

**Analisis Sentimen Kenaikan Harga Bahan Bakar Minyak
(BBM) Berdasarkan Respon Pengguna Media Sosial
Twitter Di Indonesia Menggunakan Metode *Naive Bayes*.**

SKRIPSI

Diajukan untuk Memenuhi Tugas Akhir dan Melengkapi Syarat
Guna Memperoleh Gelar Sarjana Strata Satu (S-1) dalam Teknologi
Informasi



Diajukan oleh :

MUHAMMAD ZIDAN

NIM : 1908096025

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI WALISONGO SEMARANG
2022**

PERNYATAAN KEASLIAN

Yang bertandatangan dibawah ini:

Nama : Muhammad Zidan
NIM : 1908096025
Jurusan : Teknologi Informasi

Menyatakan bahwa skripsi yang berjudul:

**Analisis Sentimen Kenaikan Harga Bahan Bakar Minyak
(BBM) Berdasarkan Respon Pengguna Media Sosial
Twitter Di Indonesia Menggunakan Metode Naive Bayes**

Secara keseluruhan adalah hasil penelitian/karya saya sendiri,
kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 7 November 2022

Pembuat Pernyataan,

Muhammad Zidan

NIM : 1908096025



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Prof. Dr. Hamka Ngaliyan Semarang
Telp.024-7601295 Fax.7615387

PENGESAHAN

Naskah skripsi berikut ini:

Judul : Analisis Senimen Kenaikan Harga Bahan Bakar Minyak (BBM)
Berdasarkan Respon Pengguna Media Sosial Twitter Di
Indonesia Menggunakan Metode Naive Bayes

Penulis : **Muhammad Zidan**

NIM : 1908096025

Jurusan : Teknologi Informasi

Telah diujikan dalam sidang tugas akhir oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Teknologi Informasi.

Semarang, Desember 2022

DEWAN PENGUJI

Penguji I,

Dr. Masy Ari Ulinuha M. T.
NIP : 198108122011011007

Penguji II,

Adzhal Arwani M, M. Kom.
NIP : 199107032019031006

Penguji III,

Wenty Dwi Yuniarti, S.Pd., M.Kom
NIP : 197706222006042005

Penguji IV,

Siti Nur'aini, M.Kom.
NIP : 198401312018012001

Pembimbing I,

Dr. Khotibul Umam, S.T., M.Kom
NIP : 197908272011011007

Pembimbing II,

Adzhal Arwani M, M. Kom
NIP : 199107032019031006

NOTA PEMBIMBING

Semarang, 7 November 2022

Yth. Ketua Program Studi Teknologi Informasi
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum. Wr. Wb.

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul : Analisis Sentimen Kenaikan Harga Bahan Bakar Minyak (BBM) Berdasarkan Respon Pengguna Media Sosial Twitter Di Indonesia Menggunakan Metode *Naive Bayes*
Nama : **Muhammad Zidan**
NIM : 1908096025
Jurusan : Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqsyah.

Wassalamu'alaikum. Wr. Wb.

Pembimbing I,



Khotibul Umam, S.T., M.Kom
NIP. 197908272011011007

NOTA PEMBIMBING

Semarang, 7 November 2022

Yth. Ketua Program Studi Teknologi Informasi
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum. Wr. Wb.

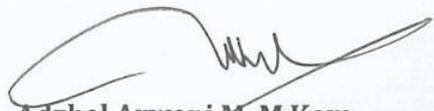
Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul : Analisis Sentimen Kenaikan Harga Bahan Bakar Minyak (BBM) Berdasarkan Respon Pengguna Media Sosial Twitter Di Indonesia Menggunakan Metode *Naive Bayes*
Nama : **Muhammad Zidan**
NIM : 1908096025
Jurusan : Teknologi Informasi

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqsyah.

Wassalamu'alaikum. Wr. Wb.

Pembimbing II,



Adzhal Arwani M, M.Kom
NIP. 199107032019031006

ABSTRAK

Bahan Bakar Minyak (BBM) memiliki peranan yang sangat penting dalam kehidupan masyarakat. Pemerintah Indonesia mengumumkan kebijakan kenaikan harga BBM sehingga memuat kotraversi masyarakat dan memicu pro dan kontra. Di era internet banyak sekali mengubah cara seseorang mengkespresikan pendapat, opini, tanggapan mereka melalui media sosial salah satunya twitter. Dalam penelitian ini analisis sentimen diterapkan untuk menganalisis sentimen dari respon pengguna media sosial twitter tentang kebijakan tersebut untuk dilakukan pengklasifikasian komentar yang didapatkan, apakah komentar tersebut bersifat positif, negatif dan netral.

Proses awal pada penelitian ini melakukan pengambilan data komentar twitter dengan kata kunci harga BBM. Selanjutnya tanggapan tersebut dilakukan proses *text preprocessing* dan pembobotan kata TFIDF (*Term Frequency Inverse Documnet Frequency*). Penelitian ini menggunakan metode *naive bayes clasiffier* dalam melakukan klasifikasi tanggapan dari media sosial twitter yang terbagi menjadi tiga kelas, yaitu positif, negatif dan netral.

Berdasarkan hasil pengujian yang telah dilakukan bahwa hasil dari model *naive bayes classifier* didapatkan sentimen negatif memiliki nilai persentase tertinggi sebesar 53.8%, pada sentimen positif sebesar 37.1% serta pada sentimen netral sebesar 9.1%. Kemudian dilakukan *split validation data* dengan perbandingan data latih dan data uji sebesar 80:20 yang diambil secara acak. Sehingga didapatkan nilai performa dari model *naive bayes* dengan klasifikasi yang **baik** dengan tingkat akurasi sebesar 81%, *presicion* sebesar 83%, *recall* sebesar 81% dan *f1 score* sebesar 79%.

Kata Kunci : analisis sentimen, BBM, twitter, *naive bayes*

KATA PENGANTAR

Alhamdulillah, Puji Syukur atas kehadiran Allah SWT atas segala nikmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi yang berjudul “Analisis Sentimen Kenaikan Harga Bahan Bakar Minyak (BBM) Berdasarkan Respon Pengguna Media Sosial Twitter Di Indonesia Menggunakan Metode *Naive Bayes*” dengan baik. Dengan tujuan dibuat skripsi ini sebagai salah satu bentuk syarat kelulusan pada program sarjana (S1) prodi teknologi informasi di Universitas Islam Negeri Walisongo Kota Semarang. Di sisi lain, penulis juga bertujuan untuk memberikan pengetahuan kepada pembaca.

Pada kesempatan ini penulis ingin mengucapkan banyak-banyak terimakasih kepada seluruh pihak yang memberi dukungan dan bantuan dari pelaksanaan skripsi hingga penyelesaian skripsi ini. Penulis mengakui bahwa apabila tanpa dibimbing, diberi arahan, serta dibina dan tanpa diberikan motivasi dari seluruh pihak, maka penulisan skripsi ini tidak akan berjalan dengan baik. Maka dari itu penulis mengucapkan terimakasih yang tak terhingga kepada :

1. Bapak Prof. Dr. Imam Taufiq, M.Ag, selaku Rektor Universitas Islam Negeri Walisongo Semarang.
2. Bapak Dr. H. Ismail, M.Ag, selaku Dekan Fakultas Teknologi Informasi Universitas Islam Negeri Walisongo Semarang.
3. Bapak Nur Cahyo Hendro Wibowo, S.T., M.Kom, selaku ketua program studi Teknologi Informasi Universitas Islam Negeri Walisongo Semarang.
4. Bapak Khotibul Umam, ST., M.Kom, dan Bapak Adzhal Arwani M, M.Kom selaku dosen pembimbing skripsi saya yang selalu memberikan dukungan, arahan, bimbingan serta motivasi dalam pelaksanaan skripsi hingga pembuatan skripsi ini.
5. Staff, karyawan dan dosen di lingkungan Universitas Islam Negeri Walisongo Semarang.
6. Orang tua tercinta dan keluarga yang selalu menemani dalam membantu penulis dan selalu mendo'akan serta memberikan dukungan baik kepada penulis.
7. Teman-teman Teknologi Informasi yang selalu memberi dukungan.
8. Semua pihak yang tidak bisa penulis sebutkan satu persatu yang terlibat dalam pembuatan skripsi ini sehingga dapat terselesaikan dengan baik.

Dalam pelaksanaan dan penyusunan skripsi, penulis menyadari bahwa tentunya masih jauh dari kata sempurna dan masih banyak kekurangan. Untuk itu, penulis sangat mengharapkan kritik serta saran yang membangun demi kesempurnaan penulisan skripsi ini, dan semoga skripsi ini dapat bermanfaat untuk semua pihak.

Semarang, 7 November 2022

Penulis

DAFTAS ISI

HALAMAN JUDUL.....	i
PERNYATAAN KEASLIAN	ii
LEMBAR PENGESAHAN.....	iii
NOTA PEMBIMBING	iv
ABSTRAK.....	vi
KATA PENGANTAR.....	vii
DAFTAS ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN	1
A. Latar Belakang.....	1
B. Rumusan Masalah	6
C. Tujuan Penelitian	6
D. Batasan Masalah	7
E. Manfaat Penelitian.....	8
F. Sistematika Penulisan	9
BAB II LANDASAN PUSTAKA.....	11
A. Kajian Pustaka	11
B. <i>Text Mining</i>	14
C. Analisis Sentimen.....	15
D. Media Sosial.....	17
E. Twitter	18
F. <i>Crawling Data</i>	19
G. <i>Text Preprocessing</i>	20
H. TFIDF	22
I. Klasifikasi.....	24
J. <i>Split Validation Data</i>	25
K. <i>Naive bayes classifier</i>	25
L. Evaluasi	27
BAB III METODOLOGI PENELITIAN.....	32
A. Metode Pengumpulan Data	32
1. Studi Pustaka	32

2. Studi Lapangan.....	32
B. Perangkat Penelitian.....	33
C. Alur Pengerjaan Penelitian	33
D. Uraian Metodologi	36
1. Pengambilan Data Twitter	37
2. <i>Preprocessing</i>	39
3. Ekstrasi Fitur.....	45
4. Klasifikasi <i>Naive Bayes</i>	46
5. Uji Model.....	47
6. Evaluasi Model	47
7. Visualisasi.....	48
BAB IV HASIL DAN PEMBAHASAN	49
A. Pengambilan Data Twitter.....	49
B. <i>Preprocessing</i>	56
C. Ekstrasi Fitur.....	71
D. Klasifikasi <i>Naive Bayes</i>	78
E. Uji Model.....	80
F. Evaluasi Model.....	82
G. Visualisasi.....	88
BAB V KESIMPULAN DAN SARAN.....	92
A. Kesimpulan	92
B. Saran.....	93
DAFTAR PUSTAKA.....	94
DAFTAR LAMPIRAN.....	99
LAMPIRAN 1 : Dokumen Hasil <i>Crawling data</i> Twitter .	99
LAMPIRAN 2 : <i>File</i> Data yang Sudah Diberi Label.....	101
LAMPIRAN 3 : <i>File</i> "colloquial-indonesian-lexicon" ...	103
LAMPIRAN 4 : <i>File</i> "FormalizationDict.txt"	104
LAMPIRAN 5 : Data Pendukung "KBBA.txt"	105
LAMPIRAN 6 : Data Pendukung "Slangword.txt"	106
LAMPIRAN 7 : Dokumen Hasil Klasifikasi.....	107

DAFTAR TABEL

Tabel 2. 1 Kajian Pustaka.....	14
Tabel 2. 2 <i>Confusion Matrix 2x2</i>	27
Tabel 2. 3 <i>Multiclass Confusion Matrix 3x3</i>	30
Tabel 3. 1 Kebutuhan Perangkat Keras.....	33
Tabel 3. 2 Kebutuhan Perangkat Lunak.....	33
Tabel 3. 3 Penerapan <i>Labelling</i>	39
Tabel 3. 4 Penerapan Pada Tahapan <i>Case Folding</i>	40
Tabel 3. 5 Penerapan Pada Tahapan <i>Cleansing</i>	41
Tabel 3. 6 Penerapan Pada Tahapan <i>Normalization</i>	42
Tabel 3. 7 Penerapan Pada Tahapan <i>Stopword Removal</i>	43
Tabel 3. 8 Penerapan Pada Tahapan <i>Stemming</i>	44
Tabel 3. 9 Penerapan Pada Tahapan <i>Tokenization</i>	45
Tabel 4. 1 Contoh Perhitungan TF	75
Tabel 4. 2 Contoh Perhitungan DF	76
Tabel 4. 3 Contoh perhitungan IDF.....	77
Tabel 4. 4 Hasil Perhitungan TFIDF.....	78
Tabel 4. 5 <i>Multiclass Confusion Matrix 3 x 3</i>	81
Tabel 4. 6 Hasil <i>Multiclass Confusion Matrix</i>	83
Tabel 4. 7 Perhitungan Nilai Pada Kelas Negatif.....	84
Tabel 4. 8 Perhitungan Nilai Pada Kelas Netral.....	84
Tabel 4. 9 Perhitungan Nilai Pada Kelas Positif.....	84
Tabel 4. 10 Hasil Perhitungan Performa	85

DAFTAR GAMBAR

Gambar 3. 1 <i>Flowchart</i> Alur Pengerjaan Penelitian.....	34
Gambar 3. 2 Twitter API.....	37
Gambar 3. 3 Hasil <i>Crawling data</i>	38
Gambar 3. 4 <i>Tweet</i> dengan opini yang sama.....	41
Gambar 4. 1 <i>Install Library Snsrape</i>	49
Gambar 4. 2 Tampilan Google Colab	50
Gambar 4. 3 <i>SC</i> Pemanggilan <i>Library</i> Proses <i>Crawling</i>	51
Gambar 4. 4 <i>Source Code</i> Proses <i>Crawling</i>	52
Gambar 4. 5 Tampilan <i>Twitter Advanced Search</i>	53
Gambar 4. 6 <i>SC</i> menyimpan data dalam bentuk <i>CSV</i>	54
Gambar 4. 7 Hasil <i>Crawling Data Tweet</i> Pada File <i>CSV</i>	54
Gambar 4. 8 Hasil Data yang sudah diberi Label	56
Gambar 4. 9 <i>Install Library</i> <i>Emoji</i>	58
Gambar 4. 10 <i>SC</i> Pemanggilan <i>Library</i> <i>emoji</i> dan <i>re</i>	58
Gambar 4. 11 <i>SC</i> Tahapan <i>Cleansing</i> dan <i>Case Folding</i>	59
Gambar 4. 12 Hasil tahapan <i>Cleansing</i> dan <i>Case Folding</i>	59
Gambar 4. 13 <i>Source Code</i> <i>Remove Dupilcate</i>	60
Gambar 4. 14 Hasil Dari Tahapan <i>Remove Duplicate</i>	61
Gambar 4. 15 <i>Import library</i> pada tahap <i>normalization</i>	61
Gambar 4. 16 <i>Source Code</i> <i>Import</i> Data Pendukung	62
Gambar 4. 17 Data pendukung kesatu dan kedua.....	63
Gambar 4. 18 Data Pendukung ketiga dan keempat.....	63
Gambar 4. 19 <i>SC</i> implementasi tahap <i>normalization</i> part 1 ..	64
Gambar 4. 20 <i>SC</i> implementasi tahap <i>normalization</i> part 2 .	65
Gambar 4. 21 Hasil dari Tahapan <i>Normalization</i>	65
Gambar 4. 22 <i>Install Library</i> <i>nlp-id</i>	66
Gambar 4. 23 Pendeklarasian <i>Library</i> <i>nlp_id</i>	66
Gambar 4. 24 <i>Source code</i> tahap <i>stopword removal</i>	67
Gambar 4. 25 Hasil dari tahapan <i>Stopword Removal</i>	67
Gambar 4. 26 <i>Install Library</i> <i>Sastrawi</i>	67

Gambar 4. 27 Pendeklarasian <i>Library</i> Sastrawi	68
Gambar 4. 28 <i>Source code</i> tahapan <i>stemming</i>	68
Gambar 4. 29 Hasil dari tahap <i>Stemming</i>	68
Gambar 4. 30 <i>Source Code</i> Tahapan <i>Tokenization</i>	69
Gambar 4. 31 Hasil dari Tahapan <i>Tokenization</i>	69
Gambar 4. 32 Hasil Setelah Melalui <i>Preprocessing</i>	70
Gambar 4. 33 Hasil Total Sentimen yang diperoleh	71
Gambar 4. 34 Tahapan <i>Split Validation Data</i>	72
Gambar 4. 35 Tahapan Proses Pembobotan dengan TFIDF...	73
Gambar 4. 36 Hasil Proses TFIDF.....	74
Gambar 4. 37 <i>Import Library Sklearn</i> pada klasifikasi.....	79
Gambar 4. 38 Proses Pengklasifikasian metode <i>naive bayes</i> .	80
Gambar 4. 39 Hasil Uji Model.....	82
Gambar 4. 40 <i>SC</i> Perhitungan performa <i>Naive Bayes</i>	86
Gambar 4. 41 Hasil Pengukuran Evaluasi Performa	87
Gambar 4. 42 Persentase Pada Sentimen.....	89
Gambar 4. 43 <i>Wordcloud</i> Pada Data Twitter	90
Gambar 4. 44 <i>Wordcloud</i> pada sentimen positif dan negatif.	91

BAB I

PENDAHULUAN

A. Latar Belakang

Bahan Bakar Minyak (BBM) memiliki peran sangat penting dalam kehidupan masyarakat. Bagi masyarakat maupun para pengusaha, BBM tidak lepas dari kebutuhan pokok sehari-hari (Sandy, 2016). Baik secara langsung atau tidak, BBM memberikan dampak yang begitu besar pada masyarakat Indonesia. Karena dampak dari perubahan harga BBM ini mempengaruhi ke berbagai sektor lainnya seperti dari segi distribusi, biaya produksi, transportasi, industri, keekonomiannya, dan demikian pula juga keharga komoditas lainnya (Sandy, 2016).

Kebijakan pemerintah melakukan pemberian informasi penting mengenai kenaikan harga BBM ini menuai banyak sekali pro dan kontra. Bahkan bisa dibilang penyakit tahunan suatu negeri Indonesia dalam kebijakan kenaikan harga BBM ini (Faiq, Ardyanto Putro, et al., 2022).

Terjadinya kenaikan harga BBM dapat menimbulkan kontroversi di kalangan masyarakat. Terutama pada masyarakat menengah ke bawah merasa dirugikan karena kebutuhan yang dibilang sudah mencukupi, ditambah lagi dari kenaikan harga BBM tersebut menjadi semakin banyak biaya

pengeluaran yang dibutuhkan dan tentunya biaya kebutuhan pokok rumah tangga pun juga semakin naik(Sandy, 2015). Di sisi lain, kenaikan BBM ini berdampak baik untuk menghematnya Anggaran Pendapatan Belanja Negara (APBN) seperti apabila harga BBM naik, maka jumlah bantuan/subsidi yang diberikan oleh pemerintah akan berkurang. Selain itu juga, dapat mengurangi pencemaran udara ketika harga BBM naik. Dengan begitu, maka masyarakat akan mengurangi pemakaian bahan bakar tersebut dan beralih ke transportasi angkutan umum.

Saat ini, era internet telah mengubah cara orang mengekspresikan pandangan, opini serta pendapat mereka (Sonawanse, 2016). Salah satu bentuknya melalui media sosial yang cukup terkenal yaitu twitter. Dengan media ini, seseorang dapat secara bebas mengutarakan ide, pendapat, opini dalam bentuk teks atau kalimat(Fajar et al., 2018). Direktur Jenderal Sumber Daya Perangkat Pos dan Informatika (SDPP) Kementerian Komunikasi dan Informatika Budi Setiawan menyatakan “perkembangan dunia teknologi berkembang sangat pesat di dunia tak terkecuali Indonesia dengan mencapai peringkat ketiga di Asia untuk jumlah pengguna internet”. Tercatat sebanyak 19,5 juta pengguna twitter di Indonesia(Mustofa & Mahfudh, 2019).

