

Klasifikasi Sentimen Masyarakat di Twitter terhadap Kenaikan Harga BBM dengan Metode K-NN

Tiara Dwi Arista¹, Yusra², Muhammad Fikry², Lola Oktavia⁴

^{1,2,3,4}Universitas Islam Negeri Sultan Syarif Kasim Riau, Jl. HR. Soebrantas No.Km.15, RW. 15, Simpang Baru, Pekanbaru, Indonesia.

E-mail : ¹11950125405@students.uin-suska.ac.id, ²yusra@uin-suska.ac.id, ³muhammad.fikry@uin-suska.ac.id, ⁴lola.oktavia@uin-suska.ac.id

Abstrak. Kenaikan harga Bahan Bakar Minyak (BBM) di Indonesia merupakan isu besar yang menjadi topik utama hingga saat ini. Kenaikan harga BBM di Indonesia telah berlangsung sejak awal September 2022. Kebijakan pemerintah menaikkan harga BBM kemudian menimbulkan banyak opini dari kalangan masyarakat. Opini masyarakat terkait kebijakan pemerintah menunjukkan adanya sentimen positif dan negatif yang dapat dilihat melalui media sosial, seperti Twitter. Tujuan dari penelitian ini adalah untuk mengklasifikasikan sentimen masyarakat terhadap kenaikan harga BBM di Twitter. Jumlah data yang digunakan adalah 3000 tweet yang dikumpulkan berdasarkan kata kunci yaitu "Kenaikan BBM" dan "BBM naik". Penerapan metode K-Nearest Neighbor (K-NN), Feature Weighting (TF-IDF), dan Feature Selection (Threshold) akan dilakukan implementasi dengan menggunakan tools yaitu Google Collab. Berdasarkan hasil pengujian metode K-NN menggunakan confusion matrix pada 10 nilai K yang berbeda (3,5,7,9,11,13,15,17,19,21) dengan mekanisme perbandingan yang digunakan 70:30, 80:20, dan 90:10 diperoleh akurasi paling tinggi sebesar 83,3% pada K=13 dan K=15 untuk perbandingan data training dan testing 90:10.

Kata Kunci : Klasifikasi Sentimen, K-NN, Twitter, Kenaikan BBM

Abstract. The rise in fuel prices in Indonesia has become a significant issue that remains a central topic today. The increase in fuel prices in Indonesia has been ongoing since early September 2022. The government's policy to raise fuel prices has generated various opinions from the public. Public opinions regarding the government's policy can be seen through social media platforms such as Twitter, where both positive and negative sentiments are expressed. The objective of this research is to classify public sentiment toward the fuel price hike on Twitter. A dataset consisting of 3000 tweets will be collected using keywords "fuel price hike" and "fuel increases". The K-Nearest Neighbor (K-NN), Feature Weighting (TF-IDF), and Feature Selection (Threshold) will be implemented using Google Colab. Based on the testing results of the K-NN method using a confusion matrix with 10 different K values (3, 5, 7, 9, 11, 13, 15, 17, 19, 21) and comparison mechanisms of 70:30, 80:20, and 90:10, the highest accuracy of 83.3% was achieved at K=13 and K=15 for the 90:10 training and testing data ratio.

Keyword : Sentiment Classification, K-NN, Twitter, Fuel Increase

PENDAHULUAN

Bahan Bakar Minyak (BBM) adalah energi yang dihasilkan dari bahan alamiah minyak dan gas bumi. BBM termasuk komoditas penting yang tidak bisa dilepaskan dari masyarakat [1]. Dalam memenuhi kebutuhan sehari-hari masyarakat membutuhkan BBM. Oleh karena itu, harga BBM mempengaruhi kenaikan harga-harga kebutuhan masyarakat [2]. Kenaikan harga BBM berdampak besar bagi masyarakat seperti keterlibatan dalam naiknya harga barang dan jasa, penurunan daya beli masyarakat, terhambatnya produksi masyarakat dan inflasi.

Gejolak harga BBM di Indonesia sebenarnya sudah terlihat sejak tahun 2013 [2]. Saat ini, harga minyak global telah mencapai US\$100 per barel, yang berdampak pada kenaikan harga BBM di Indonesia [3]. Kenaikan harga BBM di Indonesia merupakan isu besar yang menjadi topik utama hingga saat ini. Kenaikan harga BBM di Indonesia telah berlangsung sejak awal September 2022. Presiden Joko Widodo (Jokowi) memutuskan untuk menyesuaikan harga BBM bersubsidi. Jokowi mengakui bahwa keputusan untuk menyesuaikan harga BBM bersubsidi merupakan keputusan sulit. Pemerintah juga telah berusaha dengan maksimal agar masyarakat terlindungi dari fluktuasi harga minyak dunia. Tetapi menurut Jokowi, saat ini



situasi Anggaran Pendapatan Negara (APBN) dianggap tidak lagi memungkinkan untuk mempengaruhi hal tersebut. Pada tahun 2022, anggaran subsidi dan kompensasi BBM telah naik tiga kali lipat dari angka sebelumnya, yaitu Rp152,5 triliun menjadi Rp502,4 triliun dan diperkirakan akan terus bertambah. Pemerintah akhirnya menaikkan harga BBM bersubsidi yang telah diputuskan oleh Menteri ESDM Arifin Tarif di Istana Negara. Perihal kenaikan harga BBM terdapat Pertalite yang naik tarif dari Rp7.650 menjadi Rp10.000 per liter, Solar dari Rp5.150 menjadi Rp6.800 per liter, serta Pertamax dari Rp12.500 menjadi Rp14.500 per liter.

Kebijakan pemerintah menaikkan harga BBM kemudian menimbulkan banyak opini dari kalangan masyarakat. Opini mengenai kenaikan harga BBM menunjukkan adanya sentimen positif dan negatif. Sentimen masyarakat terkait kebijakan pemerintah bisa terlihat melalui platform media sosial yang populer pada kalangan masyarakat, seperti Twitter. Pada tahun 2019, sekitar 6,43 juta orang di Indonesia menggunakan Twitter yang setara dengan sekitar 52% dari seluruh pengguna media sosial [4]. Twitter menjadi petunjuk kemampuan individu dalam mengomunikasikan, menerima, dan menyebarluaskan informasi lainnya kepada pengguna lain agar masyarakat umum mengetahuinya [5]. Informasi ini dapat mengukur opini publik untuk mengetahui sentimen tentang kenaikan harga BBM yang sedang terjadi saat ini. Untuk itu maka diperlukan klasifikasi sentimen terhadap kenaikan harga BBM agar dapat memperkirakan kelas yang labelnya tidak diketahui.

