

# Analisis Sentimen Twitter Menggunakan Algoritma Naive Bayes Pada Kasus Koperasi Simpan Pinjam Indosurya

Ornensya Br Sembiring<sup>1</sup>, Elma Rulfin Tiara Kiu<sup>2</sup> and Khairun Nisa Meiah Ngafidin<sup>3</sup>

<sup>1</sup> Fakultas Informatika, Institut Teknologi Telkom Purwokerto ; ornensyasembiring@gmail.com

<sup>2</sup> Fakultas Informatika, Institut Teknologi Telkom Purwokerto ; elmakiu333@gmail.com

<sup>3</sup> Fakultas Informatika, Institut Teknologi Telkom Purwokerto ; nisa@ittelkom-pwt.ac.id

## Info Artikel:

Dikirim: 04 Februari 2023

Direvisi: 10 Mei 2023

Diterima: 26 Mei 2023

**Intisari:** Kasus KSP Indosurya kembali mencuri perhatian saat ini karena terduga pelaku selaku pendiri KSP Indosurya divonis bebas oleh Hakim Ketua Pengadilan Negeri Jakarta Barat atas segala dakwaan pada dengan pernyataan bahwa terdakwa terbukti melakukan perbuatan yang didakwakan tetapi bukan merupakan tindak pidana, melakukan perkara perdata yang kemudian menyita perhatian publik serta mengundang reaksi masyarakat diberbagai media sosial bahkan sempat menjadi trending topik dimedia sosial twitter Indonesia. Oleh karena itu dilakukan penelitian ini guna mengetahui dan mengklasifikasikan sentimen masyarakat pada kasua KSP Indosurya menggunakan metode Naive bayes. Dimana hasil tertinggi didapatkan oleh sentimen negative, sehingga didapatkan kesimpulan sentimen publik terhadap topik kasus KSP Indosurya di twitter Indonesia adalah negative dengan jumlah parameter 1200 berdasarkan analisis menggunakan metode Naive bayes.

**Kata Kunci:** Analisis sentimen; KSP Indosurya; Naive bayes; R Studio; Twitter.

## 1. Pendahuluan

Seperti yang kita ketahui dampak dari wabah virus Covid-19 menyerang segala aspek kehidupan masyarakat, bahkan di era pandemi banyak muncul kasus kopearasi simpan pinjam yang gagal membayar anggotanya kerana terkena dampak Covid-19. Salah satunya adalah Koperasi Simpan Pinjam Indosurya (KSP). KSP Indosurya merupakan Koperasi Simpan Pinjam yang berdiri berdasarkan akta pendirian Nomor 84 tertanggal 27 September 2012 oleh pendirinya Henry Surya, Mamike Hardianti, Surya Effendy, Mulyadi, Sonia yang kemudian berhasil membangun serta memiliki sebanyak 184 cabang pada 30 provinsi di Indonesia dalam kurun waktu tahun 2012 sampai dengan 2017, dengan total karyawan sebanyak 1000 orang [1].

Kasus KSP Indosurya bermula terjadi pada tahun 2020 dimana koperasi ini menawarkan beragam produk dengan iming-iming imbal hasil tinggi bagi anggota koperasi di antaranya deposito dengan bunga 9- 12% per tahun dimana jumlahnya lebih besar dari bunga deposito perbankan sebesar 5%-7% pada periode yang sama, akan tetapi koperasi ini tidak mampu membayar sehingga membuat para korban merasa rugi [2]. Kasus ini kemudian kembali mencuri perhatian saat ini karena terduga pelaku Henry Surya selaku pendiri KSP Indosurya divonis bebas oleh Hakim Ketua Pengadilan Negeri Jakarta Barat Syafrudin Ainor atas segala dakwaan pada Selasa, (24/1/2023) dengan pernyataan bahwa terdakwa Henry Surya terbukti melakukan perbuatan yang didakwakan tetapi bukan merupakan tindak pidana, melakukan perkara perdata yang kemudian menyita perhatian publik serta mengundang reaksi masyarakat diberbagai media sosial bahkan sempat menjadi trending topik dimedia sosial twitter Indonesia (Mentari Puspadini, CNBC Indonesia, 25 January 2023 19:21). Alasan mengapa dilakukan penelitian ini dikarenakan kasus KSP Indosurya sangat menyita perhatian publik serta mengundang reaksi masyarakat diberbagai media sosial bahkan sempat menjadi trending topik dimedia sosial twitter Indonesia dikarenakan terduga pelaku yang divonis bebas sehingga mendapat banyak kecaman dari berbagai pihak.

