KLASIFIKASI SENTIMEN PENGGUNA *TWITTER* TERHADAP KEBIJAKAN PEMERINTAH TERKAIT *COVID-19* MENGGUNAKAN METODE *NAIVE BAYES*

SKRIPSI

Oleh :
KHOTIBUL UMAM
NIM. 17650103

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

KLASIFIKASI SENTIMEN PENGGUNA *TWITTER* TERHADAP KEBIJAKAN PEMERINTAH TERKAIT *COVID-19* MENGGUNAKAN METODE *NAIVE BAYES*

SKRIPSI

Oleh:
KHOTIBUL UMAM
NIM. 17650103

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2022

HALAMAN PERSETUJUAN

KLASIFIKASI SENTIMEN PENGGUNA *TWITTER* TERHADAP KEBIJAKAN PEMERINTAH TERKAIT *COVID-19* MENGGUNAKAN METODE *NAIVE BAYES*

SKRIPSI

Oleh :
KHOTIBUL UMAM
NIM. 17650103

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal 9 Desember 2022

Dosen Pembimbing I	Dosen Pembimbing II
Dr. Cahyo Crysdian	Fajar Rohman Hariri, M.Kom
NIP. 19740424 2009011 008	NIP. 19890515 2018011 001

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan M.MT.,IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

KLASIFIKASI SENTIMEN PENGGUNA *TWITTER* TERHADAP KEBIJAKAN PEMERINTAH TERKAIT *COVID-19* MENGGUNAKAN METODE *NAIVE BAYES*

SKRIPSI

Oleh :
KHOTIBUL UMAM
NIM. 17650103

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal : 16 Desember 2022

Susunan Dewan Penguji

Ketua Penguji	: Dr. M. Amin Hariyadi NIP. 19670018 200501 1 001	( )
Anggota Penguji I	: Okta Qomaruddin Aziz, M.Kom NIP. 19911019 201903 1 013	( )
Anggota Penguji II	: Dr. Cahyo Crysdian NIP. 19740424 2009011 008	( )
Anggota Penguji III	: Fajar Rohman Hariri, M.Kom NIP. 19890515 2018011 001	( )

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan M.MT.,IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Khotibul Umam

NIM : 17650103

Program Studi : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencamtumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 16 Desember 2022

Yang membuat pernyataan

Khotibul Umam

NIM. 17650103

MOTTO

"Sebuah aksi paling kecil tentang kebajikan jauh lebih berharga dari pada keinginan terhebat namun belum diwujudkan."

-Kahlil Gibran

HALAMAN PERSEMBAHAN

Puji Syukur kehadirat Allah SWT, shalawat serta salam kepada Rasul-Nya
Penulis persembahkan sebuah karya tulis ini kepada:

Kepada orang tua penulis, Bapak Mohammad Makin dan Ibu Juwanah yang senantiasa mendukung dan memberi motivasi akan terwujudnya tulisan ini. Kepada seluruh keluarga besar penulis dari pihak Ayah maupun pihak Ibu yang senantiasa bangga dan menyemangati penulis dalam melaksanakan kuliah.

Kepada bapak Dr. Cahyo Crysdian selaku dosen pembimbing I dan bapak Fajar Rohman Hariri, M.Kom selaku dosen pembimbing II yang telah dengan sabar membimbing penulis agar dapat menyelesaikan skripsi. Serta seluruh dosen di prodi Teknik Informatika Universitas Islam Maulana Malik Ibrahim Malang yang telah mengajarkan ilmu-ilmu baru bagi penulis.

Teman-teman Teknik Informatika angkatan 2017 Universitas Islam Maulana Malik Ibrahim Malang yang selalu memberikan semangat dan dukungan kepada penulis. Serta seluruh orang yang pernah terlibat dan membantu baik secara langsung maupun secara tidak langsung, penulis mengucapkan banyak terima kasih.

KATA PENGANTAR

Assalamu'alaikum Wr.Wb.

Puji syukur penulis panjatkan kepada Allah SWT yang senantiasa memberikan rahmat dan kesehatan, sehingga penulis mampu menyelesaikan skripsi ini dengan baik. Penulis menyampaikan ucapan terimakasih kepada semua pihak yang pernah terlibat langsung maupun tidak langsung dalam meyelesaikan skripsi ini, bukan hanya karena usaha keras dari penulis sendiri, akan tetapi karena adanya dukungan dari berbagai pihak. Oleh karena itu penulis berterima kasih kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknolog Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan M.MT., IPM selaku Ketua Prodi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Dr. Cahyo Crysdian selaku Dosen Pembimbing I yang telah banyak meluangkan waktunya untuk memberikan bimbingan, arahan serta bantuan dalam terwujudnya karya tulis skripsi ini dari awal hingga akhir.
5. Fajar Rahman Hariri, M.Kom selaku Dosen Pembimbing II yang telah membimbing serta memberikan arahan penulisan skripsi dari awal hingga akhir.
6. Dr. M. Amin Hariyadi selaku peguji I dan Octa Qomaruddin Aziz, M.Kom selaku penguji II yang telah meluangkan waktunya untuk menguji dan dengan sabar memberi arahan dan saran dalam menyelesaikan skripsi ini.
7. Agung Teguh Wibowo Almais, S.Kom, M.T. dan Ainatul Madhiyah, M.Cs. selaku dosen wali yang dengan sabar memberikan saran dan arahan dalam menempuh masa perkuliahan.
8. Segenap civitas akademik Program Studi Teknik Informatika, dan seluruh dosen yang telah memberikan ilmu serta arahan semasa kuliah.

9. Kedua orang tua penulis, Bapak Mohammad Makin dan Ibu Juwanah yang telah memberikan banyak dukungan, doa serta bantuan sehingga penulis mampu menyelesaikan masa studi hingga mencapai gelar sarjana.
10. Jihan Roihanna Farahdhiya yang telah banyak meluangkan waktu untuk menemani dalam setiap proses yang dilalui dalam penyusunan karya tulis serta memberi motivasi dan dukungan untuk segera menyelesaikan masa studi S1.
11. Fajar Dewantara selaku sahabat semasa perkuliahan yang banyak membantu dalam terwujudnya karya tulis skripsi ini melalui kritik dan saran yang disampaikan.
12. Teman-teman Teknik Informatika angkatan 2017 yang telah mendukung penulis dalam menyelesaikan skripsi khususnya M. Yoga Pratama, Mustafa Kamal Rabbani, Arya Abimanyu, M. Rizal Sulton Aulia, Ahmad Thoriq Romi Azhar, M. Nasyitul Ibad, dan Jayanti Galuh C., yang telah bersedia membantu semasa perkuliahan hingga terwujudnya karya tulis skripsi ini.
13. Penulis sendiri yang telah berusaha dalam penyelesaian skripsi dan meringankan beban orang tua dengan menyelesaikan skripsi di semester 11.
14. Serta semua pihak yang secara langsung maupun tidak langsung turut membantu penulis menyelesaikan skripsi yang tidak bisa disebutkan satu per satu.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat kekurangan. Maka dari itu penulis dengan senang hati menerima segala saran dan kritik yang membangun. Disamping itu penulis juga berharap skripsi ini bermanfaat bagi semua orang yang membacanya.

Wassalamualaikum Wr.Wb.

Malang, 16 Desember 2020

Penulis

DAFTAR ISI

COVER
HALAMAN PERSETUJUAN ..... **iii**
HALAMAN PENGESAHAN ..... **iv**
PERNYATAAN KEASLIAN TULISAN ..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
KATA PENGANTAR ..... **viii**
DAFTAR ISI ..... **x**
DAFTAR GAMBAR ..... **xii**
DAFTAR TABEL ..... **xiii**
ABSTRAK ..... **xiv**
ABSTRACT ..... **xv**
الملخص ..... **xvi**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang Penelitian 1
1.2 Pernyataan Masalah 4
1.3 Tujuan Penelitian 5
1.4 Batasan Masalah 5
1.5 Manfaat Penelitian 6
1.6 Sistematika Penulisan 6
BAB II STUDI PUSTAKA **8**
2.1 Sentiment Analysis 8
2.2 Naïve Bayes Classifier (NBC) 10
2.3 Theoretical Framework 13
BAB III DESAIN DAN IMPLEMENTASI SISTEM **16**
3.1 Akuisisi Data 16
3.2 Rancangan Sistem 17
3.2.1 Preprocessing 19
3.2.2 Seleksi Fitur TF-IDF 28
3.2.3 Klasifikasi Naïve Bayes 32
3.2.4 Gaussian Naïve Bayes 35

3.3 Skenario Pengujian 38
BAB IV UJI COBA DAN PEMBAHASAN 39
4.1 Langkah-Langkah Uji Coba 39
4.2 Hasil Uji Coba 43
4.3 Pembahasan 53
BAB V KESIMPULAN DAN SARAN 58
5.1 Kesimpulan 58
5.2 Saran 58
DAFTAR PUSTAKA
LAMPIRAN

DAFTAR GAMBAR

Gambar 2. 1 Theoretical Framework 13
Gambar 3. 1 Rancangan Sistem 18
Gambar 3. 2 Blok Diagram Preprocessing..... 20
Gambar 3. 3 Diagram Alur Proses CaseFolding..... 21
Gambar 3. 4 Implementasi CaseFolding..... 21
Gambar 3. 5 Diagram Alur Proses Cleansing 22
Gambar 3. 6 Implementasi Cleansing 22
Gambar 3. 7 Diagram Alur Proses Stopword Removal 23
Gambar 3. 8 Implementasi Stopword Removal 23
Gambar 3. 9 Diagram Alur Proses Stemming..... 26
Gambar 3. 10 Implementasi Stemming..... 26
Gambar 3. 11 Diagram Alur Proses Tokenizing..... 28
Gambar 3. 12 Implementasi Tokenizing..... 28
Gambar 3. 13 Implementasi TF-IDF..... 32
Gambar 3. 14 Diagram Alur Klasifikasi Gaussian Naïve Bayes 37
Gambar 4. 1 Hasil Pengujian 53

DAFTAR TABEL

Tabel 3. 1 Contoh Pelabelan 17
Tabel 3. 2 Desain Nilai TF-IDF 31
Tabel 4. 1 Contoh Dataset 40
Tabel 4. 2 Hasil Pengujian Menggunakan 10% Fitur 44
Tabel 4. 3 Hasil Pengujian Menggunakan 20% Fitur 45
Tabel 4. 4 Hasil Pengujian Menggunakan 30% Fitur 46
Tabel 4. 5 Hasil Pengujian Menggunakan 40% Fitur 46
Tabel 4. 6 Hasil Pengujian Menggunakan 50% Fitur 47
Tabel 4. 7 Hasil Pengujian Menggunakan 60% Fitur 48
Tabel 4. 8 Hasil Pengujian Menggunakan 70% Fitur 49
Tabel 4. 9 Hasil Pengujian Menggunakan 80% Fitur 50
Tabel 4. 10 Hasil Pengujian Menggunakan 90% Fitur 51
Tabel 4. 11 Hasil Pengujian Confussion Matrix 52

ABSTRAK

Umam, Khotibul. 2022. **Klasifikasi Sentimen Pengguna Twitter Terhadap Kebijakan Pemerintah Terkait Covid-19 Menggunakan Metode Naive Bayes.** Skripsi. Program Studi Teknik Informatika. Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (1) Dr. Cahyo Crysdian (2) Fajar Rohman Hariri, M.Kom

Kata kunci : *Klasifikasi Sentimen, Twitter, Covid-19, Naïve Bayes, TF-IDF*

Beberapa waktu belakangan ini dunia dikejutkan dengan pandemi luar biasa hebat yang telah melanda banyak negara di dunia yaitu Covid-19 (Coronavirus Disease ’19). Hal tersebut memaksa pemerintah di dunia untuk mengeluarkan beberapa kebijakan darurat terkait Covid-19 guna menekan angka persebaran virus mematikan ini. Respon yang disampaikan publik terkait kebijakan tersebut bermacam-macam, ada yang menanggapi dengan positif dan ada yang menanggapi dengan negatif. Di masa sekarang terdapat banyak sekali cara untuk mengamati respon dari masyarakat mengenai hal tersebut. Salah satu cara yang dapat dilakukan adalah dengan mengamati statistik yang terekam di media sosial. Banyaknya sentimen yang mereka sampaikan melalui media sosial dapat dijadikan sebagai sebuah informasi penting untuk diteliti menggunakan sentimen analisis. Salah satu metode yang sering digunakan dalam analisis sentimen ini adalah metode Naïve Bayes. Terdapat beberapa tipe dari Naïve Bayes, pada penelitian ini digunakan tipe Gaussian Naïve Bayes. Hal ini dapat diamati pada penggunaan 2 parameter utama yang merupakan ciri utama dari Gaussian Naïve Bayes, yaitu mean (μ) dan varians (σ) yang digunakan untuk perhitungan data bersifat kontinu pada dua kelas yang berbeda. Oleh sebab itu penelitian ini bertujuan untuk mengetahui respon masyarakat terkait kebijakan yang diterapkan sekaligus mengetahui tingkat akurasi metode Gaussian Naïve Bayes Classifier dalam mengklasifikasikan sentimen pada media sosial. Penelitian tentang proses pengklasifkasian tweets menggunakan metode Gaussian Naïve Bayes ini menghasilkan kinerja sistem terbaik pada penggunaan fitur sebesar 70% dengan nilai accuracy sebesar 70,6%, precision sebesar 74,11%, recall sebesar 66,98%, dan nilai f-measure sebesar 70,37%. Berdasarkan nilai akurasi pengujian tersebut dapat disimpulkan bahwa penggunaan metode Gaussian Naïve Bayes dapat diimplementasikan dalam sistem klasifikasi sentimen.

ABSTRACT

Umam, Khotibul. 2022. **Sentiments Classification of Twitter Users Against Government Policies Related to Covid-19 Using Naive Bayes.** Thesis. Department of Informatics Engineering. Faculty of Science and Technology. Maulana Malik Ibrahim State Islamic University of Malang. Supervisors : (1) Dr. Cahyo Crysdian (2) Fajar Rohman Hariri, M.Kom

Keywords : *Sentiments Classification, Twitter, Covid-19, Naïve Bayes, TF-IDF*

Recently, the world has been shocked by a great pandemic that has hit many countries. It is Covid-19 (Coronavirus Disease '19) pandemic. It has forced governments worldwide to issue several emergency policies related to Covid-19 to reduce the spread of this deadly virus. The public's responses to this policy varied. Some responded positively, and others negatively. Currently, there are many ways to observe the public's response regarding this matter. One way to do this is to observe the statistics on social media. The number of sentiments they convey through social media can be used as vital information to be examined using sentiment analysis. One method that is often used in sentiment analysis is the Naïve Bayes method. There are several types of Naïve Bayes, in this study the Gaussian Naïve Bayes type was used. This can be observed in the use of 2 main parameters which are the main characteristics of Gaussian Naïve Bayes, namely the mean (μ) and variance (σ) which are used for calculating data that are continuous in two different classes. Therefore this research aims to determine the public's response to the policies implemented and the accuracy of the Naïve Bayes Classifier method in classifying sentiments on social media. Research on classifying tweets using the Gaussian Naïve Bayes method produces the best system performance on feature usage of 70% with an accuracy value of 70.6%, a precision of 74.11%, a recall of 66.98%, and an f-measure value of 70.37%. Based on the test accuracy value, it can be concluded that using the Naïve Bayes method can be implemented in a sentiment classification system.

الملخص

الأمم، خاطب. ٢٠٢٢. تصنيف مشاعر مستخدمي تويتر تجاه السياسات الحكومية المتعلقة بالكوفيد-١٩ باستخدام طريقة بايز الساذج. البحث الجامعي. قسم الهندسة المعلوماتية. كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. جاهيو كريسديان. المشرف الثاني: فجر رحمن حريري، الماجستير.

الكلمات الرئيسية : تصنيف المشاعر، تويتر، الكوفيد-١٩، بايز الساذج، TF-IDF

في الآونة الأخيرة ، صدم العالم بالوباء الاستثنائي الذي ضرب العديد من دول العالم، وبالتحديد الكوفيد-١٩ (مرض فيروس كورونا ١٩). وقد أجبر هذا الحكومات في العالم على إصدار سياسات الطوارئ المتعلقة بها للحد من انتشار هذا الفيروس القاتل. الردود المقدمة من الجمهور فيما يتعلق بالسياسة مختلفة، بعضها يستجيب بشكل إيجابي والبعض الآخر يستجيب سلبا. اليوم هناك عدة الطرق لمراقبة استجابة الجمهور حول هذا الموضوع. نستطيع مراقبة الإحصائيات المسجلة على وسائل التواصل الاجتماعي. ويمكن أن نستخدم مقدار المشاعر التي ينقلونها عبر وسائل التواصل الاجتماعي كجزء مهم من المعلومات التي يجب البحث عنها باستخدام تحليل المشاعر. إحدى الطرق المستخدمة غالبا في تحليل المشاعر هذا هي طريقة بايز الساذج. لذلك يهدف هذا البحث إلى معرفة استجابة الجمهور للسياسات المطبقة وكذلك معرفة مستوى دقة مصنف بايز الساذج في تصنيف المشاعر على وسائل التواصل الاجتماعي. نتج عن هذا البحث حول عملية تصنيف التغريدات باستخدام طريقة بايز الساذج *Gaussian Naïve Bayes* أفضل أداء للنظام في استخدام الميزات بنسبة ٧٠% بقيمة دقة ٧٠.٦%، وضبط ٧٤.١١%، واسترجاع ٦٦.٩٨%، وقيمة قياس ف ٧٠.٣٧%. بناء على قيمة دقة هذه الاختبارات، يمكن الاستنتاج أن استخدام طريقة بايز الساذج يمكن تنفيذه في نظام تصنيف المشاعر.

BAB I

PENDAHULUAN

1.1 Latar Belakang Penelitian

Beberapa waktu belakangan ini dunia dikejutkan dengan pandemi luar biasa hebat yang telah melanda banyak negara di dunia. Wabah penyakit yang diduga berawal pada tanggal 17 November 2019 di Kota Wuhan, Ibu Kota Provinsi Hubei, China ini disebut *COVID-19 (Coronavirus Disease '19)*. Terdapat sebuah jenis kelompok virus yang diketahui dapat menyebabkan penyakit pada manusia atau hewan yang disebut *Coronavirus.* Sebagian dari banyaknya jenis *Coronavirus* diketahui dapat menjadi penyebab peradangan pada saluran pernafasan manusia mulai dari demam, batuk dan pilek hingga yang lebih serius seperti *Middle East Respiratory Syndrome (MERS)* dan *Severe Acute Respiratory Syndrome (SARS).* Akhir-akhir ini telah ditemukan jenis varian baru dari *Coronavirus* yang diketahui dapat menjadi sebab dari penyakit *COVID-19* (WHO, 2019).

Untuk di Indonesia sendiri, wabah *COVID-19* tercatat pertama kali masuk pada 2 Maret 2020 di Depok, Jawa Barat. Hal ini diketahui setelah 2 orang WNI berusia 64 tahun dan 31 tahun terdeteksi positif virus ini. Dalam catatan kementerian kesehatan, dalam kurun waktu 14 hari setelah kasus pertama tercatat sudah 369 orang terdeteksi positif *COVID-19* dari 1898 orang yang diperiksa dengan kasus meninggal akibat virus ini mencapai 32 orang.

Berdasarkan statistik terkait cepatnya persebaran virus ini membuat pemerintah Indonesia tidak dapat tinggal diam, terlebih *WHO* telah mengeluarkan pernyataan bahwa virus ini telah ditetapkan sebagai sebuah pandemi sehingga kondisi ini memaksa Presiden Joko Widodo untuk segera mengesahkan status darurat nasional *COVID-19.* Semakin meningkatnya kasus kematian karena *COVID-19* ini benar-benar telah menyita perhatian pemerintah negara di seluruh dunia, tidak terkecuali pemerintah Indonesia. Kondisi ini tentu saja memaksa pemerintah untuk segera mengeluarkan beberapa kebijakan darurat terkait *COVID-19* guna menekan angka persebaran virus mematikan ini. Akan tetapi reaksi yang disampaikan publik terkait kebijakan ini ternyata bermacam-macam, ada yang menanggapi dengan positif dan ada pula yang menanggapi dengan negatif. Di masa sekarang terdapat banyak sekali cara untuk mengamati kritik, saran, serta keluhan dari masyarakat mengenai suatu hal atau kejadian. Salah satu cara yang dapat dilakukan adalah mengamati statistik yang terekam di media sosial. Termasuk mengenai perkembangan penerapan kebijakan darurat terkait pandemi di Indonesia. Berdasarkan data statistik yang dibuat oleh Kominfo, Indonesia menempati peringkat kelima terbesar jumlah pengguna *Twitter* di dunia. Berdasarkan data tersebut, dapat diketahui minat orang Indonesia pada media sosial *Twitter* sangatlah tinggi, sehingga dapat dipastikan mereka sering mengomentari hal-hal disekitar mereka ataupun kejadian yang sedang terjadi dan banyak dibicarakan orang, seperti penerapan kebijakan pemerintah khususnya yang terkait dengan pandemi saat ini. Banyaknya sentimen yang mereka sampaikan melalui *Twitter* dapat dijadikan sebagai sebuah informasi penting untuk diteliti menggunakan sentimen analisis.

Sentimen analisis merupakan studi komputasi yang membahas pendapat, perilaku dan emosi seseorang terhadap entitas. Entitas itu sendiri dapat menggambarkan individu, kejadian atau topik. Adanya sentimen analisis dapat mengatasi masalah pengelompokan opini menjadi positif atau negatif dengan cara yang lebih mudah. Sebelum dilakukan proses *sentiment analysis*, tentu saja memerlukan sebuah proses yang berfungsi untuk mengolah data teks agar siap untuk dianalisis yang biasa disebut dengan proses *text mining*. *Text mining* adalah salah satu tipe dari *data mining* yang bertujuan menemukan pola menarik dalam kumpulan data teks besar, metode ini seringkali digunakan untuk penyusunan dokumen baru dari sekumpulan dokumen (Kurniawan et al., 2012).

