

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KINERJA
PEMERINTAH DALAM MENANGGULANGI COVID-19 MENGGUNAKAN
METODE *BAGGING ENSEMBLE CLASSIFIER* DENGAN SELEKSI FITUR
*INFORMATION GAIN***

SKRIPSI

**Oleh:
MUHAMMAD IBRAM GUSTI CHILDRABAHTI
NIM. 18650094**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2023**

**KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KINERJA
PEMERINTAH DALAM MENANGGULANGI COVID-19
MENGUNAKAN METODE *BAGGING ENSEMBLE CLASSIFIER*
DENGAN SELEKSI FITUR *INFORMATION GAIN***

SKRIPSI

Oleh :

**MUHAMMAD IBRAM GUSTI CHILDRABAHTI
NIM. 18650094**

Diajukan Kepada:

**Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) 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
2023**

HALAMAN PERSETUJUAN

KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KINERJA PEMERINTAH DALAM MENANGGULANGI COVID-19 MENGUNAKAN METODE *BAGGING ENSEMBLE CLASSIFIER* DENGAN SELEKSI FITUR *INFORMATION GAIN*

SKRIPSI

Oleh :

MUHAMMAD IBRAM GUSTI CHILDRABAHTI
NIM 18650094

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal : 30 Mei 2023

Pembimbing I



Fajar Bohman Hariri, M.Kom
NIP. 19890515 201801 1 001

Pembimbing II



Dr. Cahyo Crisdian
NIP. 19740424 200901 1 008

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



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

HALAMAN PENGESAHAN

KLASIFIKASI SENTIMEN MASYARAKAT TERHADAP KINERJA PEMERINTAH DALAM MENANGGULANGI COVID-19 MENGUNAKAN METODE *BAGGING ENSEMBLE CLASSIFIER* DENGAN SELEKSI FITUR *INFORMATION GAIN*

SKRIPSI

Oleh:

MUHAMMAD IBRAM GUSTI CHILDRABAHTI
NIM. 18650094

Telah Dipertahankan di Depan Dewan Penguji
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal: 21 Juni 2023

Susunan Dewan Penguji

- Ketua Penguji : Hani Nurhayati, M.T
NIP. 19780625 200801 2 006
- Anggota Penguji I : Puspa Miladin Nuraida Safitri A
Basid, M.Kom
NIP. 19930828 201903 2 018
- Anggota Penguji II : Fajar Rohman Hariri, M.Kom
NIP. 19890515 201801 1 001
- Anggota Penguji III : Dr. Cahyo Crysdiyan
NIP. 19740424 200901 1 008



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



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

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini :

Nama : Muhammad Ibram Gusti Childrabahti
NIM : 18650094
Fakultas : Sains dan Teknologi
Jurusan : Teknik Informatika
Judul Skripsi : Klasifikasi Sentimen Masyarakat Terhadap Kinerja
Pemerintah Dalam Menanggulangi Covid-19 Menggunakan
Metode *Bagging Ensemble Classifier* Dengan Seleksi Fitur
Information Gain

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

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

Malang, 20 Juni 2023
Yang Membuat pernyataan,



Muhammad Ibram Gusti Childrabahti
NIM. 18650094

HALAMAN MOTTO

حَسْبُنَا اللَّهُ وَنِعْمَ الْوَكِيلُ

“Cukuplah Allah (menjadi penolong) bagi kami dan Dia sebaik-baik pelindung”

(Q.S. Ali-Imran : 173)

KATA PENGANTAR

Assalamualaikum Wr. Wb

Bismillaahirrohmaanirrahiim, segala ungkapan syukur penulis limpahkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis mampu menyelesaikan skripsi ini dengan baik. Sholawat serta salam tetap tercurahkan kepada junjungan kita Nabi Muhammad SAW atas syafaatnya yang telah menuntun umat islam menuju jalan yang baik. Semoga kita semua termasuk dalam golongan yang dituntun Allah SWT dan mendapat pertolongan Nabi Muhammad SAW. Aamiin.

Penulis sangat menyadari bahwa penulis masih sangat minim ilmu dan pengetahuan, sehingga tanpa adanya peran dan kontribusi dari pihak yang telah membantu meluangkan waktu dan memberikan sumbangsih pemikiran dalam membimbing penulis, penulis tidak akan mampu menyelesaikan skripsi ini dengan baik. Pada bagian ini juga segala kerendahan hati, penulis menyampaikan terima kasih kepada :

1. Prof. Dr. M. Zainuddin, M.A., Selaku rektor Universitas Islam Negeri MaulanaMalik Ibrahim Malang.
2. Dr. Sri Hariani, M.Si selalu dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.

3. Dr. Fachrul Kurniawan, M.MT., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang yang senantiasa memberikan dorongan.
4. Bapak Fajar Rohman Hariri, M.Kom selaku dosen pembimbing I yang telah bersedia meluangkan waktunya dalam membimbing dan memberi arah kepada penulis sehingga dapat menyelesaikan skripsi ini.
5. Dr. Cahyo Crysdiان selaku dosen pembimbing II yang juga bersedia meluangkan waktunya dalam membimbing dan memberi arah kepada penulis sehingga dapat menyelesaikan skripsi ini.
6. Bapak Ajib Hanani, M.T, selaku wali dosen penulis selama kuliah di Universitas Islam Negeri Maulana Malik Ibrahim Malang
7. Ibu Hani Nurhayati, M.T, selaku Dosen Penguji I dan Ibu Puspa Miladin Nuraida Safitri A Basid, M.Kom selaku Dosen Penguji II yang telah meluangkan waktu memberikan arahan untuk skripsi ini.
8. Keluarga yang selalu memberikan support saya terutama untuk Bapak Baharuddin, Ibu Yetty Sugihartitin, dan adik saya M. Raihan Novi Bahti yang selalu memberikan segala dukungan kepada penulis.
9. Segenap civitas akademika Program Studi Teknik Informatika, Terutama seluruh dosen, Terima Kasih atas ilmu dan bimbingan yang telah diberikan selama masa perkuliahan ini.
10. Teruntuk teman-teman saya dan orang yang tersayang yang menemani penulis, memberi dukungan penuh, serta memotivasi dalam mengerjakan

sehingga skripsi ini dapat terselesaikan terutama Diaz, Danu, Wahyu, dan Bella.

11. Saudara Program Studi Teknik Informatika Angkatan 2018 “Unity Of Informatics Force” yang sama-sama mengejar gelar S.Kom di Universitas yang sama.

Penulis sangat menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan dan penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca maupun bagi penulis.

Malang, 20 Juni 2023

Penulis

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN.....	iv
PERNYATAAN KEASLIAN TULISAN	iv
HALAMAN MOTTO	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiii
ABSTRAK	xiv
ABSTRACT	xv
خلاصة.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Pernyataan Masalah.....	5
1.3 Tujuan Penelitian.....	5
1.4 Batasan Penelitian	5
1.5 Manfaat Penelitian.....	5
BAB II STUDI PUSTAKA	7
2.1 Analisis Sentimen	7
2.2 <i>Text Preprocessing</i>	9
2.3 <i>Information Gain</i>	10
2.4 <i>Naïve Bayes</i>	11
2.5 <i>Ensemble Learning</i>	13
2.6 <i>Bagging Ensemble Classifier</i>	15
BAB III METODOLOGI PENELITIAN	17
3.1 Akuisisi Data	17
3.2 Perancangan Sistem.....	18
3.2.1. <i>Preprocessing</i>	19
3.2.1.1. <i>Cleansing</i>	20
3.2.1.2. <i>Lowercase</i>	20
3.2.1.3. <i>Tokenizing</i>	20
3.2.1.4. <i>Stopwords Removal</i>	21
3.2.1.5. <i>Stemming</i>	22
3.2.1.6. <i>Terms</i>	22
3.2.2. Pembobotan TF-IDF	23
3.2.3. Seleksi Fitur <i>Information Gain</i>	25
3.2.4. <i>Bagging Ensemble Classifier</i>	27
3.2.4.1. <i>Bootstrap Aggregating (Bagging)</i>	28
3.2.4.2. Klasifikasi <i>Naïve Bayes</i>	28
3.2.4.3. <i>Majority Voting</i>	32
BAB IV UJI COBA DAN PEMBAHASAN.....	34
4.1 Langkah-langkah Uji Coba	34

4.2	Hasil Uji Coba	38
4.3	Pembahasan	42
BAB V PENUTUP.....		49
5.1	Kesimpulan.....	49
5.2	Saran	49
DAFTAR PUSTAKA		48
LAMPIRAN.....		50

DAFTAR GAMBAR

Gambar 3.1 Alur Sistem Klasifikasi	18
Gambar 3.2 Alur <i>Preprocessing</i>	19
Gambar 3.3 Alur Metode Ensemble Learning	28
Gambar 4.1 Grafik hasil pengujian Bagging Ensemble Learning dengan klasifikasi dasar Decision Tree	43
Gambar 4.2 Grafik hasil pengujian Bagging Ensemble Learning dengan klasifikasi dasar KNN	43
Gambar 4.3 Grafik hasil pengujian Bagging Ensemble Learning dengan klasifikasi dasar Naive Bayes.....	44
Gambar 4.4 Grafik Hasil pengujian metode Naive Bayes dengan pembagian dataset 7:3	45
Gambar 4.5 Grafik Hasil pengujian Bagging Ensemble Classifier menggunakan Naive Bayes dengan pembagian dataset 7:3	45
Gambar 4.6 Grafik Hasil pengujian metode Naive Bayes dengan pembagian dataset 8:2	45
Gambar 4.7 Grafik Hasil pengujian Bagging Ensemble Classifier menggunakan Naive Bayes dengan pembagian dataset 8:2	46

DAFTAR TABEL

Tabel 3.1 Contoh data teks dengan label	18
Tabel 3.2 Hasil tahap <i>Cleansing</i>	20
Tabel 3.3 Hasil tahap <i>Lowecase</i>	20
Tabel 3.4 Hasil tahap <i>Tokenizing</i>	21
Tabel 3.5 Hasil tahap <i>Tokenizing</i>	21
Tabel 3.6 Hasil tahap <i>Stemming</i>	22
Tabel 3.7 Hasil <i>preprocessing</i> Dokumen 1 (D1)	22
Tabel 3.8 kata muncul dalam dokumen	23
Tabel 3.9 Perhitungan <i>Term Frequency</i>	24
Tabel 3.10 Perhitungan IDF	24
Tabel 3.11 Hasil TF-IDF	25
Tabel 3.12 Document Frequency	25
Tabel 3.13 <i>Information Gain</i> Terbesar sampai Terkecil	27
Tabel 3.14 <i>Term Frequency</i> setiap kata	29
Tabel 3.15 Perhitungan Probabilitas <i>Prior</i>	29
Tabel 3.16 Hasil Perhitungan Probabilitas Setiap Kata	30
Tabel 3.17 Contoh Data Uji	31
Tabel 3.18 Perhitungan Probabilitas Kata Uji	31
Tabel 3.19 Probabilitas Akhir	31
Tabel 3.20 Hasil <i>Bagging</i> pada setiap index perulangan	32
Tabel 4.1 Sampel dataset	34
Tabel 4.2 Pembagian Data	35
Tabel 4.3 Sampel hasil klasifikasi	36
Tabel 4.4 <i>Confusion matrix</i> hasil klasifikasi dengan fitur 20%	38
Tabel 4.5 Hasil perhitungan TP, TN, FP, FN	39
Tabel 4.6 Hasil pengujian dengan pembagian dataset 8:2	41
Tabel 4.7 Hasil pengujian dengan pembagian dataset 7:3	42

ABSTRAK

Childrabahti, Muhammad Ibram Gusti. 2023. “**Klasifikasi Sentimen Masyarakat Terhadap Kinerja Pemerintah Dalam Menanggulangi Covid-19 Menggunakan Metode *Bagging Ensemble Learning* dengan Seleksi Fitur *Information Gain***”. Skripsi. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Fajar Rohman Hariri, M.Kom, (II) Dr. Cahyo Crysdiان

Kata Kunci : *Bagging Ensemble Learning*, *Information Gain*, Covid-19, Pemerintah

Media sosial merupakan salah satu media yang digunakan untuk melakukan interaksi maupun komunikasi. Dengan adanya media sosial tentunya akan memudahkan manusia dalam menyampaikan suatu informasi terhadap orang lain atau menyampaikan kondisi yang sedang dialami. Salah satu media sosial yang biasa digunakan adalah Youtube. Youtube telah memfasilitasi pengguna untuk mendapatkan informasi dari video yang diunggah oleh seseorang melalui kanal Youtube, pengguna juga bisa memberikan komentar terhadap video yang pengguna tonton. Tidak jarang masyarakat Indonesia menyampaikan apa yang mereka rasakan, alami, atau bahkan hanya sekedar menyampaikan informasi kedalam komentar video Youtube. Pada penelitian ini mengambil kasus ketika pemerintah memberlakukan kebijakan PPKM pada masa pandemi banyak masyarakat yang memberikan komentar dengan berbagai sentimen seperti positif, negatif, dan bahkan hanya menyampaikan informasi. Dalam penelitian ini menggunakan metode *Bagging Ensemble Learning* dengan seleksi fitur *Information Gain*. Tujuan dari penelitian ini ingin mengukur pengaruh penerapan metode *Bagging Ensemble Learning* pada akurasi, presisi, *recall*, dan *f-measure* dengan metode klasifikasi dasar *Naïve Bayes*. Metode *Bagging Ensemble Learning* menerapkan proses *bootstrap aggregating* pada metode klasifikasi dasar sehingga bisa mendapatkan hasil performa lebih baik dan pada penelitian ini menggunakan klasifikasi dasar *Naive Bayes*. Sedangkan *Information Gain* digunakan untuk menghilangkan fitur-fitur yang tidak relevan sebelum memasuki proses klasifikasi. Pada klasifikasi sentimen menggunakan *Bagging Ensemble Learning* dan seleksi fitur *Information Gain* didapatkan hasil terbaik pada pembagian dataset 8:2 dengan fitur sebanyak 30% menghasilkan nilai akurasi sebesar 64%, presisi sebesar 67%, *recall* sebesar 66%, dan *f-measure* sebesar 64%. Sehingga dapat disimpulkan bahwa penggunaan *Bagging Ensemble Learning* pada metode *Naive Bayes* akan berpengaruh pada hasil klasifikasi

ABSTRACT

Childrabahti, Muhammad Ibram Gusti. 2023. “**Classification of Public Sentiment on Government Performance in Overcoming Covid-19 Using the *Bagging Ensemble Learning Method with Feature Selection Information Gain***”. Essay. Informatics Engineering Department, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang. Advisor: (I) Fajar Rohman Hariri, M.Kom, (II) Dr. Cahyo Crysdiان

Social media is one of the media used for interaction and communication. With social media, of course, it will make it easier for humans to convey information to other people or convey the conditions that are being experienced. One of the most commonly used social media is YouTube. Youtube has facilitated users to get information from videos uploaded by someone through the Youtube channel, users can also provide comments on videos that users watch. Not infrequently Indonesian people convey what they feel, experience, or even just convey information in Youtube video comments. This research takes the case when the government implemented the PPKM policy during the pandemic, many people commented with various sentiments such as positive, negative, and even just conveying information. In this study using the Bagging Ensemble Learning method with Information Gain feature selection. The purpose of this research is to measure the effect of applying the Bagging Ensemble Learning method on accuracy, precision, recall, and f-measure using the basic Naïve Bayes classification method. The Bagging Ensemble Learning method applies the bootstrap aggregating process to the basic classification method so that it can get better performance results and in this study uses the Naive Bayes basic classification. Meanwhile, Information Gain is used to eliminate irrelevant features before entering the classification process. In the sentiment classification using Bagging Ensemble Learning and Information Gain feature selection, the best results were obtained in the distribution of the 8: 2 dataset with 30% features resulting in an accuracy value of 64%, precision of 67%, recall of 66%, and f-measure of 64%. . So it can be concluded that the use of Bagging Ensemble Learning in the Naive Bayes method will affect the classification results