Oleh karena itu dilakukan penelitian ini guna mengetahui dan mengklasifikasikan sentimen masyarakat pada kasua KSP Indosurya menggunakan metode Naive bayes. Analisis sentiment sendiri merupakan sebuah ilmu bidang komputasi yang mempelajari tentang opini publik terhadap suatu topik dimana didalamnya terdapat proses mengelompokan data teks yang didalamnya mengandung opini baik itu opini positif, negatif maupun netral. Tujuannya

adalah untuk mengetahui pendapat sekelompok orang/publik terhadap suatu topik, produk, jasa maupun instansi tertentu dimana opini tersebut dapat berupa opini positif, opini negatif maupun opini netral [3].

Alasan mengapa penelitian ini menggunakan metode naïve bayes karena didasarkan pada jurnal penelitian yang ditulis oleh Xhemali, Daniela, Chris J. Hinde, dan Roger G. Stone dengan judul "Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages.", dimana pada penelitian membuktikan bahwa metode klasifikasi naïve bayes memiliki tingkat akurasi yang lebih baik dibandingkan metode klasifikasi lainnya [4]. Metode naïve bayes juga memiliki performa yang lebih baik saat melakukan pengujian tipe data kategori dibandingkan dengan metode klasifikasi lainnya, memiliki potensi kerja yang baik dan semua atribut bersifat independen atau tidak saling ketergantungan pada nilai variabel kelas yang diberikan [5].

## 2. Tinjauan Pustaka

Dalam melakukan penelitian ini dilakukan studi literatur terhadap jurnal-jurnal yang memiliki tema yang sama dengan penelitian yang akan dilakukan yang dimana dalam hal ini digunakan sebagai sumber acuan penelitian ini. Penelitian sebelumnya pernah dilakukan oleh Mila Kartika dkk, dengan judul "Analisis Sentimen Dampak Covid-19 Terhadap Pembatalan Keberangkatan Ibadah Haji Pada Tahun 2020.", dimana pada penelitian tersebut mereka menganalisis sentimen data dari Twitter menggunakan 3 metode klasifikasi yaitu metode klasifikasi KnN, Random Forest dan Naïve bayes pada penelitian tersebut menghasilkan nilai akurasi KnN sebesar 0,507, Random forest 0,531 dan Naïve bayes 0,532. Dari ke 3 metode yang digunakan tersebut maka metode yang paling unggul adalah metode Naïve bayes karena metode tersebut menggunakan probabilitas dan statistik [4].

Penelitian yang lain dilakukan oleh Muhammad Ridho Fauzi dkk yang berjudul "Analisis Big Data Sentimen Konsumen UMKM Sektor Kuliner Menggunakan Multi-Label K-Nearest Neighbor.", dimana pada penelitian tersebut mereka menganalisis sentimen konsumen terhadap data ulasan UMKM sektor Kuliner dengan menggunakan metode DataMining. Hasil dari penelitian tersebut adalah untuk mengetahui insight dari hasil analisis sentimen konsumen UMKM dan dapat membantu untuk pengambilan sebuah keputusan bisnis dalam meningkatkan daya saingnya terhadap kompetitor sektor kuliner yang berada di Kota Bogor [6].

Penelitian yang lain juga dilakukan oleh Mujaddid Izzul Fikri dkk yaitu "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter.", dimana pada penelitian tersebut mereka mengklasifikasi tweets yang berisi sentimen masyarakat mengenai Universitas Muhammadiyah Malang (UMM) menggunakan metode klasifikasi Naïve Bayes dan Support Vector Machine(SVM) [7].

Berdasarkan uraian diatas, pada penelitian ini berfokus pada analisis sentimen publik terhadap kasus KSP Indosurya, dimana data sentimen masyarakat diambil menggunakan metode scarping dari Twitter kemudian diklasifikasi menggunakan metode naïve bayes. Alasan memilih metode naïve bayes dibandingkan dengan metode klasifikasi lainnya adalah karena metode naïve bayes lebih unggul dibandingkan metode penambangan lain dalam hal pengujian tipe data kategori, tipe atribut independen dan performa lebih baik dibandingkan metode lainnya [8].

### 2.1. Naïve Bayes

Naïve Bayes merupakan metode klasifikasi yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, dimana metode ini memiliki banyak keunggulan salah satunya dapat digunakan untuk memprediksi probabilitas serta terbukti memiliki tingkat akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database walaupun menggunakan data yang sangat besar [9]. Metode Naive Bayes hanya membutuhkan sedikit data (data pelatihan dalam menentukan estimasi parameter dalam tindakan proses klasifikasi. Metode Naïve Bayes bekerja jauh lebih baik dalam situasi nyata yang kompleks seperti yang diharapkan [10]. Persamaan teorema Bayes adalah sebagai berikut :

$$P(H|X) = \frac{P(H|X)P(H)}{P(X)} \quad [11]$$

Keterangan :

X: Data dengan class yang belum diketahui

H: Hipotesis data merupakan suatu class spesifik

P(H/X): Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$ : Probabilitas hipotesis H (prior probabilitas)

$P(X/H)$ : Probabilitas X berdasarkan kondisi pada hipotesis

$H P(X)$ : Probabilitas X

## 2.2. Twitter

Twitter merupakan media sosial yang menyediakan layanan komunikasi daring dimana pada aplikasi ini pengguna dapat menyampaikan opini secara bebas kapan saja dan dimana saja yang dikenal dengan istilah Tweet, hal ini menjadi kelebihan Twitter karena mampu menyebarkan sebuah informasi secara cepat yang kemudian menjadi sebuah topik pembahasan para pengguna Twitter lainnya yang dikenal dengan istilah Trending Topik [12].