K-Nearest Neighbor (K-NN) merupakan teknik klasifikasi yang paling terkenal [6]. Metode K-NN termasuk *supervised learning*, dimana data dikelompokkan ke dalam kelas-kelas yang telah ditentukan berdasarkan jarak terpendek dari data [7]. Metode K-NN sendiri sudah banyak digunakan pada penelitian sebelumnya, seperti penelitian [8] mengenai klasifikasi komentar *bullying* pada Instagram. Penelitian ini menggunakan 1000 data yang mana diantaranya 500 tidak *bully* dan 500 *bullying*. Ditemukan akurasi paling tinggi pada K=13 di rasio 90:10 pada *fold*=6 yakni mencapai 77%. Penelitian berikutnya [9] membahas mengenai klasifikasi sentimen masyarakat terhadap *new normal* pada masa pandemi. Data yang dikumpulkan sebanyak 222 data *tweet*. Proses klasifikasi ini dilakukan dengan membandingkan dua metode yaitu *Support Vector Machine* (SVM) dan *K-Nearest Neighbor* (K-NN). Uji coba pada penelitian ini membuktikan bahwa metode K-NN mengungguli SVM dalam menangani masalah ini, dengan mencapai tingkat keakuratan sebesar 72,96%, sedangkan SVM hanya mencapai 71,19%. Penelitian lainnya yaitu [10] analisis sentimen aplikasi Halodoc dengan metode *Naïve Bayes*, SVM, dan K-NN. Data yang dipakai yakni 1000 dataset untuk ulasan positif dan negatif. Hasil penelitian ini menunjukkan bahwa metode K-NN paling akurat dalam klasifikasi sentimen karena menghasilkan akurasi tertinggi 95,00%, sementara *Naïve Bayes* dan SVM yang memiliki nilai akurasi 92,50% dan 93,00%.

Dari pemaparan diatas, penelitian ini akan mengklasifikasikan sentimen masyarakat terhadap kenaikan harga BBM di Twitter. Terdapat 3000 data *tweet* yang digunakan berdasarkan pencarian kata kunci "Kenaikan BBM" dan "BBM naik". Data-data tersebut dikumpulkan dari 1 November 2022 sampai 1 Desember 2022. Penerapan metode *K-Nearest Neighbor* (K-NN), *Feature Weighting (TF-IDF)* dan *Feature Selection (Threshold)* akan dilakukan implementasi dengan menggunakan *tools* yaitu Google Collab. Tujuan penelitian ini adalah menerapkan dan menguji keakuratan metode K-NN dalam melakukan klasifikasi sentimen masyarakat terhadap kenaikan harga BBM di Twitter.

METODOLOGI PENELITIAN

Adapun tahapan penelitian adalah seperti berikut:

2.1 Identifikasi Masalah

Pengidentifikasian masalah merupakan komponen dari proses penelitian ini, yang dapat dimaknai sebagai usaha untuk merumuskan permasalahan dan memastikan definisi tersebut dapat diukur sebagai langkah pertama pada penelitian.

2.2 Studi Pustaka

Studi pustaka adalah proses pengumpulan dan pencarian referensi yang berkaitan dengan penelitian dan dapat membantu terlaksananya penelitian sehingga mencapai tujuan yang diharapkan.

2.3 Pengumpulan Data

Terdapat 3000 data *tweet* yang digunakan dari pencarian berdasarkan kata kunci "Kenaikan BBM" dan "BBM naik". Data-data tersebut dikumpulkan terhitung dari 1 November 2022 sampai 1 Desember



2022.

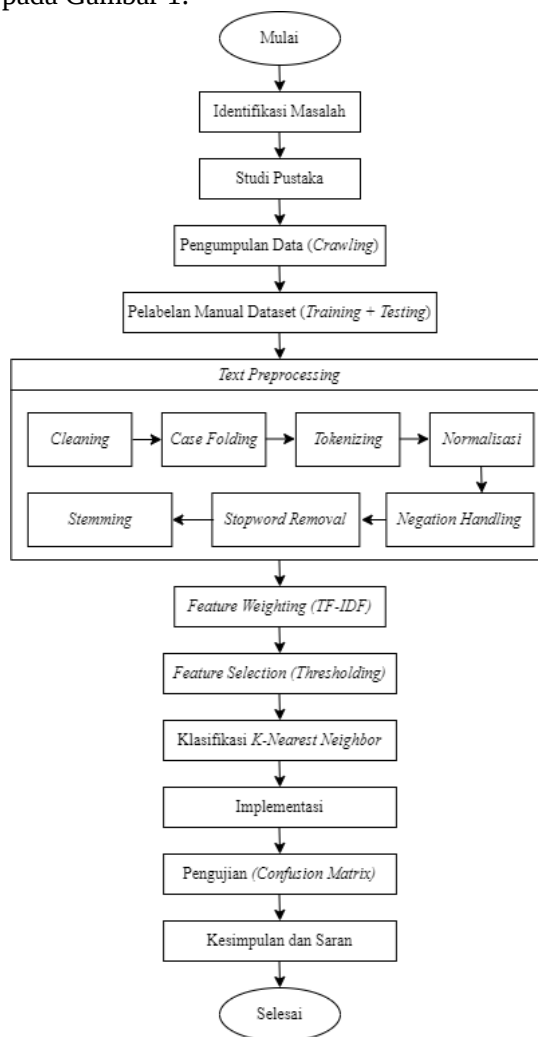
2.4 Pelabelan Manual Dataset

Tahap pelabelan dilakukan untuk mengatasi masalah pengklasifikasian pada data *tweet* agar data tersebut masuk ke dalam kelompok yang tepat sesuai dengan informasinya.

2.5 Text Preprocessing

Text preprocessing sangatlah penting dalam klasifikasi dimana dataset akan dipersiapkan dan dibersihkan untuk membuat semua bentuk kata serupa sehingga dapat mengurangi volume kata dalam kumpulan dokumen [11].