Tentunya banyak sekali pengguna twitter mengeluarkan tanggapan mereka tentang kenaikan harga BBM dalam bentuk komentar berupa tanggapan positif dan negatif. Oleh sebab itu kumpulan tanggapan, opini, serta keluhan kesah dari respon pengguna twitter terkait penyesuaian pro dan kontra pada kebijakan kenaikan harga BBM ini dapat ditampung dan dimanfaatkan untuk keperluan penelitian.

Banyaknya opini/pendapat yang dituliskan di twitter dapat diklasifikasikan sesuai sentimen yang ada agar mudah untuk mendapatkan kecenderungan tanggapan tersebut terhadap kebijakan kenaikan harga BBM ini, apakah lebih cenderung ke hal yang positif atau negatif. Tentunya dapat diketahui, sifat dari data twitter yang didapatkan memiliki karakteristik yang tidak terstruktur dan memuat banyak sekali *noise*. Sehingga diperlukan *text mining* yang memiliki peranan penting (Fajar et al., 2018).

Text mining ialah proses dari berbagai jenis data yang tidak terstruktur untuk dilakukan pencarian pola-pola informasi data serta pengetahuan yang berguna dari suatu data *text* (Turban, 2011). Salah satu bentuk dari penggunaan *text mining* yaitu analisis sentimen.

Dalam penelitian ini analisis sentimen di terapkan untuk menganalisis sentimen publik terkait kenaikan harga BBM. Analisis sentimen merupakan riset komputasional dari opini,

sentimen dan emosi yang diekspresikan secara tekstual (B, 2010). Tujuan dari analisis sentimen adalah untuk mengelompokkan polaritas dari teks atau opini yang ada dalam respon media sosial di twitter, apakah opini yang ditemukan bersifat positif, negatif atau netral.

Untuk melakukan analisis sentimen ada beberapa macam metode yang digunakan salah satunya yaitu metode algoritma *naive bayes*. *Naive bayes classifier* merupakan algoritma klasifikasi yang sederhana, efisiensi, komputasi tinggi, dan akurasi klasifikasi yang baik terutama untuk data berdimensi tinggi (Mahfudh & Mustofa, 2019). Metode *naive bayes* juga merupakan pengklasifikasian menggunakan metode probabilitas dan statistik dimana dapat memprediksi peluang berdasarkan pengalaman di masa sebelumnya (*Teorema Bayes*) dengan ciri utamanya adalah asumsi yang sangat kuat (*naif*) akan ketergantungan dari masing-masing kondisi/kejadian (Rish, 2001).

Permasalahan analisis sentimen terhadap kebijakan yang diberikan pemerintah terkait kenaikan harga BBM ini perlu untuk diperhatikan oleh kaum muslim karena Allah berfirman dalam Al-Qur'an surah Al-Hujurat ayat 6. Allah SWT, berfirman :

يَا أَيُّهَا الَّذِينَ آمَنُوا إِن جَاءَكُمْ فَاسِقٌ بِنَبَأٍ فَتَبَيَّنُوا أَن تُصِيبُوا قَوْمًا بِجَهَالَةٍ
فَتُصِيبُوا 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-Hujarat 49: Ayat 6).

Dari ayat tersebut menjelaskan bahwa jika kita memperoleh suatu opini dari seseorang yang belum diketahui kejelasannya maka diharuskan melakukan *tabayyun*. Apabila seseorang tidak melakukan *tabayyun*, maka dapat menimpa keburukan kepada diri sendiri maupun orang lain. Dari penjelasan ayat tersebut, maka penulis melakukan penelitian yang salah satu fungsinya adalah *tabayyun* dengan mengumpulkan berbagai macam opini dari masyarakat lalu menganalisisnya sehingga mendapatkan hasil yang berguna bagi masyarakat.

Sehingga pada penelitian ini, penulis melakukan analisis sentimen yaitu sentimen positif dan negatif serta netral menggunakan algoritma *naive bayes* dengan data-data yang diambil dari twitter.

B. Rumusan Masalah

Berdasarkan uraian dari latar belakang tersebut, maka pokok dari rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana mengimplementasikan algoritma *naive bayes* dalam membantu analisis sentimen kenaikan harga BBM pada respon pengguna twitter?
2. Bagaimana performa yang diberikan oleh algoritma *naive bayes classifier* pada analisis sentimen kenaikan harga BBM pada respon pengguna twitter?

C. Tujuan Penelitian

Dari rumusan masalah yang telah diberikan, penelitian ini dilakukan dengan tujuan:

1. Menerapkan algoritma *naive bayes* untuk menganalisis tanggapan positif, negatif dan netral pada respon pengguna twitter.
2. Mengetahui performa dari hasil perhitungan algoritma *naive bayes* untuk klasifikasi tanggapan positif, negatif dan netral terhadap kenaikan BBM pada respon pengguna twitter.

D. Batasan Masalah

Agar penelitian dapat dilakukan secara objektif dan jelas, maka peneliti menerapkan batasan masalah yang diperlukan pada penelitian ini. Berikut batasan masalah pada penelitian ini, adalah :

1. Data yang digunakan adalah data dari media sosial twitter yang berbahasa indonesia
2. Sentimen analisis dilakukan dengan klasifikasi dari metode *naive bayes*.
3. Tanggapan pada twitter akan diklasifikasikan menjadi tiga sentimen yaitu sentimen positif, negatif dan netral.
4. Data twitter kenaikan BBM ini diambil dari twitter pada masa pemerintahan bapak Jokowi dan KH Maruf Amin pada kebijakan tanggal 3 September 2022- 10 September 2022.
5. Kata kunci pencarian yang diteliti pada data twitter yaitu "harga BBM".
6. Menggunakan bahasa pemrograman *python* dan *software google colab (Jupyter notebook versi google)*.
7. Jumlah data yang digunakan adalah 1500 tweet.

E. Manfaat Penelitian

Manfaat dari hasil penelitian ini dibagi menjadi dua, yaitu berdasarkan manfaat secara teoritis dan manfaat secara praktis.

1. Manfaat teoritis :

- a. Membantu untuk mengklasifikasi *tweet*/ tanggapan positif, negatif dan netral.
- b. Mengetahui tingkat performa pada algoritma *naive bayes* dalam melakukan klasifikasi pada kenaikan harga BBM.
- c. Sebagai pijakan, bahan referensi dan pengembangan pada penelitian-penelitian selanjutnya yang berhubungan dengan penelitian sentimen analisis.

2. Manfaat praktis :

- a. Bagi Masyarakat
Dapat mengetahui jumlah respon/tanggapan kebijakan pemerintah mengenai kenaikan harga BBM yang lebih cenderung ke hal yang positif atau negatif.

b. Bagi Pemerintah

- Dapat menjadi bahan pertimbangan pemerintah dengan mengetahui sentimen publik mengenai suatu kebijakan kenaikan harga BBM.
- Dapat menganalisa terhadap hasil kebijakan kenaikan harga BBM dan meminimalisirkan masalah atau solusi dimasa yang akan datang berdasarkan hasil sentimen publik tersebut.

F. Sistematika Penulisan

Laporan penelitian ini secara keseluruhan terdiri dari beberapa bab, agar memudahkan para pembaca untuk memahami isi laporan maka peneliti menunjukkan sistematika penulisan. Berikut sistematika penulisan pada penelitian ini yaitu :

BAB I PENDAHULUAN

Dalam bab pendahuluan membahas latar belakang permasalahan untuk mengetahui alasan dilakukannya penelitian ini kemudian membahas rumusan masalah penelitian, tujuan penelitian, batasan masalah, manfaat penelitian dan sistematika penyusunan laporan.

BAB II LANDASAN PUSTAKA

Dalam bab landasan pustaka membahas kajian pustaka untuk mendukung dilakukan penelitian ini serta membahas teori-teori dasar yang berkaitan dengan penelitian ini yaitu analisis sentimen dengan menggunakan metode *Naive bayes*.

BAB III METODOLOGI PENELITIAN

Dalam bab metodologi penelitian membahas tentang cara peneliti memperoleh data dan kemudian membahas perangkat yang digunakan pada penelitian, kemudian membahas alur pengerjaan penelitian serta gambaran umum yang terdapat pada uraian metodologi.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab hasil dan pembahasan ini, membahas hasil penelitian yang dapat menjawab dari analisis permasalahan pada bab III metodologi penelitian.

BAB V KESIMPULAN DAN SARAN

Dalam bab kesimpulan dan saran, membahas dua point penting yaitu yang pertama, point kesimpulan yang didapatkan dari hasil bab IV yang diuraikan secara singkat dan jelas. Kemudian, yang kedua yaitu point saran yang berisikan saran-saran dari penulis untuk penelitian selanjutnya.

BAB II

LANDASAN PUSTAKA

A. Kajian Pustaka

Pada bagian kajian pustaka ini, akan membahas penelitian-penelitian terkait sebelumnya dengan tujuan untuk memperkuat pelaksanaan penelitian ini. Penelitian terkait yang mengacu pada analisis sentimen pernah dilakukan oleh Ulfa Kurniasih dan Suseno pada tahun 2022 mengenai tanggapan Bantuan Subsidi Upah (BSU) pada kenaikan harga BBM pada masyarakat Indonesia menggunakan metode klasifikasi *naive bayes* yang datanya diambil dari media sosial twitter. Dengan tujuan mengklasifikasikan sentimen berdasarkan respon dari pengguna twitter terhadap tanggapan bantuan subsidi upah pada kenaikan harga BBM. Dengan memberikan beberapa sentimen diantaranya yaitu sentimen positif dan negatif. Pada penelitian ini menggunakan dua kata kunci pencarian untuk data yang digunakan yaitu pada kata BBM dan BSU dengan diperoleh sebanyak 795 data *tweet* setiap masing-masing kata kemudian dilakukan pembobotan kata dengan TFIDF. Dengan hasil akhir diperoleh pada kata kunci BBM mendapatkan lebih banyak respon negatif sebesar 71,8% sedangkan pada kata kunci BSU mendapatkan lebih banyak respon positif sebesar 65,2%. Dari

penelitian ini didapatkan performa dari model *naive bayes* dengan rata-rata tingkat akurasi sebesar 81,64%, presisi sebesar 92,89% dan *recall* sebesar 29,91%(Kurniasih & Suseno, 2022).

Pada penelitian lainnya dilakukan oleh Mujahidin dan kawan-kawannya pada tahun 2022 melakukan implementasi analisis sentimen masyarakat mengenai kenaikan harga BBM dengan data yang diperoleh dari youtube. Dengan tujuan mengklasifikasikan tanggapan-tanggapan kenaikan harga BBM pada youtube dengan memberikan dua sentimen yaitu sentimen positif dan netral. Klasifikasi yang dipakai menggunakan metode *naive bayes* dengan data yang diperoleh sebesar 3053 *dataset* dengan membagi data latih dengan data uji sebesar 80:20. Penelitian ini menggunakan dua vektorisasi kata yaitu *word embedding fasttext* dan *bag of word* sebagai perbandingan untuk menghitung tingkat keakurasian. Dari hasil penelitian ini didapatkan jumlah respon dari sentimen negatif sebesar 66% dan sentimen positif sebesar 34% dengan tingkat akurasi pada vektorisasi kata *word embedding fasttext* sebesar 74% dan tingkat akurasi pada *bag of word* memiliki nilai yang lebih rendah sebesar 68%(Mujahidin et al., 2022).

Penelitian selanjutnya pada tahun 2016 yaitu analisis sentimen mengenai kenaikan harga BBM pada media *online* yang dilakukan oleh Sandy dan teman-temannya. Dengan data

yang diperoleh dari media *online* (kompas.com dan detik.com) dengan data diambil dari rentang waktu pada tahun 2014 dan 2015. Dengan menggunakan metode *grounded tehory*. Pada penelitian ini menggunakan dua sentimen yaitu positif dan negatif. Dari hasil yang didapatkan secara keseluruhan menunjukkan tanggapan kenaikan harga BBM pada sentimen negatif jauh lebih besar dengan angka sebesar 68% sedangkan tanggapan positif sebesar 32%(Sandy et al, 2016).

Berdasarkan penelitian yang telah dilakukan oleh beberapa penulis di atas, dapat dijadikan acuan dalam melakukan penelitian. Peneliti tertarik untuk melakukan penelitian mengenai analisis sentimen terhadap kebijakan kenaikan harga BBM pada media sosial twitter menggunakan metode *naïve bayes*. Penelitian ini berupaya menghasilkan performa yang baik dari metode *naïve bayes* dalam mengklasifikasikan sentimen kebijakan tersebut berdasarkan tiga sentimen yang dimiliki, yaitu sentimen positif, negatif dan netral dengan data yang diperoleh dari twitter pada kata kunci pencarian 'harga BBM' yang belum pernah dilakukan oleh peneliti sebelumnya. Berikut pada tabel kajian pustaka untuk melihat perbandingan penelitian-penelitian terkait dengan penelitian yang dilakukan dapat ditunjukkan pada tabel 2.1.

Tabel 2. 1 Kajian Pustaka

No	Pustaka	Topik	Metode	Objek	Klasifikasi
1	(Kurniasih & Suseno, 2022)	Tanggapan mengenai Bantuan Subsidi Upah pada kenaikan harga BBM	<i>Naive Bayes</i>	Twitter	2 Kelas (Positif dan Negatif)
2	(Mujahidin et al., 2022)	Tanggapan mengenai kenaikan harga BBM	<i>Naive Bayes</i>	Youtube	2 Kelas (Positif dan Negatif)
3	(Sandy et al., 2016).	Tanggapan mengenai kenaikan harga BBM	<i>Grounded Theory</i>	Media Online (Kompas .com dan Detik .com)	3 Kelas (Positif, Negatif dan Netral)
4	Yang sedang dikerjakan	Tanggapan mengenai kenaikan harga BBM	<i>Naive Bayes</i>	Twitter	3 Kelas (Positif, Negatif dan Netral)

B. Text Mining

Text mining adalah proses untuk memberi informasi serta pengetahuan yang bermanfaat dengan cara menambang data yang sumbernya dari teks, kutipan, dokumen dan lain sebagainya dengan maksud untuk mencari kata-kata yang menginterpretasikan setiap teks atau dokumen sehingga dapat

diketahui keterhubungan antar dokumen satu dengan lainnya (Imron, 2019). *Text mining* juga merupakan bagian dari data *mining* bedanya bentuknya lebih tidak terstruktur atau disebut sebagai *unstructured* data sedangkan data mining datanya lebih terstruktur (Sholekha et al., 2022). *Text mining* merupakan tahapan proses dari analisis dalam data yang berupa teks biasanya digunakan dalam klasifikasi dokumen tekstual dimana dokumen-dokumen tersebut akan diklasifikasikan sesuai dengan topik dokumen yang ada (Darwis et al., 2021).

Text mining bertujuan untuk menganalisis pendapat, sentimen, evaluasi, sikap, penilaian, emosi seseorang sehingga dapat diketahui apakah berkenaan dengan suatu topik, layanan, organisasi, individu, atau kegiatan tertentu. Penggunaan dari *text mining* dilakukan untuk klasterisasi, klasifikasi, *information retrieval*, dan *information extraction*.

C. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang ilmu dari *text mining*, *natural language program*, dan *artificial intelligence* (Luuqyana, 2018). Analisis Sentimen adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, dimana entitas tersebut dapat mewakili individu (Syakuro, 2017).

Analisis sentimen disebut juga sebagai *opinion mining* yang berguna dalam pengelolaan bahasa alami, komputasi linguistik, dan *text mining*. Tujuan dari analisis sentimen ialah untuk menganalisis sentimen masyarakat seperti sikap, opini, emosi, dll terhadap beberapa elemen seperti, produk, individu, topik, organisasi, dan layanan (Sonawanse, 2016). Adapun juga yang menyebutkan analisis sentimen untuk mengklasifikasikan teks sebagai point utamanya dengan memberikan hasil dalam bentuk teks positif, negatif dan netral (Luuqyana, 2018).

Untuk memberi gambaran singkat, berikut adalah berbagai jenis parameter analisis sentimen :

1. Polaritas – Inti dari jenis parameter analisis sentimen ini tertuju pada tanggapan atau komentar yang diterima pada kebijakan pemerintah tentang kenaikan harga BBM secara *online* (positif, netral, dan negatif).
2. Emosi – Intinya tertuju pada emosi yang muncul dari masyarakat pada saatnya diumumkannya kebijakan pemerintah tentang kenaikan harga BBM (senang, sedih, kecewa, bersemangat, dan lain lain)
3. Urgensi – intinya tertuju pada kedekatan pada kebijakan pemerintah terhadap kenaikan BBM atau menemukan solusi efektif untuk memecahkan masalah

pada masyarakat atas kebijakan ini (mendesak dan dapat menunggu)

4. Niat – intinya tertuju untuk mencari tahu apakah masyarakat tertarik dengan kebijakan terkait kenaikan harga BBM ini atau tidak.

D. Media Sosial

Media sosial merupakan media *online* yang dilakukan menggunakan internet, yang para pemakai media ini dapat bertukar, berbagi serta mengutarakan informasi selayaknya aktivitas didunia secara *online* melalui suatu *platform* digital. Pandangan lain mengatakan bahwa media sosial adalah media yang dapat *support* interaksi sosial serta mengubah pandangan dari komunikasi secara pasif menjadi interaktif dengan transformasinya digitalisasi (Faiq, Putro, et al., 2022). Media sosial tentu berbeda dengan media *online* karena penggunaan yang kuat dalam membentuk opini yang berkembang di masyarakat (Sandy, 2015).

Ada banyak sekali contoh media sosial yang lagi kekinian seperti instagram, facebook, twitter dan lainnya sehingga banyak sekali pengguna dari negara Indonesia yang mengakses sebuah media sosial ini hingga berjuta-juta (Sujadi, 2022). Media sosial ini gratis hanya bermodalkan gawai atau alat elektronik komunikasi lainnya dan internet. Bahkan telah menjadi bagian dari kehidupan atau kebutuhan sehari-hari

karena penggunaannya dari media sosial banyak digunakan di seluruh dunia(Luuqyana, 2018). Beberapa faktor alasan penggunaan media sosial dianggap penting yaitu karena faktor informasi yang cukup terbilang sangat cepat penyebarannya, faktor interaksi sosial antar pengguna satu dengan lainnya tanpa mengenal jarak dan waktu, lalu adapun juga karena faktor hiburan karena dapat berbagi momen-momen lucu, senang dll.

E. Twitter

Twitter merupakan media sosial yang dapat bebas mengungkapkan ide, pikiran ataupun emosi yang dituliskan dalam bentuk teks seperti layaknya *blog* tetapi versi kecilnya dengan batasan 280 karakter atau yang biasa disebut *tweet* yang dapat *dishare*, *diposting*, dikomentari serta dilihat oleh pengguna twitter lainnya(Fajar et al., 2018). Kelebihan pada media sosial ini salah satunya twitter menyediakan layanan API (*Application Programming Interface*) yang sangat baik, sehingga memudahkan setiap orang untuk mengambil data dari twitter. Pengumpulan data dari twitter dapat digunakan untuk berbagai kebutuhan seperti, mengetahui popularitas kandidat pilkada atau pemilu, mendapat informasi mengenai popularitas suatu produk atau untuk yang sederhana dapat digunakan untuk melihat semua *mention*, *retweet* atas suatu akun twitter tertentu (Mas Pintoko & Muslim, 2018).

