

Klasifikasi Keluhan Masyarakat Terhadap Sepeda Motor Listrik Dengan Menerapkan Algoritma Text Mining Dan Tf-Idf

Jupri Sinaga

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email: juprisinaga04@gmail.com
Email Penulis Korespondensi: juprisinaga04@gmail.com

Abstrak—Sepeda motor listrik merupakan kendaraan sepeda motor yang digerakkan oleh dinamo dan akumulator sehingga tidak menimbulkan emisi. Berbeda dengan sepeda motor konvensional (karburator dan injeksi) yang masih menggunakan bahan bakar bensin untuk penggerak dan menimbulkan polusi yang sangat tinggi. Dengan adanya media sosial saat ini memudahkan masyarakat mengungkapkan keluhannya lewat sosial media. Adapun masalah yang dirasakan masyarakat dalam penggunaan sepeda motor listrik yaitu masyarakat cenderung waswas baterai habis dan kesulitan mengatur pola pengisian daya, lalu kesulitan dalam menemukan stasiun pengisian baterai terdekat secara fast charging/ battery swapping jikaterjadi low battery. Hal ini menjadi penyebab berkurangnya minat masyarakat akan sepeda motor listrik. Oleh karena itu, klasifikasi dengan menerapkan algoritma Text Mining dan TF-IDF bertujuan untuk mengelompokkan masalah-masalah yang sedang terjadi didalam penggunaan sepeda motor listrik. Setelah dilakukan proses dengan menggunakan algoritma Text Mining dan TF-IDF, maka dari 20 data sampel menghasilkan data keluhan baterai yang lebih dominan dengan nilai bobot sebesar 67,004.

Kata Kunci: Klasifikasi, Sepeda Motor Listrik, Text Mining, TF-IDF

Abstract—Electric motorbikes are motorbikes that are driven by a dynamo and accumulator so they do not cause emissions. This is different from conventional motorbikes (carburetor and injection) which still use gasoline for propulsion and cause very high pollution. With the existence of social media, it is now easier for people to express their complaints via social media. The problems experienced by the public in using electric motorbikes are that people tend to be wary of the battery running out and have difficulty setting charging patterns, then have difficulty finding the nearest fast charging/battery swapping battery charging station if a low battery occurs. This is the cause of reduced public interest in electric motorbikes. Therefore, classification by applying the Text Mining and TF-IDF algorithms aims to group problems that are currently occurring in the use of electric motorbikes. After processing using the Text Mining and TF-IDF algorithms, the 20 sample data produced more dominant battery complaint data with a weight value of 67.004.

Keywords: Classification, Electric Motorcycle, Text Mining, TF-IDF

1. PENDAHULUAN

Sepeda motor listrik merupakan kendaraan sepeda motor yang digerakkan oleh dinamo dan akumulator sehingga tidak menimbulkan emisi. Berbeda dengan sepeda motor konvensional (karburator dan injeksi) yang masih menggunakan bahan bakar bensin untuk penggerak dan menimbulkan polusi yang sangat tinggi. Berdasarkan hasil penghitungan KPBB (Komite Penghapusan Bensin Bertimbang) sekitar 47% penyumbang polusi udara terbesar itu dari kendaraan bermotor. Sekitar 90 hidrokarbon dari kendaraan bermotor, 70% karbon monoksida dan 70% sulfur dioksida dari kendaraan bermotor. Namun demikian, masyarakat masih enggan menggunakan sepeda motor listrik karena dinilai masih baru dikalangan masyarakat dan belum familiar.

Banyaknya masalah sepeda motor listrik disebut sebagai faktor penyebab kurangnya daya tarik masyarakat untuk menggunakan sepeda motor listrik. Dengan adanya media sosial saat ini memudahkan masyarakat mengungkapkan keluhannya lewat sosial media. Keluhan-keluhan tersebut berupa baterai yang kurang memadai, design sepeda motor, subsidi, pengisian baterai dan masih banyak lagi. Dengan banyaknya keluhan yang terjadi sehingga kita perlu mengklasifikasikan keluhan tersebut dengan cara mengelompokkan semua keluhan yang di ungkapkan masyarakat lewat sosial media tersebut.

