

**ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TERHADAP
KINERJA BPOM RI MENGGUNAKAN ALGORITMA NAÏVE
BAYES DAN SUPPORT VECTOR MACHINE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AINA NABILLA SAFFIRA

21.21.1532

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TERHADAP
KINERJA BPOM RI MENGGUNAKAN ALGORITMA NAÏVE
BAYES DAN SUPPORT VECTOR MACHINE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

AINA NABILLA SAFFIRA

21.21.1532

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TERHADAP
KINERJA BPOM RI MENGGUNAKAN ALGORITMA NAÏVE BAYES
DAN SUPPORT VECTOR MACHINE**

yang disusun dan diajukan oleh

Aina Nabilla Saffira

21.21.1532

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 17 Juli 2023

Dosen Pembimbing



Ferian Farzi Abdolloh, M.Kom

NIK. 199302276

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TERHADAP
KINERJA BPOM RI MENGGUNAKAN ALGORITMA NAÏVE BAYES
DAN SUPPORT VECTOR MACHINE**

yang disusun dan diajukan oleh

Aina Nabilla Saffira

21.21.1532

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Juli 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Baita, M.Kom
NIK. 190302290

Arif Akbarul Huda, S.Si, M.Eng
NIK. 190302287

Ferian Fauzi Abdulloh, M.Kom
NIK. 190302276



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Juli 2023

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Aina Nabilla Saffira
NIM : 21.21.1532

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Pengguna Sosial Media Terhadap Kinerja Bpom RI Menggunakan Algoritma Naive Bayes Dan Support Vector Machine

Dosen Pembimbing : Ferian Fauzi Abdulloh, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 17 Juli 2023

Yang Menyatakan,



Aina Nabilla Saffira

HALAMAN PERSEMBAHAN

Bismillahirrahmanirrohim

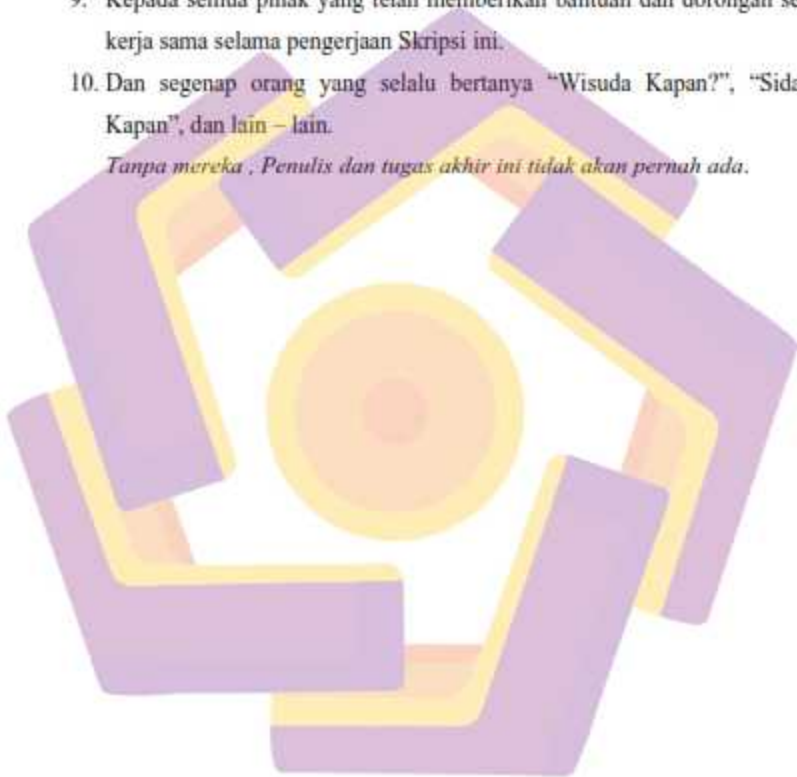
Dengan mengucapkan syukur kepada Tuhan Yang Maha Esa atas berkat, rahmat serta karunia-Nya, Penulis dapat menyelesaikan Skripsi yang berjudul **ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TERHADAP KINERJA BPOM RI MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE**

Penulis selalu mendapatkan bimbingan, dorongan, serta semangat dari banyak pihak. Oleh karena itu Penulis ingin mempersembahkan Skripsi ini kepada