Twitter dipilih karena twitter sebagian besar cocok untuk dilakukan teks *mining* karena tiga point utama (Wandani, 2021):

1. Twitter API memiliki desain yang bagus dan mudah di akses.
2. Data twitter dalam format yang nyaman untuk dianalisis.
3. Kebijakan twitter untuk data relatif liberal dibandingkan dengan API yang lain.

Keterbatasan *platform* media sosial lain seperti facebook adalah mereka tidak sembarangan mengizinkan akses mudah ke data mereka karena kebijakan privasi yang berbeda-beda. Tidak mudah mengumpulkan data secara terbuka dan otomatis dengan media lain semacam itu. Opini dari sebuah *tweet* dapat ditemukan di dalam bagian teks dari *tweet* tersebut. Penarikan data dalam bentuk teks bebas yang tidak terstruktur dan tidak berstandar(Wandani, 2021).

F. *Crawling* Data

Crawling data di twitter adalah suatu proses untuk mengambil atau mengunduh data dari server twitter dengan bantuan *Application Programming Interface* (API) twitter baik berupa data *user* maupun data *tweet*(Mas Pintoko & Muslim, 2018). Tujuan *crawling* data ini dilakukan untuk mengambil data dari twitter dimana data tersebut dibutuhkan pada

penelitian ini. Cara melakukan *crawling* data ialah dengan membuat program dengan memasukkan kata kunci untuk mencari *tweet* sesuai yang kita inginkan. Misalnya, “#kenaikanhargaBBM” program akan mengambil *tweet* yang *mention* ke hastag/tagar pada kenaikan harga BBM. Kumpulan *tweet* tersebut merupakan data yang akan digunakan.

G. Text Preprocessing

Text Preprocessing merupakan tahapan paling penting karena pada proses ini dilakukan perubahan data yang masih mentah menjadi data sesuai anjuran proses *mining* (Fikri et al., 2020). Pada analisis sentimen penting untuk dilakukan proses *text preprocessing* (Pradany & Fatichah, 2016). Tujuan dilakukannya *text preprocessing* adalah untuk menghilangkan *noise*, menyeragamkan bentuk kata dan mengurangi volume kata (Rustiana, 2017). Pada penelitian ini tahap *text preprocessing* terdiri dari beberapa proses dimulai dari proses *case folding*, *cleansing*, *remove duplicate*, *normalizaion*, *stopword removal*, *stemming*, *tokenization*.

1. Case Folding

Pada tahap ini, suatu kata, kalimat atau teks yang mengandung huruf kapitalnya disamaratakan menjadi huruf kecil. Yang bertujuan untuk menyeragamkan bentuk kata.

2. *Cleansing*

Pada tahap ini, pada suatu kata, kalimat atau teks untuk menghilangkan simbol, tanda baca, dan bilangan angka. Seperti yang diketahui dari data twitter yang khas yaitu *username* (@usernamepadatwitter123), URL (*Uniform Resource Locator*), dan "RT" (tanda *retweet*). Karena ketiga komponen itu tidak memiliki peran apapun dalam melakukan sentimen maka secara sistem akan dihapus. Yang bertujuan untuk mengurangi *noise*.

3. *Remove Duplicate*

Pada tahap ini, dilakukan penghapusan data berulang untuk menghapuskan data *tweet* yang berisi opini yang sama.

4. *Normalization*

Pada tahap ini, untuk mengganti kata tidak baku, kata singkat menjadi kata baku yang sesuai anjuran pada KBBI.

5. *Stopword Removal*

Pada tahap ini, untuk menghilangkan atau menghapus daftar kata umum yang tidak memiliki arti penting atau informasi yang dibutuhkan. Yang bertujuan untuk mengurangi volume kata yang disimpan pada sistem.

6. *Stemming*

Pada tahap ini, untuk mengubah suatu kata yang berimbuhan menjadi kata dasar, dan juga mengubah kata perulangan kata menjadi satu kata dasar.

7. *Tokenization*

Pada tahap ini, untuk memecah teks suatu kalimat menjadi beberapa bagian-bagian kata. Pada proses ini akan menghilangkan *whitespace*.(Luuqyana, 2018)

H. TFIDF

Analisis sentimen memiliki data mentah berbentuk teks, tentunya data teks ini tidak dapat dibaca langsung oleh komputer, sehingga diperlukan sebuah metode yang mengubah data yang berbentuk teks menjadi sekumpulan angka yang dapat diproses oleh komputer atau sistem(Yerzi & Sibaroni, 2021). Metode TFIDF (*Term Frequency Inverse Document Frequency*) merupakan suatu metode algoritma yang memberikan bobot terhadap teks(Fikri et al., 2020). TF ialah jumlah sebuah kata dari tiap-tiap dokumen, sedangkan IDF ialah nilai *invers* dari jumlah dokumen yang mengandung kata tersebut. Banyaknya kemunculan kata di tiap-tiap dokumen yang ada menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Banyaknya dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata-kata itu. Mudahnya, pada TFIDF untuk kata yang sering

muncul memiliki bobot nilai yang cenderung kecil, untuk kata yang jarang muncul memiliki bobot nilai yang cenderung relatif besar karena kata-kata yang sering muncul biasa disebut *stopword* yang biasanya dianggap kurang penting (Sujadi, 2022). Pada proses perhitungan TF dan IDF akan dikalikan sehingga menghasilkan nilai bobot dari kata tersebut (Fikri et al., 2020). Berikut persamaan TFIDF, yang ditunjukkan pada persamaan 2.1.

$$TFIDF(d,t) = TF(d,t) \cdot IDF(t) \quad (2.1)$$

Dimana d adalah dokumen dan t adalah kata. Sedangkan $TF(d,t)$ merupakan jumlah kata pada tiap dokumen yang dirumuskan pada persamaan 2.2.

$$TF(d,t) = \frac{\text{Jumlah kata } t \text{ pada dokumen } d}{\text{total kata pada dokumen } d} \quad (2.2)$$

Pada IDF merupakan jumlah dokumen yang mengandung kata tersebut yang dirumuskan pada persamaan 2.3.

$$IDF(t) = \log \frac{\text{total dokumen}}{\text{jumlah dokumen yang mengandung kata } t} + 1 \quad (2.3)$$

I. Klasifikasi

Klasifikasi suatu proses pengkategorian data atau suatu data-data yang dapat memenuhi syarat tertentu. Pendapat lain mengatakan klasifikasi merupakan ilmu dari data *mining* untuk melakukan prediksi penggolongan kelas dari satu data tersebut (Luuqyana, 2018). Klasifikasi kompleks dikatakan apabila pada pengkategorian objek memiliki tiga kelas atau lebih. Pada penelitian ini dilakukan tiga pengklasifikasian pada tiga *class* yaitu positif, negatif dan netral.

Cara kerjanya dimulai proses pengklasifikasian pada sistem yang harus dilakukan pelatihan atau *learning* untuk mencari analisa dari data latih agar dapat dilakukan prediksi. Setelah dilakukan proses pelatihan, selanjutnya dapat dilakukan proses pengklasifikasian. Pada proses klasifikasi dilakukan untuk memprediksi keakurasian data yang didapat dari hasil pemahaman yang digunakan pada pemberian data latih kemudian diujikan dengan data *test* atau data uji. Sehingga apabila didapatkan tingkat keakurasian yang baik dengan membandingkan antara hasil prediksi dari sistem dengan hasil sebenarnya, maka sistem dari model tersebut dapat digunakan untuk memprediksi kelas ke data-data yang belum diketahui kelasnya.

J. *Split Validation Data*

Split validation data merupakan teknik validasi yang membagi dua data bagian secara acak, bagian data pertama digunakan sebagai data latih (*training*) dan bagian data kedua digunakan sebagai data uji (*testing*) (Turmudi Zy et al., 2021).

Data latih (*training*) adalah data yang akan dipakai dalam melakukan proses pembelajaran (*learning*) sedangkan data uji (*testing*) adalah data yang belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian (Turmudi Zy et al., 2021).

Pada analisis ini akan mempartisi *dataset* dengan rasio 80:20 dengan artian 80% dilakukan sebagai data latih dan 20% dilakukan sebagai data uji. Penelitian terkait pernah dilakukan oleh Putri dan Kharisudin (2022) dengan menghasilkan tingkat akurasi yang baik (Putri & Kharisudin, 2022).

K. *Naive bayes classifier*

Naive bayes classifier merupakan sebuah metode klasifikasi dengan probabilitas sederhana yang mengaplikasikan *teorema bayes* dengan asumsi ketidaktergantungan (*independen*) yang tinggi (Nurhuda et al., 2013). Penggunaan metode *naive bayes* pada penelitian ini didasarkan pada banyaknya dataset yang dipakai sehingga membutuhkan suatu metode yang mempunyai performansi

yang cepat dalam pengklasifikasian serta keakuratan yang cukup tinggi (Nurhuda et al., 2013).

Salah satu metode dalam klasifikasi data adalah *naïve bayes*. *Naive bayes* juga merupakan salah satu metode *machine learning* yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, cara kerja *naive bayes* yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Rumus *naive bayes* dalam (Darwis et al., 2021) secara umum dapat dilihat pada persamaan 2.4.

$$P(Y | X) = \frac{P(X|Y) \cdot P(Y)}{P(X)} \quad (2.4)$$

Keterangan :

Y = Hipotesis data X dari kelas yang spesifik

X = Data dengan kelas yang belum diketahui

$P(Y|X)$ = Probabilitas hipotesis H berdasarkan kondisi X
(*posterior*)

$P(X|Y)$ = Probabilitas X berdasarkan kondisi tersebut
(*Likelihood*)

$P(Y)$ = Probabilitas hipotesis H (*prior*)

$P(X)$ = Probabilitas dari X (Evidance)

L. Evaluasi

Evaluasi merupakan langkah untuk mengukur keberhasilan suatu sistem dengan membandingkan hasil pengujian pada sistem dengan standar yang telah ditetapkan (Parikh & M.M, 2009). Secara umum tabel *confusion matrix* (matriks klasifikasi atau tabel kontigensi) digunakan untuk menentukan performa model *naive bayes* untuk mengklasifikasikan data, yang diterapkan untuk mengevaluasi hasil pengujian pada sistem dalam analisis sentimen (Apriani et al., 2019). Berikut pada tabel. 2.2 menampilkan pengklasifikasian kedalam dua kelas pada tabel *confusion matrix* (Fawcett, 2006).

Tabel 2. 2 Confusion Matrix 2x2

		<i>True Class</i>	
		<i>Positive</i>	<i>Negative</i>
<i>Predicted Class</i>	<i>Positive</i>	TP	FP
	<i>Negative</i>	FN	TN

Dalam *Confusion Matriks* tersebut, ada empat nilai yang dijadikan acuan dalam perhitungan, yaitu :

1. TP (*true positive*) , Yaitu data yang diprediksi positif dan faktanya data itu positif (Sesuai).
2. TN (*true negative*) : Yaitu data yang diprediksi negatif dan faktanya data itu negatif (Sesuai).
3. FP (*false positive*) : Yaitu data yang diprediksi positif dan faktanya data itu negatif (Tidak Sesuai).
4. FN (*false negative*) : Yaitu data yang diprediksi negatif dan faktanya data itu positif (Tidak Sesuai).

Untuk pengukuran performa model *naive bayes classifier* cara yang digunakan adalah menghitung nilai akurasi, *precision*, *recall* dan *f1 score*(Grandini et al., 2020).

Akurasi merupakan ketepatan suatu sistem melakukan pengklasifikasian yang benar (Sesuai). Cara perhitungannya dengan cara membagi jumlah yang diprediksi sistem yang sesuai dengan jumlah keseluruhan data uji. Perhitungan untuk akurasi dapat dikalkulasi dan dilihat pada persamaan 2.5.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (2.5)$$

Precision merupakan seberapa besar tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Untuk menghitung nilai *precision* dapat dilihat pada persamaan 2.6.

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (2.6)$$

Recall merupakan seberapa besar tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Untuk menghitung nilai *recall* dapat dilihat pada persamaan 2.7.

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (2.7)$$

F1 Score merupakan parameter tunggal ukuran keberhasilan retrieval yang menggabungkan *recall* dan *precision*. Untuk menghitung nilai *f1 score* dapat dilihat pada persamaan 2.8.

$$f1 \text{ Score} = 2 \times \frac{Precision \cdot Recall}{Precision + Recall} \times 100\% \quad (2.8)$$

Dalam penelitian ini, penulis menggunakan *multiclass confusion matrix* 3x3 dikarenakan *output* sentimen dari penelitian ini ada tiga *class*, yaitu negatif, netral dan positif.

Sehingga tabel *multiclass confusion matrix*-nya dapat dilihat pada tabel 2.3(Grandini et al., 2020).

Tabel 2. 3 Multiclass Confusion Matrix 3x3

		<i>True Class</i>		
		<i>Negative</i>	<i>Netral</i>	<i>Positif</i>
<i>Predicted Class</i>	<i>Negative</i>	T Neg	F NegNet	F NegPos
	<i>Netral</i>	F NetNeg	T Net	F NetPos
	<i>Positif</i>	F PosNeg	F PosNet	T Pos

Dalam *multiclass confusion matriks* tersebut, terdapat sembilan nilai yang dijadikan acuan dalam perhitungan, yaitu :

- a. T Pos (*True Positive*), Yaitu jumlah data yang diprediksi positif dan faktanya data itu positif (Sesuai).
- b. F PosNeg (*False Positive Negative*), Yaitu jumlah data yang diprediksi positif dan faktanya data itu Negatif.
- c. F PosNet (*False Positive Netral*), Yaitu jumlah data yang diprediksi positif dan faktanya data itu Netral.
- d. F NegPos(*False Negative Positive*), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu positif.

- e. T Neg (*True Negative*), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu negatif (Sesuai).
- f. F NegNet (*False Negative Netral*), Yaitu jumlah data yang diprediksi negatif dan faktanya data itu netral.
- g. F NetPos (*False Netral Positive*), Yaitu jumlah data yang diprediksi netral dan faktanya data itu positif.
- h. F NetNeg (*False Netral Negative*), Yaitu jumlah data yang diprediksi netral dan faktanya data itu negatif.
- i. T Net (*True Netral*), Yaitu jumlah data yang diprediksi netral dan faktanya data itu netral (Sesuai).

Tentunya setelah didapatkan nilai performa dari klasifikasi model *naive bayes*. Peneliti melakukan ambang batas nilai pada performa apabila diperoleh dengan nilai diatas 80% maka dinyatakan model pengklasifikasian tersebut dinyatakan layak digunakan (baik). Sebaliknya apabila menghasilkan dibawah 80% sebaiknya peneliti selanjutnya melakukan pengujian dengan metode lain agar memperoleh hasil yang lebih maksimal(Tangkelayuk & Mailoa, 2022).

BAB III

METODOLOGI PENELITIAN

A. Metode Pengumpulan Data

1. Studi Pustaka

Penulis melakukan studi pustaka dengan memanfaatkan buku-buku, jurnal, skripsi dan sejenisnya untuk mempelajari materi yang berhubungan dengan konsep, analisis, dan permasalahan lainnya yang diteliti oleh penulis. Tidak hanya itu, penulis juga melakukan pencarian data secara daring dengan menggunakan *browser* melalui internet dengan mengunjungi suatu *website* yang berhubungan dengan analisis sentimen, *text mining*, dan algoritma *naive bayes clasification*.

2. Studi Lapangan

Pada metode ini, penulis melakukan pengamatan secara langsung terhadap segala aktivitas yang sedang *trending topic* pada media sosial tentang kenaikan harga BBM, salah satunya terhadap respon pengguna media sosial twitter di Indonesia yang diteliti oleh penulis.

B. Perangkat Penelitian

Secara khusus untuk implementasi sistem membutuhkan beberapa perangkat keras dan perangkat lunak yang harus dipenuhi untuk berjalanya proses sistem. Adapun kebutuhan yang digunakan pada penelitian ini sebagai berikut:

1. Kebutuhan perangkat keras:

Tabel 3. 1 Kebutuhan Perangkat Keras

No.	Perangkat Keras	Spesifikasi
1.	Device	HP Laptop 14s-dk0xxx
2.	Processor	AMD Ryzen 3 3200U
3.	Memori (RAM)	8,00 GB
4.	Monitor	14 inch
5.	Keyboard dan Mouse	Standard

2. Kebutuhan perangkat lunak:

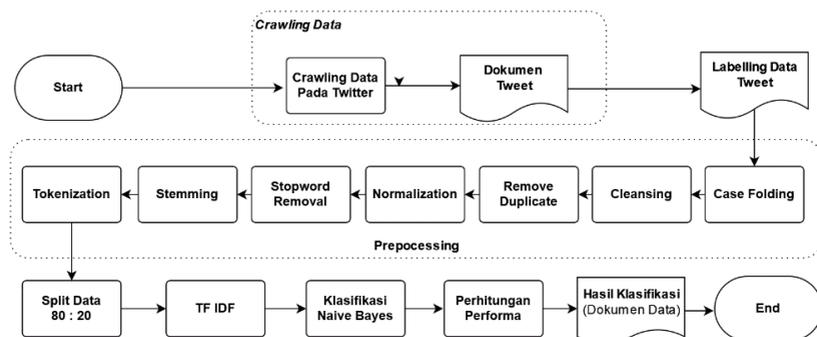
Tabel 3. 2 Kebutuhan Perangkat Lunak

No.	Perangkat Lunak	Spesifikasi
1.	Sistem Operasi	Windows 11 64-bit
2.	Bahasa Pemrograman	Python
3.	MS. Office	Ms. Word, Ms Excel 2019
4.	<i>Browser</i>	Opera
5.	<i>Google Drive</i>	Google Colab

C. Alur Pengerjaan Penelitian

Perancangan alur pengerjaan penelitian akan membahas mengenai gambaran umum yang dilakukan peneliti dalam

tahapan pengerjaan yang dimulai dari awal hingga akhir. Berikut rangkaian pengerjaan tugas akhir dapat dilihat pada gambar 3.1 yaitu *flowchart* alur pengerjaan penelitian yang ada di bawah ini (Wenty Dwi Yuniarti, 2019).



Gambar 3. 1 *Flowchart* Alur Pengerjaan Penelitian

Langkah awal dari pengerjaan penelitian ini ialah mengumpulkan data opini-opini/tanggapan dari aplikasi twitter dengan cara *crawling* data. Proses *crawling* diambil otomatis melalui API twitter yang berkaitan dengan pencarian kenaikan harga BBM. Kemudian setelah semua data sudah didapatkan lalu dilakukan proses pelabelan data untuk menentukan sentimen opini/tanggapan yang ada. Pada penelitian ini menggunakan tiga sentimen yaitu sentimen positif, negatif dan netral.

Langkah Kedua, dilakukan tahap *preprocessing* bertujuan agar data yang masih mentah dapat dibersihkan sehingga bisa dilakukan pengklasifikasian. Pada tahap *preprocessing* terdiri dari tujuh, yaitu *case folding*, *stemming*, *removal duplicate*, *stopword removal*, *stemming* dan *tokenizations*. Dimulai dari tahap *case folding*, suatu teks disamaratakan yang terdapat huruf kapitalnya menjadi huruf kecil. Kemudian setelah pada tahap *case folding* dilakukan tahapan *cleansing* tugasnya untuk melakukan penghapusan apabila ada pada URL, tanda baca, dan "@" pada *username* di twitter tujuannya agar mengurangi *noise* pada *text*. Lalu, dilakukan *removal duplicate* untuk menghapus data *tweet/opini* yang berulang atau memiliki kesamaan. Setelah itu, dilakukan *normalization* untuk mengganti kata tidak baku, kata singkat menjadi kata baku yang sesuai anjuran pada KBBI. Dilanjutkan pada tahapan *stopword* menghapus kata pada kalimatnya yang berisi kata-kata yang bersifat umum. Lanjut pada tahap *stemming* dilakukan perubahan kata yang memiliki imbuhan menjadi kata dasar. Lalu pada tahap *tokenizataion* dilakukan pemecahan suatu kalimat menjadi beberapa bagian kata-kata dan juga menghilangkan *whitespace*.