Klasifikasi merupakan kata serapan dari, classificatie, yang sendirinya berasal dari bahasa Prancis classification. Istilah ini menunjuk kepada sebuah metode untuk menyusun data secara sistematis atau menurut beberapa aturan atau kaidah yang telah ditetapkan. Di dalam KBBI, klasifikasi adalah penyusunan sistematis dalam kelompok atau golongan menurut kaidah atau standar yang di tetapkan. Maka klasifikasi yang dimaksud dalam penelitian ini adalah pengelompokan masalah-masalah yang sedang terjadi didalam penggunaan sepeda motor listrik. Adapun masalah yang dirasakan masyarakat dalam penggunaan sepeda motor listrik yaitu masyarakat cenderung waswas baterai habis dan kesulitan mengatur pola pengisian daya, lalu kesulitan dalam menemukan stasiun pengisian baterai terdekat secara fast charging/ battery swapping jikaterjadi low battery. Hal ini menjadi penyebab berkurangnya minat masyarakat akan sepeda motor listrik. Kementerian perhubungan mengungkapkan penjualan sepeda motor listrik dari tahun 2019 hingga tahun 2022 berada dikisaran 30.800 unit. Jumlah ini dinilai masih sangat minim jika dibandingkan dengan sepeda motor berbahan bakar bensin yang terjual mencapai 5 juta unit setiap tahunnya. Staf ahli utama menteri perhubungan bidang transportasi darat dan konektivitas Budi Setiyadi, menyampaikan minimnya penjualan sepeda motor listrik akibat dari faktor kepercayaan masyarakat yang masih rendah.

Text mining adalah salah satu teknologi yang mengelola data teks untuk memperoleh informasi secara otomatis. Dengan text mining, informasi baru bisa didapatkan melalui hasil analisis pada data teks semi terstruktur maupun tidak terstruktur (biasanya dalam jumlah besar). Hal ini tentu sangat membantu pekerjaan manusia seiring dengan semakin banyaknya data teks ataupun dokumen yang ada pada aplikasi web, aplikasi digital, maupun media sosial. Tentunya

data-data tersebut memiliki jumlah yang besar dan tidak terstruktur sehingga memerlukan waktu yang cukup lama untuk menganalisis informasi yang terdapat didalamnya. Text mining dikenal juga dengan istilah lain seperti Intelligent Text Analysis (ITA), Text Data Mining (TDM) atau Knowledge-Discovery in Text (KDT). Menurut Hearst, text mining diartikan sebagai penemuan informasi yang baru dan tidak diketahui sebelumnya oleh komputer dengan secara otomatis mengekstrak informasi dari sumber-sumber yang berbeda. Ditulis oleh Laili Cahyani S.Kom, M.Kom dengan judul “Aplikasi Text mining Dalam Bidang Pendidikan”[1]. Dalam buku yang ditulis oleh Amma, dkk dengan judul “Data Mining”, menjelaskan bahwa tujuan text mining adalah untuk mengolah dan menganalisis teks secara otomatis menggunakan teknik-teknik komputasi dan pembelajaran mesin. Setelah data-data tersebut dikumpulkan maka akan dilanjutkan dengan menghitung bobot setiap dokumen dengan algoritma TF-IDF[2]. Dalam buku yang ditulis oleh Dr. Ilmiyati Sari, s.si, M.si, Dr. Dina Indarti, s.si, m.si dengan judul praktek klasifikasi dan pemodelan topik dengan python, menjelaskan bahwa hasil preprocessing yang kemudian dilakukan pembobotan dengan TF-IDF di gunakan untuk mengklasifikasi perkara hukum kedalam empat kategori[3].

Penelitian yang dilakukan oleh Rudhi Ardi Sasmita, Achmad Zakki Falani dalam artikel yang berjudul pemanfaatan Algoritma TF/IDF Pada Sistem Informasi Ecomplaint Handling. Dalam penelitiannya disimpulkan bahwa algoritma TF-IDF (Term Frequency Inverse Document Frequency) dapat digunakan untuk mengukur tingkat similaritas dokumen dengan kata kunci sehingga memberi kemudahan bagi para pelanggan dalam mengajukan permasalahannya[4]. Pada penelitian yang dilakukan oleh Enda Esyudha Pratama pada penelitian tentang klasifikasi keterpakaian modul e-learning dengan menggunakan text mining dengan hasil pengujian terdiri dari 5 sampel modul dari setiap fakultas, dihasilkan bahwa kinerja algoritma dalam mengelompokkan kategori modul secara benar dengan persentase 88.9%, hal ini menyimpulkan bahwa algoritma bekerja dengan baik[5].

