ga stabil sebentar bagus sebent...
563	zoom sekolah internet mati gimana siih paket y...

1949 rows × 1 columns

Figure 2. Firstmedia Preprocessing



	Tweet
1701	
1688	abai fm ga make indihome favorit komplain neti...
1846	abang abang kakak operator yg knp ya ssh jam s...
1841	abang indihome demo benerin internet
1524	admin cantik no internet access nih ajar gmn
...	...
1821	yuk langgan indihome penuh butuh internet
240	yuk pasang wifi indihome pasang daerah malang ...
1597	yuk penuh butuh internet langgan paket indihom...
1875	yuk refresh weekendmu lagu yup lagu keren indo...
160	yupss sekolah kerja serba onlen butuh internet...

2057 rows × 1 columns

Figure 3. Preprocessing Indihome

After doing the data preprocessing step, the preprocessed tweet data becomes 2057 for Indihome tweets and 1949 for Firstmedia tweets, then stored in .csv format for labeling process.

C. Labeling

At this stage, the data that has been cleaned at the preprocessing stage is classified as positive and negative using the R language through Google Colab. The labeling of words is adjusted in the Indonesian dictionary which has been integrated in the Google Drive file. The Indonesian dictionary consists of 2 files stored in a .csv file consisting of a positive dictionary and a negative dictionary in Indonesian.

Words that are labeled positive are words that are detected according to the positive dictionary, while words that are labeled negative are words that are detected according to the negative dictionary. Each detected word is given a score to assess the sentiment class. For positive words, it is given a value of 1 while for negative words it is given a value of -1. If there is a word that is not in the positive or negative dictionary, it is assigned a value of 0.

Scoring is done by counting the number of points in each word in one sentence. If the value is >= 0 then it is labeled as a positive tweet sentiment, otherwise if the value is < 0 then it is labeled as a negative tweet sentiment. Sentiment tweets are assigned a value of 1 for positive classification and 0 for negative classification. The following is the data after the sentiment classification is given:

Table 2 Automatic Labeling Value

Provider	Positive	Negative
Indihome	980	1076
firstmedia	827	1121

In this study, experiments were also carried out using manual labeling of the two datasets. Positive labels are tweets that have praise while negative tweets are in the form of harsh and sarcastic words. Here is the data using manual labeling:

Table 3 Manual Labeling Value

Provider	Positive	Negative
Indihome	1496	560
Firstmedia	1461	487

Because there are unbalanced sentiments, resampling is carried out using a random method which is useful for balancing the data. The method used is to increase the most data, namely upsampling. The upsampling process is a resampling method by increasing the result of the minority value in proportion to the majority value. The following are the classification values after upsampling:

Table 4 Automatic Labeling Resampling Value

Provider	Positive	Negative
Indihome	1076	1076
firstmedia	1121	1121

Table 5 Manual Labeling Resampling Value

Provider	Positive	Negative
Indihome	1496	1496
firstmedia	1461	1461

UNIVERSITAS MERCU BUANA

D. TF-IDF Weighting

Word weighting is the process of assigning a weight to each word contained in a document. In searching for ranking information based on word frequency, one of the most popular methods is the TF IDF (Term Frequency - Inversed Document Frequency) method (Gunawan et al., 2018). The TF-IDF (Term Frequency - Inversed Document Frequency) method is a method of finding ranking information based on word frequency, and is one of the most frequently used methods (Gunawan et al., 2018).

To find out how important a word represents a sentence, a weighting or calculation is carried out. The scoring in the TF-IDF is based on the frequency with which words appear in the document (Arsya Monica Pravina, Imam Cholissodin, 2019). TF-IDF presents word frequency scores, especially for words of interest, such as words that often appear in one document but not all documents (Ahuja et al., 2019).

E. Classification Algorithm

This research uses the Support Vector Machine and Naïve Bayes classification algorithms. A job of assessing data objects to include them in a certain class from a number of available classes is called classification (Syadid, 2019). Support Vector Machine (SVM) is a relatively new technique for making predictions, both in the case of classification and regression. The classification concept with the Support Vector Machine is to find the best hyperplane that functions as a separator of two data classes. The Support Vector Machine algorithm works by optimally separating data using hyperplane distance measurements from the nearest point rather than looking for the maximum point to maximize the distance between class labels based on class membership limit measurements (Ramayanti & Salamah, 2018). Furthermore, this method can be used as a reliable method in solving data classification problems, the problem is solved by solving the Lagrangian equation which is a dual form of Support Vector Machine through quadratic programming (Fiska, 2017). The Support Vector Machine itself has the basic principle of a linear classifier, namely classification cases that can be separated linearly (Fitri, 2020). The Support Vector Machine is also able to work on high-dimensional datasets using kernel tricks (Rofiqoh et al., 2017). The Support Vector Machine itself has the basic principle of a linear classifier, namely classification cases that can be separated linearly (Fitri, 2020). The Support Vector Machine is also able to work on high-dimensional datasets using kernel tricks (Rofiqoh et al., 2017). The Support Vector Machine itself has the basic principle of a linear classifier, namely classification cases that can be separated linearly (Fitri, 2020). The Support Vector Machine is also able to work on high-dimensional datasets using kernel tricks (Rofiqoh et al., 2017).