Langkah Ketiga, ialah melakukan tahapan *split validation data* dengan data dibagi antara data latih dengan data uji

dengan rasio 80:20 dimana 80% dari keseluruhan jumlah data dijadikan data latih dan 20% dari keseluruhan data dijadikan data uji yang diambil secara acak. Setelah itu, masuk ke tahapan selanjutnya yaitu ekstraksi fitur yang merupakan proses mengubah kata menjadi bilangan angka serta dilakukan pembobotan nilai kata dengan menggunakan TFIDF untuk mempermudah jalannya proses pengklasifikasian *naive bayes*.

Kemudian ke tahap keempat, pada tahapan yang paling utama yaitu proses pengklasifikasian dengan metode *naive bayes*. Dengan mengklasifikasikan berdasarkan sentimen yang ada pada dokumen. Sesudah itu, masuk ke tahapan uji model untuk mengetahui ketepatan klasifikasi dengan membandingkan hasil prediksi dari suatu sistem dengan hasil dari data yang sudah diberi label sebelumnya sehingga menghasilkan *multiclass confusion matrix*. Setelah uji model dilakukan, lalu pada tahap evaluasi model untuk menentukan tingkat keakurasian, *precisson*, *recall* dan *f1 score* melalui metode *multiclass confusion matrix* pada tabel dengan tujuan mengetahui performa dari model tersebut.

D. Uraian Metodologi

Pada tahap ini dilakukan penguraian pada alur pengerjaan penelitian diatas untuk menjelaskan lebih dalam. Diantaranya sebagai berikut.

1. Pengambilan Data Twitter

Data yang digunakan dalam penelitian ini adalah data dari respon pengguna media sosial twitter yang berisikan opini/tanggapan mereka terkait kenaikan harga BBM. Data yang dikumpulkan berupa data teks yang diambil dengan teknik *crawling* menggunakan secara otomatis dengan API twitter. Cara melakukan *crawling* data ialah dengan membuat program dengan memasukkan kata kunci untuk mencari *tweet* sesuai yang kita inginkan. Pada penelitian ini, pencarian yang digunakan dengan kata kunci “hargaBBM” maka program akan mengambil *tweet* yang *mention* ke *hashtag*/tagar pada kenaikan harga BBM sehingga didapatkan data sebanyak 1500 *tweet*. Kumpulan *tweet* tersebut merupakan data yang akan digunakan. Berikut tampilan halaman twitter API pada gambar 3.2



Gambar 3. 2 Twitter API

Kemudian setelah dilakukan proses pengambilan data (*crawling data*) lalu didapatkan file dalam bentuk csv. Setelah itu, dapat dilakukan proses pelabelan secara manual untuk menentukan sentimen-sentimen dari komentar tersebut. Berikut hasil *crawling data* pada gambar 3.3.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	usernamelcontent																					
2	DitfArDm,kam mendengar dan merasa jantungku tapi raga paku mangir seluruh kadar pdmokrat maniak kenaikan harga bom																					
3	Derasand37622, hanya peduli proyek kencana cepat dan lin pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah hamparan ekonomi yg sulit dgn teganya justru menaikkan harga bom kata ad																					
4	Tham_Amura,bem dan geram gelar demo poodolan kenaikan harga bom ini tuntutan yang disampaikan																					
5	Irfpa_...kijakan ini tentu menguntungkan korporasi dan mengkan rakyat, sebab harus membuat bom dengan harga yang relatif mahal																					
6	Rochbutor,perjan ketanya ditata rayi tariff angkutan umum melawanya ga menyulitkan warga yg tinggal di disuamunya seela ditumukan, meskipun presiden yg ga ga tau siapa itulah namanya ga sudah masuk sat																					
7	Irfpa_...karena itu dibaga kuat pascabutan subsidi bom dan perlawanan bom berakan rendah adalah untuk melepas harga bom ke pasar sehingga produk bom korporasi asing dapat bersaing dengan pertama																					
8	Irfpa_...dengan kondisi masyarakat yang sebagian besar bertaraf hidup menengah ke bawah, bom dengan harga murah menjadi pilihan																					
9	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
10	Derasand37622, hanya peduli proyek kencana cepat dan lin pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah hamparan ekonomi yg sulit dgn teganya justru menaikkan harga bom kata ad																					
11	Derasand37622, hanya peduli proyek kencana cepat dan lin pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah hamparan ekonomi yg sulit dgn teganya justru menaikkan harga bom kata ad																					
12	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
13	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
14	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
15	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
16	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
17	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
18	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
19	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
20	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
21	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
22	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
23	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
24	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
25	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
26	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
27	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
28	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
29	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
30	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
31	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
32	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
33	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
34	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
35	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
36	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
37	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
38	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
39	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					
40	Irfpa_...adagiat sejak lama diketahui bahwa ada upaya pemerintah, yakni pertama untuk melepas harga bom ke pasar																					

Gambar 3. 3 Hasil *crawling data*

Pada proses pelabelan data, agar sistem dapat mudah melalui proses pengklasifikasian, maka pada kelas sentimen negatif diubah menjadi angka 0, kelas sentimen netral diubah menjadi angka 1, sedangkan untuk kelas sentimen positif diubah menjadi angka 2. Contoh penerapan *labelling* dapat dilihat pada tabel 3.3.

Tabel 3. 3 Penerapan *Labelling*

Opini	Sentimen	Label
@Hilmi28 Dulu ada yg sampai menangis karena harga BBM naik. Sekarang minta rakyat jgn cengeng menghadapi kenaikan harga. Jadi dulu tangisan kalian itu benar2 peduli rakyat atau Cuma sekedar politisi ?? #BBMnaik	Negatif	0
@Agusss_10 Gw setuju2 aja sih BBM naik, tapi harga sembakonya jadi ikutan naik.	Netral	1
@Intannn21 Lebih baik naikkan BBM dari pada langkah, salam dari kalimantan.	Positif	2

2. *Preprocessing*

Tujuan dari tahap ini agar data yang masih mentah dapat dibersihkan sehingga bisa dilakukan pengklasifikasian. Tahapan *preprocessing* dilakukan berdasarkan data yang diperoleh dari proses pengambilan data tersebut, yaitu data dari respon pengguna media sosial twitter. Adapun urutan dari tahapan *preprocessing* yang dilakukan ialah :

a. *Case Folding*

Pada tahap ini, suatu kata, kalimat atau teks yang mengandung huruf kapitalnya disamaratakan menjadi huruf kecil atau *lowercase*. Yang bertujuan

untuk menyamakan/menyeragamkan bentuk kata. Contoh proses tahap *case folding* yang terdapat pada Tabel 3.4 sebagai berikut.

Tabel 3. 4 Penerapan Pada Tahap *Case Folding*

<i>Input Process</i>	<i>Output Process</i>
@Hilmi28 Dulu ada yg sampai menangis karena harga BBM naik. Sekarang minta rakyat jgn cengeng menghadapi kenaikan harga. Jadi dulu tangisan kalian itu benar2 peduli rakyat atau Cuma sekedar politisi ?? #BBMnaik	@hilmi28 dulu ada yg sampai menangis karena harga bbm naik. sekarang minta rakyat jgn cengeng menghadapi kenaikan harga. jadi dulu tangisan kalian itu benar2 peduli rakyat atau cuma sekedar politisi ?? #bbmnaik

b. *Cleansing*

Seringnya muncul dari data twitter yang khas yaitu pada *username* (@username padatwitter123), URL (*Uniform Resource Locator*), "RT" (tanda retweet) bilangan angka, simbol dan tanda baca membuat data menjadi tidak digunakan. Karena ketiga komponen-komponen tersebut tidak memiliki peran apapun dalam melakukan sentimen maka secara sistem akan dihapus. Yang bertujuan untuk mengurangi *noise*. Contoh proses tahap *cleansing* yang terdapat pada Tabel 3.5 sebagai berikut.

Tabel 3. 5 Penerapan Pada Tahap *Cleansing*

<i>Input Process</i>	<i>Output Process</i>
@Hilmi28 dulu ada yg sampai menangis karena harga bbm naik. sekarang minta rakyat jgn cengeng menghadapi kenaikan harga. jadi dulu tangisan kalian itu benar2 peduli rakyat atau cuma sekedar politisi ?? #bbmnaik	dulu ada yg sampai menangis karena harga bbm naik sekarang minta rakyat jgn cengeng menghadapi kenaikan harga jadi dulu tangisan kalian itu benar peduli rakyat atau cuma sekedar politisi

c. *Remove Duplicate*

Tentunya dari data twitter pastinya ada para pengguna melakukan *spam* dengan maksud mengupload *tweet* opini/tanggapan secara berulang-ulang agar *tweet* tersebut dapat memperoleh perhatian. Maka dari itu, pada tahap *remove duplicate* dilakukan untuk penghapusan data berulang dan menghapuskan data *tweet* yang berisi opini yang sama. Berikut contoh *tweet* yang berisi opini yang sama yang ada pada gambar 3.4

InfoTimlo	Negatif	demo tolak kenaikan harga bbm di kota solo terus berlanjut pks pemerintah harus kreatif cari solusi
kabartiml	Negatif	demo tolak kenaikan harga bbm di kota solo terus berlanjut pks pemerintah harus kreatif cari solusi
timlovers	Negatif	demo tolak kenaikan harga bbm di kota solo terus berlanjut pks pemerintah harus kreatif cari solusi

Gambar 3. 4 Tweet dengan opini yang sama

d. *Normalization*

Tentunya pada pengguna dari data twitter tidak heran jika para penggunanya menulis kata dengan singkat-singkat, kata yang tidak baku atau yang tidak formal pada *era kidz zaman now* membuat suatu opini dengan menggabungkan dua bahasa yaitu indo-inggris sekaligus dalam satu kalimat. Nah, pada tahap *normalization* dilakukan untuk pengubahan kata yang tidak sesuai dengan KBBI. Contoh penerapan pada tahap *normalization* dapat dilihat pada tabel 3.6.

Tabel 3. 6 Penerapan Pada Tahap *Normalization*

<i>Input Process</i>	<i>Output Process</i>
dulu ada yg sampai menangis karena harga bbm naik sekarang minta rakyat jgn cengeng menghadapi kenaikan harga jadi dulu tangisan kalian itu benar peduli rakyat atau cuma sekedar politisi	dulu ada yang sampai menangis karena harga bbm naik sekarang minta rakyat jangan cengeng menghadapi kenaikan harga jadi dulu tangisan kalian itu benar peduli rakyat atau cuma sekedar politisi

e. *Stopword Removal*

Pada tahap *stopword removal* dilakukan untuk menghilangkan atau menghapus daftar kata umum yang tidak memiliki arti penting atau informasi yang dibutuhkan. Yang bertujuan untuk mengurangi jumlah kata yang disimpan oleh sistem. Contoh proses tahap *stopword removal* yang terdapat pada Tabel 3.7 sebagai berikut.

Tabel 3. 7 Penerapan Pada Tahap *Stopword Removal*

<i>Input Process</i>	<i>Output Process</i>
dulu ada yang sampai menangis karena harga BBM naik sekarang minta rakyat jangan cengeng menghadapi kenaikan harga jadi dulu tangisan kalian itu benar peduli rakyat atau cuma sekedar politisi	dulu menangis harga BBM naik sekarang minta rakyat jangan cengeng menghadapi kenaikan harga dulu tangisan benar peduli rakyat politisi

f. *Stemming*

Pada tahap *stemming* untuk mengganti suatu kata yang memiliki imbuhan menjadi bentuk kata dasar, dan juga mengubah kata perulangan kata menjadi satu kata dasar. Berikut kaidah yang telah dibuat dalam proses *stemming* pada bahasa Indonesia (Luuqyana, 2018), antara lain:

- 1) Prefiks, imbuhan yang terletak pada awal kata.
Prefiks terdiri dari “se-“, ”ke”, ”me-“ dll. Contoh:
Me-rasa
- 2) Suffiks, imbuhan yang terletak pada akhir kata.
Contoh dari suffiks adalah “-lah”, “-kah”, “-pun”
dll. Contoh: turun-kan.
- 3) Konfiks, imbuhan ini merupakan imbuhan
gabungan dari prefiks dan suffiks. Imbuhan
terdapat pada awal dan akhir kata. Contoh: per-
rasa-an.
- 4) Infiks, imbuhan yang terletak pada tengah kata.
Contoh: k-em-ilau dari kata kilau.
- 5) Perulangan kata, contoh: anak-anak.

Contoh penerapan pada tahap stemming dapat dilihat pada tabel 3.8.

Tabel 3. 8 Penerapan Pada Tahap *Stemming*

<i>Input Process</i>	<i>Output Process</i>
dulu menangis harga bbm naik sekarang minta rakyat jangan cengeng menghadapi kenaikan harga dulu tangisan benar peduli rakyat politisi	dulu nangis harga bbm naik sekarang minta rakyat jangan cengeng hadapi naik harga dulu tangis benar peduli rakyat politisi

g. Tokenization

Tokenization adalah tahap memecah kalimat menjadi bagian-bagian kata yang dinamakan token (Imron, 2019). Pada proses ini akan menghilangkan *whitespace*. Contoh proses tahap *tokenization* yang terdapat pada Tabel 3.9 sebagai berikut.

Tabel 3.9 Penerapan Pada Tahap *Tokenization*

<i>Input Process</i>	<i>Output Process</i>
dulu nangis harga bbm naik sekarang minta rakyat jangan cengeng hadapi naik harga dulu tangis benar peduli rakyat politisi	['dulu','nangis','harga','bbm','naik', 'sekarang','minta','rakyat','jangan', 'cengeng','hadapi','tangis','benar', 'peduli','politisi']

3. Ekstrasi Fitur

Setelah semua data sudah melewati tahap *preprocessing*, maka langkah selanjutnya adalah pembuatan fitur untuk mempermudah proses klasifikasi. Pada tahap pembuatan fitur pembobotan kata dilakukan dengan cara TFIDF (*Term Frequency Invers Document Frequency*). Pada proses TFIDF memberikan pembobotan terhadap setiap kata pada

setiap kalimat atau dokumen dan TFIDF menggunakan rumus yang bisa dilihat pada bab sebelumnya, maka dataset sudah siap untuk digunakan dalam *training* menggunakan perhitungan pada *naive bayes*.

4. Klasifikasi *Naive Bayes*

Dataset yang telah melewati *preprocessing* dan ekstraksi fitur, selanjutnya akan dilanjutkan dengan proses pembelajaran (*learning*) menggunakan metode klasifikasi *naive bayes*. Metode *naive bayes* merupakan metode yang digunakan untuk mengklasifikasikan data komentar untuk memperoleh sentimen analisis (Gunawan et al., 2018). Pengklasifikasian dibagi menjadi tiga *class* yaitu positif, negatif dan netral. Untuk melakukan klasifikasi sentimen akan menggunakan data hasil proses *preprocessing* sampai dengan pembobotan kata dengan TFIDF. Setelah data berhasil ditraining kemudian akan dilakukan pengujian menggunakan data *test* untuk menguji hasil ketepatan klasifikasi yang dilakukan. Kelas yang memiliki skor terbanyak atau kemunculan terbanyak pada setiap *tweet* maka *tweet* tersebut dianggap menjadi kelas tersebut (Fajar et al., 2018).

5. Uji Model

Uji model dilakukan terlebih dahulu dengan membagi data antara data uji dan data latih yaitu dengan cara *split validation*. Cara *split validation* dilakukan dengan jumlah data yang dijadikan untuk diuji diambil dari data *training* sebesar 20%(Turmudi Zy et al., 2021). Pengambilan data dilakukan secara acak dengan bantuan *library* dari *python*. Proses uji model dilakukan setelah proses *training* data dilakukan. Uji model sendiri dilakukan untuk mengetahui kinerja model. Setelah uji model dilakukan maka akan tampil seberapa besar performa metode yang dilakukan.

6. Evaluasi Model

Evaluasi model dilakukan untuk mengetahui atau mengukur kinerja suatu model (Imron, 2019). Evaluasi model dilakukan dengan cara melihat tingkat performa metode melalui *multiclass confusion matrix*. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Turmudi Zy et al., 2021). Lebih tepatnya, *data test* diujikan terhadap *data training*, maka akan menghasilkan daftar kelas-kelas dari data *test*, sebut saja prediksi kelas. Kemudian prediksi

kelas dibandingkan dengan kelas yang sebenarnya dari data test yang disembunyikan sebelumnya. Sehingga dapat dilihat pada performa model *naive bayes* yang berupa tingkat akurasi, *precision*, *recall* dan *f1 score* .

7. Visualisasi

Tahap akhir yaitu memvisualisasikan hasil dari analisis sentimen menggunakan diagram lingkaran dan *wordcloud*. *Wordcloud* akan digunakan untuk memvisualisasikan hasil analisis klasifikasi. Tujuan dari visualisasi tersebut adalah untuk mengekstrak informasi berupa topik yang sering dibahas yaitu kebijakan kenaikan harga BBM oleh pengguna twitter sehingga dapat diambil informasi yang dianggap penting dari sekian banyak teks tanggapan yang ada.

BAB IV

HASIL DAN PEMBAHASAN

A. Pengambilan Data Twitter

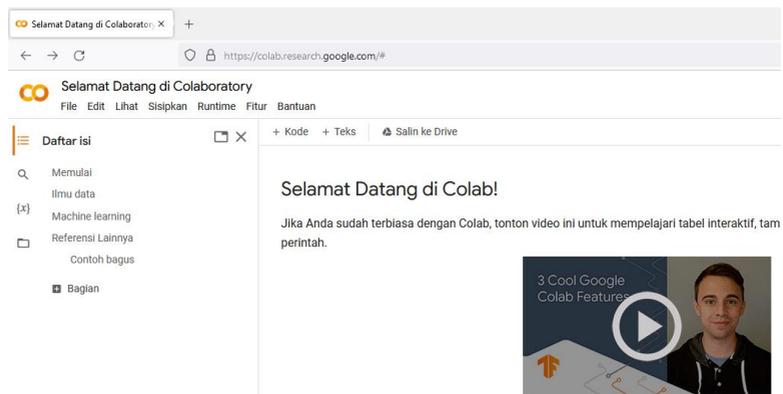
Pengambilan data atau proses *crawling data* twitter ini dilakukan menggunakan *library* yang dapat digunakan pada *python* yaitu *library snsrape*. *library snsrape* dibuat untuk mempermudah melakukan pengambilan data melalui media sosial salah satunya twitter. Tujuannya untuk mengambil data-data twitter yang ingin dicari. *Library snsrape* ini sangat efektif dikarenakan kita tidak perlu melakukan login pada twitter untuk meminta ke akun develop yaitu *API Keys* atau *API token* pada twitter. Jadi hanya perlu kata kunci pencarian yang ingin dicari pada akun twitter. Maka untuk menggunakan *library snsrape* dilakukan instalasi terlebih dahulu menggunakan perintah *pip*. Berikut cara instalasi *library snsrape* seperti gambar 4.1 di bawah.

```
[ ] pip install snsrape
```

Gambar 4. 1 *Install Library Snsrape*

Selain membutuhkan *library snsrape* untuk dapat melakukan pengambilan data twitter disini peneliti

menggunakan *tools* pendukung untuk menganalisis sentimen seperti *google colab* sebagai pendistribusi *python*. Keuntungan dari *google colab* atau *colaboratory* ini sangat efektif karena tidak perlu melakukan instalasi sehingga tidak memakan memori, hanya perlu jaringan internet dan satu akun google untuk mengkases berguna untuk menyimpan hasil codingannya. Tidak hanya itu saja, *google colab* ini digunakan untuk menyimpan, menulis, serta dapat membagikan program yang telah ditulis melalui *google drive* (Carneiro et al., 2018). *Software* ini pada dasarnya serupa dengan *jupyter notebook* gratis berbentuk *cloud* yang dijalankan menggunakan *browser*. Berikut tampilan dari *google colab* pada Gambar 4.2 di bawah.



