

**ANALISIS SENTIMEN MASYARAKAT TERHADAP  
PROGRAM MAGANG BERSERTIFIKAT PADA MEDIA  
SOSIAL TWITTER MENGGUNAKAN KLASIFIKASI  
*NAÏVE BAYES CLASSIFIER* (NBC)**



**SKRIPSI**

**FAJAR SEFULOH  
1803040079**

**TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN SAINS  
UNIVERSITAS MUHAMMADIYAH PURWOKERTO  
JANUARI 2023**

**ANALISIS SENTIMEN MASYARAKAT TERHADAP  
PROGRAM MAGANG BERSERTIFIKAT PADA MEDIA  
SOSIAL TWITTER MENGGUNAKAN KLASIFIKASI  
*NAÏVE BAYES CLASSIFIER* (NBC)**



**SKRIPSI**

diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana  
Komputer

**FAJAR SEFULOH  
1803040079**

**TEKNIK INFORMATIKA  
FAKULTAS TEKNIK DAN SAINS  
UNIVERSITAS MUHAMMADIYAH PURWOKERTO  
JANUARI 2023**

**HALAMAN PERSETUJUAN**

Skripsi yang diajukan oleh:

Nama : Fajar Sefuloh  
NIM : 1803040079  
Program Studi : Teknik Informatika  
Fakultas : Teknik dan Sains  
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto  
Judul : Analisis Sentimen Masyarakat Terhadap Program Magang Bersertifikat Pada Media Sosial Twitter Menggunakan Klasifikasi *Naïve Bayes Classifier* (NBC)

telah disetujui untuk diajukan dalam ujian skripsi

Purwokerto, 4 Januari 2023

**PEMBIMBING**

Maulida Ayu Fitriani, S.Kom, M.Cs.

NIK. 2160823



**HALAMAN PENGESAHAN**

Skripsi yang diajukan oleh:

Nama : Fajar Sefuloh  
NIM. : 1803040079  
Program Studi : Teknik Informatika  
Fakultas : Teknik dan Sains  
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto  
Judul : Analisis Sentimen Masyarakat Terhadap Program Magang Bersertifikat Pada Media Sosial Twitter Menggunakan Klasifikasi *Naïve Bayes Classifier* (NBC)  
telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer (S.Kom.) pada Program Studi Teknik Informatika, Fakultas Teknik dan Sains, Universitas Muhammadiyah Purwokerto.

**DEWAN PENGUJI**

Penguji 1 (Pembimbing) : Maulida Ayu Fitriani, S.Kom, M.Cs.  
Penguji 2 : Elindra Ambar Pambudi, S.Kom., M.Kom.  
Penguji 3 : Harjono, S.T., M.Eng.

Ditetapkan di : Purwokerto  
Tanggal :

Mengetahui  
Dekan Fakultas Teknik dan Sains

Ir. Teguh Marhendi, S.N., M.T., ASEAN.ENG., ACPE., IPM  
NIK 2160172



## HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertandatangan di bawah ini:

Nama : Fajar Sefuloh  
NIM. : 1803040079  
Program Studi : Teknik Informatika  
Fakultas : Teknik dan Sains  
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

menyatakan dengan sebenar-benarnya bahwa skripsi ini adalah hasil karya saya dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta bukan hasil penjiplakan dari karya orang lain.

Demikian pernyataan ini saya buat dan apabila kelak di kemudian hari terbukti ada unsur penjiplakan, saya bersedia mempertanggungjawabkan sesuai dengan ketentuan yang berlaku.

