

**KLASIFIKASI DAN PERAMALAN JUMLAH OPINI KENAIKAN  
HARGA BAHAN BAKAR MINYAK DENGAN  
*MULTINOMIAL NAÏVE BAYES CLASSIFIER*  
DAN *NAÏVE FORECASTING***

**SKRIPSI**

Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana pada  
Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Pakuan

**Oleh :**

**Nita Oktaviani**

**064119014**



**PROGRAM STUDI MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS PAKUAN  
BOGOR  
2023**

*Bismillahirrahmanirrahim*

*Alhamdulillahirabbilaalamiin... Puji serta syukur saya panjatkan kehadirat Allah SWT karena atas berkah rahmat serta karunia-Nya saya dapat menyelesaikan tugas akhir ini dengan lancar.*

*Terima kasih saya ucapkan kepada keluarga tercinta, terkhusus kepada kedua orang tua saya ibu Rodiah, Bapak Solikin, Kakak perempuan saya Sri Ningsih, Kakak ipar saya Shalahudin Al Ayubi, Sepupu saya Meiliana Wahyudi, serta keponakan saya yang lucu Nada Almahyra, Arsyilla Romeesa dan Abrina Falisha yang telah memberikan kasih sayang, motivasi, dan tentunya do'a yang senantiasa selalu dipanjatkan sehingga saya dapat melewati segala rintangan dan hambatan yang ada selama penyusunan tugas akhir ini. Oleh karena itu saya dapat menyelesaikan tugas akhir ini dengan baik,*

*Terima kasih saya ucapkan kepada Ibu Dr. Embay Rohaeti, M.Si dan Ibu Maya Widayastiti, M.Si yang telah memberikan bimbingan yang sangat baik serta motivasi dalam proses penyusunan tugas akhir ini. Tidak lupa kepada seluruh dosen dan staf Program Studi Matematika Fakultas MIPA Universitas Pakuan yang telah memberikan ilmu yang sangat bermanfaat. Semoga apa yang sudah diberikan dibalas dengan kebaikan oleh Allah SWT.*

*Terima kasih saya ucapkan kepada teman istimewaiku Ahmad Fauzan Al Fajrian yang selalu rela menanggapi setiap keluh kesahku, teman belajar serta memberi motivasi dan semangat. Semoga Allah SWT memudahkan jalan kita menuju kesuksesan di masa depan. Terakhir saya ucapkan terima kasih kepada pihak-pihak yang tidak dapat saya sebutkan satu persatu atas bantuan yang diberikan sehingga saya dapat menyelesaikan tugas akhir ini. Semoga karya ini dapat memberikan manfaat bagi pembacanya.*

*Aamiin Ya Rabbal Aalamiin*

## HALAMAN PENGESAHAN

**JUDUL** : **KLASIFIKASI DAN PERAMALAN JUMLAH OPINI  
KENAIKAN HARGA BAHAN BAKAR MINYAK DENGAN  
*MULTINOMIAL NAÏVE BAYES CLASSIFIER* DAN  
*NAÏVE FORECASTING***

**NAMA** : **NITA OKTAVIANI**

**NPM** : **064119014**

**Bogor, Agustus 2023**

**Menyetujui,**

**Pembimbing Pendamping**



**Maya Widyastiti, M.Si**

**Pembimbing Utama**



**Dr. Embay Rohaeti, M.Si**

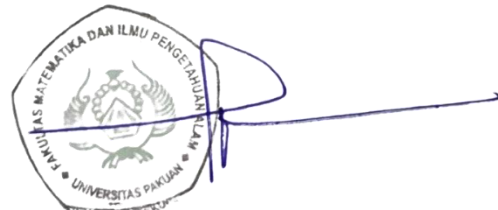
**Mengetahui,**

**Ketua Program Studi Matematika**



**Dr. Ir. Fitria Virgantari, M.Si**

**Dekan FMIPA**



**Asep Denih, S.Kom., M.Sc., Ph.D.**

**SURAT PERNYATAAN MENGENAI SKRIPSI DAN SUMBER  
INFORMASI SERTA PELIMPAHAN KEKAYAAN  
INTELEKTUAL DI UNIVERSITAS PAKUAN**

Saya yang bertanda tangan di bawah ini :

Nama : Nita Oktaviani

NPM : 064119014

Judul Skripsi : Klasifikasi dan Peramalan Jumlah Opini Kenaikan Harga Bahan Bakar Minyak dengan *Multinomial Naïve Bayes Classifier* dan *Naïve Forecasting*

Dengan ini menyatakan bahwa skripsi di atas adalah benar karya saya dengan arahan dari komisi pembimbing dan belum pernah diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Universitas Pakuan.

Bogor, Agustus 2023



Nita Oktaviani

064119014

## RIWAYAT HIDUP



Nita Oktaviani lahir di Bogor pada tanggal 3 Oktober 2000. Anak kedua dari dua bersaudara dan merupakan anak dari pasangan Ibu Rodiah dan Bapak Solikin.

Pada tahun 2007 penulis memulai pendidikan formal pada jenjang Sekolah Dasar di SD Negeri Dewi Sartika 1. Pada tahun 2013 penulis melanjutkan pendidikan di SMP Negeri 12 Bogor dan lulus pada tahun 2016. Pada tahun 2016 hingga 2019 penulis melanjutkan pendidikan di SMA Negeri 7 Bogor. Pada Tahun 2019 penulis melanjutkan pendidikan jenjang Strata Satu di Universitas Pakuan Bogor Fakultas Matematika dan Ilmu Pengetahuan Alam dengan pilihan Program Studi Matematika.

Selama menempuh pendidikan di Universitas Pakuan, penulis melaksanakan Praktek Kerja Lapangan di Dinas kependudukan dan Pencatatan Sipil Kota Bogor pada Tahun 2021.

Penulis aktif dalam kegiatan organisasi Himpunan Mahasiswa Matematika (HIMATIKA) Universitas Pakuan. Pada periode 2022 penulis menjabat sebagai anggota Departemen Pengembangan Sumber Daya Mahasiswa dan Agama (PSDMA). Pengalaman organisasi yang didapat yaitu menjadi *master of ceremony* pada acara Logika Matematika pada tahun 2020, anggota divisi editor dan menjadi *master of ceremony* pada acara Lomba Cepat Tepat Matematika pada tahun 2022. Penulis juga aktif dalam komunitas *Virtual Education Academy* sebagai anggota.

## RINGKASAN

**NITA OKTAVIANI**, Klasifikasi dan Peramalan Jumlah Opini Kenaikan Harga Bahan Bakar Minyak dengan *Multinomial Naïve Bayes Classifier* dan *Naïve Forecasting*. Dibimbing oleh EMBAY ROHAETI dan MAYA WIDYASTITI.

Data opini pengguna Twitter terhadap topik kenaikan harga bahan bakar minyak memiliki nilai sentimen yang dapat menentukan kelas opini dominan terhadap topik kenaikan harga bahan bakar minyak. Opini pengguna Twitter akan diklasifikasi dalam tiga kelas yaitu kelas opini positif, negatif dan netral menggunakan model *multinomial naïve bayes classifier*. Hasil klasifikasi yang diperoleh dilanjutkan pada tahapan peramalan jumlah opini dengan metode *naïve forecasting*. Tujuan dari penelitian ini yaitu mengklasifikasikan opini dengan *multinomial naïve bayes classifier*, meramalkan jumlah opini untuk jangka waktu satu minggu kedepan dengan *naïve forecasting*, dan mengevaluasi hasil klasifikasi serta hasil peramalan. Hal tersebut dapat dilakukan dengan penggunaan bantuan *software* R. Data yang digunakan berupa data sekunder sebanyak 2500 data. Periode pengambilan data selama satu minggu dimulai dari 20 Oktober 2022 hingga 27 Oktober 2022. Hasil dari pemodelan *multinomial naïve bayes classifier* didapatkan tiga kelas yaitu sebanyak 775 dokumen diklasifikasikan sebagai opini negatif, 475 dokumen diklasifikasikan sebagai opini netral, dan 581 dokumen diklasifikasikan sebagai opini positif. Akurasi model *multinomial naïve bayes classifier* dikategorikan sangat baik sebesar 92% untuk keseluruhan kelas. Hasil peramalan tiga kelas klasifikasi dengan *naïve forecasting* yaitu jumlah opini kelas positif sebanyak 44 opini dengan nilai RMSE sebesar 8,96, jumlah opini kelas negatif sebanyak 25 opini dengan nilai RMSE sebesar 14,87, dan jumlah opini kelas netral sebanyak 21 opini dengan nilai RMSE sebesar 11,45. Hal ini menunjukkan Peramalan dengan *naïve forecasting* dikategorikan cukup baik.

Kata kunci : Opini, Klasifikasi, Peramalan, *Multinomial naïve bayes classifier*, *Naïve forecasting*

## KATA PENGANTAR

Puji serta syukur penulis panjatkan kepada Allah SWT karena atas berkah rahmat dan karunia-Nya penulis dapat menyelesaikan laporan penelitian yang berjudul **“Klasifikasi dan Peramalan Jumlah Opini Kenaikan Harga Bahan Bakar Minyak dengan *Multinomial Naïve Bayes Classifier* dan *Naïve Forecasting*”**.

Penelitian ini ditujukan sebagai salah satu syarat dalam memperoleh gelar Sarjana Matematika Program Studi Matematika Fakultas MIPA Universitas Pakuan. Dalam penyusunan laporan penelitian ini, banyak pihak yang berkontribusi membantu penulis untuk dapat terselesaikannya laporan penelitian ini. Pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Dr. Embay Rohaeti, M.Si selaku pembimbing utama.
2. Maya Widyastiti, M.Si selaku pembimbing pendamping.
3. Dr. Ir. Fitria Virgantari, M.Si selaku Ketua Program Studi Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Pakuan.
4. Keluarga yang senantiasa memberikan dukungan dan semangat.
5. Teman istimewa di Program Studi Matematika.
6. Pihak lain yang telah membantu dan tidak bisa disebutkan satu persatu.

Diharapkan laporan penelitian ini dapat bermanfaat baik bagi pembaca terkhusus kepada mahasiswa Program Studi Matematika dan Ilmu Pengetahuan Alam Universitas Pakuan. Oleh karena itu, kritik dan serta saran yang membangun diharapkan demi penelitian ini menjadi lebih baik.

Bogor, Agustus 2023



Penulis

# DAFTAR ISI

	<b>Halaman</b>
<b>HALAMAN PENGESAHAN</b> .....	<b>i</b>
<b>SURAT PERNYATAAN</b> .....	<b>ii</b>
<b>RIWAYAT HIDUP</b> .....	<b>iii</b>
<b>RINGKASAN</b> .....	<b>iv</b>
<b>KATA PENGANTAR</b> .....	<b>v</b>
<b>DAFTAR ISI</b> .....	<b>vi</b>
<b>DAFTAR GAMBAR</b> .....	<b>ix</b>
<b>DAFTAR TABEL</b> .....	<b>x</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xi</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Tujuan .....	3
1.3 Ruang Lingkup .....	3
1.4 Manfaat .....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>5</b>
2.1 <i>Machine Learning</i> .....	5
2.2 <i>Crawling Data</i> .....	5
2.3 <i>Lexicon Based</i> .....	5
2.4 <i>Data Preprocessing</i> .....	6
2.5 <i>Hold Out Validation</i> .....	7
2.6 Frekuensi dan Pembobotan Istilah.....	7
2.6.1 <i>Term Frequency (TF)</i> .....	7
2.6.2 <i>Inverse Document Frequency (IDF)</i> .....	8
2.6.3 <i>Term Frequency - Inverse Document Frequency (TF – IDF)</i> .....	8
2.7 <i>Naïve Bayes Classifier</i> .....	8
2.7.1 <i>Definisi dan Konsep Dasar Naïve Bayes Classifier</i> .....	8
2.7.2 <i>Multinomial Naïve Bayes Classifier</i> .....	9



2.7.3	Bentuk Persamaan <i>Multinomial Naïve Bayes Classifier</i> .....	9
2.7.4	<i>Maximum A Posteriori (MAP)</i> .....	9
2.7.5	<i>Maximum Likelihood (MLE)</i> .....	10
2.8	<i>Naïve Forecasting</i> .....	11
2.9	Evaluasi Hasil Klasifikasi .....	11
2.9.1	<i>Confusion Matrix</i> .....	11
2.10	Evaluasi Hasil Peramalan.....	13
2.10.1	<i>Root Mean Squared Error (RMSE)</i> .....	13
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>14</b>
3.1	Data.....	14
3.2	Tahapan Analisis .....	14
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>		<b>18</b>
4.1	Pelabelan <i>Tweet</i> .....	18
4.2	Pra Proses Data.....	18
4.2.1	Pembersihan Data .....	19
4.2.2	<i>Case Folding</i> .....	20
4.2.3	Penghapusan <i>Stopwords</i> .....	21
4.2.4	<i>Stemming</i> .....	22
4.2.5	Tokenisasi.....	23
4.3	Pembagian Data.....	24
4.4	Pembentukan <i>Term Frequency - Inverse Document Frequency (TF-IDF)</i> ..	24
4.5	Model <i>Multinomial Naïve Bayes Classifier</i> .....	25
4.5.1	Perhitungan Probabilitas <i>Prior</i> .....	25
4.5.2	Perhitungan Probabilitas Kelas.....	26
4.6	Proses Klasifikasi Model <i>Multinomial Naïve Bayes</i> .....	27
4.7	Evaluasi Model <i>Multinomial Naïve Bayes Classifier</i> .....	30
4.7.1	Evaluasi Kelas Klasifikasi.....	31
4.8	Peramalan Tiga Kelas Klasifikasi Dengan <i>Naïve Forecasting</i> .....	33
4.8.1	Tabulasi Data Jumlah Opini .....	33
4.8.2	Proses Peramalan <i>Naïve Forecasting</i> .....	34