Dalam teknik pengklasifikasian teks tersebut terdapat banyak metode yang dapat digunakan, salah satunya adalah *Naïve Bayes Classifier.* Dalam hal akurasi dan perhitungan komputasi, metode *Naive Bayes Classifier* dapat dipertimbangkan sebagai sebuah metode klasifikasi data yang berpotensi lebih baik daripada metode klasifikasi lainnya (Ting et al., 2014).

Metode klasifikasi *Naive Bayes* adalah sebuah metode pengklasifikasi yang bersumber pada hasil karya *Thomas Bayes* mengenai metode probabilitas dan statistik. Metode ini memprediksi kemungkinan di masa depan berdasarkan pengalaman masa lalu. Ciri utama metode ini adalah asumsi yang sangat kuat (naif) terhadap ketidakterikatan setiap kondisi atau peristiwa.

Masalah terkait analisis sentimen ini seharusnya sudah menjadi sebuah perhatian khusus bagi seorang cendikiawan muslim, dikarenakan metode analisis sentimen ini dapat mengurangi resiko adanya penyebaran berita palsu yang

dianggap merugikan, terlebih pada kasus ini berhubungan dengan kesehatan dan kepentingan banyak orang. Penyebaran berita serta informasi palsu ini memiliki keterkaitan dengan yang telah difirmankan Allah di dalam Alquran pada Surat Al-Hujurat ayat 6 :

يَٰٓأَيُّهَا ٱلَّذِينَ ءَامَنُوٓاْ إِن جَآءَكُمْ فَاسِقُۢ بِنَبَإٖ فَتَبَيَّنُوٓاْ أَن تُصِيبُواْ قَوْمَۢا بِجَهَٰلَةٖ فَتُصْبِحُواْ عَلَىٰ مَا فَعَلْتُمْ نَٰدِمِينَ

"*Hai orang-orang yang beriman, jika datang kepadamu orang fasik membawa suatu berita, maka periksalah dengan teliti agar kamu tidak menimpakan suatu musibah kepada suatu kaum tanpa mengetahui keadaannya yang menyebabkan kamu menyesal atas perbuatanmu itu*" (QS. Al-Hujurat : 6).

Pada *Tafsir Al-Muyassar* karya Al-Qarni (2008) ayat tersebut ditafsirkan sebagai berikut, kepada orang-orang yang menngakui kebenaran Allah dan Rasul-Nya serta memenuhi dan menjalankan syariat-Nya, ketika pelaku kejahatan datang kepada kalian dengan suatu kabar berita, maka sebelum kalian menyetujui kebenaran berita tersebut dan mengutipnya maka periksalah berita tersebut agar kalian mengetahui kebenaran sesungguhnya dan kalian tidak melakukan perbuatan zalim seperti melawan orang yang tidak bersalah sehingga kalian menyesali perbuatan tersebut.

Oleh karena begitu pentingnya mengklasifikasi sentimen pada masyarakat mengenai suatu kejadian ataupun objek tertentu supaya mendapati suatu hasil yang akurat maka peneliti memutuskan untuk melakukan penelitian yang berjudul "Klasifikasi Sentimen Pengguna *Twitter* Terhadap Kebijakan Pemerintah Terkait *COVID-19* Menggunakan Metode *Naïve Bayes.*"

1.2 Pernyataan Masalah

1. Seberapa tinggi nilai *accuracy, precision, recall, dan f-measure* metode *Gaussian Naïve Bayes Classifier* dalam pengklasifikasian teks pada

klasifikasi sentimen pengguna *Twitter* terhadap kebijakan pemerintah terkait *COVID-19*.

2. Tahapan apa yang mempengaruhi nilai *accuracy, precision, recall, dan f-measure* metode *Gaussian Naïve Bayes Classifier* dalam pengklasifikasian teks pada klasifikasi sentimen pengguna *Twitter* terhadap kebijakan pemerintah terkait *COVID-19*.

1.3 Tujuan Penelitian

1. Mengukur tingkat *accuracy, precision, recall,* dan *f-measure* metode *Naïve Bayes Classifier* dalam pengklasifikasian teks pada klasifikasi sentimen pengguna *Twitter* terhadap kebijakan pemerintah terkait *COVID-19.*
2. Menganalisa tahapan yang mempengaruhi nilai *accuracy, precision, recall,* dan *f-measure* metode *Naïve Bayes Classifier* dalam pengklasifikasian teks pada klasifikasi sentimen pengguna *Twitter* terhadap kebijakan pemerintah terkait *COVID-19.*

1.4 Batasan Masalah

Data yang digunakan bersumber dari media sosial *Twitter* berbahasa Indonesia dengan rentang waktu posting yang dimulai tanggal 1 Mei 2020 hingga tanggal 30 Juni 2020 (bertepatan dengan beberapa hari sebelum dan sesudah perayaan Hari Raya Idul Fitri 2020) dan pada tanggal 1 Mei 2021 hingga tanggal 30 Juni 2021 (bertepatan dengan beberapa hari sebelum dan sesudah perayaan Hari Raya Idul Fitri 2021) dan berfokus pada *tweets* dari wilayah DKI Jakarta.

1.5 Manfaat Penelitian

Output dari penelitian ini diharapkan dapat memberikan manfaat pada pihak-pihak berikut :

a. Bagi Kementrian Kesehatan diharapkan mampu menjadi bahan pertimbangan diterima tidaknya suatu kebijakan terkait masalah pandemi *COVID-19* di Indonesia

b. Bagi pengamat hukum dan kesehatan sebagai acuan dan bahan pembelajaran terkait diterima tidaknya suatu kebijakan oleh masyarakat

1.6 Sistematika Penulisan

Penelitian skripsi ini terbagi dalam 5 bab dengan runtutan penulisan sebagai berikut :

BAB I Pendahuluan

Bab Pendahuluan membahas mengenai latar belakang masalah, pernyataan masalah, tujuan masalah, batasan masalah, manfaat penelitian, serta sistematika penulisan hasil penelitian.

BAB II Studi Pustaka

Bab Tinjauan pustaka membahas mengenai teori-teori mendasar yang dijadikan acuan serta hasil-hasil penelitian terkait sebelumnya yang dapat mendasari penelitian ini.

BAB III Desain dan Implementasi Sistem

Bab Desain dan Implementasi Sistem membahas mengenai analisa serta perancangan dasar pada kebutuhan sistem yang dibutuhkan untuk menunjang penelitian serta pembuatan aplikasi sebagai hasil dari penelitian.

BAB IV Uji Coba dan Pembahasan

Bab Uji Coba dan Pembahasan membahas mengenai pengujian sistem dari penelitian yang dilakukan serta pembahasan dari hasil pengujian yang dilakukan.

BAB V Kesimpulan dan Saran

Bab Kesimpulan dan Saran membahas mengenai kesimpulan yang dapat diambil dari penelitian yang telah dilakukan serta saran yang dapat dijadikan dasar untuk perkembangan penelitian di masa depan.

BAB II

STUDI PUSTAKA

2.1 *Sentiment Analysis*

Teknologi informasi saat ini menawarkan banyak kesempatan untuk dapat mengembangkan sebuah sistem yang secara otomatis dapat memahami serta mengevaluasi pendapat orang di internet. Analisis terkait pendapat tersebut dikenal juga sebagai analisis sentimen, yang mengacu pada analisis otomatis teks evaluatif dengan berfokus pada pengklasifikasian teks berdasarkan polaritasnya (Lee & Pang, 2008).

Menurut Medhat et al. (2014) analisis sentimen atau penambangan opini adalah sebuah materi perhitungan komputasi terkait opini, sikap, dan emosi pada manusia terhadap suatu objek. Objek disini dapat mewakili sebuah individu, peristiwa, atau berupa topik.

Sentiment analysis atau *opinion mining* merujuk pada bidang pemrosesan bahasa alami, linguistik komputer, dan *text mining* untuk tujuan menganalisis pemikiran, emosi, sikap, ulasan, dan perasaan manusia terkait suatu topik, hasil karya, layanan, organisasi, personal maupun aktivitas lainnya. (Liu, 2010).

Analisis sentimen ini merupakan cabang dari sebuah bidang ilmu yang disebut *Natural Language Processing (NLP)*. Untuk bidang pengolahan bahasa natural atau *NLP* ini memang sudah sering kali dibahas dalam penelitian-penelitian dan telah memiliki sejarah yang panjang dalam pengaplikasiannya pada banyak hal, akan tetapi untuk penganalisaan opini orang dan sentimen belum banyak penelitian

yang membahasnya sebelum tahun 2000. Setelah bahasan mengenai bidang ilmu ini mulai dikenalkan, kini telah banyak penelitian yang aktif memberikan pengembangan dari metode-metode terkait analisis sentimen. Terdapat beberapa alasan mengapa bidang ini banyak diminati, pertama terdapat banyak masalah penelitian yang sebelumnya belum banyak dipelajari dan kurang diperhatikan secara konvensional. Kedua, bahasan terkait pengolahan ini memiliki sumber data yang besar, sebagai contoh dapat diambil dari media sosial ataupun *web*. Ketiga, banyaknya bidang yang sangat memerlukan hasil dari pengolahan opini publik terhadap bidang tersebut, seperti dunia bisnis, ekonomi, politik, dll.

Menganalisis data menggunakan media sosial telah menjadi topik yang menarik bagi para peneliti text mining karena media sosial memudahkan para peneliti untuk mengumpulkan data dan terdapat beberapa alasan mengapa media sosial dapat dikatakan lebih baik dari platform lain, yaitu adaptabilitas dan kemudahan penggunaan, dapat diandalkan dan alat penelitian real-time, mesin pencari dan analisis data lebih cepat. Analisis sentimen adalah salah satu bentuk analisis data yang paling banyak diminati untuk dilakukan (Styawati & Mustofa, 2019).

Analisis sentimen telah menjadi sebuah media bisnis intelijen. Kemampuan analisis sentimen termasuk mengekstraksi pendapat publik terkait isu, produk, atau layanan tertentu yang berisi kumpulan teks tidak terstruktur (Saidah & Mayary, 2020).

Efek dan keuntungan dari analisis sentimen telah menyebabkan pesatnya pertumbuhan penelitian maupun perkembangan aplikasi terkait analisis sentimen,

bahkan di Amerika Serikat sekitar 20-30 perusahaan menggunakan analisis sentimen untuk mengumpulkan informasi tentang opini publik terkait layanan perusahaan (Sumartini, 2011). Cara kerja mendasar dari analisis sentimen adalah proses pengelompokan sebuah kata yang terkandung pada sebuah teks panjang atau naskah panjang dan selanjutnya dilakukan proses pengelolaan teks sehingga dapat ditentukan pendapat yang disampaikan merupakan sentimen positif atau negatif. Penggunaan analisis sentimen juga dapat menyimpulkan tinjauan emosi manusia pada suatu objek. Pendapat dari orang lain inilah yang sangat penting dan mempengaruhi kelanjutan bagi objek yang sedang dibahas.

Berdasarkan karya tulis yang disampaikan oleh Alita dan Rahman (2020) disebutkan jika analisis sentimen bisa dikatakan sebagai semacam tahapan klasifikasi dan kenyataannya tidak sesederhana tahapan klasifikasi pada umumnya dikarenakan adanya pemanfaatan bahasa yang memuat kata ambigu, tidak adanya tanda baca dalam sebuah kalimat, dan perkembangan dari bahasa itu sendiri. Oleh karena itu, pada penelitian tersebut dilakukan sebuah pengembangan terkait analisis sentimen dengan melakukan pendeteksian pada kalimat-kalimat yang mengandung kata-kata sarkasme.

2.2 *Naïve Bayes Classifier (NBC)*

Naïve Bayes Classifier adalah sebuah algoritma pengklasifikasian *Supervised Multiclass* yang didasarkan pada penerapan teorema *Bayes* dengan asumsi " naif " independensi bersyarat antara setiap pasangan variabel (Valdiviezo-Diaz et al., 2019). Ini dapat diartikan bahwa probabilitas satu atribut tidak mempengaruhi probabilitas atribut lainnya. Namun demikian, hasil dari pengklasifikasian *Naïve*

Bayes seringkali benar (Mukherjee & Sharma, 2012). *Naïve Bayes Classifier* dikenal lebih baik dari beberapa metode klasifikasi lainnya. Karena pertama, syarat paling penting dari *Naïve Bayes* adalah anggapan akan ketidakterikatan yang kuat dari setiap kejadian atau peristiwa. Kedua, modelnya sederhana dan mudah dibuat. Ketiga, model dapat diimplementasikan untuk kumpulan data yang besar (Salmi & Rustam, 2019).

Dalam melakukan pengkategorian menggunakan algoritma *Naïve Bayes* memerlukan proses pencarian peluang yang berguna untuk melihat prediksi di masa mendatang berdasarkan data dari masa lampau. Dalam menjalankan proses tersebut tentunya diperlukan sebuah *dataset.* Pada algoritma *Naïve Bayes, dataset* tersebut dibagi menjadi dua yaitu data latih dan data uji (Sidiq et al., 2020).

Keuntungan dalam penggunaan algoritma Naive Bayes adalah metode ini hanya membutuhkan sedikit data pelatihan untuk menentukan perkiraan parameter yang dibutuhkan pada tahapan klasifikasi. *Naive Bayes* berulang kali diketahui dapat menghasilkan hasil yang jauh lebih baik dalam kebanyakan kondisi di dunia nyata bahkan pada sesuatu yang lebih rumit dari yang diperkirakan (Dulhare, 2018).

Pada penelitian terkait mengenai *Naïve Bayes Classifier* disebutkan, jika pada penelitian tersebut telah mampu menghasilkan *classifier* berbasis *Naive Bayes* yang menggunakan *TF-IDF*. Setelah dilatih dengan 2.784 *tweet* beranotasi, validasi silang *K-Fold* menunjukkan bahwa pengklasifikasi memiliki akurasi 72%. Pengklasifikasi kemudian dapat memprediksi polaritas dari 24.273 *tweet*. Hasil prediksi melaporkan bahwa 52% *tweet* positif, dan 48% negatif, yang

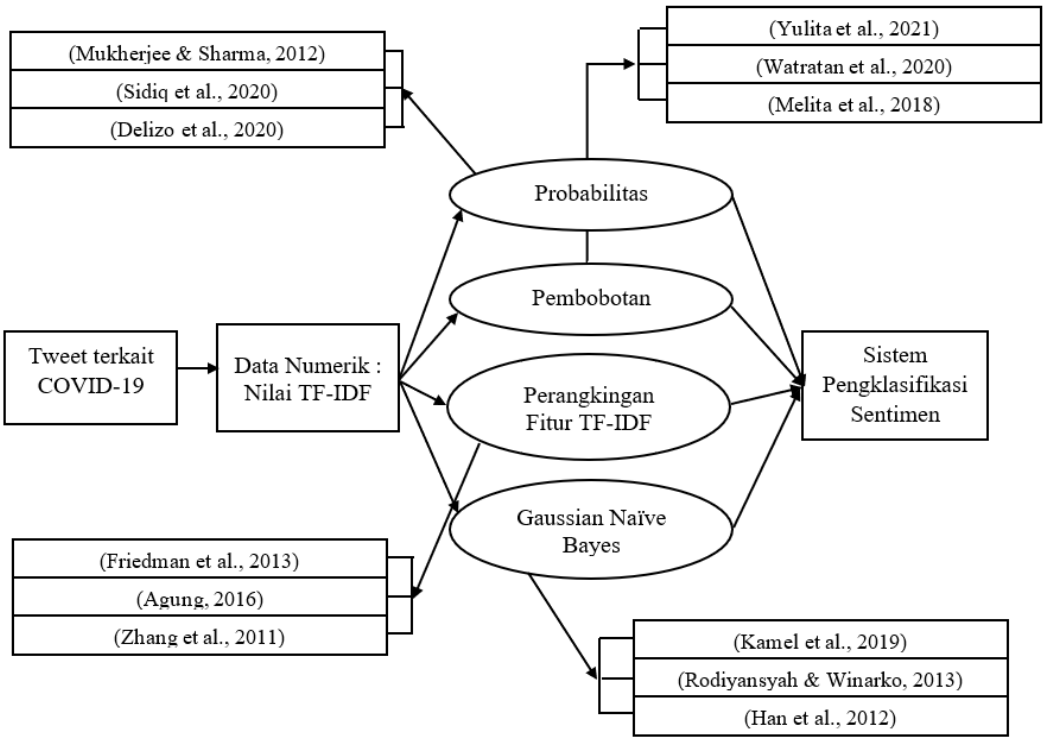
mengungkapkan sentimen keseluruhan *tweet* terkait *COVID-19* dari Januari hingga Maret 2020. Setelah melakukan tinjauan yang cermat, hasil pada penelitian ini dapat disimpulkan bahwa *classifier* yang dihasilkan dapat secara efektif melakukan analisis sentimen pada *tweet* bahasa Inggris dan Tagalog. Hasil penelitian ini menyarankan bahwa perbaikan lebih lanjut harus mencakup penambahan lebih banyak data untuk meningkatkan akurasi dan presisi pengklasifikasi *Naïve Bayes* (Delizo et al., 2020).

Pada karya tulis yang ditulis oleh Yulita et al. (2021) disampaikan jika hasil dari penelitian tersebut adalah rata – rata masyarakat Indonesia merespon positif mengenai program vaksinasi *COVID-19* dengan persentase data sebesar 60,3% dan total data sebesar 2278 data. Respon negatif pada program ini lebih kecil jika dibanding dengan respon netralnya, hal ini menunjukkan jika ternyata hanya sebagian kecil masyarakat yang kontra dengan program vaksinasi ini. Kondisi ini divalidasi dengan didapatkannya persentase respon negatif sebesar 5,4% (203 data), dan respon netral sebesar 34,4% (1299 data). Pemanfaatan metode *Naive Bayes* untuk menganalisis opini ini dapat dikatakan sudah sangat baik, hal ini didukung dengan didapatkannya hasil akurasi yang tinggi sebesar 93%.

Pada penelitian terkait *Naïve Bayes Classifier* mengenai *COVID-19* yang dilakukan oleh Watratan et al. (2020) disebutkan jika algoritma *Naive Bayes* membutuhkan data latih yang berfungsi untuk menghasilkan nilai probabilitas setiap variabel pada kelas yang berbeda, sehingga nilai probabilitas dari variabel tersebut dapat dimaksimalkan untuk meninjau tingkat persebaran *COVID-19* di Indonesia berdasarkan tahapan klasifikasi yang dijalankan menggunakan algoritma

Naive Bayes itu sendiri. Algoritma Naive Bayes didasarkan pada ilmu probabilitas dan statistik, terutama dalam penggunaan data panduan untuk mendukung hasil klasifikasi. Dalam algoritma Naive Bayes, semua variabel akan berkontribusi pada tahapan penentuan hasil klasifikasi, dengan bobot variabel yang sama dan setiap variabel saling tidak terikat. Pada penelitian tersebut yang memanfaatkan data kejadian terkait *COVID-19* pada masing-masing provinsi di Indonesia yang digunakan sebagai data latih, algoritma *Naive Bayes* telah sukses mengklasifikasikan 17 data dari 33 data uji. Sehingga algoritma ini terbilang sukses memperkirakan besaran kasus *COVID-19* pada setiap provinsi dengan nilai persentase akurasi 48,48%.

2.3 *Theoretical Framework*



Gambar 2. 1 Theoretical Framework

Analisa sentimen merupakan sebuah cabang ilmu dari *Natural Language Processing (NLP)* yang dapat merealisasikan sebuah sistem yang berfungsi mengidentifikasi serta memvisualisasikan sebuah opini dari masyarakat mengenai suatu objek yang dibahas dalam bentuk teks. Berdasarkan fakta yang terjadi di lapangan, saat ini informasi mengenai suatu hal serta dampak baik buruknya pada masyarakat telah sangat mudah didapatkan dengan bantuan beberapa *platform* sosial media di internet. Melalui *platform-platform* ini informasi tersebut dapat dikumpulkan untuk selanjutnya dapat diolah dengan analisa sentimen. Melalui analisa sentimen inilah informasi yang tadinya tidak terstruktur dan terkesan tidak ada gunanya dapat diolah menjadi sebuah data yang lebih terstruktur dan memiliki nilai guna yang tinggi. Data-data tersebutlah yang nantinya akan merefleksikan opini masyarakat mengenai produk, kejadian, maupun objek lain sehingga dapat dijadikan tolak ukur kepuasan maupun dampak dari objek tersebut. Dalam analisa sentimen, informasi yang didapat dari masyarakat dapat dibedakan menjadi dua hal, yaitu fakta dan opini. Fakta merupakan pandangan masyarakat secara objektif dan jujur pada suatu objek yang dibahas, sedangkan opini merupakan reaksi subjektif yang merefleksikan emosi masyarakat terkait objek yang dibahas. Hal ini tentunya sangat bermanfaat dalam mengenali respon masyarakat terkait objek yang sedang dibahas.

Pada penelitian ini, sentimen atau opini masyarakat didapatkan melalui *Twitter* dengan menggunakan metode *Scrapping Data* dan yang dimanfaatkan sebagai data primer pada penelitian ini merupakan reaksi pengguna *Twitter* yang divisualisasikan dalam bentuk *tweets* mereka terkait objek khusus yang sedang

dibahas pada penelitian ini. Metode *Scrapping Data* ini adalah metode yang berfungsi untuk mengekstraksi suatu atribut yang biasanya disajikan dalam format dokumen teks untuk kemudian hasilnya disimpan dalam suatu format yang diinginkan. Informasi yang berhasil dikumpulkan serta tertuang pada *tweets* ini dapat dijadikan acuan respon masyarakat terkait suatu objek yang sedang dibahas.