Tujuan dari metodologi penelitian ini adalah untuk mencapai hasil penelitian yang diinginkan. Tahapan metodologi penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahap Metodologi Penelitian Feature Weighting

Tahap terakhir merupakan tahap pembobotan kata yaitu dengan TF-IDF atau *Term Frequency Inverse Document Frequency*, dimana tujuannya untuk memberi nilai bobot pada setiap kata serta mengindikasikan seberapa unik kemunculan kata [12]. Perhitungan hasil TF-IDF digunakan untuk memperhitungkan jarak terpendek dengan menggunakan rumus *Euclidean Distance*. Dalam tahap pembobotan kata, berikut langkah-langkah yang dilakukan :

1. Menghitung frekuensi kemunculan setiap kata pada seluruh dokumen
2. Menghitung nilai IDF menggunakan rumus =

$$IDF = \log \frac{d}{df}$$

Keterangan :



IDF = Bobot IDF

d = Jumlah dokumen keseluruhan

df = Jumlah dokumen yang memiliki *term*

3. Melakukan perhitungan bobot TF-IDF dari semua kata pada dokumen dengan menggunakan rumus =

$$W_{dt} = tf_t \times IDF_t$$

Keterangan :

W_{dt} = Bobot *term* ke t pada dokumen d

tf_t = Jumlah *term* t yang muncul pada dokumen t

IDF_t = Bobot IDF yang telah dihitung sebelumnya

4. Hasil yang didapatkan akan digunakan sebagai patokan untuk menentukan ukuran *vector* sebelum melakukan perhitungan jarak dengan *Euclidean Distance* =

$$d_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

Keterangan :

d_{ij} = Jarak antar i dan j

x_i = Kata ke i di dokumen x

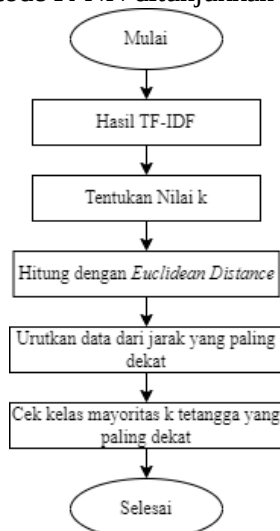
y_i = Kata ke i di dokumen y

2.6 Feature Selection

Tahap ini merupakan tahap pemilihan fitur yang bertujuan untuk mengurangi jumlah variabel input menjadi variabel yang paling berguna. Adapun pada penelitian ini pemilihan fitur menggunakan *threshold* yang merupakan nilai batas relevan untuk pemilihan kategori fitur optimal [13].

2.7 Klasifikasi K-Nearest Neighbor

Metode *K-Nearest Neighbor* (K-NN) merupakan salah satu teknik yang dapat dipakai untuk melakukan klasifikasi. Cara kerjanya adalah dengan mencari jarak terdekat antara *training sample* dan *query instance* untuk mendapatkan KNN-nya. Kelas yang kemunculannya paling banyak adalah hasil klasifikasinya [14]. Tahapan-tahapan metode K-NN ditunjukkan pada Gambar 2.



Gambar 2. Flowchart Metode K-NN

2.8 Implementasi

Tahap implementasi adalah tahap penerapan dari hasil analisis yang telah dilakukan. Pengimplementasian memiliki tujuan untuk memastikan apakah tahap-tahap yang telah dilakukan telah sesuai dengan yang diharapkan. Implementasi dilakukan dengan menggunakan *tools* yaitu Google Collab.

2.9 Pengujian

Pengujian dilaksanakan dengan maksud untuk memahami bagaimana kinerja suatu metode. Dalam penelitian ini, pengujian tingkat keakuratan metode dilakukan dengan menggunakan *confusion matrix*. Pengujian *confusion matrix* ditunjukkan pada tabel berikut :



Tabel 1. Confusion Matrix

Predicted Class	Positive	Negative
Positive	TP (True Positive)	FP (False Positive)
Negative	FN (False Negative)	TN (True Negative)

Parameter yang digunakan yaitu *accuracy*, *precision*, *recall*, dan *f1 score*.

$$\text{Accuracy} = \frac{(TP+TN)}{TP+TN+FP+FN} \times 100\%$$

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\%$$

Keterangan =

TP = Jumlah data yang memiliki kelas positif yang diklasifikasi positif

TN = Jumlah data yang memiliki kelas negatif yang diklasifikasi negatif

FP = Jumlah data yang memiliki kelas positif yang diklasifikasi negatif

FN = Jumlah data yang memiliki kelas negatif yang diklasifikasi positif

2.10 Kesimpulan dan Saran

Inti dari hasil penelitian adalah bagian kesimpulan. Sementara itu, saran berisikan rekomendasi yang diperlukan terkait penelitian ini, untuk memungkinkan pengembangan lebih lanjut untuk kedepannya dan mendapatkan hasil yang lebih optimal.

HASIL DAN PEMBAHASAN

3.1 Data Selection

Data *tweet* yang didapat melalui *crawling* di Twitter akan melalui seleksi data, atribut yang digunakan untuk membersihkan data dan untuk klasifikasi selanjutnya akan dipilih pada tahapan ini. Berikut adalah gambar data sebelum dilakukan seleksi data.