4.9 Evaluasi Hasil peramalan.....	36
<b>BAB V PENUTUP .....</b>	<b>38</b>
5.1 Kesimpulan.....	38
5.2 Saran .....	39
<b>DAFTAR PUSTAKA.....</b>	<b>40</b>
<b>LAMPIRAN .....</b>	<b>43</b>

## DAFTAR GAMBAR

	<b>Halaman</b>
1. Diagram Alir Tahapan Analisis .....	15
2. Grafik Hasil Pelabelan <i>Tweet</i> .....	18
3. Visualisasi <i>Term</i> Terpopuler dalam <i>Wordcloud</i> .....	23
4. Grafik Hasil Pembagian Data .....	24
5. Grafik Pola Data Runtun Waktu Periode Satu minggu.....	34
6. Grafik Hasil Peramalan .....	35

## DAFTAR TABEL

	<b>Halaman</b>
1. <i>Confusion Matrix</i> Klasifikasi Multikelas .....	11
2. Proses Pembersihan Data .....	19
3. Proses <i>Case Folding</i> .....	20
4. Proses Penghapusan <i>Stopwords</i> .....	21
5. Proses <i>Stemming</i> .....	22
6. Hasil TF-IDF pada Data Latih.....	25
7. Hasil Perhitungan Probabilitas <i>Prior</i> .....	26
8. Hasil Perhitungan Probabilitas kelas.....	27
9. Tiga Dokumen Data Latih sebelum Proses Klasifikasi.....	28
10. Hasil Klasifikasi Pada Tiga Data Latih .....	29
11. Hasil <i>Confusion Matrix</i> Keseluruhan Kelas .....	30
12. <i>Confusion Matrix</i> Pada Kelas Positif .....	31
13. <i>Confusion Matrix</i> Pada Kelas Negatif.....	31
14. <i>Confusion Matrix</i> Pada Kelas Netral .....	32
15. Performa Model <i>Multinomial Naive Bayes Classifier</i> Setiap Kelas .....	32
16. Distribusi Frekuensi Data Jumlah Opini .....	33
17. Hasil Perhitungan <i>Naïve Forecasting</i> .....	36
18. Nilai <i>Root Mean Square Error</i> (RMSE) Setiap Dokumen Hasil Peramalan .....	37

## DAFTAR LAMPIRAN

	<b>Halaman</b>
1. Hasil <i>Crawling</i> Data Twitter .....	44
2. Data Hasil Pelabelan dengan <i>Lexicon Based</i> .....	45
3. Data Bersih Hasil Pra Proses Data pada 20 Oktober 2022 - 27 Oktober 2022 (Periode Satu Minggu) .....	50
5. Hasil <i>Output Term Frequency</i> Pada Data Latih .....	52
6. Hasil <i>Output Model Multinomial Naïve Bayes Classifier</i> .....	52
7. Hasil <i>Output Confusion Matrix</i> dan Evaluasi Akurasi Hasil Klasifikasi <i>Multinomial Naïve Bayes Classifier</i> .....	52
8. Hasil <i>Output Akurasi Naïve Forecasting</i> .....	53

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Subsidi bahan bakar minyak merupakan penurunan harga bahan bakar minyak yang ditetapkan oleh pemerintah. Hadirnya subsidi bahan bakar berpengaruh pada beberapa aspek ekonomi seperti harga kebutuhan pokok melalui biaya produksi, dan tarif angkutan umum. Sasaran subsidi bahan bakar minyak yaitu untuk membantu masyarakat kurang mampu, namun di atas 70% subsidi bahan bakar minyak dipakai oleh masyarakat mampu dengan mobil pribadi. Kondisi ini berakibat pada penyaluran subsidi bahan bakar minyak yang tidak tepat sasaran. Oleh karena itu melalui siaran pers yang digelar di Istana Negara, pemerintah menetapkan kebijakan pengalihan subsidi bahan bakar minyak. Kebijakan yang ditetapkan yaitu menaikkan harga bahan bakar minyak sejak 3 September 2022 untuk mengatasi permasalahan tersebut, (Pribadi 2022).

Perubahan harga bahan bakar minyak menjadi ramai dibicarakan masyarakat terutama melalui media sosial salah satunya Twitter. Media sosial Twitter menjadi incaran masyarakat Indonesia dalam menyebarkan opini, maupun memverifikasi ketepatan informasi yang sedang terjadi secara *real time*. Oleh karena itu Indonesia menempati peringkat keenam sebagai negara dengan jumlah pengguna Twitter terbesar di dunia. Jumlah pengguna Twitter aktif di Indonesia mencapai 14,8 juta pada April 2023 (Kemp, 2023). Masyarakat yang biasa saja akan perubahan harga bahan bakar minyak menyuarakan opini netral, namun ada pula masyarakat yang menyuarakan rasa kecewa atau penolakan melalui opini negatif, ataupun menyatakan dukungan melalui opini positif yang saling memancing komentar dari pengguna lain untuk berpendapat. Kumpulan opini masyarakat baik positif, negatif maupun netral mengenai kenaikan harga bahan bakar minyak dapat memberikan gambaran berupa dampak dari sisi masyarakat, namun tidak bisa dinilai secara langsung dikarenakan keberagaman opini yang ada. Oleh karena itu perlu dilakukan penelitian untuk mengetahui topik pembicaraan kenaikan harga bahan bakar minyak didominasi oleh tanggapan

masyarakat yang positif, negatif atau netral berdasarkan opini di media sosial Twitter, serta dilakukan peramalan terkait persebaran opini di waktu yang akan datang. Penelitian ini didasari oleh opini masyarakat pengguna Twitter dan melalui tahapan *machine learning* dengan metode *multinomial naïve bayes classifier* dan *naïve forecasting*.

Menurut Kurniawan (2020), *machine learning* merupakan suatu teknik matematika yang dapat membuat komputer atau mesin mempelajari suatu algoritma berdasarkan data. Hasil dari proses *machine learning* berupa pola untuk membuat model matematis yang sesuai guna memprediksi kejadian yang akan datang. Algoritma *machine learning* yang akan digunakan pada penelitian ini adalah *naïve bayes classifier* berdistribusi *multinomial*. Metode peramalan jumlah opini yang akan digunakan pada penelitian ini yaitu *naïve forecasting*. Menurut Sanubari dkk (2020), *naïve bayes classifier* adalah salah satu metode dalam pengklasifikasian dengan dasar metode probabilitas bersyarat dan metode statistik. Asumsi *naïve bayes classifier* bersifat independen, namun *naïve bayes classifier* tidak perlu memerlukan data pelatihan yang besar untuk melakukan pengklasifikasian, cocok untuk input kategorikal serta mudah diimplementasikan menjadi keunggulan dari *naïve bayes classifier*. Menurut Kirana & Bhawiyuga (2021), *naïve forecasting* merupakan metode peramalan *time series* yang berfokus pada variabel masa lalu untuk memprediksi nilai masa kini. Kinerja metode *naïve forecasting* efektif dalam melakukan peramalan walaupun metode ini termasuk dalam kategori sederhana.

Beberapa penelitian terdahulu mengenai *multinomial naïve bayes classifier* dan *naïve forecasting* diantaranya yaitu Azmi dkk (2021), telah meramalkan curah hujan di Kabupaten Banyuwangi menggunakan *naïve bayes classifier* berdistribusi multinomial dan *pursuit regression* (PPR). Hidayatillah dkk (2019) telah menerapkan algoritma *multinomial naïve bayes classifier* dalam klasifikasi level reaksi publik terhadap kandidat pemilihan umum. Kumila dkk (2019) telah membandingkan metode *moving average* dan *naïve forecasting* pada peramalan data kemiskinan. Hasil dari penelitian tersebut diperoleh bahwa metode *naïve forecasting* lebih akurat. Berdasarkan beberapa penelitian terdahulu yang telah dilakukan menunjukkan bahwa metode *multinomial*

*naïve bayes classifier* dan *naïve forecasting* dapat diaplikasikan pada berbagai kasus klasifikasi dan peramalan. Hal ini menunjukkan bahwa metode tersebut sangat bermanfaat

Pada penelitian ini, klasifikasi opini pengguna Twitter terhadap topik kenaikan harga bahan bakar minyak akan dibagi menjadi tiga kelas yaitu opini kelas positif, opini kelas negatif dan opini kelas netral. Metode yang digunakan yaitu *multinomial naïve bayes classifier* dengan melihat kinerja algoritma merujuk pada penelitian (Azmi dkk, 2021). Penggunaan metode peramalan *naïve forecasting* mengacu pada penelitian (Kumila dkk, 2019). Berdasarkan uraian latar belakang yang dipaparkan, maka judul yang akan diangkat untuk penelitian ini adalah **“Klasifikasi dan Peramalan Jumlah Opini Kenaikan Harga Bahan Bakar Minyak dengan *Multinomial Naïve Bayes Classifier* dan *Naïve Forecasting*”**.

## 1.2 Tujuan

Adapun tujuan penelitian yang ingin dicapai yaitu :

1. Mengklasifikasikan opini pengguna Twitter dalam kelas sentimen positif, negatif atau netral terhadap topik kenaikan harga bahan bakar minyak.
2. Meramalkan opini pengguna Twitter berdasarkan model *multinomial naïve bayes classifier* untuk jangka waktu satu minggu kedepan.
3. Evaluasi hasil pengklasifikasian dan peramalan opini pengguna Twitter terkait topik kenaikan harga bahan bakar minyak.

## 1.3 Ruang Lingkup

Ruang lingkup dari penelitian ini yaitu kumpulan opini pengguna Twitter di Indonesia mengenai kebijakan pemerintah dalam menaikkan harga bahan bakar minyak selama periode satu minggu dengan menggunakan metode *multinomial naïve bayes classifier* dan *naïve forecasting*.



#### **1.4 Manfaat**

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut :

1. Sebagai referensi bagi pemerintah dalam menganalisa opini masyarakat dominan di media sosial Twitter terhadap kebijakan menaikkan harga bahan bakar minyak, serta peramalan jumlah opini masyarakat diwaktu yang akan datang .
2. Metode ini dapat digunakan untuk mendapatkan respon masyarakat sebagai studi pendahuluan ketika akan dikeluarkan suatu kebijakan baru oleh pemerintah.
3. Penelitian ini diharapkan dapat menambah pengetahuan serta sebagai rujukan dalam pengembangan penelitian selanjutnya.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 *Machine Learning***

*Machine learning* merupakan suatu bidang pengetahuan dimana saat kecerdasan buatan sudah melekat dalam kehidupan sehari-hari. Pemanfaatan matematika dan statistika menjadi basis dalam pembuatan serta pengujian model *machine learning* menggunakan data dengan volume yang besar. Dalam algoritma *machine learning*, terdapat tiga kategori pembelajaran berdasarkan karakteristik data yaitu *supervised learning*, *unsupervised learning* dan *semi-supervised learning*. (Behrman, et al., 2020).

Pada penelitian ini termasuk dalam karakteristik data *supervised learning*. Menurut Elisawati dkk (2023), pada *supervised learning* diberikan label pada data untuk dipelajari oleh suatu algoritma *machine learning*. Algoritma akan mempelajari data berlabel kemudian akan ditemukan pola dan hubungan berdasarkan data berlabel. Pelatihan data dilakukan dengan memetakan *input* ke *output* sampai model mencapai tingkat akurasi yang diinginkan.

#### **2.2 *Crawling Data***

*Crawling* data merupakan suatu teknik pengumpulan data yang dilakukan dengan cara pemindaian data secara menyeluruh melalui internet. Hasil yang akan didapatkan melalui *crawling* data dapat berupa teks, audio, video maupun gambar dengan ukuran data yang besar. *Crawling* data dilakukan menggunakan *uniform resource locator* (url) dengan memasukan *application programming interface* (API) untuk memperoleh dataset sesuai kebutuhan (Tinegas, 2021).

#### **2.3 *Lexicon Based***

Menurut Han dkk (2018), *lexicon based* atau pendekatan berbasis leksikon merupakan suatu metode dalam penentuan nilai kecenderungan sentimen sebuah ulasan. Suatu ulasan akan dinilai berdasarkan akumulasi kecenderungan sentimen dari setiap *term* dengan leksikon sentimen. Leksikon sentimen merupakan kumpulan *term*

yang merujuk pada suatu kategori seperti positif, negatif dan netral. Pada pembobotan *lexicon based*, Kumpulan *term* positif diberi nilai +1, *term* negatif diberi nilai -1 dan jika suatu *term* tidak ada dalam leksikon sentimen maka nilai sentimennya adalah 0. Suatu ulasan akan dikategorikan sebagai ulasan positif jika hasil akumulasi sentimen bernilai positif, jika hasil bernilai negatif maka ulasan akan dikategorikan negatif, dan dikategorikan sebagai ulasan netral jika hasil akumulasi sentimen bernilai 0. Bentuk persamaan dalam perhitungan nilai sentimen sebagai berikut :

$$NilaiPos(u) = \sum_{i=1}^m NilaiTerm(T_i) \quad (1)$$

$$NilaiNeg(u) = \sum_{j=1}^n NilaiTerm(T_j) \quad (2)$$

$$NilaiSentimen(u) = NilaiPos(u) + NilaiNeg(u) \quad (3)$$

Keterangan :

u = Ulasan dengan m *term* positif dan n *term* negatif

## 2.4 Data Preprocessing

Sebelum dilakukan pengklasifikasian menggunakan algoritma *machine learning*, data akan melalui langkah persiapan agar proses pengklasifikasian mendapatkan hasil yang optimal. Proses tersebut dinamakan *preprocessing* data. Alur dari *preprocessing* data yaitu dimana kumpulan data mentah akan diubah menjadi data yang siap dipakai sebagai bahan pelatihan model *machine learning* (Hidayatillah dkk, 2019).