Keywords : *Bagging Ensemble Learning, Information Gain, Covid-19, Government*

خلاصة

شلدراباهتي ، محمد إبراهيم جوستي .2023". تصنيف الرأي العام حول الأداء الحكومي في التغلب على Covid-19 باستخدام طريقة تعلم مجموعة التعبئة مع اختيار ميزات اكتساب المعلومات أطرؤحة .قسم هندسة المعلوماتية بكلية العلوم والتكنولوجيا .الجامعة الإسلامية الحكومية مولانا مالك إبراهيم فجر رحمان حريري ، محمد كوم (2) د . كاهو كريسيديان (1)مالانج .المشرفون

الحكومة ، Covid-19 ، الكلمات المفتاحية :تعلم مجموعة التعبئة ، اكتساب المعلومات

وسائل التواصل الاجتماعي هي إحدى الوسائط المستخدمة للتفاعل والتواصل .مع وسائل التواصل الاجتماعي ، بالطبع ، ستسهل على البشر نقل المعلومات إلى أشخاص آخرين أو نقل الظروف التي يمرون بها .يعد موقع للمستخدمين الحصول على Youtube أحد أكثر وسائل التواصل الاجتماعي شيوعًا .سهّل موقع YouTube ويمكن للمستخدمين أيضًا ، Youtube معلومات من مقاطع الفيديو التي تم تحميلها بواسطة شخص ما عبر قناة تقديم تعليقات على مقاطع الفيديو التي يشاهدها المستخدمون .ليس من النادر أن ينقل الإندونيسيون ما يشعرون يأخذ هذا البحث الحالة عندما نفذت . Youtube به أو يختبرونه أو حتى ينقلوا المعلومات فقط في تعليقات فيديو أثناء الوباء ، أعطى العديد من الأشخاص تعليقات بمشاعر مختلفة مثل الإيجابية PPKM الحكومة سياسات والسلبية وحتى مجرد نقل المعلومات .في هذه الدراسة ، يتم استخدام أسلوب تعليمي مجموعة التعبئة مع اختيار ميزة اكتساب المعلومات .الغرض من هذا البحث هو قياس تأثير تطبيق طريقة تعلم مجموعة التعبئة على الدقة والدقة الأساسية .تطبق طريقة تعلم مجموعة التعبئة في Naïve Bayes والاستدعاء والقياس باستخدام طريقة تصنيف عملية التجميع على طريقة التصنيف الأساسية بحيث يمكنها الحصول على نتائج أداء أفضل وفي هذه الدراسة الأساسية .وفي الوقت نفسه ، يتم استخدام كسب المعلومات للتخلص من Naive Bayes تستخدم تصنيف الميزات غير ذات الصلة قبل الدخول في عملية التصنيف .تم الحصول على أفضل النتائج في توزيع مجموعة البيانات مع ميزات 30٪ مما أدى إلى قيمة دقة 64٪ ودقة 67٪ واستدعاء 66٪ في تصنيف المشاعر باستخدام ميزة 2 :8 اختيار مجموعة التعلم واكتساب المعلومات .، و قياس 64٪ .لذلك يمكن استنتاج أن استخدام تعلم مجموعة سيؤثر على نتائج التصنيف Naive Bayes التعبئة في طريقة

BAB I

PENDAHULUAN

1.1 Latar Belakang

Media sosial merupakan salah satu media yang digunakan untuk berinteraksi dan berkomunikasi. Tentunya media sosial memudahkan untuk berbagi informasi dengan orang lain atau menyampaikan keadaan yang dialami. Salah satu media sosial yang paling banyak digunakan adalah Youtube. Menurut laporan Kepios, pada Juli 2022, jumlah pengguna YouTube di Indonesia menempati urutan ketiga dengan 139 juta pengguna aktif, mengungguli Amerika Serikat dan India. (Kepios, 2022). Youtube telah memfasilitasi pengguna untuk mendapatkan informasi dari video yang diunggah oleh seseorang melalui kanal Youtube, pengguna juga bisa memberikan komentar terhadap video yang pengguna tonton. Tidak jarang masyarakat Indonesia menyampaikan apa yang mereka rasakan, alami, atau bahkan hanya sekedar menyampaikan informasi kedalam komentar video Youtube.

Salah satu contoh ketika pertama kali diberlakukan kebijakan PPKM dan video berita tersebut diunggah di Youtube, pengguna Youtube yang melihat video tersebut berbondong-bondong mengungkapkan emosi atau apa yang mereka rasakan kedalam komentar video dan tak jarang memberikan *hashtag* sehingga topik yang dibicarakan bisa menempati puncak *trending* di Youtube. Virus Covid-19 mulai memasuki Indonesia pada bulan Maret 2020 kemudian kasus positif semakin melonjak setiap harinya, pemerintah selalu berupaya menekan laju penularan virus dengan mengeluarkan berbagai kebijakan seperti penggunaan

masker, PPKM, penggunaan aplikasi peduli lindungi, dan sebagainya. Berita mengenai penanganan Covid diberitakan hampir diseluruh media termasuk Youtube yang ditonton jutaan warga Indonesia. Selama pandemi Covid-19, konten komentar YouTube menjadi lebih beragam, termasuk pertanyaan, pendapat, komentar, dan kritik. Salah satu yang menjadi perhatian akhir-akhir ini adalah kinerja pemerintah dalam menangani pandemi Covid-19.

Masyarakat cenderung meluapkan emosi yang mereka rasakan ketika melihat video pemberitaan pemerintah menangani pandemi berupa komentar, ada yang berupa opini positif, dan tak jarang sejumlah orang menuliskan sesuatu yang bisa menimbulkan emosi atau reaksi negatif, tidak membangun dan hanya diperuntukkan mencaci. Contohnya dengan terus melonjaknya angka kematian yang disebabkan oleh Covid-19 masyarakat mempertanyakan bahkan mengkritik kinerja serta kebijakan pemerintah dalam menanggulangi pandemi. Dengan berbagai komentar dari masyarakat, maka diperlukan sistem untuk mengklasifikasi sentimen masyarakat terhadap kebijakan pemerintah dalam menanggulangi Covid-19. Dengan itu penulis berusaha membuat sistem untuk mengklasifikasi sentimen dari masyarakat yaitu sentimen positif, negatif, dan netral. Penelitian bisa membantu pemerintah dalam mengukur tingkat kepuasan masyarakat terhadap kebijakan pemerintah menangani pandemi dan diharapkan pemerintah bisa lebih bijak dalam mengambil keputusan serta menjaga media sosial agar menjadi tempat yang aman untuk semua orang.

Dalam Al-Qur'an telah dijelaskan mengenai perkara tutur kata yang baik dan lembut dalam menyampaikan kebenaran kepada seseorang bahkan kepada pemimpin seperti pemerintah, seperti firman Allah pada surah Thahaa ayat 42-44:

أَذْهَبْ أَنْتَ وَأَخُوكَ بِآيَاتِي وَلَا تَنِيَا فِي ذِكْرِي (٤٢) أَدْهَبَا إِلَى فِرْعَوْنَ إِنَّهُ طَغَىٰ (٤٣)

فَقُولَا لَهُ ، قَوْلًا لِّبْنَا لَعَلَّهُ ، يَتَذَكَّرُ أَوْ يَخْشَىٰ (٤٤)

Artinya: “Pergilah engkau dan juga saudaramu dengan membawa tanda-tanda (kekuasaan)-Ku, dan janganlah kamu berdua lalai mengingat-Ku; (42) pergilah kamu berdua kepada Fir'aun, karena sungguh dia benar-benar telah melampaui batas; (43) maka berbicaralah kamu berdua kepadanya (Fir'aun) dengan kata-kata yang lemah lembut, mudah-mudahan dia sadar atau takut. (44)” (Q.S. Thahaa : 42-44).

Berdasarkan tafsir Ibnu Katsir jilid 5, dapat disimpulkan ayat ini memerintahkan kepada nabi Musa dan Harun untuk menyerukan kebenaran tentang Allah kepada Fir'aun dengan lemah lembut, agar seruannya menyentuh jiwa, lebih mendalam, dan tepat sasaran. Sama halnya ketika masyarakat ingin mengutarakan sesuatu mengenai kinerja pemerintah dalam penanganan pandemi alangkah baiknya disampaikan dengan kata-kata yang tertata dan membangun.

Beberapa metode yang populer digunakan untuk mengklasifikasikan sentimen antara lain *Naïve Bayes (NB)*, *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)*, dan *Decision Tree*. Metode-metode klasifikasi tersebut seringkali digunakan di penelitian dengan topik kesehatan, bisnis, dan industri. Ketika membangun model klasifikasi, data yang digunakan untuk membangun model mungkin memiliki *noise* atau informasi yang tidak seimbang. Untuk meningkatkan akurasi klasifikasi maka digunakanlah metode *ensemble learning*. Penulis dapat menggabungkan beberapa model yang mengarah pada pengurangan

data yang biasa. Metode *ensemble* seperti *bagging* dan *boosting* telah dikembangkan oleh ahli statistik bernama Leo Breiman (1928-2005), metode tersebut dapat diterapkan pada pengklasifikasi dasar. Penulis memilih menggunakan metode *Bagging Ensemble Classifier* dengan metode klasifikasi dasar *Naïve Bayes*, karena dalam penerapan metode ini ditujukan mendapatkan hasil klasifikasi yang lebih baik dari nilai akurasi, *recall*, presisi, dan *f-measure* dibandingkan dengan hanya menggunakan metode klasifikasi dasar. Penulis juga menggunakan metode *Information Gain* dalam melakukan seleksi fitur yang diharapkan dapat meningkatkan akurasi dalam klasifikasi serta performa dengan membuang fitur yang tidak relevan sehingga mempercepat proses klasifikasi.

Dalam Al-Qur'an telah dijelaskan pula bahwa Allah menciptakan segala sesuatu sesuai dengan ukurannya masing-masing, seperti firman Allah pada surah Al-Qamar ayat 49:

إِنَّا كُلَّ شَيْءٍ خَلَقْنَاهُ بِقَدَرٍ

Artinya: “*Sesungguhnya Kami menciptakan segala sesuatu sesuai dengan ukuran*” (Q.S Al-Qamar: 49).

Merujuk pada tafsir Kementerian Agama Republik Indonesia surat Al-Qamar ayat 49 menjelaskan mengenai apa yang terjadi pada semua makhluk sudah ditetapkan oleh Allah menurut ukuran, yaitu suatu sistem dan ketentuan yang telah ditetapkan. Penulis menggunakan ayat dan tafsir tersebut sebagai dasar mengapa penelitian ini perlu dilakukan, yaitu mengukur pengaruh dari metode *Bagging Ensemble Learning* pada akurasi, *recall*, presisi dan *f-measure* dengan metode klasifikasi dasar *Naïve Bayes*.

1.2 Pernyataan Masalah

Bagaiman pengaruh implementasi metode *Bagging Ensemble Classifier* pada akurasi, *recall*, presisi, dan *f-measure* dengan metode klasifikasi dasar *Naïve Bayes* pada sistem klasifikasi sentimen masyarakat terhadap kinerja pemerintah dalam menanggulangi Covid-19 menggunakan metode *Bagging Ensemble Classifier* dengan seleksi fitur *Information Gain*.

1.3 Tujuan Penelitian

Tujuan dari penelitian ini ingin mengukur pengaruh implementasi metode *Bagging Ensemble Learning* pada akurasi, *recall*, presisi, dan *f-measure* dengan metode klasifikasi dasar *Naïve Bayes* pada sistem klasifikasi sentimen masyarakat terhadap kinerja pemerintah dalam menanggulangi Covid-19 menggunakan metode *Bagging Ensemble Classifier* dengan seleksi fitur *Information Gain*.

1.4 Batasan Penelitian

Penelitian ini memiliki Batasan sebagai berikut:

1. Data yang digunakan dalam penelitian ini diambil dari komentar Youtube yang diambil melalui *Youtube API* dari 3 video pemberitaan Covid-19 dengan kata kunci “pemerintah” dan “Covid” pada bulan Juli 2022.
2. Data yang akan diklasifikasikan menjadi 3 kelas yaitu positif, negatif, dan netral.

1.5 Manfaat Penelitian

Hasil dari penelitian ini menjelaskan seberapa besar pengaruh metode *Bagging Ensemble Classifier* ketika diaplikasikan pada metode klasifikasi dasar

Naïve Bayes pada sistem klasifikasi sentimen masyarakat terhadap kinerja pemerintah dalam menanggulangi Covid-19 menggunakan metode *Bagging Ensemble Classifier* dengan seleksi fitur *Information Gain*. Sehingga dapat memberikan manfaat bagi beberapa pihak, antara lain:

1. Pihak Pemerintah, membantu menilai kinerja dan kebijakan pemerintah dalam penanggulangan pandemi Covid-19.
2. Pihak Peneliti, sebagai rujukan untuk penelitian lebih lanjut.

BAB II

STUDI PUSTAKA

2.1 Analisis Sentimen

Penulis mengangkat sentimen analisis sebagai topik penelitian. Pada sub bab ini akan membahas penelitian sebelumnya yang mengangkat topik sentiment analisis serta penjelasan sentiment analisis menurut peneliti terdahulu.

Penelitian dengan topik analisis sentimen dilakukan oleh Maulana (2016) dalam mengklasifikasikan *tweet* berdasarkan emosi manusia. Data *tweet* yang diambil berjumlah delapan ratus ribu *tweet* selama 2 minggu dan akan diklasifikasikan kedalam enam emosi yaitu senang, sedih, marah, terkejut, takut, dan jijik/muak. Dalam penelitian ini didapati sistem bisa dengan baik mendeteksi emosi neural, senang, dan sedih dengan *fscore* masing-masing 77%, 75% dan 65%, sedangkan peforma pada kelas lain mulai menurun drastis bahkan pada kelas emosi jijik tidak terdeteksi sama sekali (Maulana, 2016).

Penelitian yang dilakukan oleh Kosasih dan Alberto (2021) yaitu Analisis sentimen produk game di shopee menggunakan metode TF-IDF dan Pengklasifikasi *Naïve Bayes* meneliti seberapa baik performa algoritma *Naïve Bayes* dengan pembobotan fitur TF-IDF. Pada penelitian ini mengambil 1000 data review dan dibagi menjadi data latih dan data uji berturut-turut sebanyak 700 dan 300. Berdasarkan hasil klasifikasi diperoleh tingkat akurasi sebesar 80,2%. Dari penelitian ini menunjukkan penggunaan algoritma klasifikasi *Naïve Bayes*

ditambah dengan pembobotan fitur TF-IDF memiliki performa yang cukup tinggi (Kosasih & Alberto, 2021).

Penelitian selanjutnya dilakukan oleh Pratiwi dan Adiwijaya (2018) mengenai penggunaan seleksi fitur pada sentimen analisis dalam sebuah *review* film. Dengan menggunakan seleksi fitur *Information Gain* mengurangi lebih dari 90% fitur yang tidak perlu sementara skema klasifikasi yang diusulkan mencapai akurasi klasifikasi sentimen 96% (Pratiwi & Adiwijaya, 2018). Dari penelitian ini menunjukkan penggunaan seleksi fitur *Information Gain* akan sangat membantu dalam menghilangkan fitur yang tidak relevan sehingga perhitungan algoritma klasifikasi dapat menunjukkan hasil yang tinggi (Pratiwi & Adiwijaya, 2018).

Penelitian serupa juga dilakukan oleh Ridho (2021). Metode *Support Vector Machine* digunakan untuk mengklasifikasi pengguna layanan indihome dengan kelas pengaduan layanan, pemasangan, tagihan, lambat, dan *disconnect*. Data yang digunakan adalah dari Twitter berupa teks tweet dengan total 1000 tweet, dan pemanfaatan fitur terbaik 80%, menghasilkan akurasi 88,9%, presisi 74,11, *recall* 69,22% dan *f-measure* 71,1%. % di kelas layanan (C1), akurasi 95%, presisi 87,2%, *recall* 87,8% *f-measure* 87,4% di kelas fasilitas (C2), akurasi 94,4%, presisi 85,67%, *recall* 86,55% *f-measure* 86,07% di kelas penagihan (C3), skor akurasi 93%, presisi 82,47%, *recall* 82,62%, *f-measure* 82,48% di kelas lambat (C4).