Naïve Bayes Classifier is a data mining algorithm learning technique that utilizes probability and statistical methods. Naïve Bayes Classifier in classifying there are two important processes, namely learning (training) and testing (AFSHOH, 2017). The Naive Bayes classification algorithm according to quoting is an algorithm used to find the highest probability value to classify test data in the most appropriate category, Bayesian classifier has higher accuracy and speed, especially when applied to large datasets (Juanita, 2020). Naïve Bayes is one of the data mining algorithms that is easy to use and has a fast processing time, is easy to implement with a fairly simple structure and has a high level of effectiveness (Antinasari et al., 2017).

III. RESULTS AND DISCUSSION

https://lib.mercubuana.ac.id/

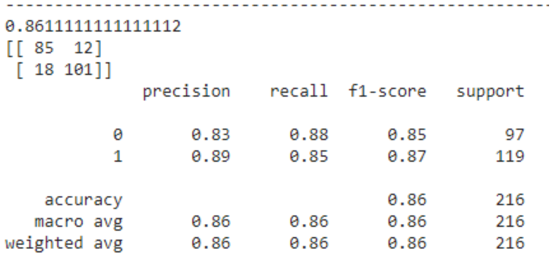
In this study, the authors conducted experiments on two operator case studies, namely Indihome and Firstmedia. Experiments were carried out using two algorithms, namely Support Vector Machine and Naïve Bayes. There are three scenarios where the percentage of data is divided. The distribution of data presentation is the separation of training and testing data based on the percentage, for example 90%: 10% means 90% is training data and 10% is testing data. In this study, the percentage distribution was divided into three experimental scenarios, namely the first experiment using data sharing of 90%: 10%, the second experiment using data sharing of 80%: 20% and the third experiment using data sharing of 70%: 30%. This study also compares the accuracy values carried out using the manual labeling method and the automatic labeling method.

The accuracy of each scenario can be different, because the model that has been formed is evaluated and tested using the concept of a confusion matrix. The formula used in determining accuracy is:

Accuracy = (TP + TN) / (TP + FP + TN + FN) * 100

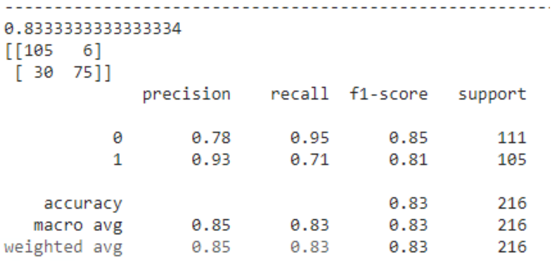
TP = True Positive
TN = True Negative
FP = False Positive
FN = False Negative

The results of the model performance using the automatic labeling method and data sharing or percentage split 90% : 10% that the Indihome case study has an accuracy value of 86% for the SVM algorithm and 83% for the Naïve Bayes algorithm. Meanwhile, the Firstmedia case study has an accuracy value of 89% for the SVM algorithm and 88% for the Naïve Bayes algorithm. The following are the results of the percentage split performance of 90%: 10%.



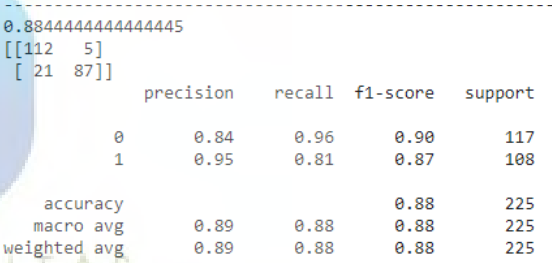
```
------------------------------------------------------------
0.8611111111111112
[[ 85  12]
 [ 18 101]]
              precision    recall  f1-score   support

           0       0.83      0.88      0.85        97
           1       0.89      0.85      0.87       119

    accuracy                           0.86       216
   macro avg       0.86      0.86      0.86       216
weighted avg       0.86      0.86      0.86       216
```

Figure 4. SVM Indihome Split 90:10 Automatic Labeling



```
------------------------------------------------------------
0.8333333333333334
[[105   6]
 [ 30  75]]
              precision    recall  f1-score   support

           0       0.78      0.95      0.85       111
           1       0.93      0.71      0.81       105

    accuracy                           0.83       216
   macro avg       0.85      0.83      0.83       216
weighted avg       0.85      0.83      0.83       216
```