Gambar 4. 2 Tampilan Google Colab

Dengan adanya layanan *google colab* kini peneliti dapat melakukan proses pengkodean menggunakan bahasa pemrograman *python*. Untuk melakukan proses *crawling data* maka harus disiapkan atau meng-*import* data yang diperlukan seperti Gambar 4.3 di bawah.

```
import snsctrpe.modules.twitter as sntwitter
import pandas as pd
import csv
```

Gambar 4. 3 *Source Code* Pemanggilan *Library* Proses *Crawling*

Pada Gambar 4.3 terdapat proses pendeklarasian *library* *snsctrpe*, *library csv*, dan *library pandas*. *Library snsctrpe* adalah suatu *library* dengan mengambil data-data yang ada di dalam twitter menggunakan bahasa pemrograman *python*. *Library Csv (Command Separated Values)* adalah *library phyton* yang menyediakan layanan baca dan menulis suatu data bertipe *file csv* atau excel. *Library pandas* adalah *library* pada *phyton* yang berguna untuk pengolahan data. Pada *library csv* dan *pandas* sudah disediakan di dalam bawaan *google colab*, jadi tidak perlu dilakukan instalasi lagi. Setelah memasukkan kode tersebut, selanjutnya memasukkan kode proses *crawling data*. Berikut pada gambar 4.4 *source code* penerapan proses *crawling data*.

```
[4] query="Harga BBM (BBM) lang:id until:2022-09-10 since:2022-09-03"
tweets=[]
limit=1500
%time
try:
    print("Star crawling")
    for tweet in sntwitter.TwitterSearchScrapper(query=query).get_items():
        if len(tweets)==limit:
            break
        else:
            tweets.append([tweet.username, tweet.content])
            df=pd.DataFrame(tweets,columns=['username','content'])

except Exception as e:
    print(e)

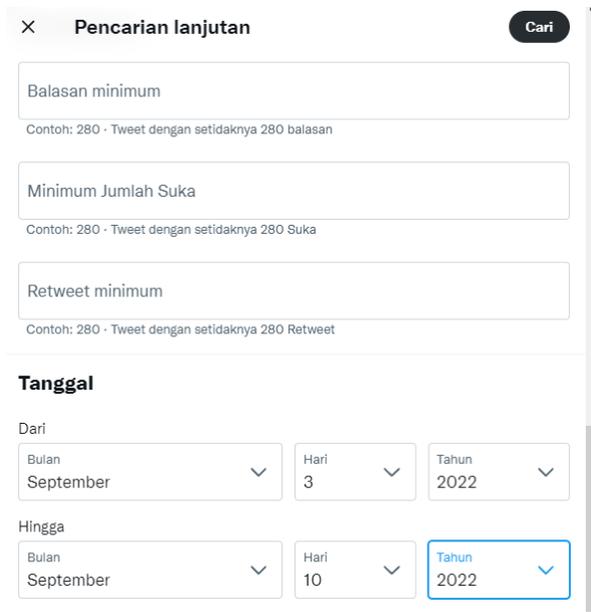
print("Finished")
print("-----")
```

Gambar 4. 4 Source Code Proses Crawling

Untuk proses kata kunci pencarian atau *query* menggunakan bantuan twitter *advanced search*. Twitter *advanced search* merupakan sebuah fitur pada twitter untuk mencari *tweet* menggunakan parameter tertentu, terdapat kata-kata yang ingin dicari, tanggal yang ingin dicari serta bahasa yang ingin dicari juga (Matt Holand, 2017). Seperti pada gambar 4.4 diatas, pada penelitian ini menggunakan *query* "harga BBM (BBM) lang: id until: 2022-09-10 since:2022-09-03".

Pada saat proses pengambilan data twitter, peneliti mengambil kata kunci yang pada saat proses *crawling data* berada pada posisi *trending topic* twitter atau pada posisi pembahasan terbanyak pada *tweet* yaitu "harga BBM" pada

masa pemerintahan presiden Joko Widodo dan KH. Maruf Amin yang berjumlah 1500 *tweet*. Data yang diambil menggunakan data yang berbahasa Indonesia dan diambil dari tanggal 3 September 2022 – 10 September 2022. Berikut tampilan twitter *advanced search* pada gambar 4.5.



The image shows the Twitter Advanced Search interface. At the top, there is a search bar with the text "Pencarian lanjutan" and a "Cari" button. Below the search bar, there are three input fields for filters:

- Balasan minimum**: Input field with a placeholder "Balasan minimum" and a hint "Contoh: 280 · Tweet dengan setidaknya 280 balasan".
- Minimum Jumlah Suka**: Input field with a placeholder "Minimum Jumlah Suka" and a hint "Contoh: 280 · Tweet dengan setidaknya 280 Suka".
- Retweet minimum**: Input field with a placeholder "Retweet minimum" and a hint "Contoh: 280 · Tweet dengan setidaknya 280 Retweet".

Below the filters, there is a section titled "Tanggal" (Date) with two rows of date selection:

- Dari** (From): Three dropdown menus for "Bulan" (Month) set to "September", "Hari" (Day) set to "3", and "Tahun" (Year) set to "2022".
- Hingga** (Until): Three dropdown menus for "Bulan" (Month) set to "September", "Hari" (Day) set to "10", and "Tahun" (Year) set to "2022".

Gambar 4. 5 Tampilan Twitter *Advanced Search*

Setelah dilakukan proses *crawling data* yang ada pada proses gambar 4.4 selanjutnya data disimpan dalam bentuk file

csv. Berikut cara menyimpan hasil pengambilan data twitter pada gambar 4.6.

```
[5] df.to_csv("Data_Kotor_HargaBBM.csv",index=False)
```

Gambar 4. 6 *Source Code* menyimpan data dalam bentuk CSV

Setelah menjalankan kode pada gambar 4.6 di atas maka didapatkan file csv yang belum diolah. Berikut pada gambar 4.7 hasil dari proses *crawling* yang belum diolah pada file csv.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
1		username	content																			
2		cliffordki	"sana mendengar dan merasa jertanmu Tapi nga pake Nangsi,Seluruh Kader PDemokrat																			
3			Menolak kenaikan Harga BBM,																			
4																						
5		@PDemokratJat																				
6		@AgusYudiHoyono																				
7		@BAMujiyono																				
8																						
9		#K1TahunPartaiDemokrat																				
10		#PerjuangkanHarapanRakyat	https://t.co/KYgfjbnHy																			
11		Darwan4537622	@OposisiCerdas Hanya Peduli: Proyek Kereta Cepat dan IKN.																			
12			Memastikan amat kejan di tengah kondisi masyarakat yang berada di bawah himpitan ekonomi yg sulit dgn teganya justru menaikkan harga BBM,K kata Achmad Nur Padahal saat ini harga minyak dunia sedang turun																			
13			https://t.co/ta0Dynaaw																			
14		Thavit_Amoura	"BEM DAN GERAM GELAR DEMO PENOLAKAN KENAIKAN HARGA BBM, INI TUNTUTAN YANG DISAMPAIKAN.																			
15																						
16		#BangkitAtauTertindas																				
17		#BangkitAtauTertindas																				
18																						
19		https://t.co/gW4R0eHt4																				
20		#Fp	"Kedua ini tentu menguntungkan korporat dan merugikan rakyat, sebab harus membeli BBM dengan harga yang relatif mahal."																			
21		dukusterr	"jalan kotanya ditata rapi, tarif angkutan umum massanya ga menyulitkan warga yg tinggal di DKI semuanya serba dimudahkan, meskipun presiden yg gw ga tau siapa itulah namanya gw sudah muak k																			
22		#Fp	"Karena itu, diduga kuat pencabutan subsidi BBM dan peltangan BBM berakibat rendah adalah untuk melepas harga BBM ke pasar, sehingga produk BBM korporasi asing dapat bersaing dengan Pertamina @pe																			
23		#Fp	"Dengan kondisi masyarakat yang sebagian besar bertaraf hidup menengah ke bawah, BBM dengan harga murah masih menjadi pilihan."																			
24		#Fp	"Api yakni Pertamina untuk melepas harga BBM ke pasar."																			

Gambar 4. 7 Hasil *Crawling Data Tweet* Pada File csv

Setelah proses *crawling data* selesai dan berhasil disimpan dalam bentuk file dokumen csv pada excel, selanjutnya masuk pada proses pelabelan data *tweet*. Dalam proses pelabelan data *tweet* ini, peneliti menentukan nilai dari *tweet* pada twitter yang bernilai positif dalam artian setuju

atau pro dengan kebijakan kenaikan harga BBM, pada bernilai negatif yaitu yang kontra atau menolak kebijakan tersebut sedangkan netral yaitu ditengah-tengah dalam artian tidak pro ataupun kontra biasanya memuat informasi. Dalam proses pelabelan data ini seharusnya untuk menentukan nilai tersebut dilakukan oleh pakar bahasa atau yang terkait dibidangnya dan juga setidaknya membutuhkan dua orang atau lebih untuk menghindari perbedaan pendapat yang sama serta menghindari subjektivitas dalam menentukan sentimen *tweet* yang terdapat dari hasil data *crawling* (Imron, 2019) .

Namun demikian, disini pada saat proses pelabelan data peneliti hanya dilakukan dengan dua orang yaitu mahasiswa lainnya yang bukan pakar dari bahasa ataupun psikologi dibidangnya. Konsekuensi dari penelitian dalam pelabelan data ini adalah terdapatnya beberapa *tweet* yang tidak sama ataupun tidak sesuai dengan sentimen analisis sebenarnya. Karena proses pelabelan data twitter ini dilakukan secara manual, dan juga dalam proses pelabelan secara manual ini membutuhkan waktu yang tidak cepat, hal tersebut menjadi kelemahan pada saat jumlah data yang akan di proses pelabelan sangat banyak. Berikut pada gambar 4.8 menunjukkan hasil data yang sudah proses *labelling*.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2		Lisuanamandjantiment content																			
3		GiffFlartid Negatif	kami mendengar dan merasa jertanmu lagi raga pale nangis seluruh kader pedemokrat menolak kenaikan harga bbm																		
4		Darwardi Negatif	hanya peduli proyek kereta cepat dan ikn pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah hingintan ekonomi yg sulit dgn teganya justru menaikkan harga bbm																		
5		Irfqn_ Negatif	kejajikan ini tentu menguntungkan korporat dan mengigit rakyat sebab harus membeli bbm dengan harga yang relatif mahal																		
6		duckabuste Netral	jajanannya kotanya ditata rapi tarif angkutan umum massalnya ga menyulitkan warga yg tinggal di di sisi semuanya serba dimudahkan meskipun presiden yg gva tau siapa itulah namanya ger sudah																		
7		Irfqn_ Positif	karena itu diduga kuat pencabutan subsidi bbm dan palarangan bbm berikatan rendah adalah untuk melepas harga bbm ke pasar sehingga produk bbm korporasi asing dapat bersaing dengan pe																		
8		Irfqn_ Positif	dengan kondisi masyarakat yang sebagian besar bertaraf hidup menengah ke bawah bbm dengan harga murah masih menjadi pilihan																		
9		Irfqn_ Negatif	apalagi sejak lama diketahui bahwa ada upaya pemerintah yakni pertama untuk melepas harga bbm ke pasar																		
10		Darwardi Negatif	hanya peduli proyek kereta cepat dan ikn pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah hingintan ekonomi yg sulit dgn teganya justru menaikkan harga bbm																		
11		ZattazAlri Netral	basi sus kas gimna bbm yg naik di tengah harga minyak dunia yang turun gimna kinc yg super lokal dan presiden yg bianya cuma naik naikin harga pajak kedudukan kulan pamerkan kama mer																		
12		aki_slawi Negatif	sebelum pertamina menaikkan harga bbm sudah memurunkan harga revivo pertamina menaikkan harga bbm konsumen beralih ke vivo revivo dirjen esdm meminta vivo meny																		
13		Sinyopi Negatif	strugling dgn harga bbm trs nepugibildi data my pertamina tuh dr mana hubungannya yg ada org mah pusing udh hrg pd naik kudu rajin ganti pass trs ngingit td apaan ya kwkw lagan pejabat																		
14		oktavinia_ Negatif	tolak kenaikan harga bbm serikat buruh akan gelar aksi sebulan penuh																		
15		purnomono Negatif	tolak kenaikan harga bbm aliansi buruh beksis melawan rencanakan unjuk rasa september																		
16		sukaridi_ri Negatif	tolak kenaikan harga bbm aliansi buruh beksis melawan rencanakan unjuk rasa september																		
17		Rungu_Oi Negatif	tuntutan mahasiswa dalam aksi penolakan kenaikan harga bbm dibantem																		
18		Tribun_Rc Negatif	berlaku mulai hari ini tarif opol resmi naik skuti harga bbm catat ini perubahan harganya levat																		
19		Nettsambli Negatif	massa aksi jebot pager geruduk dipid indahnya tolak kenaikan harga bbm sepi																		
20		S_Ahmadl Negatif	mau pilih yg mana mari kita bangkit bersama tolak kenaikan harga bbm																		
21		Memosis Negatif	mana parpol lain yg pro rakyat tolak kenaikan harga bbm bersubsidi																		
22		infoboks_ Netral	pengelola jasa ekspedisi di makassar sesuaikan tarif setelah harga bbm naik																		
23		Saufadder Netral	pemerintah malaysia menghitng dg patokan dolar amp harga minyak dunia yg sama dpt harga bbm ron kualitas pertamax plus tanpa subsidi sebesar rp knp simple si jawabnya petromas																		
24		Thaviri_An Negatif	penolakan kenaikan harga bbm berlanjut, massa mahasiswa unjuk rasa di monas																		

Gambar 4. 8 Hasil Data yang sudah diberi Label

B. *Preprocessing*

Tahapan ini terdiri dari beberapa proses karena data komentar memiliki karakteristik yang tidak terstruktur yang sangat memuat *noise*. Maka, pada tahapan ini bertujuan untuk mengubah suatu data yang masih mentah/kotor lalu diolah menjadi data bersih sehingga dapat dilakukan pengklasifikasian.

Penerapan tahap *preprocessing* data pada penelitian ini dilakukan dengan melakukan 7 proses secara urut, di antaranya:

1. *Case Folding*

Proses *case folding* bertujuan untuk menyemarakatkan suatu teks yang terdapat huruf kapital diubah menjadi huruf kecil. Adapun kode program yang memperlihatkan implementasi dari *case folding* disatukan dalam tahapan proses *cleansing*.

2. *Cleansing*

Proses *cleansing* pada tahap ini bertujuan untuk membersihkan data komentar dari hal yang tidak diperlukan seperti tanda baca, *hashtag*, *username* dan sebagainya. Dalam melakukan proses *cleansing* tersebut dilakukan 9 tahapan untuk mendapatkan hasil yang maksimal, di antaranya ialah:

- a. Menghapus kata RT (*Retweet*)
- b. Menghapus tanda “@nama” pada *username*
- c. Mengubah emoji menjadi teks
- d. Menghapus URL (<http://> atau <https://>)
- e. Menghapus kata *hashtag* “#*hashtag*”
- f. Menyeragamkan huruf menjadi huruf kecil semua (*Case Folding*)
- g. Menghapus kata yang berulang menjadi tunggal (‘oooooo’ menjadi ‘o’)
- h. Mengubah tanda baca yang berulang menjadi tunggal (‘!!!!!!’ menjadi ‘!’)
- i. Menghapus angka serta tanda baca

Pada proses *cleansing library* yang harus disiapkan yaitu *library* re dan emoji. *Library* re yang bertujuan untuk melakukan tahapan *Regular Expression* (regex) atau deretan karakter yang digunakan untuk

pencarian teks dengan menggunakan pola (*pattern*). *Library* emoji yang bertujuan untuk mengubah emoji menjadi suatu teks. Maka dari itu, dilakukan terlebih dahulu instalasi *library* emoji menggunakan perintah `pip` sedangkan pada *library* re sudah disiapkan pada google *colab*. Berikut pada gambar 4.9 *source code install library* emoji.

```
[ ] pip install emoji
```

Gambar 4. 9 *Install Library* Emoji

Kemudian *import library* emoji dan re untuk melakukan proses *cleansing*. Berikut pada gambar 4.10 *source code library* yang dibutuhkan pada proses *cleansing*.

```
import emoji  
import re
```

Gambar 4. 10 *SC* Pemanggilan *Library emoji* dan *re*

Adapun kode program yang memperlihatkan implementasi dari tahap *cleansing* data ditunjukkan oleh gambar 4.11.

```

def cleansing(tweet):
    # Menghapus kata RT
    t1 = re.sub('RT\s+', '', tweet)
    # Menghapus tanda @_username
    t2 = re.sub('\B@\w+', '', t1)
    # Mengubah emoji menjadi text
    t3 = emoji.demojize(t2)
    # Menghapus URL (http:// or https://)
    t4 = re.sub('(http|https):\/\/\S+', '', t3)
    # Menghapus kata Hastag "#hashtag"
    t5 = re.sub('#[A-Za-z0-9]+', '', t4)
    # Mengubah menjadi huruf kecil
    t6 = t5.lower()
    # Mengubah huruf pada kata yang berulang menjadi tunggal ('oooooooo' menjadi 'oo')
    t7 = re.sub(r'(\.)\1+', r'\1\1', t6)
    # Mengubah tanda baca yang berulang menjadi tunggal ('!!!!!!!' menjadi '!')
    t8 = re.sub(r'[?!\.!\!]+(?=[\?!\!])', '', t7)
    # Menghapus angka dan tanda baca
    t9 = re.sub(r'[a-zA-Z]', ' ', t8)
    return t9

for i, r in df.iterrows():
    y=cleansing(r['content'])
    df.loc[i,'content']=y

df.head()

```

Gambar 4. 11 Source Code Tahapan *Cleansing* dan *Case Folding*

Berikut hasil dari penerapan hasil dari *cleansing* dan *case folding* yang ditunjukkan pada gambar 4.12 dibawah ini.

	username	Sentiment	content
0	GhiffariDKI	Negatif	kami mendengar dan merasa jeritanmu tapi nga pe...
1	Darwan45537622	Negatif	hanya peduli proyek kereta cepat dan ikn pe...
2	Thavit_Amoura	Negatif	bem dan geram gelar demo penolakan kenaikan ha...
3	Irfqn_	Negatif	kebijakan ini tentu menguntungkan korporat dan...
4	duckbusterr	Netral	jalanan kotanya ditata rapi tarif angkutan um...

Gambar 4. 12 Hasil Penerapan tahap *Cleansing* dan *Case Folding*

3. *Removing Duplicate*

Pada proses ini tujuannya untuk menghapus tanggapan/opini yang memiliki kesamaan persis atau terduplikat sehingga yang diperlukan satu tanggapan saja. Tentunya sangat mempengaruhi pada jumlah analisis sentimen karena data yang digunakan sama atau terduplikasi.