1. Bapak Rasyid Surono (Alm.) dan Ibu Listiyani Noorrahmah Hanief serta saudara – saudara tercinta terutama kakak saya Giesse Khairina yang telah memberikan begitu banyak dorongan dan dukungan yang begitu besar.
2. Keluarga saya Pakde, Om, Bude, Tante, dan Sepupu – sepupu saya yang selalu memberikan dukungan dan doa dalam menyelesaikan perkuliahan ini.
3. Partner saya selama lanjut Transfer S1 ini Aprillia Nur Safitri, Nur Karimah, dan Azzahra Labibah S.F, yang sudah mau bekerja keras dan berjuang bersama-sama untuk menyelesaikan perkuliahan ini.
4. Dosen Pembimbing saya Bapak Ferian Fauzi Abdulloh, M.Kom yang telah meluangkan waktunya dan memberikan arahan serta bimbingan sehingga skripsi ini dapat terselesaikan dengan baik.
5. Seluruh dosen, pengajar, staff, serta karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu pengetahuan dan membantu dalam kegiatan perkuliahan selama ini.
6. Partner saya Gelang Surya Semedi yang telah menemani, menyemangati, dan mendengar keluh kesah saya selama masa perkuliahan ini hingga dapat menyelesaikan skripsi ini.

7. Sahabat - sahabat saya di Timuranz, Soda Gembira, teman rumah yang sudah mau mendengar keluhan saya dan sahabat lainnya yang tidak bisa disebutkan satu per satu, selalu menyemangati dan mendoakan saya.
8. Perusahaan – Perusahaan tempat saya bekerja yang sudah memberikan dukungan dan bantuan selama perkuliahan ini sampai skripsi saya selesai.
9. Kepada semua pihak yang telah memberikan bantuan dan dorongan serta kerja sama selama pengerjaan Skripsi ini.
10. Dan segenap orang yang selalu bertanya “Wisuda Kapan?”, “Sidang Kapan”, dan lain – lain.

Tanpa mereka , Penulis dan tugas akhir ini tidak akan pernah ada.



KATA PENGANTAR

Dengan mengucapkan syukur kepada Tuhan Yang Maha Esa atas berkat, rahmat serta karunia-Nya, Penulis dapat menyelesaikan Skripsi yang berjudul **ANALISIS SENTIMEN PENGGUNA SOSIAL MEDIA TERHADAP KINERJA BPOM RI MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE**.

Skripsi ini ditujukan untuk memenuhi salah satu persyaratan ujian guna memperoleh gelar Sarjana Komputer (S.Kom) pada Jurusan Informatika Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.

Penulis menyadari dalam penulisan Skripsi ini masih jauh dari sempurna, dan banyak kekurangan baik dalam metode penulisan maupun dalam pembahasan materi. Hal tersebut dikarenakan keterbatasan kemampuan Penulis. Sehingga Penulis mengharapkan saran dan kritik yang bersifat membangun sehingga dapat memperbaiki segala kekurangannya di kemudian hari.

Dalam penulisan Skripsi ini, Penulis selalu mendapatkan bimbingan, dorongan, serta semangat dari banyak pihak. Oleh karena itu Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada pembimbing yang terhormat, yakni Yth. Bapak Ferian Fauzi Abdulloh, M.Kom selaku Dosen Pembimbing, yang telah meluangkan waktunya, tenaga dan pikirannya untuk membimbing Penulis dalam penulisan Skripsi ini. Selain pembimbing Penulis juga ingin mengucapkan banyak rasa terima kasih kepada Tim Dosen Penguji serta kepada :

1. Yth. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas Amikom Yogyakarta;
2. Yth. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta;
3. Yth. Ibu Windha Mega Pradnya Duhita, M.Kom selaku Ketua Program Studi SI Informatika Universitas Amikom Yogyakarta.
4. Seluruh Staff Dosen dan Karyawan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

5. Kedua orangtua yang telah memberikan dorongan dan doa sehingga Penulis dapat menyelesaikan tugas akhir ini.
6. Semua teman-teman penulis.

Akhirnya, Penulis mengucapkan rasa terima kasih kepada semua pihak dan apabila ada yang tidak disebutkan Penulis mohon maaf, dengan besar harapan semoga skripsi yang ditulis oleh Penulis ini dapat bermanfaat khususnya bagi Penulis sendiri dan umumnya bagi pembaca. Bagi para pihak yang telah membantu dalam penulisan skripsi ini semoga segala amal dan kebbaikannya mendapatkan balasan yang berlimpah dari Tuhan YME. Amiin.