Purwokerto, 4 Januari 2023

Yang membuat pernyataan



Fajar Sefuloh

## HALAMAN PERSEMBAHAN

Dengan segala kerendahan hati dan rasa hormat melalui tulisan ini, maka penyusun ingin mengucapkan persembahan tugas akhir ini kepada:

1. Allah Subhanahu wata'ala yang senantiasa melimpahkan rahmat dan karunia-Nya.
2. Almarhum Bapak Kastam dan Ibu Wasirah selaku kedua Orang Tua saya yang selama ini telah membantu dalam bentuk doa, kasih sayang, semangat, serta perhatian yang tidak henti-hentinya mengalir demi kelancaran dan kesuksesan dalam menyelesaikan tugas akhir ini. Semoga menjadi kebahagiaan bagi mereka.
3. Kepada seluruh keluarga dari Bapak Kastam yang senantiasa mendukung dan memberikan semangat untuk menyelesaikan perkuliahan.
4. Bapak Feri Wibowo, S.Kom., M.Cs. selaku Ketua Program Studi Teknik Universitas Muhammadiyah Purwokerto
5. Bapak Mukhlis Prasetyo Aji, S.T., M.Kom selaku dosen pembimbing akademik yang selalu memberi bimbingan kepada saya dan teman-teman.
6. Kepada Ibu Maulida Ayu Fitriani, S.Kom, M.Cs. selaku dosen pembimbing sekaligus dosen akademik yang selalu memberikan bimbingan, arahan, dorongan sehingga tugas akhir ini dapat terselesaikan.
7. Kepada seluruh dosen program studi Teknik Informatika Universitas Muhammadiyah Purwokerto yang telah memberikan banyak ilmunya.
8. Kepada seluruh Staf Tata Usaha dan Staf pegawai yang telah memberikan kemudahan dalam kelancaran administrasi kepada saya dalam penyusunan tugas akhir ini.
9. Kepada seluruh sahabat perjuangan angkatan 2018 Teknik Informatika.
10. Kepada semua sivitas akademika Universitas Muhammadiyah Purwokerto yang telah membantu penyusunan tugas akhir ini baik secara langsung maupun tidak langsung.
11. Terimakasih juga kepada teman – teman saya yang telah membantu penyusun menyelesaikan laporan ini.

## MOTTO

“Karena, sesungguhnya sesudah kesulitan itu ada kemudahan. Sesungguhnya, sesudah kesulitan itu ada kemudahan.”

(QS al-Insyirah: 5-6)

“Usaha dan keberanian tidak cukup tanpa tujuan dan arah perencanaan.”

- John F. Kennedy



## KATA PENGANTAR

Puji syukur penyusun ucapkan atas kehadiran Allah Subhanahu wa Ta'ala yang telah melimpahkan rahmat, hidayah dan anugerah-Nya sehingga penyusun dapat menyelesaikan penulisan Laporan Tugas Akhir dengan judul **“ANALISIS SENTIMEN MASYARAKAT TERHADAP PROGRAM MAGANG BERSERTIFIKAT PADA MEDIA SOSIAL TWITTER MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES CLASSIFIER* (NBC)”** dengan baik. Tugas Akhir ini merupakan salah satu persyaratan kurikulum untuk menyelesaikan pendidikan sarjana pada Jurusan Teknik Informatika Fakultas Teknik dan Sains Universitas Muhammadiyah Purwokerto.

Namun, menyadari laporan Tugas Akhir ini masih ada hal-hal yang belum sempurna dan luput dari perhatian. Baik itu dari bahasa yang digunakan maupun dari teknik penyajiannya. Oleh karena itu, dengan segala kekurangan dan kerendahan hati, sangat mengharapkan kritik dan saran dari para pembaca sekalian demi perbaikan laporan ini kedepannya.

Akhirnya, besar harapan agar kehadiran laporan tugas akhir ini dapat memberikan manfaat yang berarti untuk para pembaca. Dan yang terpenting adalah semoga dapat turut serta memajukan ilmu pengetahuan.