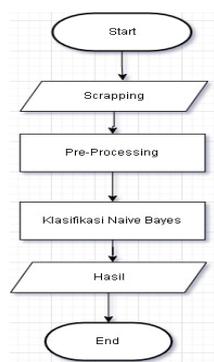
## 2.3. Analisis Sentimen

Analisis sentimen merupakan metode penambangan data yang berfungsi untuk mengekstrak data teks komentar yang baik bersifat opini, sentimen maupun emosi sehingga dapat diketahui kecenderungan opini seseorang terhadap sebuah masalah atau objek [4]. Analisis sentimen juga dapat diartikan sebagai aktivitas penambangan data/opini karena didalamnya terdapat aktivitas pengumpulan data opini yang bersifat positif atau negatif. Dalam serangkaian aktivitas analisis sentimen yang dilakukan dimulai dari menganalisis, mengolah, dan menghasilkan data bersifat opini dalam entitas, dimana pada aktifitas ini sering digunakan untuk mengetahui opini dari sebuah layanan, produk, orang, fenomena, atau topik tertentu. Pada analisis ini data yang ingin diolah dapat diambil dari ulasan teks, forum komunitas, tweet atau blog setelah itu dilakukan preprosesing data, tokenisasi, stopword, penghapusan, derivasi, deteksi emosi, dan peringkat sentimeter [5].

## 3. Metode Penelitian Pembahasan

Dalam penelitian ini dilakukan pengambilan data Twitter di Indonesia yang khusus membahas respon dari pengguna Twitter di Indonesia mengenai kasus Koperasi Simpan Pinjam. Data diambil menggunakan website Jupyter Notebook kemudian dilakukan filter data dan preprocessing text. Setelah itu akan diketahui data mana saja yang memiliki sentimen positif maupun sentimen negatif yang kemudian siap dianalisis dengan menggunakan website Rstudio dan dianalisis data tersebut menggunakan metode Analisis sentimen.

Kemudian dilanjutkan dengan melakukan klasifikasi menggunakan metode Algoritma naïve bayes, dimana algoritma ini menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes[4]. Metode ini merupakan metode pengklasifikasian probabilistik sederhana dimana dilamnya dilakukan perhitungan sekumpulan probabilitas dan kombinasi nilai dari dataset yang diberikan untuk menentukan probabilitas hasil [13] - [14].



Gambar 1. Diagram alir penelitian

### 3.1. Data Crawling/Scrapping

Merupakan proses pengambilan data mentah menggunakan tools Jupyter Notebook yang berkaitan dengan topik penelitian yang dipilih dalam hal ini Keyword yang dipakai ialah "KSP Indosurya". Pengambilan data dari tools Jupyter Notebook menggunakan library Twitter dimana proses cara kerja program dimulai dari mencari kata kunci kemudian kemudian setelah didapatkan data akan disimpan atau diunduh sesuai dengan ekstensi yang diinginkan. Dalam penelitian ini data diambil jenis ekstensi CSV.

### 3.2. Pre-Processing

Pre-procesing merupakan tahap lanjutan dari tahap sebelumnya dimana pada tahap ini terdapat proses menghilangkan atau menghapuskan beberapa atribut yang terdapat pada tabel data yang diambil berdasarkan proses sebelumnya contohnya seperti data tweet yang tidak lengkap, data berlebihan (double data) maupun format data yang tidak sesuai yang berpotensi menurunkan kinerja proses penambangan selanjutnya. Setelah itu dilanjutkan dengan proses normalisasi data yang didalamnya terdapat beberapa sub proses yakni filtering (menghilangkan tanda baca), stopword removal (menghapus karakter atau atribut yang tidak memiliki atri/tidak diperlukan), stemming (pembentukan kata dasar guna menghindari error pada performa tambang data), tokenization (penghapusan tanda baca pada data sehingga menghasilkan kumpulan kata dari data yang ada).