A	B	C	D	E
time	location	id	username	text
10/11/22 12:13	Bekasi, Jawa Barat	1.59E+23	cep_yanaA	b'@kennakerrri gw ga dpt tapi spv gw dapet, kan konyol... sekecil kecilnya gaji spv \xf0\x9f\x98\x82\n\
10/11/22 11:31		1.59E+23	ogah	b'ahlan wa sahan ramadhan dan kenaikan BBM semoga semakin sabarrrrrrr #tolakbbmnaik' #tolakbbmnaik'
10/11/22 10:03	ÅA Å@ÅA Å-ÅA Å	1.59E+23	ÅA Å@ÅA Å	b'@gumpnhell yg jadi korban rakyat...alasan dukung demo kenaikan BBM, data2 dijual,lak supir cikal'
10/11/22 8:59		1.59E+23	gebby	b'kenaikan BBM dikatakan mahal jika kita sehari hari naik mobil mewah beli pertamax'
10/11/22 7:28	Duren Sawit, Indonesia	1.59E+23	Johnsingle	b'pak presiden yang terhormatttt... mana janji mu? mana blt dari kenaikan harga BBM itu? saya tidak melihat dan tidak merasakan blt
10/11/22 7:16	Lubuk Pakam, Indonesia	1.59E+23	Jarangpulangsi	b'hey masyarakat Sumatera utara tak perlu panik dengan kenaikan harga BBM. yang pastinya klen semua akan dapat bansos #bantuanbb
10/11/22 6:48	Lubuk Pakam, Indonesia	1.59E+23	Jarangpulangsi	b'wajar jika penyesuaian harga BBM dilakukan pemerintah. seluruh masyarakat Sumatera utara dukung kenaikan harga BBM #bantuanbb
10/11/22 4:11		1.59E+23	Vanvan	b'@carastalk han, anaknya lebih suka ngobrolin jumlah lubang sedotan daripada kenaikan BBM'
10/11/22 3:40		1.59E+23	Fikri Haikal	b'@arifus80056962 @jokowi mau siapa aja pemimpinnya yg namanya kehidupan itu maju, dan harga pokok gaakan stuck segitu\ue2\x8
10/11/22 1:50		1.59E+23	Weyrun	b'@rm14duadua @s_b_darmawan @mouldie_sep lah pdi dulu nolak kenaikan BBM pas udah berkuasa malah nerima. tunggu aja ntar pa
10/11/22 0:12	Indonesia	1.59E+23	IF	b'sektor riil pun terhantam. terlebih setelah kenaikan harga BBM.\n\ntiap ketemu kawan2 ukm, keluhannya sama: sepi banget. ga maki
9/11/22 23:52	Duren Sawit, Indonesia	1.59E+23	ÅA Å Å%ÅCÅiÅCÅ	b'@flowber4n_hanya di jaman jkw kenaikan BBM sampai 13x. jaman cebong hukum tebang pilih'
9/11/22 18:26		1.59E+23	Visioning Prapatan	b'ehehehe...selalu ada alasan untuk menaikkan cukai rokok dan target pajak yes @kemenkeuri ??\nbaru mau bangkit pasca kopit sudah
9/11/22 15:20		1.59E+23	Boge als Mursid	b'@ramlirizal @kedaiPENACOM mengingatkan kpd kaum tua dan kaum muda, marilah bangkit melawan kedololiman, sebelum kita tamba
9/11/22 14:20		1.59E+23	fiona23	b'@pksjabar kami menolak kenaikan BBM hemm pro rakyat, kami menolak #trotoardepok bodo amat... \xf0\x9f\x85\x99 mengsyedihh
9/11/22 14:09		1.59E+23	New_pengguna	b'@muhammad_saewad tidakkah tercium bau korupsi di kenaikan harga BBM ini....?'
9/11/22 12:29	Lampung	1.59E+23	Nothing !!	b'@makinhits itu kalo elo ikut demo kenaikan BBM tapi bawaanya spanduk slank, ama nebakar foto mantan bukan bakar ban :))'
9/11/22 7:19		1.59E+23	waspada manusia mu	b'@clarapu19417438 harga minyak dunia turun disaat BBM naik di Indonesia loe cerna pakai otak kau jgn lonte aja kau.pidato jokowi bil
9/11/22 6:58		1.59E+23	Suara_Fals_Indonesia	b'@tan_mar3m hari ini pdip pro jokowi cuman haso kek nya... \n\nkader lain udah gatel, \nkenaikan BBM susah nya ekonomi rakyat, api
9/11/22 5:02	Papua Barat, Indonesia	1.59E+23	Wajah Indonesia	b'kenaikan harga BBM tak berpengaruh terhadap daya beli masyarakat. \n#bantuanbbmuntukrakyat \n#bltbbmtepatasasan \n#subsidi
8/11/22 6:42	MALUKU UTARA	1.59E+23	Sabri hadad	b'mantab sejumlah pemuda Papua dukung kenaikan BBM \n#bantuanbbmuntukrakyat \n#bltbbmtepatasasan \n#subsidi

Gambar 3. Data Sebelum Tahap Data Selection

Pada penelitian ini, terdapat 3000 data *tweet* yang dikelompokkan ke dalam dua kategori, yaitu *tweet* opini dan *tweet* non opini. Data non opini diperoleh sebanyak 1500 data *tweet*. Data opini diperoleh sebanyak 1500 data *tweet*, terdiri dari 1170 opini negatif dan 330 opini positif.

3.2 Pelabelan Data

Proses pemberian label dataset dilakukan secara manual. Pelabelan dilakukan tokoh ahli yaitu dosen Bahasa Indonesia Universitas Islam Negeri Sultan Syarif Kasim Riau bernama Ibu Elfina, S.Pd.,M.Pd. Hasil pelabelan



ditunjukkan pada Gambar 4.