Purwokerto, 4 Januari 2023

Yang membuat pernyataan

Fajar Sefuloh

## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Purwokerto dan demi pengembangan ilmu pengetahuan, saya yang bertanda tangan di bawah ini:

Nama : Fajar Sefuloh  
NIM. : 1803040079  
Program Studi : Teknik Informatika  
Fakultas : Teknik dan Sains  
Perguruan Tinggi : Universitas Muhammadiyah Purwokerto

menyetujui untuk memberikan Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*) kepada Universitas Muhammadiyah Purwokerto atas karya ilmiah saya yang berjudul:

Analisis Sentimen Masyarakat Terhadap Program Magang Bersertifikat Pada Media Sosial Twitter Menggunakan Klasifikasi *Naïve Bayes Classifier* (NBC) beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Muhammadiyah Purwokerto berhak menyimpan, mengalihmedia/ mengalihformatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan skripsi saya dengan tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Purwokerto

Pada tanggal : 4 Januari 2023

Yang menyatakan,

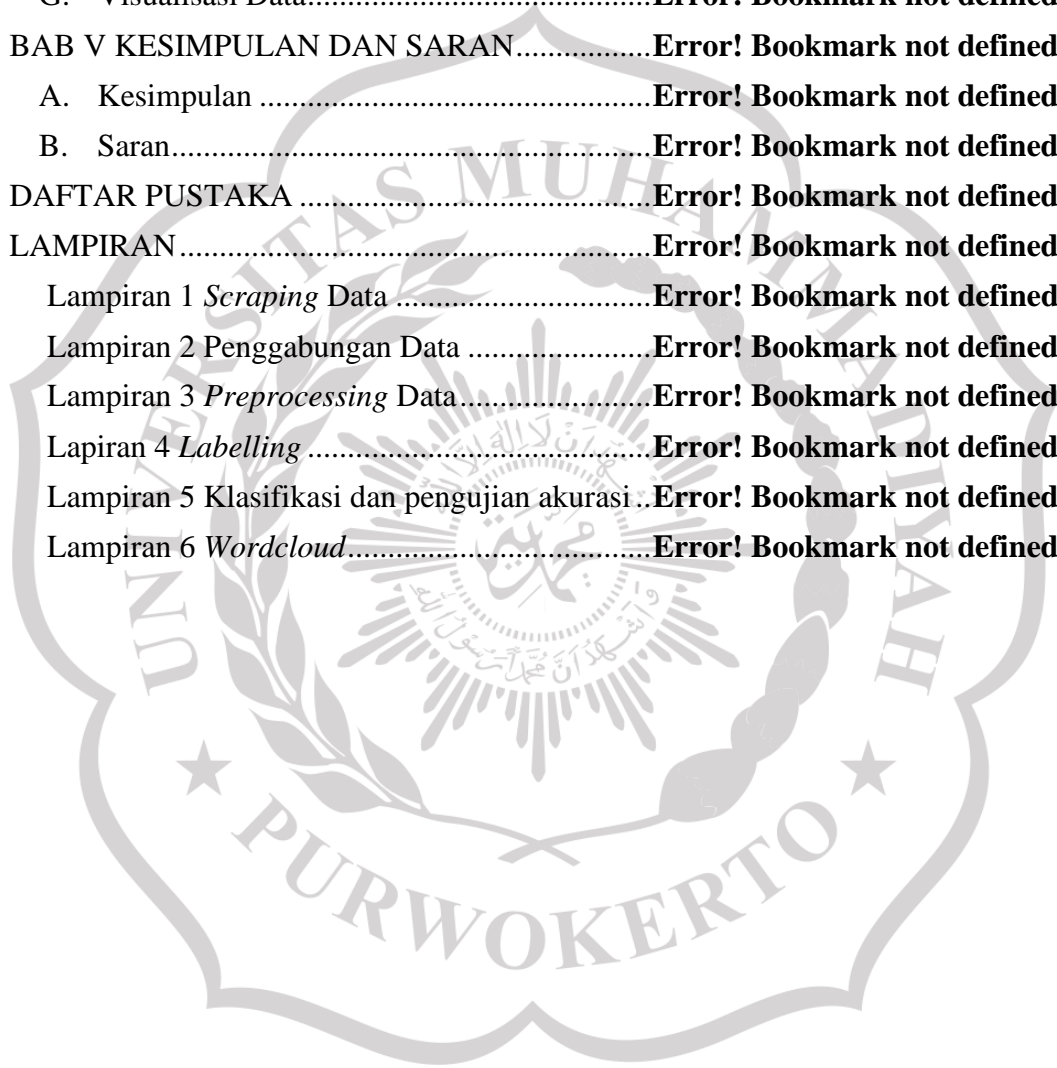
The image shows an official stamp of Universitas Muhammadiyah Purwokerto. The stamp is yellow and features the university's logo (a bird) and the text 'SEKOLAH BILU BILU PURWOKERTO' on the left, 'UNIVERSITAS MUHAMMADIYAH PURWOKERTO' in the center, and 'MENERANGI TERAPEL' at the bottom. Below the stamp is a handwritten signature in black ink.