Analisis sentimen digunakan untuk mengklasifikasikan opini secara otomatis terhadap produk komersial atau isu politik berdasarkan sikap. Hal ini harus

dibedakan dari penempatan topik di mana teks diklasifikasikan berdasarkan satu atau beberapa topik. Pada bagian ini Jo (2019) menggambarkan analisis sentimental sebagai contoh khusus dari kategorisasi teks (Jo, 2019).

Menurut Jo (2019) terdapat tiga kategori dalam analisis sentimental. Positif berarti opini yang menggambarkan sesuatu atau seseorang dengan ekspresi positif seperti bagus, bagus, dan hebat. Netral berarti yang menggambarkan sesuatu secara objektif tanpa positif atau negatif, atau dengan campuran keduanya. Negatif berarti yang menggambarkan sesuatu dengan ekspresi negatif, seperti buruk, mengerikan, dan sedih. Netral dapat dibagi menjadi dua jenis: satu tanpa ekspresi sentimental dan satu dengan campuran positif dan negatif.

2.2 Text Preprocessing

Tahap *text preprocessing* dalam data mining secara umum dibutuhkan demi mendapatkan data yang siap pakai, akurat, bebas dari *noise*, *missing value* atau nilai yang hilang sehingga data dapat diprediksi dengan benar dan akurat (Mishra et al., 2020). Pada penelitian ini akan menggunakan 5 langkah *preprocessing*, yaitu *cleansing*, *lowercase*, *tokenizing*, *stopwords removal*, *stemming*.

Tahap *text preprocessing* diawali dengan mengambil data mentah dari sumber yang akan digunakan, dalam penelitian ini penulis menggunakan data dari twitter sebagai data sekunder. data mentah yang telah diambil kemudian dilakukan tahap *preprocessing* sehingga mengembalikan token atau kata yang telah dibersihkan dan siap untuk dipakai. Penggunaan tahap *preprocessing* juga diterapkan dalam penelitian yang dilakukan oleh Ismail Sunni (Sunni & Widyantoro, 2012) yaitu,

Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik menggunakan tahapan eksperimen praproses dengan melakukan klasifikasi dengan menggunakan metode praproses secara terpisah, salah satu metode yang diuji coba adalah *Convert Emoticon* bisa mendapatkan akurasi tertinggi yaitu 71.8%.

2.3 Information Gain

Information gain merupakan algoritma pemilihan fitur yang digunakan untuk mengurangi fitur-fitur yang diketahui tidak signifikan sehingga fitur-fitur yang digunakan pada proses klasifikasi selanjutnya menjadi signifikan atau berdampak besar pada pelabelan. (Putra *et al.*, 2016).

Information Gain mengukur berapa banyak informasi kehadiran dan ketidakhadiran dari suatu kata yang berperan untuk membuat keputusan klasifikasi yang benar dalam label apapun. Dalam menghitung *Information Gain* dibutuhkan nilai *entropy* dengan Persamaan 2.4 berikut:

$$Entropy(A) = \sum_i^c -p_i \log_2 p_i \quad (2.4)$$

Pada Persamaan 2.4 merupakan jumlah kelas klasifikasi, sedangkan $P(i)$ merupakan proporsi sampel pada kelas i dengan sampel data. Setelah nilai *entropy* ditemukan, kemudian menghitung nilai *Information Gain* dengan A merupakan atribut, dan v adalah suatu nilai yang mungkin untuk atribut A seperti pada Persamaan 2.5.

$$Inf.Gain(A) = Entropy(S) + p(A) \sum_j^v p(v|A) \log_2 p(v|A) + p(\bar{A}) \sum_j^v p(v|\bar{A}) \log_2 p(v|\bar{A}) \quad (2.5)$$

Pada Persamaan 2.5 penulis mengetahui bahwa j adalah himpunan nilai yang mungkin untuk atribut A , nilai v adalah kelas yang mungkin untuk atribut A . $p(v|A)$ merupakan jumlah proporsi dari sebuah dokumen yang berada di label kelas, sedangkan $p(v|\bar{A})$ merupakan jumlah proporsi dari sebuah dokumen yang tidak berada di label kelas.

Penelitian yang dilakukan oleh Shaltout *et al* (2014) tentang klasifikasi Influenza A berdasarkan inang virus menggunakan metode *Decision Tree* dan *Neural Network* serta menggunakan seleksi fitur *Information Gain*. Ukuran kinerja dari penelitian ini menghitung akurasi, sensitifitas, spesifisitas, presisi dan waktu. Didapati dengan menggunakan seleksi fitur *Information Gain* dapat meningkatkan efisiensi dari klasifikasi sebesar 90% untuk kedua pengklasifikasi, tanpa mengurangi kinerja secara signifikan (Shaltout *et al.*, 2014).

2.4 Naïve Bayes

Naïve Bayes adalah algoritma yang cukup kuat untuk mengklasifikasikan data berdasarkan probabilitas. Dalam banyak penelitian juga algoritma ini bekerja dengan sangat baik. *Naïve Bayes* hanya bekerja pada teorema Bayes dan menggunakan berbagai probabilitas untuk mengklasifikasikan data. Pada kelas *Naïve Bayes* dengan probabilitas maksimum dianggap sebagai kelas yang diprediksi. *Naïve Bayes* juga dikenal sebagai *Maximum a Posterior Naïve Bayes* memiliki berbagai kelebihan dan kekurangan di berbagai domain. *Naïve Bayes* adalah algoritma yang cepat dan sangat skalabel. Ini digunakan pada *Multiclass* dan *Binary Classification* (Ahuja *et al.*, 2019). Hal ini juga dapat digunakan pada dataset kecil dan dengan demikian juga memberikan hasil yang baik.

Naïve Bayes adalah algoritma klasifikasi yang menggunakan probabilitas untuk menghitung probabilitas suatu dokumen terpilih ke dalam kelas tertentu. Untuk menghitung probabilitas tergantung pada jumlah kelas dan jumlah fitur. Misalnya, ada tiga kelas; positif (C1), negatif (C2) dan netral (C3) dengan n fitur kata t_1, t_2, \dots, t_n (Kosasih & Alberto, 2021). Jika terdapat dokumen x , maka peluang x untuk masuk ketiga kelas dapat dihitung dengan mengkalikan probabilitas *prior* dari kelas ke n (C_n) dengan jumlah hasil pengkalian probabilitas kata ke n (t_n) yang diketahui memiliki kelas ke n (C_n) seperti pada Persamaan 2.6.

$$V(C_n|x) = P(C_n) * P(t_1|C_n) * P(t_2|C_n) * \dots * P(t_n|C_n) \quad (2.6)$$

Dimana :

$V(C_n|x)$ = Probabilitas dokumen x termasuk pada kelas C_n .

$P(C_n)$ = Probabilitas *prior* dari kelas C_n .

$P(t_n|C_n)$ = Probabilitas kata t_n yang diketahui memiliki kelas C_n .

Untuk mencari nilai dari probabilitas *prior* ($P(C_n)$) dan probabilitas kata ($P(t_n|C_n)$) menggunakan Persamaan 2.7 dan 2.8.

$$P(C_n) = \frac{N_c}{N} \quad (2.7)$$

$$P(t_n|C_n) = \frac{\text{count}(tn,c)+1}{\text{count}(c)+|V|} \quad (2.8)$$

Dimana :

N_c = Jumlah kelas c pada seluruh dokumen.

N = Jumlah seluruh dokumen.

$\text{count}(tn, c)$ = jumlah kata tn yang ditemukan pada seluruh dokumen dengan kelas c .

$count(c)$ = jumlah kata di seluruh dokumen dengan kelas c .

V = Jumlah total term pada seluruh dokumen.

Berdasarkan Persamaan 2.1 untuk mendapatkan probabilitas dokumen x termasuk kedalam kelas C_n dilakukan perhitungan dengan mengkalikan probabilitas *prior* kelas C_n dengan hasil total pengkalian probabilitas kata yang ada di dokumen x . Untuk mengetahui kelas dari dokumen x maka perlu dilakukan perhitungan probabilitas dokumen sebanyak jumlah kelas (C_n) kemudian kelas dari dokumen x akan diketahui dengan memilih probabilitas dengan nilai tertinggi.

2.5 Ensemble Learning

Ensemble Learning adalah paradigma pembelajaran mesin di mana banyak algoritma dasar dilatih untuk memecahkan masalah yang sama. Berbeda dengan pendekatan pembelajaran mesin biasa yang mencoba mempelajari satu hipotesis dari data pelatihan, metode *ensemble* mencoba membangun serangkaian hipotesis dan menggabungkannya untuk digunakan (Wang *et al.*, 2014).

Alur kerja khas *ensemble learning* adalah melatih satu model terlebih dahulu dan kemudian menggabungkannya melalui beberapa strategi, di mana suatu model biasanya dilatih oleh algoritma klasifikasi yang ada, seperti algoritma C4.5 dan algoritma BP *Neural Network*. Suatu *ensemble* dikatakan homogen jika semua model dilatih dengan algoritma yang sama, misalnya, "*decision tree ensemble*" hanya berisi pohon keputusan, sedangkan "*neural network ensemble*" hanya berisi jaringan saraf. Untuk *ensemble* homogen, model disebut *base learners*, dan algoritma pembelajaran yang sesuai disebut *base learning algorithms*. Sebaliknya,

ensemble heterogen berisi model dan algoritma pembelajaran yang berbeda, dan tidak ada model tunggal atau algoritma pembelajaran dasar. Untuk *ensemble* yang heterogen, model biasanya disebut *compoment learners* (Liu, 2021).

Berdasarkan penelitian yang dilakukan oleh Wang *et al* (2014), melakukan penilaian komparatif kinerja tiga metode *ensemble* populer (*Bagging*, *Boosting*, dan *Random Subspace*) berdasarkan lima metode dasar (*Naive Bayes*, *Maximum Entropy*, *Decision Tree*, *K Nearest Neighbor*, and *Support Vector Machine*) untuk klasifikasi sentimen. Selain itu, sepuluh dataset analisis sentimen publik diselidiki untuk memverifikasi efektivitas pembelajaran *ensemble* untuk analisis sentimen. Berdasarkan total 1200 percobaan kelompok komparatif, hasil empiris mengungkapkan bahwa metode *ensemble* secara substansial meningkatkan kinerja metode dasar untuk klasifikasi sentimen. Di antara tiga metode *ensemble*, *Random Subspace* memiliki hasil perbandingan yang lebih baik, meskipun jarang dibahas dalam literatur. Hasil ini menggambarkan bahwa metode pembelajaran *ensemble* dapat digunakan sebagai metode yang layak untuk klasifikasi sentimen. secara substansial meningkatkan kinerja metode dasar untuk klasifikasi sentimen (Wang *et al.*, 2014).

Penelitian yang dilakukan oleh Nai-Arun dan Punnee (2014) mengenai klasifikasi diabetes menggunakan metode *ensemble learning* menggunakan algoritma klasifikasi dasar yang menurutnya populer yaitu *Naïve Bayes*, *k-Nearest Neighbor*, dan *Decision Tree* kemudian jenis *ensamble learning* yang digunakan adalah *bagging* dan *boosting*. Hasil penelitian yang dilakukan oleh Nai-Arun dan Punnee menunjukkan bahwa model terbaik dengan akurasi tertinggi adalah *bagging*

dengan algoritma klasifikasi dasar *Decision Tree* sebesar 95,312% dibandingkan dengan hanya menggunakan algoritma dasarnya saja sebesar 94,621%. Dari penelitian ini menunjukkan bahwa pengaplikasian metode *bagging ensemble learning* dapat meningkatkan performa dari algoritma klasifikasi dasar (Nai-Arun & Punnee, 2014).

2.6 *Bagging Ensemble Classifier*

Bagging (kependekan dari *bootstrap aggregating*) adalah salah satu algoritma pembelajaran *ensemble* yang paling awal. Algoritma ini juga salah satu yang paling intuitif dan paling sederhana untuk diterapkan, dengan kinerja yang sangat bagus. Keragaman dalam *Bagging* diperoleh dengan menggunakan replika *bootstrap* dari data pelatihan, subset data pelatihan yang berbeda diambil secara acak dengan penggantian dari seluruh dataset pelatihan. Setiap subset data pelatihan digunakan untuk melatih model dasar yang berbeda dari tipe yang sama. Strategi kombinasi metode dasar untuk *Bagging* adalah *majority voting* (Wang *et al.*, 2014).

Penelitian dengan topik klasifikasi menggunakan metode *Bagging Ensemble Classifier* dilakukan oleh Subasi *et al* (2020) mengenai Klasifikasi Data Kardiotokogram untuk Antisipasi Risiko Janin Menggunakan *Bagging Ensemble Classifier*. Tujuan penelitian ini untuk mengevaluasi performansi klasifikasi algoritma *machine learning ensemble* pada data CTG *antepartum*. Oleh karena itu, makalah ini difokuskan pada algoritma pembelajaran mesin *Ensemble Bagging* untuk mengklasifikasikan sinyal detak jantung janin sebagai normal atau abnormal. Akurasi, *F-measure* dan area ROC digunakan sebagai metrik kinerja

untuk menilai keberhasilan pengklasifikasi. Hasil eksperimen menunjukkan bahwa pengklasifikasi *Ensemble Bagging* menghasilkan hasil yang memuaskan, dan *Bagging* dengan *Random Forest* mencapai hasil yang lebih baik dengan akurasi 99,02% (Subasi et al., 2020).

BAB III

METODOLOGI PENELITIAN

3.1 Akuisisi Data

Penulis menggunakan data primer berupa dataset komentar video yang didapatkan dari platform Youtube. Dataset tersebut berisi komentar masyarakat diambil dari video pemberitaan mengenai Covid dan kebijakan pemerintah mengenai Covid. Dataset komentar yang baru didapatkan masih berstatus data sekunder karena penulis belum mengolah data tersebut. Data sekunder akan diubah menjadi data primer dengan dilabeli manual oleh tenaga ahli. Data primer digunakan sebagai masukan untuk membangun model klasifikasi dan menguji proses klasifikasi.

Tahap pengumpulan data diambil dari Youtube API sebanyak 3 video dengan kata kunci “covid” dan “pemerintah”. Dalam tahap pengumpulan data dilakukan dengan menggunakan bahasa pemrograman python dan hasil dari pengumpulan data yaitu berupa *file* dengan tipe csv yang berisikan komentar dan nama akun yang memberikan komentar.

Data penelitian berjumlah 191 yang terakumulasi sejak Juli 2022. Data penelitian dibagi menjadi data latih dan data uji, data latih nantinya akan diberikan 3 kelas label sesuai dengan emosi manusia yaitu positif (C1), negatif (C2), dan netral (C3).

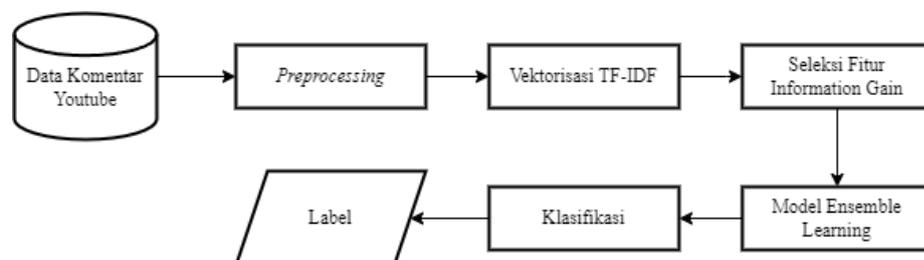
Berikut contoh data yang telah diambil dari Youtube API dan telah dilabeli secara manual seperti pada Tabel 3.1.

Tabel 3.1 Contoh data teks dengan label

Dokumen	Komentar	Label
D1	Apa yang dilakukan pemerintah, sebetulnya sudah cukup baik, terlepas dari kamu pro/kontra dengan jokowi (karna cuma 2 kubu)	Positif(C1)
D2	Miris ngelihat perkembangan Covid-19 di Indonesia yang ngga kunjung mereda	Negatif (C2)
D3	emosi bgt, lg kondisi begini gue disuruh liputan tentang Covid. ini dosen gue bercanda ga si?	Negatif (C2)
D4	Heran dengan carut marutnya pemerintah yg semakin gak sinkron satu sm lain.	Negatif (C2)
D5	Bikin was-was banget berita Covid-19 akhir” ini	Negatif (C2)
D6	penyaluran bantuan beras warga terdampak pandemi Covid 19 pemerintah kabupaten kediri	Netral(C3)

3.2 Perancangan Sistem

Diagram pada Gambar 3.1 merupakan tahapan yang akan dilakukan penulis untuk menuntaskan penelitian, diantara lain *preprocessing*, pembobotan TF-IDF, seleksi fitur, penerapan metode klasifikasi, kemudian pengujian.