	A	B	C	D	E	F
1	time	location		username	text	sentimen
2	10/11/22 12:13	Bekasi, Jawa	1.59E+23	@cepYanAisib	b/kemnakkerri gw ga dpt tapi spv gw dapet, kan konyol... sekecil kecilnya gajii spv\xf0\x9f\x98\x82\n\nnalhamdulillah sih, ga dpt uang sua	negatif
3	10/11/22 11:31		1.59E+23	@gah	b'ahlan wa sahan ramadhan dan kenaikan bhm semoga semakin sabarrrrr #tlokabkbnmaik #tlokabkbnmaik	positif
4	10/11/22 10:03	Å&Å&Å&	1.59E+23	Å&Å&Å&	b/gumpnhell yg jadi korban rakyat....alasan dungemo denaikn bhm, dated dijual,lakur sipir cika'	negatif
5	10/11/22 8:59		1.59E+23	@gebbby	b'kenaikan bhm dikatakan malah jika kita sehari hari naik mobil mewah bel pertamax'	negatif
6	10/11/22 7:28	Duren Sawit,	1.59E+23	@johnshingle	b/pak presiden yang terhormattttttt... mana jani mu? mana blt dari kenaikan harga bhm itu? saya tidak melihat dan tidak merasakan blt yg	negatif
7	10/11/22 7:16	Lubuk Pakar	1.59E+22	@Jarangunglp	yg mayasarakat Sumatera utara tak perlu panik dengan kenaikan harga bhm, yang pastinya klin semua akan dapat bansos #bantuanbhm	positif
8	10/11/22 6:48	Lubuk Pakar	1.59E+23	@Jarangunglp	b/wajar jika penyesuaian harga bhm dilakukan pemerintah, seluruh masyarakat Sumatera utara dengan dugaan kenaikan harga bhm #bantuanbhm	positif
9	10/11/22 4:11		1.59E+23	@Vanvan	b/@carstalkn han, ankanya lebih suka ngobrolin lubang lubang sendatan daripada kenaikan bhm	positif
10	10/11/22 3:40		1.59E+23	@Fikri Haikal	b/@arifatusso6962 @jokowi mau siapa aja pemimpinya yg namanya kehidupan itu maju, dan harga pokok gaakan stuck segitu\%x2D\%	positif
11	10/11/22 1:50		1.59E+23	@Weyrun	b/@m14duadua @s_darmawan @moldiee, sep lah di duga nolak kenaikan harga bhm pas udah berkuasa malah nerima, tunggu aja ntar pas	negatif
12	10/11/22 1:02	Indonesia	1.59E+22	@IF	b/sektor rilau pun terhantam, terlebih sektor kenaikan harga bhm,\nn\tintapi ketemu kawan2 ukm, kolahunya sama: sepi banget, ga makan	negatif
13	9/11/22 23:52	Duren Sawit,	1.59E+23	Å&Å&Å&	b/\%0F0elwelaâ, hanya di jamann jugk kenaikan bhm sampai 13%.. jaman cebong huhum tebang pilhh!	negatif
14	9/11/22 23:26		1.59E+23	@Visioning Pr	b/@hehehe...seben ada alasan untuk menaikkan cukup rokok dan target pajak yes @kemkennerri ??nbaru mau bangkit pasca kopit sudah c	negatif
15	9/11/22 15:20		1.59E+23	@Boge als Mu'	b/@ramlirzal @kedaiapencom mengingatn kp'd kaum tua dan kaum muda, marilah bangkit melawan kedolmalan, sebelum kita tambah	negatif
16	9/11/22 14:42		1.59E+23	@fionaz23	b/@ksjabar kami menolak kenaikan bhm hemm pr rakyat, kami menolak ttrottoardepek dogo amat... \xf0\x9f\x9a\x9b mengsyedihhh'	negatif
17	9/11/22 14:09		1.59E+23	@New_pengg	b/@muhammad_saewad tidakkah terciut bau korupsi di kenaikan harga bhm ini....?	negatif
18	9/11/22 12:19	Lampung	1.59E+23	@Nothing !!	b/@makinhini itu kalo elo ikutan demo kenaikan bhm tapi bawanya spanduk slank, ama ngebakar fonta baksa bak ban :))'	positif
19	9/11/22 7:29		1.59E+23	@waspadaa !!	b/@cakra19417438 harga minyak dunia turun disaat bnm naik di indonesia loe cerna pakai otak kau jgn lonte aja kau-pidato jokowi bilan	positif
20	9/11/22 6:58		1.59E+23	@Suara_Fals_	b/@tan_mar3m hari ini pdip pr Jokowi kamu hasto kek ya...n\knader lain udah gate!,\n'kenaikan bhm susah nya ekonomi rakyat, apa i	negatif
21	9/11/22 6:02	Papua Barat,	1.59E+23	@Wajah Indon	b/Wajah indon b'kenaikan harga bhm tak berpengaruh terhadap daya beli masyarakat.\n\n#bantuanbhm#subsidi#praktik#nbtlbttbmtpatsasaran#\n#nbtlbttbmtpatsasaran#\n#	positif
22	8/11/24 6:42	MALUKU UTA	1.59E+23	@Sabri hadad	b/mantab sejumliah pemuda papua dugaan kenaikan bhm \n#bantuanbhm#ntunakrayat#\n#btltbmtpatsasaran#\n#btltbmtpatsasaran#\n#	positif

Gambar 4. Pelabelan Manual Dataset

3.3 Text Preprocessing

Berikut contoh hasil dari tahapan *text preprocessing*.

3.3.1 Cleaning

Pada proses pembersihan, atribut-atribut seperti kata-kata atau karakter yang tidak berkaitan dengan informasi akan dibersihkan dengan tujuan meminimalisir *noise* dalam proses klasifikasi. Atribut-atribut tersebut berupa tanda baca, URL, *mention*, *hashtag*, dan *emoticon* [15].

Tabel 2. *Cleaning*

Sebelum <i>Cleaning</i>	Hasil <i>Cleaning</i>
b'@m45broo Ga setuju..di pertamina juga janjinya berantas mafia ? harga bbm trs naik.'	Ga setuju di pertamina juga janjinya berantas mafia harga bbm trs naik

3.3.2 Case Folding

Pada proses *case folding*, seluruh abjad yang terdapat didalam dokumen akan diubah menjadi *lowercase* atau huruf kecil [16].

Tabel 3. Case Folding

Sebelum Case Folding	Hasil Case Folding
Ga setuju..di pertamina juga janjinya berantas mafia ? harga bbm trs naik	ga setuju di pertamina juga janjinya berantas mafia harga bbm trs naik

3.3.3 Tokenizing

Pada proses *tokenizing*, tujuannya adalah memisahkan kalimat atau memotong string input berdasarkan setiap kata yang membentuk dokumen tersebut [17].

Tabel 4.*Tokenizing*

Sebelum <i>Tokenizing</i>	Hasil <i>Tokenizing</i>
ga setuju di pertamina juga janjinya berantas mafia harga bbm trs naik	['ga', 'setuju', 'di', 'pertamina', 'juga', 'janjinya', 'berantas', 'mafia', 'harga', 'bbm', 'trs', 'naik']

3.3.4 Normalisasi

Pada proses *normalisasi*, setiap kata yang terdapat kesalahan ejaan atau tidak baku akan diubah menjadi bentuk yang baku [16], *normalisasi* penelitian ini memanfaatkan kamus *normalisasi*.