### 3.3. Klasifikasi Sentimen

Dalam mengklasifikasi sentimen pada penelitian ini menggunakan tools RStudio dengan library syuzhet untuk menampilkan jenis-jenis emosi yg didalamnya menggunakan fungsi `get_nrc_sentiment` untuk mengakumulasi 7 kategori emosi berdasarkan data bersih dari proses Pre-Processing yang dijabarkan sebagai berikut :

1. Anger, merupakan tipe sentimen yang memiliki opini cenderung ke arah negatif.
2. Disgust, merupakan tipe sentimen yang memiliki opini cenderung ke arah negatif yang membedakannya dengan sentimen Anger ialah levelnya berbeda.
3. Fear, merupakan tipe sentimen yang memiliki opini cenderung diantar negatif dan positif.
4. Joy, merupakan tipe sentimen yang memiliki opini cenderung ke arah positif sedikit dengan negatif. dan netral
5. Mixed, merupakan tipe sentimen yang memiliki opini cenderung ke arah positif sedikit dengan negatif. dan netral yang membedakannya dengan sentimen Joy ialah levelnya berbeda.
6. Sadness, merupakan tipe sentimen yang memiliki opini cenderung ke arah negatif sedikit dengan netral.
7. Surprised, merupakan tipe sentimen yang memiliki opini cenderung cenderung kearah negatif sedikit dengan positif.

## 4. Metode dan Hasil Penelitian

Berikut merupakan hasil implementasi dari alur penelitian dimulai dari tahap Crawling Data dilanjutkan dengan tahap Pre-Processing dan diakhiri dengan tahap Klasifikasi Sentimen yang dijabarkan sebagai berikut :Pembahasan merupakan bagian terpenting dari naskah publikasi. harus mengandung hasil-hasil simulasi atau pengukuran sebagai validasi metode. Pembahasan dapat berupa tabel hasil, narasi yang didapat dari perhitungan suatu rumus maupun prosentase dari grafik perhitungan.

### 4.1. Tahap Crawling Data

Berikut merupakan tampilan dari hasil pengambilan data tweet dari Twitter mengenai indosurya yang sudah diupload di jupyter-notebook. Data yang diambil terdiri dari 3 bagian yaitu bagian *datetime* yang berisi data waktu tweet yang diambil diupload di Twitter, kemudian data *username* yang berisi data *username* pemilik akun dari data tweet yang diambil, dan yang terakhir data *content* yang berisi data keseluruhan tweet yang diambil berdasarkan *datetime* dan *username*.

	<i>datetime</i>	<i>username</i>	<i>content</i>
0	2023-01-28 08:36:37+00:00	DonnySyalendra	"saudara sekalian sore ini kami mengadakan rap...
1	2023-01-28 08:35:56+00:00	DonnySyalendra	"saudara sekalian sore ini kami mengadakan rap...
2	2023-01-28 08:35:11+00:00	DonnySyalendra	'saudara sekalian sore ini kami mengadakan rap...
3	2023-01-28 08:34:38+00:00	evandirgan	@yellowlesskin Baju 3 warna, kumpul bawa palu...
4	2023-01-28 08:34:02+00:00	DonnySyalendra	saudara sekalian sore ini kami mengadakan rapa...
5	2023-01-28 08:33:28+00:00	Cakrabirawa201	@AdiputeriC @mohmahfudmd Lanjutkan pak sampai...
6	2023-01-28 08:29:42+00:00	DalilaApsarian1	"saudara sekalian sore ini kami mengadakan rap...
7	2023-01-28 08:29:18+00:00	evandirgan	@abu_waras Hakim nya kebanyakan wara wiri jadi...
8	2023-01-28 08:28:55+00:00	DalilaApsarian1	saudara sekalian sore ini kami mengadakan rapa...
9	2023-01-28 08:28:07+00:00	DalilaApsarian1	"saudara sekalian sore ini kami mengadakan rap...
10	2023-01-28 08:27:05+00:00	DalilaApsarian1	'saudara sekalian sore ini kami mengadakan rap...
11	2023-01-28 08:25:16+00:00	dndtirani	Hakim Lepas Terdakwa KSP Indosurya, Pakar: Bis...

Gambar 2. Tampilan data tweet

Berikut merupakan tampilan proses selanjutnya dimana pada proses ini menampilkan kembali data setelah dilakukan pembersihan pada data yang berada dalam bagain *username* diaman didalamnya dilakukan proses pembersihan tag *username* (@) pada bagian *username* selaku bagian yang menampung data akun pengguna.