Penelitian ini mengusulkan sebuah metode untuk pengklasifikasian opini dari masyarakat yaitu *Naïve Bayes Classifier (NBC).* Berdasarkan algoritma *NBC* disusunlah sebuah tujuan dari penelitian ini yaitu merancang sebuah sistem pelatihan pada mesin supaya dapat mengklasifikasikan opini dalam bentuk *tweets*.

BAB III

DESAIN DAN IMPLEMENTASI SISTEM

3.1 Akuisisi Data

Data yang nantinya akan dipergunakan sebagai objek utama pada penelitian ini merupakan data primer yang berbentuk komentar-komentar masyarakat yang disampaikan dalam bentuk *tweet* yang tersebar di media sosial *Twitter*. Teknik yang digunakan untuk mengumpulkan data adalah teknik *scraping* pada *Twitter* dengan kata kunci *"ppkm, wfh, psbb, vaksinasi"* dan mengunakan tagar *"#ppkm, #wfh, #psbb, #vaksinasi"* kemudian setiap data yang berhasil diambil tersebut disimpan dan dirapikan dalam bentuk file dengan ekstensi *.csv*. Langkah selanjutnya yang dilakukan pada data tersebut adalah proses pelabelan untuk menentukan label aktual masing-masing kelas sentimen dari setiap komentar tersebut.

Total data komentar yang dikumpulkan dari tahapan *scraping* data yang dimulai pada tanggal 2 Mei 2020 hingga tanggal 29 Juni 2020 dan pada tanggal 1 Mei 2021 hingga tanggal 30 Juni 2021 adalah 1500 *tweet. Tweet* tersebut adalah *tweet* yang didapatkan dari media sosial *Twitter* dengan berfokus pada *tweet* yang berasal dari Provinsi DKI Jakarta. Kemudian dilakukan proses pelabelan secara manual pada data yang telah terkumpul, akan tetapi tidak semua data memasuki tahap pelabelan ini, dilakukan proses pemilihan sampel secara acak untuk kemudian dilakukan proses pelabelan seperti yang terlihat pada Tabel 3.1. Proses pelabelan dilakukan oleh Ibu Baiq Hamidah,S.Pd selaku alumni S1 FBPS, Bahasa Indonesia IKIP Budi Utomo Malang sekaligus Guru Bahasa Indonesia SDN Madyopuro 4 Malang.

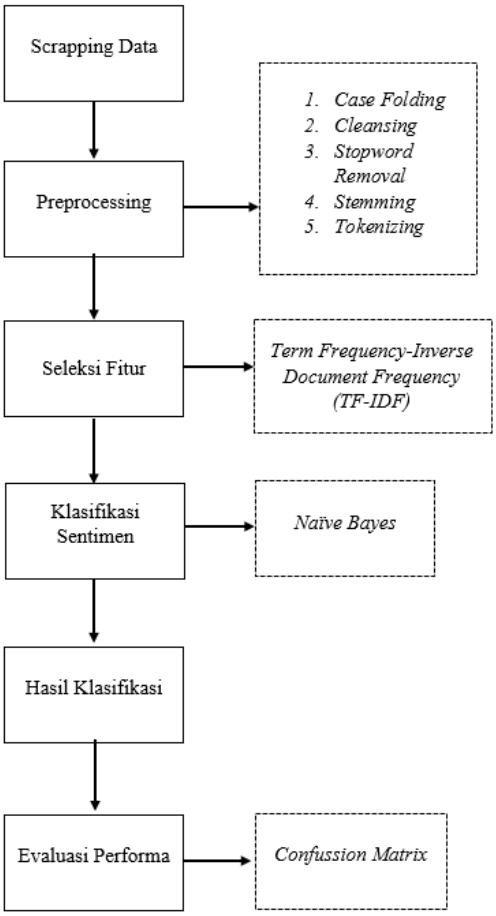
Pelabelan manual yang dilakukan adalah pelabelan dengan cara mengamati opini : pelabelan menggunakan metode ini dikerjakan dengan cara meninjau opini satu persatu, kemudian akan diputuskan opini tersebut dapat dikategorikan pada sentimen positif atau negatif. Keunggulan dari metode pelabelan ini adalah hasil yang diperoleh akan lebih tepat dengan sentimen aslinya dan tentu saja tidak membutuhkan tahapan komputasi (Febrianti, 2020).

Tabel 3. 1 Contoh Pelabelan

DOKUMEN	TWEET	LABEL
D1	@detikcom PPKM sudah terbilang efektif di kota2 besar. Kayak Jakarta turunnya sudah cukup drastis.	Positif
D2	Semoga ppkm tidak diperpanjang, sudah muak asli	Negatif
D3	PPKM sudah selesai ? Semoga bisa menaikkan tingkat kesembuhan	Positif
D4	PPKM jan panjang lagi napa	Negatif

3.2 Rancangan Sistem

Dalam pembuatan sistem klasifikasi ini mempunyai rancangan alur bagaimana langkah demi langkah dalam tiap masing-masing proses mulai dari awal pengumpulan data hingga data terklasifikasi sehingga dapat dilakukan perhitungan tingkat akurasi pada proses pengujian. Berdasarkan pada apa yang disampaikan Villavicencio et al. (2021) terkait rancangan alur sistem pada penelitian tersebut, telah disebutkan untuk masing-masing langkah pada alur sistem terkait analisis sentimen dan metode pengklasifikasiannya. Oleh karena itu, penulis menyusun sebuah rancangan alur sistem yang dapat diamati pada Gambar 3.1.



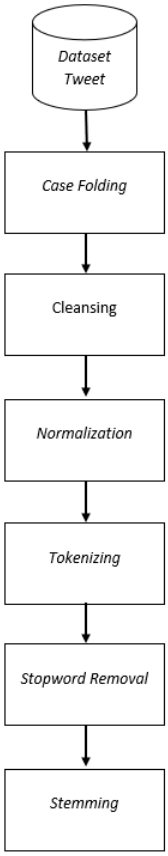
Gambar 3. 1 Rancangan Sistem

Sistem ini dimulai dengan tahapan mengumpulkan data memanfaatkan metode *scraping* data *training* dan *testing*, kemudian proses dilanjutkan dengan klasifikasi manual untuk data *training* dengan menentukan kelas positif dan negatif pada data *training*, kemudian dilakukan proses *Preprocessing* dan proses seleksi fitur pada data *training* maupun data *testing*, kemudian data yang telah diolah sebelumnya disimpan pada model klasifikasi *Naïve Bayes* untuk kemudian dilanjutkan pada tahapan klasifikasi dengan algoritma *Naïve Bayes*, dan tahapan terakhir dan juga output dari sistem ini adalah hasil klasifikasi.

3.2.1 *Preprocessing*

Preprocessing adalah sebuah tahapan menyiapkan dokumen teks yang tidak sistematis menjadi sebuah dokumen yang lebih runtut dan siap untuk diproses pada tahapan berikutnya (Ipmawati et al., 2017). Penggunaan *Preprocessing* sangat berperan pada *Sentiment Analysis*, terutama jika data yang digunakan bersumber dari media sosial yang kebanyakan penggunanya menggunakan kata yang tidak formal dan tidak baku sehingga kondisi ini dapat mempengaruhi tahapan analisis selanjutnya oleh sistem.

Preprocessing adalah sebuah tahap yang penting dalam proses penelitian ini dikarenakan pada proses ini bertujuan untuk meminimalisir atribut yang kurang dibutuhkan dengan cara menghilangkan karakter ataupun merubah karakter yang kurang berpengaruh tersebut sehingga proses klasifikasi bisa menampilkan hasil yang maksimal dan lebih akurat. Gambaran proses yang terjadi pada tahapan *Preprocessing* ini dapat dilihat pada Gambar 3.2.

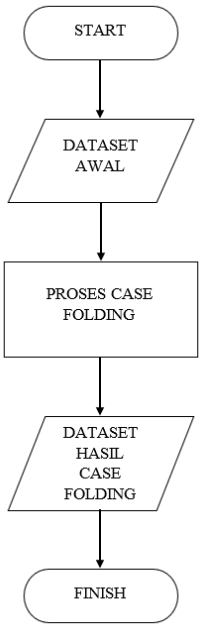


Gambar 3. 2 Blok Diagram *Preprocessing*

a) CaseFolding

CaseFolding adalah sebuah tahapan mengkonversi setiap karakter pada teks menjadi huruf kecil (Mustaqhfiri et al., 2012). Tahap awal yang dilakukan pada tahapan *preprocessing* ini adalah proses *casefolding.* Proses ini bertujuan untuk merubah setiap karakter pada dokumen teks menjadi bentuk huruf kecil semua (*lowercase*). Hal ini berfungsi untuk menyamakan beberapa karakter yang biasanya ditulis menggunakan huruf kapital atau angka menjadi sama semua dalam bentuk huruf kecil semua dan setiap karakter yang bukan huruf akan dihapus karena dianggap sebagai sebuah *delimiter*. Diagram alur dan implementasi

pseudocode proses *CaseFolding* terdapat pada Gambar 3.3 dan Gambar 3.4.



Gambar 3. 3 Diagram Alur Proses *CaseFolding*

```
lowercase = tweet.lower()
print (lowercase)
```

Gambar 3. 4 Implementasi *CaseFolding*

b) Cleansing

Tahapan selanjutnya adalah proses *Cleansing*. Tahapan ini sebenarnya adalah tahapan tambahan dari tahapan sebelumnya, pada tahap ini bertujuan untuk menghapus karakter selain huruf yang tidak berhubungan dengan tema penelitian seperti "#, @, angka, tautan, dan tanda baca lainnya". Proses ini bermanfaat untuk mengurangi karakter-karakter yang kurang dibutuhkan pada proses klasifikasi nantinya.

Diagram alur dan implementasi *pseudocode* proses *Cleansing* dapat dilihat pada Gambar 3.5 dan Gambar 3.6.

Gambar 3. 5 Diagram Alur Proses Cleansing

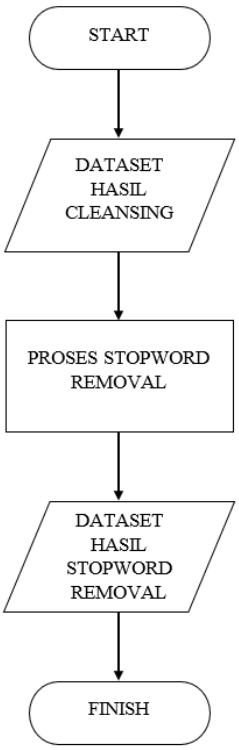
```
hasil = re.sub(r"\d+","",kalimat)
hasil1 = re.sub("@[A-Za-z0-9]+","",hasil)
hasil2 = re.sub(r'(?i)\b((?:https?://|www\d{-,3}[.]|
         [a-z0-9.\-]+[.][a-z]{2,4}/)(?:[^\s()<>]+|\
         (([^\s()<>]+|(\([^\s()<>]+\)))*\))+(?:\(([^\s()<>]+|(\
         ([^\s()<>]+\)))*\)|[^\s`!()\[\]{};:\'".,<>?<<>>""'']))
         ','',hasil1)
hasil3 = re.sub(r'[^\w]|_',' ',hasil2)
hasil4 = re.sub("\S*\d\S*", "",hasil3).strip()
hasil5 = re.sub(r"\b\d+\b"," ",hasil4)
hasil6 = re.sub('[\s]+',' ',hasil5)
hasil7 = hasil6.strip()
print (hasil7)
```

Gambar 3. 6 Implementasi *Cleansing*

c) Stopword Removal

Proses selanjutnya adalah proses *Stopword Removal*. Proses ini bertujuan untuk menghapus kata yang kemungkinan keluar dalam

jumlah besar dalam sebuah dokumen teks yang tidak memiliki makna penting pada tema yang dibahas dan hilangnya kata-kata ini tidak akan mempengaruhi hasil klasifikasi. Kata-kata ini dalam bahasa Indonesia biasanya merupakan kata hubung seperti “di, dari, pada, kah, kan, dengan, yang, dll.”. Hal ini berfungsi untuk mengurangi jumlah kata yang dianalisa sehingga mempermudah proses analisa dan klasifikasi nantinya. Diagram alur dan implementasi *pseudocode* proses *Stopword Removal* digambarkan pada Gambar 3.7 dan Gambar 3.8.



Gambar 3. 7 Diagram Alur Proses *Stopword Removal*

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

factory = StopWordRemoverFactory()
stopword = factory.create_stop_word_remover()
stop = stopword.remove(kalimat)
print (stop)
```

Gambar 3. 8 Implementasi *Stopword Removal*

d) Stemming

Menurut Zaman dan Winarko (2011) *Stemming* adalah tahapan yang berfungsi untuk mengubah berbagai bentuk kata seperti *prefix, sufix,* maupun gabungan antara *prefix* dan *sufix (confix),* menjadi akar kata dari kata tersebut. Proses berikutnya adalah proses *Stemming.* Proses ini memungkinkan kita untuk mengubah kata dalam sebuah dokumen teks menjadi bentuk akar katanya berdasarkan aturan tertentu. Pada teks dalam bahasa Indonesia, teknik ini digunakan untuk menghapus imbuhan pada kata tertentu. Pada tahap ini digunakan sebuah algoritma karya Bobby Nazief dan Mirna Adriani, dengan rincian algoritma sebagai berikut :

1. Kata awal dikoreksi pada daftar perbendaharaan akar kata, jika mendapati kata tersebut di dalamnya maka kata berikut dilabeli sebagai akar kata dan algoritma dapat berhenti.
2. Menghapus *Inflectional suffixes* dengan cara menghapus atribut (“-lah”, ”-kah”, “-tah” atau “-pun”), selanjutnya menghapus *inflectional possessive pronoun suffixes* (“-ku”, “-mu” atau ”-nya”). Koreksi kata pada daftar perbendaharaan akar kata, jika terdapat kata tersebut, algortima dapat dihentikan, jika kata belum didapatkan maka dilanjutkan pada tahapan selanjutnya.
3. Menghapus *Derivational Suffix* (“-i” atau ”-an”,”). Jika kata sudah ditemukan pada daftar perbendaharaan akar kata, maka

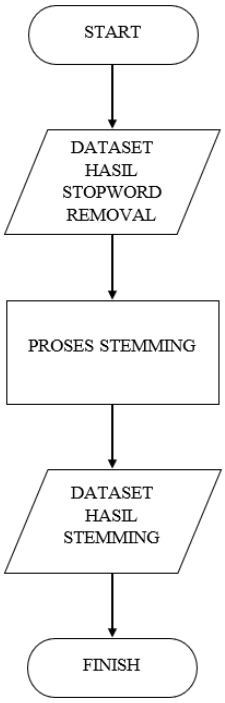
algoritma dihentikan, jika kata belum ditemukan maka dilanjutkan pada langkah 3a :

a. Apabila akhiran “-an” telah dihapus tetapi huruf akhir kata tersebut adalah “-k”, maka huruf “-k” akan dihapus juga. Apabila kata tersebut telah ditemukan pada daftar perbendaharaan akar kata maka algoritma dapat dihentikan, jika kata belum ditemukan maka dilanjutkan pada langkah 3b.
b. Akhiran yang sebelumnya telah terhapus (“-i”, “- an” atau “-kan”) dapat dikembalikan dan proses dilanjutkan pada langkah 4.

4. Menghapus *Derivational Prefix* (“be-”,”di-”,”ke-”,”me-”,”pe-“,”se-” dan “te-“). Apabila kata telah ditemukan pada daftar perbendaharaan akar kata, maka algoritma dihentikan, jika kata belum ditemukan maka dilanjutkan pada tahapan *recoding*. Tahapan ini akan berhenti apabila telah mencapai syarat-syarat berikut :
   a. Mengandung kombinasi awalan dan akhiran yang tidak diijinkan.
   b. Awalan yang ditemukan mirip dengan awalan yang telah dihapus sebelumnya.
   c. Tiga awalan sudah dihapus.

5. Jika setiap tahapan telah dijalankan tetapi akar kata belum ditemukan pada daftar perbendaharaan akar kata maka algoritma ini akan menampilkan kata yang awal sebelum memasuki tahapan *stemming*.

Diagram alur dan implementasi *pseudocode* proses *Stemming* dapat dilihat pada Gambar 3.9 dan Gambar 3.10.



Gambar 3. 9 Diagram Alur Proses *Stemming*

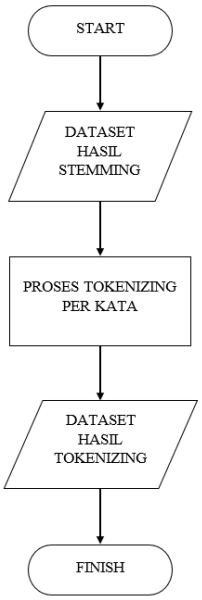
```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

stem_factory = StemmerFactory()
stemmer = stem_factory.create_stemmer()
stem = stemmer.stem(kalimat)
print (stem)
```

Gambar 3. 10 Implementasi *Stemming*

e) *Tokenizing*

Tokenizing adalah proses pemisahan *string* dalam sebuah kalimat sehingga didapat potongan tunggal dalam bentuk kata yang menyusun kalimat tersebut. Pemisahan kalimat yaitu proses memisah *string* teks panjang menjadi sekumpulan kata penyusunnya. Pemisahan kalimat menjadi kumpulan kata menggunakan fungsi split (), dengan beberapa tanda pemisah seperti titik “.”, tanda tanya ”?” dan tanda seru “!” (Mustaqhfiri et al., 2012). Proses berikutnya merupakan proses *Tokenizing*. Proses ini dilakukan untuk memotong dokumen yang berbentuk kalimat atau paragraf menjadi bentuk per kata atau yang biasa disebut sebagai satu entitas. Pemisahan ini didasarkan pada kata dalam sebuah kalimat atau paragraf yang dipisahkan oleh spasi. Setiap kata yang telah dipisah tersebut dapat dikatakan sebagai sebuah *token*. *Token* inilah yang nantinya dapat membantu pada proses pembobotan. Diagram alur dan implementasi *pseudocode* proses *Tokenizing* terdapat pada Gambar 3.11 dan Gambar 3.12.



Gambar 3. 11 Diagram Alur Proses *Tokenizing*