Yogyakarta, 17 Juli 2023

Penulis

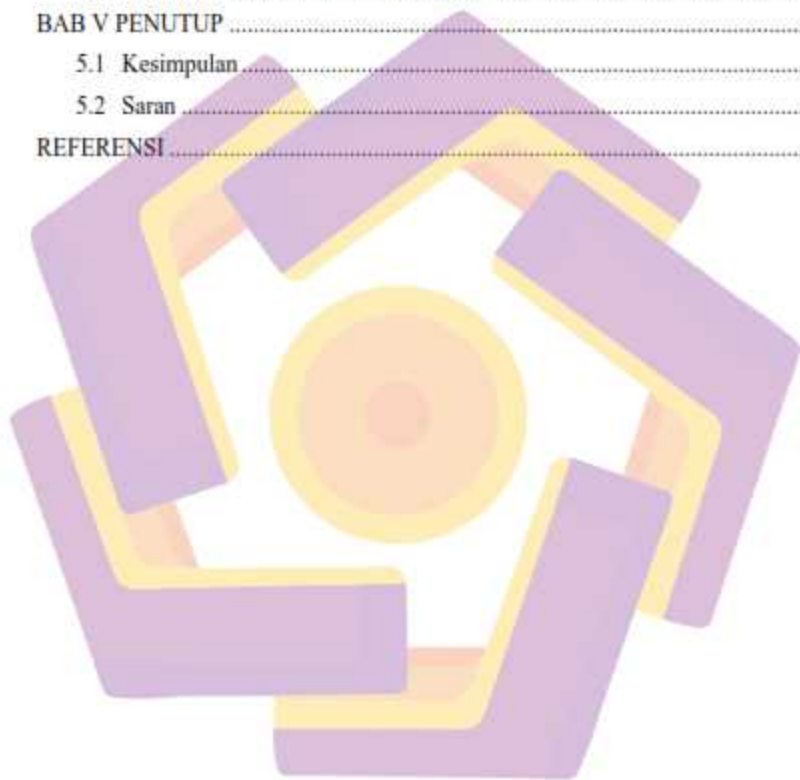


DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	13
2.2.1 Analisis Sentimen	13
2.2.2 Text Mining	13
2.2.3 <i>Preprocessing</i>	14
2.2.4 Klasifikasi	16
2.2.5 Metode <i>Naïve Bayes</i>	17
2.2.6 Metode <i>Support Vector Machine</i>	18
2.2.7 TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	21

2.2.8	<i>Confusion Matrix</i>	22
2.2.9	Youtube.....	24
2.2.10	<i>Python</i>	24
2.2.11	<i>K-Fold Cross Validation</i>	25
2.2.12	<i>Google Collabs</i>	26
BAB III METODE PENELITIAN		27
3.1	Objek Penelitian.....	27
3.2	Alur Penelitian.....	29
3.2.1	Studi Literatur.....	31
3.2.2	Pengumpulan Data.....	31
3.2.3	<i>Pre-Processing</i>	32
3.2.4	<i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	33
3.2.5	Klasifikasi.....	39
3.2.5.1	<i>Naïve Bayes</i>	39
3.2.5.2	<i>Support Vector Machine (SVM)</i>	40
3.2.6	Evaluasi.....	42
3.2.7	Validasi.....	42
3.3	Alat dan Bahan.....	42
3.3.1	Alat Penelitian.....	42
3.3.2	Bahan Penelitian.....	43
BAB IV HASIL DAN PEMBAHASAN		44
4.1	Proses Pengumpulan Data.....	44
4.2	Pelabelan Data (<i>Labeling Manual</i>).....	46
4.3	<i>Preprocessing</i>	47
4.3.1	<i>Remove Duplicate</i>	47
4.3.2	<i>Case Folding & Cleaning</i>	48
4.3.2	<i>Stopwords</i>	49
4.3.3	<i>Tokenizing</i>	51
4.3.4	<i>Stemming</i>	52

4.4	Proses Pembobotan TF-IDF.....	54
4.5	Klasifikasi.....	55
4.5.1	Klasifikasi Support Vector Machine (SVM).....	56
4.5.2	Klasifikasi <i>Naïve Bayes</i>	59
4.6	Validasi menggunakan K-Fold.....	63
BAB V PENUTUP.....		65
5.1	Kesimpulan.....	65
5.2	Saran.....	65
REFERENSI.....		66