	<b>datetime</b>	<b>username</b>	<b>content</b>
0	2023-01-28 08:36:37+00:00	DonnySyalendra	"saudara sekalian sore ini kami mengadakan rap...
1	2023-01-28 08:35:56+00:00	DonnySyalendra	"saudara sekalian sore ini kami mengadakan rap...
2	2023-01-28 08:35:11+00:00	DonnySyalendra	'saudara sekalian sore ini kami mengadakan rap...
3	2023-01-28 08:34:38+00:00	evandirgan	baju 3 warna, kumpul bawa palu, ujung - ujung ...
4	2023-01-28 08:34:02+00:00	DonnySyalendra	saudara sekalian sore ini kami mengadakan rapa...
5	2023-01-28 08:33:28+00:00	Cakrabirawa201	lanjutkan pak sampaikan jarit hati nuraninya p...
6	2023-01-28 08:29:42+00:00	DallilaApsarian1	"saudara sekalian sore ini kami mengadakan rap...
7	2023-01-28 08:29:18+00:00	evandirgan	hakim nya kebanyakan wara wiri jadinya masuk a...
8	2023-01-28 08:28:55+00:00	DallilaApsarian1	saudara sekalian sore ini kami mengadakan rapa...
9	2023-01-28 08:28:07+00:00	DallilaApsarian1	"saudara sekalian sore ini kami mengadakan rap...

**Gambar 3.** Tampilan data setelah dibersihkan *username* dan huruf kapital

#### 4.2. Tahap Pre-Processing

Berikut tampilan data hasil dari Case Folding dimana pada tahap ini dilakukan proses pembersihan karakter yang tadinya huruf kapital menjadi huruf kecil. Pembersihan yang dilakukan ini hanya dilakuakn bada data bagian content sedangkan pada bagain datetime dan username tidak dilakukan proses Case Folding ini.

```
Case Folding Result :
0 "saudara sekalian sore ini kami mengadakan rap...
1 "saudara sekalian sore ini kami mengadakan rap...
2 "saudara sekalian sore ini kami mengadakan rap...
3 baju 3 warna, kumpul bawa palu, ujung - ujung ...
4 saudara sekalian sore ini kami mengadakan rapa...
5 lanjutkan pak sampaikan jarit hati nuraninya p...
6 "saudara sekalian sore ini kami mengadakan rap...
7 hakim nya kebanyakan wara wiri jadinya masuk a...
8 saudara sekalian sore ini kami mengadakan rapa...
9 "saudara sekalian sore ini kami mengadakan rap...
Name: content, dtype: object
```

**Gambar 4.** Hasil *case folding*

Berikut merupakan tampilan data dari hasil pembersihan *content* yang sama sehingga tidak terdapat tweet/komentar yang serupa sehingga data lebih bersih dan tidak terjadi pendoublean data.

	<b>datetime</b>	<b>username</b>	<b>content</b>
0	2023-01-28 08:36:37+00:00	DonnySyalendra	saudara sekalian sore ini kami mengadakan rapa...
1	2023-01-28 08:35:56+00:00	DonnySyalendra	saudara sekalian sore ini kami mengadakan rapa...
3	2023-01-28 08:34:38+00:00	evandirgan	baju warna kumpul bawa palu ujung ujung nya pu...
5	2023-01-28 08:33:28+00:00	Cakrabirawa201	lanjutkan pak sampaikan jarit hati nuraninya p...
7	2023-01-28 08:29:18+00:00	evandirgan	hakim nya kebanyakan wara wiri jadinya masuk a...
...	...	...	...
994	2022-12-19 03:28:38+00:00	NinukHartomo	kalau indosurya emang bangke bgt mas boed saya...
995	2022-12-19 02:07:56+00:00	KompasData	total kerugian korban ksp indosurya disebut me...
997	2022-12-19 00:54:06+00:00	MediaKormando	ratusan korban ksp indosurya menangis minta ga...
998	2022-12-18 23:08:59+00:00	inilahrakyat	yg mengawasi ksp indosurya bukan ojk perlu dit...
999	2022-12-18 23:06:12+00:00	hdsambodo	misalnya gugatan dikabulka infosurya divonis b...

**Gambar 5.** Hasil data yang sudah dibersihkan dari konten duplikat

Berikut merupakan tampilan hasil *stopword* dari data tweet yang diambil yang dimana proses ini dilakukan untuk menghapus kata-kata yang tidak memiliki makna.

```
HASIL STOPWORDS :
0 [saudara, sore, mengadakan, rapat, kordinasi, ...
1 [saudara, sore, mengadakan, rapat, kordinasi, ...
3 [baju, warna, kumpul, bawa, palu, ujung, ujung...
5 [lanjutkan, jarit, hati, nuraninya, korban, in...
7 [hakim, kebanyakan, wara, wiri, masuk, angin, ...
11 [hakim, lepas, terdakwa, ksp, indosurya, pakar...
15 [dukung, md, naikkn, sampe, selesai, hakim, gi...
18 [korban, mengajukan, gugatan, pmh, perdata, pe...
21 [sampe, clear]
23 [dukung, upaya, pemerintah, kejagung, upaya, h...
24 [gerak, cepat, kawal, trus, indosurya, sampe, ...
25 [udah, loh, dakwaan, masuk, pencucian, uang, n...
26 [sampe, turun, tangan, bs, biarkan, indosurya,...
27 [besok, besok, milih, hakim, muda, udah, pensi...
28 [pemerintah, ajukan, kasasi, ksp, indosurya, b...
29 [mahfud, kabareskrim, vonis, bebas, bos, ksp, ...
30 [kasasi, hakim, memutus, bebas, tersangka, ind...
32 [menko, polhukam, mahfud, md, pemerintah, mem...
33 [mahfud, md, vonis, bebas, indosurya, kalah]
34 [hukuman, terdakwa, lucu, korban, barang, bukt...
35 [indosurya, emang, bener, kawal]
36 [mahfud, pemerintah, buka, investasi, bodong, ...
37 [setuju, kalah, menegakkan, hukum, kebenaran, ...
38 [kemana, rakyat, keadilan]
39 [dmn, keadilan, korban, indosuryakerugian, kor...
40 [mahfud, md, pemerintah, revisi, uu, koperasi,...
41 [tersangka, indosurya, bebasin, keturunan, tio...
```