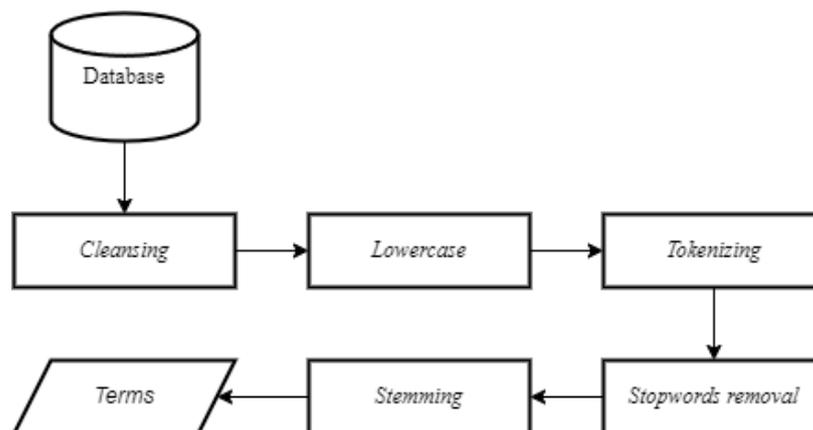
Gambar 3.1 Alur Sistem Klasifikasi

Berdasarkan Gambar 3.1 pada tahapan pertama dari penelitian ini mengubah data yang tidak terstruktur menjadi terstruktur pada tahap *preprocessing*, kemudian

melakukan pembobotan pada setiap *term* menggunakan pembobotan TF-IDF, data teks yang ditokenisasi kemudian diseleksi menggunakan seleksi fitur *Information Gain*, setelah itu dilakukan pemodelan menggunakan metode *Bagging Ensemble Classifier* dengan algoritma klasifikasi dasar klasifikasi *Naïve Bayes*. Dari model yang telah dibuat kemudian dilakukan pengujian untuk mengetahui pengaruh metode *Bagging Ensemble Classifier* dengan algoritma dasar *Naïve Bayes* dan akurasi, presisi, *recall*, dan *f-measure* sebagai output dari penelitian ini.

3.2.1. Preprocessing

Tahap *Preprocessing* pada penelitian ini mengubah data teks mentah yang umumnya menggunakan kata yang tidak baku menjadi data teks yang bersih dan telah ditokenisasi. Sebagai contoh proses pre-processing maka data pada Tabel 3.1 dokumen 1 (D1) akan digunakan untuk tahap *preprocessing*. Tahapan yang akan dilakukan sebagai berikut seperti Gambar 3.2



Gambar 3.2 Alur *Preprocessing*

3.2.1.1. *Cleansing*

Cleansing secara Bahasa berarti membersihkan, dalam proses *cleansing* data teks akan dibersihkan dari kata atau data teks yang tidak diperlukan seperti tautan, *mention*, *hashtag*, dan *emoticon*.

Berikut contoh hasil tahap *cleansing* dari dokumen D1 yang akan ditunjukkan pada Tabel 3.2

Tabel 3.2 Hasil tahap *Cleansing*

Sebelum <i>Cleansing</i>	Setelah <i>Cleansing</i>
Apa yang dilakukan pemerintah, sebetulnya sudah cukup baik, terlepas dari kamu pro/kontra dengan jokowi (karna cuma 2 kubu)	Apa yang dilakukan pemerintah sebetulnya sudah cukup baik terlepas dari kamu pro kontra dengan Jokowi karna cuma 2 kubu

3.2.1.2. *Lowercase*

Tahapan *lowercase* mengubah data hasil *cleansing* yang semula memiliki huruf kapital menjadi huruf kecil secara keseluruhan. Tahapan ini diperlukan karena dalam program, kata “Aku” dan “aku” dianggap berbeda atau bisa diartikan *Case Sensitive*.

Berikut contoh hasil tahap *lowecase* dari dokumen D1 yang akan ditunjukkan pada Tabel 3.3

Tabel 3.3 Hasil tahap *Lowecase*

Sebelum <i>Lowercase</i>	Setelah <i>Lowercase</i>
Apa yang dilakukan pemerintah sebetulnya sudah cukup baik terlepas dari kamu pro kontra dengan Jokowi karna cuma 2 kubu	apa yang dilakukan pemerintah sebetulnya sudah cukup baik terlepas dari kamu pro kontra dengan jokowi karna cuma 2 kubu

3.2.1.3. *Tokenizing*

Umumnya dalam tahapan *preprocessing* sebuah analisis sentimen selalu dilakukan tahapan *tokenizing*. *Tokenizing* sendiri merupakan tahapan untuk

memecah sebuah kalimat menjadi kata atau token. Hasil dari proses ini berupa list atau array yang berisi kata dari kalimat sebelumnya.

Berikut contoh hasil tahap *tokenizing* dari dokumen D1 yang akan ditunjukkan pada Tabel 3.4

Tabel 3.4 Hasil tahap *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
apa yang dilakukan pemerintah sebetulnya sudah cukup baik terlepas dari kamu pro kontra dengan jokowi karna cuma 2 kubu	["apa", "yang", "dilakukan", "pemerintah", "sebetulnya", "sudah", "cukup", "baik", "terlepas", "dari", "kamu", "pro", "kontra", "dengan", "Jokowi", "karna", "cuma", "2", "kubu"]

3.2.1.4. *Stopwords Removal*

Stopwords dalam Bahasa berarti kata-kata penghenti, kata penghenti adalah kata umum yang sering kali muncul dalam kalimat yang memiliki fungsi tetapi tidak memiliki arti. Karakteristik umum dari *stopwords* seringkali adalah kata yang memiliki frekuensi kemunculan dalam kalimat yang cukup tinggi seperti artikel, konjungsi, dan preposisi. Karena dianggap tidak memiliki arti maka kata-kata penghenti tersebut harus dihilangkan karena dianggap tidak relevan.

Berikut contoh hasil tahap *stopword removal* dari dokumen D1 yang akan ditunjukkan pada Tabel 3.5

Tabel 3.5 Hasil tahap *Tokenizing*

Sebelum <i>Stopwords Removal</i>	Setelah <i>Stopwords Removal</i>
["apa", "yang", "dilakukan", "pemerintah", "sebetulnya", "sudah", "cukup", "baik", "terlepas", "dari", "kamu", "pro", "kontra", "dengan", "Jokowi", "karna", "cuma", "2", "kubu"]	["dilakukan", "pemerintah", "cukup", "baik", "pro", "kontra", "kubu"]

3.2.1.5. Stemming

Stemming adalah proses untuk mengganti sebuah kata yang sebelumnya memiliki kata imbuhan seperti “mem-”, “di-”, “ter-”, “-kan” menjadi kata dasar. Pada tahapannya data teks yang sebelumnya sudah ditokenisasi akan dihilangkan *sufiks*, *prefiks* dan *konfiks*.

Berikut contoh hasil tahap *stemming* dari dokumen D1 yang akan ditunjukkan pada Tabel 3.6

Tabel 3.6 Hasil tahap *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
[“dilakukan”, “pemerintah”, “cukup”, “baik”, “pro”, “kontra”, “kubu”]	[“lakukan”, “pemerintah”, “cukup”, “baik”, “pro”, “kontra”, “kubu”]

3.2.1.6. Terms

Setelah melewati beberapa tahap dalam proses *preprocessing* maka hasil akhir dari proses ini adalah hasil dari tahap akhir dari *preprocessing* yaitu *stemming* pada Tabel 3.7.

Tabel 3.7 Hasil *preprocessing* Dokumen 1 (D1)

Keterangan	Terms
<i>Cleansing</i>	Apa yang dilakukan pemerintah sebetulnya sudah cukup baik terlepas dari kamu pro kontra dengan Jokowi karna cuma 2 kubu
<i>Lowercase</i>	apa yang dilakukan pemerintah sebetulnya sudah cukup baik terlepas dari kamu pro kontra dengan jokowi karna cuma 2 kubu
<i>Tokenizing</i>	[“apa”, “yang”, “dilakukan”, “pemerintah”, “sebetulnya”, “sudah”, “cukup”, “baik”, “terlepas”, “dari”, “kamu”, “pro”, “kontra”, “dengan”, “Jokowi”, “karna”, “cuma”, “2”, “kubu”]
<i>Stopwords Removal</i>	[“dilakukan”, “pemerintah”, “cukup”, “baik”, “pro”, “kontra”, “kubu”]
<i>Stemming</i>	[“lakukan”, “pemerintah”, “cukup”, “baik”, “pro”, “kontra”, “kubu”]

3.2.2. Pembobotan TF-IDF

Data yang telah didapatkan dari proses *preprocessing* akan dijadikan *term* pada proses ini yaitu pembobotan TF-IDF. Sebelumnya data tersebut akan dihitung tingkat kemunculan di setiap dokumen yang akan ditunjukkan pada Tabel 3.8 sebagai berikut.

Tabel 3.8 kata muncul dalam dokumen

<i>Term</i>	Jumlah Kata					
	D1	D2	D3	D4	D5	D6
lakukan	1	0	0	0	0	0
pemerintah	1	0	0	1	0	1
cukup	1	0	0	0	0	0
baik	1	0	0	0	0	0
pro	1	0	0	0	0	0
kontra	1	0	0	0	0	0
kubu	1	0	0	0	0	0
miris	0	1	0	0	0	0
lihat	0	1	0	0	0	0
berkembang	0	1	0	0	0	0
Covid	0	1	1	0	1	1
tidak	0	1	1	1	0	0
reda	0	1	0	0	0	0
emosi	0	0	1	0	0	0
kondisi	0	0	1	0	0	0
suruh	0	0	1	0	0	0
liputan	0	0	1	0	0	0
bercanda	0	0	1	0	0	0
heran	0	0	0	1	0	0
carut	0	0	0	1	0	0
marut	0	0	0	1	0	0
sinkron	0	0	0	1	0	0
waswas	0	0	0	0	1	0
berita	0	0	0	0	1	0
salur	0	0	0	0	0	1
bantu	0	0	0	0	0	1
beras	0	0	0	0	0	1
warga	0	0	0	0	0	1
dampak	0	0	0	0	0	1
pandemi	0	0	0	0	0	1
Total	7	6	7	6	3	8

Langkah selanjutnya yaitu menghitung *Term Frequency* (TF) dari *term* yang muncul dalam sebuah dokumen dengan menggunakan Persamaan 3.1.

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

Berikut contoh perhitungan *Term Frequency* dari kata “pemerintah” pada Tabel 3.9.

Tabel 3.9 Perhitungan *Term Frequency*

Dokumen	Term	Jumlah terms	Frekuensi	Term Frequency (TF)
D1	pemerintah	7	1	1/7
D2		6	0	0/6
D3		6	0	0/7
D4		6	1	1/6
D5		3	0	0/3
D6		8	1	1/8

Selanjutnya menghitung *Inverse Document Frequency* (IDF) dengan menggunakan Persamaan 3.2.

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

Contoh perhitungan IDF pada kata “pemerintah” pada Tabel 3.10.

Tabel 3.10 Perhitungan IDF

Dokumen	Term	Frekuensi	IDF
D1	pemerintah	1	$\log \frac{6}{3} + 1 = 1,301$
D2		0	
D3		0	
D4		1	
D5		0	
D6		1	

Setelah nilai dari TF dan IDF ditemukan maka akan lebih mudah untuk menghitung TF-IDF dengan Persamaan 3.3.

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

Berikut merupakan contoh hasil perhitungan TF-IDF pada kata “pemerintah” pada Tabel 3.11.

Tabel 3.11 Hasil TF-IDF

Dokumen	Term	TF	IDF	TF-IDF
D1	pemerintah	1/7	1,301	0,08601
D2		0/6		0
D3		0/6		0
D4		1/6		0,10034
D5		0/3		0
D6		1/8		0,16262

3.2.3. Seleksi Fitur *Information Gain*

Tahap seleksi fitur *Information Gain* akan menurunkan jumlah fitur yang tidak relevan untuk tahap perhitungan selanjutnya. Untuk mempermudah perhitungan selanjutnya, *Document Frequency* dari setiap kata sangat dibutuhkan pada setiap dokumen di seluruh kelas. *Document Frequency* dari setiap kata akan ditampilkan pada Tabel 3.12.

Tabel 3.12 Document Frequency

Term	DF	Kata Muncul (Kelas)			Kata Tidak Muncul (Kelas)		
		Positif	Negatif	Netral	Positif	Negatif	Netral
pemerintah	3	1	1	1	0	3	0
Covid	4	0	3	1	1	1	0
pandemi	1	0	0	1	1	4	0
bercanda	1	0	1	0	1	3	1
emosi	1	0	1	0	1	3	1
heran	1	0	1	0	1	3	1
sinkron	1	0	1	0	1	3	1
waswas	1	0	1	0	1	3	1
...
Total	13	1	9	3	7	23	5

Langkah berikutnya menghitung nilai *entropy* dari setiap kata. Berikut contoh perhitungan *entropy* dari kata “pemerintah” menggunakan Persamaan 2.4.

$$\begin{aligned} \text{Entropy (pemerintah)} &= -\left(\left(\frac{3}{6}\right) \log\left(\frac{3}{6}\right)\right) \\ &= 0,1505 \end{aligned}$$

Setelah mendapatkan nilai *entropy* selanjutnya menghitung nilai *Information Gain* dari kata “pemerintah” berdasarkan Persamaan 2.5

$$\begin{aligned} \text{Information Gain (pemerintah)} &= 0,1505 + \frac{3}{13} \times \left(\left(\frac{1}{6}\right) \log\left(\frac{1}{6}\right) + \right. \\ &\quad \left. \left(\frac{1}{6}\right) \log\left(\frac{1}{6}\right) + \left(\frac{1}{6}\right) \log\left(\frac{1}{6}\right)\right) + \\ &\quad \frac{3}{13} \times \left(0 + \left(\frac{3}{6}\right) \log\left(\frac{3}{6}\right) + 0\right) \\ &= 0,1505 + (-0,0897423) + (-0,034720) \end{aligned}$$

$$\text{Information Gain (pemerintah)} = 0,0268$$

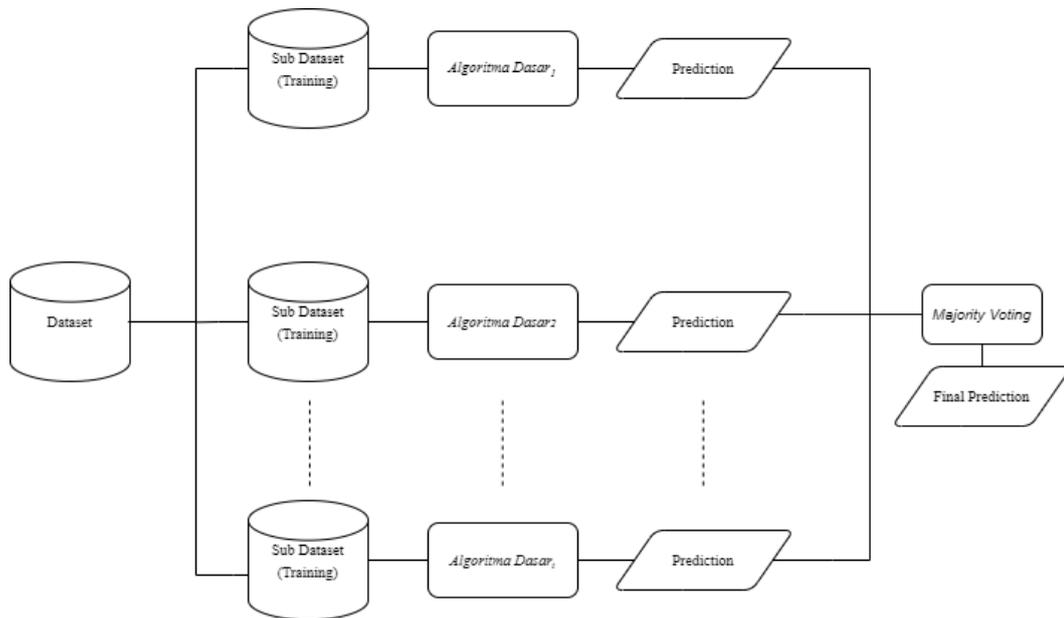
Ulangi perhitungan *entropy* dan *Information Gain* pada seluruh kata sesuai pada contoh perhitungan pada kata “pemerintah”. Setelah seluruh kata telah memiliki nilai *Information Gain*, selanjutnya akan diurutkan secara *descending* sehingga *term* dengan nilai *Information Gain* tertinggi akan berada pada urutan teratas seperti pada Tabel 3.13.