Fajar Sefuloh

## DAFTAR ISI

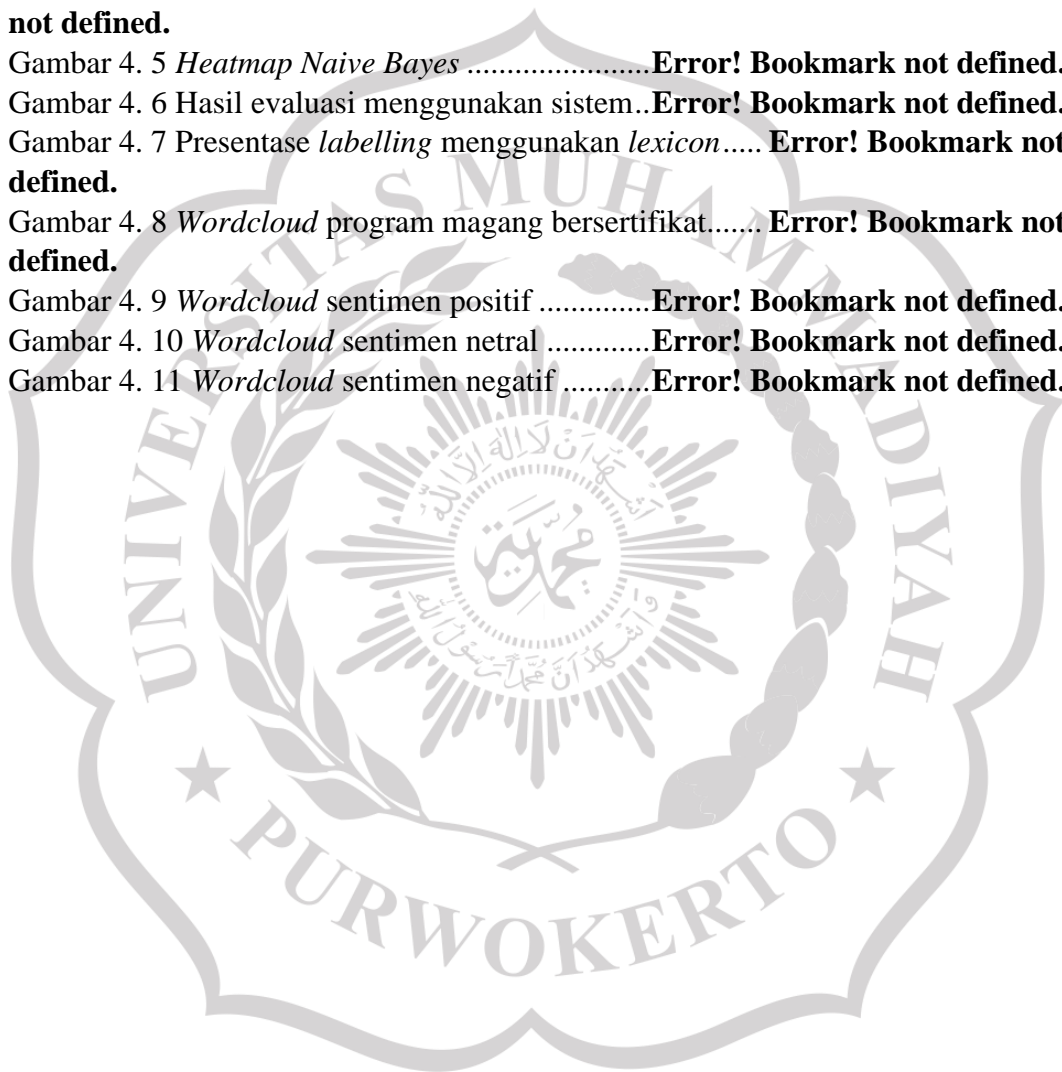
HALAMAN SAMPUL .....	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	1
HALAMAN PENGESAHAN.....	2
HALAMAN PERNYATAAN ORISINALITAS.....	3
HALAMAN PERSEMBAHAN .....	4
MOTTO .....	5
KATA PENGANTAR .....	6
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	7
DAFTAR ISI.....	8
DAFTAR GAMBAR .....	10
DAFTAR TABEL.....	11
DAFTAR LAMPIRAN.....	12
ABSTRAK.....	13
ABSTRACT.....	14
BAB I PENDAHULUAN.....	<b>Error! Bookmark not defined.</b>
A. Latar Belakang Masalah.....	<b>Error! Bookmark not defined.</b>
B. Perumusan Masalah .....	<b>Error! Bookmark not defined.</b>
C. Batasan Masalah.....	<b>Error! Bookmark not defined.</b>
D. Tujuan Penelitian .....	<b>Error! Bookmark not defined.</b>
E. Manfaat Penelitian .....	<b>Error! Bookmark not defined.</b>
BAB II TINJAUAN PUSTAKA.....	<b>Error! Bookmark not defined.</b>
A. Hasil Penelitian Terdahulu.....	<b>Error! Bookmark not defined.</b>
B. Landasan Teori.....	<b>Error! Bookmark not defined.</b>
BAB III METODE PENELITIAN.....	<b>Error! Bookmark not defined.</b>
A. Jenis Penelitian.....	<b>Error! Bookmark not defined.</b>
B. Metode Penelitian.....	<b>Error! Bookmark not defined.</b>
C. Waktu dan Tempat Penelitian .....	<b>Error! Bookmark not defined.</b>
D. Alat Dan Bahan .....	<b>Error! Bookmark not defined.</b>
E. Alur Penelitian .....	<b>Error! Bookmark not defined.</b>
BAB IV HASIL DAN PEMBAHASAN .....	<b>Error! Bookmark not defined.</b>