Menurut Locarso (2022), suatu data sebelum diolah menjadi sebuah model perlu dilakukan tahapan *preprocessing* untuk membersihkan data dari *noise* dan merubah data menjadi lebih terstruktur. tahapan *preprocessing* data yang digunakan dalam klasifikasi opini ialah :

1. Pembersihan data berupa penghapusan *uniform resource locator* (url), simbol-simbol, dan angka yang tidak diperlukan.
2. *Case folding* merupakan penyetaraan data dengan mengubah huruf kapital menjadi huruf kecil serta dilakukan penghapusan tanda baca.
3. Penghapusan *stopwords* atau kumpulan kata sambung yang tidak memiliki banyak arti, seperti “di”, “ke”, “dari”, “pada”, dan lainnya. Tujuan dilakukan penghapusan *stopwords* agar proses klasifikasi menjadi efektif (Purbo, 2020).

4. *Stemming* ialah tahap perubahan kata atau *term* yang berimbuhan akan diubah menjadi bentuk dasar pada tahapan ini.
5. Tokenisasi merupakan tahap dilakukan pemenggalan kalimat menjadi kumpulan kata atau *term* yang terpisah.

## 2.5 *Hold Out Validation*

Agustina dkk (2022), validasi *hold out* merupakan suatu metode dalam pemecahan data dengan membagi *dataset* menjadi data latih dan data uji. Data latih digunakan dalam pelatihan pembuatan model, sementara data uji digunakan untuk menguji seberapa baik performa dari model. Secara umum proporsi pembagian dataset menggunakan metode *hold out* ialah 80% total keseluruhan data dipakai untuk data latih sementara 20% sisanya untuk data uji.

## 2.6 **Frekuensi dan Pembobotan Istilah**

Pembobotan istilah merupakan mekanisme penilaian atas kecocokan setiap kata atau *term* pada suatu dokumen. Salah satu skema pembobotan *term* yang banyak digunakan ialah dengan pendekatan TF-IDF (Manning dkk, 2008).

### 2.6.1 *Term Frequency (TF)*

Manning dkk (2008), *term frequency* (TF) adalah statistik yang mengukur frekuensi kemunculan suatu kata atau *term* dalam dokumen. *Term frequency* (TF) dilakukan dengan cara memberikan bobot pada setiap yang tergantung pada jumlah kemunculan *term* pada dokumen, dengan tujuan untuk memetakan jumlah kemunculan *term* pada dokumen. Bentuk persamaan *term frequency* (TF) sebagai berikut :

$$tf_{(t,d)} = \begin{cases} 1 & ,if\ tf_{(t,d)} > 0 \\ 0 & ,lainnya \end{cases} \quad (4)$$

Keterangan :

$tf_{(t,d)}$  = Nilai *term frequency* dari suatu *term* t pada dokumen d

### 2.6.2 *Inverse Document Frequency (IDF)*

Manning dkk (2008), suatu kata kunci akan memiliki relevansi terhadap kumpulan dokumen jika kumpulan dokumen merujuk pada kategori tertentu. Kata kunci tersebut tersebar dalam setiap istilah pada dokumen. Pada perhitungan *term frequency*, sebuah istilah yang sering muncul akan bernilai tinggi dan istilah bernilai rendah jika jarang muncul. Untuk mengatasi ketimpangan frekuensi pada suatu dokumen, digunakan statistik tingkat dokumen yaitu *inverse document frequency (IDF)*. *Inverse document frequency (IDF)* bekerja dengan melemahkan efek istilah yang sering muncul sehingga bobotnya berkurang. Bentuk persamaan *inverse document frequency (IDF)* sebagai berikut :

$$idf_t = \log \frac{N}{df_t} \quad (5)$$

Keterangan :

$df_t$  = Nilai *document frequency* dari suatu *term t*

$idf_t$  = Nilai *inverse document frequency* dari suatu *term t*

$N$  = Jumlah dokumen pada dataset

### 2.6.3 *Term Frequency - Inverse Document Frequency (TF – IDF)*

Manning dkk (2008), skema pembobotan *term frequency – inverse document frequency* merupakan gabungan antara hasil *term frequency* dengan hasil *inverse document frequency* yang akan menghasilkan bobot komposit untuk setiap istilah dalam dokumen. Bentuk persamaan *term frequency – inverse document frequency* sebagai berikut :

$$tf - idf_{t,d} = tf_{t,d} \times idf_t \quad (6)$$

## 2.7 *Naïve Bayes Classifier*

### 2.7.1 *Definisi dan Konsep Dasar Naïve Bayes Classifier*

Effendy dkk (2022) menyatakan bahwa *naïve bayes classifier* merupakan metode klasifikasi statistik berdasarkan teorema bayes yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Metode ini bekerja dengan baik

dalam menangani *dataset* yang berukuran besar dan dapat menangani data yang tidak relevan. Terdapat beberapa distribusi model *naïve bayes classifier* yang digunakan yaitu distribusi Gaussian (normal), multinomial, dan bernoulli.

### 2.7.2 Multinomial Naïve Bayes Classifier

*Multinomial naïve bayes classifier* merupakan salah satu distribusi model *naïve bayes classifier* apabila data bersifat multi-kategori dengan varians model memiliki fitur diskret. Pada model *multinomial naïve bayes classifier* mengasumsikan setiap fitur dari atribut data memiliki nilai multinomial atau beberapa kemungkinan nilai ( Heryadi & Wahyono, 2020).

### 2.7.3 Bentuk Persamaan Multinomial Naïve Bayes Classifier

Menurut Manning dkk (2008), bentuk persamaan *multinomial naïve bayes classifier* untuk suatu dokumen  $d$  yang berada di kelas  $c$  sebagai berikut :

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c) \quad (7)$$

Keterangan :

$P(c|d)$  = Probabilitas dokumen  $d$  pada kelas  $c$

$P(c)$  = Probabilitas suatu dokumen yang berada di kelas  $c$  (*prior probability*)

$P(t_k|c)$  = Probabilitas bersyarat *term*  $t_k$  yang berada dalam dokumen pada kelas  $c$

### 2.7.4 Maximum A Posteriori (MAP)

Menurut Di dkk (2021), berdasarkan persamaan probabilitas bayes, dilakukan estimasi parameter untuk mendapatkan kemungkinan nilai parameter terbaik yang diberikan oleh data. *Maximum a posteriori* (MAP) merupakan aturan yang dapat digunakan dalam pengestimasian parameter untuk mencari kelas terbaik dalam klasifikasi *naïve bayes* dengan menggabungkan pengetahuan sebelumnya dan data yang diamati. Bentuk persamaan MAP sebagai berikut.

$$C_{map} = \arg \max_{c \in C} \hat{P}(c|d) = \arg \max_{c \in C} \hat{P}(c) \prod_{1 \leq k \leq n_d} \hat{P}(t_k|c) \quad (8)$$

Keterangan :

$C_{map}$  = Nilai *Maximum a posteriori* (MAP) dari kelas  $c$

### 2.7.5 Maximum Likelihood (MLE)

*Maximum likelihood* merupakan metode pengestimasi parameter dengan memperkirakan kemampuan probabilitas sebagai frekuensi relatif dari sebuah kejadian. Pengaplikasian *maximum likelihood* pada *naïve bayes classifier* dapat digunakan dalam perhitungan  $\hat{P}(c)$  dan  $\hat{P}(t|c)$  dengan persamaan sebagai berikut (Manning dkk, 2008).

$$\hat{P}(c) = \frac{N_c}{N} \quad (9)$$

Keterangan :

$\hat{P}(c)$  = Probabilitas suatu dokumen yang berada di kelas  $c$  (*prior probability*)

$N_c$  = Jumlah dokumen kelas  $c$

$N$  = Jumlah seluruh dokumen

$$\hat{P}(t|c) = \frac{T_{ct}}{\sum_{t' \in vT_{ct'}}} \quad (10)$$

Keterangan :

$\hat{P}(t|c)$  = Probabilitas bersyarat *term*  $t$  yang berada didokumen pada kelas  $c$

$T_{ct}$  = Jumlah kemunculan *term*  $t$  pada dokumen dengan kategori kelas  $c$

$t' \in vT_{ct'}$  = Jumlah frekuensi seluruh *term* pada kelas  $c$

Kelemahan dalam perhitungan *maximum likelihood* yaitu hasil probabilitas dokumen  $d$  pada kelas  $c$  akan bernilai 0 jika suatu *term* dalam kelas  $c$  tidak terlihat pada data latih. Oleh karena itu diaplikasikan teknik *laplace smoothing* atau *add-one* dalam mengatasi permasalahan tersebut. Bentuk persamaan 10 berubah menjadi sebagai berikut (Manning dkk, 2008).

$$\hat{P}(t|c) = \frac{T_{ct}+1}{\sum_{t' \in v(T_{ct'}+1)} = \frac{T_{ct}+1}{(\sum_{t' \in vT_{ct'}})+B'} \quad (11)$$

Keterangan :

$B'$  = Jumlah seluruh *term*

Untuk perhitungan *multinomial naïve bayes* menggunakan sistem pembobotan tf-idf, bentuk persamaan yang digunakan menjadi sebagai berikut.

$$\hat{P}(t|c) = \frac{w_{ct}+1}{(\sum_{w' \in vW_{ct'}})+B'} \quad (12)$$

Keterangan :

$W_{ct}$  = Bobot TF-IDF *term* t pada dokumen dengan kelas c

$(\sum w' \in vW_{ct'})$  = Jumlah bobot TF-IDF seluruh *term* pada kelas c

$B'$  = Jumlah IDF seluruh *term*

## 2.8 Naïve Forecasting

Kumila dkk, (2019), naïve merupakan salah satu metode peramalan sederhana dengan asumsi bahwa kejadian sebelumnya merupakan peramalan terbaik untuk hari ini. Bentuk persamaan dari metode naïve sebagai berikut:

$$Naïve = X_{t-1} \quad (13)$$

Keterangan :

$X_t$  = Data aktual pada periode t

## 2.9 Evaluasi Hasil Klasifikasi

### 2.9.1 Confusion Matrix

Menurut Locarso (2022), *confusion matrix* merupakan matriks yang berisi nilai prediksi benar dan salah. *Confusion matrix* pada penelitian ini digunakan dalam mengukur performa metode *naïve bayes classifier* dalam melakukan klasifikasi data. Dihasilkan nilai *accuracy*, *precision*, *recall* dan *f1-score* dalam mengukur performa model yang akan dievaluasi. Kharis dkk (2019), model *confusion matrix* untuk multikelas klasifikasi disajikan pada Tabel 1.

Tabel 1. *Confusion Matrix* Klasifikasi Multikelas

		Kelas Prediksi			
		Kelas 1	Kelas 2	...	Kelas k
Kelas Aktual	Kelas 1	$f_{11}$	$f_{12}$	...	$f_{1k}$
	Kelas 2	$f_{21}$	$f_{22}$	...	$f_{2k}$
	⋮	⋮	⋮	⋮	
	Kelas k	$f_{k1}$	$f_{k2}$	...	$f_{kk}$



Terdapat empat komponen dalam *confusion matrix* yaitu *true positives* (TP), *false positives* (FP), *true negatives* (TN), dan *false negatives* (FN). *True positives* (TP) merupakan jumlah prediksi benar positif yang diklasifikasikan sebagai positif, *true negatives* (TN) merupakan jumlah prediksi benar negatif yang diklasifikasikan sebagai negatif, *false positives* (FP) merupakan jumlah prediksi negatif yang diklasifikasikan sebagai positif, dan *false negatives* (FN) merupakan jumlah prediksi positif yang diklasifikasikan sebagai negatif (Mohajon, 2020).