Berikut kode program yang memperlihatkan implementasi dari menghapus baris yang terduplikasi ditunjukkan oleh gambar 4.13.

```
df = df.drop_duplicates()
df = df.reset_index(drop=True)
df
```

Gambar 4. 13 *Source Code Remove Dupilcate*

Berikut hasil dari penerapan dari tahapan *remove duplicate* menghasilkan data menjadi terurai dari 1500 data menjadi 1483 data yang dapat ditunjukkan pada gambar 4.14

	username	Sentiment	content
0	GhiffariDKI	Negatif	kami mendengar dan merasa jeritanmu tapi nga p...
1	Darvan45537622	Negatif	hanya peduli proyek kereta cepat dan lkn pe...
2	Thavit_Amoura	Negatif	bem dan geram gelar demo penolakan kenaikan ha...
3	Irfqn_	Negatif	kebijakan ini tentu menguntungkan korporat dan...
4	duckbusterr	Netral	jalanannya ditata rapi tarif angkutan um...
...
1478	LantasPBM	Positif	kasat lantas potres prabumulih apk muthemainah...
1479	HajarAs32215348	Negatif	partai demokrat instruksikan kadernya turun ke...
1480	KangMasroer	Positif	karor gembira menteri esdm ada potensi harga...
1481	LantasPBM	Positif	kasat lantas potres prabumulih apk muthemainah...
1482	detik_jabar	Negatif	terima audiensi driver ojol pemkab garut tola...

1483 rows x 3 columns

Gambar 4. 14 Hasil Dari Tahapan *remove duplicate*

4. Normalization

Pada proses ini tujuannya mengubah kata tidak baku atau kata-kata gaul menjadi baku.

Pada tahapan ini harus dipersiapkan *import* data *pandas*, *re*, *csv* serta *string*. Berikut *source code* pada gambar 4.15 yang diperlukan pada tahapan *normalization*.

```
import pandas as pd
import re
import csv
import string
```

Gambar 4. 15 *Import library* pada tahap *normalization*

Tentunya pada tahapan *normalization* di *python* tidak menyediakan pengubahan kata tidak baku menjadi

baku dalam berbahasa Indonesia, maka dari itu peneliti membuat serta melakukan kumpulan kata-kata yang tidak formal menjadi formal dalam bentuk teks dan excel. Jadi, diperlukan beberapa data pendukung pada tahapan *normalization* ini. Berikut pada gambar 4.16 untuk *import* data pendukung.

```

slang_dictionary = pd.read_csv('colloquial-indonesian-lexicon.csv')
slang_dict = pd.Series(slang_dictionary['formal'].values, index=slang_dictionary['slang']).to_dict()

slang_dictionary1 = pd.read_csv('kbba.txt', sep='\t')
slang_dict1 = pd.Series(slang_dictionary1['tujuan'].values, index=slang_dictionary1['7an']).to_dict()

slang_dictionary2 = pd.read_csv('slangword.txt', sep=':')
slang_dict2 = pd.Series(slang_dictionary2['dan'].values, index=slang_dictionary2['&']).to_dict()

slang_dictionary3 = pd.read_csv('formalizationDict.txt', sep='\t')
slang_dict3 = pd.Series(slang_dictionary3['tujuan'].values, index=slang_dictionary3['7an']).to_dict()

```

Gambar 4. 16 *Source Code Import* Data Pendukung

Peneliti melakukan sebanyak empat *file* data pendukung pada tahapan perubahan bentuk kata yang tidak baku menjadi baku yaitu pada *file* pertama bernama "*colloquial-indonesian-lexicon.csv*", *file* kedua "*kbba.txt*", *file* ketiga "*slangword.txt*" dan *file* keempat "*formalizationDict.txt*". Berikut hasil dari *source code* data pendukung *file* kesatu dan kedua yang terdapat pada gambar 4.17.

```
[ ] slang_dictionary.head()
```

	slang	formal	In-dictionary	context	category1	category2	category3
0	woww	wow	1		wow	elongasi	0
1	aminn	amin	1	Selamat ulang tahun kakak tulus semoga panjang...	elongasi		0
2	met	selamat	1	Met hari netaas kak? Wish you all the best @t...	abreviasi		0
3	netaas	menetas	1	Met hari netaas kak? Wish you all the best @t...	afiksasi	elongasi	0
4	keberpa	keberapa	0	Birthday yg keberpa kak?	abreviasi		0


```
slang_dictionary1.head()
```

	7an	tujuan
0	@	di
1	ababil	abg labil
2	abis	habis
3	acc	accord
4	ad	ada

Gambar 4. 17 Data pendukung kesatu dan kedua

Adapun ada data pendukung yang ketiga dan keempat berikut hasil *source code* yang ditujukan pada gambar 4.18.

```
[ ] slang_dictionary2.head()
```

	&	dan
0	+	tambah
1	/	atau
2	=	sama dengan
3	ababil	anak ingusan
4	abal2	palsu


```
slang_dictionary3.head()
```

	7an	tujuan
0	@	di
1	ababil	abg labil
2	abis	habis
3	acc	accord
4	ad	ada

Gambar 4. 18 Data Pendukung ketiga dan keempat

Tentunya yang ditampilkan hanya lima data teratas saja dalam pemrograman sebenarnya banyak sekali isi *file* data pendukung tersebut. Setelah dilakukan *peng-importan file* dokumen, kemudian dilakukan proses pengubahan kata menjadi kata baku pada keempat *file* tersebut dengan menggunakan proses *codingan* dengan fungsi *def*. Berikut *source code* pada tahap *normalization part 1* yang ditunjukkan pada gambar 4.19.

```
def Slangwords(text):
    for word in text.split():
        if word in slang_dict.keys():
            text = text.replace(word, slang_dict[word])
            text = re.sub('@[\w]+', '', text)
    return text

def Slangwords1(text):
    for word in text.split():
        if word in slang_dict1.keys():
            text = text.replace(word, slang_dict1[word])
            text = re.sub('@[\w]+', '', text)
    return text

def Slangwords2(text):
    for word in text.split():
        if word in slang_dict2.keys():
            text = text.replace(word, slang_dict2[word])
            text = re.sub('@[\w]+', '', text)
    return text

def Slangwords3(text):
    for word in text.split():
        if word in slang_dict3.keys():
            text = text.replace(word, slang_dict3[word])
            text = re.sub('@[\w]+', '', text)
```

Gambar 4. 19 *Source Code* implementasi tahapan *normalization part 1*

Selanjutnya proses implementasi dengan memanggil keempat file data pendukung tersebut serta peneliti juga melakukan sedikit penambahan untuk perubahan kata *typo* secara manual pada tahapan *normalization*. Berikut kode program yang memperlihatkan implementasi *normalization* part 2 ditunjukkan oleh gambar 4.20.

```
df['Text_Clean'] = df['content'].apply(Slangwords)
df['Text_Clean'] = df['Text_Clean'].apply(Slangwords1)
df['Text_Clean'] = df['Text_Clean'].apply(Slangwords2)
df['Text_Clean'] = df['Text_Clean'].apply(Slangwords3)
df['Text_Clean'] = df['Text_Clean'].str.replace('mendeenggakr', 'mendengar')
df['Text_Clean'] = df['Text_Clean'].str.replace('naikx', 'naiknya')
df['Text_Clean'] = df['Text_Clean'].str.replace('harenggak', 'ngga')
df['Text_Clean'] = df['Text_Clean'].str.replace('nga', 'ngga')
df['Text_Clean'] = df['Text_Clean'].str.replace('naikx', 'naiknya')
df['Text_Clean'] = df['Text_Clean'].str.replace('sleepy', 'tidur')
df['Text_Clean'] = df['Text_Clean'].str.replace('healing', 'libur')
df['Text_Clean'] = df['Text_Clean'].str.replace('face', 'wajah')
df['Text_Clean'] = df['Text_Clean'].str.replace('nggak', 'ngga')
df['Text_Clean'] = df['Text_Clean'].str.replace('mendenggar', 'mendengar')
df['Text_Clean'].head()
```

Gambar 4. 20 SC implementasi tahapan *normalization* part 2

Hasil dari tahapan *normalization* ini dapat ditunjukkan pada gambar 4.21 di bawah ini.

```
0   kami mendengar dan merasa jeritanmu tapi engga...
1   hanya peduli proyek kereta cepat dan ikn pe...
2   bem dan geram gelar demo penolakan kenaikan ha...
3   kebijakan ini tentu menguntungkan korporat dan...
4   jalanan kotanya ditata rapi tarif angkutan um...
Name: Text_Clean, dtype: object
```

Gambar 4. 21 Hasil dari Tahapan *Normalization*

5. *Stopword Removal*

Tahap *stopword removal* merupakan tahapan yang dilakukan untuk menghapus kata-kata yang bersifat umum yang tidak memiliki arti penting atau informasi yang dibutuhkan. Pada proses tahapan *stopword removal* perlu disiapkan *library* yang akan digunakan yaitu *library nlp_id*. Maka untuk menggunakan *library nlp_id* perlu dilakukan instalasi terlebih dahulu dengan cara memakai perintah `!pip` sebagaimana ditunjukkan pada gambar 4.22.

```
!pip install nlp-id
```

Gambar 4. 22 *Install Library nlp-id*

Setelah dilakukan *install nlp_id*, selanjutnya *men-import library nlp_id* dengan memanggil kelas *stopword* seperti yang ada pada Gambar 4.23.

```
from nlp_id.stopword import StopWord
```

Gambar 4. 23 *Pendeklarasian Library nlp_id*

Lalu, memasukkan kode proses *stopword removal* seperti yang ada pada gambar 4.24.

```
stopword = StopWord()

df['Text_Clean'] = df['Text_Clean'].apply(stopword.remove_stopword)
df['Text_Clean'].head()
```

Gambar 4. 24 Source code tahap *stopword removal*

Kemudian hasil dari tahapan proses *stopword removal* dapat ditunjukkan pada gambar 4.25 di bawah ini.

```
0   mendengar jeritanmu engga pakai menangis kader...
1   peduli proyek kereta cepat ikn pemerintah keja...
2   bem geram gelar demo penolakan kenaikan harga ...
3   kebijakan menguntungkan korporat merugikan rak...
4   jalanan kotanya ditata rapi tarif angkutan mas...
Name: Text_Clean, dtype: object
```

Gambar 4. 25 Hasil dari tahapan *Stopword Removal*.

6. *Stemming*

Tahap *stemming* adalah tahap mengubah kata yang memiliki imbuhan menjadi bentuk kata dasar. Pada proses tahapan *stemming* perlu disiapkan *library* yang akan digunakan yaitu *library* sastrawi. Maka untuk menggunakan *library* sastrawi perlu dilakukan instalasi terlebih dahulu dengan cara memakai perintah `!pip` sebagaimana ditunjukkan pada gambar 4.26.

```
!pip install Sastrawi
```

Gambar 4. 26 *Install Library Sastrawi*

Setelah dilakukan *install* sastrawi, selanjutnya men-*import* library sastrawi dengan memanggil kelas *stemmerfactory* seperti yang ada pada Gambar 4.27.

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

Gambar 4. 27 Pendeklarasian *Library* Sastrawi

Kemudian, memasukkan kode proses *stemming* seperti yang ada pada gambar 4.28.

```
# Membuat stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

df['Text_Clean'] = df['Text_Clean'].apply(stemmer.stem)
df['Text_Clean'].head()
```

Gambar 4. 28 *Source code* tahapan *stemming*

Sehingga, hasil dari tahapan *stemming* ini dapat di tunjukkan pada gambar 4.29 di bawah ini.

```
0    dengar jerit engga pakai menang kader pdemokra...
1    peduli proyek kereta cepat ikn perintah kejam ...
2    bem geram gelar demo tolak naik harga bbm tuntutan
3    bijak untung korporat rugi rakyat beli bbm den...
4    jalan kota tata rapi tarif angkut massal engen...
Name: Text_Clean, dtype: object
```

Gambar 4. 29 Hasil dari tahap *Stemming*

7. Tokenization

Tahap *tokenization* merupakan tahap dimana memecahkan kalimat/teks menjadi beberapa bagian-bagian kata. Berikut *source code* yang dari tahap *tokenization* ditunjukkan pada gambar 4.30.

```
def split_word(teks):
    list_teks = []
    for txt in teks.split(" "):
        list_teks.append(txt)
    return list_teks

df['Text_Clean_token'] = df['Text_Clean'].apply(split_word)
df['Text_Clean_token'].head()
```

Gambar 4. 30 Source Code Tahapan Tokenization

Sehingga, hasil dari tahapan *tokenization* dapat ditunjukkan pada gambar 4.31 dibawah ini.

```
0    [dengar, jerit, engga, pakai, menang, kader, p...
1    [peduli, proyek, kereta, cepat, ikn, perintah,...
2    [bem, geram, gelar, demo, tolak, naik, harga, ...
3    [bijak, untung, korporat, rugi, rakyat, beli, ...
4    [jalan, kota, tata, rapi, tarif, angkut, massa...
Name: Text_Clean_token, dtype: object
```

Gambar 4. 31 Hasil dari Tahapan *tokenization*

Setelah melalui ketujuh tahapan *preprocessing*, lalu didapatkan hasil dari data yang sudah siap digunakan. Selanjutnya peneliti ingin menampilkan bentuk hasil yang sudah dilakukan proses *preprocessing* yang ditunjukkan pada gambar 4.32.

	username	Sentiment	content	Text_Clean	Text_Clean_token
0	GhiffarIDKI	Negatif	kami mendengar dan merasa jentanmu tapi nga p...	dengar jerit engga pakai menang kader pdemokra...	[dengar, jerit, engga, pakai, menang, kader, p...
1	Darwan45537622	Negatif	hanya peduli proyek kereta cepat dan ikon pe...	peduli proyek kereta cepat ikon perintah kejam ...	[peduli, proyek, kereta, cepat, ikon, perintah,...
2	Thavit_Amoura	Negatif	berm dan geram gelar demo penotakan kenaikan ha...	berm geram gelar demo tolak naik harga BBM tuntut ...	[berm, geram, gelar, demo, tolak, naik, harga, ...
3	lrfqn_	Negatif	kebijakan ini tentu menguntungkan korporat dan...	bijak untung korporat rugi rakyat beli BBM den...	[bijak, untung, korporat, rugi, rakyat, beli, ...
4	duckbusterr	Netral	jalan kota ternyata ditata rapi tarif angkutan um...	jalan kota tata rapi tarif angkutan massal enggan...	[jalan, kota, tata, rapi, tarif, angkutan, massa...
...
1478	LantasPBM	Positif	kasat lantas Polres prabumulih akip muthemainah...	kasiat lantas Polres prabumulih akip muthemain...	[kasiat, lantas, Polres, prabumulih, akip, mut...
1479	HajarAs32215348	Negatif	partai demokrat instruksikan kadernya turun ke...	partai demokrat instruksi kader turun jalan ak...	[partai, demokrat, instruksi, kader, turun, ja...
1480	KangMasroer	Positif	kabar gembira menteri esdm ada potensi harga...	kabar gembira menteri esdm potensi harga BBM T...	[kabar, gembira, menteri, esdm, potensi, harga...
1481	LantasPBM	Positif	kasat lantas Polres prabumulih akip muthemainah...	kasiat lantas Polres prabumulih akip muthemain...	[kasiat, lantas, Polres, prabumulih, akip, mut...
1482	detik_jabar	Negatif	terima audiensi driver ojol pembak garut tola...	terima audiensi driver ojol pembak garut tolak...	[terima, audiensi, driver, ojol, pembak, garut...

1483 rows x 5 columns

Gambar 4. 32 Hasil Setelah Melalui *Preprocessing*

Sehingga setelah dilakukan tahapan *preprocessing*, data berubah menjadi 1483 data *tweet* karena terdapat komentar yang terduplikat/sama. Tentunya peneliti ingin mengetahui jumlah sentimen tiap data komentar mengenai tanggapan-tanggapan positif, negatif dan netral yang didapatkan setelah melalui *preprocessing*. Namun dalam bahasa pemrograman sentimen yang terdiri dari sentimen positif, negatif dan netral diurutkan sesuai dengan huruf alfabet sehingga sentimen negatif diurutkan menjadi urutan pertama pada bahasa

pemrograman dimulai dari angka nol, sentimen netral diurutkan pada kedua pada pemrograman disebut angka satu, dan sentimen positif diurutkan pada ketiga pada pemrograman disebut angka dua pada *array*, yang terdapat pada gambar di bawah ini. Berikut gambar 4.33 hasil total setiap sentimen pada data *tweet* setelah di *preprocessing*.

```
[35] df['Sentiment'].unique()

array(['Negatif', 'Netral', 'Positif'], dtype=object)

[36] df.Sentiment.value_counts()

Negatif    798
Positif    550
Netral     135
Name: Sentiment, dtype: int64
```

Gambar 4. 33 Hasil Total Sentimen yang diperoleh pada data *tweet*

Dan didapatkan sesuai urutan jumlah terbesar pada sentimen negatif didapatkan sebanyak 798 data *tweet*, pada sentimen positif didapatkan sebanyak 550 data *tweet*, pada sentimen netral didapatkan sebanyak 135 data *tweet*.

C. Ekstrasi Fitur

Sesudah dilakukan *preprocessing*, pada proses ekstrasi fitur perlu disiapkan *library* yang akan digunakan yaitu *library sklearn* atau *scikit-learn*. *Library sklearn* merupakan sebuah modul dari bahasa pemrograman *python* yang dibangun

berdasarkan *numpy*, *sciPy*, dan *matplotlib*. Fungsi dari modul ini adalah untuk membantu melakukan *processing* data ataupun melakukan *training* data untuk kebutuhan *machine learning* atau *data science* (Pedregosa et al., 2012). Pada ekstraksi fitur menggunakan *library sklearn* yang diantaranya adalah pada kelas *train_test_split*, *Label_encoder*, dan *TfidfVectorizer*.

Data selanjutnya dilakukan *split validation data* yaitu pembagian data uji dengan data latih untuk mempermudah proses klasifikasi data. Pada proses pembagian digunakan data uji sebanyak 20% dari data keseluruhan. Proses pembagian data dapat dilihat pada gambar 4.34.

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test = train_test_split(df['Text_Clean'],df['Sentiment'],test_size = 0.2,random_state = 42)
```

Gambar 4. 34 Tahapan *Split Validation Data*

Pada proses ekstraksi fitur, proses awal yang dilakukan oleh sistem adalah mengubah kumpulan data menjadi suatu representasi vektor dengan meng-*import labelencoder* untuk mengubah suatu teks menjadi bilangan angka dengan tujuan agar dapat diproses oleh sistem dalam pengklasifikasian. Lalu, selanjutnya ketahapan kedua yaitu TFIDF untuk pemrosesan ekstraksi fitur dengan tujuan membobot suatu kata. Pada proses

pembuatan *word vector* dan pembobotan kata menggunakan bantuan *library sklearn* yaitu dengan meng-*import* kelas *TfidfVectorizer*, adapun *source code* dari pembobotan TFIDF dapat dilihat pada gambar 4.35.