A. Pengumpulan Data .....	<b>Error! Bookmark not defined.</b>
B. <i>Preprocessing</i> Data .....	<b>Error! Bookmark not defined.</b>
C. <i>Labelling</i> Data .....	<b>Error! Bookmark not defined.</b>
D. Ekstraksi Fitur .....	<b>Error! Bookmark not defined.</b>
E. Klasifikasi <i>Naïve Bayes Classifier</i> .....	<b>Error! Bookmark not defined.</b>
F. Pengujian Nilai Akurasi .....	<b>Error! Bookmark not defined.</b>
G. Visualisasi Data.....	<b>Error! Bookmark not defined.</b>
BAB V KESIMPULAN DAN SARAN.....	<b>Error! Bookmark not defined.</b>
A. Kesimpulan .....	<b>Error! Bookmark not defined.</b>
B. Saran.....	<b>Error! Bookmark not defined.</b>
DAFTAR PUSTAKA .....	<b>Error! Bookmark not defined.</b>
LAMPIRAN.....	<b>Error! Bookmark not defined.</b>
Lampiran 1 <i>Scraping</i> Data .....	<b>Error! Bookmark not defined.</b>
Lampiran 2 Penggabungan Data .....	<b>Error! Bookmark not defined.</b>
Lampiran 3 <i>Preprocessing</i> Data.....	<b>Error! Bookmark not defined.</b>
Lampiran 4 <i>Labelling</i> .....	<b>Error! Bookmark not defined.</b>
Lampiran 5 Klasifikasi dan pengujian akurasi ..	<b>Error! Bookmark not defined.</b>
Lampiran 6 <i>Wordcloud</i> .....	<b>Error! Bookmark not defined.</b>