Pada Tabel 1 nilai *true positives* (TP) pada *confusion matrix* multikelas ditunjukkan oleh diagonal kolom  $f_{11}$  hingga  $f_{kk}$ , nilai *false positives* (FP) ditunjukkan oleh jumlah nilai kolom yang bersesuaian dengan *true positives* (TP) kecuali nilai *true positives* (TP), nilai *false negatives* (FN) ditunjukkan oleh jumlah nilai baris yang bersesuaian dengan *true positives* (TP) kecuali nilai *true positives* (TP), dan nilai *true negatives* (TN) ditunjukkan oleh jumlah nilai semua kolom dan baris kecuali nilai yang telah dihitung (Jurafsky & Martin, 2023)

*Accuracy* adalah perbandingan data yang diidentifikasi benar dengan jumlah seluruh data. Bentuk persamaan yang digunakan dalam perhitungan *accuracy* sebagai berikut (Kharis dkk, 2019).

$$Accuracy = \frac{\sum_{j=1}^m f_{jj}}{\sum_{k=1}^m \sum_{j=1}^m f_{jk}} \quad (14)$$

*Precision* adalah penggambaran akurasi antara data dengan hasil prediksi, untuk menemukan peringkat yang paling relevan. Bentuk persamaan yang digunakan dalam perhitungan *precision* sebagai berikut (Cahyani, 2022).

$$Precision = \frac{true\ positives}{true\ positives + false\ positives} \quad (15)$$

*Recall* merupakan proporsi data benar positif yang diprediksi benar. Bentuk persamaan yang digunakan dalam perhitungan *recall* sebagai berikut.

$$Recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad (16)$$

*F1-score* merupakan ukuran keseimbangan antara *precision* dan *recall*. Nilai *f1-score* berada diantara 0 dan 1, semakin tinggi nilai *f1-score* maka semakin baik kualitas

klasifikasi yang dihasilkan. Bentuk persamaan yang digunakan dalam perhitungan *f1-score* sebagai berikut (Cahyani, 2022).

$$F1 - Score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \quad (17)$$

## 2.10 Evaluasi Hasil Peramalan

### 2.10.1 *Root Mean Squared Error* (RMSE)

Menurut Hodson (2022), *root mean squared error* (RMSE) merupakan akar kuadrat dari *mean squared error* untuk menghasilkan nilai yang mewakili kesalahan suatu kinerja model. Nilai *root mean squared error* (RMSE) berkisar dari nol hingga positif tak hingga, semakin besar nilai *root mean squared error* (RMSE) menunjukkan kesalahan model dalam memprediksi semakin tinggi. Adapun bentuk persamaan *root mean squared error* (RMSE) sebagai berikut.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

Keterangan :

$y_i$  = Nilai data sebenarnya

$\hat{y}_i$  = Nilai hasil peramalan

n = Jumlah observasi

## **BAB III**

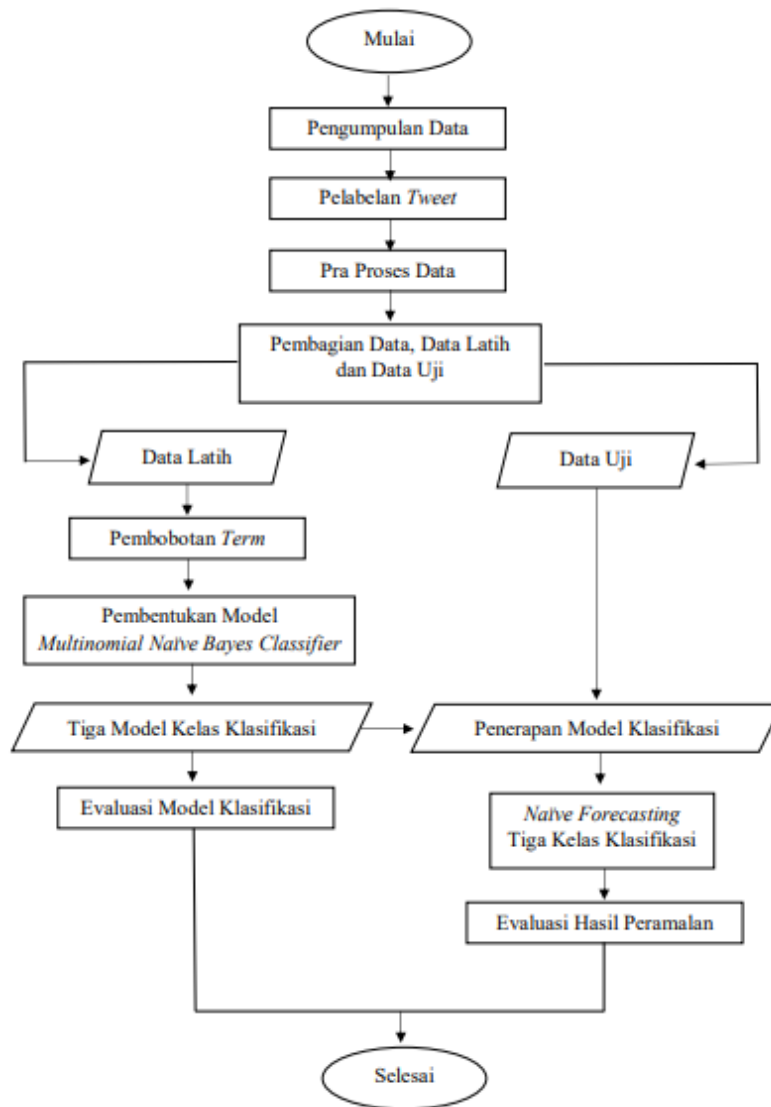
### **METODOLOGI PENELITIAN**

#### **3.1 Data**

Data yang digunakan dalam penelitian klasifikasi dan peramalan opini pengguna Twitter terkait topik kenaikan harga bahan bakar minyak menggunakan metode *multinomial naïve bayes classifier* dan *naïve forecasting* merupakan data sekunder. Data tersebut berupa *tweet* pengguna Twitter pada tanggal 20 sampai 27 Oktober 2022 sebanyak 2500 data yang mengandung kata kunci kenaikan harga bahan bakar minyak dan diambil melalui *crawling* data Twitter. Terdapat dua peubah yang ada pada penelitian ini, yaitu peubah bebas dan peubah tak bebas. Peubah bebas yaitu opini pengguna Twitter terhadap topik kenaikan harga bahan bakar minyak, sedangkan peubah tak bebas yaitu label positif, negatif, dan netral dari setiap *tweet*.

#### **3.2 Tahapan Analisis**

Rangkaian tahapan dalam penelitian klasifikasi dan peramalan opini Twitter terhadap topik kenaikan harga bahan bakar minyak menggunakan metode *multinomial naïve bayes classifier* dan *naïve forecasting* digambarkan secara skematik pada Gambar 1.



Gambar 1. Diagram Alir Tahapan Analisis

Keterangan diagram alir tahapan analisis pada Gambar 1 sebagai berikut :

1. Pengumpulan Data

Data yang digunakan yaitu *tweet* pengguna Twitter pada tanggal 20 sampai 27 Oktober 2022 yang mengandung kata kunci kenaikan harga bahan bakar minyak melalui *crawling* data Twitter.

2. Pelabelan *Tweet*  
Pada penelitian ini proses pelabelan *tweet* dilakukan dengan *lexicon based*. Pelabelan akan dibagi kedalam tiga kelas, yaitu kelas data positif, kelas data negatif dan kelas data netral.
3. Pra Proses Data  
Pada tahap ini dilakukan pra proses data dengan tahapan yaitu pembersihan data, *case folding*, penghapusan *stopwords*, dan tokenisasi. *Output* yang akan dihasilkan dari proses ini adalah data bersih yang siap dipakai untuk penelitian.
4. Pembagian Data Latih dan Data Uji  
Pada penelitian ini, pembagian data dilakukan dengan metode *hold out validation* dimana data akan dibagi kedalam proporsi 80% data latih dan 20% data uji.
5. Pembobotan Data  
Pada tahap ini dilakukan pembobotan data menggunakan *term frequency - inverse document frequency* (TF-IDF) dengan memberikan nilai berupa bobot pada suatu kata atau *term*.
6. Pembentukan Model *Multinomial Naïve Bayes Classifier*  
Data latih digunakan untuk pembuatan model *multinomial naïve bayes classifier*. Tahapan dalam pembentukan model yaitu dilakukan perhitungan untuk mencari probabilitas kelas serta probabilitas kemunculan kata atau *term*. Hasil dari pembentukan model *multinomial naïve bayes classifier* ialah tiga model kelas klasifikasi.
7. *Naïve Forecasting* Tiga Kelas Klasifikasi  
Pada tahap ini, data uji diterapkan model klasifikasi untuk melakukan peramalan. Peramalan tiga kelas klasifikasi menggunakan *naïve forecasting* akan menghasilkan peramalan untuk opini positif, negatif dan netral.
8. Evaluasi Model Klasifikasi  
Evaluasi model klasifikasi dilakukan untuk mengukur performa model *multinomial naïve bayes classifier*. Performa dari model akan dianalisis menggunakan *confusion matrix*. Ukuran akurasi yang dipakai yaitu *accuracy*, *precision*, *recall* dan *f1-score*.

9. Evaluasi Hasil Peramalan

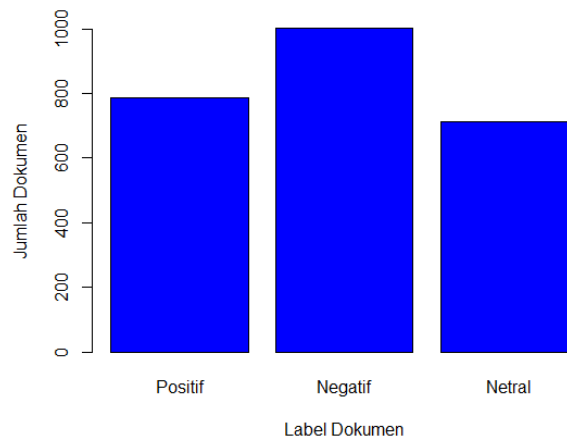
Pada tahap ini dilakukan perhitungan nilai error pada hasil peramalan menggunakan *root mean squared error* (RMSE). Hasil perhitungan ini akan menjadi tolak ukur keakuratan model *naïve forecasting* dalam melakukan peramalan.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Pelabelan *Tweet*

Pelabelan *tweet* hasil *crawling* data dilakukan dengan *lexicon based*. Proses pelabelan bertujuan untuk memberikan nilai sentimen berupa kelas positif, negatif atau netral terhadap kumpulan *tweet*. Hasil pelabelan *tweet* disajikan pada Gambar 2.



Gambar 2. Grafik Hasil Pelabelan *Tweet*

Gambar 2 menunjukkan frekuensi *tweet* hasil pelabelan untuk setiap kelas. *Tweet* direpresentasikan sebagai dokumen, sementara label dokumen merupakan kelas sentimen pada setiap dokumen. Hasil pelabelan dokumen diperoleh bahwa label dokumen terbanyak ialah kelas negatif sebanyak 1001, lalu kelas positif sebanyak 786 dan kelas netral sebanyak 713.

#### 4.2 Pra Proses Data

Data setelah melalui tahap pelabelan akan dilakukan pra proses data. Tahapan dalam pra proses data yaitu pembersihan data, *case folding*, penghapusan *stopwords*, *stemming* dan tokenisasi. Tahap pembersihan data dilakukan setelah diperoleh data berlabel untuk menunjang tahapan pra proses yang akan dilakukan selanjutnya.

#### 4.2.1 Pembersihan Data

Pembersihan data yang dilakukan berupa penghapusan *uniform resource locator (url)*, simbol-simbol dan angka. Hasil dari pembersihan data untuk tiga dokumen disajikan pada Tabel 2.

Tabel 2. Proses Pembersihan Data

Tanggal	Label	Contoh Hasil Pembersihan Data	
		Dokumen Sebelum	Dokumen Sesudah
27/10/2022	Positif	Polda Jatim bagi 250 Paket Sembako kepada Nelayan Kalisari Surabaya yang terdampak kenaikan BBM di Kelurahan Kalisari Surabaya pada hari Kamis, 27 Oktober 2022. <a href="https://t.co/AofjhGfZAe">https://t.co/AofjhGfZAe</a>	Polda Jatim bagi Paket Sembako kepada Nelayan Kalisari Surabaya yang terdampak kenaikan BBM di Kelurahan Kalisari Surabaya pada hari Kamis Oktober
27/10/2022	Negatif	Kenaikan bbm ini bikin ongkir paketan naik juga [macil; marah kecil]. <a href="https://t.co/flYXzX3fAh">https://t.co/flYXzX3fAh</a>	Kenaikan bbm ini bikin ongkir paketan naik juga macil marah kecil
27/10/2022	Netral	Soal Bantalan Sosial Kenaikan BBM, Pemprov NTB Belum Patuhi Inpres <a href="https://t.co/ryEbQry45U">https://t.co/ryEbQry45U</a>	Soal Bantalan Sosial Kenaikan BBM Pemprov NTB Belum Patuhi Inpres



Tabel 2 menunjukkan perubahan dokumen setelah melalui tahap pembersihan data. Perubahan yang terjadi berupa penghapusan *uniform resource locator (url)*, tanda baca dan simbol pada ketiga dokumen serta hilangnya angka pada dokumen pertama.

#### 4.2.2 Case Folding

*Case folding* merupakan tahapan dalam pra proses data yang dilakukan setelah pembersihan data. *Case folding* bertujuan untuk mengubah huruf kapital menjadi huruf kecil agar data menjadi seragam. Berikut tiga dokumen hasil dari *case folding* disajikan pada Tabel 3.

Tabel 3. Proses *Case Folding*

Tanggal	Label	Contoh Hasil Proses <i>Case Folding</i>	
		Dokumen Sebelum	Dokumen Sesudah
27/10/2022	Positif	Polda Jatim bagi Paket Sembako kepada Nelayan Kalisari Surabaya yang terdampak kenaikan BBM di Kelurahan Kalisari Surabaya pada hari Kamis Oktober	polda jatim bagi paket sembako kepada nelayan kalisari surabaya yang terdampak kenaikan bbm di kelurahan kalisari surabaya pada hari kamis oktober
27/10/2022	Negatif	Kenaikan bbm ini bikin ongkir paketan naik juga macil marah kecil	kenaikan bbm ini bikin ongkir paketan naik juga macil marah kecil
27/10/2022	Netral	Soal Bantalan Sosial Kenaikan BBM Pemprov NTB Belum Patuhi Inpres	soal bantalan sosial kenaikan bbm pemprov ntb belum patuhi inpres

Pada Tabel 3 menunjukkan hasil dari proses *case folding*. Hasil proses *case folding* berupa perubahan huruf dari beberapa *term* pada ketiga dokumen. Terdapat beberapa *term* pada ketiga dokumen yang ditandai dengan warna merah, hal ini menunjukkan adanya perubahan setelah proses *case folding*.

#### 4.2.3 Penghapusan *Stopwords*

Tahap penghapusan *stopwords* dilakukan setelah tahap *case folding* selesai. Penghapusan *stopwords* dilakukan dengan menghilangkan kosakata yang terdapat dalam kamus *stopwords*. Berikut tiga dokumen hasil dari penghapusan *stopwords* disajikan pada Tabel 4.