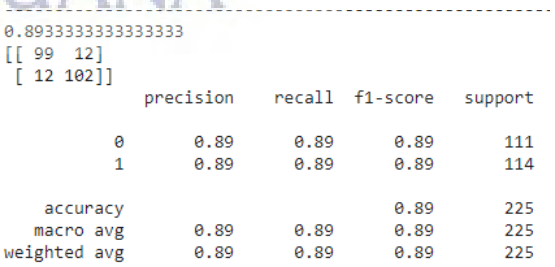
Figure 5. Naïve Bayes Indihome Split 90:10 Automatic Labeling



```
------------------------------------------------------------
0.8844444444444445
[[112   5]
 [ 21  87]]
              precision    recall  f1-score   support

           0       0.84      0.96      0.90       117
           1       0.95      0.81      0.87       108

    accuracy                           0.88       225
   macro avg       0.89      0.88      0.88       225
weighted avg       0.89      0.88      0.88       225
```

Figure 6. Naïve Bayes Firstmedia Split 90:10 Automatic Labeling

UNIVERSITAS MERCU BUANA

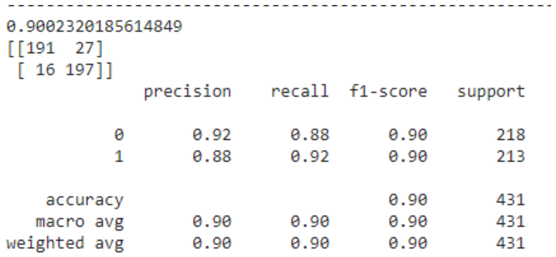


```
------------------------------------------------------------
0.8933333333333333
[[ 99  12]
 [ 12 102]]
              precision    recall  f1-score   support

           0       0.89      0.89      0.89       111
           1       0.89      0.89      0.89       114

    accuracy                           0.89       225
   macro avg       0.89      0.89      0.89       225
weighted avg       0.89      0.89      0.89       225
```

Figure 7. SVM Firstmedia Split 90:10 Automatic Labeling

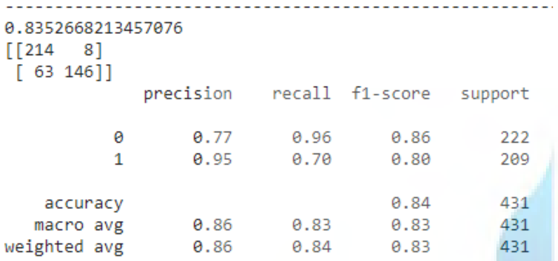
The results of the model performance using the automatic labeling method and data sharing or percentage split 80% : 20% that the Indihome case study has an accuracy value of 90% for the SVM algorithm and 84% for the Naïve Bayes algorithm. Meanwhile, the Firstmedia case study has an accuracy value of 92% for the

SVM algorithm and 91% for the Naïve Bayes algorithm. The following are the results of the 80% : 20% percentage split performance.



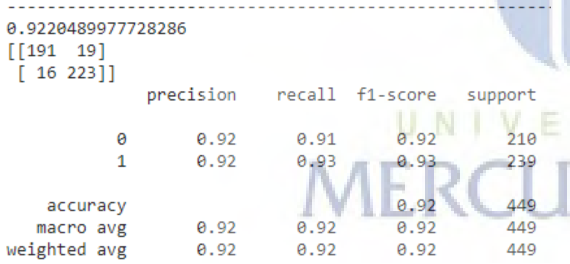
```
------------------------------------------------------------
0.9002320185614849
[[191  27]
 [ 16 197]]
              precision    recall  f1-score   support

           0       0.92      0.88      0.90       218
           1       0.88      0.92      0.90       213

    accuracy                           0.90       431
   macro avg       0.90      0.90      0.90       431
weighted avg       0.90      0.90      0.90       431
```

Figure 8. SVM Indihome Split 80:20 Automatic Labeling



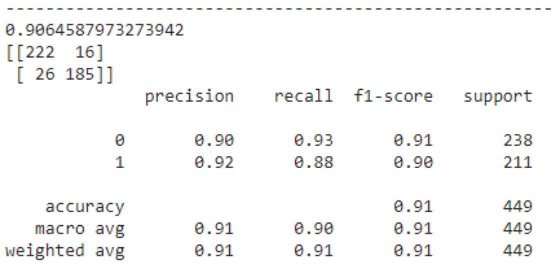
```
------------------------------------------------------------
0.8352668213457076
[[214   8]
 [ 63 146]]
              precision    recall  f1-score   support

           0       0.77      0.96      0.86       222
           1       0.95      0.70      0.80       209

    accuracy                           0.84       431
   macro avg       0.86      0.83      0.83       431
weighted avg       0.86      0.84      0.83       431
```