## DAFTAR GAMBAR

- Gambar 3. 1 *Flowchart* penelitian .....**Error! Bookmark not defined.**  
Gambar 4. 1 Penggabungan file .....**Error! Bookmark not defined.**  
Gambar 4. 2 Jumlah keseluruhan *tweets* .....**Error! Bookmark not defined.**  
Gambar 4. 3 Data akhir hasil *preprocessing* .....**Error! Bookmark not defined.**  
Gambar 4. 4 Hasil akurasi menggunakan *classification report* ..**Error! Bookmark not defined.**  
Gambar 4. 5 *Heatmap Naive Bayes* .....**Error! Bookmark not defined.**  
Gambar 4. 6 Hasil evaluasi menggunakan sistem..**Error! Bookmark not defined.**  
Gambar 4. 7 Presentase *labelling* menggunakan *lexicon*..... **Error! Bookmark not defined.**  
Gambar 4. 8 *Wordcloud* program magang bersertifikat..... **Error! Bookmark not defined.**  
Gambar 4. 9 *Wordcloud* sentimen positif .....**Error! Bookmark not defined.**  
Gambar 4. 10 *Wordcloud* sentimen netral .....**Error! Bookmark not defined.**  
Gambar 4. 11 *Wordcloud* sentimen negatif .....**Error! Bookmark not defined.**