**Gambar 6.** Hasil *stopword*

Berikut merupakan tampilan dari hasil *stemming* data yang berfungsi untuk mengubah bentuk kata menjadi kata dasar atau dapat diartikan sebagai mengubah kata berhimpunan menjadi kata bentuk dasar. Pada tampilan ini menampilkan sebagian hasil *stemming* data dari total 2488 data kata yang berhasil dilakukan *stemming* data.

```
2488
=====
saudara : saudara
sore : sore
mengadakan : ada
rapat : rapat
kordinasi : kordinasi
kejaksaan : jaksa
agung : agung
mabes : mabes
polri : polri
mentri : tri
ukm : ukm
kantor : kantor
staff : staff
kepresidenan : presiden
membahas : bahas
keterkejutan : kejut
indonesia : indonesia
```

**Gambar 7.** Hasil *stemming* data

Berikut merupakan hasil dari *tokenisasi* data twitter mengenai indosurya yang dimana pada proses *tokenisasi* dilakukan untuk memecah kalimat menjadi beberapa bagian tertentu. Dimana pada bagian *content* seluruh data tweet yang tadi berupa sebuah kalimat paragraph diubah menjadi bentuk potongan kata yang membentuk paragraph pada data *content*.

HASIL DARI TOKENISASI :

	datetime	username	content
0	2023-01-28 08:36:37+00:00	DonnySyalendra	[saudara, sore, ini, kami, mengadaka...
1	2023-01-28 08:35:56+00:00	DonnySyalendra	[saudara, sore, ini, kami, mengadaka...
3	2023-01-28 08:34:38+00:00	evandirgan	[baju, warna, kumpul, bawa, palu, ujung, ujung...
5	2023-01-28 08:33:28+00:00	Cakrabirawa201	[lanjutkan, pak, sampaikan, jarit, hati, nuran...
7	2023-01-28 08:29:18+00:00	evandirgan	[hakim, nya, kebanyakan, wara, wiri, jadinya, ...

Gambar 8. Hasil tokenisasi data

Berikut merupakan tampilan data yang tadi sudah dibersihkan dari beberapa proses diantaranya adalah menghapus *content* yang duplicate, melakukan *sropword*, *tokenisasi* dan *stemming* pada data.

Unnamed: 0	datetime	username	content
0	2023-01-28 08:36:37+00:00	DonnySyalendra	[saudara, 'sore', 'ada', 'rapat', 'kordinasi...
1	2023-01-28 08:35:56+00:00	DonnySyalendra	[saudara, 'sore', 'ada', 'rapat', 'kordinasi...
2	2023-01-28 08:34:38+00:00	evandirgan	[baju, 'warna', 'kumpul', 'bawa', 'palu', 'u...
3	2023-01-28 08:33:28+00:00	Cakrabirawa201	[lanjut, 'jarit', 'hati', 'nuran', 'korban...
4	2023-01-28 08:29:18+00:00	evandirgan	[hakim, 'banyak', 'wara', 'wiri', 'masuk', '...
5	2023-01-28 08:25:16+00:00	dndtirani	[hakim, 'lepas', 'dakwa', 'ksp', 'Indosurya...
6	2023-01-28 08:22:23+00:00	Cakrabirawa201	[dukung, 'md', 'naikn', 'sampe', 'selesai', '...
7	2023-01-28 08:18:12+00:00	Syaman59	[korban, 'aju', 'gugat', 'pmh', 'perdata', '...
8	2023-01-28 08:15:44+00:00	Cakrabirawa201	[sampe, 'clear]
9	2023-01-28 08:12:44+00:00	Lululimakrunn_	[dukung, 'upaya', 'perintah', 'jagung', 'upa...
10	2023-01-28 08:06:37+00:00	AlganHendrakus1	[gerak, 'cepat', 'kawat', 'trus', 'Indosurya...
11	2023-01-28 07:59:14+00:00	AdiputeriC	[udah, 'toh', 'dakwa', 'masuk', 'cuci', 'uan...
12	2023-01-28 07:59:13+00:00	AdiputeriC	[sampe, 'turun', 'tangan', 'bs', 'biar', 'In...
13	2023-01-28 07:47:51+00:00	bahayainformasi	[besok, 'besok', 'milih', 'hakim', 'muda', '...
14	2023-01-28 07:42:32+00:00	beritaekonomi24	[perintah, 'aju', 'kasasi', 'ksp', 'Indosury...