2. METODE PENELITIAN

2.1 Text Mining

Text mining adalah salah satu bidang khusus dari data mining. Text mining dapat didefinisikan sebagai proses pengumpulan informasi dimana pengguna berinteraksi dengan kumpulan dokumen dengan alat analisis yang merupakan bagian dari penambangan data dan yang termasuk adalah pengalokasian dokumen[6]. Tujuan text mining adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Cara yang digunakan dalam mempelajari suatu data teks, dengan terlebih dahulu menentukan fitur-fitur yang ada pada dokumen. Sebelum menentukan fitur-fitur yang mewakili, dibutuhkan tahap pra-proses yang dilakukan secara umum dalam teks mining, yaitu case folding, tokenizing, filtering, stemming, tagging dan analyzing.

2.2 TF-IDF (Term Frequency - Inverse Document Frequency)

TF-IDF (term Frequency-inverse document frequency) merupakan proses untuk melakukan transformasi data dari data tekstual ke dalam data numerik untuk dilakukan pembobotan pada tiap kata atau fitur. TF – IDF merupakan sebuah ukuran statistik yang digunakan untuk mengevaluasi seberapa penting sebuah kata didalam sebuah dokumen. Dibawah ini merupakan persamaan TF-IDF:

$$W_{dt} = tf_{dt} * idf_t = tf_{dt} * \log\left(\frac{N}{df_t}\right) \quad (1)$$

Keterangan:

W_{dt} = Nilai pembobotan TF-IDF atau W dari term t dikategorikan c

tf_{dt} = Jumlah kemunculan term t dalam dokumen d

N = Jumlah dokumen secara keseluruhan

df_t = Jumlah dokumen yang mengandung term t[7]

Term Frequency adalah cara untuk mencari bobot dari sebuah dokumen. Dimana akan dicari jumlah kemunculan term pada dokumen. Semakin besar kemunculan sebuah term, akan memengaruhi besarnya bobot dan nilai kesesuaian. Berikut adalah persamaan dari Term Frequency:

$$W(d, t) = TF(d, t) \quad (2)$$

Dimana $TF(d, t)$ merupakan frekuensi dari term t pada teks d.

Pendekatan IDF hampir sama dengan TF, yaitu menghitung frekuensi kemunculan suatu term. Namun, jika TF menghitung kemunculan suatu term hanya di satu dokumen teks, maka IDF menghitung kemunculan suatu term di keseluruhan corpus dokumen[8].

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, penulis menganalisa opini ataupun keluhan yang di sampaikan oleh masyarakat pada media sosial twitter tentang sepeda motor listrik dan mencoba menganalisa data yang didapatkan dari pencarian data twitter. Pada tahapan ini penulis melakukan penarikan data dengan menggunakan aplikasi Rapidminer. Kemudian data diolah dengan

menggunakan algoritma text mining melalui tahapan pra-processing seperti case folding, tokenizing, filtering, dan juga stemming dengan maksud pengolahan data dapat menjadi lebih efisien.

Setelah proses text mining dilakukan proses selanjutnya ialah dengan menerapkan algoritma TF-IDF untuk menentukan bobot dari data komentar yang dimiliki. TF-IDF (Term Frequency-Inverse Document Frequency) merupakan metode yang digunakan untuk mengukur pentingnya suatu term atau kata dalam sebuah dokumen yang besar. Setelah proses text mining dan juga tf-idf selesai dilakukan, penelitian ini juga menerapkan algoritma Naïve Bayes untuk mengklasifikasikan atau mengelompokkan keluhan tersebut kedalam beberapa kelas keluhan.

Dari penjelasan diatas dapat disimpulkan bahwa hasil dari pada penelitian ini dapat berupa pengelompokan suatu keluhan yang disampaikan oleh masyarakat pada media sosial twitter kedalam kelas-kelas keluhan. Sehingga diharapkan dapat memudahkan dalam mendapatkan informasi keluhan tentang sepeda motor listrik. Untuk mengolah kata hasil text mining dengan cara membandingkan terhadap data target.