Tabel 3.13 *Information Gain* Terbesar sampai Terkecil

No	Term	Information Gain
1	Covid	0,1176
2	pemerintah	0,0268
3	pandemi	-0,0727
4	bercanda	-0,0727
5	emosi	-0,0727
6	heran	-0,0727
7	sinkron	-0,0727
8	waswas	-0,0727

3.2.4. *Bagging Ensemble Classifier*

Ide dari model model *ensemble* adalah untuk menghasilkan beberapa versi prediktor (pengklasifikasi dasar) dan menggunakan hasil ini untuk mendapatkan prediktor agregat. *Output* dari model ini digabungkan dengan cara *majority voting* untuk membuat keputusan *output* terakhir. Model ini juga mengurangi tingkat varian dan membantu menghindari *overfitting*. Hal ini biasanya diterapkan pada algoritma *Decision Tree*, namun dapat digunakan dengan algoritma lain. *Bagging Ensemble Classifier* memiliki alur seperti Gambar 3.3 dibawah.



Gambar 3.3 Alur Metode Ensemble Learning

3.2.4.1. *Bootstrap Aggregating (Bagging)*

Terdapat tiga tahap dalam proses perhitungan menggunakan *bagging* diantaranya yang pertama pembagian data training, kedua pemodelan menggunakan metode klasifikasi dasar, dan terakhir *majority voting*. Dataset dibagi menjadi 2 yaitu data latih dan data uji, sub dataset yang akan digunakan untuk menguji model dasar diambil dari data training yang dipilih dengan cara *random sampling* yang masing-masing fitur memiliki probabilitas yang sama untuk masuk ke sub dataset. Tahap pertama dan kedua akan diulangi sebanyak N kali dan hasil dari pengujian setiap tahap klasifikasi algoritma dasar akan di *voting* untuk mendapatkan hasil prediksi utama.

3.2.4.2. *Klasifikasi Naïve Bayes*

Pada tahap klasifikasi, algoritma dasar menggunakan algoritma Naive Bayes yang terdiri dari dua tahap yaitu tahap pelatihan dan tahap pengujian. Pada

tahap pelatihan, dilakukan proses pelatihan terhadap data pelatihan yang diperoleh dengan proses bagging menggunakan algoritma Naive Bayes, Ketika proses latihan selesai selanjutnya akan dilakukan tahap uji dengan memasukan data uji yang sudah diberikan kedalam data yang sudah terlatih untuk diklasifikasi.

Langkah pertama adalah menghitung frekuensi kata individual di setiap dokumen untuk menentukan frekuensi kata (term frequency). Karena pada langkah sebelumnya kita menghitung TF-IDF dan Information Gain untuk setiap kata , maka pada tahap training sebagai contoh akan menggunakan 30% dari data pada Tabel 3.3 yaitu sebanyak 8 fitur yang akan ditunjukkan pada Tabel 3.14.

Tabel 3.14 *Term Frequency* setiap kata

No.	Term	Positif	Negatif	Netral
1.	Covid	0	3	1
2.	pemerintah	1	1	1
3.	pandemi	0	0	1
4.	bercanda	0	1	0
5.	emosi	0	1	0
6.	heran	0	1	0
7.	sinkron	0	1	0
8.	waswas	0	1	0

Kemudian menghitung probabilitas *prior* dari setiap kelas yang ditunjukkan pada tabel 3.15.

Tabel 3.15 Perhitungan Probabilitas *Prior*

Kelas	Probabilitas
Positif	1/6
Negatif	4/6
Netral	1/6

Selanjutnya menghitung probabilitas yang dimiliki oleh setiap kata menggunakan Persamaan 2.8 sebagai berikut.

$$P(\text{pemerintah}|\text{positif}) = \frac{|1 + 1|}{|1 + 8|} = 0,11111$$

Karena dalam penelitian ini menggunakan pembobotan kata TF-IDF, maka dalam penghitungan probabilitas term menggunakan persamaan multinomial seperti pada Persamaan 3.10.

$$P(t_n|C_n) = \frac{W_{ct}+1}{\sum W' \in V W'_{ct} + B'} \quad (3.10)$$

Dimana:

W_{ct} = Nilai bobot term t pada kelas c

$\sum W' \in V W'_{ct}$ = Jumlah total nilai W dari keseluruhan term yang memiliki kelas c.

B' = jumlah nilai IDF pada seluruh dokumen.

$$P(\text{pemerintah}|\text{positif}) = \frac{0,25402 + 1}{1,77815 + 51,7882}$$

$$P(\text{pemerintah}|\text{positif}) = 51,92757$$

Tabel 3.16 Hasil Perhitungan Probabilitas Setiap Kata

No.	Term	Positif	Negatif	Netral
1.	Covid	52,35062	52,10354	51,98353
2.	pemerintah	51,92757	52,01263	51,98353
3.	pandemi	51,89935	51,96717	51,98353
4.	bercanda	51,89935	52,01263	51,85853
5.	emosi	51,89935	52,01263	51,85853
6.	heran	51,89935	52,01263	51,85853
7.	sinkron	51,89935	52,01263	51,85853
8.	waswas	51,89935	52,01263	51,85853

Tabel 3.16 adalah hasil probabilitas setiap kata. Hasil perhitungan probabilitas setiap kata kemudian digunakan pada tahap pengujian sebagai acuan dalam proses klasifikasi.

Pada proses pengujian data uji yang telah diberikan akan dimasukkan kedalam data yang sudah berlabel, pada prosesnya sendiri setiap kosa kata (fitur) dari data testing akan dijumlah kemunculan dari setiap kata, apabila kata dari data uji tidak ada dalam data pelatihan maka kata tersebut tidak akan digunakan dalam proses pengujian. Berikut contoh data uji pada tabel 3.17 dan hasil pengujian dari data uji pada Tabel 3.18.

Tabel 3.17 Contoh Data Uji

No.	Komentar	Label
1.	Emosi banget liat orang2 ga patuh udah tau Covid masi aja keluyuran, heran.	?

Tabel 3.18 Perhitungan Probabilitas Kata Uji

No.	Term	Positif	Negatif	Netral
1.	Covid	52,35062	52,10354	51,98353
2.	emosi	51,89935	52,01263	51,85853
3.	heran	51,89935	52,01263	51,85853

Untuk memperoleh hasil klasifikasi yaitu nilai total probabilitas tertinggi akan menggunakan Persamaan 2.6 sebagai berikut.

$$V_j(\text{positif}) = P(\text{positif}) * P(\text{covid}|\text{positif}) * P(\text{emosi}|\text{positif}) \\ * P(\text{heran}|\text{positif})$$

$$V_j(\text{positif}) = 0,1666 * 52,35062 * 51,89935 * 51,89935$$

$$V_j(\text{positif}) = 23501,43$$

Tabel 3.19 Probabilitas Akhir

Kelas	Probabilitas Akhir
Positif	23501,43
Negatif	93970,94
Netral	23299,95

Berdasarkan perhitungan diatas menghasilkan probabilitas akhir pada Tabel 3.19 dengan kelas positif dengan nilai probabilitas tertinggi yaitu 93970,94 jika dibulatkan. Contoh data uji yang diberikan dengan begitu memiliki label negatif.

3.2.4.3. Majority Voting

Setelah perhitungan pengklasifikasi dasar terakhir selesai, selanjutnya hasil klasifikasi label akan dikumpulkan dan dilakukan *majority voting* untuk menentukan label final dari data uji. Dengan kata lain hasil perhitungan pengklasifikasi dasar yang dilakukan sebanyak N kali dengan sub data latih yang dibagikan melalui proses *bagging* akan dipilih label dengan kuantitas terbanyak. dengan Persamaan 3.11 berikut.

$$H(x) = \arg \max_{y \in Y} \sum_{t=1}^T 1(y = h_t(x)) \quad (3.11)$$

Masukan dari persamaan diatas ada D sebagai dataset, L sebagai algoritma klasifikasi dasar, dan T adalah berapa kali dataset akan di *Bagging*. Berdasarkan persamaan diatas h merupakan model klasifikasi dari algoritma dasar dengan D_t sebagai dataset.

Tabel 3.20 Hasil *Bagging* pada setiap index perulangan

t	Kelas
1	Negatif
2	Positif
3	Netral
4	Negatif

Berdasarkan Tabel 3.20 Proses *Bagging Ensemble Learning* dilakukan sebanyak T = 4 dengan hasil masing-masing perulangan adalah Negatif, Positif, Netral, Negatif dengan begitu bisa dikatakan data test yang diklasifikasikan

dengan *Bagging Ensemble Learning* menggunakan metode dasar *Naïve Bayes* memiliki label negatif.

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Langkah-langkah Uji Coba

Langkah-langkah uji coba pada sistem yang telah dibuat adalah sebagai berikut:

1. *Input dataset*

Dataset yang terkumpul dari komentar youtube sebanyak 192 komentar yang disimpan dalam file berformat *csv*. *Dataset* terdiri dari 56 komentar dengan label positif yang ditandai dengan angka 1, 88 komentar dengan label negative yang ditandai dengan angka 2, dan 48 komentar berlabel netral yang ditandai dengan angka 3 seperti yang ditunjukkan pada Tabel 4.1.

Tabel 4.1 Sampel dataset

No	Komentar	Label
1	"SEBULAN SETAHUN PANDEMI COVID 19"	1
2	"Zehahahahaha"	3
3	"Senetron suda mulai tayang teman teman.. ada yg mau nonton..ini sinetron baru loh.."	3
4	"Bosen"	2
5	"Tolong pemerintah segera lakukan lockdown untuk mencegah penularan lebih banyak."	1
6	"Hadeeuhhhhh"	3
7	"SEMOGA SLLU DILINDUNGI D BERI KSEHATAN SELALU AMIIN DIJAUHKAN DRI SAKIT INI 🙏🙏 AMIIN"	1
8	"Pergantian cuaca... Di libatkan"	1
9	"berita gk bermutu terus bermunculan ,,celek ngasuuu"	2
10	"Kantor pemerintah di sidak dan swiping krn ikut dlm persekongkolan dan utang negara yg rakyat wajib utuk membayar.."	3

2. Pembagian *dataset*

Sebelum masuk dalam pengujian, pertama kali data akan dibagi menjadi 2 yaitu data latih dan data uji dengan prosentase pada Tabel 4.2.

Tabel 4.2 Pembagian Data

Data Latih	Data Uji
80%	20%
70%	30%

Pembagian dataset pada Tabel 4.2 diambil dengan tujuan menganalisis manakah dari pembagian dataset tersebut yang memiliki performa terbaik merujuk pada penelitian yang dilakukan oleh Sriyano dan Setiawan (2021) yang memiliki performa terbaik dengan menggunakan data latih sebesar 80% pada pendeteksi hoax menggunakan Naive Bayes dengan pembobotan TF-IDF (Sriyano & Setiawan, 2021) dan penelitian yang dilakukan oleh Kosasih dan Alberto (2021) yaitu Analisis sentimen produk game di shopee menggunakan metode TF-IDF dan Pengklasifikasi *Naïve Bayes* yang menggunakan perbandingan data latih 70% dan data uji 30% (Kosasih & Alberto, 2021).

3. Proses klasifikasi

Pada tahap ini akan dilakukan klasifikasi menggunakan metode *Bagging Ensemble Classifier* dengan metode dasar *Naïve Bayes* sebagaimana telah dijabarkan pada bab 3.

4. Menampilkan hasil klasifikasi

Tahap terakhir adalah menampilkan hasil klasifikasi yang berupa hasil klasifikasi kelas dan kelas aktual berdasarkan persentase pembagian

fitur dengan skenario fitur 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%.

Tabel 4.3 Sampel hasil klasifikasi

No	Aktual	Jumlah Fitur								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
1	1	1	1	1	1	1	1	1	1	1
2	3	3	3	3	3	3	3	3	3	3
3	3	2	2	2	2	2	2	2	2	2
4	2	3	2	2	2	2	2	2	2	2
5	1	2	2	2	2	2	2	2	2	2
6	3	3	3	3	3	3	3	3	3	3
7	1	1	1	1	1	1	1	1	1	1
8	1	3	2	1	1	1	1	1	1	1
9	2	2	2	2	1	1	1	1	1	1
10	3	2	2	2	2	2	2	2	2	2

5. Evaluasi Sistem

Pada tahap evaluasi sistem akan dilakukan perhitungan untuk mengetahui nilai akurasi, presisi, *recall*, dan *F-measure*. Nilai-nilai tersebut dapat dihasilkan dengan membandingkan kelas hasil klasifikasi dengan kelas aktual menggunakan *confusion marix* sehingga dapat diketahui TP, TN, FP, dan FN sebagai berikut.

1. TP (*True Positive*) merupakan data aktual berlabel positif yang diklasifikasikan sebagai positif.
2. TN (*True Negative*) merupakan data aktual berlabel negative yang diklasifikasikan sebagai negative.
3. FP (*False Positive*) merupakan data aktual berlabel positif yang diklasifikasikan sebagai negative.

4. FN (*False Negative*) merupakan data aktual berlabel negative yang diklasifikasikan sebagai positif.

Berikut persamaan untuk menghitung akurasi, *recall*, presisi, dan *F-measure*.