Tabel 5. Normalisasi

Sebelum <i>Normalisasi</i>	Hasil <i>Normalisasi</i>
ga setuju	tidak setuju



di	di
pertamina	pertamina
juga	juga
janjinya	janjinya
berantas	berantas
mafia	mafia
harga	harga
bbm	bbm
trs	terus
naik	naik

3.3.5 Negation Handling

Pada proses *negation handling*, jika didapatkan kata yang memiliki sifat negasi, maka akan mengakibatkan arti kata berikutnya menjadi bertentangan dengan makna yang sebenarnya [18].

Tabel 6. *Negation Handling*

Sebelum <i>Negation Handling</i>	Hasil <i>Negation Handling</i>
['tidak', 'setuju', 'di', 'pertamina', 'juga', 'janjinya', 'berantas', 'mafia', 'harga', 'bbm', 'terus', 'naik']	['menolak', 'di', 'pertamina', 'juga', 'janjinya', 'berantas', 'mafia', 'harga', 'bbm', 'terus', 'naik']

3.3.6 Stopword Removal

Pada proses *stopword removal*, dilakukan penghilangan kata penghubung yang dianggap tidak memiliki arti dalam klasifikasi sentimen sehingga tidak akan berdampak [19].

Tabel 7. *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Hasil <i>Stopword Removal</i>
['menolak', 'di', 'pertamina', 'juga', 'janjinya', 'berantas', 'mafia', 'harga', 'bbm', 'terus', 'naik']	['menolak', 'pertamina', 'janjinya', 'berantas', 'mafia', 'harga', 'bbm', 'naik']

3.3.7 Stemming

Pada proses *stemming*, semua kata akan diubah ke bentuk dasarnya [20]. Algoritma stemming yang dipakai adalah *Enhanced Confix Stripping* (ECS).

Tabel 8. *Stemming*

Sebelum <i>Stemming</i>	Hasil <i>Stemming</i>
['menolak', 'pertamina', 'janjinya', 'berantas', 'mafia', 'harga', 'bbm', 'naik']	['tolak', 'pertamina', 'janji', 'berantas', 'mafia', 'harga', 'bbm', 'naik']

3.4 Feature Weighting

Setelah tahap *text preprocessing*, setiap kata atau *term* dari hasil *preprocessing* akan diberi bobot menggunakan TF-IDF. Berikut tampilan hasil perhitungan TF-IDF menggunakan *tools* Google Collab.

	000	09	10	100	1001	100k	100km	100persen	100rb	1075	...	ywdh	zainal	zaim	zaman	zefanya	zerr	zholimi	zico	zona	Kelas
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	0.0	0.0	0.0	0.0	0.0	0.0	negatif
1	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	0.0	0.0	0.0	0.0	0.0	0.0	positif
2	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	0.0	0.0	0.0	0.0	0.0	0.0	negatif
3	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	0.0	0.0	0.0	0.0	0.0	0.0	positif
4	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	0.0	0.0	0.0	0.0	0.0	0.0	negatif
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
95	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	0.0	0.0	0.0	0.0	0.0	0.0	negatif
96	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	0.0	0.0	0.0	0.0	0.0	0.0	negatif
97	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	0.0	0.0	0.0	0.0	0.0	0.0	positif
98	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	0.0	0.0	0.0	0.0	0.0	0.0	negatif
99	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	0.0	0.0	0.0	0.0	0.0	0.0	negatif

100 rows × 3912 columns

Gambar 5. Hasil TF-IDF



3.5 Feature Selection

Selanjutnya dilakukan pemilihan fitur terbaik dengan menggunakan *threshold*. Pada penelitian ini ambang batas yang digunakan yaitu 0,001. Berikut tampilan hasil *threshold* menggunakan *tools* Google Collab.

Hasil Feature Selection

	000	10	2022	abang	aceh	ahok	akibat	alas	allah	angkut	...	untung	upah	usaha	utang	utara	wajar	warga	wkwk	yogyakarta	Kelas
0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif
1	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
2	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.279229	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif
3	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
4	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1495	0.0	0.0	0.0	0.0	0.0	0.0	0.167579	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
1496	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif
1497	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif
1498	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif
1499	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.000000	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif

1500 rows x 186 columns

Gambar 6. Hasil *Threshold*

3.6 Klasifikasi dengan K-Nearest Neighbor

Tahap berikutnya adalah melakukan pengelompokan menggunakan metode K-NN. Dalam tahap ini, data akan di proses melalui fase *training* dan *testing*, kemudian dilakukan *modelling* dengan metode K-NN untuk menghasilkan nilai *accuracy*, *precision*, *recall*, dan *f1 score*. Berikut tampilan hasil klasifikasi metode K-NN menggunakan *tools* Google Collab.

Hasil Klasifikasi

	000	10	2022	abang	aceh	ahok	akibat	alas	allah	angkut	...	upah	usaha	utang	utara	wajar	warga	wkwk	yogyakarta	Kelas Target	Hasil Klasifikasi
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif	negatif
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.340121	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
145	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	positif	negatif
146	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
147	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
148	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif
149	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	negatif	negatif

150 rows x 187 columns

Gambar 7. Hasil Klasifikasi Metode K-NN

3.7 Pengujian

Pengujian dilakukan untuk menguji data *training* dan data *testing* dengan memanfaatkan *confusion matrix*. Pada penelitian ini dilakukan 3 pengujian yaitu data opini – non opini, data opini, dan data *balance*. Pengujian menggunakan 10 nilai K yang berbeda (3,5,7,9,11,13,15,17,19,21) dengan mekanisme perbandingan yang diterapkan yakni 70:30, 80:20, dan 90:10.

3.7.1 Pengujian Data Opini – Non Opini

Pengujian ini dilakukan dengan 3000 data yang terdiri dari kelas opini dan non opini. Berikut tabel pengujian yang telah dilakukan.