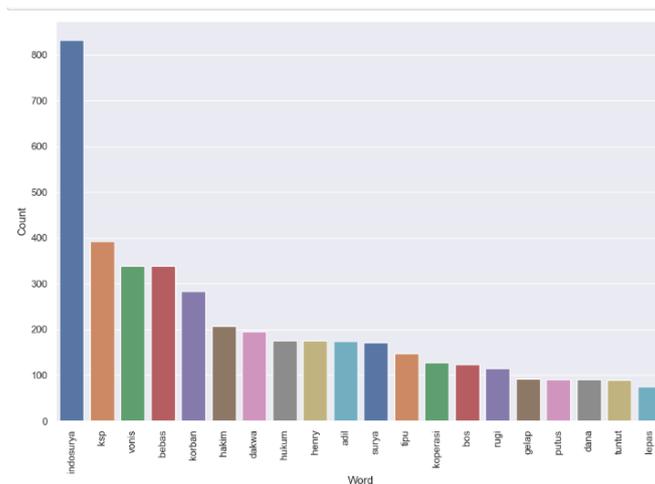
Gambar 9. Tampilan data bersih

Berikut merupakan tampilan data yang sudah dibersihkan *whitespace*, *punctuation* dan *multiple whitespace*. Pada tampilan tersebut akan dihapus kalimat yang memiliki spasi yang berlebihan.

Unnamed: 0	datetime	username	content
0	2023-01-28 08:36:37+00:00	DonnySyalendra	saudara sore ada rapat kordinasi jaksa agung m...
1	2023-01-28 08:35:56+00:00	DonnySyalendra	saudara sore ada rapat kordinasi jaksa agung m...
2	2023-01-28 08:34:38+00:00	evandirgan	baju warna kumpul bawa palu ujung ujung pulang...
3	2023-01-28 08:33:28+00:00	Cakrabirawa201	lanjut jarit hati nurani korban indosurya
4	2023-01-28 08:29:18+00:00	evandirgan	hakim banyak wara wiri masuk angin dah

Gambar 10. Tampilan data hasil proses *whitespace*

Berikut merupakan tampilan dari berapa banyak jumlah kata yang muncul pada data. Dimana pada tampilan hasi menunjukkan kata yang paling sering muncul pada data adalah kata indosurya sedangkan data kata yang paling jarang muncul pada data adalah kata lepas.



Gambar 11. Hasil persentasi kata

Berikut merupakan tampilan *wordcloud* atau dapat disebut dengan kata yang seringkali muncul pada data yang diolah. Dimana pada tampilan hasil *wordcloud* menunjukkan sejumlah kata yang berukuran lebih besar dari yang lain yang berarti kata tersebut merupakan kata yang paling sering muncul pada data dimana pada hasil ini kata indosurya merupakan kata yang paling sering muncul pada data.