1. Akurasi

Akurasi adalah skor seberapa mirip nilai prediksi dengan nilai aktual. Nilai akurasi ditujukan untuk mengukur seberapa besar tingkat kesalahan yang dapat terjadi dalam pengujian. Berikut persamaan untuk menghitung nilai akurasi seperti pada Persamaan 4.1

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.1)$$

2. Presisi

Presisi adalah skor seberapa tepat prediksi yang dilakukan oleh sistem dengan menggunakan perulangan yang dilakukan dengan cara yang sama. Oleh karena itu nilai presisi diperlukan untuk mengukur berapa persen prediksi yang tepat sasaran. Berikut persamaan untuk menghitung nilai presisi seperti pada Persamaan 4.2

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (4.2)$$

3. *Recall*

Recall adalah skor tingkat seberapa sukses sistem untuk menemukan kembali informasi. Berbeda dengan nilai presisi yang mencocokkan nilai aktual dengan nilai prediksi, *recall* mencocokkan nilai

yang diprediksi dengan nilai aktualnya. Berikut persamaan untuk menghitung nilai *recall* seperti pada Persamaan 4.3

$$Recall = \frac{TP}{TP+FP} \quad (4.3)$$

4. *F-Measure*

Nilai *f-measure* merupakan hasil perhitungan evaluasi dengan mengkombinasikan nilai *recall* dan *presisi*. Nilai *f-measure* berfungsi untuk menentukan tingkat efektivitas suatu pengujian. Pengukuran *f-measure* dapat dihitung menggunakan Persamaan 4.4.

$$F\text{-measure} = 2 \times \frac{\text{presisi} \times \text{recall}}{\text{presisi} + \text{recall}} \quad (4.4)$$

4.2 Hasil Uji Coba

Pada sub bab ini akan memfokuskan hasil uji coba sesuai dengan skenario yang telah dijelaskan pada sub bab 4.1. Setelah dilakukan klasifikasi dan membandingkan hasil klasifikasi dengan data aktual menggunakan *confusion matrix*, maka didapatkan hasil sebagai berikut pada Tabel 4.4:

Tabel 4.4 *Confusion matrix* hasil klasifikasi dengan fitur 20%

Prediksi \ Aktual	Positif	Negatif	Netral
Positif	5	8	3
Negatif	2	9	0
Netral	0	4	8

Hasil *confusion matrix* pada Tabel 4.4 dapat diketahui nilai TP, TN, FP, dan FN dengan melakukan perhitungan pada kelas Positif sebagai berikut.

$$TP = 5$$

$$TN = (9 + 0) + (4 + 8) = 21$$

$$FP = 2 + 0 = 2$$

$$FN = 8 + 3 = 11$$

Tabel 4.5 Hasil perhitungan TP, TN, FP, FN

Jumlah Fitur	Kelas	TP	TN	FP	FN
20%	Positif	5	21	2	11
	Negatif	9	13	12	2
	Netral	8	24	3	4

Berdasarkan hasil perhitungan TP, TN, FP, dan FN, maka dapat dilakukan perhitungan akurasi, presisi, *recall*, dan *f-measure* pada masing-masing kelas sebagai berikut.

a. Kelas Positif

$$Akurasi = \frac{5 + 21}{39} 100\% = 67\%$$

$$Presisi = \frac{5}{5 + 2} 100\% = 71,4\%$$

$$Recall = \frac{5}{5 + 11} 100\% = 31,25\%$$

$$f - measure = 2 \frac{71,4 \times 31,25}{71,4 + 31,25} = 43,47\%$$

b. Kelas Negatif

$$Akurasi = \frac{9 + 13}{39} 100\% = 56,4\%$$

$$Presisi = \frac{9}{9 + 12} 100\% = 42,8\%$$

$$Recall = \frac{9}{9 + 2} 100\% = 81,81\%$$

$$f - measure = 2 \frac{42,8 \times 81,81}{42,8 + 81,81} = 56,2\%$$

c. Kelas Netral

$$Akurasi = \frac{8 + 24}{39} 100\% = 82\%$$

$$Presisi = \frac{8}{8 + 3} 100\% = 72,7\%$$

$$Recall = \frac{8}{8 + 4} 100\% = 67\%$$

$$f - measure = 2 \frac{72,7 \times 67}{72,7 + 67} = 69,73\%$$

Nilai presisi, *recall*, dan *f-measure* dari seluruh kelas bisa didapatkan dengan menghitung rata-rata dari masing-masing nilai dan nilai akurasi didapatkan dengan menghitung nilai TP seluruh kelas sebagai berikut.

$$Akurasi = \frac{5 + 9 + 8}{39} 100\% = 56,4\%$$

$$Presisi = \frac{71,4 + 42,8 + 72,7}{3} = 62,3\%$$

$$Recall = \frac{31,25 + 81,81 + 67}{3} = 60,02\%$$

$$f - measure = \frac{43,47 + 56,2 + 69,73}{3} = 56,46\%$$

Berdasarkan hasil perhitungan diatas, maka dapat dilakukan perhitungan akurasi, presisi, *recall*, dan *f-measure* pada masing-masing skenario pengujian sebagai berikut:

1. Skenario Pembagian Dataset 8:2

Tabel 4.6 Hasil pengujian dengan pembagian dataset 8:2

Fitur	Metode	Akurasi	Presisi	Recall	F-Measure
10%	Bagging Ensemble Classifier	51%	55%	54%	51%
	Naïve Bayes	46%	48%	48%	46%
20%	Bagging Ensemble Classifier	56%	62%	60%	56%
	Naïve Bayes	49%	51%	51%	49%
30%	Bagging Ensemble Classifier	64%	67%	66%	64%
	Naïve Bayes	56%	57%	58%	57%
40%	Bagging Ensemble Classifier	64%	66%	65%	64%
	Naïve Bayes	56%	57%	58%	57%
50%	Bagging Ensemble Classifier	64%	66%	65%	64%
	Naïve Bayes	56%	57%	58%	57%
60%	Bagging Ensemble Classifier	64%	66%	65%	64%
	Naïve Bayes	56%	57%	58%	57%
70%	Bagging Ensemble Classifier	64%	66%	65%	64%
	Naïve Bayes	56%	57%	58%	57%
80%	Bagging Ensemble Classifier	64%	66%	65%	64%
	Naïve Bayes	56%	57%	58%	57%
90%	Bagging Ensemble Classifier	64%	66%	65%	64%
	Naïve Bayes	56%	57%	58%	57%

Pada Tabel 4.6 pengujian dengan menggunakan perbandingan 8:2 pada pembagian dataset, nilai terbaik dihasilkan dari menggunakan fitur sebanyak 30% dan menggunakan metode *Bagging Ensemble Classifier* yang menghasilkan akurasi sebesar 64%, presisi sebesar 67%, *recall* sebesar 66%, dan *f-measure* sebesar 64%.

2. Skenario Pembagian Dataset 7:3

Tabel 4.7 Hasil pengujian dengan pembagian dataset 7:3

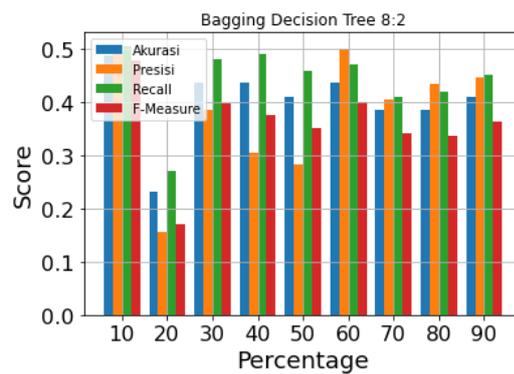
Fitur	Metode	Akurasi	Presisi	Recall	F-Measure
10%	Bagging Ensemble Classifier	45%	45%	45%	45%
	Naïve Bayes	43%	43%	43%	43%
20%	Bagging Ensemble Classifier	46%	48%	47%	47%
	Naïve Bayes	45%	47%	43%	45%
30%	Bagging Ensemble Classifier	45%	47%	46%	45%
	Naïve Bayes	45%	48%	46%	45%
40%	Bagging Ensemble Classifier	50%	52%	51%	50%
	Naïve Bayes	48%	52%	50%	49%
50%	Bagging Ensemble Classifier	50%	52%	51%	50%
	Naïve Bayes	48%	52%	50%	49%
60%	Bagging Ensemble Classifier	50%	52%	51%	50%
	Naïve Bayes	48%	52%	50%	49%
70%	Bagging Ensemble Classifier	50%	52%	51%	50%
	Naïve Bayes	48%	52%	50%	49%
80%	Bagging Ensemble Classifier	50%	52%	51%	50%
	Naïve Bayes	48%	52%	50%	49%
90%	Bagging Ensemble Classifier	50%	52%	51%	50%
	Naïve Bayes	48%	52%	50%	49%

Pada Tabel 4.7 pengujian dengan menggunakan perbandingan 7:3 pada pembagian dataset, nilai terbaik dihasilkan dari menggunakan fitur sebanyak 40% dengan menggunakan metode *Bagging Ensemble Classifier* yang menghasilkan akurasi sebesar 50%, presisi sebesar 52%, *recall* sebesar 51%, dan *f-measure* sebesar 50%.

4.3 Pembahasan

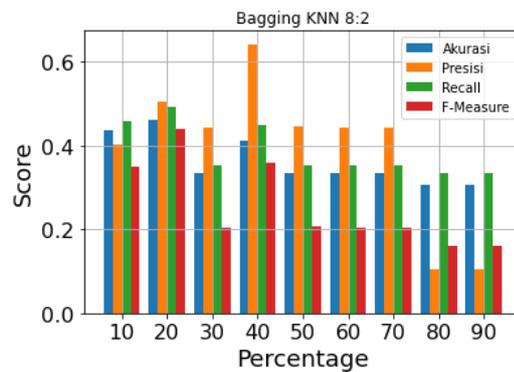
Merujuk pada penelitian yang dilakukan oleh Nai-Arun dan Punnee (2014) pada penerapan metode *Bagging Ensemble Learning* menggunakan metode dasar

Naive Bayes, *k-Nearest Neighbor* (KNN), dan *Decision Tree*. Penulis mencoba menggunakan 3 klasifikasi dasar pada penelitian ini demi menemukan manakah dari 3 klasifikasi dasar tersebut yang memiliki performa terbaik pada kasus penelitian ini (Nai-Arun & Punnee, 2014).

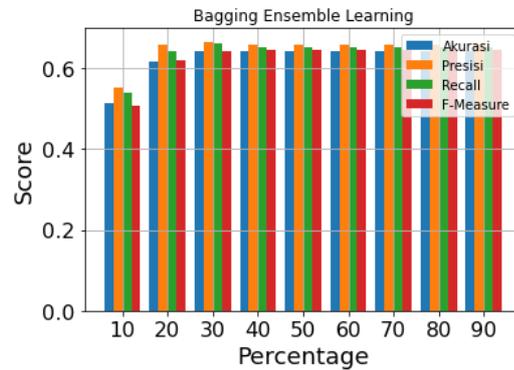


Gambar 4.1 Grafik hasil pengujian Bagging Ensemble Learning dengan klasifikasi dasar

Decision Tree



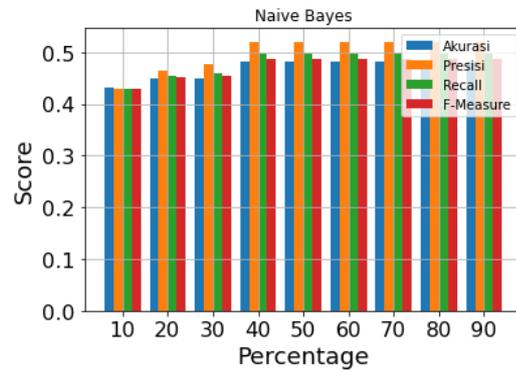
Gambar 4.2 Grafik hasil pengujian Bagging Ensemble Learning dengan klasifikasi dasar KNN



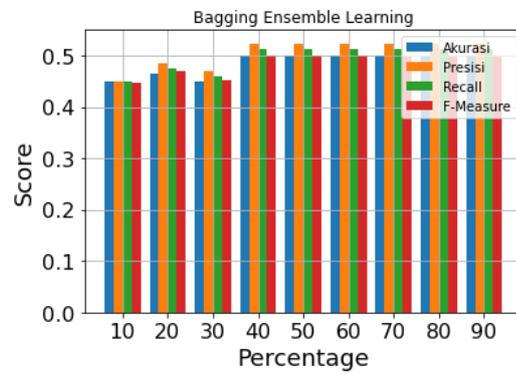
Gambar 4.3 Grafik hasil pengujian Bagging Ensemble Learning dengan klasifikasi dasar Naive Bayes

Berdasarkan grafik hasil percobaan metode *Bagging Ensemble Learning* yang telah dilakukan dengan menggunakan klasifikasi dasar *Decision Tree*, KNN, dan *Naive Bayes* didapatkan hasil performa terbaik dihasilkan dengan menggunakan klasifikasi dasar *Naive Bayes* seperti yang ditampilkan pada Gambar 4.3. Maka dari itu pada penelitian ini penulis menggunakan metode *Naive Bayes* sebagai metode klasifikasi dasarnya.

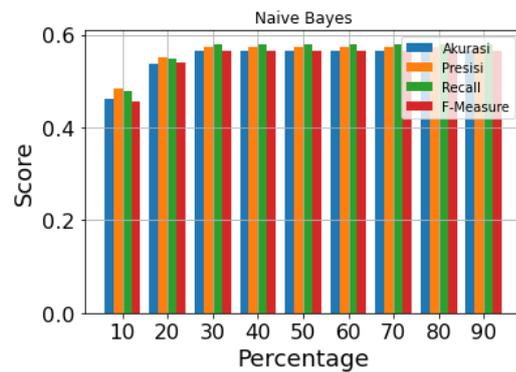
Berdasarkan hasil dari skenario pengujian yang telah dilakukan pada pembagian dataset dan pembagian jumlah fitur, informasi yang didapatkan bahwa penggunaan metode *Bagging Ensemble Classifier* mempengaruhi hasil klasifikasi menggunakan metode dasar *Naive Bayes*. Berikut merupakan grafik hasil dari Skenario Pengujian.



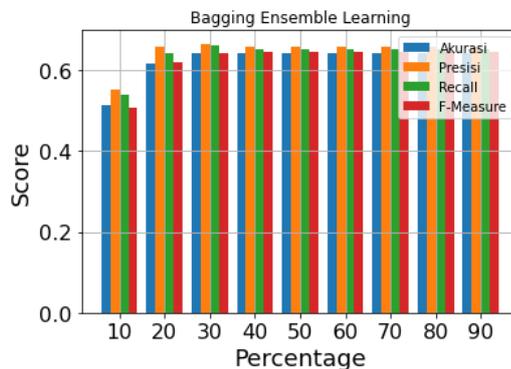
Gambar 4.4 Grafik Hasil pengujian metode Naive Bayes dengan pembagian dataset 7:3



Gambar 4.5 Grafik Hasil pengujian Bagging Ensemble Classifier menggunakan Naive Bayes dengan pembagian dataset 7:3



Gambar 4.6 Grafik Hasil pengujian metode Naive Bayes dengan pembagian dataset 8:2



Gambar 4.7 Grafik Hasil pengujian Bagging Ensemble Classifier menggunakan Naive Bayes dengan pembagian dataset 8:2

Setelah dilakukan serangkaian percobaan dengan 2 skenario pembagian dataset, dapat dilihat bahwa penggunaan metode *Naive Bayes* sebagai metode klasifikasi dasar mempunyai performa yang konsisten dengan menghasilkan nilai yang sama dari fitur sebanyak 40% hingga 90%. Pada sisi performa hasil percobaan dengan nilai tertinggi dihasilkan dari pembagian dataset 8:2 dengan menggunakan metode *Bagging Ensemble Learning* sesuai pada Gambar 4.7 dengan percobaan yang menggunakan fitur sebanyak 30% dengan nilai akurasi sebesar 64%, presisi sebesar 67%, *recall* sebesar 66%, dan *f-measure* sebesar 64%.

Pada skenario pembagian dataset 7:3, nilai performa tertinggi juga didapatkan dengan menggunakan metode *Bagging Ensemble Learning* pada percobaan yang menggunakan fitur sebanyak 40% dengan nilai akurasi sebesar 50%, presisi sebesar 52%, *recall* sebesar 51%, dan *f-measure* sebesar 50%. Jika melihat nilai performa maka menggunakan metode *Bagging Ensemble Learning* bisa meningkatkan performa metode klasifikasi *Naive Bayes*, tetapi nilai performa masih dikisaran 50%. Hal ini bisa terjadi dikarenakan jumlah data latih yang

berbeda dan proses *bootstrap* yang mengacak data latih pada setiap perulangannya sehingga bisa berpengaruh bahkan mengurangi performa terhadap hasil klasifikasi.

Jika melihat skenario terbaik pada hasil uji coba yaitu pada skenario pembagian dataset 8:2 dengan menggunakan metode *Bagging Ensemble Learning* dan memakai 30% jumlah fitur, bisa ditetapkan bahwa penggunaan metode *Bagging Ensemble Learning* pada metode klasifikasi dasar *Naive Bayes* bisa mempengaruhi nilai akurasi, presisi, *recall*, dan *f-measure* dengan peningkatan akurasi dari 56% ke 64%, peningkatan presisi dari 57% ke 67%, peningkatan *recall* dari 58% ke 65%, dan peningkatan *f-measure* dari 56% ke 64%. Bila dihitung rata-rata peningkatan performa maka penggunaan metode *Bagging Ensemble Learning* bisa meningkatkan performa klasifikasi *Naive Bayes* yaitu akurasi meningkat sebesar 14%, presisi meningkat sebesar 17,5%, *recall* meningkat sebesar 12,1%, dan *f-measure* meningkat sebesar 14%. Hal ini disebabkan dengan adanya proses *bootstrap* dan *agregating* yang memvalidasi hasil klasifikasi yang dilakukan dengan menggunakan metode dasar *Naive Bayes*.