Tabel 4. Proses Penghapusan *Stopwords*

Tanggal	Label	Contoh Hasil Penghapusan <i>Stopwords</i>	
		Dokumen Sebelum	Dokumen Sesudah
27/10/2022	Positif	polda jatim bagi paket sembako kepada nelayan kalisari surabaya yang terdampak kenaikan bbm di kelurahan kalisari surabaya pada hari Kamis oktober	polda jatim paket sembako nelayan kalisari surabaya terdampak kenaikan bbm kelurahan kalisari surabaya Kamis oktober
27/10/2022	Negatif	kenaikan bbm ini bikin ongkir paketan naik juga macil marah kecil	kenaikan bbm bikin ongkir paketan naik macil marah
27/10/2022	Netral	soal bantalan sosial kenaikan bbm pemprov ntb belum patuhi inpres	bantalan sosial kenaikan bbm pemprov ntb patuhi inpres

Berdasarkan Tabel 4 menunjukkan perubahan pada ketiga dokumen hasil dari proses penghapusan *stopwords*. Perubahan yang terjadi berupa penghapusan kosakata

yang termasuk dalam kamus *stopwords*. Penghapusan dilakukan karena kosakata dalam kamus *stopwords* dianggap tidak memiliki banyak arti, sehingga tidak berpengaruh dalam proses klasifikasi.

#### 4.2.4 Stemming

Data setelah dilakukan penghapusan *stopwords* akan melalui tahap *stemming*. Pada tahap *stemming* akan dilakukan perubahan kata atau *term* berimbuhan menjadi bentuk dasar. Hasil *stemming* untuk tiga dokumen disajikan pada Tabel 5.

Tabel 5. Proses *Stemming*

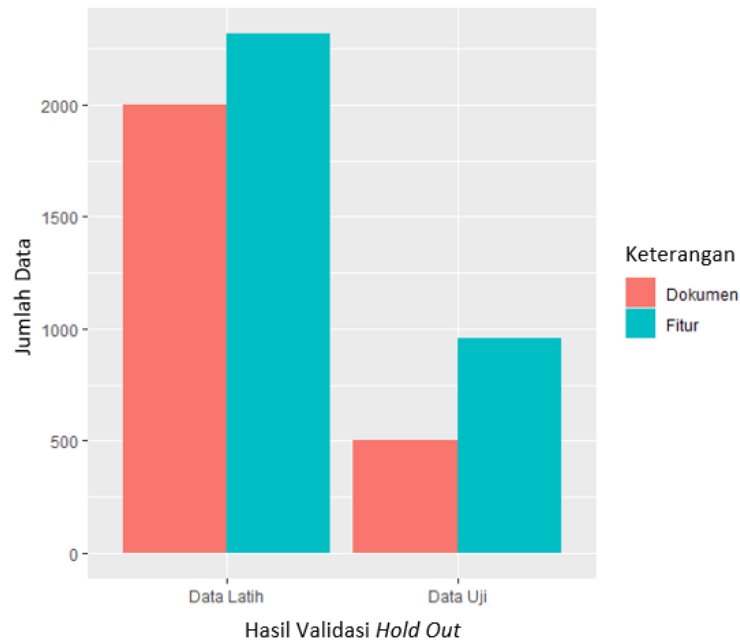
Tanggal	Label	Contoh Hasil Proses <i>Stemming</i>	
		Dokumen Sebelum	Dokumen Sesudah
27/10/2022	Positif	polda jatim paket sembako nelayan kalisari surabaya terdampak kenaikan bbm kelurahan kalisari surabaya Kamis oktober	polda jatim paket sembako nelayan kalisari surabaya <b>dampak naik</b> bbm <b>lurah</b> kalisari surabaya Kamis oktober
27/10/2022	Negatif	kenaikan bbm bikin ongkir paketan naik macil marah	<b>naik</b> bbm bikin ongkir <b>paket</b> naik macil marah
27/10/2022	Netral	bantalan sosial kenaikan bbm pemprov ntb patuhi inpres	<b>bantal</b> sosial <b>naik</b> bbm pemprov ntb patuh inpres

Pada Tabel 5 ditunjukkan perubahan terhadap kata atau *term* berimbuhan menjadi dasar setelah melalui tahap *stemming*. Perubahan kata atau *term* berimbuhan setelah melalui tahap *stemming* ditandai dengan warna merah. Data setelah melalui tahap *stemming* akan dipecah menjadi kumpulan *term* terpisah pada tahap tokenisasi.



### 4.3 Pembagian Data

Pembagian data dilakukan menggunakan validasi *hold out*. Data bersih akan dipecah menjadi data latih dan data uji dengan proporsi pembagian data 80:20. Berikut hasil dari pembagian data untuk 2500 dokumen disajikan pada Gambar 4.



Gambar 4. Grafik Hasil Pembagian Data

Gambar 4 menunjukkan grafik jumlah dokumen setelah dibagi menjadi data uji dan data latih. Jumlah data latih sebanyak 2001 dokumen dengan 2317 fitur atau *term*, sementara jumlah data uji sebanyak 499 dokumen dengan 957 fitur atau *term*

### 4.4 Pembentukan *Term Frequency - Inverse Document Frequency (TF-IDF)*

Pembobotan data dilakukan setelah data dibagi menjadi data latih dan data uji. Pembobotan data hanya dilakukan terhadap data latih menggunakan *term frequency – inverse document frequency (tf-idf)*. Hasil dari pembobotan data disajikan pada Tabel 6.

Tabel 6. Hasil TF-IDF pada Data Latih

Dokumen	Term								
	polda	jatim	paket	sembako	nelayan	...	dampak	naik	bbm
1	0,071217	0,072076	0,07088	0,064994	0,088541	...	0,027915	0,033141	0,000364
2	0	0	0	0	0	...	0	0	0,000228
3	0,071217	0,072076	0,07088	0,064994	0,088541	...	0,027915	0,033141	0,000364
4	0,071217	0,072076	0,07088	0,064994	0,088541	...	0,027915	0,033141	0,000364
...	...	...	...	...	...	...	...	...	...
1998	0	0	0	0	0	...	0	0	0,000273
1999	0	0	0	0	0	...	0,017447	0,020713	0,000228
2000	0	0	0	0	0	...	0	0	0,00091
2001	0	0	0	0	0	...	0	0	0,000321

Berdasarkan Tabel 6 menunjukkan seluruh *term* pada data latih telah memiliki bobot sesuai dengan frekuensi kemunculan *term* pada dokumen. Hasil pembobotan *term frequency – inverse document frequency (tf-idf)* bernilai rendah ketika *term* muncul di hampir semua dokumen dan bernilai tinggi ketika frekuensi kemunculan relatif sedikit.

#### 4.5 Model *Multinomial Naïve Bayes Classifier*

Model *multinomial naïve bayes classifier* merupakan model probabilistik. Model *multinomial naïve bayes* yang akan diperoleh terdiri dari probabilitas *prior* dan probabilitas kelas. Probabilitas *prior* dan probabilitas kelas merupakan parameter berbeda yang akan diestimasi dari data latih.

##### 4.5.1 Perhitungan Probabilitas *Prior*

Probabilitas *prior* merupakan kemungkinan dokumen termasuk dalam kelas tertentu. Probabilitas *prior* yang akan dihitung yaitu untuk kelas positif, negatif dan netral. Probabilitas *prior* dihitung menggunakan persamaan 9 dengan hasil disajikan pada Tabel 7.

Tabel 7. Hasil Perhitungan Probabilitas *Prior*

Label	Probabilitas <i>Prior</i>
	$\hat{P}(c) = \frac{N_c}{N}$
Positif	$\hat{P}(positif) = \frac{629}{2001}$ = 0,3143
Negatif	$\hat{P}(negatif) = \frac{801}{2001}$ = 0,4003
Netral	$\hat{P}(netral) = \frac{571}{2001}$ = 0,285357

Tabel 7 menunjukkan hasil perhitungan probabilitas *prior* untuk ketiga kelas. Kelas negatif memiliki hasil perhitungan tertinggi sebesar 0,4003 ,sementara kelas netral memiliki hasil perhitungan terendah sebesar 0,285357.

#### 4.5.2 Perhitungan Probabilitas Kelas

Probabilitas kelas dihitung setelah perhitungan probabilitas *prior* selesai. Probabilitas kelas merupakan probabilitas bersyarat suatu *term* termasuk dalam kelas positif, negatif dan netral. Probabilitas kelas dihitung menggunakan persamaan 12 dengan hasil disajikan pada Tabel 8.

Tabel 8. Hasil Perhitungan Probabilitas kelas

<i>Term</i>	<b>Probabilitas Kelas</b>		
	Negatif	Netral	Positif
polda	0,000293	0,002936	0,001943
jatim	0,000282	0,00285	0,001962
paket	0,000351	0,002866	0,001935
sembako	0,000991	0,002803	0,00202
nelayan	0,000328	0,000369	0,002821
kalisari	0,000281	0,000405	0,004717
surabaya	0,00031	0,002869	0,003592
dampak	0,001252	0,003849	0,003238
naik	0,001354	0,004427	0,002886
bbm	0,000391	0,00046	0,000432
lurah	0,000306	0,000408	0,002503
kamis	0,000281	0,000372	0,002545
oktober	0,00093	0,002851	0,001969
⋮	⋮	⋮	⋮
warna	0,000282	0,000339	0,000451
saksi	0,000282	0,000339	0,000388

Pada Tabel 8 menunjukkan hasil perhitungan probabilitas kelas pada data latih. Hasil probabilitas *prior* dan probabilitas kelas akan diaplikasikan sebagai model probabilistik *multinomial naïve bayes* dalam melakukan klasifikasi.

#### 4.6 Proses Klasifikasi Model *Multinomial Naïve Bayes*

Proses klasifikasi dengan model probabilistik *multinomial naïve bayes* dilakukan terhadap data data latih sejumlah 2001 dokumen. Proses ini dilakukan untuk menentukan keakuratan model dalam melakukan klasifikasi. Berikut merupakan tiga dokumen data latih disajikan pada Tabel 9.



Tabel 9. Tiga Dokumen Data Latih sebelum Proses Klasifikasi

	<b>Dokumen</b>	<b>Label</b>	<b>Klasifikasi</b>
1	‘polda’, ‘jatim’, ‘paket’, ‘sembako’, ‘nelayani’, ‘kalisari’, ‘surabaya’, ‘dampak’, ‘naik’, ‘bbm’, ‘lurah’, ‘kalisari’, ‘surabaya’, ‘kamis’, ‘oktober’	Positif	?
2	‘efek’, ‘naik’, ‘bbm’	Netral	?
3	‘laju’, ‘inflasi’, ‘ri’, ‘relatif’, ‘kendali’, ‘bbm’, ‘alami’	Negatif	?

Tabel 9 menunjukkan tiga dokumen yang akan melalui proses klasifikasi. Pada proses ini, dihitung probabilitas setiap kategori kelas dan dicari probabilitas tertingginya menggunakan persamaan 8. Hasil perhitungan yang didapatkan disajikan pada Tabel 10 .

Tabel 10. Hasil Klasifikasi Pada Tiga Data Latih

Dokumen	$\hat{P}(c d)$	Klasifikasi
1	<b>Positif</b> $\hat{P}(c_{positif} d_1) = 1,91886e^{-40}$	Positif
	Negatif $\hat{P}(c_{negatif} d_1) = 1,41293e^{-51}$	
	Netral $\hat{P}(c_{netral} d_1) = 3,23138e^{-44}$	
2	Positif $\hat{P}(c_{positif} d_2) = 1,40131e^{-10}$	Netral
	Negatif $\hat{P}(c_{negatif} d_2) = 1,78656e^{-10}$	
	<b>Netral</b> $\hat{P}(c_{netral} d_2) = 7,5791e^{-10}$	
3	Positif $\hat{P}(c_{positif} d_3) = 6,72351e^{-24}$	Negatif
	<b>Negatif</b> $\hat{P}(c_{negatif} d_3) = 5,19979e^{-20}$	
	Netral $\hat{P}(c_{netral} d_3) = 8,04461e^{-21}$	

Tabel 10 menunjukkan proses klasifikasi terhadap tiga data latih. Proses klasifikasi dilakukan dengan memilih nilai probabilitas kelas dokumen terbesar sebagai hasil klasifikasi. Hasil prediksi menunjukkan kesesuaian antara label data aktual dengan label hasil klasifikasi terpilih. Probabilitas kelas dokumen terpilih ditandai dengan warna merah pada penulisan.

#### 4.7 Evaluasi Model *Multinomial Naïve Bayes Classifier*

Evaluasi hasil klasifikasi model *multinomial naïve bayes classifier* dibentuk dalam *confusion matrix*. Model *confusion matrix* yang digunakan adalah *matrix*  $3 \times 3$  yang terdiri dari kelas prediksi dan kelas aktual. Hasil dari model *confusion matrix* keseluruhan dokumen terdapat pada Tabel 11.

Tabel 11. Hasil *Confusion Matrix* Keseluruhan Kelas

Kelas Prediksi	Kelas Aktual		
	Negatif	Netral	Positif
Negatif	775	16	10
Netral	80	475	16
Positif	36	12	581

Tabel 11 menyajikan hasil klasifikasi dalam *confusion matrix* untuk keseluruhan kelas. *Confusion matrix* berisi perbandingan antara hasil prediksi oleh model dengan kelas data sebenarnya. Hasil prediksi benar atau terklasifikasi sama dengan kelas data sebenarnya ditandai dengan warna biru, sementara untuk kolom lainnya menunjukkan kesalahan hasil prediksi yang tidak sesuai dengan kelas data sebenarnya.