```
token = kalimat.split()
print (token)
```

Gambar 3. 12 Implementasi *Tokenizing*

3.2.2 Seleksi Fitur *TF-IDF*

Term Frequency-Inverse Document Frequency (*TF-IDF*) adalah algoritma statistik numerikal yang menghasilkan nilai bobot untuk masing-masing atribut (atau kata) dalam sebuah dokumen teks (Masuda et al., 2011). Metode ini sering digunakan dalam pemrosesan bahasa alami (*NLP*) atau dalam pencarian informasi dan penambangan teks (Friedman et al., 2013). Metode ini menentukan bobot kata dalam teks, mengukur tingkat kepentingan suatu istilah (kata) pada sekumpulan dokumen. Tingkat kepentingan suatu kata meningkat sebanding dengan total kata muncul pada sebuah teks dokumen (Zhang et al., 2011). Algoritma Term Frequency-Inverse Document Frequency (TFIDF) adalah salah satu cara pemberian

bobot pada hubungan sebuah kata (term) dengan sebuah dokumen. TFIDF adalah ukuran statistik yang dimanfaatkan untuk mengevaluasi seberapa penting kata pada sebuah dokumen atau sekumpulan kata. Tingkat kemunculan sebuah kata pada sebuah dokumen menampilkan tingkat kepentingan suatu kata pada dokumen tersebut. Seberapa banyak dokumen yang memiliki kata tersebut menentukan seberapa umum kata tersebut pada sekumpulan dokumen. Bobot kata akan tinggi jika kata tersebut sering muncul pada sebuah dokumen dan akan rendah jika kata sering muncul pada beberapa dokumen (Agung, 2016). Algoritma *TF-IDF* ini dimanfaatkan pada penelitian ini, dikarenakan algoritma ini dikenal sebagai sebuah algoritma pembobotan yang terkenal efektif, mudah diterapkan dan menghasilkan akurasi yang baik (Melita et al., 2018). Hal ini dapat diketahui hanya dengan waktu pemrosesan yang cepat bahkan tanpa menggunakan proses komputasi yang banyak sudah dapat menghasilkan hasil yang akurat, terlebih dengan konsep *TF-IDF* yang memberikan nilai rendah pada kata yang sering muncul pada semua dokumen yang digunakan sehingga metode ini juga dapat digunakan sebagai metode penentuan kata-kata untuk *StopList* yang pada bahasa Indonesia biasanya berupa kata-kata hubung yang tidak dibutuhkan pada penelitian ini dapat dihilangkan sehingga mampu mempermudah untuk proses-proses pengolahan kata selanjutnya

Tahapan selanjutnya dalam proses penelitian ini adalah tahapan seleksi fitur. Pada penelitian ini penulis memanfaatkan algoritma *TF-IDF (Term Frequency-Inverse Document Frequency)*. *TF* adalah total kata muncul pada sebuah dokumen. Besar kecilnya nilai *TF* bergantung pada seberapa sering sebuah kata muncul dalam dokumen, semakin sering kemunculan sebuah kata maka nilai *TF* akan semakin

besar. Sedangkan *IDF* adalah perhitungan seberapa banyak sebuah kata terdistribusikan pada semua dokumen tersedia. Berbanding terbalik dengan *TF*, nilai *IDF* akan semakin tinggi jika frekuensi kemunculan kata semakin sedikit.

a) Pembobotan Fitur

Langkah berikutnya adalah menghitung besaran bobot *TF* dan *IDF* dari masing-masing kata menggunakan persamaan *TF-IDF* berikut :

$$TF_{ij} = \frac{f_{ij}}{max_k f_{kj}} \quad \textbf{(3.1)}$$

Dimana f_{ij} merupakan jumlah banyaknya kemunculan *(i) term* atau kata dalam suatu dokumen *(j)*, sedangkan $max_k f_{kj}$ adalah jumlah total *(i) term* atau kata yang muncul pada suatu dokumen *(j)*.

$$IDF_i = \log\left(\frac{N}{n_i}\right) + 1 \quad \textbf{(3.2)}$$

Dimana N merupakan total dari semua dokumen yang digunakan, sedangkan n_i merupakan total dokumen yang memiliki *(i) term* atau kata tersebut. Besaran nilai *TF-IDF* dapat dihitung menggunakan persamaan (3.3).

$$TFIDF_{ij} = TF_{ij} \times IDF_i \quad \textbf{(3.3)}$$

b) Perangkingan Fitur

Langkah selanjutnya pada bagian seleksi fitur adalah perangkingan fitur berdasarkan besaran bobot total dari nilai tertinggi hingga terendah. Perhitungan bobot total dilakukan dengan menjumlah keseluruhan bobot

yang dimiliki oleh masing-masing term pada keseluruhan dokumen yang tersedia.

Sebagai gambaran, implementasi nilai TF-IDF disampaikan dalam bentuk Tabel 3.2. Implementasi *pseudo-code* dari proses TF-IDF dapat diamati pada gambar 3.13.

Tabel 3. 2 Desain Nilai *TF-IDF*

TERM	D_1	D_2	D_3	…	D_n	***Bobot Total***
$Term_1$	$TF\text{-}IDF_{1,1}$	$TF\text{-}IDF_{2,1}$	$TF\text{-}IDF_{3,1}$	…	$TF\text{-}IDF_{n,1}$	$TF\text{-}IDF_{Total,1}$
$Term_2$	$TF\text{-}IDF_{1,2}$	$TF\text{-}IDF_{2,2}$	$TF\text{-}IDF_{3,2}$	…	$TF\text{-}IDF_{n,2}$	$TF\text{-}IDF_{Total,2}$
$Term_3$	$TF\text{-}IDF_{1,3}$	$TF\text{-}IDF_{2,3}$	$TF\text{-}IDF_{3,3}$	…	$TF\text{-}IDF_{n,3}$	$TF\text{-}IDF_{Total,3}$
$Term_4$	$TF\text{-}IDF_{1,4}$	$TF\text{-}IDF_{2,4}$	$TF\text{-}IDF_{3,4}$	…	$TF\text{-}IDF_{n,4}$	$TF\text{-}IDF_{Total,4}$
$Term_5$	$TF\text{-}IDF_{1,5}$	$TF\text{-}IDF_{2,5}$	$TF\text{-}IDF_{3,5}$	…	$TF\text{-}IDF_{n,5}$	$TF\text{-}IDF_{Total,5}$
$Term_6$	$TF\text{-}IDF_{1,6}$	$TF\text{-}IDF_{2,6}$	$TF\text{-}IDF_{3,6}$	…	$TF\text{-}IDF_{n,6}$	$TF\text{-}IDF_{Total,6}$
$Term_7$	$TF\text{-}IDF_{1,7}$	$TF\text{-}IDF_{2,7}$	$TF\text{-}IDF_{3,7}$	…	$TF\text{-}IDF_{n,7}$	$TF\text{-}IDF_{Total,7}$
$Term_8$	$TF\text{-}IDF_{1,8}$	$TF\text{-}IDF_{2,8}$	$TF\text{-}IDF_{3,8}$	…	$TF\text{-}IDF_{n,8}$	$TF\text{-}IDF_{Total,8}$
$Term_9$	$TF\text{-}IDF_{1,9}$	$TF\text{-}IDF_{2,9}$	$TF\text{-}IDF_{3,9}$	…	$TF\text{-}IDF_{n,9}$	$TF\text{-}IDF_{Total,9}$
$Term_{10}$	$TF\text{-}IDF_{1,10}$	$TF\text{-}IDF_{2.10}$	$TF\text{-}IDF_{3,10}$	…	$TF\text{-}IDF_{n,10}$	$TF\text{-}IDF_{Total,10}$
…	…	…	…	…	…	…
$Term_m$	$TF\text{-}IDF_{1,m}$	$TF\text{-}IDF_{2,m}$	$TF\text{-}IDF_{3,m}$	…	$TF\text{-}IDF_{n,m}$	$TF\text{-}IDF_{Total,m}$

```
//Pemisahan Kata Unik
bagOfWordsA = documentA.split(' ')
bagOfWordsB = documentB.split(' ')
bagOfWordsC = documentC.split(' ')

uniqueWords = set(bagOfWordsA).union(set(bagOfWordsB).union(set(bagOfWordsC)))
print(uniqueWords)

//Jumlah Kata pada Dokumen
numOfWordsA = dict.fromkeys(uniqueWords, 0)
for word in bagOfWordsA:
    numOfWordsA[word] += 1

numOfWordsB = dict.fromkeys(uniqueWords, 0)
for word in bagOfWordsB:
    numOfWordsB[word] += 1

numOfWordsC = dict.fromkeys(uniqueWords, 0)
for word in bagOfWordsC:
    numOfWordsC[word] += 1
```

```
//Perhitungan TF
def computeTF(wordDict, bagOfWords):
    tfDict = {}
    bagOfWordsCount = len(bagOfWords)

    for word, count in wordDict.items():
        tfDict[word] = count / float(bagOfWordsCount)

    return tfDict

tfA = computeTF(numOfWordsA, bagOfWordsA)
tfB = computeTF(numOfWordsB, bagOfWordsB)
tfC = computeTF(numOfWordsC, bagOfWordsC)
```

```
//Perhitungan IDF
def computeIDF(documents):
    import math
    N = len(documents)

    idfDict = dict.fromkeys(documents[0].keys(), 0)
    for document in documents:
        for word, val in document.items():
            if val > 0:
                idfDict[word] += 1

    for word, val in idfDict.items():
        idfDict[word] = (math.log10(N / float(val))+1)
    return idfDict

idfs = computeIDF([numOfWordsA, numOfWordsB, numOfWordsC])
```

```
//Perhitungan TF-IDF
def computeTFIDF(tfBagOfWords, idfs):
    tfidf = {}
    for word, val in tfBagOfWords.items():
        tfidf[word] = val * idfs[word]
    return tfidf