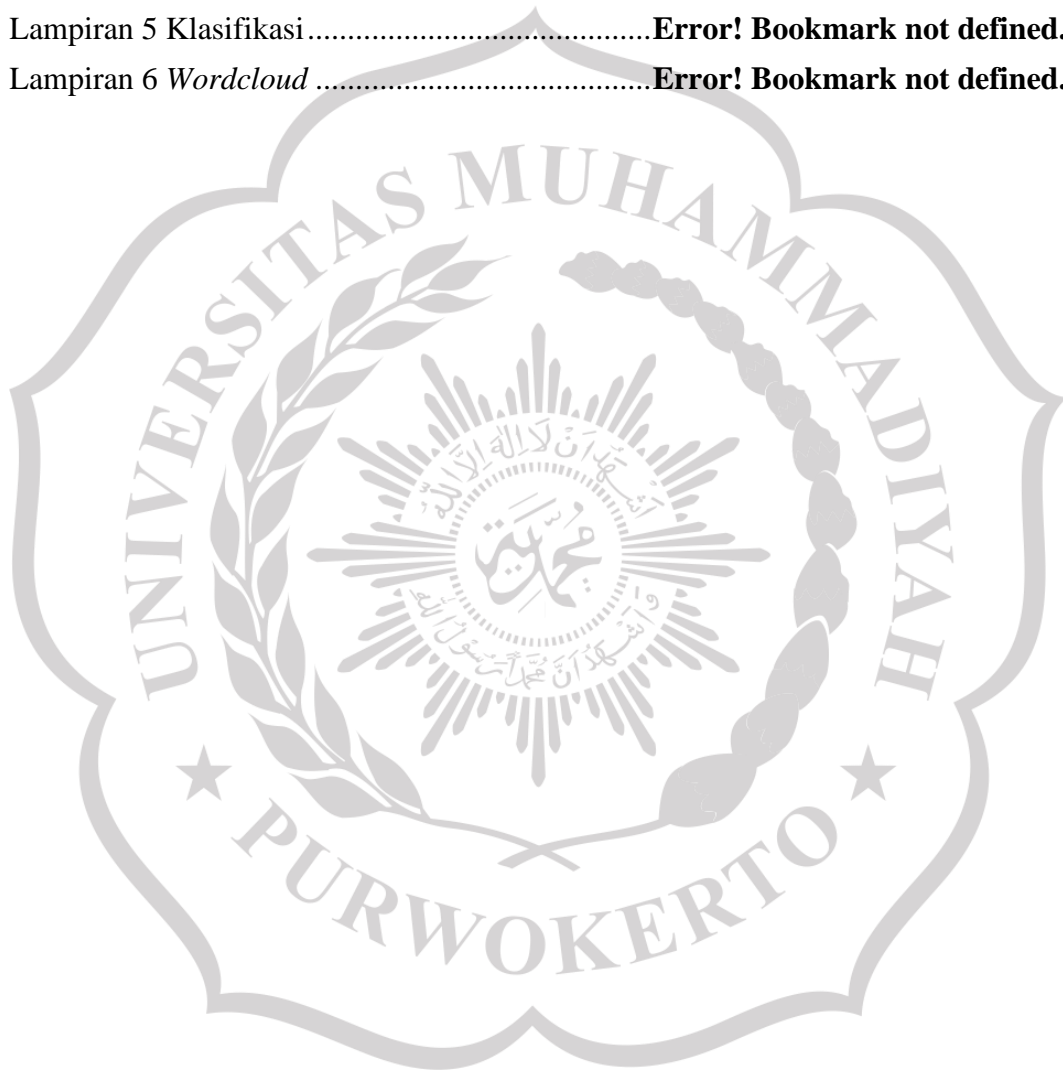


## DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu .....	<b>Error! Bookmark not defined.</b>
Tabel 2. 2 <i>Confusion Matrix</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 3. 1 Proses <i>cleansing</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 3. 2 Proses <i>Case folding</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 3. 3 Proses <i>tokenizing</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 3. 4 Proses <i>filtering slangword</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 3. 5 Proses <i>stopword removal</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 3. 6 Proses <i>to sentence</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 3. 7 Proses <i>stemming</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 1 Hasil penggabungan data <i>tweets</i> keseluruhan .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 2 Proses <i>cleansing</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 3 Proses <i>Case folding</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 4 Proses <i>tokenizing</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 5 Proses <i>filtering slangword</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 6 Proses <i>stopword removal</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 7 Proses <i>to sentence</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 8 Proses <i>stemming</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 9 Hasil <i>labelling</i> data .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 10 Contoh hasil <i>Countvectorizer</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 11 Perbandingan data latih dan uji .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 12 Data pendekatan <i>Naive Bayes Multinomial</i> secara manual .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 13 Data <i>Confusion Matrix</i> .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 14 <i>Confusion Matrix</i> kelas positif .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 15 <i>Confusion Matrix</i> kelas netral .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 16 <i>Confusion Matrix</i> kelas negatif .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 17 Contoh data <i>tweet</i> pada <i>wordcloud</i> positif .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 18 Contoh data <i>tweet</i> pada <i>wordcloud</i> netral .....	<b>Error! Bookmark not defined.</b>
Tabel 4. 19 Contoh data <i>tweet</i> pada <i>wordcloud</i> negatif .....	<b>Error! Bookmark not defined.</b>

## DAFTAR LAMPIRAN

Lampiran 1 <i>Scraping Data</i> .....	<b>Error! Bookmark not defined.</b>
Lampiran 2 Penggabungan Data .....	<b>Error! Bookmark not defined.</b>
Lampiran 3 <i>Preprocessing Data</i> .....	<b>Error! Bookmark not defined.</b>
Lampiran 4 <i>Labelling</i> .....	<b>Error! Bookmark not defined.</b>
Lampiran 5 Klasifikasi .....	<b>Error! Bookmark not defined.</b>
Lampiran 6 <i>Wordcloud</i> .....	<b>Error! Bookmark not defined.</b>