*Confusion matrix* digunakan dalam perhitungan performa model *multinomial naïve bayes classifier*. Ukuran akurasi yang digunakan yaitu *accuracy*, *precision*, *recall* dan *f1-score*. *Accuracy* dihitung untuk keseluruhan kelas, sementara *precision*, *recall* dan *f1-score* dihitung untuk masing-masing kelas. Perhitungan *accuracy* model *multinomial naïve bayes classifier* menggunakan persamaan 14 dengan hasil perhitungan *accuracy* sebagai berikut:

$$Accuracy = \frac{\sum_{j=1}^m f_{jj}}{\sum_{k=1}^m \sum_{j=1}^m f_{jk}}$$

$$Accuracy = \frac{775 + 475 + 581}{775 + 16 + 10 + 80 + 475 + 16 + 36 + 12 + 581}$$

$$Accuracy = 0,915042 \approx 92\%$$

Pada perhitungan *accuracy* diperoleh hasil sebesar 92%. Hasil *accuracy* yang diperoleh menunjukkan ketepatan akurasi yang dihasilkan model *multinomial naïve bayes classifier* dalam melakukan klasifikasi sebesar 92%.

#### 4.7.1 Evaluasi Kelas Klasifikasi

Hasil *confusion matrix* keseluruhan kelas akan dibagi berdasarkan masing-masing kelas. Pembagian *confusion matrix* bertujuan untuk memperjelas komposisi *true positives*, *false positives*, *true negatives* dan *false negatives*, sehingga dapat dilakukan perhitungan *precision*, *recall* dan *f1-score* untuk setiap kelas. Hasil pembagian *confusion matrix* untuk setiap kelas disajikan pada tabel berikut:

Tabel 12. *Confusion Matrix* Pada Kelas Positif

Kelas Prediksi	Kelas Aktual		
	Negatif	Netral	Positif
Negatif	775	16	10
Netral	80	475	16
Positif	36	12	581

Tabel 13. *Confusion Matrix* Pada Kelas Negatif

Kelas Prediksi	Kelas Aktual		
	Negatif	Netral	Positif
Negatif	775	16	10
Netral	80	475	16
Positif	36	12	581

Tabel 14. *Confusion Matrix* Pada Kelas Netral

Kelas Prediksi	Kelas Aktual		
	Negatif	Netral	Positif
Negatif	775	16	10
Netral	80	475	16
Positif	36	12	581

Berdasarkan ketiga tabel diatas hasil pemecahan *confusion matrix* pada setiap kelas berpusat pada *true positives* (TP) ditunjukkan dengan warna biru. Kolom nilai *true negatives* (TN) ditunjukkan dengan warna abu-abu, sementara *false positives* ditunjukkan dengan warna jingga dan *false negatives* ditunjukkan dengan warna hijau. Untuk hasil perhitungan performa model dengan ukuran *precision*, *recall* dan *f1-score* pada setiap kelas disajikan pada Tabel 15.

Tabel 15. Performa Model *Multinomial Naive Bayes Classifier* Setiap Kelas

Ukuran	Kelas		
	Negatif	Netral	Positif
<i>Precision</i>	0,9675	0,8319	0,9237
<i>Recall</i>	0,8698	0,9443	0,9572
<i>F1-score</i>	0,9161	0,8845	0,9401

Berdasarkan Tabel 15 menunjukkan hasil pengukuran performa model *multinomial naïve bayes classifier* pada setiap kelas. Hasil perhitungan *precision* menunjukkan tingkat kebaikan model dapat melakukan klasifikasi dengan benar, sementara *recall* menunjukkan tingkat kebaikan model dapat melakukan klasifikasi dengan benar untuk kelas positif, dan *f1-score* menunjukkan keseimbangan antara *recall* dan *precision*.

#### 4.8 Peramalan Tiga Kelas Klasifikasi Dengan *Naïve Forecasting*

Data yang akan digunakan dalam peramalan yaitu data uji sejumlah 499 dokumen. Peramalan akan dibagi menjadi tiga kelas yaitu peramalan opini kelas positif, kelas negatif dan kelas netral. Tahapan dalam peramalan jumlah opini dengan *naïve forecasting* yaitu tabulasi data, proses peramalan dan evaluasi akurasi hasil peramalan.

##### 4.8.1 Tabulasi Data Jumlah Opini

Data uji akan melalui proses tabulasi data sebelum dilakukan peramalan. Tabulasi data akan dilakukan dengan pembuatan tabel frekuensi. Hasil dari tabulasi data disajikan pada Tabel 16.

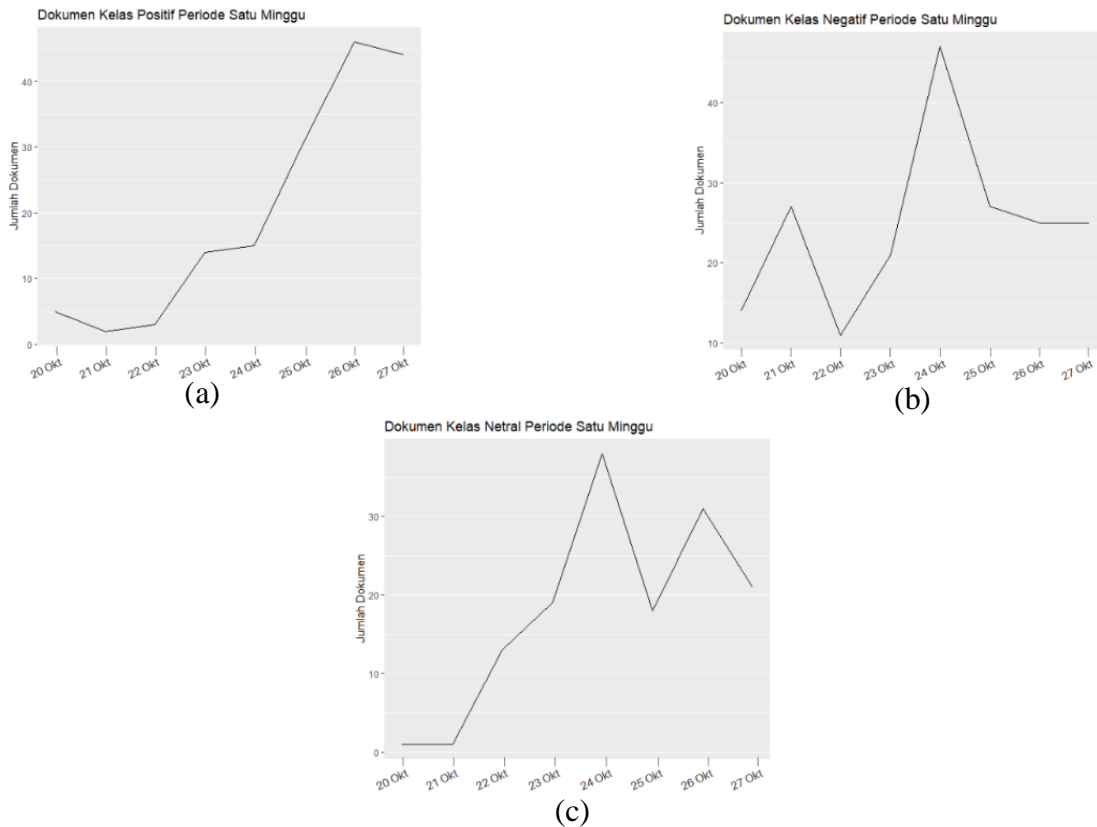
Tabel 16. Distribusi Frekuensi Data Jumlah Opini

Tanggal	Jumlah Dokumen		
	Positif	Negatif	Netral
20/10/2022	5	14	1
21/10/2022	2	27	1
22/10/2022	3	11	13
23/10/2022	14	21	19
24/10/2022	15	47	38
25/10/2022	31	27	18
26/10/2022	46	25	31
27/10/2022	44	25	21

Pada Tabel 16 ditunjukkan jumlah dokumen dari setiap kelas hasil dari tabulasi data. Hasil tabulasi data menunjukkan frekuensi kemunculan dokumen yang termasuk pada kelas tertentu. Data setelah ditabulasi akan melalui proses peramalan dengan *naive forecasting*. Peramalan akan dilakukan pada masing-masing dokumen.

#### 4.8.2 Proses Peramalan *Naïve Forecasting*

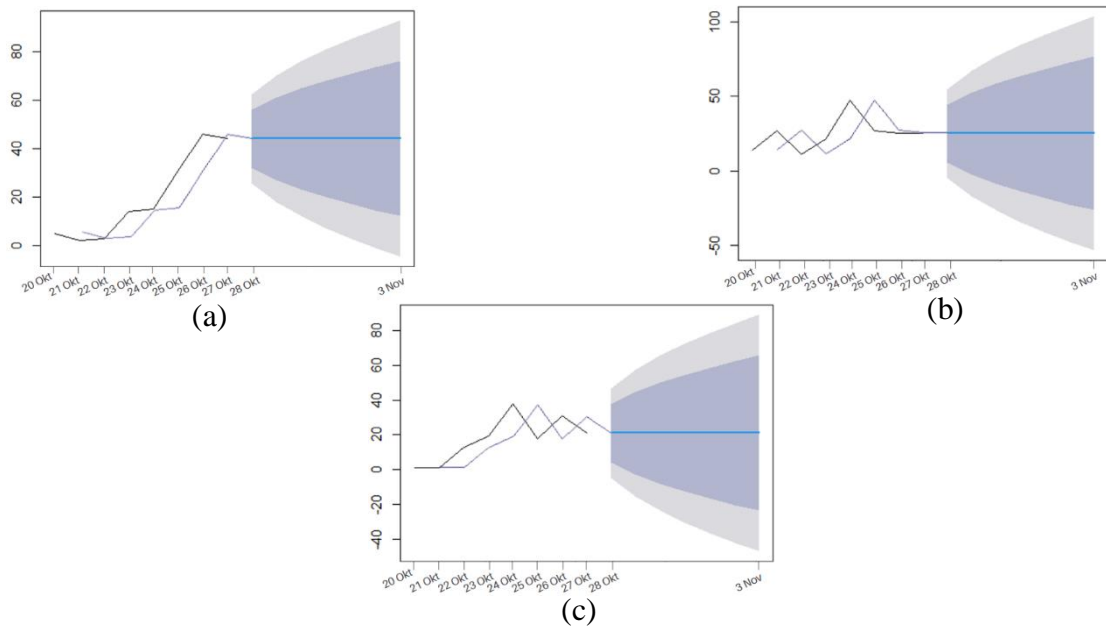
*Naïve forecasting* merupakan salah satu metode peramalan runtun waktu. Data yang telah ditabulasi akan diubah kedalam data runtun waktu sebelum dilakukan peramalan. Data hasil klasifikasi setelah menjadi data runtun waktu disajikan pada Gambar 5.



Gambar 5. Grafik Pola Data Runtun Waktu Periode Satu minggu

Gambar 5 menunjukkan grafik pola data runtun waktu periode waktu satu minggu. Pada grafik a) menunjukkan pola data *trend*, sementara grafik b) dan c) tidak menunjukkan pola tertentu. Frekuensi kemunculan dokumen dari setiap kelas selama periode waktu satu minggu memiliki kecenderungan meningkat.

Data setelah menjadi data runtun waktu akan melalui tahap peramalan. Proses peramalan dihitung menggunakan persamaan 13. Hasil peramalan menggunakan *naïve forecasting* disajikan pada gambar berikut:



Gambar 6. Grafik Hasil Peramalan

Berdasarkan Gambar 6 menunjukkan grafik hasil *naïve forecasting*. Garis berwarna hitam menunjukkan data aktual, sementara garis berwarna biru menunjukkan data hasil peramalan. Grafik a) merupakan hasil peramalan kelas positif, b) merupakan hasil peramalan kelas negatif dan c) merupakan hasil peramalan kelas netral. Berdasarkan ketiga grafik hasil peramalan tersebut, metode *naïve forecasting* tidak memperhitungkan *trend*. *Naïve forecasting* hanya mengambil nilai pada siklus terakhir. Untuk hasil perhitungan *naïve forecasting* menggunakan persamaan 13 disajikan pada Tabel 17.



Tabel 17. Hasil Perhitungan *Naïve Forecasting*

Tanggal	Data Ramal ( $X_{t-1}$ )			Data Aktual ( $X_t$ )		
	Positif	Negatif	Netral	Positif	Negatif	Netral
20/10/2022	NA	NA	NA	5	14	1
21/10/2022	5	14	1	2	27	1
22/10/2022	2	27	1	3	11	13
23/10/2022	3	11	13	14	21	19
24/10/2022	14	21	19	15	47	38
25/10/2022	15	47	38	31	27	18
26/10/2022	31	27	18	46	25	31
27/10/2022	46	25	31	44	25	21
28/10/2022	44	25	21			
29/10/2022	44	25	21			
30/10/2022	44	25	21			
31/10/2022	44	25	21			
01/11/2022	44	25	21			
02/11/2022	44	25	21			
03/11/2022	44	25	21			

Tabel 17 menunjukkan hasil perhitungan *naïve forecasting*. Peramalan tiga kelas klasifikasi dengan *naïve forecasting* pada waktu  $X_t$  dihitung berdasarkan jumlah opini pada waktu  $X_{t-1}$ .

#### 4.9 Evaluasi Hasil peramalan

Evaluasi hasil peramalan dihitung menggunakan *root mean square error* (RMSE) dengan persamaan 13. Hasil perhitungan *root mean square error* menjadi penilaian keakuratan model *naïve forecasting* dalam melakukan peramalan. Hasil Perhitungan *root mean square error* disajikan pada Tabel 18.

Tabel 18. Nilai *Root Mean Square Error* (RMSE) Setiap Dokumen Hasil Peramalan

<b>Kelas Dokumen</b>	<b>Nilai <i>root mean square error</i> (RMSE)</b>
Positif	8,958236
Negatif	14,86607
Netral	11,45098

Tabel 18 menunjukkan akurasi hasil *naïve forecasting* menggunakan ukuran *root mean square error* (RMSE). Hasil yang diperoleh menunjukkan besaran *error* yang dihasilkan dari proses peramalan. Semakin kecil nilai *root mean square error* (RMSE) menunjukkan bahwa akurasi peramalan semakin baik.