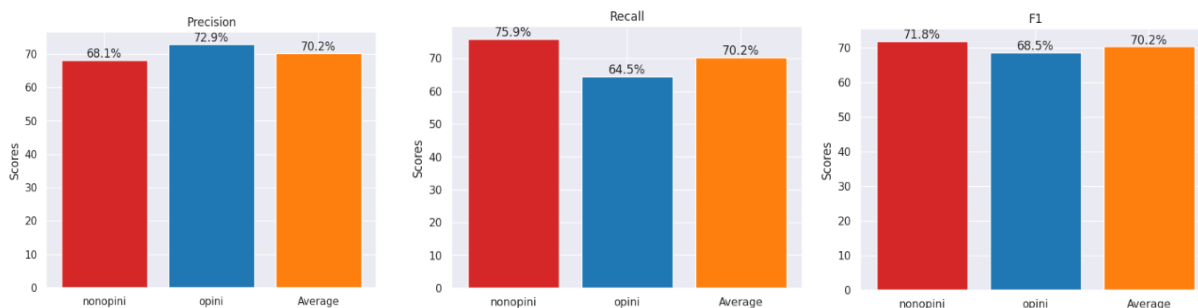
Tabel 9. Pengujian Data Opini – Non Opini

Nilai K	Accuracy		
	70:30	80:20	90:10
3	67.22	67.0	66.33
5	67.78	68.67	66.66
7	68.89	68.16	66.33
9	68.89	66.66	64.33



11	69.89	68.5	66.0
13	70.22	69.17	66.66
15	67.89	69.5	68.0
17	69.33	69.17	68.67
19	69.22	70.0	66.66
21	68.44	68.83	67.0

Dari pengujian yang dilakukan, didapatkan hasil akurasi paling tinggi di rasio 70:30 pada K=13 dengan *accuracy* 70,2% , *precision* 70,2% , *recall* 70,2% , dan *f1 score* 70,2%. Berikut perbandingan grafik *confusion matrix* pada K=13.



Grafik 1. Perbandingan *Confusion Matrix* Data Opini – Non Opini

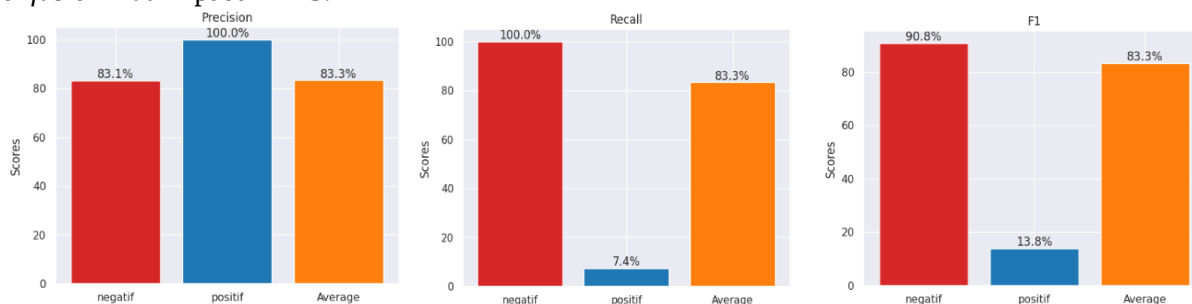
3.7.2 Pengujian Data Opini

Pengujian ini dilakukan dengan 1500 data opini yang terdiri dari 1170 kelas negatif dan 330 kelas positif. Berikut tabel pengujian yang telah dilakukan.

Tabel 10. Pengujian Data Opini

Nilai K	Accuracy		
	70:30	80:20	90:10
3	80.22	79.66	82.67
5	81.33	82.0	82.67
7	81.33	82.0	82.67
9	80.44	82.67	82.67
11	81.33	83.0	82.67
13	81.33	82.67	83.33
15	81.33	82.67	83.33
17	81.11	82.67	82.67
19	81.11	82.0	82.67
21	81.33	82.33	82.67

Dari pengujian yang dilakukan, didapatkan hasil akurasi paling tinggi di rasio 90:10 pada K=13 dan K=15 dengan *accuracy* 83,3% , *precision* 83,3%, *recall* 83,3%, dan *f1 score* 83,3%. Berikut perbandingan grafik *confusion matrix* pada K=13.



Grafik 2. Perbandingan *Confusion Matrix* Data Opini

3.7.3 Pengujian Data Balance

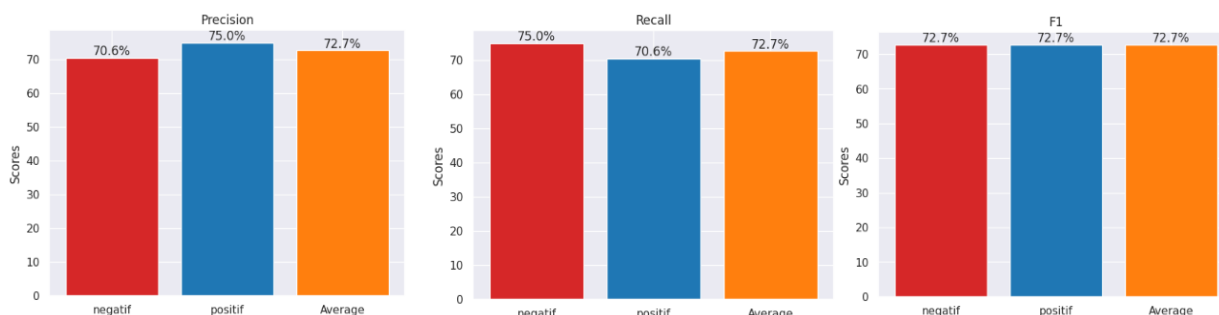


Pengujian ini dilakukan dengan 660 yang diambil dari data opini, terdiri dari 330 kelas negatif dan 330 kelas positif. Berikut tabel pengujian yang telah dilakukan.

Tabel 11. Pengujian Data *Balance*

Nilai K	Accuracy		
	70:30	80:20	90:10
3	61.11	65.16	65.16
5	60.60	65.16	63.63
7	58.09	57.58	66.66
9	62.12	62.87	72.72
11	61.61	61.37	68.19
13	57.08	62.87	62.12
15	58.59	59.84	60.60
17	58.59	60.60	62.12
19	57.08	57.58	65.15
21	56.57	59.84	65.15

Dari pengujian yang dilakukan, didapatkan hasil akurasi paling tinggi di rasio 90:10 pada K=9 dengan *accuracy* 72,7% , *precision* 72,7% , *recall* 72,7% , dan *f1 score* 72,7%. Berikut perbandingan grafik *confusion matrix* pada K=9.