## ABSTRAK

Indonesia adalah salah satu dari lima negara dengan pengguna Twitter terbanyak di dunia, sebanyak 77% adalah pengguna aktif. Twitter mempermudah masyarakat menyampaikan opini atau informasi melalui sebuah *tweets*. Topik yang dibahas masyarakat di Twitter bervariasi, salah satunya adalah program Magang Bersertifikat yang merupakan salah satu dari program Magang Studi Independen Bersertifikat (MSIB) dimana program ini dirancang oleh Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi dengan tujuan untuk memberikan kesempatan kepada mahasiswa untuk belajar dan mengembangkan diri di dunia industri profesional selama 1-2 semester. Program ini memunculkan sentimen dimasyarakat sehingga perlu dianalisis sentimennya, sehingga informasi dari sentimen pada *tweets* Twitter dapat berguna dengan diklasifikasikan kedalam sentimen positif, netral dan negatif. Pada penelitian ini, data didapatkan dari hasil *web scarping* menggunakan bahasa pemrograman python. Data dari hasil *scraping* didapatkan sebesar 14,586, selanjutnya data tersebut dilakukan *preprocessing* dengan melalui tahap *cleansing*, *case folding*, *tokenizing*, *filtering slangword*, *stopword removal*, *to sentence*, *stemming* dan yang terakhir adalah *drop duplicate and empty tweet*. Hasil *preprocessing* data menghasilkan data akhir sejumlah 11.919 data yang digunakan untuk proses *labelling*. *Labelling* data dilakukan menggunakan metode *Lexicon Based* dengan mendapatkan jumlah sentimen positif 9.595 data, sentimen netral sebesar 14 data dan sentimen negatif sebesar 2310 data. Kemudian diterapkan algoritma klasifikasi *Naïve Bayes* dengan model *Naïve Bayes Multinomial* dengan rasio data latih dan uji sebesar 8:2. Hasil pengujian yang dilakukan terhadap *tweets* magang bersertifikat memperoleh *Acurracy Score* sebesar 85%, dengan nilai *Precision* tertinggi 87%, *Recall* tertinggi 95% dan nilai *F1-Score* tertinggi 91%.

**Kata Kunci:** magang bersertifikat, msib, analisis sentimen, *naïve bayes classifier*, *lexicon based*

## ABSTRACT

Indonesia is one of the five countries with the most Twitter users in the world, as many as 77% are active users. Twitter makes it easy for people to convey opinions or information through tweets. The topics discussed by the public on Twitter vary, one of which is the Certified Internship program which is one of the Certified Independent Study Internship programs (MSIB) where this program was designed by the Ministry of Education, Culture, Research and Technology with the aim of providing opportunities for students to learn and develop yourself in the world of professional industry for 1-2 semesters. This program raises sentiments in the community so that sentiment needs to be analysed, so that information from sentiments on Twitter tweets can be useful by classifying sentiments into positive, neutral and negative. In this study, data was obtained from web scarping using the Python programming language. The data obtained from scraping was 14,586, then the data was preprocessed by going through the stages of cleansing, case folding, tokenizing, filtering slangwords, removing stopwords, to sentences, stemming and finally dropping duplicate and empty tweets. The results of the data preprocessing produce the final data of 11,919 data used for the labeling process. Data labeling was carried out using the Lexicon Based method by obtaining a total of 9,595 positive sentiments, 14 data of neutral sentiment and 2,310 negative sentiments. Then the Naïve Bayes classification algorithm is applied with the Naïve Bayes Multinomial model with a ratio of training and test data of 8:2. The test results carried out on certified apprentice tweets obtained an Accuracy Score of 85%, with the highest Precision score of 87%, the highest Recall of 95% and the highest F1-Score value of 91%.

**Keywords:** certified internship, msib, sentiment analysis, naïve bayes classifier, lexicon based