3.1 Pengumpulan Data

Pengumpulan data merupakan tahapan yang berfungsi untuk mendapatkan data yang digunakan dalam penelitian. Jumlah data yang digunakan dalam penelitian ini adalah sebanyak 20 data sampel. Data awal dalam penelitian ini dapat dilihat pada tabel berikut:

Tabel 1. Data Awal Komentar Masyarakat Dari Twitter

No.	Komentar
1.	@sepedamotorlistrik, baterai sepeda motor listrik ini beneran bikin frustrasi! Cepat habis dan sulit diisi ulang. Harusnya lebih tahan lama dan efisien!
2.	Sedih banget, baterai sepeda motor listrikku ga kuat buat jarak jauh. Jadi harus sering-sering nyari stasiun pengisian daya. Gimana caranya bisa lebih tahan lama, ya?
3.	@sepedamotorlistrik, baterai sepeda motor listrik yang dulu awet, sekarang kok jadi kurang maksimal ya? Kenapa harus sering-sering nge-charge? Kurang puas deh!
4.	Beneran deh, baterai sepeda motor listrik ini nyebelin banget. Tiap kali diisi penuh, cuma tahan sebentar. Ga puas sama performanya!
5.	@sepedamotorlistrik, jarak tempuh sepeda motor listrik ini beneran bikin kecewa! Ga bisa ngelewat beberapa kilometer aja udah abis. Harusnya bisa lebih jauh dong!
6.	Penggunaan sepeda motor listrik jadi ribet gara-gara baterainya yang sering lemah. Harus mantengin indikator sambil khawatir baterai mati pas lagi di jalan. Fix bikin stres!
7.	Mau jalan sehari-hari pakai sepeda motor listrik, tapi kok baterainya cepat abis? Efisiensinya kurang banget nih! Harus sering-sering ngisi daya, capek!
8.	Ketika butuh baterai yang awet buat sepeda motor listrik, yang aku dapat cuma yang cepat habis. Belum jauh, udah minta diisi lagi. Kurang memuaskan!
9.	@sepedamotorlistrik, aku ngeluh dong. Baterai sepeda motor listrikku ini kok kurang awet ya? Belum jauh, udah kempes. Mesti bawa kabel charger kemana-mana.
10.	Mungkin ini keluhan umum, tapi baterai sepeda motor listrik ini bikin frustrasi banget. Ga awet dan sering abis saat dibutuhkan. Harusnya diperbaiki dong!
11.	@sepedamotorlistrik, kok harga sepeda motor listriknya mahal banget ya? Jadi pengen beralih ke alternatif lain yang lebih terjangkau. Harganya perlu dipertimbangkan ulang nih!
12.	Mau liburan pake sepeda motor listrik, tapi jarak tempuhnya sangat terbatas. Sama kayak jalan kaki! Kasian banget yang mau jalan jauh.
13.	Baterai sepeda motor listrikku udah ga bisa diandalkan. Udah berasa lowbat, eh mati tengah jalan! Aduh, jangan gitu dong.
14.	Kok harga sepeda motor listrik segini mahal ya? Rasanya kayak beli mobil aja. Sulit buat masyarakat biasa bisa afford.
15.	Ada yang merasakan rem sepeda motor listriknya agak berdecit? Kadang suaranya cukup mengganggu dan bikin was-was. Perlu perbaikan nih!
16.	Kecewa banget sama jarak tempuh sepeda motor listrik ini. Belum jauh udah minta diisi daya lagi. Buat perjalanan sehari-hari kurang memadai!
17.	Rem sepeda motor listrikku ini kurang ngegrip, jadi perlu tenaga ekstra buat berhenti. Padahal rem yang efektif itu penting buat keselamatan pengendara.
18.	@sepedamotorlistrik, seriusan deh, kenapa jarak tempuh sepeda motor listrik ini sependek itu? Ga nyaman buat perjalanan jauh.
19.	Rem sepeda motor listrik ini kurang responsif, agak bikin khawatir saat berkendara. Harusnya lebih baik dan bisa memberikan keamanan yang lebih.
20.	Gimana sih @sepedamotorlistrik, harga sepeda motor listrik ini bikin kantong jebol. Padahal mau berkontribusi ke lingkungan juga.