```
[ ] from sklearn.preprocessing import LabelEncoder
Encoder = LabelEncoder()
y_train = Encoder.fit_transform(y_train)
y_test = Encoder.fit_transform(y_test)

[ ] from sklearn.feature_extraction.text import TfidfVectorizer
Tfidf_vect = TfidfVectorizer(max_features=510)
Tfidf_vect.fit(df['Text_Clean'])
Train_X_Tfidf = Tfidf_vect.transform(X_train)
Test_X_Tfidf = Tfidf_vect.transform(X_test)
```

Gambar 4. 35 Tahapan Proses Pembobotan dengan TFIDF

Sehingga setelah tahap proses pembobotan dengan TFIDF menghasilkan kalimat yang sudah menjadi kumpulan *array* yang menjadi suatu matriks, dimana setiap baris mewakili setiap dokumen, sedangkan pada setiap kolom mewakili seluruh kata yang ada pada seluruh teks. Berikut hasil TFIDF ditunjukkan pada gambar 4.36.

```

Train_X_Tfidf.toarray()
array([[0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.46969016, ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       ...,
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ],
       [0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.         ]])

```

Gambar 4. 36 Hasil Proses TFIDF

Sebagai contoh penelitian menggunakan tiga komentar untuk perhitungan manualisasi untuk pembobotan TFIDF sebagai berikut :

(Doc1)="saya menolak kenaikan harga BBM ini"

(Doc2)=" Setuju harga bbm naik, agar APBN kembali stabil"

(Doc3)="BBM naik, sembako jadi ikutan naik dong"

Setelah sistem melakukan *preprocessing* maka menjadi seperti ini :

(Doc1)=[‘tolak’,’naik’,’harga’,’bbm’]

(Doc2)=[‘setuju’,’harga’,’bbm’,’naik’,’apbn’,’kembali’,’stabil’]

(Doc3)=[‘bbm’,’naik’,’sembako’,’ikut’,’naik’,’dong’]

Pada tahap selanjutnya dilakukan perhitungan dengan metode TFIDF dengan tujuan membentuk *word vector* yang telah diberi bobot nilai. Pada metode TFIDF terdapat dua kata yaitu TF (*Term Frequency*) dan IDF (*Invers Document*

Frequency), dimana TF sendiri merupakan jumlah sebuah kata dari tiap dokumen, sedangkan IDF berfungsi mengurangi bobot suatu kata apabila kemunculannya banyak tersebar diseluruh dokumen. Dimana pada perhitungan pembobotan TFIDF langkah awalnya dilakukan pencarian nilai TF terlebih dahulu. Berikut contoh perhitungan TF secara manualisasi yang terdapat pada tabel 4.1

Tabel 4. 1 Contoh perhitungan TF

Token	Tf		
	D1	D2	D3
tolak	1	0	0
naik	1	1	2
harga	1	1	0
bbm	1	1	1
setuju	0	1	0
apbn	0	1	0
kembali	0	1	0
stabil	0	1	0
sembako	0	0	1
ikut	0	0	1
dong	0	0	1

Setelah didapatkan nilai TF, supaya dapat masuk ke perhitungan IDF maka harus menentukan DF (*Document Frequency*) terlebih dahulu. DF sendiri merupakan banyaknya

jumlah dokumen yang mengandung kata tersebut. Berikut contoh perhitungan DF yang terdapat pada tabel 4.2

Tabel 4. 2 contoh Perhitungan DF

Token	TF			DF
	D1	D2	D3	
tolak	1	0	0	1
naik	1	1	2	3
harga	1	1	0	2
bbm	1	1	1	3
setuju	0	1	0	1
apbn	0	1	0	1
kembali	0	1	0	1
stabil	0	1	0	1
sembako	0	0	1	1
ikut	0	0	1	1
dong	0	0	1	1

Setelah didapatkan nilai DF, selanjutnya masuk ke perhitungan IDF. Diketahui jumlah dokumen (D) yang digunakan sebagai contoh disini yaitu sebanyak tiga komentar, sehingga dapat diketahui dokumen menjadi $D=3$. Dimana perhitungan IDF dan rumusnya dapat dilihat pada tabel 4.3 sebagai berikut.

Tabel 4. 3 contoh perhitungan IDF

Token	Df	D/Df (D=3)	IDF (Log (D/Df))	IDF + 1
tolak	1	3	Log 3= 0,477	1,477
naik	3	1	Log 1= 0	1
harga	2	1,5	Log 1,5=0,176	1,176
bbm	3	1	Log 1= 0	1
setuju	1	3	Log 3= 0,477	1,477
apbn	1	3	Log 3= 0,477	1,477
kembali	1	3	Log 3= 0,477	1,477
stabil	1	3	Log 3= 0,477	1,477
sembako	1	3	Log 3= 0,477	1,477
ikut	1	3	Log 3= 0,477	1,477
dong	1	3	Log 3= 0,477	1,477

Setelah nilai TF dan IDF didapatkan, langkah selanjutnya menghitung TFIDF dengan mengkalikan hasil nilai TF dengan nilai IDF. Berikut pada tabel 4.4 ditunjukkan perhitungan TFIDF.

Tabel 4. 4 Hasil Perhitungan TFIDF

Token	Tf			IDF (LogD/Df)	IDF+1	TF*IDF		
	D1	D2	D3			D1	D2	D3
tolak	1	0	0	0,477	1,477	1,477	0	0
naik	1	1	2	0	1	1	1	2
harga	1	1	0	0,176	1,176	1,176	1,176	0
bbm	1	1	1	0	1	1	1	1
setuju	0	1	0	0,477	1,477	0	1,477	0
apbn	0	1	0	0,477	1,477	0	1,477	0
kembali	0	1	0	0,477	1,477	0	1,477	0
stabil	0	1	0	0,477	1,477	0	1,477	0
sembako	0	0	1	0,477	1,477	0	0	1,477
ikut	0	0	1	0,477	1,477	0	0	1,477
dong	0	0	1	0,477	1,477	0	0	1,477

Sehingga apabila hasil TFIDF dituliskan dalam *array* menjadi seperti dibawah ini :

```
Array([
[1.477, 1, 1.176, 1, 0, 0, 0, 0, 0, 0, 0],
[0, 1, 1.176, 1, 1.477, 1.477, 1.477, 1.477, 0, 0, 0],
[0, 2, 0, 1, 0, 0, 0, 0, 1.47, 1.477, 1.477]
])
```

Dimana setiap barisnya menginterpretasikan tiap dokumen, dan tiap kolomnya menginterpretasikan tiap kata pada seluruh teks.

D. Klasifikasi *Naive Bayes*

Dataset yang telah melewati *preprocessing* dan ekstrasi fitur, selanjutnya akan dilanjutkan dengan proses pembelajaran (*learning*) menggunakan metode klasifikasi

naive bayes. Setelah diberi data latih sebanyak 80% kepada sistem, selanjutnya sistem dapat melakukan pengujian menggunakan data uji dengan tujuan menguji ketepatan suatu sistem dalam mengklasifikasikan data. Pada proses pengklasifikasian perlu disiapkan *library* yang akan digunakan yaitu *library sklearn* atau *scikit learn* sama seperti proses ekstraksi fitur dalam menggunakan *library* tersebut.

Adapun *library sklearn* yang dipakai pada proses klasifikasi di antaranya adalah *MultinomialNB*, *accuracy_score*, *precision_score*, *recall_score*, *f1_score*, *classification_report* dan *confusion matrix*.

Proses awal mulanya sama seperti sebelumnya yaitu siapkan *library sklearn* terlebih dahulu. Pada *google colab* sudah menyediakan *library sklearn* ini maka tidak perlu dilakukan instalasi. Jadi, langsung saja melakukan peng-*importan library* tersebut dengan memanggil beberapa kelas yang dibutuhkan. Berikut *source code* meing-*import library sklearn* pada proses klasifikasi dapat dilihat pada gambar 4.37.

```
#klasifikasi naive bayes
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
```

Gambar 4. 37 *Import Library Sklearn* pada proses klasifikasi

Setelah itu, selanjutnya proses pengklasifikasian dengan metode *naive bayes*. Adapun kode program proses klasifikasi *naive bayes* dapat dilihat pada gambar 4.38.

```
clf = MultinomialNB().fit(Train_X_Tfidf.toarray(), y_train)
predicted = clf.predict(Test_X_Tfidf.toarray())
print("Multinomial NB Accuracy:" , accuracy_score(y_test,predicted))
print("Multinomial NB Precision:", precision_score(y_test,predicted,average='weighted'))
print("Multinomial NB Recall :",recall_score(y_test,predicted,average='weighted'))
print("Multinomial NB F1 Score:",f1_score(y_test,predicted,average='weighted'))
print(f'confusion matrix:\n {confusion_matrix(y_test, predicted)}')
print('-----\n')
print(classification_report(y_test,predicted,zero_division=0))
```

Gambar 4. 38 Proses Pengklasifikasian dengan metode *naive bayes*

E. Uji Model

Pada tahap uji model dilakukan dengan tujuan agar dapat ketepatan model dalam pengklasifikasian dan mendapatkan hasil klasifikasi yang ditunjukkan pada tabel *multiclass confusion matrix*. Pada *multiclass confusion matrix* sendiri terdapat dua kelas, yaitu kelas prediksi (*predicted class*) dan kelas sebenarnya (*true class*). Pada penelitian ini menggunakan *multiclass confusion matrix* karena terdapat tiga kelas yang digunakan yaitu kelas negatif, netral dan positif. Berikut tabel dari *multiclass confusion matrix* 3 x 3 dapat ditunjukkan pada tabel 4.5.

Tabel 4. 5 *Multiclass Confusion Matrix 3 x 3*

		<i>Predicted Class</i>		
		Negative	Netral	Positif
<i>True Class</i>	Negative	T Neg	F NegNet	F NegPos
	Netral	F NetNeg	T Net	F NetPos
	Positif	F PosNeg	F PosNet	T Pos

Untuk mengetahui ketepatan model dalam pengklasifikasian maka harus ditentukan nilai akurasi. Dengan cara, menjumlahkan total data yang diprediksi sesuai dengan data sebenarnya, kemudian dibagi dengan jumlah keseluruhan data yang di uji. Berikut rumus untuk menghitung nilai akurasi dapat dilihat pada persamaan 4.1.

$$\text{Akurasi} = \frac{TPos+T Neg+T Net}{\text{Total data yang di Uji}} \quad (4.1)$$

Setelah dilakukan tahapan uji model, sehingga didapatkan hasil nilai akurasi dan *multiclass confusion matrix* 3 x 3 yang ditunjukkan pada gambar 4.39.

```
MultinomialNB Accuracy: 0.8148148148148148  
confusion matrix:  
[[151  0 13]  
 [ 19  2  5]  
 [ 18  0 89]]
```

Gambar 4. 39 Hasil Uji Model

Dari gambar 4.39 dapat disimpulkan hasil uji model diperoleh nilai akurasi sebesar 0.81 atau 81% yang didapatkan dari *multiclass confusion matrix* 3x3. Namun, menghitung nilai akurasi saja tidaklah cukup karena jumlah data tiap kelas pada data latih tidaklah seimbang maka perlu dilakukan evaluasi.

F. Evaluasi Model

Setelah melalui proses uji model, selanjutnya dilakukan tahap evaluasi model untuk mengetahui performa pada model tersebut. Pada penelitian ini, peneliti melakukan perhitungan performa yang terdiri dari nilai akurasi, *presicion*, *recall* dan *f1 score* yang dihitung dari *multiclass confusion matrix* 3x3. Diketahui sebelumnya dari tahapan uji model didapatkan hasil *multiclass confusion matrix* 3x3 yang di tunjukkan pada tabel 4.6.

Tabel 4. 6 Hasil *Multiclass Confusion Matrix*

		<i>Predicted Class</i>		
		Negative	Netral	Positif
<i>True Class</i>	Negative	151	0	13
	Netral	19	2	5
	Positif	18	0	89

Kemudian, peneliti melakukan perhitungan secara manualisasi untuk menghitung nilai performa dari model *naive bayes* pada nilai akurasi yang dihitung dari tabel *multiclass confusion matrix* di atas. Sehingga, hasil perhitungan nilai akurasi secara manualisasi dapat dijabarkan sebagai berikut :

$$\text{Akurasi} = \frac{T \text{ Neg} + T \text{ Net} + T \text{ Pos}}{\text{Total data yang di Uji}} \times 100\%$$

$$\text{Akurasi} = \frac{151+2+89}{151+0+13+19+2+5+18+0+89} \times 100\%$$

$$\text{Akurasi} = \frac{242}{297} \times 100\%$$

$$\text{Akurasi} = 0,81 \times 100\%$$

$$\text{Akurasi} = 81\%$$

Selanjutnya untuk menghitung performa model lainnya pada nilai *presisi*, *recall* serta *f1 score* maka perlu ditentukan nilai *true positif*, *true negatif*, *false positif* dan *false negatif* pada masing-masing kelas yang terdapat pada *multiclass confusion matrix* dengan berukuran 3x3. Pada ukuran 3x3 sangatlah sulit untuk memperoleh nilai *true positif*, *true negatif*, *false positif*

dan *false* negatif pada tiap-tiap kelasnya. Maka dari itu, untuk mempermudah mencari nilai-nilai tersebut dilakukanlah perubahan menjadi kolom 2x2 terlebih dahulu. Sehingga menjadi seperti berikut :

- Kelas Negatif

Tabel 4. 7 Perhitungan nilai pada kelas Negatif

True\Pred	Negatif	Bukan
Negatif	TP = 151	FN=(0+13)= 13
Bukan	FP=(19+18) = 37	TN=(2+5+0+89)= 96

- Kelas Netral

Tabel 4. 8 Perhitungan nilai pada kelas Netral

True\Pred	Netral	Bukan
Netral	TP = 2	FN=(19 +5)= 24
Bukan	FP=(0+0) = 0	TN=(151+13+18+89)= 271

- Kelas Positif

Tabel 4. 9 Perhitungan nilai pada Kelas Positif

True\Pred	Positif	Bukan
Positif	TP = 89	FN=(18 +0)= 18
Bukan	FP=(13+5)= 18	TN=(151+15+19+2)= 187

Setelah didapatkan nilai *true* positif, *true* negatif, *false* positif dan *false* negatif. Sehingga berdasarkan rumus pada bab

sebelumnya nilai *precision*, *recall*, serta *f1 score* dari tabel hasil *multiclass confusion matrix* pada tabel di atas dapat disimpulkan menjadi pada tabel 4.10 sebagai berikut :

Tabel 4. 10 Hasil Perhitungan Performa

<i>Class</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>	<i>False Negative (FN)</i>	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
Negatif	151	37	13	80%	92%	86%
Netral	2	0	24	100%	8%	14%
Positif	89	18	18	83%	83%	83%

Sehingga berdasarkan tabel 4.10 nilai *precision* dapat dihitung pada nilai negatif sebesar 80%, nilai netral sebesar 100% serta nilai positif sebesar 83%. Untuk nilai *recall* dapat dihitung pada nilai negatif sebesar 92%, nilai netral sebesar 8% serta nilai positif sebesar 83%. Sedangkan untuk nilai *f1 score* pada nilai negatif sebesar 86%, nilai netral sebesar 14% dan nilai positif sebesar 83%. Untuk perhitungan nilai akurasi, *precision*, *recall* dan *f1 score* pada suatu sistem dapat menggunkan *source code* yang tunjukkan pada gambar 4.40 sebagai berikut.

```

#klasifikasi naive bayes
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix

clf = MultinomialNB().fit(Train_X_Tfidf.toarray(), y_train)
predicted = clf.predict(Test_X_Tfidf.toarray())
print("Multinomial NB Accuracy:" , accuracy_score(y_test,predicted))
print("Multinomial NB Precision:", precision_score(y_test,predicted,average='weighted'))
print("Multinomial NB Recall :",recall_score(y_test,predicted,average='weighted'))
print("Multinomial NB F1 Score:",f1_score(y_test,predicted,average='weighted'))
print(f'confusion matrix:\n {confusion_matrix(y_test, predicted)}')
print('-----\n')
print(classification_report(y_test,predicted,zero_division=0))

```

Gambar 4. 40 Source Code Perhitungan performa *Naive Bayes*

Kemudian, selanjutnya mencari nilai akuarasi, *precision*, *recall* dan *f1 score* pada pemrograman bahasa *python* yang telah dibuat pada kode program pada gambar 4.40 di atas.

Sehingga hasil dari kode program tersebut didapatkan hasil performa metode *naive bayes* dari tiap-tiap kelasnya melalui nilai *precision*, *recall*, dan *f1 score*. Hasil dari nilai *precision*, *recall*, dan *f1 score* berupa angka desimal dengan rentang nilai sebesar 0-1 yang artinya, semakin tinggi angka yang diperoleh semakin baik juga hasil dari nilai tersebut. Berikut hasil dari keseluruhan proses evaluasi model dapat ditunjukkan pada gambar 4.41.