Kajian yang dilakukan memberikan opini positif, negatif, atau netral tentang inisiatif pemerintah untuk memerangi pandemi Covid-19. Analisis sentimen yang dilakukan dapat menjadi parameter bagi pemerintah dalam menghadapi isu-isu kritis. Dengan demikian, pemerintah juga dapat merefleksikan prestasi dan pencapaiannya dalam melawan pandemi Covid-19 dan masyarakat dapat menyampaikan pendapatnya dengan lebih hati-hati. Sebagaimana pada Al-Qur'an surat Az-Zumar Ayat 6:

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ فَتُصْبِحُوا عَلَىٰ مَا فَعَلْتُمْ نَادِمِينَ

Artinya: “*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)

Berdasarkan tafsir Ibnu Katsir menjelaskan bahwa Allah SWT memerintahkan kaum mukmin untuk selalu memeriksa informasi yang mereka dapatkan dengan teliti dan tidak menerima secara langsung berita yang disampaikan yang akibatnya membalikkan kenyataan, sehingga barang siapa menerima berita tanpa diteliti terlebih dahulu dan langsung menyampaikannya maka sama dengan mengikuti jejaknya. Seorang Muslim harus menjadi sumber informasi yang dapat dipercaya, kita tidak boleh menjadi orang jahat yang mengubah keraguan menjadi keputusan. Ketika kita menjadi Muslim yang amanah, segala sesuatu menjadi stabil dan teratur. Sebagai seorang muslim, jangan bertindak tergesa-gesa. karena sifat ini bisa membuat umat Islam bertindak zalim terhadap manusia, sehingga nantinya mereka menyesalinya.

BAB V

PENUTUP

5.1 Kesimpulan

Setelah dilakukan penelitian terhadap klasifikasi sentimen masyarakat terhadap kinerja pemerintah dalam menanggulangi Covid-19 menggunakan metode *Bagging Ensemble Classifier* dengan seleksi fitur *Information Gain* menghasilkan performa terbaik dengan menggunakan pembagian dataset 8:2 dan fitur sebesar 30% yang menghasilkan nilai akurasi sebesar 64%, presisi sebesar 67%, *recall* sebesar 66%, dan *f-measure* sebesar 64%.

Berdasarkan hasil percobaan klasifikasi menggunakan metode *Naive Bayes* dan metode *Bagging Ensemble Classifier* dengan klasifikasi dasar *Naive Bayes*, penggunaan *Bootstrap Aggregating (Bagging)* berpengaruh terhadap performa dari metode *Naive Bayes* dengan meningkatkan akurasi meningkat sebesar 14%, presisi meningkat sebesar 17,5%, *recall* meningkat sebesar 12,1%, dan *f-measure* meningkat sebesar 14%. Sehingga dapat disimpulkan bahwa penggunaan *Bootstrap Aggregating* pada metode *Naive Bayes* akan berpengaruh pada hasil klasifikasi.

5.2 Saran

Berdasarkan hasil percobaan penelitian ini, diharapkan penelitian selanjutnya dapat meningkatkan hasil klasifikasi yang lebih akurat. Oleh karena itu, penulis memiliki saran untuk pengembangan penelitian selanjutnya, sebagai berikut :

1. Menambah jumlah kata informal pada kamus sehingga dapat mempengaruhi pembacaan *term* pada sebuah dokumen.

2. Menambah jumlah dataset sehingga jumlah term pada data training lebih beragam.
3. Melakukan percobaan menggunakan metode seleksi fitur lain seperti algoritma genetika, *Chi Square*, *Mutual Information* dan sebagainya.

DAFTAR PUSTAKA

- Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, 152, 341–348. <https://doi.org/10.1016/j.procs.2019.05.008>
- Jo, V. (2019). Text Mining Concept, Implementation, and Big Data Challenge. In *Seminars in Diagnostic Pathology* (Vol. 36, Issue 2). Springer. <https://doi.org/10.1053/j.semmp.2019.02.002>
- Kepios. (2022). *YOUTUBE STATISTICS AND TRENDS*. <https://datareportal.com/essential-youtube-stats>
- Kosasih, R., & Alberto, A. (2021). Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier. *ILKOM Jurnal Ilmiah*, 13(2), 101–109. <https://doi.org/10.33096/ilkom.v13i2.721.101-109>
- Liu, S. (2021). Machine Learning. In *Machine learning*. Tsinghua University Press.
- Maulana, M. (2016). *SKRIPSI Deteksi Emosi Manusia Pada Tweet Bahasa Indonesia Dengan Klasifikasi Naive Bayes*. <http://repository.its.ac.id/48885/>
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC - Trends in Analytical Chemistry*, 132, 116045. <https://doi.org/10.1016/j.trac.2020.116045>
- Nai-Arun, N., & Punnee, S. (2014). Ensemble learning model for diabetes classification. *Advanced Materials Research*, 931–932(May 2014), 1427–1431. <https://doi.org/10.4028/www.scientific.net/AMR.931-932.1427>
- Pratiwi, A. I., & Adiwijaya. (2018). On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis. *Applied Computational Intelligence and Soft Computing*, 2018. <https://doi.org/10.1155/2018/1407817>
- PUTRA, I. B. G. W., SUDARMA, M., & KUMARA, I. N. S. (2016). Klasifikasi Teks Bahasa Bali dengan Metode Supervised Learning Naive Bayes Classifier. *Teknologi Elektro*, 15(2), 81–86. <https://ojs.unud.ac.id/index.php/JTE/article/view/ID21577>
- Putranti, N. D., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91. <https://doi.org/10.22146/ijccs.3499>
- Shaltout, N. A., El-Hefnawi, M., Rafea, A., & Moustafa, A. (2014). Information gain as a feature selection method for the efficient classification of influenza based on viral hosts. *Lecture Notes in Engineering and Computer Science*, 1, 625–631.

- Sriyano, C. S., & Setiawan, E. B. (2021). Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF. *E-Proceeding of Engineering: Vol.8, No.2, 8(2)*, 3396–3405.
- Subasi, A., Kadasa, B., & Kremic, E. (2020). Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier. *Procedia Computer Science*, 168(2019), 34–39. <https://doi.org/10.1016/j.procs.2020.02.248>
- Sunni, I., & Widyantoro, D. H. (2012). Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik. *Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro Dan Informatika*, 1(2), 200–206.
- Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*, 57(1), 77–93. <https://doi.org/10.1016/j.dss.2013.08.002>

26	Nipu rakyat blum aja Lo kena azab dialam kubur jgn mikirin uang terus cri keuntungan covid diadakan LG dsar gila Lo ingat hdup d dunia hnya sementara	2
27	beginian masih laku? ☐	3
28	Gak ngefek boss	3
29	Jangan 2 ada laboratoriumnya nih seperti di Ukraina, masker laku lagi dah 🙅🙅🙅	2
30	Luhut Binsar panjaringan	2
31	yang sekolah pda seneng, mikirin yang kerja broo!	3
32	Yang kek gini nih gw demen bismillah kopid tambah naik pengen bdr lagi awoakwok biar peringkat 1	2
33	Dah nganggur covit ga ilang2	2
34	Lebih baik di tutup dseluruh Indonesia tapi pemerintah harus berimakan	2
35	Apa2 covik .semua covik ngk ad kerjaan pemerintah ini kalian d gaji dr rakyat tp nyusahin rakyatmu .	2
36	Bosan dg kasus covid yg mengsensarakan rakyat semua omong kosong	2
37	Amin	3
38	ah...yg bner nih	3
39	Saya juga sedih ketika RAKYAT harus 3M... MENTRI NYA 17 M...	2
40	Demi Prokes Sekolah Di liburkan ...	1
41	BODOH DISINI AMAN AJA YG PENTING GK ADA TEST NYA	2
42	terus ada yg percaya ??	3
43	pakai masker vaksin 3x jgan lupa kalau keluar rumah pakai sepatu jgan pakai sandal ntr di tilang.,	2
44	BISMILLAH SEMOGA CORONA HILANG DAN TIDAK LIBUR 2 MINGGU DAN TIBA TIBA JADI 2 TAHUN DAN SAYA BENAR BENAR MENJADI BODOH SAAT LIBUR PANJANG 2 TAHUN ITU YANG DULUNYA SAYA MENDAPATKAN RANGKING 1 TERUS SEKARANG MALAH SAYA TURUN DRASTIS DAN SAYA TIDAK NAIK KELAS MELULU	1
45	Astagfirullah haladziiiiim..corone corone	1
46	Semoga sekolah online lg biar bisa cepat lulus	1
47	Dnger tuuuuh	3
48	Woiiii tolong lah .. Baru ekonomi berputar jngn d matiin lagi dnk kami rakyat kecil	1
49	Hox lah	3
50	APAKAH ADA MAFIA ALAT MEDIS YG BERMAIN SUPAYA MENDAPAT CUAN ??????????	2
51	KaPan ya dunia kiamat ,, Dunia ini sudah kacau ,, semoga 2035 kiamat terjadi amin ,,	2

52	SEMOGA ONLINE LAGI 🙏	3
53	Baru juga dapat kerja masak mau dirumahkan lagi. Nasib2, corona bener2 menghancurkan hidupku.	1
54	Ppkm akal akalan si misikin kalau orang kaya gak mungkin mau mendekati desember nguber bonus natalnya mangkanya diseting ppkm jilid.3 😄😄😄😄😄😄	2
55	Nih lo catet omongan w ya, kalau masuk bulan desember covid19 nya hilang w yakin bener kalau emang tetjadi inibyg ke 3x nya di bulan yg sama berarti biang keroknya dah jelas siapa kah dia???????????? Yg rakus proyek and rakuas kekuasaan (PPKM jilid.3)	2
56	Bisnis moyet	2
57	Ayooo vaksin ke 3 bosteer	1
58	Besok omicron varian B4B1 😄😄😄...	2
59	MAU SEKOLAH JUGAA	3
60	Baru ye bernafas ya ppkm lg gmna ni ya 😄	3
61	mau idul adha biasa, GA boleh solat id	2
62	Mendingan satgas covid dialih fungsikan untuk ngurusin sembako, biar ga melonjak terus harganya.	3
63	Horeeeee nganggur lagi	2
64	Udah ga perduli	3
65	Produsen masker bahagia mendengar ini 😄	2
66	Asik bisnis nya jalan lagii 🙏	1
67	Setiap mendekati hari raya selalu saja meningkat sampai"; curiga kalau covid ini sebenarnya iblis yg mengganggu umat muslim	2
68	REZIM KACAUUU	2
69	Moto GP mandalika udah terlaksana Formula E udah terlaksana juga PRJ stengah jalan udah terlaksana juga 🙏 Nih skarang mao memasuki idul adha ada PPKM lagi ada COVID lagi 😄 Manteep pak ANIES 🙏🙏🙏	2
70	Kalo gua sih bodo amat 😄	3
71	Tambahi lg!	3
72	Hadeeuhhhhh	3
73	dis kau jangan jauh dari ku 😄	3
74	seharus orang orang selalu pakai masker ditempat";umum walopun telah dipaksin,karena Aku merasakan korona masih ada,masih banyak yg batuk";dan batuk";itu hampir sama 😄	1
75	Zehahahahaha	3
76	Plis naik dong.. pengen sidang skripsi online 🙏	2
77	Sontekkk manee..	3
78	Csri dong itu warga yg kena kovid itu daerah mana kp apa jln nya nah klg nya diberi bantuan.jgn sapu rata sj menibgkat covid jkrt yg mana.kami di klr p gadung kec p gadung selama ini	1

	algandulillah nyaman sj itu bebar yg ke wisma atlit itu nggak disuruh lg nt dibayar pura2 sakit.capek dgr nya.cb lah tangkap copid itu memdng bentuk nya lkaya biji kedong dong	
79	Ngerti kok wajar meningkat udah mau lebaran haji...masyarakat sudah paham	2
80	Udah basi preet	2
81	Selama dipimpin oleh pemimpin yang saat ini semua kacau ☹️ semua serba mahal minyak goreng, BBM,bahan baku,terus banyak peraturan yang gak masuk akal,mbok y di pikir yang kena dampaknya rakyat kecil dan semua rakyat,Ojo mikir weteng e dewe,gaji juga gak sesuai dengan kebutuhan pokok,	2
82	Jual vaksin di tiap warung sekalian di pedagang kaki lima..sengaja banget covid di naik naikan trus ..cari uang yg halal supaya berkah	2
83	Faksin booster masih banyak makanya masih diungkit ²	2
84	Klo saya perhatikan klo cuaca y hujan Mulu pasti angka covid naik...cuma sekedar info ditempat kerja saya pernah satu keluarga kena covid semua , Boss , baby sister , pembantu , anak , cucu , mantu , sampai akhir y Boss saya yg laki meninggal.	1
85	Hnya Indonesia Yg masih Sibuk dgn Covid.... LOL sampa Harus memaksa Masyakarot harus Booster. Trus yg nggk bisa booster bgmna dong..?? Apa pemerintah menjamin kehidupannya	2
86	Lagu lama	2
87	ya elah	3
88	waktu itu saya pergi ke pasar hanya untuk membeli tahu dan tempe.. terus saya goreng..	3
89	Bangsa ini terpuruk krn menangani pandemic berlebihan ya bos? Justru krn kulakan vaksin kebanyakan dan yg berlebihan jadinya semua rakyat hidupnya susah.. : Ppn naik Bpjs dipaksakan Pertamina dipaksakan Listrik naik Pdam naik Banyak hal jadi dipaksakan krn negara butuh uang, apalagi ngotot IKN juga... Nasib nasib...	2
90	Padahal pcr sdh tidak diwajibkan. Strategi indonesia sdh bagus. tapi berpotensi memperburuk kasus covid...	1
91	Bismillah daring lagi aminn 🙏	1
92	Rs di indo sepi job karena banyak yang berobat ke luar negeri, jadinya covid disalahkan. Kasian covid.	2
93	BISMILLAH SEMOGA COVID MENINGKAT TRUSS AAMIIN	2
94	Surabaya aman losssss gk ono opo2	3
95	ajg lah	3
96	Hidup WAKANDA	3
97	Ngga bosan atas kebohongan ini apa anggaranya msh ada pak cari topik lain dong bosen saya.....	2
98	Bosen	2

99	Yessss si copid masuk lagi,, mudah2an MASKER saya cepet laku	2
100	Haha gara2 Pak pres sowan ke ukraina ini ningkat lagi	2
101	Stlh PMK kini kopet naik lg, mnjelang idul adha. Cma cocokologi aja gan 🙄	2
102	Semoga kovid meningkat lagi	2
103	mbel gedes cuk , bosen	2
104	Aku muak di rumah tross...	2
105	PPKM?? Gimana dng orang yg datang dari luar negri??bisa bebas keliling indonesia?	1
106	Yaaa iya lah udah di datangkan 74 juta dosis.klu gak di suntikan ya tekor bandar,,,, Pengen sehat kok tambah dosis mikir dong.....rakyat.	2
107	ayo bisa yuk kita lewat pandemi ini dengan memakai masker, mencuci tangan, vaksin lengkap, dan menjaga jarak. Bersama kita bisa mewujudkan Indonesia sehat.	1
108	Jadul di perbaharui ... 🙄🙄🙄🙄🙄 Janji jgn susahkn orang cari nafkah Itu ngk apa2	1
109	Naikin lg biar ada pemasukan lagi	2
110	Peduli lindungi	1
111	🙄🙄🙄	3
112	Kasus makin naik .. tapi orang yang baru pulang dari luar negeri bebas berkeliaran ia ... Habis tuh bebas pcr lagi .	2
113	Saya kira udh selesai covid nya	3
114	Stok booster masih banyak 😊	1
115	🙄🙄🙄🙄🙄🙄🙄🙄	3
116	Selesai pemilu apa selesai juga cerita soal covid??? Kita liat bersama sama	2
117	Selesai pemilu apa selesai juga cerita soal covid??? Kita liat bersama sama	2
118	Selesai pemilu apa selesai juga cerita soal covid??? Kita liat bersama sama	2
119	MotoGp oh MotoGp cepat lah ke Indo	3
120	Kompas lagi kompass lagi kompass lagi..dan lagi lagi kompas... Heran gue sma kmpas	3
121	cuan	3
122	Covid d indonesia ternyata masih tinggi tapi kenapa yang pulang dari luar negeri sudah bebas pcr ya	1
123	Sesuai prediksi lagu lama terulang kembali	2
124	ayo Prokes Ketat... utk Indonesia Sehat Rek...!!!!	1
125	Corona masih berkembang,,kirain sudah habiss	1
126	Waspada!!	1
127	Tolong pemerintah segera lakukan lockdown untuk mencegah penularan lebih banyak.	1
128	Bhulsheeettttt...	2