## BAB V

### PENUTUP

#### 5.1 Kesimpulan

Berdasarkan hasil klasifikasi menggunakan *multinomial naïve bayes classifier* dan peramalan dengan *naïve forecasting* yang telah dilakukan, terdapat beberapa hal yang dihasilkan yaitu :

1. Berdasarkan pemodelan *multinomial naïve bayes classifier* diperoleh tiga klasifikasi yaitu kelas positif, kelas negatif dan kelas netral. Hasil klasifikasi yang diperoleh sebanyak 775 dokumen diklasifikasikan sebagai opini negatif, 475 dokumen diklasifikasikan sebagai opini netral, dan 581 dokumen diklasifikasikan sebagai opini positif. Hal ini menunjukkan bahwa opini negatif lebih banyak dibandingkan opini positif dan netral.
2. Hasil peramalan tiga kelas klasifikasi selama periode satu minggu sejumlah 44 opini kelas positif, 25 opini kelas negatif dan 21 opini kelas netral. Hal ini menunjukkan bahwa bertambahnya jumlah opini positif selama periode satu minggu kedepan akan lebih banyak dibandingkan dengan opini negatif dan opini netral.
3. Akurasi model *multinomial naïve bayes classifier* untuk kelas positif sebesar 94% dengan tingkat kesalahan model *naïve forecasting* pada kelas positif sebesar 9%. Akurasi model *multinomial naïve bayes classifier* untuk kelas negatif sebesar 92% dengan tingkat kesalahan model *naïve forecasting* pada kelas negatif sebesar 15%. Akurasi model *multinomial naïve bayes classifier* untuk kelas netral sebesar 88% dengan tingkat kesalahan model *naïve forecasting* pada kelas netral sebesar 11%.

## 5.2 Saran

Adapun saran yang dapat dikembangkan berdasarkan penelitian ini diantaranya yaitu :

1. Pelabelan data dilakukan dengan lebih baik serta dilakukan pelabelan ulang agar didapatkan data berkualitas tinggi.
2. Jangkauan waktu pengambilan data diperpanjang sehingga jangka waktu untuk peramalan lebih panjang dan pola data runtun waktu terdeteksi jelas.
3. Metode peramalan ditambah dengan metode peramalan yang lain sehingga dapat dilakukan perbandingan dan menemukan metode peramalan terbaik.

## DAFTAR PUSTAKA

- Agustina, N., Citra, D. H., Purnama, W., Nisa, C., Kurnia, A. R. 2022. The Implementation of Naïve Bayes Algorithm for Sentiment Analysis of Shopee Reviews On Google Play Store. *Malcom: Indonesian Journal of Machine Learning and Computer Science*, **2(1)** : 47-54.
- Azmi, A. U., Hadi, A. F., Dewi, Y. S., Tirta, I. M., Ubaidillah, F., Anggraeni, D. 2021. Naïve Bayes Classifier (NBC) for Forecasting Rainfall in Banyuwangi District Using Projection Pursuit Regression. *Proceedings of the International Conference on Mathematics, Geometry, Statistics, and Computation (IC-MaGeStiC 2021)*. **96** : 190-195. <https://doi.org/10.2991/acsr.k.220202.036>
- Behrman, E., Bhattacharyya, S., Chakraborti, S., De, S., Mani, A., Pan, I. 2020. *Quantum Machine Learning*. Jerman: De Gruyter.
- Cahyani, D. E. 2022. Penerapan Machine Learning Untuk Prediksi Penyakit Stroke. *Jurnal Kajian Matematika dan Aplikasinya*. **3(1)** : 15-22. <http://dx.doi.org/10.17977/um055v3i1p15-22>
- Di, R., Wang, P., He, C., Guo, Z. 2021. Constrained Adjusted Maximum a Posteriori Estimation of Bayesian Network Parameters. *Entropy*. **23(10)** : 1283. <https://doi.org/10.3390/e23101283>
- Effendy, M. H., Anggraeni, D., Dewi, Y. S., Hadi, A. F. 2022. Classification of Bank Deposit Using Naïve Bayes Classifier (NBC) and K-Nearest Neighbor (K-NN). *Proceedings of the International Conference on Mathematics, Geometry, Statistics, and Computation (IC-MaGeStiC 2021)*. **96** : 163-166. <https://dx.doi.org/10.2991/acsr.k.220202.031>
- Elisawati., Permana, A. A., Rahmaddeni., Santoso, L. W., Wahidin, A. J., Wahyuddin, S., Wardhani, A. K., Wibowo, G. W. N., Wijayanti, R. R., Yanto, A. A., Yuliasuti, G. E. 2023. *Machine Learning*. Padang : PT Global Eksekutif Teknologi.
- Han, H., Zhang, Y., Zhang, J., Yang, J., Zou, X. 2018. Improving the performance of lexicon-based review sentiment analysis method by reducing additional introduced sentiment bias. *PLoS One*. **13(8)** : e0202523. <https://doi.org/10.1371/journal.pone.0202523>
- Heryadi, Y., Wahyono, T. 2020. *Machine Learning (Konsep dan Implementasi)*. Yogyakarta: Gava Media.

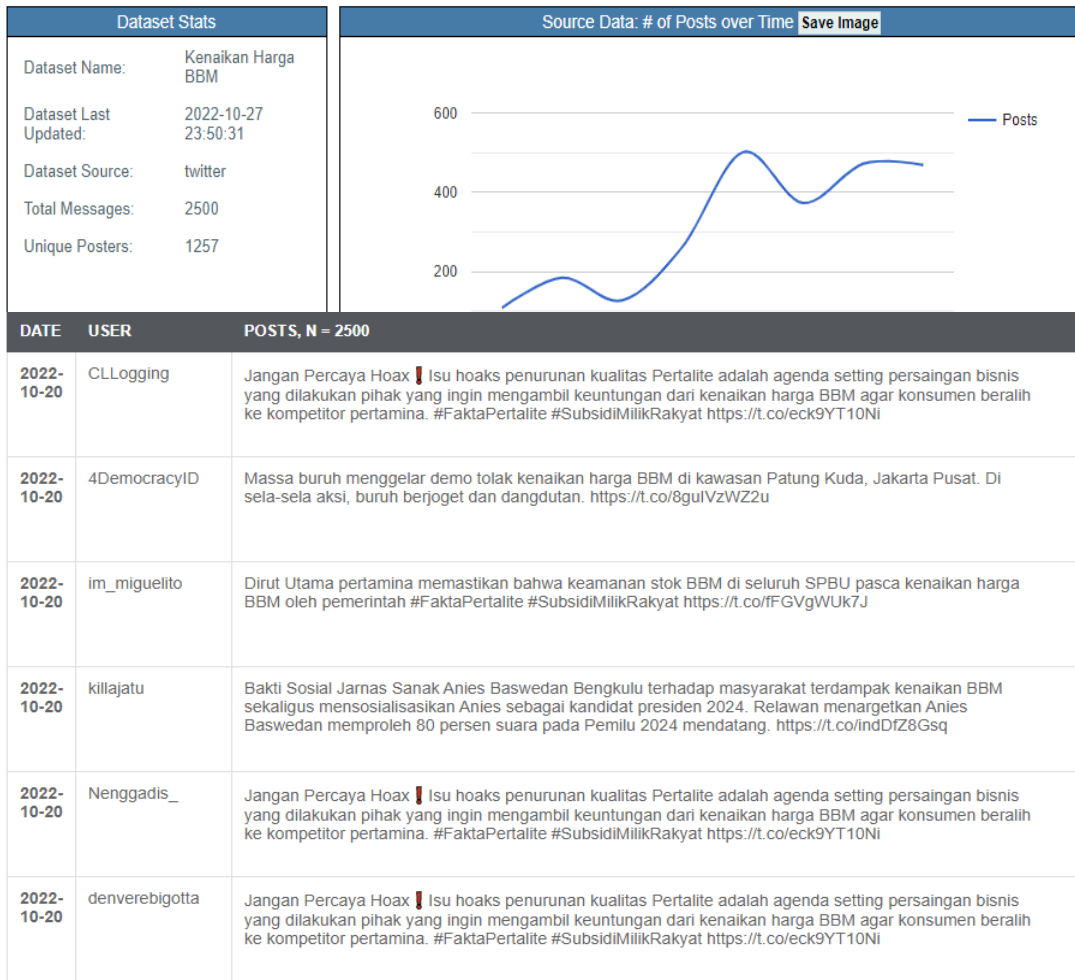
- Hidayatillah, R., Mirwan, M., Hakam, M., Nugroho, A. 2019. Levels of Political Participation Based on Naïve Bayes Classifier. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*. **13(1)** : 73-82. <https://doi.org/10.22146/ijccs.42531>
- Hodson, T. O. 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*. **15(14)** : 5481-5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Jurafsky, D., Martin, J. H. 2023. *Speech and Language Processing (3<sup>rd</sup> ed.draft)* . Britania Raya: Pearson Prentice Hall.
- Kemp, S. 2023. Data Reportal : Twitter Users, Stats, Data & Trends. [https://datareportal.com/essential-twitter-stats?utm\\_source=DataReportal&utm\\_medium=Country Article Hyperlink&utm\\_campaign=Digital 2023&utm\\_term=Indonesia&utm\\_content=Facebook Stats Link](https://datareportal.com/essential-twitter-stats?utm_source=DataReportal&utm_medium=Country Article Hyperlink&utm_campaign=Digital 2023&utm_term=Indonesia&utm_content=Facebook Stats Link) [diakses 11 Mei 2023]
- Kharis, S. A. A., Hadi, I., Hasanah, K. A. 2019. Multiclass Classification of Brain Cancer with Multiple Multiclass Artificial Bee Colony Feature Selection and Support Vector Machine. *Journal of Physics Conference Series*. **1417(1)** : 012015.
- Kirana, A. P., Bhawiyuga, A. 2021. Novel Coronavirus Pandemic in Indonesia: Cases Overview and Daily Data Time Series using Naïve Forecast Method. *Indonesian Journal of Electronics, Electromedical, and Medical Informatics (IJEEMI)*. **3(1)**:1-8.
- Kumila, A., Sholihah, B., Evizia, E., Safitri, N., Fitri, S. 2019. Perbandingan Metode Moving Average dan Metode Naïve Dalam Peramalan Data Kemiskinan. *Jurnal Teori dan Aplikasi Matematika (JTAM)*. **3(1)** : 65-73.
- Kurniawan, D. 2020. *Pengenalan Machine Learning dengan Python*. Jakarta: PT Elex Media Komputindo.
- Locarso, G. K. 2022. Analisis Sentimen Review Aplikasi Peduli Lindungi Pada Google Play Store Menggunakan NBC. *Jurnal Teknik Informatika Kaputama (JTIK)*. **6(2)**. 353-361.
- Manning, C.D., Raghavan, P., Schütze, H. 2008. *An Introduction To Information Retrieval*. England: Chambridge University Press.
- Mohajon, J. 2020. Confusion Matrix for Your Multi-Class Machine Learning Model <https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826> [diakses 29 Mei 2020]

- Pribadi, A. 2022. *SIARAN PERS: Pengalihan Subsidi BBM, Demi Subsidi Tepat Sasaran dan Antisipasi Gejolak Minyak Dunia*. <https://migas.esdm.go.id/post/read/siaran-pers-pengalihan-subsidi-bbm-demi-subsidi-tepat-sasaran-dan-antisipasi-gejolak-minyak-dunia> [diakses 27 Desember 2022]
- Purbo, O. W. 2020. *Text Mining Analisis Medsos, Kekuatan Brand, dan Intelijen di Internet*. Sleman: Andi.
- Sanubari, T., Prianto, C., Riza, N. 2020. *Odol (One Desa One Product Unggulan Online) Penerapan Metode Naive Bayes Pada Pengembangan Aplikasi E-commerce Menggunakan Codeigniter*. Bandung: Kreatif.
- Tinegas, R. 2021. *Dqlab AI-Poweres Learning : Crawling Sebagai Teknik Pengumpulan Data Sekunder* <https://dqlab.id/crawling-sebagai-teknik-pengumpulan-data-sekunder> [diakses 29 Juli 2021]

# **LAMPIRAN**



## Lampiran 1. Hasil *Crawling* Data Twitter



**Lampiran 2. Data Hasil Pelabelan dengan *Lexicon Based***

Tanggal	Tweet	Sentimen	
		Nilai	Label
20/10/2022	6 Kritik PKS di Tiga Tahun Pemerintahan Jokowi-Ma'ruf Juru Bicara PKS @kholidmuhammad menyampaikan 6 kritik untuk pemerintah, mulai dari kemunduran demokrasi hingga inflasi dan kenaikan BBM. <a href="https://t.co/IQEDFMIu54">https://t.co/IQEDFMIu54</a> <a href="https://t.co/6uFP2BKnhf">https://t.co/6uFP2BKnhf</a>	-5	Negatif
20/10/2022	Kegiatan Bansos Dalam Rangka Membantu Masyarakat yang kurang mampu yang terdampak kenaikan harga BBM, Kapolsek Samarinda Ulu Kumpul H. Kustiana, SH., <a href="https://t.co/srwBqgwD4I">https://t.co/srwBqgwD4I</a> ., melalui Kanit Binmas Iptu Roesdi, SH., PS. Panit 2 Binmas Aipda Iman Tawakal, SE., dan Bhabinkamtibmas <a href="https://t.co/rYThjwJfER">https://t.co/rYThjwJfER</a>	-2	Negatif
20/10/2022	Kenaikan harga BBM serta pembelian Pertalite ternyata memengaruhi penjualan mobil bekas di Jateng. Kini mobil di bawah 2.000 cc jadi incaran, Lur! <a href="https://t.co/UQqRDNwrpk">https://t.co/UQqRDNwrpk</a>	0	Netral
20/10/2022	@erickthohir Di hadapi, rakyat yg kena atasi solusi bumh, pln & bbn naikkk, meh..... Yg jelas d hadapi itu kenaikan gaji karyawan BUMN, asekkkk	0	Netral