3.2 Penerapan Algoritma Text Mining

Pada penelitian ini, peneliti melakukan pengolahan data terhadap sample data yang digunakan dengan menerapkan algoritma text mining. Adapun tahapan algoritma text mining adalah sebagai berikut:

1. Case Folding

Tahap ini dilakukan untuk mengkonversi seluruh teks menjadi bentuk standart atau lowercase. Hasil pada tahap ini dapat dilihat pada tabel berikut:

Tabel 2. Tahap Case Folding

Komentar	Case Folding
@sepedamotorlistrik, baterai sepeda motor listrik ini beneran bikin frustrasi! Cepat habis dan sulit diisi ulang. Harusnya lebih tahan lama dan efisien!	baterai sepeda motor listrik ini beneran bikin frustrasi cepat habis dan sulit diisi ulang harusnya lebih tahan lama dan efisien

2. Tokenizing

Tahap ini dilakukan untuk melakukan pemotongan string input berdasarkan kata penyusunnya. Hasil tahapan ini dapat dilihat pada tabel berikut:

Tabel 3. Tahap Tokenizing

Komentar	Tokenizing
baterai sepeda motor listrik ini beneran bikin frustrasi cepat habis dan sulit diisi ulang harusnya lebih tahan lama dan efisien	[baterai] [sepeda] [motor] [listrik] [ini] [beneran] [bikin] [frustrasi] [cepat] [habis] [dan] [sulit] [diisi] [ulang] [harusnya] [lebih] [tahan] [lama] [dan] [efisien]

3. Filtering

Tahap ini bertujuan untuk mengambil kata-kata penting dengan menerapkan stopword by tala. Hasil tahap ini dapat dilihat pada tabel berikut:

Tabel 4. Tahap Filtering

Komentar	Filtering
baterai sepeda motor listrik ini beneran bikin frustrasi cepat habis dan sulit diisi ulang harusnya lebih tahan lama dan efisien	[baterai] [sepeda] [motor] [listrik] [ini] [beneran] [bikin] [frustrasi] [cepat] [habis] [dan] [sulit] [diisi] [ulang] [harusnya] [lebih] [tahan] [lama] [dan] [efisien]

4. Stemming

Tahapan ini dilakukan untuk mencari bentuk kata dasar dari tiap kata dari hasil filtering. Hasil pada tahap ini dapat dilihat pada tabel berikut:

Tabel 5. Tahap Stemming

Komentar	Stemming
baterai sepeda motor listrik ini beneran bikin frustrasi cepat habis dan sulit diisi ulang harusnya lebih tahan lama dan efisien	[baterai] [sepeda] [motor] [listrik] [bikin] [frustrasi] [cepat] [habis] [sulit] [isi] [ulang] [tahan] [efisien]

3.3 Data Resource

Data Resource digunakan sebagai keyword atau yang akan dijadikan sebagai kata kunci dalam proses hitung kemiripan yang berupa kata dasar. Kata kunci ini yang akan mewakili masing-masing kategori dokumen. Keyword atau kata kunci di peroleh dari perbandingan masing-masing isi dokumen, token yang memiliki nilai bobot yang paling tinggi akan dijadikan sebagai keyword. Adapun untuk menentukan bobot dari masing-masing kategori, penulis menggunakan alat bantu atau software yaitu Orange versi 3.35.0. Adapun langkah-langkah yang dilakukan penulis adalah:

1. Sebelum melakukan kerja dengan Orange, terlebih dahulu dipastikan untuk melakukan install terhadap package list yaitu dengang mengklik tools option lalu pilih add-on kemudian centang list "Text". Kegiatan ini berfungsi untuk mengaktifkan text mining pada Orange.
2. Kemudian, Import file dokumen ke dalam canvas atau layar kerja Orange. Penulis menggunakan file.csv.
3. Drag pilihan Corpus pada menu text mining. Sambungkan file kerja dengan Corpus.
4. Selanjutnya adalah melakukan pre-process text dengan melakukan drag menu pre-process text dan menghubungkan Corpus dengan pre-process text kemudian pada menu ini dapat dipilih proses apa saja yang diinginkan, pada bagian ini penulis hanya melakukan proses tokenizing dan stopword dikarenakan penulis ingin mengambil token dari setiap kalimat. Hasil dari pre-process text dapat dilihat pada table hasil.