Figure 9. Naïve Bayes Indihome Split 80:20 Automatic Labeling



```
------------------------------------------------------------
0.9220489977728286
[[191  19]
 [ 16 223]]
              precision    recall  f1-score   support

           0       0.92      0.91      0.92       210
           1       0.92      0.93      0.93       239

    accuracy                           0.92       449
   macro avg       0.92      0.92      0.92       449
weighted avg       0.92      0.92      0.92       449
```

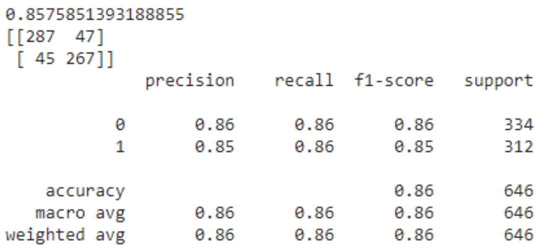
Figure 10. SVM Firstmedia Split 80:20 Automatic Labeling



```
------------------------------------------------------------
0.9064587973273942
[[222  16]
 [ 26 185]]
              precision    recall  f1-score   support

           0       0.90      0.93      0.91       238
           1       0.92      0.88      0.90       211

    accuracy                           0.91       449
   macro avg       0.91      0.90      0.91       449
weighted avg       0.91      0.91      0.91       449
```

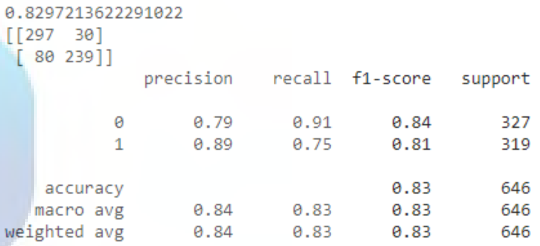
Figure 11. Naïve Bayes Firstmedia Split 80:20 Automatic Labeling

The results of the model performance using the automatic labeling method and data sharing or percentage split 70% : 30% that the Indihome case study has an accuracy value of 86% for the SVM algorithm and 83% for the Naïve Bayes algorithm. Meanwhile, the Firstmedia case study has an accuracy value of 90% for the SVM algorithm and 88% for the Naïve Bayes algorithm. The following are the results of the 70% : 30% percentage split performance.



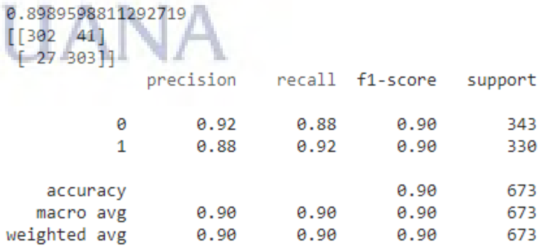
```
0.8575851393188855
[[287  47]
 [ 45 267]]
              precision    recall  f1-score   support

           0       0.86      0.86      0.86       334
           1       0.85      0.86      0.85       312

    accuracy                           0.86       646
   macro avg       0.86      0.86      0.86       646
weighted avg       0.86      0.86      0.86       646
```

Figure 12. SVM Indihome Split 70:30 Automatic Labeling



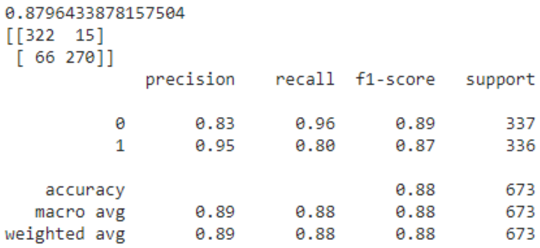
```
0.8297213622291022
[[297  30]
 [ 80 239]]
              precision    recall  f1-score   support

           0       0.79      0.91      0.84       327
           1       0.89      0.75      0.81       319

    accuracy                           0.83       646
   macro avg       0.84      0.83      0.83       646
weighted avg       0.84      0.83      0.83       646
```

Figure 13. Naïve Bayes Indihome Split 70:30 Automatic Labeling



```
0.8989598811292719
[[302  41]
 [ 27 303]]
              precision    recall  f1-score   support

           0       0.92      0.88      0.90       343
           1       0.88      0.92      0.90       330

    accuracy                           0.90       673
   macro avg       0.90      0.90      0.90       673
weighted avg       0.90      0.90      0.90       673
```