129	Dikira varian popcron atau pop ice,ee malah varian BA4 N BA5 ky komplek	2
130	Mungkin sampai akhir zaman baru selesai ga da covid,nnti yg membohongi rakyat demi kepentingan perut y akan di adili oleh malaikat lg ga da kata2 toleran?	2
131	Bagaikan sinetron ga ada ujungnya 🤔	3
132	Lg sibuk nyapres semua, bilang aja "msh terkendali"..	2
133	Bikin pusing aja dan mresahkan msyarakat lgi	2
134	KAMI SUDAH BOSAN DISUGUHI BERITA COVID.COVID BULLSHT GAK BERBAHAYA UDH KAGA LAKU DAH JUALNA COVID KAMI SUDAH TIDAK TAKUT	2
135	Berita Dancuuuuukkkkk!!!!	2
136	CIEE CIEE MULAI KEHABISAN KUCURAN DUIT COVID NIIHHH NAKES . UP TEROOOSSS BIAR TURUN TERUSS CUANNYA	2
137	Males udah cape rasanya. Jaga aja stamina, jiwa sehat,.	1
138	Kepala satgas covid ajja dah mampus kok ini masih diterusin lg..!! Mau tuh yg jadi kepala satgas covid yg baru nyusul mampus..!!	2
139	Croot,ketemu berita begini langsung blokir aja,biar gk rame 😏,berita sampah ini	1
140	Bukan meningkat...tapi sengaja di tingkat kan...maklum ntar lg idul adha...	2
141	Terus vaksin apalagi	3
142	Dokter Loeis kok bs meninggal?heran sy, apa segampang itu?	1
143	Kendalikan dulu harga sembako yg harganya kian meroket.....	1
144	Ingat : DRAKOR COVID blm selesai ..msh ada Episode baru yg akan Tayang .. artinya Bisnis Vaksin akan mulai menggeliat lagi ditengah - tengah harga sembako jg semakin naik ..	2
145	PUNCAK KOMEDI	3
146	berita gk bermutu terus bermunculan ,,celek ngasuuu	2
147	Bosannnn ahhh	3
148	Mao cari sesuap makan aj di belain sebar brita bohong.	2
149	Masih ada kopet ya 😊	1
150	sedia payung sebelum hujan	1
151	cuan cuancuan	3
152	Peningkatan pengangguran tak terkendali Peningkatan kerugian BUMN-BUMN tak terkendali Peningkatan import sangat tak terkendali di Negeri Tanah Surga, Negeri Zamrud Katulistiwa, Negeri sempalan surga yg jatuh ke bumi	2
153	Pergantian cuaca di manfaat kan seprti ini ..pdhl dah dri dulu batuk pilek pd pergantian cuaca dah biasa tpi skrng tidk	1
154	Sy mencari berita penemuan ribuan amunisi dan bahan peledak djie kian han kenapa tidak menjadi headline? Dan itu tanda bahaya secara gamblang dan nyata bagi NKRI.	3

155	Lah laboratorium nya aja udah dihancurin RUSIA. covid yg sekarang ada itu cuma sisa2 doang kan kemaren yg meninggal udah jutaan orang seluruh dunia. Asal ga dikembangin lagi tuh virus udah aman kok kan udah pada divaksin semua warga Indonesia.. inget ya asal ga dikembang biakan lagi dilaboratorium. Harusnya pemerintah berani bilang ke PBB tuh Ukraina penjahat kemanusiaan harus dituntut karena dengan sengaja membuat virus mematikan dan disebarkan ke seluruh dunia.	2
156	Rakyat sudah semakin pintar!! Dan Rakyat adalah hakikatnya yang punya kuasa!! Namun berhubung Rakyat tak berkuasa gini ni jadinya,, Bukan masalah harus jaga kesehatan, Jaga kesehatan mah masing2 juga insyaallah masyarakat cerdas, Cuma <i>greget</i> aja kalo udah ada berita2 gini,	1
157	Kumat	2
158	Pergantian cuaca... Di libatkan	1
159	Haduhhhhh puyeng kyk liat sinetron	3
160	Dengan adanya varian baru proyek vaksin baru perut yg gendut semakin gendut.	2
161	Msh byk ga di vaksin ada juga cuma ktp nya di vaksin krn mau bebas msk mall 😊😊😊	1
162	Basi basi dah basi covid 19	3
163	Seng meningkat kuwi beritane... Kopit proyek bisnis..... Ora Gawe berita ben semangat rakyate malah di takut takuti di apusi ae....	2
164	Senetron suda mulai tayang teman teman.. ada yg mau nonton..ini sinetron baru loh..	3
165	Ujung"nya kluarin vaksin lagi	2
166	Bagaimana. Tarif Listrik di naikan 7 golongan. Biar sama per kwh nya. Biar tidak rugi pemerintah. Berikut daftar kenaikan tarif listrik golongan tersebut : Mulai, 1 Juli 2022. R1M: 900 VA: Rp 1.699,53 R1: 1.300 VA - 2.200 VA: Rp 1.699,53 R2: 3.500 VA - 5.500 VA: Rp 1.699,53 R3: 6.600 VA ke atas: Rp 1.699,53 P1: 6.600 VA - 200 KVA: Rp 1.699,53 P2: 200 KVA ke atas: 1.522,88 P.3/TR : 16.99,53 Direktur Utama PLN Mulai, 1 Juli 2022. Listrik Non Subsidi. Mengikuti Tarif Adjustment. Biar tidak ada yang turun daya. Daya tetap / tambah daya. Tarif Listrik Rumah Tangga. - Daya 4 A / 900 W, R1M = 1.699,53 / kwh. - Daya 6 A / 1.300 W, R1 = 1.699,53 / kwh. - Daya 10 A / 2.200 W, R1 = 1.699,53 / kwh. - Daya 16 A / 2.200 W, R2 = 1.699,53 / kwh. - Daya 20 A / 4.400 W, R2 = 1.699,53 / kwh. - Daya 25 A / 5.500 W, R2 = 1.699,53 / kwh. - Daya 35 A / 7.700 W, R3 = 1.699,53 / kwh + PPN 11 % = 1.886,48 / kwh. - Daya 50 A / 11.000 W, R3 = 1.699,53 / kwh + PPN 11 % = 1.886,48 / kwh.	2

167	Goreng terus sampai gosong buat bisnis. Negara lain sdh gk mikirin covid. Indonesia kok sibuk mikirin covid. Apa lg midia dibesar besarkan dasar dancok.	2
168	Tidak ada yg lain apa kok mulai lagi	2
169	Dulu juga bilng nya gtu wkwwk	2
170	Jangan buru buru di longgar kan soal masker. Hadeeeeeeh	1
171	Bagaimana, tarif listrik, non subsidi, orang mampu, golongan orang-orang kaya, di naikan aja, biar tidak rugi pemerintah, biar sama per kwh nya. Tarif Listrik Naik, Non Subsidi. Listrik Non Subsidi. Tarif Listrik Rumah Tangga. - Daya 4 A / 900 W, R1M = 1.352 / kwh, di naikan 25,7 %, jadi 1.699,53 / kwh. Biar sama per kwh nya = 1.699,53 / kwh. - Daya 4 A / 900 W, R1M = 1.352 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 6 A / 1.300 W, R1 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 10 A / 2.200 W, R1 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 16 A / 3.500 W, R2 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 20 A / 4.400 W, R2 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 25 A / 5.500 W, R2 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 35 A / 7.700 W, R3 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh + PPN 11 % = 1.886,48 / kwh. - Daya 50 A / 11.000 W, R3 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh + PPN 11 % = 1.886,48 / kwh.	1
172	Juli 2017 s/d Juni 2022. Selama 5 tahun. Pemerintah Bohong. Katanya. Tarif Listrik Rumah Tangga. Daya 4 A / 900 W, R1M, mengikuti tarif adjusment = 1.467,28 / kwh. Sekalinya tidak, beda 115,28 / kwh. - Daya 4 A / 900 W, R1M = 1.352 / kwh. - Daya 6 A / 1.300 W, R1 = 1.467,28 / kwh. - Daya 10 A / 2.200 W, R1 = 1.467,28 / kwh. - Daya 16 A / 3.500 W, R2 = 1.467,28 / kwh. - Daya 20 A / 4.400 W, R2 = 1.467,28 / kwh. - Daya 25 A / 5.500 W, R2 = 1.467,28 / kwh. - Daya 35 A / 7.700 W, R3 = 1.467,28 / kwh + PPN 10 % = 1.614,28 / kwh. - Daya 50 A / 11.000 W, R3 = 1.467,28 / kwh + PPN 10 % = 1.614,28 / kwh. Mulai Juli 2022. Bagaimana, tarif listrik, non subsidi, orang mampu, golongan orang-orang kaya, di naikan aja, biar tidak rugi pemerintah, biar sama per kwh nya. Pemerintah Jangan Bohong. Priseden Jokowi setuju naikan. Tarif Listrik Rumah Tangga. Daya 4 A / 900 W, R1M = 1.352 / kwh, di naikan 25,7 %, jadi 1.699,53 / kwh. Biar sama per kwh nya = 1.699,53 / kwh. Biar tidak selisih. 1.699,53 - 1.444,70 = 254,83 / kwh. Biar tidak beda = 254,83 / kwh. Tarif Listrik Naik, Non Subsidi. Listrik Non Subsidi. Tarif Listrik Rumah Tangga. - Daya 4 A / 900 W, R1M = 1.352 / kwh, di naikan 25,7 %, jadi 1.699,53 / kwh. Biar sama per kwh nya = 1.699,53 / kwh. Biar ikut tarif adjusment. - Daya 4 A / 900 W, R1M = 1.352 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 6 A / 1.300 W, R1 = 1.444,70 / kwh, di naikan jadi	2

	1.699,53 / kwh. - Daya 10 A / 2.200 W, R1 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 16 A / 3.500 W, R2 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 20 A / 4.400 W, R2 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh. - Daya 25 A / 5.500 W, R2 = 1.444,70 / kwh, di naikan jadi 1.699,53 / kwh.	
173	Tidak lama lagi ada pundi pundi vaksin covid Betul gak lut 😊😊 dasar otak haram	2
174	Wakwaw lah	3
175	halah ujungnya pemerintah mengeluarkan uang untuk covid dan uang di korupsi separo,kan lumayan rihi oleh2 1triliun	2
176	proses sangat penting harga sembako yg mhl jg penting gimana caranya supaya gk mhl	1
177	Pake logika aj..bner gx tu covid,knp org yg dh vaksin bs kena..mn byk varian lg,dh kyk mkanan aj,jd ap gunany vaksin,👀👀bkin kaya pnjual vaksin👛👛	2
178	Banyak skrang yg longgar gk mematuhi proses utamanya masker... Yuk jaga selalu proses diluar maupun dalam ruangan	1
179	Covid Varian B4 & B5 akan secepatnya meningkat. Pemerintah harus siap sedia menghadapi lonjakan yg tidak terduga ini	1
180	Klo ngBayangin & ngRasain Kondisi Kereta & Stasiun Manggarai saat ini yg Dempet ² an kagak Karuan+Macet nya jakarta Rasanya WFH selalu yg gw harapkan, wkwk.. yg ptg Ada Laptop, Wifi, Ms Team, Kerjaan & Gajian lancar tiap bulan, No Problemo!!	1
181	HINDARI JUDI ...!! USAHAKU HANCUR KARENA JUDI ...	1
182	Covid : *rise wkwkland government : this is fine wkwklander : this is fine Economy : this is not fine	1
183	😊😊	3
184	scam cabinet ,penetrated by WEF,destroyer of the world.	3
185	Masyarakat produktif dan aman dari ancaman COVID-19 dengan antusias mengikuti vaksinasi dosis lengkap dan dosis penguat di sentra pelayanan kesehatan terdekat. Bersatu Lawan COVID-19.ID	1
186	Bayar 1m10jt 😊	3
187	semoga indonesia jaya selalu aamin	1
188	Pak tolong Hentikan sementara pelanggaran aktivitas dan hentikan juga penggunaan masker yang tlah dikurangi Jika tidak pandemi meledak 300 Kali lipat pakkk Tolong dengar masyarakat ya pak	2
189	saya kira udah ilang, tp walaupun begitu saya masih pakai masker	1
190	Semoga bangsa Indonesia (semua masyarakat nya)di seluruh wilayah Indonesia selalu dalam lindungan Allah SWT (penguasa	1

	alam jagat raya bumi ini) hingga di jauh kan dari seluruh wabah penyakit (Corona dll) beserta di jauh kan dari seluruh bencana alam	
191	Terima kasih atas informasinya 🙏 Pak	1
192	Selamat menikmati.. https://youtu.be/FGXLw2AipKM	3

Lampiran 2

Hasil pengujian metode klasifikasi dasar *Naive Bayes* menggunakan perbandingan

Dataset 7:3

No.	Jumlah Fitur	Akurasi	Presisi	<i>Recall</i>	<i>F-Measure</i>
1	10%	0,431034	0,42963	0,430331	0,428723
2	20%	0,448276	0,465278	0,453964	0,452893
3	30%	0,465517	0,497666	0,477598	0,470984
4	40%	0,482759	0,520254	0,497206	0,486772
5	50%	0,482759	0,520254	0,497206	0,486772
6	60%	0,482759	0,520254	0,497206	0,486772
7	70%	0,482759	0,520254	0,497206	0,486772
8	80%	0,482759	0,520254	0,497206	0,486772
9	90%	0,482759	0,520254	0,497206	0,486772

Lampiran 3

Hasil pengujian metode *Bagging Ensemble Classifier* dengan metode klasifikasi dasar *Naive Bayes* menggunakan perbandingan Dataset 7:3

No.	Jumlah Fitur	Akurasi	Presisi	<i>Recall</i>	<i>F-Measure</i>
1	10%	0,448276	0,449123	0,449938	0,447229
2	20%	0,465517	0,483971	0,473572	0,469328
3	30%	0,465517	0,488735	0,478687	0,468337
4	40%	0,5	0,523302	0,512788	0,501032
5	50%	0,5	0,523302	0,512788	0,501032
6	60%	0,5	0,523302	0,512788	0,501032
7	70%	0,5	0,523302	0,512788	0,501032
8	80%	0,5	0,523302	0,512788	0,501032
9	90%	0,5	0,523302	0,512788	0,501032

Lampiran 4

Hasil pengujian metode klasifikasi dasar *Naive Bayes* menggunakan perbandingan

Dataset 8:2

No.	Jumlah Fitur	Akurasi	Presisi	<i>Recall</i>	<i>F-Measure</i>
1	10%	0,461538	0,484687	0,477904	0,455556
2	20%	0,487179	0,50796	0,508207	0,489524
3	30%	0,564103	0,572805	0,580177	0,56566
4	40%	0,564103	0,572805	0,580177	0,56566
5	50%	0,564103	0,572805	0,580177	0,56566
6	60%	0,564103	0,572805	0,580177	0,56566
7	70%	0,564103	0,572805	0,580177	0,56566
8	80%	0,564103	0,572805	0,580177	0,56566
9	90%	0,564103	0,572805	0,580177	0,56566

Lampiran 5

Hasil pengujian metode *Bagging Ensemble Classifier* dengan metode klasifikasi dasar *Naive Bayes* menggunakan perbandingan Dataset 8:2

No.	Jumlah Fitur	Akurasi	Presisi	<i>Recall</i>	<i>F-Measure</i>
1	10%	0,512821	0,550595	0,53851	0,508243
2	20%	0,564103	0,623377	0,599116	0,564312
3	30%	0,641026	0,665359	0,661616	0,641636
4	40%	0,641026	0,659091	0,652146	0,643701
5	50%	0,641026	0,659091	0,652146	0,643701
6	60%	0,641026	0,659091	0,652146	0,643701
7	70%	0,641026	0,659091	0,652146	0,643701
8	80%	0,641026	0,659091	0,652146	0,643701
9	90%	0,641026	0,659091	0,652146	0,643701