Adapun keyword yang diperoleh dari hasil melakukan pembobotan dengan software Orange dapat dilihat pada tabel berikut:

Tabel 7. Data Resource

No	Jenis Dokumen	Kata Kunci
----	---------------	------------

1	Baterai	Baterai, Kapasitas, Habis, daya, Menurun, isi, Guna, Tahan, Performa, Awet, Pakai, Penuh, Lemah, indikator, Lowbat
2	Jarak Tempuh	Jarak, Perjalanan, Tempuh, isi, Daya, Pendek, kehabisan, Stasiun, Jelajah, Kapasitas, Mobilitas, Mampu, Capai.
3	Rem	Rem, Keluhan, Terasa, Reponsif, Nyaman, Licin, Khawatir, Respon, Longgar, Performa, Tajam, Berdecit, Terkunci.
4	Harga	Lebih, Harga, Terjangkau, Opsi, Tinggi, Pembiayaan, Fleksibel, Jangkauan, Kebanyakan, Mahal, Besar, Penawaran, Bayar, Uang, Anggaran.

3.4 Penerapan TF-IDF

TF-IDF (Term Frequency – Invers Document Frequency) merupakan metode yang digunakan dalam pengembalian informasi dan pemrosesan teks. Metode ini digunakan untuk memberikan bobot pada kata-kata dalam sebuah dokumen di dalam kumpulan dokumen yang lebih besar. Frekuensi kemunculan sebuah kata dalam sebuah dokumen disebut term frequency (TF) dan inverse frekuensi dokumen yang mengandung kata yang disebut inverse document frequency (IDF). Tujuan Dari algoritma ini adalah unrtuk melihat seberapa penting kata (term) tersebut di dalam dokumen.

Pada tahap ini Pencocokan term dilakukan terhadap dokumen sampel, yaitu dengan menghitung kemunculan tiap term pada setiap dokumen, jika tedapat dokumen yang memiliki term dalam isi dokumen tersebut maka akan diberi angka 1 dan jika tidak maka akan diberi nilai angka 0, kegiatan ini adalah salah satu bentuk dari TF binary. Berikut adalah hasil dari pencocokan term pada tiap dokumen. Adapun term dibagi menjadi 4 kategori term yaitu term1 merupakan kata keyword baterai, term2 merupakan keyword jarak tempuh, term3 merupakan keyword rem, dan term4 merupakan keyword harga. Untuk hasil penentuan Term Frequency (TF) nya dapat dilihat pada tabel berikut:

Tabel 8. Term Frequency dan Document Frequency Pada Term1

No.	term1	D1	D2	D3	D4	D5	D6	...	D18	D19	D20	DF
1	Baterai	1	1	1	1	0	2	...	0	0	0	10
2	Kapasitas	0	0	0	0	0	0	...	0	0	0	0
3	Habis	1	0	0	0	1	0	...	0	0	0	5
4	Daya	0	1	0	0	0	0	...	0	0	0	3
5	Menurun	0	0	0	0	0	0	...	0	0	0	0
6	isi	1	1	0	1	0	0	...	0	0	0	6
7	Guna	0	0	0	0	0	1	...	0	0	0	1
8	Tahan	1	1	0	1	0	0	...	0	0	0	3
9	Performa	0	0	0	1	0	0	...	0	0	0	1
10	Awet	0	0	1	0	0	0	...	0	0	0	4
11	Pakai	0	0	0	0	0	0	...	0	0	0	2
12	Penuh	0	0	0	1	0	0	...	0	0	0	1
13	Lemah	0	0	0	0	0	1	...	0	0	0	1
14	Indikator	0	0	0	0	0	1	...	0	0	0	1
15	Lowbat	0	0	0	0	0	0	...	0	0	0	1

Tabel 9. Inverse Document Frequency (IDF) Term1

Term1	DF	N	IDF
			$\log(n/df) + 1$
Baterai	10	20	1,3010
Kapasitas	0	20	0
Habis	5	20	1,6020
Daya	3	20	1,8239
Menurun	0	20	0
Pengisian	6	20	1,5228
Guna	1	20	2,3010
Tahan	3	20	1,8239
Performa	1	20	2,3010
Awet	4	20	1,6989
Pakai	2	20	2
Penuh	1	20	2,30102
Lemah	1	20	2,3010
Indikator	1	20	2,3010
Lowbat	1	20	2,3010