Figure 14. SVM Firstmedia 70:30 Automatic Labeling

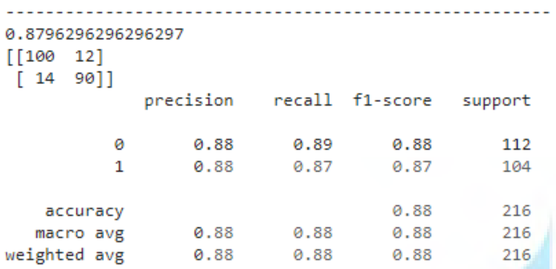


```
0.8796433878157504
[[322  15]
 [ 66 270]]
              precision    recall  f1-score   support

           0       0.83      0.96      0.89       337
           1       0.95      0.80      0.87       336

    accuracy                           0.88       673
   macro avg       0.89      0.88      0.88       673
weighted avg       0.89      0.88      0.88       673
```

Figure 15. Naïve Bayes Firstmedia 70:30 Automatic Labeling

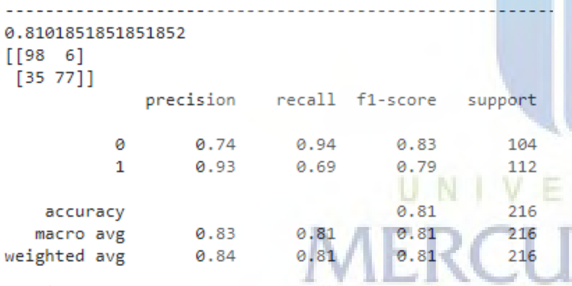
UNIVERSITAS MERCU BUANA

The results of the model performance using the manual labeling method and data sharing or percentage split 90% : 10% that the Indihome case study has an accuracy value of 88% for the SVM algorithm and 81% for the Naïve Bayes algorithm. Meanwhile, the Firstmedia case study has an accuracy value of 91% for the SVM algorithm and 84% for the Naïve Bayes algorithm. The following are the results of the 90% : 10% percentage split performance.



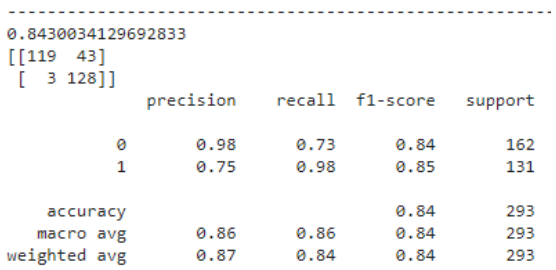
```
-------------------------------------------------------
0.8796296296296297
[[100  12]
 [ 14  90]]
              precision    recall  f1-score   support

           0       0.88      0.89      0.88       112
           1       0.88      0.87      0.87       104

    accuracy                           0.88       216
   macro avg       0.88      0.88      0.88       216
weighted avg       0.88      0.88      0.88       216
```

Figure 16. SVM Indihome Split 90:10 Manual Labeling



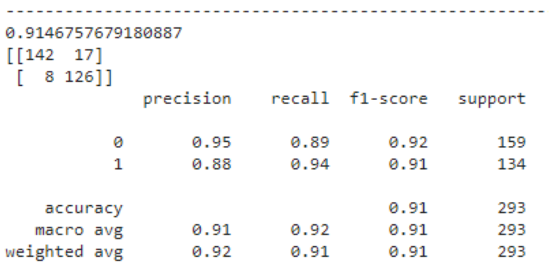
```
-------------------------------------------------------
0.8101851851851852
[[98  6]
 [35 77]]
              precision    recall  f1-score   support

           0       0.74      0.94      0.83       104
           1       0.93      0.69      0.79       112

    accuracy                           0.81       216
   macro avg       0.83      0.81      0.81       216
weighted avg       0.84      0.81      0.81       216
```

Figure 17. Naïve Bayes Indihome Split 90:10 Manual Labeling



```
-------------------------------------------------------
0.8430034129692833
[[119  43]
 [  3 128]]
              precision    recall  f1-score   support

           0       0.98      0.73      0.84       162
           1       0.75      0.98      0.85       131

    accuracy                           0.84       293
   macro avg       0.86      0.86      0.84       293
weighted avg       0.87      0.84      0.84       293
```

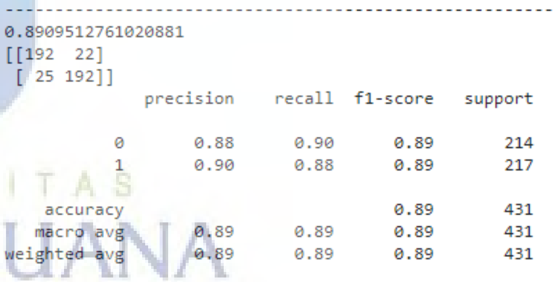
Figure 18. Naïve Bayes Firstmedia Split 90:10 Labeling Manual



```
-------------------------------------------------------
0.9146757679180887
[[142  17]
 [  8 126]]
              precision    recall  f1-score   support

           0       0.95      0.89      0.92       159
           1       0.88      0.94      0.91       134

    accuracy                           0.91       293
   macro avg       0.91      0.92      0.91       293
weighted avg       0.92      0.91      0.91       293
```