tfidfA = computeTFIDF(tfA, idfs)
tfidfB = computeTFIDF(tfB, idfs)
tfidfC = computeTFIDF(tfC, idfs)
```

Gambar 3. 13 Implementasi *TF-IDF*

3.2.3 Klasifikasi *Naïve Bayes*

Proses selanjutnya, yang merupakan proses utama dari penelitian ini adalah proses klasifikasi dan pada penelitian kali ini algoritma yang dimanfaatkan adalah algoritma pengklasifikasian menggunakan algoritma *Naïve Bayes Classifier*. Ciri khas yang sangat kuat dari metode ini adalah hipotesis yang kuat akan

ketidakterikatan pada setiap variabel. Proses pengklasifikasian ini membutuhkan dua proses pengerjaan, yaitu proses *training data* dan proses *testing data*. Proses *training* digunakan untuk pelatihan dan pengenalan *dataset* pada sistem, sedangkan proses *testing* digunakan untuk pengujian *dataset* dengan mengacu pada perhitungan probabilitas dari proses *training*.

Umumnya teorema *Bayes* dituliskan dengan persamaan (3.4) (Han et al., 2012) :

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad \textbf{(3.4)}$$

Dimana:

$P(A|B)$ = *Posterior probability*

$P(B|A)$ = *Likelihood*

$P(A)$ = *Class prior probability*

$P(B)$ = Predictor prior probability

Tahapan untuk menentukan probabilitas suatu data diawali dengan menentukan nilai *likelihood* berdasarkan pada *dataset* yang disiapkan. Nilai *Likelihood* yang didapatkan selanjutnya akan dikalikan dengan setiap nilai probabilitas dari data yang digunakan. Hasil dari perhitungan tersebut selanjutnya akan digunakan sebagai panduan untuk mengklasifikasi data uji baru. Pada praktiknya, nilai P(B) sering diabaikan, dikarenakan nilai P(B) konstan (Saputro & Sari, 2019).

Tahapan pengklasifikasian menggunakan *Naïve Bayes Classifier* pada dokumen merepresentasikan setiap dokumen dengan atribut “ $X1$,$X2$,$X3$, … , Xn ”

yang memiliki arti bahwasanya variabel $X1$ untuk kata pertama, $X2$ untuk kata kedua, dan seterusnya. Kumpulan dokumen *tweet* direpresentasikan dengan V. Ketika menjalankan tahapan klasifikasi, *Naïve Bayes Classifier* akan menentukan nilai probabilitas paling tinggi dari setiap kelas dokumen diuji (V_{MAP}) dengan memasukkan atribut $(a_1, a_2, a_3, \dots, a_n)$. Sedang persamaan untuk V_{MAP} ditulis dalam bentuk persamaan (3.5) berikut :

$$V_{MAP} = \underset{V_j \in V}{argmax} P(V_j | a_1, a_2, a_3, \dots, a_n) \quad \textbf{(3.5)}$$

Berdasarkan teorema *Bayes*, maka persamaan (3.5) dapat ditulis menjadi,

$$V_{MAP} = \underset{V_j \in V}{argmax} \frac{P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j)}{P(a_1, a_2, a_3, \dots, a_n)} \quad \textbf{(3.6)}$$

$P(a_1, a_2, a_3, \dots, a_n)$ nilainya tetap untuk setiap dokumen V_j sehingga persamaan (3.6) dinyatakan dengan persamaan berikut :

$$V_{MAP} = \underset{V_j \in V}{argmax} P(a_1, a_2, a_3, \dots, a_n | V_j) P(V_j) \quad \textbf{(3.7)}$$

Naïve Bayes Classifier meringankan perhitungan ini dengan asumsi bahwa pada masing-masing kelas, setiap variabel saling tidak terikat.

Sehingga dapat diartikan sebagai berikut :

$$V_{MAP} = \underset{V_j \in V}{argmax} P(V_j) \prod_i P(a_i | V_j) \quad \textbf{(3.8)}$$

Kemudian jika pada persamaan (3.7) disubstitusikan pada persamaan (3.8), maka dapat menghasilkan persamaan (3.9) berikut :

$$V_{MAP} = \underset{V_j \in V}{argmax} P(V_j) \prod_i P(a_i | V_j) \quad \textbf{(3.9)}$$

$P(V_j)$ dan nilai probabilitas dari kata a_i pada setiap kelas $P(a_i | V_j)$ yang dihitung dirumuskan sebagai berikut,

$$P(V_j) = \frac{docs_j}{training} \quad \textbf{(3.10)}$$

$$P(a_i|V_j) = \frac{n_i + 1}{n + kosakata} \quad \textbf{(3.11)}$$

Dimana $docs_j$ adalah total dokumen pada kelas *j* dan *training* adalah total dokumen. Variabel n_i merepresentasikan nilai total kata a_i muncul dalam kelas V_j, *n* adalah total kata yang muncul pada kelas V_j dan *kosakata* adalah total kata yang berbeda pada semua dokumen (Rodiyansyah & Winarko, 2013).

Terkait dengan kemungkinan adanya probabilitas nol pada masing-masing probabilitas maka perlu menambahkan nilai 1 pada masing-masing perkiraan perhitungan probabilitas. Teknik berikut disebut juga dengan teknik *Laplace Smoothing.*

3.2.4 *Gaussian Naïve Bayes*

Algoritma yang penulis gunakan pada sistem adalah *Gaussian Naïve Bayes*. Algoritma ini biasanya digunakan untuk pengklasifikasian yang memiliki atribut bersifat kontinu. Algoritma ini memanfaatkan fungsi dasar distribusi normal *Gaussian.* Ciri utama algoritma ini adalah penggunaan 2 parameter utama, yaitu mean (μ) dan varians (σ^2). Mean adalah nilai rata-rata dari nilai atribut yang bersifat kontinu. Rumus perhitungan untuk mencari nilai rata-rata dapat dilihat pada persamaan (3.12). Varians adalah nilai kuadrat dari simpangan nilai data pada atribut yang bersifat kontinu. Rumus perhitungan dari varians tertulis pada persamaan (3.13).

$$\mu = \frac{\in X}{N} \quad \textbf{(3.12)}$$

Dimana :

- μ = mean
- $\in X$ = jumlah nilai data TF-IDF
- N = jumlah data

$$\sigma = \frac{\in (X - \mu)^2}{N - 1} \quad \textbf{(3.13)}$$

Dimana :

- σ = varians dari data TF-IDF
- X = nilai data TF-IDF
- μ = nilai rata-rata data TF-IDF
- N = jumlah data

Ketika berhadapan dengan data kontinu, terdapat anggapan umum bahwa nilai kontinu yang berkorelasi dengan setiap kelas didistribusikan sesuai dengan distribusi *Gaussian*. Perhitungan pada data pelatihan dilakukan dengan membagi berdasarkan kelas yang tersedia, kemudian mean dan standar deviasi pada masing-masing kelas dihitung. Selanjutnya dilakukan proses untuk menghitung probabilitas dari kumpulan data kontinu menggunakan persamaan (3.14) berikut (Kamel et al., 2019) :

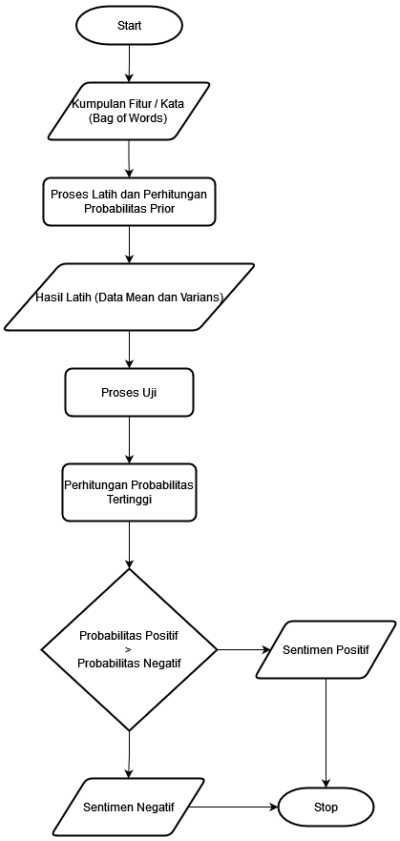
$$P(B|A) = \frac{1}{\sqrt{2}.\pi.\sigma} exp^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \textbf{(3.14)}$$

Dimana :

- $P(B|A)$ = probabilitas *likelihood*
- σ = varians dari data TF-IDF
- π = 3.14

- μ = nilai rata-rata data TF-IDF

Sebagai gambaran, berikut ditampilkan sebuah diagram alur dari tahapan klasifikasi menggunakan algoritma *Gaussian Naïve Bayes* yang dapat diamati pada Gambar 3.14.



Gambar 3. 14 Diagram Alur Klasifikasi *Gaussian Naïve Bayes*

Proses awal pada tahap ini adalah penginputan kumpulan kata (*bag of words)* yang telah diberi bobot pada proses sebelumnya, kemudian data tersebut digunakan pada proses pelatihan. Langkah awal pada proses pelatihan ini adalah dengan membagi dokumen latih berdasarkan kelasnya, yaitu kelas positif dan negatif. Langkah berikutnya adalah dengan menghitung nilai *mean* dan *varians* pada setiap kata pada setiap kelas menggunakan persamaan (3.12) dan persamaan (3.13).

Langkah selanjutnya pada proses ini adalah menghitung probabilitas *prior* dari masing-masing kelas menggunakan persamaan (3.10). Proses berikutnya adalah proses pengujian.

Perhitungan probabilitas pada tahap pengujian ini memanfaatkan persamaan (3.14) dengan nilai *mean* dan *varians* memanfaatkan nilai yang didapat dari data *training* sebelumnya, akan tetapi jika menemui kata yang nilai *mean* dan *varians* tidak terdapat pada data *training* atau bernilai 0 maka probabilitas dari kata tersebut bisa tidak digunakan atau bisa juga dianggap 1.

Langkah selanjutnya dari proses ini adalah mengkalikan hasil probabilitas dari masing-masing kata pada kelas yang sama dengan probabilitas *prior* kelas tersebut. Proses inilah yang nantinya akan menentukan kelas hasil klasifikasi dari data *testing*, dimana kelas tersebut ditentukan oleh hasil perhitungan dengan total nilai probabilitas tertinggi.

3.3 Skenario Pengujian

Berikut penulis berikan langkah-langkah untuk proses pengujian yang akan dilakukan nantinya pada penelitian ini :

1. *Input Dataset*
2. Pembagian *Dataset*
3. Penentuan Penggunaan Persentase Jumlah Fitur
4. Pelatihan *Dataset*
5. Pengujian *Dataset*
6. Penyajian Hasil
7. Evaluasi Kinerja Sistem

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Langkah-Langkah Pengujian

Pengujian dilakukan dengan mengacu pada langkah-langkah pengujian yang disusun secara runtut dan sistematis sehingga proses pengujian pada sistem yang dibangun dapat dilakukan dengan baik dan mudah untuk dipahami. Skenario pengujian yang digunakan pada pengujian sistem kami adalah sebagai berikut.

1. *Input* Data Uji

   Langkah pertama yang dilakukan pada pengujian ini adalah mengisikan data pada sistem. Data yang yang digunakan merupakan data hasil *scraping* pada media sosial *Twitter* yang telah tersimpan pada file dengan ekstensi *.csv*. Data tersebut berjumlah 1500 dengan rincian 796 *tweets* negatif dan 719 *tweets* positif yang sebelumnya telah dilakukan pelabelan oleh tenaga ahli. Contoh *dataset* yang digunakan dapat dilihat pada Tabel 4.1 berikut.

Tabel 4. 1 Contoh *Dataset*

No.	Label	Text
1	Negatif	PPKM panjang deh :(
2	Negatif	gaenak banget njrit wkwkwkw gara gara PPKM gabisa makan begituan
3	Positif	Belajar dari PSBB sebelumnya, ada baiknya Anda juga memiliki persiapan matang dalam menghadapi PPKM darurat kali ini. Tentunya akan lebih bijak apabila Anda mampu mengelola keuangan di masa-masa PPKM seperti sekarang ini. #PDIPerjuangan #SolidBergerak
4	Positif	Berakhir Besok, Anies Baswedan: Ini Bukan Perkara PPKM Level 4 Tapi Soal Keselamatan https://t.co/p3Y2AGEvq4
5	Negatif	PPKM ayodong berakhir sedih banget gw
6	Positif	Detik detik kemerdekaan ppkm beres bos
7	Negatif	ppkm ini kapan sih kelarnya? capek bgt gw njir
8	Positif	#BesokSenin apa ppkm bakal di perpanjang, kalo dilihat hasil PPKM kemaren kasus covid udh turun padahal
9	Positif	aaaa makasih, km juga harus jaga kesehatan ya selagi ppkm manfaatin buat hidup sehat
10	Negatif	Besok PPKM diperpanjang lg ga? Kalo diperpanjang, siap2 mau ngurus dokumen pindah ke Bruney soalnya

2. Pembagian *Dataset*

Langkah selanjutnya pada pengujian sistem kami pembagian *dataset* menjadi dua bagian, yaitu data latih dan data uji dengan jumlah persentase sebesar 60% dari dataset menjadi data *train* dan 40% dari dataset menjadi data *test*. Dataset juga dibagi berdasarkan jumlah persentase penggunaan fitur sesuai dengan yang akan disebutkan pada poin selanjutnya. Tahapan ini bertujuan untuk menentukan jumlah data yang digunakan pada masing-masing proses, baik pada proses pelatihan maupun proses pengujian.

3. Penentuan Penggunaan Persentase Jumlah Fitur

Kemudian langkah selanjutnya adalah dengan menentukan persentase penggunaan fitur untuk pengujian dengan besaran nilai persentase 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% dari total fitur. Penggunaan jumlah fitur pada penelitian ini didasarkan pada penelitian yang dilakukan oleh ButarButar et al. (2019) menurut penelitian tersebut pembagian ini berfungsi untuk mengukur tingkat akurasi sistem dengan lebih baik melalui beberapa variasi pengujian.

4. Pelatihan Dataset

Proses pelatihan diawali dengan pembagian dataset tergnatung pada jumlah kelas yang diinginkan, pada pengujian ini dataset dibagi menjadi 2 kelas yaitu positif dan negatif. Setelah dataset terkumpul pada masing-masing kelas maka langkah selanjutnya adalah dengan menghitung nilai mean dan standar deviasi pada masing-masing kata.

5. Pengujian Dataset

Tahapan ini merupakan tahap akhir dari algoritma yang digunakan pada sistem. Tahapan ini dilakukan dengan menghitung hasil klasifikasi akhir dari data uji dengan menggunakan rumus dasar probabilitas *Gaussian Naïve Bayes* seperti yang ditunjukkan pada persamaan (3.14). Nilai x yang digunakan merupakan bobot data uji, sedangkan nilai *mean* dan *varians* nya menggunakan nilai yang didapat dari perhitungan pada data pelatihan. Kemudian hasil probabilitas pada masing-masing kata yang didapat dikalikan dengan nilai probabilitas prior masing-masing kelas untuk mendapatkan hasil

probabilitas masing-masing dokumen pada masing-masing kelas. Hasil probabilitas tertinggi dijadikan hasil klasifikasi dari dokumen tersebut.

6. Penyajian Hasil

Tahapan ini menyajikan hasil klasifikasi dari pengujian yang dilakukan berupa perbandingan antara kelas aktual dengan kelas hasil klasifikasi sistem. Hasil yang ditampilkan merupakan hasil klasifikasi yang telah dilakukan dengan menggunakan persentase jumlah fitur seperti yang telah disebutkan pada poin 3. Berdasarkan pada nilai hasil yang disajikan ini nantinya dapat diketahui besaran nilai *accuracy, precision, recall,* dan *f-measure.*

7. Evaluasi Kinerja Sistem

Kemudian setelah didapatkan hasil dari data yang telah terklasifikasi maka dilanjutkan pada proses perhitungan tingkat akurasi pada klasifikasi menggunakan metode klasifikasi *Naïve Bayes*. Hasil dari skenario pengujian sebelumnya akan dilakukan perhitungan pada nilai *accuracy, precision, recall,* dan *f-measure* menggunakan metode *Confussion Matrix*.

Perhitungan nilai *accuracy, precision, recall,* dan *f-measure* masing-masing dilakukan dengan menggunakan persamaan (4.1), (4.2), (4.3), (4.4).

$$Accuracy : \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \,.\, 100\% \qquad \textbf{(4.1)}$$

$$Precision : \frac{\mathrm{TP}}{\mathrm{TP+FP}} \,.\, 100\% \qquad \textbf{(4.2)}$$

$$Recall : \frac{\mathrm{TP}}{\mathrm{TP+FN}} \,.\, 100\% \qquad \textbf{(4.3)}$$

$$F\text{-}measure : 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad \textbf{(4.4)}$$

Dimana :

- TP *(True Positive)* adalah data aktual berlabel positif (1) diprediksi sistem berlabel positif (1)
- TN *(True Negative)* adalah data aktual berlabel negatif (0) diprediksi sistem berlabel negatif (0)
- FP *(False Positive)* adalah data aktual berlabel negatif (0) diprediksi sistem berlabel positif (1)
- FN *(False Negative)* adalah data aktual berlabel positif (1) diprediksi sistem berlabel negatif (0)

4.2 Hasil Uji Coba

Hasil pada pengujian yang menggunakan persentase fitur seperti yang telah disebutkan pada poin 3 ini berupa kelas positif atau negatif pada tiap dokumen, sehingga dapat diketahui ketepatan dari sistem pengklasifikasian ini dengan kelas aktual yang telah diberikan sebelumnya. Hasil ditampilkan berdasarkan masing-masing persentase jumlah fitur yang digunakan. Bersumber pada data hasil tersebut dapat diketahui jumlah kesesuaian antara kelas aktual dan kelas prediksi sistem, sehingga dapat dilakukan perhitungan untuk evaluasi kinerja sistem menggunakan *confussion matrix*.

Tabel 4.2 menyajikan jumlah nilai hasil pengujian menggunakan 10% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 2 Hasil Pengujian Menggunakan 10% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	1	0	0	1	0
8	0	0	0	1	0	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			203	217	71	109

Accuracy : $\frac{203+217}{600} . 100\% = 70\%$

Precision : $\frac{203}{203+71} . 100\% = 74{,}08\%$

Recall : $\frac{203}{203+109} . 100\% = 65{,}06\%$

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall} = 69{,}28\%$

Tabel 4.3 menyajikan jumlah nilai hasil pengujian menggunakan 20% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan

persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 3 Hasil Pengujian Menggunakan 20% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	1	0	0	1	0
8	0	0	0	1	0	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			205	218	70	107

Accuracy : $\frac{205+218}{600} . 100\% = 70{,}5\%$

Precision : $\frac{205}{205+70} . 100\% = 74{,}54\%$

Recall : $\frac{205}{205+107} . 100\% = 65{,}70\%$

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall} = 69{,}84\%$

Tabel 4.4 menyajikan jumlah nilai hasil pengujian menggunakan 30% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 4 Hasil Pengujian Menggunakan 30% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	1	0	0	1	0
8	0	1	0	0	1	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			204	216	72	108

Accuracy : $\frac{204+216}{600} . 100\% = 70\%$

Precision : $\frac{204}{204+72} . 100\% = 73{,}91\%$

Recall : $\frac{204}{204+108} . 100\% = 65{,}38\%$

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall} = 69{,}38\%$

Tabel 4.5 menyajikan jumlah nilai hasil pengujian menggunakan 40% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 5 Hasil Pengujian Menggunakan 40% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0

NO	Aktual	Prediksi	TP	TN	FP	FN
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	1	0	0	1	0
8	0	1	0	0	1	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			202	216	72	110

Accuracy : $\frac{202+216}{600}$. 100% = 69,6%

Precision : $\frac{202}{202+72}$. 100% = 73,72%

Recall : $\frac{202}{202+110}$. 100% = 64,74%

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall}$ = 68,94%

Tabel 4.6 menyajikan jumlah nilai hasil pengujian menggunakan 50% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 6 Hasil Pengujian Menggunakan 50% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	0	0	1	0	0

NO	Aktual	Prediksi	TP	TN	FP	FN
8	0	0	0	1	0	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			202	218	70	110

Accuracy : $\frac{202+218}{600}$. 100% = 70%

Precision : $\frac{202}{202+70}$. 100% = 74,26%

Recall : $\frac{202}{202+110}$. 100% = 64,74%

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall}$ = 69,17%

Tabel 4.7 menyajikan jumlah nilai hasil pengujian menggunakan 60% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 7 Hasil Pengujian Menggunakan 60% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	0	0	1	0	0
8	0	0	0	1	0	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0

NO	Aktual	Prediksi	TP	TN	FP	FN
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			202	218	70	110

Accuracy : $\frac{202+218}{600}$. 100% = 70%

Precision : $\frac{202}{202+70}$. 100% = 74,26%

Recall : $\frac{202}{202+110}$. 100% = 64,74%

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall}$ = 69,17%

Tabel 4.8 menyajikan jumlah nilai hasil pengujian menggunakan 70% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 8 Hasil Pengujian Menggunakan 70% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	0	0	1	0	0
8	0	0	0	1	0	0