DAFTAR TABEL

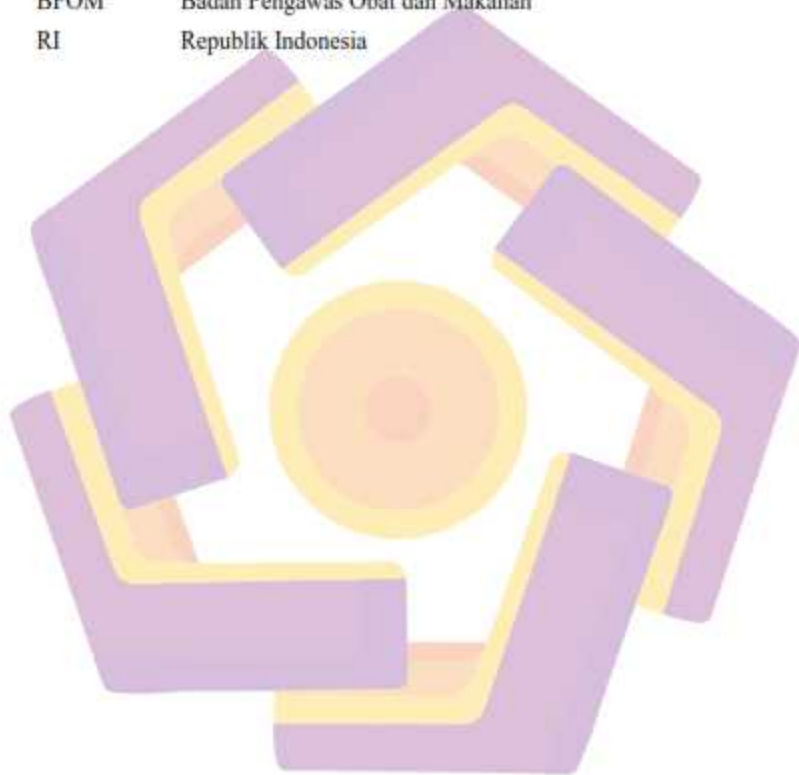
Tabel 2.1 Keaslian Penelitian	9
Tabel 2.2 Confusion Matrix	22
Tabel 3.1 Labeling Manual	31
Tabel 3.2 Perhitungan Manual Setiap Kata (Term)	35
Tabel 3.3 Hasil Nilai DF dan IDF	36
Tabel 3.4 Nilai TF-IDF	38
Tabel 3.5 Spesifikasi perangkat keras yang digunakan	42
Tabel 3.6 Spesifikasi perangkat lunak yang digunakan	43
Tabel 4.1 Pelabelan Data	46
Tabel 4.2 Perbandingan Hasil Remove Duplicate	47
Tabel 4.3 Hasil dari Case Folding & Cleaning	48
Tabel 4.4 Hasil dari Stopwords	50
Tabel 4.5 Hasil dari Tokenizing	51
Tabel 4.6 Hasil dari Stemming	52
Tabel 4.7 Perbandingan Data Latih dan Data Uji	56
Tabel 4.8 Hasil Evaluasi dari SVM	56
Tabel 4.9 Hasil Evaluasi dari <i>Naïve Bayes</i>	59

DAFTAR GAMBAR

Gambar 2.1 Alur Preprocessing	14
Gambar 2.2 Ilustrasi Algoritma SVM dengan Hyperplane.....	19
Gambar 3.1 Alur Kerja Penelitian.....	30
Gambar 3.2 Diagram Alur Preprocessing	33
Gambar 3.3 Diagram Alur klasifikasi naïve bayes	40
Gambar 3.4 Diagram Alur Klasifikasi SVM	41
Gambar 4.1 Implementasi kode program import library pandas	44
Gambar 4.2 Fungsi Crawling Pengumpulan Data	45
Gambar 4.3 Contoh salah satu Kode Program Crawling data	46
Gambar 4.4 Kode Program Remove Duplicate	48
Gambar 4.5 Kode Program Case Folding & Cleaning.....	49
Gambar 4.6 Kode Program Stopwords	51
Gambar 4.7 Kode Program Tokenizing	52
Gambar 4.8 Kode Program Stemming.....	54
Gambar 4.9 Kode Program TF-IDF	54
Gambar 4.10 Kata yang sering muncul dalam dokumen.....	55
Gambar 4.11 Statistik Precision SVM dengan 9x pengujian.....	57
Gambar 4.12 Statistik Recall SVM dengan 9x pengujian	58
Gambar 4.13 Statistik F1-Score SVM dengan 9x pengujian.....	58
Gambar 4.14 Statistik Accuracy SVM dengan 9x pengujian	59
Gambar 4.15 Statistik Precision Naïve Bayes dengan 9x pengujian.....	60
Gambar 4.16 Statistik Recall Naïve Bayes dengan 9x pengujian.....	61
Gambar 4.17 Statistik F1-Score Naïve Bayes dengan 9x pengujian	62
Gambar 4.18 Statistik Accuracy Naïve Bayes dengan 9x pengujian.....	63
Gambar 4.19 Kode Program dan Hasil Validasi Naïve Bayes	63
Gambar 4.20 Kode Program dan Hasil Validasi SVM	64

DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
NB	Naïve Bayes
TF-IDF	Term Frequency-Inverse Document Frequency
BPOM	Badan Pengawas Obat dan Makanan
RI	Republik Indonesia



INTISARI

Media sosial menjadi salah satu wadah masyarakat dalam mengungkapkan pendapat salah satu yang digunakan adalah media sosial youtube yang sering menjadi sumber opini sebagai respon masyarakat di Indonesia. Youtube merupakan media sosial yang berfokus pada media berbagi video yang banyak diminati dari berbagai kalangan. Pengguna dapat memberikan *like*, komen, dan *subscribe*. Komentar yang diberikan pada konten video di youtube dapat berupa opini, saran, dan kritikan selain itu pengguna juga dapat melihat *trending* yang terjadi dari Indonesia dan seluruh dunia. Dari komentar video yang ditulis pengguna memberikan informasi mengenai opini dan kritikan serta saran yang sangat beragam, informasi tersebut akan menjadi dataset yang dapat diolah menjadi analisis sentimen.

Selanjutnya dataset tersebut akan diolah dan diproses dengan menggunakan metode *Naïve Bayes Classifier* dan *Support Vector Machine* dengan tujuan penelitian ini untuk menganalisis sentimen masyarakat mengenai Pengedaran Obat Sirup BPOM RI berupa sentimen positif atau negatif dengan membandingkan kedua metode klasifikasi tersebut sehingga mendapatkan hasil evaluasi berupa *F1-Score*, *Precision*, *Recall*, dan *Accuracy* yang terbaik. Langkah pertama untuk melakukan klasifikasi dengan *crawling* data menggunakan Youtube API v3 dan Bahasa pemrograman python, setelah data berhasil dikumpulkan akan diberikan label secara manual oleh peneliti setelah itu akan dilakukan tahapan *preprocessing* terhadap data dan diberikan pembobotan pada kata menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Untuk Langkah selanjutnya dengan menggunakan klasifikasi *Multinomial Naïve Bayes* dan *Support Vector Machine (SVM)*. Dan langkah terakhir dilakukan validasi menggunakan *K-Fold Cross Validation* dengan nilai $K = 10$. Dari keseluruhan tahapan tersebut didapatkan bahwa metode *Naïve Bayes Classifier* lebih akurat daripada *Support Vector Machine*.

Kata Kunci: Analisis Sentimen, Komentar Youtube, *Naïve Bayes Classifier*, *Support Vector Machine*, Obat Sirup BPOM RI.

ABSTRACT

Social media has become a forum for the community to express opinions, one of which is social media YouTube, which is often a source of opinion as a response to society in Indonesia. Youtube is a social media that focuses on video-sharing media that is in great demand from various groups. Users can give likes, comments, and subscribe. Comments given on video content on YouTube can be in the form of opinions, suggestions, and criticism. In addition, users can also see trends that are happening from Indonesia and around the world. From video comments written by users providing information regarding opinions and criticisms as well as very diverse suggestions, this information will become a dataset that can be processed into sentiment analysis.

Furthermore, the dataset will be processed and processed using the Naïve Bayes Classifier and Support Vector Machine methods with the aim of this study to analyze public sentiment regarding the Distribution of BPOM RI Syrup Drug in the form of positive or negative sentiment by comparing the two classification methods so as to obtain an evaluation result in the form of F1-Score The best , Precision, Recall and Accuracy. The first step is to classify by crawling data using the Youtube API v3 and Python programming language, after the data has been successfully collected the researcher will label it manually after that the preprocessing stage will be carried out on the data and the words will be weighted using the Term Frequency-Inverse Document Frequency method (TF-IDF). For the next step, use the Multinomial Naïve Bayes classification and Support Vector Machine (SVM). And the last step is validation using K-Fold Cross Validation with a value of K = 10. From all these stages it is found that the Support Vector Machine method is more accurate than the Naïve Bayes Classifier.

Keywords: *Sentiment Analysis, Youtube Comment, Naïve Bayes Classifier, Support Vector Machine, BPOM RI Syrup.*