Tabel 10. Keseluruhan TF-IDF Term1

Term1	D1	D2	D3	D4	D5	...	D20
Baterai	1,3010	1,3010	1,3010	1,3010	0	...	0
Kapasitas	0	0	0	0	0	...	0
Habis	1,6020	0	0	0	1,6020	...	0
Daya	0	1,8239	0	0	0	...	0
Menurun	0	0	0	0	0	...	0
isi	1,5228	1,5228	0	1,5228	0	...	0
Guna	0	0	0	0	0	...	0
Tahan	1,8239	1,8239	0	1,8239	0	...	0
Performa	0	0	0	2,3010	0	...	0
Awet	0	0	1,6989	0	0	...	0
Pakai	0	0	0	0	0	...	0
Penuh	0	0	0	2,3010	0	...	0
Lemah	0	0	0	0	0	...	0
Indikator	0	0	0	0	0	...	0
Lowbat	0	0	0	0	0	...	0
Nilai	6,2498	6,4717	3	9,2498	1,6020	...	0
Jumlah bobot Keseluruhan			67,0044				

3.5 Hasil

Setelah dilakukan pembobotan terhadap keseluruhan keluhan maka didapatkan hasil dari masing- masing keluhan yaitu sebagai berikut:

Tabel 11. Hasil Klasifikasi TF-IDF

No	Term	Jenis Data Keluhan	Jumlah Hasil Bobot dari Data Sampel
1	Term1	Baterai	67,0044
2	Term2	Jarak Tempuh	52,6506
3	Term3	Rem	26,8007
4	Term4	Harga	13.5966

Berdasarkan tabel di atas maka hasil yang diperoleh dari 20 data sampel yang diolah dengan menerapkan algoritma Text Mining dan TF-IDF (Term Frequency-Inverse Document Frequency), menghasilkan data keluhan baterai yang lebih dominan dengan nilai bobot sebesar 67,0044.

4. KESIMPULAN

Berdasarkan pengujian 20 sampel data Twitter keluhan masyarakat terhadap sepeda motor listrik menggunakan bahasa pemrograman Python dengan memanfaatkan Google Colaboratory, bahwa 20 sampel data Twitter dapat diproses secara baik oleh bahasa pemrograman Python. Bahasa pemrograman Python mampu memproses dari tahapan-tahapan yang ada pada algoritma text mining dan Term Frequency-Inverse Document Frequency (IDF), sehingga hasil klasifikasi dengan penerapan algoritma text mining dan Term Frequency-Inverse Document Frequency (IDF) dengan hasil pengujian menggunakan sistem mendapatkan hasil yang sama yaitu menjadi angka klasifikasi keluhan tertinggi dengan nilai 67.004.

REFERENCES

- [1] S. K. M. K. Laili Cahyani, APLIKASI TEXT MINING DI BIDANG PENDIDIKAN. CV Literasi Nusantara Abadi, 2023.
- [2] W. S et al., Data Mining. Global Eksekutif Teknologi, 2023. [Online]. Available: <https://books.google.co.id/books?id=xmqvEAAAQBAJ>
- [3] S. S. M. S. Dr. Ilmiyati Sari and S. S. M. S. Dr. Dina Indarti, TEXT MINING: Praktek Klasifikasi dan Pemodelan Topik dengan Phyton. uwais inspirasi indonesia, 2023.
- [4] R. A. Sasmita and A. Z. Falani, "PEMANFAATAN ALGORITMA TF/IDF PADA SISTEM INFORMASI ECOMPLAINT HANDLING".
- [5] E. E. Pratama, "JIP (Jurnal Informatika Polinema) KLASIFIKASI KETERPAKAIAN MODUL E-LEARNING BERBASIS MOODLE DENGAN PENDEKATAN TEXT MINING".
- [6] H. Sari, G. L. Ginting, and T. Zebua, "Penerapan Algoritma Text Mining dan TF-IDF Untuk Pengelompokan Topik Skripsi Pada Aplikasi Repository STMIK Budi Darma," Terap. Inform. Nusant., vol. 2, no. 7, pp. 414–432, 2021.
- [7] Y. S. Hartini et al., Prosiding Seminar Nasional Sanata Dharma Berbagi "Pengembangan, Penerapan Dan Pendidikan 'Sains Dan Teknologi' Pasca Pandemi." Sanata Dharma University Press, 2022.
- [8] D. Marta, G. L. Ginting, and A. M. H. Sihite, "Deteksi Berita Palsu Tentang Vaksinasi Covid-19 Dengan Menggunakan Text Mining Dan Algoritma Cosine Similarity," KOMIK (Konferensi ..., vol. 6, no. November, pp. 129–139, 2023, doi: 10.30865/komik.v6i1.5738.