Grafik 3. Perbandingan *Confusion Matrix* Data *Balance*

KESIMPULAN

Berdasarkan tahapan penelitian dan pengujian dengan metode K-NN dapat disimpulkan bahwa metode K-NN dapat digunakan dalam proses klasifikasi sentimen pada kasus kenaikan harga BBM di Twitter. Dengan dilakukan 3 pengujian yang berbeda, diperoleh pengujian data opini yang paling terbaik dengan nilai akurasi tertinggi dari proses klasifikasi yaitu 83,3% pada K=13 dan K=15 dengan *precision* 83,3% , *recall* 83,3%, dan *f1 score* 83,3% pada perbandingan data *training* dan *testing* 90:10. Dari pengujian menunjukan bahwa pengujian menggunakan data opini menghasilkan akurasi yang lebih tinggi, namun apabila jumlah data dikurangi maka akurasi yang dihasilkan lebih rendah. Saran untuk penelitian berikutnya dapat dikembangkan dengan menggunakan *feature weighting* dan *feature selection* lainnya.

DAFTAR PUSTAKA

- [1] N. Puspitasari, A. Tejawati, and F. Prakoso, "Estimasi Stok Penerimaan Bahan Bakar Minyak Menggunakan Metode Fuzzy Tsukamoto," *JRST (Jurnal Ris. Sains dan Teknol.*, vol. 3, no. 1, p. 9, 2019.
- [2] A. M. Iqbal and R. V. Salomo, "Analisis Kebijakan Pengenaan Pajak atas Bahan Bakar Minyak Bersubsidi Minyak Solar," *Transparansi J. Ilm. Ilmu Adm.*, vol. 1, no. 1, pp. 1–11, 2018.
- [3] A. Setiyowati, "Kenaikan Harga Bahan Bakar Minyak (BBM) Atas Nama Rakyat (Tinjauan Kritis Konsepsi Keadilan Sosial Ekonomi Ibnu Taimiyah)," *JESI (Jurnal Ekon. Syariah Indones.*, vol. 8, no. 2, p. 107, 2019.



- [4] N. A. Azmi, A. T. Fathani, D. P. Sadayi, I. Fitriani, and M. R. Adiyaksa, "Social Media Network Analysis (SNA): Identifikasi Komunikasi dan Penyebaran Informasi Melalui Media Sosial Twitter," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1422, 2021.
- [5] A. P. Rodrigues and N. N. Chiplunkar, "A New Big Data Approach for Topic Classification and Sentiment Analysis of Twitter Data," *Evol. Intell.*, vol. 15, no. 2, pp. 877–887, 2022.
- [6] I. A. Angreni, S. A. Adisasmita, M. I. Ramli, and S. Hamid, "Pengaruh Nilai K pada Metode K-Nearest Neighbor (KNN) terhadap Tingkat Akurasi Identifikasi Kerusakan Jalan," *Rekayasa Sipil*, vol. 7, no. 2, p. 63, 2019.
- [7] L. M. Sinaga, Sawaluddin, and S. Suwilo, "Analysis of classification and Naïve Bayes Algorithm K-Nearest Neighbor in Data Mining," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 725, no. 1, 2020.
- [8] R. M. Candra and A. Nanda Rozana, "Klasifikasi Komentar Bullying pada Instagram Menggunakan Metode K-Nearest Neighbor," *IT J. Res. Dev.*, vol. 5, no. 1, pp. 45–52, 2020.
- [9] D. Muhidin and A. Wibowo, "Perbandingan Kinerja Algoritma Support Vector Machine dan K-Nearest Neighbor terhadap Analisis Sentimen Kebijakan New Normal," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 5, no. 2, p. 153, 2020.
- [10] E. Indrayuni, A. Nurhadi, and D. A. Kristiyanti, "Implementasi Algoritma Naive Bayes, Support Vector Machine, dan K-Nearest Neighbors untuk Analisa Sentimen Aplikasi Halodoc," *Fakt. Exacta*, vol. 14, no. 2, p. 64, 2021.
- [11] R. Kosasih and A. Alberto, "Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF dan Algoritma K-Nearest Neighbor," *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 1, pp. 134–139, 2021.
- [12] R. Damarta, A. Hidayat, and A. S. Abdullah, "The Application of K-Nearest Neighbors Classifier for Sentiment Analysis of PT PLN (Persero) Twitter Account Service Quality," *J. Phys. Conf. Ser.*, vol. 1722, no. 1, 2021.
- [13] A. M. Pravina, "Sentiment Analysis of Delivery Service Opinions on Twitter Documents using K-Nearest Neighbor," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 996–1012, 2022.
- [14] R. R. A. Siregar, Z. U. Siregar, and R. Arianto, "Klasifikasi Sentiment Analysis pada Komentar Peserta Diklat Menggunakan Metode K-Nearest Neighbor," *Kilat*, vol. 8, no. 1, pp. 81–92, 2019.
- [15] A. Yoga Pratama, Y. Umaidah, and Voutama, "Analisis Sentimen Media Sosial Twitter dengan Algoritma K-Nearest Neighbor dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja)," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 897–910, 2021.
- [16] I. Romli, S. Prameswari R, and A. Z. Kamalia, "Sentiment Analysis About Large-Scale Social Restrictions in Social Media Twitter Using Algoritma K-Nearest Neighbor," *J. Online Inform.*, vol. 6, no. 1, p. 96, 2021.
- [17] G. K. Pati and E. Umar, "Analisis Sentimen Komentar Pengunjung terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier dan K- Nearest Neighbor," vol. 6, pp. 2309–2315, 2022.
- [18] P. K. Singh and S. Paul, "Deep Learning Approach for Negation Handling in Sentiment Analysis," *IEEE Access*, vol. 9, pp. 102579–102592, 2021, doi: 10.1109/ACCESS.2021.3095412.
- [19] A. Deviyanto and M. D. R. Wahyudi, "Penerapan Analisis Sentimen pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 3, no. 1, p. 1, 2018.
- [20] S. Ernawati and R. Wati, "Penerapan Algoritma K-Nearest Neighbors pada Analisis Sentimen Review Agen Travel," *J. Khatulistiwa Inform.*, vol. 6, no. 1, pp. 64–69, 2018.