9	0	0	0	1	0	0
10	1	1	1	0	0	0
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0

16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			209	215	73	103

$Accuracy : \frac{209+215}{600} . 100\% = 70{,}6\%$

$Precision : \frac{209}{209+73} . 100\% = 74{,}11\%$

$Recall : \frac{209}{209+103} . 100\% = 66{,}98\%$

$F\text{-}measure : 2 \times \frac{Precision \times Recall}{Precision + Recall} = 70{,}37\%$

Tabel 4.9 menyajikan jumlah nilai hasil pengujian menggunakan 80% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 9 Hasil Pengujian Menggunakan 80% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	0	0	1	0	0
8	0	0	0	1	0	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0

…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			201	212	76	111

Accuracy : $\frac{201+212}{600}$. 100% = 68,8%

Precision : $\frac{201}{201+76}$. 100% = 72,56%

Recall : $\frac{201}{201+111}$. 100% = 64,42%

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall}$ = 68,25%

Tabel 4.10 menyajikan jumlah nilai hasil pengujian menggunakan 90% fitur. Nilai yang dituliskan tersebut selanjutnya masing-masing dapat dihitung dengan persamaan (4.1), (4.2), (4.3), dan (4.4) untuk mengukur persentase kinerja sistem yang dibangun.

Tabel 4. 10 Hasil Pengujian Menggunakan 90% Fitur

NO	Aktual	Prediksi	TP	TN	FP	FN
1	0	0	0	1	0	0
2	0	1	0	0	1	0
3	0	0	0	1	0	0
4	0	0	0	1	0	0
5	1	1	1	0	0	0
6	0	1	0	0	1	0
7	0	0	0	1	0	0
8	0	0	0	1	0	0
9	0	1	0	0	1	0
10	1	0	0	0	0	1
11	1	1	1	0	0	0
12	0	1	0	0	1	0
13	0	1	0	0	1	0
14	0	0	0	1	0	0
15	1	1	1	0	0	0
16	0	0	0	1	0	0
17	0	1	0	0	1	0
18	0	0	0	1	0	0
19	0	1	0	0	1	0
20	1	1	1	0	0	0
…	…	…	…	…	…	…
600	1	1	1	0	0	0
Total			202	210	78	110

Accuracy : $\frac{202+210}{600}$. 100% = 68,6%

Precision : $\frac{202}{202+78}$. 100% = 72,14%

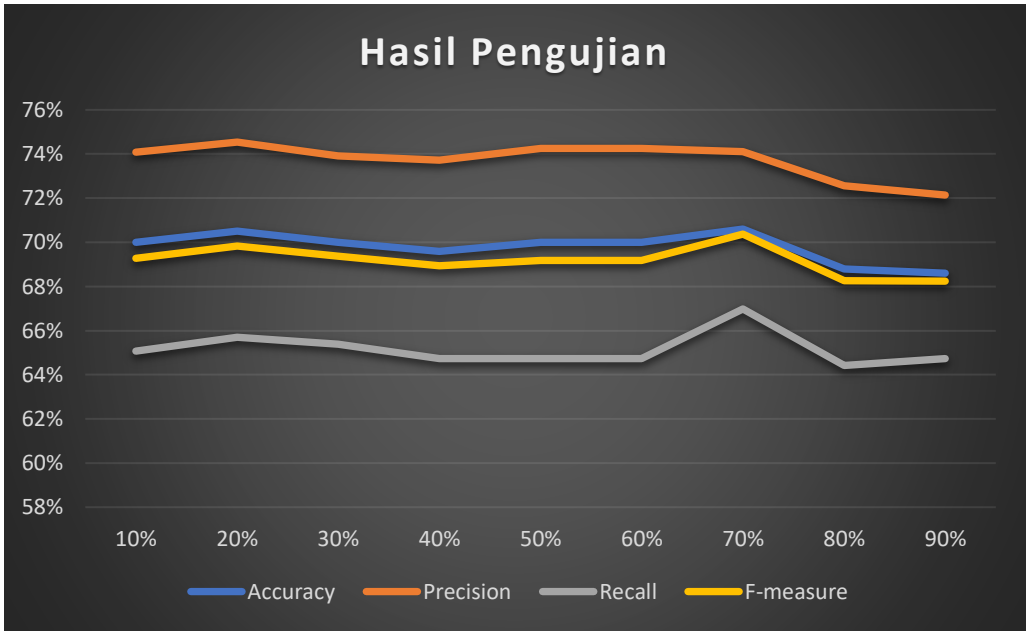
Recall : $\frac{202}{202+110}$. 100% = 64,74%

F-measure : $2 \times \frac{Precision \times Recall}{Precision + Recall}$ = 68,24%

Hasil perhitungan *accuracy*, *precision*, *recall*, dan *f-measure* untuk setiap persentase fitur disajikan pada Tabel 4.11 dan Gambar 4.1. Berdasarkan data tersebut dapat diketahui tingkat persentase kinerja sistem yang telah dibangun.

Tabel 4. 11 Hasil Pengujian Confussion Matrix

Jumlah Fitur	***Accuracy***	***Precision***	***Recall***	***F-Measure***
10%	70%	74,08%	65,06%	69,28%
20%	70,5%	74,54%	65,70%	69,84%
30%	70%	73,91%	65,38%	69,38%
40%	69,6%	73,72%	64,74%	68,94%
50%	70%	74,26%	64,74%	69,17%
60%	70%	74,26%	64,74%	69,17%
70%	70,6%	74,11%	66,98%	70,37%
80%	68,8%	72,56%	64,42%	68,25%
90%	68,6%	72,14%	64,74%	68,24%



Gambar 4. 1 Hasil Pengujian

4.3 Pembahasan

Berdasarkan pada skenario dan hasil pengujian yang dipaparkan pada sub bab sebelumnya, dapat diketahui jika tiap-tiap langkah yang dilakukan pada pengujian ini berpengaruh pada hasil klasifikasi. Tahapan *preprocessing* berpengaruh pada kata yang akan diolah pada pengujian, semakin baik *preprocessing* yang diterapkan akan menghasilkan *dataset* yang lebih baik untuk proses pengujian. Tahapan seleksi fitur menggunakan metode *TF-IDF* juga berpengaruh pada hasil pengklasifikasian, disebabkan pada tahapan ini dilakukan pengurutan fitur berdasarkan besaran nilai bobot sehingga proses ini dapat mengurangi jumlah fitur yang digunakan. Penggunaan jumlah fitur yang tepat dan sesuai dengan konteks yang dibahas dapat mempengaruhi hasil klasifikasi.

Berdasarkan pada Tabel 4.11 dapat diketahui jika proses pengklasifkasian *tweets* menggunakan metode *Gaussian Naïve Bayes* menghasilkan kinerja sistem terbaik pada penggunaan fitur sebesar 70% dengan nilai *accuracy* sebesar 70,6%,

precision sebesar 74,11%, *recall* sebesar 66,98%, dan nilai *f-measure* sebesar 70,37%. Nilai *accuracy* sebesar 70,6% menandakan jika persentase keakuratan sistem dalam memprediksi data aktual dapat dikategorikan baik. Nilai *precision* sebesar 74,11% juga mengindikasikan jika persentase prediksi benar pada keseluruhan hasil prediksi dapat dikategorikan baik. Sedangkan nilai *recall* sebesar 66,98% menandakan jika persentase dari hasil prediksi terhadap data aktual dapat dikategorikan cukup. Sehingga berdasarkan data tersebut dapat dikatakan jika *harmonic mean* dari nilai *recall* dan *precision* dapat dikategorikan baik dengan nilai *f-measure* sebesar 70,37%. Berdasarkan data hasil evaluasi kinerja sistem yang telah disebutkan, dapat diketahui jika data uji yang digunakan lebih banyak berlabel positif, sehingga sistem lebih banyak mendeteksi nilai *FN (False Negative)* dibandingkan dengan nilai *FP (False Positive)*. Kondisi tersebut dapat diketahui pada nilai *precision* yang lebih tinggi dan terpaut jauh dari nilai *recall.*

Hasil perhitungan evaluasi kinerja sistem yang didapat tersebut disebabkan penggunaan persentase fitur yang tepat dan sesuai antara data latih dan data uji sehingga dapat mempengaruhi hasil klasifikasi lebih baik. Penggunaan fitur yang terlalu sedikit menyebabkan banyak fitur penting yang terlewat dan pada akhirnya tidak masuk pada proses pengujian, sedangkan penggunaan fitur terlalu banyak akan menyebabkan banyak fitur yang tidak tepat dan tidak sesuai dengan konteks pengujian akan masuk pada proses pengujian. Kedua hal tersebut dapat mempengaruhi hasil klasifikasi dan kinerja sistem yang dibangun.

Pengujian ini mendapati kebanyakan kesalahan atau *error* dikarenakan pada proses *preprocessing* gagal memproses data dengan kata yang dituliskan tanpa

spasi seperti : “benerbener, efeknegatif, euforiaberlebihan”, perulangan huruf yang sama pada satu kata seperti : “nyebelinnn, kerenn, capeee”, penggunaan bahasa indonesia yang bercampur dengan bahasa lain seperti : “ppkm bikin kangen dolan rumah”, dan *typo* dalam penulisan seperti : “nheriii, nprmal, posutif”. Kejadian tersebut terjadi dikarenakan *dataset* yang digunakan bersumber dari data *tweets* yang sifat penulisannya menggunakan bahasa informal dan bebas.

Pada proses seleksi fitur juga mendapati beberapa kata penting yang seharusnya dapat menunjang kinerja sistem lebih baik tetapi pada perangkingan fitur mendapati rangking yang rendah sehingga hal ini dapat mempengaruhi hasil pengklasifikasian dikarenakan pada proses pengujian membutuhkan persentase penggunaan fitur. Hal ini menjadi salah satu penyebab penggunaan fitur terbaik berada pada penggunaan fitur sebesar 70%.

Sistem pengklasifikasian yang dibangun ini diharapkan dapat memberikan salah satu referensi dalam melakukan validasi terhadap berita dan juga komentar orang lain terkait hal yang penting dalam masyarakat seperti kebijakan pemerintah Indonesia dalam penanggulangan pandemi *COVID-19.* Adanya kebijakan pemerintah untuk menanggulangi pandemi tersebut mungkin tidak sesuai dengan apa yang kita harapkan, akan tetapi dalam Islam kita diajarkan untuk selalu menyampaikan argumen kita dengan cara yang baik. Hal tersebut sesuai dengan yang telah difirmankan pada Surat An-Nahl ayat 125 :

ادْعُ إِلَىٰ سَبِيلِ رَبِّكَ بِالْحِكْمَةِ وَالْمَوْعِظَةِ الْحَسَنَةِ ۖ وَجَادِلْهُمْ بِالَّتِي هِيَ أَحْسَنُ ۚ إِنَّ رَبَّكَ هُوَ أَعْلَمُ بِمَنْ ضَلَّ عَنْ سَبِيلِهِ ۖ وَهُوَ أَعْلَمُ بِالْمُهْتَدِينَ

"Serulah (manusia) kepada jalan Tuhan-mu dengan hikmah dan pelajaran yang baik dan bantahlah mereka dengan cara yang baik. Sesungguhnya Tuhanmu Dialah yang lebih mengetahui tentang siapa yang tersesat dari jalan-Nya dan Dialah yang lebih mengetahui orang-orang yang mendapat petunjuk" (QS. An-Nahl : 125)

Berdasarkan pada *Tafsir Al-Muyassar* karya Al-Qarni (2008) Surat An-Nahl ayat 125 ditafsirkan sebagai berikut, serukan (wahai Rasul) melalui kamu dan pengikutmu ke agama Tuhanmu dan ke jalan lurus-Nya di jalan kebijaksanaan yang telah diwahyukan Allah kepadamu dalam Al-Qur'an dan Sunnah. Dan katakanlah pada orang-orang dengan cara yang sesuai dengan keinginan diantara mereka, lalu tegurlah mereka dengan cara yang baik sehingga mampu mendorong mereka mencintai kebaikan dan menjauhi kejahatan. Dan berdebatlah dengan mereka dengan cara yang terbaik, argumentatif, bijaksana, dan lemah lembut. Karena sesungguhnya engkau tidak memiliki kewajiban selain untuk menyampaikan, dan memang telah menyampaikan, karena hanya Allah yang dapat membimbing mereka. Dia lebih mengetahui siapa yang tersesat dan Dia lebih mengetahui siapa yang akan mendapat petunjuk.

Berdasarkan tafsiran tersebut, dapat diambil kesimpulan jika kita sebagai muslim diperintahkan untuk selalu menyampaikan kebaikan dengan cara yang baik. Sedangkan ketika menemui suatu perbedaan pendapat terkait sesuatu maka sampaikan pendapat tersebut dengan cara yang terbaik. Hal ini sesuai dengan keadaan yang dihadapi di masa sekarang, kita seringkali menemui perbedaan pendapat terkait sesuatu, tidak terkecuali dengan kebijakan yang diterapkan pemerintah. Oleh karena itu, berdasarkan ayat tersebut kita diperintahkan untuk

selalu menyampaikan pendapat tersebut dengan cara yang baik, termasuk ketika kita berkomentar pada sosial media.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Proses pengklasifkasian *tweets* memanfaatkan algoritma *Gaussian Naïve Bayes* ini menghasilkan kinerja sistem terbaik pada penggunaan fitur sebesar 70% dengan nilai *accuracy* sebesar 70,6%, *precision* sebesar 74,11%, *recall* sebesar 66,98%, dan nilai *f-measure* sebesar 70,37%. Hasil ini didapat setelah melakukan berbagai macam proses pengujian berdasarkan skenario pengujian yang telah disusun sebelumnya.

Berdasarkan pengujian yang telah dilakukan, dapat diketahui jika proses *preprocessing* dan seleksi fitur menggunakan metode *TF-IDF* berpengaruh pada proses pengklasifikasian dan hasilnya. Berdasarkan pernyataan tersebut dapat disimpulkan jika penggunaan *preprocessing* yang tepat akan mempengaruhi kata yang nantinya akan diproses, dan penggunaan fitur yang tepat dan sesuai dapat mempengaruhi hasil pengklasifikasian menggunakan *Gaussian Naïve Bayes* ini.

5.2 Saran

Penelitian tentang proses pengklasifkasian *tweets* menggunakan metode *Gaussian Naïve Bayes* ini masih memiliki beberapa kekurangan, sehingga diharapkan pada penelitian selanjutnya dapat menghasilkan data klasifikasi yang lebih akurat. Berdasarkan pernyataan tersebut, penulis mencoba memberikan saran untuk perkembangan penelitian di masa depan sebagai berikut :

1. Penambahan total *dataset* sehingga menambah jumlah fitur kata pada perbendaharaan kata pada data training, sehingga diharapkan hal ini dapat menambah kemampuan sistem dalam mengklasifikasi.

2. Mengoptimalkan kinerja *preprocessing* dalaam mengenali kata-kata dalam bahasa informal dengan cara menambahkan proses tambahan pada tahapan *cleansing* dan *normalization* untuk pengenalan kata-kata informal (kata singkatan, kata gaul, dan kata dalam bahasa asing) yang umum.

3. Memanfaatkan metode lain pada tahapan seleksi fitur seperti *Hybrid Particle Swarm, Information Gain, F-score dan Rough Set, Chi-Square* dan sebagainya.

DAFTAR PUSTAKA

Agung, A. A. P. (2016). *Implementasi Text Summarization Menggunakan Metode Vector Space Model pada Artikel Berita Berbahasa Indonesia* [Skripsi]. Universitas Komputer Indonesia.

Alita, D., & Rahman, A. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputasi : Ilmu Komputer Unila Publishing Network*, *8*(No : 2).

Al-Qarni, A. (2008). *Tafsir Al-Muyassar* (Terj. Tim Penerjemah Qisthi Press). Qisthi Press.

ButarButar, T. M. E. Y., Fauzi, M. A., & Indriati. (2019). Penentuan Rating Review Film Menggunakan Metode Multinomial Naïve Bayes Classifier dengan Feature Selection berbasis Chi-Square dan Galavotti-Sebastiani-Simi Coefficient. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *Vol. 3*(No. 1).

Delizo, J. P. D., Abisado, M. B., & De Los Trinos, Ma. I. P. (2020). Philippine Twitter Sentiments during Covid-19 Pandemic using Multinomial Naïve-Bayes. *International Journal of Advanced Trends in Computer Science and Engineering*, *Vol. 9*(No. 1.3).

Dulhare, U. N. (2018). Prediction System for Heart Disease Using Naive Bayes and Particle Swarm Optimization. *Biomedical Research*, *Vol. 29*(12).

Febrianti, A. V. (2020). *Analisis Sentimen Data Ulasan Pengunjung Objek Wisata Lawang Sewu Kota Semarang pada Situs Tripadvisor.* UNIVERSITAS NEGERI SEMARANG.

Friedman, C., Rindflesch, T. C., & Corn, M. (2013). Natural Language Processing: State of the Art and Prospects for Significant Progress, a Workshop Sponsored by the National Library of Medicine. *Journal of Biomedical Informatics*, *46*, 765–773.

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques 3rd Edition* (Third Edition). The Morgan Kaufmann.

Ipmawati, J., Kusrini, & Luthfi, E. T. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, *6(1)*, 28–36. http://dx.doi.org/10.2311/ijns.v6i1.1444

Kamel, H., Abdulah, D., & M.Al-Tuwaijari, J. (2019). *Cancer Classification Using Gaussian Naive Bayes Algorithm*. 165–170. https://doi.org/10.1109/IEC47844.2019.8950650

Kurniawan, B., Effendi, S., & Salim Sitompul, O. (2012). Klasifikasi Konten Berita Dengan Metode Text Mining. *JURNAL DUNIA TEKNOLOGI INFORMASI*, *Vol. 1*, *No. 1*, 14–19.

Lee, L., & Pang, B. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *1–2*. https://doi.org/10.1561/1500000011

Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing, Second Edition*.

Masuda, K., Matsuzaki, T., & Tsujii, J. (2011). Semantic Search based on the Online Integration of NLP Techniques. *Procedia - Social and Behavioral Sciences*, *27*, 281–290.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*. https://doi.org/10.1016/j.asej.2014.04.011

Melita, R., Amrizal, V., Suseno, H. B., & Dirjam, T. (2018). Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) dan Cosine Similarity pada Sistem Temu Kembali Informasi untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Syarah Umdatil Ahkam). *JURNAL TEKNIK INFORMATIKA*, *Vol. 11*(No. 2).

Mukherjee, S., & Sharma, N. (2012). Intrusion Detection using Naive Bayes Classifier with Feature Reduction. *Procedia Technology*, *4*, 119–128.

Mustaqhfiri, M., Abidin, Z., & Kusumawati, R. (2012). Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance. *MATICS*. https://doi.org/DOI:10.18860/mat.v0i0.1578

Rodiyansyah, S. F., & Winarko, E. (2013). Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification. *Indonesian Journal of Computing and Cybernetics Systems*.

Saidah, S., & Mayary, J. (2020). Analisis Sentimen Pengguna Twitter Terhadap Dompet Elektronik dengan Metode Lexicon Based dan K–Nearest Neighbour. *Jurnal Ilmiah Informatika Komputer*, *Vol. 25*(1).

Salmi, N., & Rustam, Z. (2019). Naïve Bayes Classifier Models for Predicting the Colon Cancer. *IOP Conference Series: Materials Science and Engineering*, *Vol. 546*(5).

Saputro, I. W., & Sari, B. W. (2019). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa. *Citec Journal*, *Vol. 6*.

Sidiq, Y. N. S., Fathonah, N. S., & Riza, N. (2020). *Metode Klasifikasi Menentukan Kenaikan Level UKM Bandung Timur dengan Algoritma Naïve Bayes Pada Sistem JURAGAN Berbasis Komunitas*. Kreatif Industri Nusantara.

Styawati, & Mustofa, K. (2019). A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems), Vol. 13*(No. 3).

Sumartini, N. W. (2011). *Text Mining Classifier dengan Metode Naïve Bayes dan Support Vector Machines untuk Sentiment Analysis* [Thesis]. Universitas Udayana.

Ting, S. L., Ip, W. H., & Tsang, A. H. C. (2014). Is Naïve Bayes a Good Classifier for Document Classification ? *International Journal of Software Engineering and Its Applications 5, 3*.

Valdiviezo-Diaz, P., Ortega, F., Cobos, E., & Lara-Cabrera, R. (2019). A Collaborative Filtering Approach Based on Naïve Bayes Classifier. *IEEE Access, 7*.

Villavicencio, C., Jerison Macrohon, J., Inbaraj, X. A., Jeng, J.-H., & Hsieh, J.-G. (2021). Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes. *Information*, *12*(5).

Watratan, A. F., Puspita. B, A., & Moeis, D. (2020). Implementasi Algoritma Naive Bayes untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia. *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, *1*(1), 7–14.

WHO. (2019). Pertanyaan dan Jawaban Terkait Coronavirus. *Pertanyaan Dan Jawaban Terkait Coronavirus*. https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-for-public

Yulita, W., Nugroho, E. D., & Algifari, M. H. (2021). Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier. *JURNAL DATA MINING DAN SISTEM INFORMASI, Vol. 2*(No. 2).

Zaman, B., & Winarko, E. (2011). Analisis Fitur Kalimat untuk Peringkas Teks Otomatis pada Bahasa Indonesia. *Indonesian Journal of Computing and Cybernetics Systems*, *5*(2).

Zhang, W., Yoshida, T., & Tang, X. (2011). A Comparative Study of TF-IDF, LSI and Multi-Words for Text Classification. *Expert Systems with Applications*, *38*(3), 2758–2765.

LAMPIRAN

LAMPIRAN I

Hasil Klasifikasi Sentimen

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	0	0	0	0	0	0	0	0	0	0
2	0	1	1	1	1	1	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1	1
6	0	1	1	1	1	1	1	1	1	1
7	0	1	1	1	1	0	0	0	0	0
8	0	0	0	1	1	0	0	0	0	0
9	0	1	1	1	1	1	1	1	1	1
10	1	0	0	0	0	0	0	0	0	0
11	1	1	1	1	1	1	1	1	1	1
12	0	1	1	1	1	1	1	1	1	1
13	0	1	1	1	1	1	1	1	1	1
14	0	0	0	0	0	0	0	0	0	0
15	1	1	1	1	1	1	1	1	1	1
16	0	0	0	0	0	0	0	0	0	0
17	0	1	1	1	1	1	1	1	1	1
18	0	0	0	0	0	0	0	0	0	0
19	0	1	1	1	1	1	1	1	1	1
20	1	1	1	1	1	1	1	1	1	1
21	0	0	0	0	0	0	0	0	0	0
22	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1
24	1	1	1	1	1	1	1	1	1	1
25	0	0	0	0	0	0	0	0	0	0
26	1	1	1	1	1	1	1	1	1	1
27	1	1	1	1	1	1	1	1	1	1
28	0	1	1	1	1	1	1	1	1	1
29	0	0	0	0	0	0	0	0	0	0
30	0	1	1	1	1	1	1	1	1	1
31	0	1	1	1	1	1	1	1	1	1
32	0	1	1	1	1	1	1	1	1	1
33	1	1	1	1	1	1	1	1	1	1
34	1	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
36	1	1	1	1	1	1	1	1	1	1
37	1	1	1	1	1	1	1	1	1	1
38	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0
41	1	1	1	1	1	1	1	1	1	1
42	1	1	1	1	1	1	1	1	1	1
43	0	0	0	0	0	0	0	0	0	0
44	0	1	1	1	1	1	1	1	1	1