20/10/2022	<p><a href="https://t.co/J0o4ncXb4W">https://t.co/J0o4ncXb4W</a> “ Polsek Batu Sopang menggelar pembagian sembako sebagai bentuk kepedulian dan perhatian Polri kepada masyarakat yang terdampak akibat kenaikan BBM (Bahan Bakar Minyak). <a href="https://t.co/hxrOogtYOK">https://t.co/hxrOogtYOK</a></p>	-5	Negatif
20/10/2022	<p>Yang mengadu ke Balai Kota Dibatasi Tidak Boleh Mengadu soal Kenaikan Harga Sembako, BBM dan Mahalnya Kebutuhan Hidup.</p> <p>ðŸŒœ</p>	-1	Negatif
20/10/2022	<p>Pastikan Ketersediaan BBM Cukup, Babinsa Pengamanan Dan Pengawasan SPBU. Babinsa Koramil 01/Laratama Serda Abraham Tunliu melaksanakan pengamanan dan Pengawasan wilayah serta monitoring aktifitas SPBU paska terjadinya kenaikan BBM. <a href="https://tco/7SBuvhbXqx">https://tco/7SBuvhbXqx</a></p>	-2	Negatif
20/10/2022	<p>Jgn takut dengan kenaikan BBM karena Indonesia ada mobil tak kasat mata yaitu ESEMKA.</p> <p>Tak kasat mata ini maksudnya mobil ini tak membutuhkan energi. Hanya pikirkan Pak Joko maka mobilnya bisa langsung jalan <a href="https://t.co/SyDuW6BGhy">https://t.co/SyDuW6BGhy</a></p>	-3	Negatif

20/10/2022	<p><a href="https://t.co/J0o4ncXb4W">https://t.co/J0o4ncXb4W</a> “ Personil Polres Paser gelar pengamanan dan monitoring di SPBU dalam upaya mendisiplinkan warga dampak kenaikan harga Bahan Bakar Minyak (BBM) dan guna mengantisipasi kemacetan akibat antrian serta penyalahgunaan BBM di wilayah hukum Polres Paser.</p> <p><a href="https://t.co/BqweMBVLx2">https://t.co/BqweMBVLx2</a></p>	-5	Negatif
20/10/2022	<p>Tapi kok alasannya dari dulu sama aja ya? Untuk analisis lebih lanjut, yok kita simak video yang bisa langsung ditonton di <a href="https://t.co/NYXnZ2xZgm">https://t.co/NYXnZ2xZgm</a> <a href="#">ðŸ˜‰</a></p> <p>Kalau dari sobat teknik sendiri, bagaimana nih dengan adanya kenaikan harga BBM ini? Setuju atau ngga setuju?<a href="#">ðŸ˜‰</a>”</p>	-1	Negatif
:	:	:	:
27/10/2022	<p>[Lamongan] Harga Rajungan Terjun Bebas Akibat Kenaikan Harga BBM   Blusukan Kru MMC</p> <p>Di saat harga BBM yang tidak kunjung turun, harga rajungan di pasaran Lamongan malah melesat terjun ke bawah. Bagaimana respon masyarakat ?</p> <p><a href="https://t.co/gT2ZefEa1K">https://t.co/gT2ZefEa1K</a></p> <p><a href="https://t.co/AfpQNw993L">https://t.co/AfpQNw993L</a></p>	-2	Negatif
27/10/2022	<p>Memberikan bantuan sembako kepada warga yang membutuhkan yang terdampak kenaikan harga BBM atas nama bapak M.</p>	1	Positif

	AMIN di Jalan Prof. Dr. Hamka Lk.3 Kel. Durian Kec. Bajenis Kota Tebing Tinggi <a href="https://t.co/xIeWX8BhtW">https://t.co/xIeWX8BhtW</a>		
27/10/2022	Melaksanakan sambang serta memberi bantuan berupa sembako kepada warga dsn I Desa Gunung Para II Kec.Dolok Merawan an: Manahan Manurung yg terdampak Kenaikan BBM guna meringankan kebutuhan ekonomi sehari hari. Serta mengajak warga untuk tetap menjaga situasi kamtibmas yang aman <a href="https://t.co/QuG5ycwDH4">https://t.co/QuG5ycwDH4</a>	5	Positif
27/10/2022	Polda Jatim bagikan 250 Paket Sembako kepada masyarakat yang terdampak kenaikan BBM yang dilaksanakan pada Rabu, 26 Oktober 2022 DI Pasar Jambangan Surabaya <a href="https://t.co/r6EwwUXY9h">https://t.co/r6EwwUXY9h</a>	0	Netral
27/10/2022	Bank Dunia mengatakan menyusutnya nilai mata uang di sebagian besar negara berkembang telah mendorong kenaikan harga pangan dan bahan bakar yang mengancam akan memperburuk krisis pangan dan energi global. via @voaindonesia <a href="https://t.co/wphhwAJPTk">https://t.co/wphhwAJPTk</a> <a href="https://t.co/IUt6Cx0A1i">https://t.co/IUt6Cx0A1i</a>	-6	Negatif

27/10/2022	Anggaran subsidi dan kompensasi energi akan kembali membengkak sebesar Rp. 198 Triliun, jika tidak ada kenaikan harga BBM Pertalite dan Solar. Saat ini anggaran subsidi dan kompensasi energi untuk 2022 dipatok sebesar Rp 502 triliun. Angka itu sudah membengkak Rp. 349,9 triliun. <a href="https://t.co/GkNHfU7BCs">https://t.co/GkNHfU7BCs</a>	-1	Negatif
27/10/2022	Polda Jatim bagikan 250 Paket Sembako kepada masyarakat yang terdampak kenaikan BBM yang dilaksanakan pada Rabu, 26 Oktober 2022 DI Pasar Jambangan Surabaya <a href="https://t.co/8kJQanZ8cp">https://t.co/8kJQanZ8cp</a>	0	Netral
27/10/2022	Polda Jatim bagikan 250 Paket Sembako kepada masyarakat yang terdampak kenaikan BBM yang dilaksanakan pada Rabu, 26 Oktober 2022 DI Pasar Jambangan Surabaya <a href="https://t.co/JRObni4YEk">https://t.co/JRObni4YEk</a>	0	Netral
27/10/2022	Pasokan BBM Aman Penanganan Inflasi Indonesia Diapresiasi Pemerintah memastikan tekanan inflasi akibat kenaikan harga energi global yang sangat tinggi tidak terlampaui berpengaruh terhadap inflasi di dalam negeri. Pemerintah, menjaga stabilisasi harga minyak, BBM bersubsidi. <a href="https://t.co/eiOAP6vVzq">https://t.co/eiOAP6vVzq</a>	0	Netral
27/10/2022	Polda Jatim bagikan 250 Paket Sembako kepada masyarakat yang terdampak kenaikan BBM yang dilaksanakan pada Rabu, 26	0	Netral

	Oktober 2022 DI Pasar Jambangan Surabaya <a href="https://t.co/Fz4Vj4xnuV">https://t.co/Fz4Vj4xnuV</a>		
--	---	--	--

**Lampiran 3. Data Bersih Hasil Pra Proses Data pada 20 Oktober 2022 - 27 Oktober 2022 (Periode Satu Minggu)**

<b>Tanggal</b>	<b><i>Tweet</i></b>	<b>Label Sentimen</b>
20/10/2022	kritik pks pemerintahan jokowimaruf juru bicara pks kritik pemerintah mundur demokrasi inflasi bbm	Negatif
20/10/2022	giat bansos rangka bantu masyarakat dampak naik harga bbm kapolsek samarinda ulu kompol kustiana kanit binmas iptu roesdi panit binmas aipda iman tawakal bhabinkamtibmas	Negatif
20/10/2022	harga bbm beli pertalite nyata pengaruh jual mobil bekas jateng mobil incar	Netral
20/10/2022	hadap rakyat kena solusi bumh pln bbm naikk hadap gaji karyawan bumh aseki	Netral
20/10/2022	polsek batu sopang gelar sembako bentuk peduli hati polri masyarakat dampak akibat bbm bahan bakar minyak	Negatif
20/10/2022	adu balai kota dibatasi adu harga sembako bbm mahal butuh hidup	Negatif
20/10/2022	sedia bbm babinsa aman awas spbu babinsa koramil laratama serda abraham tunliu laksana aman awas wilayah monitoring aktifitas spbu paska bbm	Negatif
20/10/2022	rakyat cerdas terima bbm non subsidi	Positif

20/10/2022	alangkah properti gambar suai konteks postingan bapak ibu isi kecewa bapak ibu parpol hadap harga bbm jabat presiden wakil presiden	Positif
20/10/2022	adu balai kota batas adu harga sembako bbm mahal butuh hidup	Negatif
:	:	:
27/10/2022	polda jatim paket sembako masyarakat dampak naik bbm laksana rabu oktober pasar jombang surabaya	Netral
27/10/2022	polres madiun bagi beras warga kurang mampu dampak naik harga bbm	Netral
27/10/2022	tolak tunjang transportasi anggota dprd mamuju pasca bbm	Netral
27/10/2022	lebay pasca naik harga bbm ternyata harga sembako masih lihat stabil	Netral
27/10/2022	pemerintah beri blt warga dampak naik harga bbm blt bbm efektif bantu masyarakat sumatera barat	Negatif
27/10/2022	harga bbm subsidi akibat alokasi anggaran subsidi bentuk blt banding dampak positif terima rakyat putar roda ekonomi beri adil	Positif
27/10/2022	alokasi anggaran dana subsidi bbm wujud bentuk blt dampak naik bbm subsidi alami imbang dengan benefit rakyat	Positif
27/10/2022	dampak naik bbm inflasi blt subsidi bbm bukti inflasi kendali indonesia	Netral
27/10/2022	blt bbm warga miskin pemerintah jaga angka miskin ekstrem lonjak bijak pemerintah berupa naik harga bbm paham masyarakat tata subsidi	Negatif



27/10/2022	pemerintah kenal blt bbm oktober kpm jaga daya beli masyarakat butuh distribusi blt bbm oktober roda ekonomi putar tahan harga global	Negatif
------------	---	---------

#### Lampiran 4. Hasil *Output Term Frequency* Pada Data Latih

```
> #train.tokens.dfm = dfm_trim(train.tokens.dfm,min_termfreq = 5)
> train.tokens.dfm
Document-feature matrix of: 2,001 documents, 2,317 features (99.41% sparse) and 0 docvars.
  features
docs   polda jatim paket sembako nelayan kalisari surabaya dampak naik bbm
text1  1  1  1  1  1  1  2  2  1  1  1
text2  0  0  0  0  0  0  0  0  0  0  1
text3  1  1  1  1  1  1  2  2  1  1  1
text4  1  1  1  1  1  1  2  2  1  1  1
text5  1  1  1  1  1  1  2  2  1  1  1
text6  1  1  1  1  1  1  2  2  1  1  1
[ reached max_ndoc ... 1,995 more documents, reached max_nfeat ... 2,307 more features ]
> |
```

#### Lampiran 5. Hasil *Output Model Multinomial Naïve Bayes Classifier*

```
> summary(mnb)
===== Multinomial Naive Bayes =====
- Call: multinomial_naive_bayes(x = train.tokens.matrix, y = train$label, laplace = 1)
- Laplace: 1
- Classes: 3
- Samples: 2001
- Features: 2317
- Prior probabilities:
  - Negatif: 0.4003
  - Netral: 0.2854
  - Positif: 0.3143
```

#### Lampiran 6. Hasil *Output Confusion Matrix* dan Evaluasi Akurasi Hasil Klasifikasi *Multinomial Naïve Bayes Classifier*

```
-----
Confusion Matrix and Statistics

          Reference
Prediction Negatif Netral Positif
Negatif    775     16     10
Netral      80    475     16
Positif     36     12    581

Overall statistics

Accuracy : 0.915
95% CI : (0.902, 0.9269)
No Information Rate : 0.4453
P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.8702

McNemar's Test P-Value : 1.624e-12
```

```

Statistics by Class:
                Class: Negatif
Sensitivity      0.8698
Specificity      0.9766
Pos Pred Value   0.9675
Neg Pred Value   0.9033
Precision        0.9675
Recall           0.8698
F1               0.9161
Prevalence       0.4453
Detection Rate   0.3873
Detection Prevalence 0.4003
Balanced Accuracy 0.9232
                Class: Netral
Sensitivity      0.9443
Specificity      0.9359
Pos Pred Value   0.8319
Neg Pred Value   0.9804
Precision        0.8319
Recall           0.9443
F1               0.8845
Prevalence       0.2514
Detection Rate   0.2374
Detection Prevalence 0.2854
Balanced Accuracy 0.9401
                Class: Positif
Sensitivity      0.9572
Specificity      0.9656
Pos Pred Value   0.9237
Neg Pred Value   0.9810
Precision        0.9237
Recall           0.9572
F1               0.9401

```

**Lampiran 7. Hasil *Output* Akurasi *Naïve Forecasting***

```

R 4.2.2 . ~/
> rmse(Y,forecastPositif)
[1] 8.958236
> rmse(Y,forecastNegatif)
[1] 14.86607
> rmse(Y,forecastNetral)
[1] 11.45098
> |

```