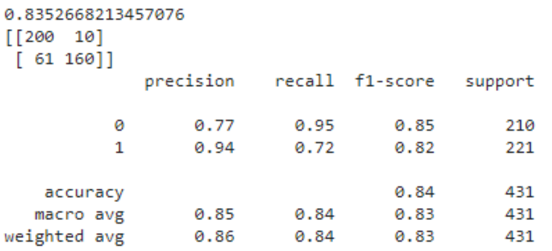
Figure 19. SVM Firstmedia Split 90:10 Manual Labeling

The results of the model performance using manual labeling methods and data sharing or percentage split 80% : 20% that the Indihome case study has an accuracy value of 89% for the SVM algorithm and 84% for the Naïve Bayes algorithm. Meanwhile, the Firstmedia case study has an accuracy value of 93% for the SVM algorithm and 85% for the Naïve Bayes algorithm. The following are the results of the 80% : 20% percentage split performance.



```
-------------------------------------------------------
0.8909512761020881
[[192  22]
 [ 25 192]]
              precision    recall  f1-score   support

           0       0.88      0.90      0.89       214
           1       0.90      0.88      0.89       217

    accuracy                           0.89       431
   macro avg       0.89      0.89      0.89       431
weighted avg       0.89      0.89      0.89       431
```

Figure 20. SVM Indihome Split 80:20 Manual Labeling

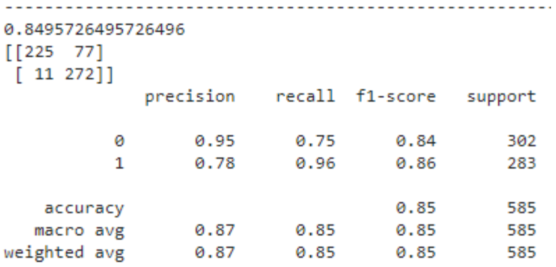


```
0.8352668213457076
[[200  10]
 [ 61 160]]
              precision    recall  f1-score   support

           0       0.77      0.95      0.85       210
           1       0.94      0.72      0.82       221

    accuracy                           0.84       431
   macro avg       0.85      0.84      0.83       431
weighted avg       0.86      0.84      0.83       431
```

Figure 21. Naïve Bayes Indihome Split 80:20 Labeling Manual

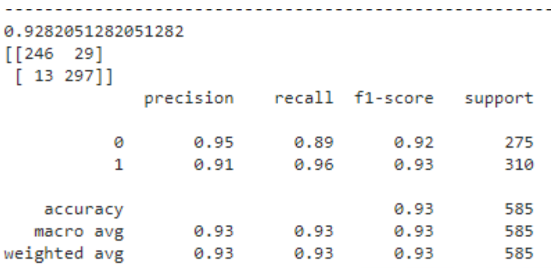
UNIVERSITAS MERCU BUANA

https://lib.mercubuana.ac.id/



```
------------------------------------------------------
0.8495726495726496
[[225  77]
 [ 11 272]]
              precision    recall  f1-score   support

           0       0.95      0.75      0.84       302
           1       0.78      0.96      0.86       283

    accuracy                           0.85       585
   macro avg       0.87      0.85      0.85       585
weighted avg       0.87      0.85      0.85       585
```

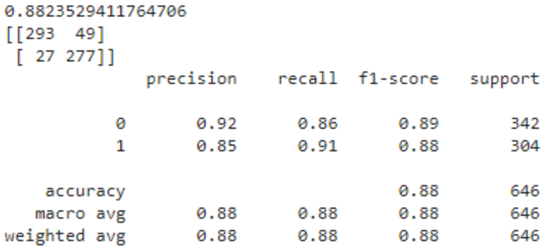
Figure 22. Naïve Bayes Firstmedia Split 80:20 Labeling Manual



```
------------------------------------------------------
0.9282051282051282
[[246  29]
 [ 13 297]]
              precision    recall  f1-score   support

           0       0.95      0.89      0.92       275
           1       0.91      0.96      0.93       310

    accuracy                           0.93       585
   macro avg       0.93      0.93      0.93       585
weighted avg       0.93      0.93      0.93       585
```

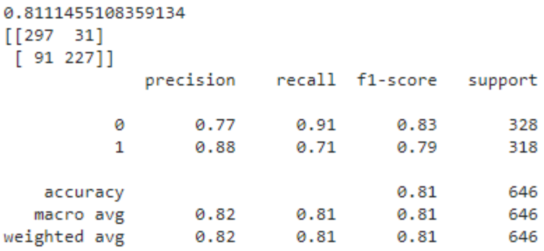
Figure 23. SVM Firstmedia Split 80:20 Labeling Manual

The results of the model performance using manual labeling methods and data sharing or percentage split 70% : 30% that the Indihome case study has an accuracy value of 88% for the SVM algorithm and 81% for the Naïve Bayes algorithm. Meanwhile, the Firstmedia case study has an accuracy value of 89% for the SVM algorithm and 82% for the Naïve Bayes algorithm. The following are the results of the 70% : 30% percentage split performance.