45	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0
48	0	1	1	1	1	1	1	1	1	1
49	0	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0	0
51	0	1	1	1	1	1	1	1	1	1
52	1	1	1	1	1	1	1	1	1	1
53	0	1	1	1	1	1	1	1	1	1
54	1	1	1	1	1	1	1	1	1	1
55	1	1	1	1	1	1	1	1	1	1
56	0	0	0	0	0	0	0	0	0	0
57	0	0	0	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0
62	1	1	1	1	1	1	1	1	1	1
63	1	1	1	1	1	1	1	1	1	1
64	0	1	1	1	1	1	1	1	1	1
65	1	1	1	1	1	1	1	1	1	1
66	1	1	1	1	1	1	1	1	1	1
67	1	1	1	1	1	1	1	1	1	1
68	0	1	1	1	1	1	1	1	1	1
69	0	1	1	1	1	1	1	1	1	1
70	0	1	1	1	1	1	1	1	1	1
71	0	0	0	0	0	0	0	0	0	0
72	0	0	0	0	0	0	0	0	0	0
73	0	0	0	0	0	0	0	0	0	0
74	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	0	0	0	0	0	0
76	0	0	0	0	0	0	0	0	0	0

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
77	0	0	0	0	0	0	0	0	0	0
78	1	1	1	1	1	1	1	1	1	1
79	1	1	1	1	1	1	1	1	1	1
80	1	1	1	1	1	1	1	1	1	1
81	1	1	1	1	1	1	1	1	1	1
82	1	1	1	1	1	1	1	1	1	1
83	0	1	1	1	1	1	1	1	1	1
84	1	1	1	1	1	1	1	1	1	1
85	1	1	1	1	1	1	1	1	1	1
86	1	1	1	1	1	1	1	1	1	1
87	0	0	0	0	0	0	0	0	0	0
88	1	1	1	1	1	1	1	1	1	1
89	1	1	1	1	1	1	1	1	1	1
90	0	1	1	1	1	1	1	1	1	1
91	1	1	1	1	1	1	1	1	1	1
92	1	1	1	1	1	1	1	1	1	1
93	1	1	1	1	1	1	1	1	1	1
94	1	1	1	1	1	1	1	1	1	1
95	1	1	1	1	1	1	1	1	1	1
96	0	1	1	1	1	1	1	1	1	1
97	1	1	1	1	1	1	1	1	1	1
98	1	1	1	1	1	1	1	1	1	1
99	1	0	0	0	0	0	0	0	0	0
100	1	1	1	1	1	1	1	1	1	1
101	0	1	1	1	1	1	1	1	1	1
102	1	1	1	1	1	1	1	1	1	1
103	1	1	1	1	1	1	1	1	1	1
104	1	1	1	1	1	1	1	1	1	1
105	0	1	1	1	1	1	1	1	1	1
106	1	1	1	1	1	1	1	1	1	1
107	1	1	1	1	1	1	1	1	1	1
108	0	0	0	0	0	0	0	0	0	0
109	0	0	0	0	0	0	0	0	0	0
110	0	0	0	0	0	0	0	0	0	0
111	0	0	0	0	0	0	0	0	0	0
112	1	1	1	1	1	1	1	1	1	1
113	1	1	1	1	1	1	1	1	1	1
114	1	0	0	0	0	0	0	0	0	0
115	1	0	0	0	0	0	0	0	0	0
116	1	1	1	1	1	1	1	1	1	1
117	1	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
118	1	1	1	1	1	1	1	1	1	1
119	1	1	1	1	1	1	1	1	1	1
120	1	1	1	1	1	1	1	1	1	1
121	1	1	1	1	1	1	1	1	1	1
122	1	1	1	1	1	1	1	1	1	1
123	1	1	1	1	1	1	1	1	1	1
124	1	1	1	1	1	1	1	1	1	1
125	1	1	1	1	1	1	1	1	1	1
126	1	1	1	1	1	1	1	1	1	1
127	1	1	1	1	1	1	1	1	1	1
128	1	1	1	1	1	1	1	1	1	1
129	1	1	1	1	1	1	1	1	1	1
130	1	0	0	0	0	0	0	0	0	0
131	1	1	1	1	1	1	1	1	1	1
132	1	0	0	0	0	0	0	0	0	0
133	1	1	1	1	1	1	1	1	1	1
134	1	1	1	1	1	1	1	1	1	1
135	0	1	1	1	1	1	1	1	1	1
136	1	1	1	1	1	1	1	1	1	1
137	1	1	1	1	1	1	1	1	1	1
138	1	1	1	1	1	1	1	1	1	1
139	1	1	1	1	1	1	1	1	1	1
140	1	1	1	1	1	1	1	1	1	1
141	1	1	1	1	1	1	1	1	1	1
142	1	1	1	1	1	1	1	1	1	1
143	0	1	1	1	1	1	1	1	1	1
144	1	1	1	1	1	1	1	1	1	1
145	1	1	1	1	1	1	1	1	1	1
146	1	1	1	1	1	1	1	1	1	1
147	1	1	1	1	1	1	1	1	1	1
148	1	1	1	1	1	1	1	1	1	1
149	1	0	0	0	0	0	0	0	0	0
150	1	1	1	1	1	1	1	1	1	1
151	1	0	0	0	0	0	0	0	0	0
152	1	1	1	1	1	1	1	1	1	1
153	1	1	1	1	1	1	1	1	1	1
154	1	1	1	1	1	1	1	1	1	1
155	0	0	0	0	0	0	0	0	0	0
156	1	1	1	1	1	1	1	1	1	1
157	1	1	1	1	1	1	1	1	1	1
158	0	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
159	1	0	0	0	0	0	0	0	0	0
160	1	1	1	1	1	1	1	1	1	1
161	1	1	1	1	1	1	1	1	1	1
162	1	1	1	1	1	1	1	1	1	1
163	1	1	1	1	1	1	1	1	1	1
164	1	1	1	1	1	1	1	1	1	1
165	1	1	1	1	1	1	1	1	1	1
166	1	1	1	1	1	1	1	1	1	1
167	1	1	1	1	1	1	1	1	1	1
168	1	1	1	1	1	1	1	1	1	1
169	0	1	1	1	1	1	1	1	1	1
170	1	1	1	1	1	1	1	1	1	1
171	0	0	0	0	0	0	0	1	1	1
172	0	0	0	0	0	0	0	0	0	0
173	0	0	0	0	0	0	0	0	0	0
174	1	1	1	1	1	1	1	1	1	1
175	1	1	1	1	1	1	1	1	1	1
176	1	1	1	1	1	1	1	1	1	1
177	1	1	1	1	1	1	1	1	1	1
178	1	1	1	1	1	1	1	1	1	1
179	1	1	1	1	1	1	1	1	1	1
180	1	1	1	1	1	1	1	1	1	1
181	1	1	1	1	1	1	1	1	1	1
182	0	1	1	1	1	1	1	1	1	1
183	0	0	0	0	0	0	0	0	0	1
184	1	1	1	1	1	1	1	1	1	1
185	1	1	1	1	1	1	1	1	1	1
186	1	1	1	1	1	1	1	1	1	1
187	1	1	1	1	1	1	1	1	1	1
188	1	1	1	1	1	1	1	1	1	1
189	0	0	0	0	0	0	0	0	0	0
190	0	1	1	1	1	1	1	1	1	1
191	0	1	0	0	0	0	0	0	1	1
192	0	0	0	0	0	0	0	0	0	0
193	0	0	0	0	0	0	0	0	0	0
194	1	1	1	1	1	1	1	1	1	1
195	1	0	0	0	0	0	0	0	0	0
196	1	1	1	1	1	1	1	1	1	1
197	1	1	1	1	1	1	1	1	1	1
198	1	1	1	1	1	1	1	1	1	1
199	1	0	0	0	0	0	0	0	0	0

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
200	1	1	1	1	1	1	1	1	1	1
201	0	0	0	0	0	0	0	0	0	0
202	0	0	0	0	0	0	0	0	0	0
203	0	0	0	0	0	0	0	0	0	0
204	0	0	0	0	0	0	0	0	0	0
205	0	0	0	0	0	0	0	0	0	0
206	1	0	0	0	0	0	0	0	0	0
207	1	1	1	1	1	1	1	1	1	1
208	1	0	0	0	0	0	0	0	0	0
209	0	0	0	0	0	0	0	0	0	0
210	0	0	0	0	0	0	0	0	0	0
211	0	0	0	0	0	0	0	0	0	0
212	0	0	0	0	0	0	0	0	0	0
213	1	1	1	1	1	1	1	1	1	1
214	1	1	1	1	1	1	1	1	1	1
215	1	1	1	1	1	1	1	1	1	1
216	1	0	0	0	0	0	0	0	0	0
217	1	1	1	1	1	1	1	1	1	1
218	0	1	1	1	1	1	1	1	1	1
219	0	0	0	0	0	0	0	0	0	0
220	1	1	1	1	1	1	1	1	1	1
221	1	0	0	0	0	0	0	0	0	0
222	1	1	1	1	1	1	1	1	1	1
223	1	1	1	1	1	1	1	1	1	1
224	1	1	1	1	1	1	1	1	1	1
225	1	1	1	1	1	1	1	1	1	1
226	0	1	1	1	1	1	1	1	1	1
227	0	0	0	0	0	0	0	0	0	0
228	0	0	0	0	0	0	0	0	0	0
229	0	0	0	0	0	0	0	0	0	0
230	0	0	0	0	0	0	0	0	0	0
231	0	0	0	0	0	0	0	0	0	0
232	0	0	0	0	0	0	0	0	0	0
233	0	0	0	0	0	0	0	0	0	0
234	0	0	0	0	0	0	0	0	0	0
235	0	0	0	0	0	0	0	0	0	0
236	0	0	0	0	0	0	0	0	0	0
237	0	0	0	0	0	0	0	0	0	0
238	0	0	0	0	0	0	0	0	0	0
239	1	1	1	1	1	1	1	1	1	1
240	0	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
241	1	1	1	1	1	1	1	1	1	1
242	1	0	0	0	0	0	0	0	0	0
243	0	0	0	0	0	0	0	0	0	0
244	0	0	0	0	0	0	0	0	0	0
245	1	0	0	0	0	0	0	0	0	0
246	0	1	1	1	1	1	1	1	1	1
247	0	0	0	0	0	0	0	0	0	0
248	0	0	0	0	0	0	0	0	0	0
249	1	0	0	0	0	0	0	0	0	0
250	1	0	0	0	0	0	0	0	0	0
251	1	0	0	0	0	0	0	0	0	0
252	1	0	0	0	0	0	0	0	0	0
253	0	0	0	0	0	0	0	0	0	0
254	0	0	0	0	0	0	0	0	0	0
255	1	0	0	0	0	0	0	0	0	0
256	0	0	0	0	0	0	0	0	0	0
257	1	0	0	0	0	0	0	0	0	0
258	0	0	0	0	0	0	0	0	0	0
259	0	0	0	0	0	0	0	0	0	0
260	0	0	0	0	0	0	0	0	0	0
261	0	0	0	0	0	0	0	0	0	0
262	0	0	0	0	0	0	0	0	0	0
263	1	0	0	0	0	0	0	0	0	0
264	1	0	0	0	0	0	0	0	0	0
265	0	0	0	0	0	0	0	0	0	0
266	1	0	0	0	0	0	0	0	0	0
267	0	0	0	0	0	0	0	0	0	0
268	0	0	0	0	0	0	0	0	0	0
269	0	0	0	0	0	0	0	0	0	0
270	1	0	0	0	0	0	0	0	0	0
271	1	0	0	0	0	0	0	0	0	0
272	0	0	0	0	0	0	0	0	0	0
273	0	0	0	0	0	0	0	0	0	0
274	1	0	0	0	0	0	0	0	0	0
275	1	0	0	0	0	0	0	0	0	0
276	0	0	0	0	0	0	0	0	0	0
277	0	0	0	0	0	0	0	0	0	0
278	1	0	0	0	0	0	0	0	0	0
279	0	0	0	0	0	0	0	0	0	0
280	1	0	0	0	0	0	0	0	0	0
281	0	0	0	0	0	0	0	0	0	0

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
282	1	0	0	0	0	0	0	0	0	0
283	0	0	0	0	0	0	0	0	0	0
284	1	0	0	0	0	0	0	0	0	0
285	0	0	0	0	0	0	0	0	0	0
286	0	0	0	0	0	0	0	0	0	0
287	0	0	0	0	0	0	0	0	0	0
288	1	0	0	0	0	0	0	0	0	0
289	0	0	0	0	0	0	0	0	0	0
290	0	0	0	0	0	0	0	0	0	0
291	1	0	0	0	0	0	0	0	0	0
292	0	0	0	0	0	0	0	0	0	0
293	1	0	0	0	0	0	0	0	0	0
294	0	0	0	0	0	0	0	0	0	0
295	0	0	0	0	0	0	0	0	0	0
296	1	0	0	0	0	0	0	0	0	0
297	0	0	0	0	0	0	0	0	0	0
298	1	0	0	0	0	0	0	0	0	0
299	1	0	0	0	0	0	0	0	0	0
300	1	0	0	0	0	0	0	0	0	0
301	1	0	0	0	0	0	0	0	0	0
302	0	0	0	0	0	0	0	0	0	0
303	1	0	0	0	0	0	0	0	0	0
304	0	0	0	0	0	0	0	0	0	0
305	0	0	0	0	0	0	0	0	0	0
306	0	0	0	0	0	0	0	0	0	0
307	0	0	0	0	0	0	0	0	1	1
308	1	0	0	0	0	0	0	0	0	0
309	1	0	0	0	0	0	0	0	0	0
310	0	0	0	0	0	0	0	0	1	1
311	0	0	0	0	0	0	0	0	0	0
312	1	0	0	0	0	0	0	0	0	0
313	1	0	0	0	0	0	0	0	0	0
314	1	0	0	0	0	0	0	0	0	0
315	0	0	0	0	0	0	0	0	0	0
316	0	0	0	0	0	0	0	0	0	0
317	0	0	0	0	0	0	0	0	0	0
318	0	0	0	0	0	0	0	0	0	0
319	0	0	0	0	0	0	0	0	0	0
320	0	0	0	0	0	0	0	0	0	0
321	0	0	0	0	0	0	0	0	0	0
322	0	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
323	0	0	0	0	0	0	0	0	1	1
324	0	0	0	0	0	0	0	0	1	1
325	0	0	0	0	0	0	0	0	0	0
326	1	0	0	0	0	0	0	0	0	0
327	1	1	1	1	1	1	1	1	1	1
328	0	0	0	0	0	0	0	0	0	0
329	1	0	0	0	0	0	0	0	0	0
330	0	0	0	0	0	0	0	0	0	0
331	0	0	0	0	0	0	0	0	0	0
332	0	0	0	0	0	0	0	0	0	0
333	0	0	0	0	0	0	0	0	0	0
334	0	0	0	0	0	0	0	0	0	0
335	0	0	0	0	0	0	0	0	0	0
336	0	0	0	0	0	0	0	0	0	0
337	0	0	0	0	0	0	0	0	0	0
338	1	0	0	0	0	0	0	0	0	0
339	0	0	0	0	0	0	0	0	0	1
340	1	0	0	0	0	0	0	0	0	0
341	0	0	0	0	0	0	0	0	0	0
342	0	0	0	0	0	0	0	0	0	0
343	0	0	0	0	0	0	0	0	0	0
344	1	0	0	0	0	0	0	0	0	0
345	1	1	1	1	1	1	1	1	1	1
346	1	1	0	0	0	0	0	0	0	0
347	1	0	0	0	0	0	0	0	0	0
348	0	0	0	0	0	0	0	0	0	0
349	0	0	0	0	0	0	0	0	0	0
350	0	0	0	0	0	0	0	0	0	0
351	0	0	0	0	0	0	0	0	0	0
352	0	0	0	0	0	0	0	0	0	0
353	0	0	0	0	0	0	0	0	0	0
354	0	0	0	0	0	0	0	0	0	0
355	0	0	0	0	0	0	0	0	0	0
356	1	0	0	0	0	0	0	0	0	0
357	1	0	0	0	0	0	0	0	0	0
358	0	0	0	0	0	0	0	0	0	0
359	1	0	0	0	0	0	0	0	0	0
360	0	0	0	0	0	0	0	0	0	0
361	1	0	0	0	0	0	0	0	0	0
362	1	1	1	1	1	1	1	1	1	1
363	1	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
364	0	1	1	1	1	1	1	1	1	1
365	0	0	0	0	0	0	0	0	0	0
366	0	0	0	0	0	0	0	0	0	0
367	0	0	0	0	0	0	0	0	0	0
368	0	0	0	0	0	0	0	0	0	0
369	0	0	0	0	0	0	0	0	0	0
370	0	0	0	0	0	0	0	0	0	0
371	0	0	0	0	0	0	0	0	0	0
372	1	0	0	0	0	0	0	0	0	0
373	1	0	0	0	0	0	0	0	0	0
374	0	0	0	0	0	0	0	0	0	0
375	0	0	0	0	0	0	0	0	0	0
376	1	0	0	0	0	0	0	0	0	0
377	1	0	0	0	0	0	0	0	0	0
378	1	0	0	0	0	0	0	0	0	0
379	1	1	1	1	1	1	1	1	1	1
380	1	0	0	0	0	0	0	0	0	0
381	1	1	1	1	1	1	1	1	1	1
382	0	0	0	0	0	0	0	0	0	0
383	1	0	0	0	0	0	0	0	0	0
384	1	1	1	1	1	1	1	1	1	1
385	1	0	0	0	0	0	0	0	0	0
386	1	1	1	1	1	1	1	1	1	1
387	1	1	1	1	1	1	1	1	1	1
388	1	0	0	0	0	0	0	0	0	0
389	1	0	0	0	0	0	0	0	0	0
390	0	0	0	0	0	0	0	0	0	0
391	1	0	0	0	0	0	0	0	0	0
392	0	0	0	0	0	0	0	0	0	0
393	1	0	0	0	0	0	0	0	0	0
394	0	0	0	0	0	0	0	0	0	0
395	0	0	0	0	0	0	0	0	0	0
396	0	0	0	0	0	0	0	0	0	0
397	0	0	0	0	0	0	0	0	0	0
398	0	0	0	0	0	0	0	0	0	0
399	0	0	0	0	0	0	0	0	0	0
400	0	0	0	0	0	0	0	0	0	0
401	0	0	0	0	0	0	0	0	0	0
402	0	0	0	0	0	0	0	0	0	0
403	0	0	0	0	0	0	0	0	0	0
404	1	0	0	0	0	0	0	0	0	0

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
405	1	0	0	0	0	0	0	0	0	0
406	1	0	0	0	0	0	0	0	0	0
407	1	0	0	0	0	0	0	0	0	0
408	1	0	0	0	0	0	0	0	0	0
409	0	0	0	0	0	0	0	0	0	0
410	1	0	0	0	0	0	0	0	0	0
411	1	0	0	0	0	0	0	0	0	0
412	0	0	0	0	0	0	0	0	0	0
413	0	0	0	0	0	0	0	0	0	0
414	1	0	0	0	0	0	0	0	0	0
415	0	0	0	0	0	0	0	0	0	0
416	1	1	1	1	1	1	1	1	1	1
417	0	0	0	0	0	0	0	0	0	0
418	0	0	0	0	0	0	0	0	0	0
419	0	0	0	0	0	0	0	0	0	0
420	0	0	0	0	0	0	0	0	0	0
421	1	0	0	0	0	0	0	0	0	0
422	0	0	0	0	0	0	0	0	0	0
423	1	0	0	0	0	0	0	0	0	0
424	1	1	1	1	1	1	1	1	1	1
425	1	1	1	1	1	1	1	1	1	1
426	1	0	0	0	0	0	0	0	0	0
427	1	0	0	0	0	0	0	0	0	0
428	1	1	1	0	0	0	0	0	0	0
429	1	0	0	0	0	0	0	0	0	0
430	1	0	0	0	0	0	0	0	0	0
431	1	1	1	1	1	1	1	1	0	1
432	0	0	0	0	0	0	0	0	0	0
433	1	0	0	0	0	0	0	0	0	0
434	0	0	0	0	0	0	0	0	0	0
435	0	0	0	0	0	0	0	0	0	0
436	0	0	0	0	0	0	0	0	0	0
437	0	0	0	0	0	0	0	0	0	0
438	0	0	0	0	0	0	0	0	0	0
439	0	0	0	0	0	0	0	0	0	0
440	0	0	0	0	0	0	0	0	0	0
441	0	0	0	0	0	0	0	0	0	0
442	0	0	0	0	0	0	0	0	0	0
443	0	0	0	0	0	0	0	0	0	0
444	0	0	0	0	0	0	0	0	0	0
445	1	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
446	1	1	1	1	1	1	1	1	1	1
447	1	1	1	1	1	1	1	1	1	1
448	1	0	0	0	0	0	0	0	0	0
449	1	1	1	1	1	1	1	1	1	1
450	0	0	0	0	0	0	0	0	0	0
451	1	0	0	0	0	0	0	0	0	0
452	0	0	0	0	0	0	0	0	0	0
453	0	0	0	0	0	0	0	0	0	0
454	0	0	0	0	0	0	0	0	0	0
455	0	0	0	0	0	0	0	0	0	0
456	0	0	0	0	0	0	0	0	0	0
457	0	0	0	0	0	0	0	0	0	0
458	1	0	0	0	0	0	0	0	0	0
459	1	0	0	0	0	0	0	0	0	0
460	1	0	0	0	0	0	0	0	0	0
461	1	0	0	0	0	0	0	0	0	0
462	1	0	0	0	0	0	0	0	0	0
463	1	0	0	0	0	0	0	0	0	0
464	1	0	0	0	0	0	0	0	0	0
465	1	0	0	0	0	0	0	0	0	0
466	1	1	1	1	1	1	1	1	1	1
467	1	0	0	0	0	0	0	0	0	0
468	1	0	0	0	0	0	0	0	0	0
469	1	0	0	0	0	0	0	0	0	0
470	1	1	1	1	1	1	1	1	1	1
471	1	0	0	0	0	0	0	0	0	0
472	1	0	1	1	1	1	1	1	1	1
473	1	1	1	1	1	1	1	1	1	1
474	0	1	1	1	1	1	1	1	1	1
475	0	0	0	0	0	0	0	0	0	0
476	1	0	0	0	0	0	0	0	0	0
477	1	1	1	1	1	1	1	1	1	1
478	0	0	0	0	0	0	0	0	0	0
479	1	0	0	0	0	0	0	0	0	0
480	1	0	0	0	0	0	0	0	0	0
481	1	1	1	1	1	1	1	1	1	1
482	1	1	1	1	1	1	1	1	1	1
483	1	1	1	1	1	1	1	1	1	1
484	1	1	1	1	1	1	1	1	1	1
485	0	0	0	0	0	0	0	0	0	0
486	1	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
487	0	0	0	0	0	0	0	0	0	0
488	0	1	1	1	1	1	1	1	1	1
489	0	1	1	1	1	1	1	1	1	1
490	0	1	1	1	1	1	1	1	1	1
491	0	0	0	0	0	0	0	0	0	0
492	0	0	0	0	0	0	0	0	0	0
493	0	0	0	0	0	0	0	0	0	0
494	0	0	0	0	0	0	0	0	0	0
495	0	0	0	0	0	0	0	0	0	0
496	0	0	0	0	0	0	0	0	0	0
497	0	0	0	0	0	0	0	0	0	0
498	0	0	0	1	1	1	1	1	1	1
499	0	0	0	0	0	0	0	0	0	0
500	0	0	0	0	0	0	0	0	0	0
501	0	0	0	0	0	0	0	0	0	0
502	1	1	1	1	1	1	1	1	1	1
503	1	1	1	1	1	1	1	1	1	1
504	1	1	1	1	1	1	1	1	1	1
505	1	1	1	1	1	1	1	1	1	1
506	0	0	0	0	0	0	0	0	0	0
507	0	0	0	0	0	0	0	0	0	0
508	0	1	1	1	1	1	1	1	1	1
509	0	1	1	1	1	1	1	1	1	1
510	0	0	0	0	0	0	0	0	0	0
511	1	1	1	1	1	1	1	1	1	1
512	1	1	1	1	1	1	1	1	1	1
513	1	0	0	0	0	0	0	0	0	0
514	1	1	1	1	1	1	1	1	1	1
515	0	0	0	0	0	0	0	0	0	0
516	0	1	1	1	1	1	1	1	1	1
517	1	1	1	1	1	1	1	1	1	1
518	1	1	1	1	1	1	1	1	1	1
519	1	1	1	1	1	1	1	1	1	1
520	1	1	1	1	1	1	1	1	1	1
521	1	1	1	1	1	1	1	1	1	1
522	1	1	1	1	1	1	1	1	1	1
523	1	1	1	1	1	1	1	1	1	1
524	1	1	1	1	1	1	1	1	1	1
525	1	1	1	1	1	1	1	1	1	1
526	1	1	1	1	1	1	1	1	1	1
527	1	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
528	1	1	1	1	1	1	1	1	1	1
529	1	1	1	1	1	1	1	1	1	1
530	0	1	1	1	1	1	1	1	1	1
531	1	1	1	1	1	1	1	1	1	1
532	1	1	1	1	1	1	1	1	1	1
533	1	1	1	1	1	1	1	1	1	1
534	1	1	1	1	1	1	1	1	1	1
535	1	1	1	1	1	1	1	1	1	1
536	1	1	1	1	1	1	1	1	1	1
537	0	1	1	1	1	1	1	1	1	1
538	1	1	1	1	1	1	1	1	1	1
539	1	1	1	1	1	1	1	1	1	1
540	1	1	1	1	1	1	1	1	1	1
541	1	1	1	1	1	1	1	1	1	1
542	1	1	1	1	1	1	1	1	1	1
543	1	1	1	1	1	1	1	1	1	1
544	1	1	1	1	1	1	1	1	1	1
545	1	1	1	1	1	1	1	1	1	1
546	1	1	1	1	1	1	1	1	1	1
547	0	1	1	1	1	1	1	1	1	1
548	0	1	1	1	1	1	1	1	1	1
549	0	0	0	0	0	0	0	0	0	0
550	1	0	1	1	0	0	0	0	0	0
551	1	1	1	1	1	1	1	1	1	1
552	0	1	1	1	1	1	1	1	1	1
553	0	1	1	1	1	1	1	1	1	1
554	0	1	1	1	1	1	1	1	1	1
555	0	1	1	1	1	1	1	1	1	1
556	0	0	0	0	0	0	0	0	0	0
557	1	0	1	1	0	0	0	0	0	0
558	1	1	1	1	1	1	1	1	1	1
559	0	1	1	1	1	1	1	1	1	1
560	0	1	1	1	1	1	1	1	1	1
561	0	1	1	1	1	1	1	1	1	1
562	0	1	1	1	1	1	1	1	1	1
563	1	0	0	0	0	0	0	0	0	0
564	0	1	1	1	1	1	1	1	1	1
565	0	0	0	0	0	0	0	0	0	0
566	0	1	1	1	1	1	1	1	1	1
567	1	1	1	1	1	1	1	1	1	1
568	0	1	1	1	1	1	1	1	1	1

NO	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
569	0	1	1	1	1	1	1	1	1	1
570	1	0	0	0	0	0	0	0	0	0
571	0	1	1	1	1	1	1	1	1	1
572	0	1	1	1	1	1	1	1	1	1
573	1	1	1	1	1	1	1	1	1	1
574	1	1	1	1	1	1	1	1	1	1
575	1	1	1	1	1	1	1	1	1	1
576	1	1	1	1	1	1	1	1	1	1
577	1	1	1	1	1	1	1	1	1	1
578	1	1	1	1	1	1	1	1	1	1
579	1	1	1	1	1	1	1	1	1	1
580	1	1	1	1	1	1	1	1	1	1
581	0	1	1	1	1	1	1	1	1	1
582	0	0	0	0	0	0	0	0	0	0
583	1	1	1	1	1	1	1	1	1	1
584	0	0	0	0	0	0	0	0	0	0
585	0	1	1	1	1	1	1	1	1	1
586	0	1	1	1	1	1	1	1	1	1
587	0	1	1	1	1	1	1	1	1	1
588	1	1	1	1	1	1	1	1	1	1
589	0	1	1	1	1	1	1	1	1	1
590	0	1	1	1	1	1	1	1	1	1
591	1	1	1	1	1	1	1	1	1	1
592	1	1	1	1	1	1	1	1	1	1
593	0	1	1	1	1	1	1	1	1	1
594	0	1	1	1	1	1	1	1	1	1
595	0	0	0	0	0	0	0	0	0	0
596	0	0	0	0	0	0	0	0	0	0
597	0	0	0	0	0	0	0	0	0	0
598	1	1	1	1	1	1	1	1	1	1
599	1	1	1	1	1	1	1	1	1	1
600	1	1	1	1	1	1	1	1	1	1

LAMPIRAN II

Ranking Term

NO	TERM	Bobot Total
1	ppkm	23.24630
2	panjang	7.695258

NO	TERM	Bobot Total
3	vaksinasi	5.774660
4	level	3.715277
5	iya	3.193280
6	moga	3.189600
7	banget	3.103352
8	covid	3.065840
9	bikin	2.894266
10	selesai	2.771697
11	deh	2.760166
12	rumah	2.701884
13	sih	2.592071
14	ga	2.535000
15	psbb	2.304968
16	minggu	2.272988
17	enak	2.216569
18	orang	2.196521
19	besok	2.174348
20	vaksin	2.059499
21	jalan	2.058691
22	plis	2.053001
23	masyarakat	2.037215
24	kangen	1.921792
25	gara	1.855298
26	turun	1.809638
27	jam	1.777430
28	bantu	1.732724
29	tutup	1.731527
30	udahan	1.719808
31	jilid	1.627636
32	ayodong	1.627636
33	rebah	1.627636
34	memuakkannnnn	1.627636
35	tai	1.627636
36	warga	1.600917
37	hari	1.585568
38	capek	1.561056
39	makan	1.501160
40	si	1.481681
41	kelar	1.471901
42	gin	1.469059
43	laku	1.424881
44	nya	1.419355

NO	TERM	Bobot Total
45	bosen	1.415534
46	pelan	1.406782
47	mati	1.389076
48	giat	1.330777
49	daerah	1.312669
50	indonesia	1.309598
51	presiden	1.309598
52	rakyat	1.290061
53	hasil	1.260276
54	biar	1.227270
55	bahagia	1.195765
56	pandemi	1.142129
57	merdeka	1.112463
58	agustus	1.105546
59	lv	1.088109
60	detik	1.085091
61	ppkmpelan	1.085091
62	lebar	1.085091
63	mundur	1.085091
64	bgst	1.085091
65	resah	1.085091
66	fix	1.085091
67	bubar	1.085091
68	hibur	1.085091
69	eeyoooo	1.085091
70	kota	1.064377
71	semenjak	1.046317
72	perintah	1.032134
73	program	1.024268
74	gak	1.013620
75	udh	1.012564
76	allah	1.005426
77	to	0.992849
78	lengkap	0.984748
79	nama	0.957790
80	ayo	0.951923
81	gila	0.949578
82	terap	0.941259
83	lockdown	0.937952
84	pintu	0.930078
85	saffarudin	0.926050
86	kakek	0.926050

NO	TERM	Bobot Total
87	sehat	0.923063
88	reda	0.919098
89	sedia	0.914408
90	buru	0.912340
91	batas	0.899719
92	wkwk	0.889764
93	alhamdulillah	0.886363
94	dikit	0.872844
95	ku	0.870427
96	negara	0.867006
97	cuman	0.867006
98	juta	0.861654
99	anies	0.844069
100	xixi	0.844069
101	jaya	0.844069
102	lengang	0.844069
103	aamiin	0.837035
104	nongkrong	0.820216
105	sepi	0.816824
106	muak	0.814804
107	wkwkwkwkwk	0.813818
108	gakuatt	0.813818
109	bebb	0.813818
110	gegara	0.813818
111	atuhhh	0.813818
112	makin	0.813818
113	ati	0.813818
114	bioskop	0.813818
115	muluuuuu	0.813818
116	menang	0.813818
117	mentally	0.813818
118	bukappkm	0.813818
119	bener	0.813818
120	pakai	0.811395
121	cepat	0.798871
122	kak	0.791315
123	hilang	0.787798
124	motor	0.783322
125	nunggu	0.778649
126	mari	0.774947
127	jadi	0.760941
128	harga	0.760941

NO	TERM	Bobot Total
129	putih	0.754616
130	lapar	0.751217
131	angka	0.750284
132	bansos	0.748658
133	salah	0.740840
134	sembuh	0.738561
135	ancur	0.717459
136	jarang	0.717459
137	x	0.715584
138	ken	0.707767
139	dukung	0.701623
140	tambah	0.699196
141	lihat	0.699196
142	terima	0.697529
143	positif	0.696357
144	tanggal	0.681728
145	tingkat	0.678923
146	kumpul	0.675255
147	kaya	0.664705
148	dagang	0.662667
149	amp	0.655175
150	kena	0.653196
151	capeq	0.651055
152	dispensasi	0.651055
153	bangetttt	0.651055
154	bank	0.651055
155	capee	0.651055
156	donggg	0.651055
157	c	0.651055
158	pinjam	0.651055
159	jelek	0.651055
160	gbk	0.651055
161	berita	0.651055
162	rek	0.651055
163	cok	0.651055
164	dolan	0.651055
165	lari	0.651055
166	selamat	0.649284
167	yuk	0.649284
168	prokes	0.648603
169	tinggal	0.646015
170	ekonomi	0.637848

NO	TERM	Bobot Total
171	tular	0.628391
172	kerja	0.618984
173	fakta	0.606675
174	solusi	0.606675
175	capai	0.606343
176	bisa	0.596817
177	nonton	0.596817
178	hidup	0.596145
179	jokowi	0.595318
180	ratus	0.591868
181	beres	0.587672
182	beras	0.579149
183	ri	0.578782
184	sukses	0.574436
185	dampak	0.570252
186	keluarga	0.566379
187	pas	0.566154
188	kasih	0.562245
189	putus	0.558362
190	ny	0.553920
191	harap	0.544436
192	susah	0.543164
193	bijak	0.543059
194	dukcapil	0.542545
195	unduh	0.542545
196	targetkanx	0.542545
197	taut	0.542545
198	jaki	0.542545
199	batam	0.542545
200	ngopi	0.542545
201	semimgguyu	0.542545