```

Multinomial NB Accuracy: 0.8148148148148148
Multinomial NB Precison: 0.8307185328461923
Multinomial NB Recall : 0.8148148148148148
Multinomial NB F1 Score: 0.785921990467445
confusion matrix:
[[151  0 13]
 [ 19  2  5]
 [ 18  0 89]]
-----

```

	precision	recall	f1-score	support
0	0.80	0.92	0.86	164
1	1.00	0.08	0.14	26
2	0.83	0.83	0.83	107
accuracy			0.81	297
macro avg	0.88	0.61	0.61	297
weighted avg	0.83	0.81	0.79	297

Gambar 4. 41 Hasil Pengukuran Evaluasi Performa

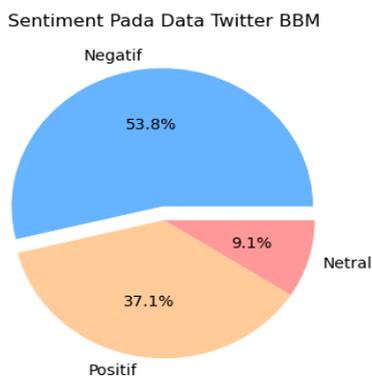
Hasil dari evaluasi model dapat dilihat bahwa nilai *precision*, *recall*, *f1 score* disetiap kelas dapat dikatakan memiliki tingkat kemampuan yang baik pada model *naive bayes*. Sehingga dari gambar diatas dapat disimpulkan tingkat keberhasilan sistem dalam mencari ketepatan antara informasi yang diminta oleh pengguna atau disebut *precision* didapatkan nilai *precision* untuk kelas negatif sebesar 80%, untuk kelas netral sebesar 100%, untuk kelas positif sebesar 83% sehingga dari angka tersebut dapat diartikan bahwa proporsi label yang diprediksi dengan netral dari total prediksi sangatlah tinggi diantara kelas positif dan negatif. Sedangkan tingkat keberhasilan sistem dalam menemukan kembali

sebuah informasi atau disebut *recall* didapatkan kelas negatif sebesar 92%, untuk kelas netral sebesar 0.08% dan kelas positif sebesar 83%. Sehingga hal ini berarti kinerja keberhasilan sistem dalam menemukan kembali informasi yang bernilai netral dalam dokumen sangat rendah dibandingkan dengan menemukan informasi kembali yang bernilai positif dan negatif. Sehingga diperoleh total keseluruhan dari nilai *precision*, *recall* serta *f1 score* pada gambar 4.35 didapatkan nilai *precision* sebesar 83%, nilai *recall* sebesar 81% dan nilai *f1 score* sebesar 79%.

G. Visualisasi

Tahap terakhir yaitu memvisualisasikan hasil dari analisis sentimen menggunakan diagram lingkaran dan *wordcloud*. *Wordcloud* akan digunakan untuk memvisualisasikan hasil analisis klasifikasi. Tujuan dari visualisasi tersebut untuk mengetahui jumlah sentimen dan kata yang sering dibicarakan pada topik kenaikan harga BBM sehingga dapat diperoleh informasi yang ditampilkan dalam bentuk gambar.

Diketahui sebelumnya hasil akhir dari proses *preprocessing* menghasilkan data sebanyak 798 komentar bernilai negatif, 550 komentar bernilai positif dan 135 komentar bernilai netral. Persentase data bernilai positif, negatif dan netral dapat dilihat pada gambar 4.36.



Gambar 4. 42 Persentase Pada Sentimen

Pada gambar 4.36 dapat disimpulkan sentimen negatif memiliki nilai persentase tertinggi sebesar 53.8%, pada sentimen positif sebesar 37.1% serta pada sentimen netral sebesar 9.1% sehingga dari angka tersebut diperoleh bahwa hampir setengah tanggapan mengenai kebijakan pemerintah pada kenaikan harga BBM di media sosial twitter masih kurang setuju atau kontra. Sisanya sebesar 37.1% menyatakan setuju atau pro terhadap kebijakan tersebut dan 9.1% lainnya berisikan sentimen netral atau memuat informasi saja.

Setelah itu, peneliti ingin mengetahui juga kata-kata yang banyak diperbincangkan oleh respon pengguna media sosial twitter di Indonesia, untuk itu peneliti mengvisualisasikannya dalam bentuk *wordcloud* pada gambar 4.37.

BAB V

KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan penelitian yang dilaksanakan, dapat disimpulkan bahwa :

1. Metode *Naive Bayes Classifier* terbukti dapat menganalisis sentimen dengan klasifikasi yang baik. Dalam penelitian digunakan komposisi data sebesar 80% pada data latih dan 20% data uji dari jumlah keseluruhan data diambil secara acak. Proses yang dilakukan untuk melakukan analisis sentimen pada penelitian ini mulanya melakukan ekstraksi teks dengan melalui tahapan *preprocessing*, kemudian pembobotan dengan TFIDF, dan klasifikasian teks dengan metode *naive bayes*. Hasil klasifikasi yang diberikan dapat berupa kelas dengan sentimen positif, negatif dan netral.
2. Dari penelitian ini didapatkan sebanyak 1500 data *tweet* kemudian setelah melalui tahap *preprocessing* menjadi 1483 data *tweet* dengan menghasilkan data sebanyak 550 komentar bersentimen positif, 798 komentar sentimen negatif dan 135 komentar sentimen netral. Sehingga diketahui sentimen negatif

memiliki nilai persentase tertinggi sebesar 53.8%, pada sentimen positif sebesar 37.1% serta pada sentimen netral sebesar 9.1%.

3. Berdasarkan hasil performa pada metode *Naive Bayes* terbukti algoritma yang akurat karena menghasilkan nilai akurasi sebesar 81%, *precision* sebesar 83%, *recall* sebesar 81% serta *f1 score* sebesar 79%.

B. Saran

Berdasarkan penelitian yang dilaksanakan, penulis berharap kepada peneliti selanjutnya untuk dapat dikembangkan dan terdapat beberapa saran-saran, yaitu:

1. Menggunakan metode klasifikasi yang berbeda seperti *Support Vector Machine (SVM)*, *K-NN*, *Decision Tree* dengan begitu dapat membandingkan hasil performa yang dilakukan untuk mencari metode klasifikasi yang terbaik.
2. Pada penelitian ini, data diambil dari twitter pada penelitian berikutnya diharapkan data dapat diperoleh dari media sosial lainnya seperti youtube, facebook, instagram atau tik tok dll.
3. Penambahan koleksi kamus pada kata yang tidak baku/gaul, karena pada media sosial banyak komentar yang berisikan bahasa yang kurang baku.

DAFTAR PUSTAKA

- Apriani, R., Gustian, D., Program, S., Sistem, I., Putra, U. N., Indonesia, S., Raya, J., Kaler, C., 21, N., & Sukabumi, K. (2019). ANALISIS SENTIMEN DENGAN NAÏVE BAYES TERHADAP KOMENTAR APLIKASI TOKOPEDIA. In *Jurnal Rekayasa Teknologi Nusa Putra* (Vol. 6, Issue 1).
- B, L. (2010). Sentiment Analysis and Subjectivity,. In *Handbook Of Natural Language Processing* (2nd ed., pp. 1-704). Chapman & Hall/CRC.
- Carneiro, T., da Nobrega, R. V. M., Nepomuceno, T., Bian, G. bin, de Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*, 6, 61677-61685. <https://doi.org/10.1109/ACCESS.2018.2874767>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, 15(1), 131-145.
- Faiq, M., Ardyanto Putro, & Erwin Budi Setiawan. (2022). Analisis Sentimen Terhadap Kebijakan Pemerintah dengan Feature Expansion Metode GloVe pada Media sosial Twitter. *E-Proceeding of Engineering*, 9(1), 54-66.
- Fajar, R., Program, S., Rekayasa, P., Lunak, N., & Bengkalis, R. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *JURNAL INOVTEK POLBENG*, 3(1), 50-59.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fikri, M. I., Sabrila, T. S., Azhar, Y., & Malang, U. M. (2020). Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA Jurnal*, 10(2), 71-76.

- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. <http://arxiv.org/abs/2008.05756>
- Gunawan, B., Sasty, H., #2, P., Esyudha, E., & #3, P. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika*, 4(2), 17–29. www.femaledaily.com
- Imron, A. (2019). *ANALISIS SENTIMEN TERHADAP TEMPAT WISATA DI KABUPATEN REMBANG MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER*. Universitas Islam Indonesia.
- Kurniasih, U., & Suseno, A. T. (2022). Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(4), 2335–2340. <https://doi.org/10.30865/mib.v6i4.4958>
- Luuqyana, W. A. (2018). *ANALISIS SENTIMEN CYBERBULLYING PADA KOMENTAR INSTAGRAM DENGAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE*. Universitas Brawijaya.
- Mahfudh, A. A., & Mustofa, H. (2019). Klasifikasi Pemahaman Santri Dalam Pembelajaran Kitab Kuning Menggunakan Algoritma Naive Bayes Berbasis Forward Selection. *Walisongo Journal of Information Technology*, 1(2), 101. <https://doi.org/10.21580/wjit.2019.1.2.4529>
- Mas Pintoko, B., & Muslim, K. (2018). Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naive Bayes Classifier. *E-Proceeding of Engineering*, 5(3), 8121–8130.
- Matt Holand. (2017). *Twitter. A guide to searching for research purposes*. NWA LKS.
- Mujahidin, S., Prasetyo, B., & Utomo, M. C. C. (2022). Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan

- Metode Gaussian naïve bayes. *Jurnal Vocational Teknik Elektronika Dan Informatika*, 10(3), 17–24.
<http://ejournal.unp.ac.id/index.php/voteknika/index>
- Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes. *Walisongo Journal of Information Technology*, 1(1), 1.
<https://doi.org/10.21580/wjit.2019.1.1.3915>
- Nurhuda, F., Sihwi, S. W., & Doewes, A. (2013). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal ITSMART*, 2(2), 35–42.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). *Scikit-learn: Machine Learning in Python*. <http://arxiv.org/abs/1201.0490>
- Pradany, L. N., & Faticah, C. (2016). ANALISA SENTIMEN KEBIJAKAN PEMERINTAH PADA KONTEN TWITTER BERBAHASA INDONESIA MENGGUNAKAN SVM DAN K-MEDOID CLUSTERING. *SCAN*, 11(1), 59–66.
- Putri, M. I., & Kharisudin, I. (2022). Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Analisis Sentimen Data Review Pengguna Aplikasi Marketplace Tokopedia. *PRISMA, Prosiding Seminar Nasional Matematika*, 5, 759–766.
<https://journal.unnes.ac.id/sju/index.php/prisma/>
- Rish, Irina. (2001). Empirical Study of The Naive Bayes Classifier. *IJCAI 2001 Work Empir Methods Artif Intell*, 3, 41–46.
- Rustiana, D. N. R. (2017). ANALISIS SENTIMEN PASAR OTOMOTIF MOBIL: TWEET TWITTER MENGGUNAKAN NAÏVE BAYES. *Jurnal SIMETRIS*, 8.

- Sandy, B. C. D. M. A. I. (2015). *ANALISIS SENTIMEN TERHADAP KENAIKAN HARGA BAHAN BAKAR MINYAK (BBM) PADA MEDIA ONLINE* (R. Kiuk, Ed.). Sekolah Tinggi Manajemen Informatika & Komputer (STIKOM) Uyelindo Kupang. <http://www.uyelindo.ac.id>.
- Sandy, B. C. D. M. A. I. (2016). *ANALISIS TOPIK-TOPIK YANG MEMPENGARUHI TERJADINYA SENTIMEN TERKAIT KENAIKAN HARGA BAHAN BAKAR MINYAK (BBM) PADA MEDIA ONLINE*. <http://www.uyelindo.ac.id>.
- Sholekha, I., Faqih, A., & Bahtiar, A. (2022). Sentiment Analysis of Public Opinion Covid-19 Vaccine Using Naïve Bayes and Random Forest Methods. *JURNAL TEKNIK INFORMATIKA*, 15(1), 34–43. <https://doi.org/10.15408/jti.v15i1.24847>
- Sonawanse, S. & V. A. K. (2016). *Teknik Analisis Sentimen Data Twitter: Survei*. <http://ai.stanford>.
- Sujadi, H. (2022). ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP WABAH COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE. *INFOTECH Journal*, 8(1), 22–27. <https://doi.org/10.31949/infotech.v8i1.1883>
- Syakuro, A. (2017). *Analisis Sentimen Terhadap Masyarakat Terhadap E-COMMERCE Pada Media Sosial Menggunakan Metode Naive Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)*. Universitas Islam Negeri Maulana Malik Ibrahim.
- Tangkelayuk, A., & Mailoa, E. (2022). Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree. *Jurnal Teknik Informatika Dan Sistem Informasi*, 9(2), 1109–1119. <http://jurnal.mdp.ac.id>
- Turban, E.; et. al. (2011). *Decision Support and Business Intelligence Systems* (9th ed.). Pearson Education.
- Turmudi Zy, A., Adji Ardiansyah, L., & Maulana, D. (2021). Implementasi Algoritma Naïve Bayes Dalam

- Mendiagnosa Penyakit Angin Duduk. *Jurnal Pelita Teknologi*, 16(1), 52–65.
- Wandani, A. (2021). Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN, Random Forest, dan Naive Bayes. In *Jurnal Sains Komputer & Informatika (J-SAKTI)* (Vol. 5, Issue 2).
- Wenty Dwi Yuniarti. (2019). *Dasar-Dasar Pemrograman dengan Python*. Deepublish.
- Yerzi, F. S., & Sibaroni, Y. (2021). ANALISIS SENTIMEN TERHADAP KEBIJAKAN PEMERINTAH DALAM MENANGANI COVID-19 DENGAN PENDEKATAN LEXICON BASED. *E-Proceeding of Engineering*, 8(5), 11354–11366.

DAFTAR LAMPIRAN

LAMPIRAN 1 : Contoh Dokumen Hasil *Crawling data* Twitter

No	Username	Content
1	GhiffariDKI	kami mendengar dan merasa jeritanmu tapi nga pake nangis seluruh kader pdemokrat menolak kenaikan harga bbm
2	Darwan45537622	hanya peduli proyek kereta cepat dan ikn pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah himpitan ekonomi yg sulit dgn teganya justru menaikkan harga bbm kata achmad nur padahal saat ini harga minyak dunia sedang turun
3	Thavit_Amoura	bem dan geram gelar demo penolakan kenaikan harga bbm ini tuntutan yang disampaikan
4	Irfqn_	kebijakan ini tentu menguntungkan korporat dan merugikan rakyat sebab harus membeli bbm dengan harga yang relatif mahal
5	duckbusterr	jalanan kotanya ditata rapi tarif angkutan umum massalnya ga menyulitkan warga yg tinggal di dki semuanya serba dimudahkan meskipun presiden yg gw ga tau siapa itulah namanya gw sudah muak sebut namanya menaikkan harga bbm tp pak anies mendengar suara hati warganya salut clapping hands medium light skin tone
6	Irfqn_	karena itu diduga kuat pencabutan subsidi bbm dan pelarangan bbm beroktan rendah adalah untuk melepas harga bbm ke pasar sehingga produk

		bbm korporasi asing dapat bersaing dengan pertamina
7	Irfqn_	dengan kondisi masyarakat yang sebagian besar bertaraf hidup menengah ke bawah bbm dengan harga murah masih menjadi pilihan
8	Irfqn_	apalagi sejak lama diketahui bahwa ada upaya pemerintah yakni pertamina untuk melepas harga bbm ke pasar
9	Darwan45537622	hanya peduli proyek kereta cepat dan ikn pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah himpitan ekonomi yg sulit dgn teganya justru menaikkan harga bbm kata achmad nur padahal saat ini harga minyak dunia sedang turun
10	ZettaZahra2	vivo revvo sebelum pertamina menaikkan harga bbm sudah menurunkan harga revvo pertamina menaikkan harga bbm konsumen beralih ke vivo revvo dirjen esdm meminta vivo menyesuaikan menaikkan harga

.....

1498	KangMasroer	kabar gembira menteri esdm ada potensi harga bbm turun
1499	LantasPBM	kasat lantas polres prabumulih akp muthemainah s h melaksanakan kegiatan bakti sosial memberikan bantuan sembako terhadap masyarakat yang terdampak kenaikan harga bbm prabumulih september
1500	detik_jabar	terima audiensi driver ojol pemkab garut tolak kenaikan harga bbm

LAMPIRAN 2 : Contoh Dokumen Data yang Sudah Diberi Label

no	username	Sentiment	content
1	GhiffariDKI	Negatif	kami mendengar dan merasa jeritanmu tapi nga pake nangis seluruh kader pdemokrat menolak kenaikan harga bbm
2	Darwan45537622	Negatif	hanya peduli proyek kereta cepat dan ikn pemerintah amat kejam di tengah kondisi masyarakat yang berada di bawah himpitan ekonomi yg sulit dgn teganya justru menaikkan harga bbm kata achmad nur padahal saat ini harga minyak dunia sedang turun
3	Thavit_Amoura	Negatif	bem dan geram gelar demo penolakan kenaikan harga bbm ini tuntutan yang disampaikan
4	Irfqn_	Negatif	kebijakan ini tentu menguntungkan korporat dan merugikan rakyat sebab harus membeli bbm dengan harga yang relatif mahal
5	duckbusterr	Netral	jalanan kotanya ditata rapi tarif angkutan umum massalnya ga

			menyulitkan warga yg tinggal di dki semuanya serba dimudahkan meskipun presiden yg gw ga tau siapa itulah namanya gw sudah muak sebut namanya menaikkan harga bbm tp pak anies mendengar suara hati warganya salut clapping hands medium light skin tone
6	Irfqn_	Positif	karena itu diduga kuat pencabutan subsidi bbm dan pelarangan bbm beroktan rendah adalah untuk melepas harga bbm ke pasar sehingga produk bbm korporasi asing dapat bersaing dengan pertamina

.....

1498	KangMasroer	Positif	kabar gembira menteri esdm ada potensi harga bbm turun
1499	LantasPBM	Positif	kasat lantas polres prabumulih akp muthemainah s h melaksanakan kegiatan bakti sosial memberikan bantuan sembako terhadap masyarakat yang terdampak kenaikan harga bbm prabumulih september
1500	detik_jabar	Negatif	terima audiensi driver ojol pemkab garut tolak kenaikan harga bbm

LAMPIRAN 3 : Contoh Data Pendukung "colloquial-indonesian-lexicon"

No	slang	formal
1	woww	wow
2	aminn	amin
3	met	selamat
4	netaas	menetas
5	keberpa	keberapa
6	eeeehhhh	eh
7	kata2nyaaa	kata-katanya
8	hallo	halo
9	kaka	kakak
10	ka	kak
11	daah	dah
12	aaaaahhhh	ah
13	yaa	ya
14	smga	semoga
15	slalu	selalu
16	amiin	amin
17	kk	kakak
18	trus	terus
19	kk	kakak

.....

15396	wktu	waktu
15397	hr	hari
15398	gatau	enggak tau

**LAMPIRAN 4 : Contoh Data Pendukung
"FormalizationDict.txt"**

7an	tujuan
@	di
ababil	abg labil
abis	habis
acc	accord
ad	ada
adlah	adalah
adlh	adalah
adoh	aduh
afaik	as far as i know
aha	tertawa
ahaha	haha
aing	saya
aj	saja
aja	saja
.....	
wkwk	tertawa
xixixi	tertawa
y	kenapa
ya	iya
ybs	yang bersangkutan
yg	yang
yi	yaitu
yl	yang lain
yg	yang
yo	iya
yoha	iya
yowes	ya sudah
yup	iya
ywdh	ya sudah

LAMPIRAN 5 : Contoh Data Pendukung “KBBA.txt”

7an	tujuan
@	di
ababil	abg labil
abis	habis
acc	accord
ad	ada
adlah	adalah
adlh	adalah
adoh	aduh
afaik	as far as i know
aha	tertawa
ahaha	haha
aing	saya
aj	saja
aja	saja
ajep-ajep	dunia gemerlap
ajj	saja
ak	saya
aka	dikenal juga sebagai
akika	aku
akko	aku
akkoh	aku
.....	
wkwkkw	tertawa
soft	halus
setting	atur
angis	nangis
benarjujur	benar
benarjujur	jujur
sayan	sayang
mgkin	mungkin

LAMPIRAN 6 : Contoh Data Pendukung “Slangword.txt”

&:dan
+:tambah
/:atau
=:sama dengan
ababil:anak ingusan
abal2:palsu
abal:palsu
ad:ada
akooh:aku
alay:norak
albm:album
ampe:sampai
anjir:waw
anyway:ngomong-ngomong
aq:aku
asap:secepatnya
.....
tpii:tapi
trs:terus
trus:terus
ttg:tentang
ttp:tetap
tw:tau
typo:salah tulis
u:you
unyu:menggemaskan
with:dengan
woles:santai
wtf:apa-apaan
x:kali
y:ya
yg:yang

LAMPIRAN 7 : Contoh Dokumen Hasil Klasifikasi

no	Data Uji	Label Aktual	Label Prediksi
1	ngga,bbm,mending,turu,tidur, wajah,engga,libur,libur ,libure,turu,thumbs,yawning, wajah,sleeping,wajah	0	0
2	resiko,pertamina,sasar,marah, rakyat,harga,bbm,dipungkiri,nasib ,bumn,negeri,jadi,sapi,perah,peelihara ,rakyat,harga,bbm,tentu,negara	0	0
3	ntt,cap,provinsi,miskin,ganggu, denggan,naik,harga,bbm,makax, janggan,rendah,orang,kerja,kerja,kerja	2	0
4	bingung,faham,dprd,gorontalo,tolak, amp,lokasi,demo,tolak,hasil,efek, ngga,bbm,hasil,man,biar,engga, anarkis,demo,pulang,denggan,bangengga, hal,anak,menang,nak,permen,diam	0	0
5	perintah,naik,harga,bbm,demokrat, buru,kebun,binatang,kreatif,hbd,surabaya	0	0
6	inflasi,dampak,naik,harga,harga,butuh ,pokok,bbm,masyarakat,bantu,subsidi, tajam,kuat,data,manfaat,teknologi,informasi, mari,laku,upaya,upaya,meringgankan,biaya,hidup	2	0
7	ekonom,senior,faisal,basri,nilai,bijak,perintah, naik,harga,bbm,subsidi,paham,denggan,sesuai, harga,minyak,fenomena,global,bansosharus, tepatsasaran	2	2
.....			
296	pasca,naik,harga,bbm,polres,sampang ,ratus,paket,sembako,fit	2	2
297	semesta,menggasih,ajar,harga,bbm,mikir, hitung,ongkos,bolak,rumah,kantor,anyway, jarak,rumah,kantor	0	0