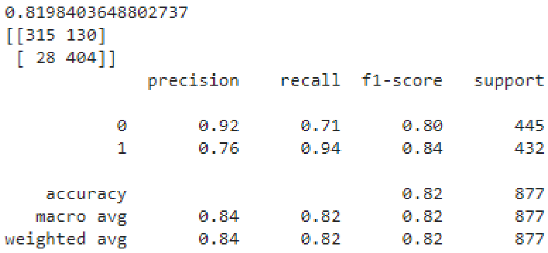
```
0.8823529411764706
[[293  49]
 [ 27 277]]
              precision    recall  f1-score   support

           0       0.92      0.86      0.89       342
           1       0.85      0.91      0.88       304

    accuracy                           0.88       646
   macro avg       0.88      0.88      0.88       646
weighted avg       0.88      0.88      0.88       646
```

Figure 24. SVM Indihome Split 70:30 Manual Labeling



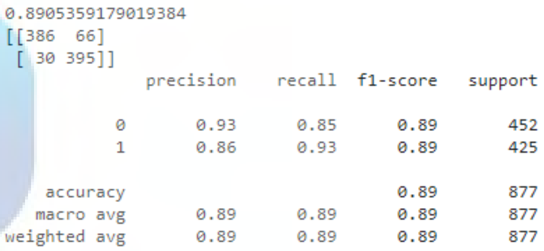
```
0.8111455108359134
[[297  31]
 [ 91 227]]
              precision    recall  f1-score   support

           0       0.77      0.91      0.83       328
           1       0.88      0.71      0.79       318

    accuracy                           0.81       646
   macro avg       0.82      0.81      0.81       646
weighted avg       0.82      0.81      0.81       646
```

Figure 25. Naïve Bayes Indihome Split 70:30 Labeling Manual



```
0.8198403648802737
[[315 130]
 [ 28 404]]
              precision    recall  f1-score   support

           0       0.92      0.71      0.80       445
           1       0.76      0.94      0.84       432

    accuracy                           0.82       877
   macro avg       0.84      0.82      0.82       877
weighted avg       0.84      0.82      0.82       877
```

Figure 26. Naïve Bayes Firstmedia Split 70:30 Labeling Manual



```
0.8905359179019384
[[386  66]
 [ 30 395]]
              precision    recall  f1-score   support

           0       0.93      0.85      0.89       452
           1       0.86      0.93      0.89       425

    accuracy                           0.89       877
   macro avg       0.89      0.89      0.89       877
weighted avg       0.89      0.89      0.89       877
```

Figure 27. SVM Firstmedia Split 70:30 Manual Labeling

IV. CONCLUSION

The results of this study are experiments using the Support Vector Machine algorithm have the highest accuracy values in the three percentage split experimental scenarios on the two data models tested, namely Indihome and Firstmedia using automatic and manual labeling. The results of the discussion are summarized in the following table:

Table 6 Comparison of Automatic Labeling Accuracy

Test	Indihome		Firstmedia	
	SVM	NB	SVM	NB
90:10	86%	83%	89%	88%
80:20	90%	84%	92%	91%
70:30	86%	83%	90%	88%

UNIVERSITAS MERCU BUANA

Table 7 Comparison of Manual Labeling Accuracy

Test	Indihome		Firstmedia	
	SVM	NB	SVM	NB
90:10	88%	81%	91%	84%
80:20	89%	84%	93%	85%
70:30	88%	81%	89%	82%

From the following table, it is concluded that experiments using manual labeling are more representative in determining data classes. The 80:20 percentage split model has the best accuracy value on the data tested by both Indihome and Firstmedia with an accuracy value of 90% and 92% in the automatic labeling experiment and 89% and 93% in the manual labeling experiment. The proportion of 80:20 data from the training and testing datasets has better results because it provides an evaluation value that is close to balance after the tuning process. This experiment shows that the greater the number of training datasets, the better the evaluation value will be obtained because there will be a lot of learning processes that occur in the training dataset. So it can be concluded that the 80:20 percentage split model using the Support Vector Machine algorithm is the best experimental scenario in this study.