202	mimika	0.542545
203	shopeefoodpaylater	0.542545
204	kantor	0.542545
205	kepri	0.542545
206	sertifikat	0.542545
207	aman	0.542545
208	semester	0.542545
209	nkri	0.542545
210	himbauan	0.542545
211	tunda	0.542545
212	sudah	0.542545

NO	TERM	Bobot Total
213	ra	0.542545
214	rampungrampung	0.542545
215	koyo	0.542545
216	fast	0.542545
217	furious	0.542545
218	naikin	0.542545
219	lapor	0.542545
220	prestasi	0.542545
221	sarawak	0.542545
222	takat	0.542545
223	julai	0.542545
224	lingkung	0.542545
225	bersih	0.542545
226	asli	0.542545
227	ges	0.542545
228	wae	0.542545
229	transportasi	0.542545
230	imbas	0.542545
231	bos	0.542545
232	lancar	0.542545
233	logistik	0.542545
234	olahraga	0.542545
235	kebut	0.542545
236	fitness	0.542545
237	bukti	0.542545
238	bin	0.541611
239	donk	0.539266
240	bilang	0.536325
241	coba	0.530594
242	data	0.527543
243	see	0.525199
244	karet	0.524767
245	paket	0.522674
246	normal	0.519463
247	ajar	0.514754
248	aje	0.514754
249	tetangga	0.509959
250	kg	0.506442
251	depan	0.503564
252	nder	0.500811
253	badan	0.500608
254	bendera	0.497726

NO	TERM	Bobot Total
255	sektor	0.495817
256	ribut	0.495817
257	gabut	0.495817
258	mudah	0.494765
259	laksana	0.492374
260	i	0.483736
261	stress	0.479585
262	kayak	0.473436
263	ganti	0.469348
264	kemarin	0.467755
265	jangkau	0.465039
266	suka	0.465039
267	melulu	0.465039
268	halu	0.465039
269	oke	0.465039
270	denger	0.465039
271	berantakan	0.465039
272	penderima	0.465039
273	doubble	0.465039
274	sukany	0.465039
275	signifikan	0.465039
276	p	0.465039
277	diilangin	0.465039
278	foto	0.465039
279	sii	0.465039
280	kadrun	0.465039
281	pcr	0.465039
282	mmengalami	0.465039
283	kecuali	0.465039
284	tobat	0.465039
285	pa	0.465039
286	tidur	0.465039
287	masuk	0.465039
288	perkara	0.465039
289	koo	0.465039
290	mobilisasi	0.465039
291	gerai	0.465039
292	realitanya	0.465039
293	kaget	0.465039
294	sok	0.465039
295	baswedan	0.465039
296	mana	0.465039

NO	TERM	Bobot Total
297	dibatasin	0.465039
298	mbak	0.465039
299	stay	0.465039
300	atur	0.464644
301	huhu	0.457204
302	serta	0.453208
303	akibat	0.450911
304	gantung	0.436102
305	tvone	0.436102
306	akim	0.436102
307	hippi	0.436102
308	nasib	0.436102
309	update	0.433758
310	my	0.433758
311	alas	0.429952
312	gitu	0.426645
313	time	0.420569
314	sasar	0.418518
315	org	0.411834
316	hamil	0.410311
317	balada	0.406909
318	manjang	0.406909
319	melar	0.406909
320	tanah	0.406909
321	minyak	0.406909
322	bunuh	0.406909
323	ciduk	0.406909
324	rendem	0.406909
325	dunia	0.406909
326	kibar	0.406909
327	balong	0.406909
328	pantai	0.406909
329	pengelolapedagang	0.406909
330	hantam	0.406909
331	proses	0.406909
332	deadline	0.406909
333	gimanaaaaaaaaaaaaa	0.406909
334	kabar	0.406909
335	ultah	0.406909
336	belanja	0.406909
337	longgar	0.406909
338	justifikasi	0.406909

NO	TERM	Bobot Total
339	sbg	0.406909
340	turunbisa	0.406909
341	trend	0.406909
342	ibun	0.406909
343	mcd	0.406909
344	tagih	0.406909
345	scroll	0.406909
346	tiktok	0.406909
347	asik	0.406909
348	deket	0.406909
349	ngerayain	0.406909
350	sm	0.406909
351	extend	0.406909
352	sebel	0.406909
353	ketat	0.402809
354	new	0.401673
355	darurat	0.401673
356	jual	0.400933
357	pikir	0.395657
358	syukur	0.391661
359	pondok	0.390722
360	pesantren	0.390722
361	in	0.388920
362	sekolah	0.387744
363	buka	0.386865
364	eh	0.386865
365	yaa	0.385303
366	mas	0.382192
367	bawa	0.382192
368	doain	0.380471
369	untung	0.379706
370	yaudahlah	0.374962
371	jakarta	0.367928
372	pdhl	0.367928
373	one	0.361697
374	ncsm	0.361697
375	bagirata	0.361697
376	pelanpelan	0.361697
377	pakdhe	0.361697
378	klt	0.361697
379	menunggucc	0.361697
380	blackpink	0.361697

NO	TERM	Bobot Total
381	alam	0.361697
382	kongsi	0.361697
383	beban	0.361697
384	buk	0.361697
385	the	0.361697
386	ndalu	0.361697
387	ringan	0.361697
388	sugeng	0.361697
389	bidan	0.361697
390	nurunin	0.361697
391	di	0.361697
392	ayooo	0.361697
393	lanjooot	0.361697
394	more	0.361697
395	movie	0.361697
396	Cuma	0.361697
397	kapolri	0.361697
398	kembang	0.361697
399	hitung	0.361697
400	isi	0.361697
401	kegabutan	0.361697
402	bruney	0.361697
403	manggung	0.361697
404	want	0.361697
405	fck	0.361697
406	via	0.361697
407	again	0.361697
408	wib	0.361697
409	hut	0.361697
410	persen	0.361697
411	massal	0.361697
412	gakalo	0.361697
413	urus	0.361697
414	strategi	0.361697
415	dokumen	0.361697
416	siap	0.361697
417	pindah	0.361697
418	bil	0.361697
419	potong	0.361533
420	kali	0.359812
421	menit	0.358419
422	rezeki	0.350343

NO	TERM	Bobot Total
423	kasar	0.342660
424	pedes	0.342660
425	habis	0.340127
426	pingin	0.337628
427	drastis	0.332758
428	riset	0.325527
429	pagi	0.325527
430	sawojajar	0.325527
431	nd	0.325527
432	manfaatin	0.325527
433	selagi	0.325527
434	jaga	0.325527
435	makasihkm	0.325527
436	kemensos	0.325527
437	bulog	0.325527
438	bagi	0.325527
439	uang	0.325527
440	produktif	0.325527
441	zoom	0.325527
442	pentol	0.325527
443	doang	0.325527
444	ukur	0.325527
445	mending	0.325527
446	cuk	0.325527
447	strawberry	0.325527
448	pilih	0.325527
449	aaaa	0.325527
450	tala	0.325527
451	fisik	0.325527
452	gini	0.325527
453	hmm	0.325527
454	murah	0.325527
455	climaheat	0.325527
456	jepang	0.325527
457	oleh	0.325527
458	secondannya	0.325527
459	utama	0.325527
460	prioritas	0.325527
461	adidas	0.325527
462	tuh	0.325318
463	jk	0.325318
464	ribu	0.324642

NO	TERM	Bobot Total
465	aksi	0.310026
466	gera	0.310026
467	dine	0.308063
468	manusia	0.307734
469	berat	0.303337
470	sakit	0.303337
471	produksi	0.295934
472	muara	0.295934
473	faktor	0.295934
474	rebut	0.295934
475	full	0.295934
476	pacar	0.295934
477	bulk	0.295934
478	bio	0.295934
479	farma	0.295934
480	antusias	0.295934
481	nasional	0.295934
482	bucin	0.295934
483	cuek	0.295934
484	angke	0.295934
485	brdoa	0.295934
486	tan	0.295934
487	bebek	0.295934
488	komunikasi	0.295934
489	pfft	0.295934
490	indo	0.295934
491	meanwhile	0.295934
492	stadium	0.295934
493	emirates	0.295934
494	maksimal	0.295934
495	bagus	0.295934
496	picu	0.295934
497	uji	0.295934
498	gemuk	0.295934
499	neng	0.295934
500	jauhjauh	0.295934
501	sebelumsebelumnya	0.295934
502	tua	0.295934
503	ponsel	0.295934
504	integrasi	0.295934
505	layar	0.295934
506	males	0.295934

NO	TERM	Bobot Total
507	begituan	0.295934
508	smoga	0.295934
509	ajak	0.295934
510	njrit	0.295934
511	status	0.295934
512	uda	0.295934
513	jakartakamu	0.295934
514	aktivitas	0.295934
515	syarat	0.295934
516	bgor	0.295934
517	pngumuman	0.295934
518	bsok	0.295934
519	kemenkes	0.287218
520	siti	0.287218
521	bicara	0.287218
522	juru	0.287218
523	nadia	0.287218
524	tarmizi	0.287218
525	kain	0.283067
526	mrk	0.283067
527	memang	0.281356
528	ambil	0.279288
529	desa	0.273541
530	dr	0.272254
531	cm	0.272254
532	hadiah	0.271273
533	ngelanggar	0.271273
534	ampir	0.271273
535	kerasin	0.271273
536	an	0.271273
537	engga	0.271273
538	wiryawati	0.271273
539	tina	0.271273
540	gerindra	0.271273
541	politis	0.271273
542	keras	0.271273
543	diem	0.271273
544	nii	0.271273
545	post	0.271273
546	kepala	0.271273
547	ajeee	0.271273
548	tetep	0.271273

NO	TERM	Bobot Total
549	ntr	0.271273
550	gada	0.271273
551	beneran	0.271273
552	rugi	0.271273
553	boro	0.271273
554	tempuh	0.271273
555	pulang	0.271273
556	lelah	0.271273
557	tegas	0.271273
558	sangka	0.271273
559	waw	0.271273
560	umur	0.271273
561	online	0.271273
562	kai	0.271273
563	lokal	0.271273
564	akuuutu	0.271273
565	kapolda	0.271273
566	gampang	0.271273
567	covidbaca	0.271273
568	tinjau	0.271273
569	daftar	0.271273
570	erti	0.271273
571	beliau	0.271273
572	tim	0.271273
573	forkopimda	0.271273
574	maaf	0.269124
575	langgar	0.269124
576	tuhan	0.263772
577	gendut	0.263772
578	kakak	0.260995
579	ngeluh	0.257377
580	upaya	0.257377
581	usaha	0.254303
582	skrg	0.250406
583	tok	0.250406
584	info	0.250406
585	hhhmmm	0.250406
586	nasi	0.250406
587	sambal	0.250406
588	plus	0.250406
589	lalap	0.250406
590	fenomena	0.250406

NO	TERM	Bobot Total
591	kaaku	0.250406
592	bablaaass	0.250406
593	kait	0.250406
594	pengin	0.250406
595	periksa	0.250406
596	dokter	0.250406
597	emang	0.250406
598	enaakk	0.250406
599	efektif	0.250406
600	nutup	0.250406
601	gede	0.250406
602	anjing	0.250406
603	baliho	0.250406
604	wkwkwkw	0.250406
605	direfund	0.250406
606	case	0.250406
607	lgaku	0.250406
608	hangat	0.250406
609	closed	0.250406
610	inget	0.250406
611	ppkmnya	0.250406
612	muolor	0.250406
613	ya	0.250406
614	informal	0.250406
615	jemput	0.250406
616	ros	0.250406
617	profesi	0.250406
618	joko	0.250406
619	bablas	0.250406
620	los	0.250406
621	widodo	0.250406
622	dol	0.250406
623	trus	0.250406
624	ganjur	0.250406
625	loh	0.250406
626	latih	0.248532
627	resto	0.247908
628	wilayah	0.246187
629	jawa	0.246187
630	kunjung	0.246187
631	tahan	0.246187
632	banting	0.246187

NO	TERM	Bobot Total
633	hadap	0.246187
634	besar	0.246187
635	total	0.246187
636	bal	0.246187
637	karantina	0.246187
638	sebab	0.246187
639	semi	0.246187
640	aku	0.246187
641	pandemiakhirnya	0.246187
642	menjeritjokowi	0.246187
643	stm	0.241131
644	salur	0.241131
645	hilir	0.241131
646	camat	0.241131
647	gaji	0.232759
648	lha	0.232519
649	kah	0.232519
650	corona	0.232519
651	virus	0.232519
652	pesawat	0.232519
653	gmn	0.232519
654	alodokter	0.232519
655	layang	0.232519
656	hepi	0.232519
657	ae	0.232519
658	efektivitas	0.232519
659	mahal	0.232519
660	cah	0.232519
661	infeksi	0.232519
662	wey	0.232519
663	krna	0.232519
664	kacau	0.232519
665	imunity	0.232519
666	bentuk	0.232519
667	herd	0.232519
668	suamik	0.232519
669	bobo	0.232519
670	lockdownpsbb	0.232519
671	lockdownmenurut	0.232519
672	opsiserukan	0.232519
673	disiplin	0.232519
674	sane	0.232519

NO	TERM	Bobot Total
675	vaksinasiagar	0.232519
676	tembok	0.232519
677	kesel	0.232519
678	mutlak	0.232519
679	subuh	0.232519
680	ilaaang	0.232519
681	gamau	0.232519
682	bosan	0.232519
683	please	0.232519
684	rbbln	0.232519
685	hidung	0.232519
686	suer	0.232519
687	coronces	0.232519
688	pinokio	0.232519
689	efek	0.232519
690	target	0.228602
691	ngeri	0.225085
692	lain	0.221568
693	live	0.219133
694	sekda	0.217018
695	truss	0.217018
696	rb	0.217018
697	insentif	0.217018
698	bahas	0.217018
699	bupatiwali	0.217018
700	tujuh	0.217018
701	nakes	0.217018
702	aceh	0.217018
703	click	0.217018
704	also	0.217018
705	anak	0.212099
706	langsung	0.203983
707	nanyak	0.203455
708	nantitentu	0.203455
709	om	0.203455
710	samapai	0.203455
711	wkkwkwkw	0.203455
712	thru	0.203455
713	rasa	0.203455
714	vibes	0.203455
715	timbang	0.203455
716	mikir	0.203455

NO	TERM	Bobot Total
717	chelsea	0.203455
718	plongo	0.203455
719	planga	0.203455
720	larang	0.203455
721	oh	0.203455
722	tb	0.203455
723	dingin	0.203455
724	bau	0.203455
725	sholat	0.203455
726	kerumun	0.203455
727	drive	0.203455
728	ngebiayain	0.203455
729	bulan	0.203455
730	dana	0.203455
731	setanvaksin	0.203455
732	lingkar	0.203455
733	pola	0.203455
734	turunppkm	0.203455
735	tertawa	0.203455
736	alloh	0.203455
737	pinter	0.203455
738	struknya	0.203455
739	yoon	0.203455
740	lucu	0.203455
741	sehatsehat	0.203455
742	jeplak	0.203455
743	dasar	0.203455
744	suara	0.203455
745	ohh	0.203455
746	verguso	0.203455
747	max	0.203455
748	hodob	0.203455
749	ultra	0.203455
750	sekali	0.203455
751	liatliatan	0.203455
752	malam	0.198032
753	mikro	0.191487
754	sapi	0.191487
755	kenyataanya	0.191487
756	daging	0.191487
757	daruratsmpe	0.191487
758	krusuhan	0.191487

NO	TERM	Bobot Total
759	all	0.191487
760	you	0.191487
761	eat	0.191487
762	cpatttttttttttt	0.191487
763	ayam	0.191487
764	slesaikn	0.191487
765	nmanya	0.191487
766	iniiiii	0.191487
767	geprek	0.191487
768	sunti	0.180848
769	asa	0.180848
770	nyaman	0.180848
771	sunyi	0.180848
772	toa	0.180848
773	kenceng	0.180848
774	umumin	0.180848
775	teknis	0.180848
776	suntik	0.180848
777	masjid	0.180848
778	belah	0.180848
779	sosialisasi	0.180848
780	pergi	0.176727
781	cahyono	0.171330
782	aipda	0.171330
783	mangkubumen	0.171330
784	lurah	0.171330
785	bhabinkamtibmas	0.171330
786	sambang	0.171330
787	kira	0.171330
788	istilah	0.171330
789	pukimak	0.171330
790	spesial	0.171330
791	ngomong-ngomong	0.171330
792	iki	0.171330
793	ancen	0.171330
794	gaplek	0.171330
795	ikate	0.171330
796	ngko	0.171330
797	penjara	0.171330
798	marioalah	0.171330
799	hate	0.171330
800	this	0.171330

NO	TERM	Bobot Total
801	system	0.171330
802	banjarsari	0.171330
803	damping	0.171330
804	sahar	0.171330
805	kapolsek	0.171330
806	satrio	0.171330
807	djoko	0.171330
808	kompol	0.171330
809	melarat	0.171330
810	nambah	0.171330
811	miskin	0.171330
812	tolong	0.171330
813	kontrak	0.171330
814	kasi	0.171330
815	utomo	0.171330
816	krl	0.162764
817	gojek	0.162764
818	strp	0.162764
819	surat	0.162764
820	busway	0.162764
821	kanan	0.155013
822	kiri	0.155013
823	siangppkm	0.155013
824	indoensia	0.155013
825	meratahadeeh	0.155013
826	pusiingyang	0.155013
827	pusing	0.155013
828	gambar	0.155013
829	teriak	0.155013
830	zolim	0.155013
831	dongok	0.155013
832	titis	0.155013
833	tuhanya	0.155013
834	nyoba	0.155013
835	gantiin	0.155013
836	medsos	0.155013
837	lawan	0.155013
838	redam	0.155013
839	senjata	0.155013
840	tangkis	0.155013
841	bulu	0.155013
842	olimpiade	0.155013

NO	TERM	Bobot Total
843	ngeyelan	0.155013
844	konspirasi	0.155013
845	kopid	0.155013
846	percaya	0.155013
847	diongkosin	0.155013
848	sentra	0.155013
849	dateng	0.155013
850	emoh	0.155013
851	fokus	0.155013
852	long	0.155013
853	kalah	0.155013
854	misi	0.155013
855	korban	0.155013
856	ginting	0.155013
857	dalih	0.155013
858	kontrol	0.155013
859	chen	0.155013
860	tuntas	0.155013
861	berani	0.155013
862	naahkalau	0.155013
863	receh	0.155013
864	petis	0.155013
865	bakso	0.155013
866	pantaiduduk	0.155013
867	mppkm	0.155013
868	mppkmj	0.155013
869	dugan	0.155013
870	diselingin	0.155013
871	obrolanobrolan	0.155013
872	mmppkmj	0.155013
873	senin	0.155013
874	wisata	0.155013
875	cpt	0.155013
876	es	0.155013
877	pusat	0.147967
878	medan	0.147967
879	taat	0.147967
880	lipat	0.147967
881	pinggir	0.147967
882	cafe	0.147967
883	heboh	0.147967
884	dirumahin	0.147967

NO	TERM	Bobot Total
885	engapan	0.147967
886	engap	0.147967
887	wkkkk	0.147967
888	phk	0.147967
889	dekat	0.147967
890	pemkab	0.147967
891	persuasif	0.147967
892	humanis	0.147967
893	tega	0.147967
894	langkah	0.147967
895	sadar	0.147967
896	tanggulang	0.147967
897	mknya	0.147967
898	sebar	0.147967
899	bantul	0.147967
900	silalahi	0.141534
901	nyadar	0.141534
902	ngadain	0.141534
903	hajat	0.141534
904	g	0.141534
905	kafan	0.141534
906	brwarna	0.141534
907	fucksin	0.135636
908	elo	0.135636
909	boong	0.135636
910	noh	0.135636
911	ude	0.135636
912	alattas	0.135636
913	block	0.135636
914	jenis	0.135636
915	rekomendasi	0.135636
916	sinovac	0.135636
917	dosis	0.135636
918	realisasi	0.135636
919	berisikin	0.135636
920	unfaedah	0.135636
921	kosankontrakancicilan	0.135636
922	bayar	0.135636
923	buncit	0.135636
924	duhh	0.135636
925	tl	0.135636
926	wis	0.125203

NO	TERM	Bobot Total
927	triliun	0.125203
928	kuli	0.125203
929	bandara	0.125203
930	stasiun	0.125203
931	paktiap	0.125203
932	muncul	0.125203
933	kucurin	0.125203
934	bakar	0.125203
935	mafia	0.125203
936	dluan	0.125203
937	jenggot	0.125203
938	kuat	0.125203
939	tenaga	0.125203
940	covidnya	0.125203
941	tolak	0.125203
942	supervisor	0.125203
943	minum	0.125203
944	b	0.125203
945	dl	0.125203
946	setan	0.125203
947	barusan	0.125203
948	anterga	0.125203
949	ending	0.125203
950	patuh	0.120566
951	tunai	0.120566
952	sosial	0.120566
953	apm	0.120566
954	t	0.120566
955	hesron	0.120566
956	pantau	0.120566
957	girsang	0.120566
958	jatah	0.116260
959	kendali	0.116260
960	rolls	0.116260
961	kecayangan	0.116260
962	hihisambil	0.116260
963	ajah	0.116260
964	smg	0.116260
965	and	0.116260
966	healthy	0.116260
967	safe	0.116260
968	luck	0.116260

NO	TERM	Bobot Total
969	me	0.116260
970	wish	0.116260
971	sapatauu	0.116260
972	sandal	0.116260
973	nitip	0.116260
974	royce	0.116260
975	beli	0.116260
976	berattt	0.116260
977	lele	0.116260
978	pecel	0.116260
979	suruh	0.116260
980	fortuner	0.116260
981	bodo	0.116260
982	ribet	0.116260
983	drpd	0.116260
984	sabar	0.116260
985	papar	0.116260
986	delta	0.116260
987	varian	0.116260
988	daya	0.116260
989	bayang	0.116260
990	simpul	0.116260
991	kasus	0.116260
992	kgn	0.116260
993	susul	0.116260
994	dki	0.116260
995	nuice	0.116260
996	media	0.116260
997	cov	0.116260
998	indonesiahehe	0.116260
999	humble	0.116260
1000	opinion	0.116260
1001	shotsemoga	0.116260
1002	first	0.116260
1003	diapresiasiturun	0.116260
1004	revisi	0.108509
1005	tip	0.108509
1006	pegang	0.108509
1007	luar	0.108509
1008	pos	0.108509
1009	rbsedih	0.108509
1010	tunggal	0.108509

NO	TERM	Bobot Total
1011	hard	0.108509
1012	no	0.108509
1013	skripsi	0.108509
1014	mikirin	0.108509
1015	wisuda	0.108509
1016	sidang	0.108509
1017	feelings	0.108509
1018	uring	0.108509
1019	yah	0.108509
1020	betah	0.108509
1021	gtu	0.108509
1022	like	0.108509
1023	literally	0.108509
1024	parah	0.108509
1025	ngmgin	0.108509
1026	situ	0.108509
1027	ulama	0.105009
1028	peran	0.105009
1029	alim	0.105009
1030	barang	0.105009
1031	mahfud	0.105009
1032	saudara	0.105009
1033	bui	0.105009
1034	keluar	0.105009
1035	kawan	0.105009
1036	pimpin	0.105009
1037	asuh	0.105009
1038	laki	0.105009
1039	rehabilitasi	0.105009
1040	implementasi	0.105009
1041	agama	0.105009
1042	jahat	0.101727
1043	twt	0.101727
1044	july	0.101727
1045	dar	0.101727
1046	jenguk	0.101727
1047	ucapin	0.101727
1048	pgn	0.101727
1049	sungkawa	0.101727
1050	quality	0.101727
1051	mma	0.101727
1052	he	0.101727

NO	TERM	Bobot Total
1053	bela	0.101727
1054	serius	0.093008
1055	sebenernya	0.093008
1056	ngurusin	0.093008
1057	event	0.093008
1058	kampus	0.093008
1059	jujur	0.093008
1060	kpan	0.093008
1061	jugaa	0.093008
1062	becanda	0.093008
1063	rapat	0.093008
1064	keadaanya	0.093008

LAMPIRAN III

Jumlah Penggunaan Fitur

NO	Persentase Fitur	Jumlah Fitur
1	10%	107 Term
2	20%	213 Term
3	30%	319 Term
4	40%	426 Term
5	50%	532 Term
6	60%	639 Term
7	70%	745 Term
8	80%	851 Term
9	90%	956 Term