REFERENCES

AFSHOH, F. (2017). Analisis Sentimen Menggunakan Naive Bayes Untuk Melihat Persepsi Masyarakat Terhadap Kenaikan Harga Jual Rokok Pada Media Sosial Twitter. *Informatika, Program Studi Komunikasi, Fakultas Informatika, D A N Surakarta, Universitas Muhammadiyah*, 1–17.

Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, *152*(January), 341–348. https://doi.org/10.1016/j.procs.2019.05.008

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Ilmiah*, *10*(2), 160–165. https://doi.org/10.33096/ilkom.v10i2.303.160-165

Antinasari, P., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tentang Opini Film Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes Dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *1*(12), 1718–1724. http://j-ptiik.ub.ac.id

APJII. (2020). Laporan Survei Internet APJII 2019 – 2020. *Asosiasi Penyelenggara Jasa Internet Indonesia*, *2020*, 1–146. https://apjii.or.id/survei

Arsya Monica Pravina, Imam Cholissodin, P. P. A. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(3), 2789–2797. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793

de Oliveira, G. A., Albuquerque, R. de O., de Andrade, C. A. B., de Sousa, R. T., Orozco, A. L. S., & Villalba, L. J. G. (2020). Anonymous real-time analytics monitoring solution for decision making supported by sentiment analysis. *Sensors (Switzerland)*, *20*(16), 1–29. https://doi.org/10.3390/s20164557

Fiska, R. R. (2017). Penerapan Teknik Data Mining dengan Metode Support Vector Machine. *Sains Dan Teknologi Informasi (SATIN)*, *3*(1).

Fitri, E. (2020). Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive

https://lib.mercubuana.ac.id/

Bayes, Random Forest Dan Support Vector Machine. *Jurnal Transformatika*, *18*(1), 71. https://doi.org/10.26623/transformatika.v18i1.2317

Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *4*(2), 113. https://doi.org/10.26418/jp.v4i2.27526

Juanita, S. (2020). Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *Jurnal Media Informatika Budidarma*, *4*(3), 552. https://doi.org/10.30865/mib.v4i3.2140

Nur Habibi, M., & Sunjana. (2019). Analysis of Indonesia Politics Polarization before 2019 President Election Using Sentiment Analysis and Social Network Analysis. *International Journal of Modern Education and Computer Science*, *11*(11), 22–30. https://doi.org/10.5815/ijmecs.2019.11.04

Oyebode, O., & Orji, R. (2019). Social Media and Sentiment Analysis: The Nigeria Presidential Election 2019. *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2019*, *October*, 140–146. https://doi.org/10.1109/IEMCON.2019.8936139

Prager, J. (n.d.). *Open-Domain Open-Domain Question – Answering*.

Putra Nuansa, E. (2017). Analisis Sentimen Pengguna Twitter Terhadap Pemilihan Gubernur Dki Jakarta Dengan Metode Naïve Bayesian Classification Dan Support Vector Machine. *Institut Teknologi Sepuluh Nopember Surabaya*, 1–101.

Ramayanti, D., & Salamah, U. (2018). Complaint Classification Using Support Vector Machine for Indonesian Text Dataset. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *3*(7), 179–184.

Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *1*(12), 1725–1732. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628

Syadid, F. (2019). Analisis Sentimen Komentar Netizen Terhadap Calon Presiden Indonesia 2019 Dari Twitter Menggunakan Algoritma Term Frequency-Invers Document Frequency (Tf- Idf) Dan Metode Multi Layer Perceptron (Mlp) Neural Network. *Skripsi Universitas Islam Negeri Syarif Hidayatullah Jakarta*, 1–89.

Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, *4*(3), 650. https://doi.org/10.30865/mib.v4i3.2181

UNIVERSITAS MERCU BUANA

https://lib.mercubuana.ac.id/

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah terlampir sebelumnya dengan judul "Analisis Sentimen Mengenai Pelayanan Internet Service Provider di Indonesia Pada Media Sosial Dengan Membandingkan Hasil Kinerja Algoritma Klasifikasi Naïve Bayes dan Support Vector Machine". Kertas kerja ini berisi semua material hasil penelitan Tugas Akhir. Di dalam kertas kerja ini disajikan beberapa bagian yang terdiri dari literature review, dataset yang digunakan, tahapan eksperimen, dan hasil eksperimen secara keseluruhan.

Bagian I membahas mengenai literature review yang berisi artikel jurnal sebelumnya yang menjadi dasar atau landasan dalam penelitian ini. Bagian II menjelaskan tentang source code yang digunakan pada penelitian ini. Bagian III menjelaskan mengenai dataset yang digunakan. Bagian IV memuat tahapan eksperimen yang disajikan dalam gambar beserta penjelasan dari tiap tahapan. Bagian V merupakan bagian terakhir dari kertas kerja ini yang menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan, meliputi penjelasannya.

UNIVERSITAS
MERCU BUANA

https://lib.mercubuana.ac.id/