Gambar 12. Hasil *wordcloud*

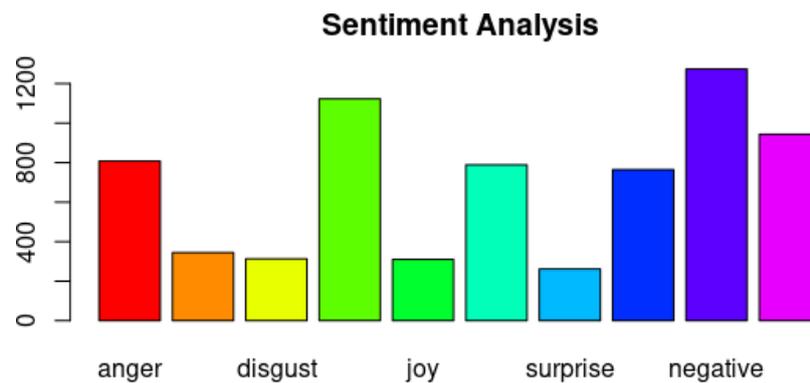
Berikut merupakan tampilan hasil tokenisasi dari bagian *content* dimana pada tahap ini dilakukan proses pemecahan seluruh paragraph data tweet dari bagian *content* menjadi beberapa bagian/token tertentu.

```
[ 'saudara', 'sore', 'ada', 'rapat', 'kordinasi', 'jaksa', 'agung', 'mabes', 'polri', 'tri', 'ukm', 'kantor', 'staff',
  'presiden', 'bahas', 'kejut', 'indonesia', 'perintah', 'rakyat', 'indosurya', 'next' ]
[ 'saudara', 'sore', 'ada', 'rapat', 'kordinasi', 'jaksa', 'agung', 'mabes', 'polri', 'tri', 'ukm', 'kantor', 'staff',
  'presiden', 'bahas', 'kejut', 'indonesia', 'perintah', 'rakyat', 'indosurya' ]
[ 'baju', 'warna', 'kumpul', 'bawa', 'palu', 'ujung', 'ujung', 'pulang', 'bawa', 'berkat', 'hasil', 'tangis', 'ribu',
  'korban', 'yak', 'hakim', 'pensiun' ]
[ 'lanjut', 'jarit', 'hati', 'nurani', 'korban', 'indosurya' ]
[ 'hakim', 'banyak', 'wara', 'wiri', 'masuk', 'angin', 'dah' ]
[ 'hakim', 'lepas', 'dakwa', 'ksp', 'indosurya', 'pakar', 'gerus', 'percaya', 'publik' ]
[ 'dukung', 'md', 'naikkn', 'sampe', 'selesai', 'hakim', 'gimana' ]
[ 'korban', 'aju', 'gugat', 'pmh', 'perdata', 'adil', 'ksp', 'indosurya', 'kembali', 'rugi', 'korban', 'sita', 'aset',
  'ksp', 'indosurya', 'sbg', 'jamin', 'vonis', 'perkara', 'pidana', 'ybs', 'tdak', 'merta', 'kembali', 'rugi', 'korba
  n' ]
[ 'sampe', 'clear' ]
[ 'dukung', 'upaya', 'perintah', 'jagung', 'upaya', 'hukum', 'kasasi', 'indosurya', 'vonis', 'bebas', 'onslag', 'putu
  s', 'salah', 'audit', 'hakim' ]
```

Gambar 13. Hasil tokenisasi

### 4.3. Tahap Klasifikasi Sentimen

Berikut merupakan hasil sentimen *analysis* dengan metode *naïve bayes* menggunakan *tools* R Studio didapatkan hasil terbesar pada sentimen *negative* dengan total 1200 parameter. Pada bagian ini menampilkan hasil berupa grafik dari beberapa sentiment dari hasil olah data tweet Indosurya yang diambil dari twitter, dimana pada grafik terdapat hasil sentimen yang berhasil dideteksi yaitu sentimen *Anger*, *Disgust*, *Fear*, *Joy*, *Mixed*, *Sadness*, dan *Surprised*.



Gambar 14. Tampilan hasil sentimen analysis

## 5. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dengan pengolahan data melalui Jupyter-Notebook untuk crawling data dan preprocessing data, lalu untuk melakukan proses sentimen analysis dilakukan dengan menggunakan tools R Studio. Hasil dari penelitian ialah didapatkan beberapa jenis sentimen yaitu anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative, positive. Dimana hasil tertinggi didapatkan oleh sentimen negative, sehingga didapatkan kesimpulan sentimen publik terhadap topik kasus KSP Indosurya di twitter Indonesia adalah negative dengan jumlah parameter 1200 berdasarkan analisis menggunakan metode naïve bayes, hal ini sebanding dengan keresahan masyarakat di media sosial terkait kasus KSP Indosurya. Saran dari penelitian ini adalah diharapkan kedepannya dilakukan penelitian kembali dengan menggunakan metode yang berbeda.

## Daftar Pustaka

- [1] S. B. Y. G. C. Purba, "Kriteria Proposal Perdamaian PKPU yang Cukup Terjamin Dalam Kasus KSP INDOSURYA CIPTA," vol. 4, no. 3, pp. 607–616, 2022.
- [2] Britney Azzahra Wiguna and Yeti Sumiyati, "Pelayanan Koperasi Simpan Pinjam Kepada Masyarakat (Menurut UU Cipta Kerja & PP No. 7 Tahun 2021)," Bandung Conf. Ser. Law Stud., vol. 2, no. 1, pp. 208–214, 2022, doi: 10.29313/bcsls.v2i1.703.
- [3] P. S. M. Suryani, L. Linawati, and K. O. Saputra, "Penggunaan Metode Naïve Bayes Classifier pada Analisis Sentimen Facebook Berbahasa Indonesia," Maj. Ilm. Teknol. Elektro, vol. 18, no. 1, p. 145, 2019, doi: 10.24843/mite.2019.v18i01.p22.
- [4] M. Kartika, S. Saepudin, and D. Gustian, "Analisis Sentimen Dampak Covid-19 Terhadap Pembatalan Keberangkatan Ibadah Haji Pada Tahun 2020," J. Sains Komput. Inform. (J-SAKTI), vol. 5, no. 2, pp. 964–972, 2021.
- [5] R. W. Samsir, Ambiyar, Unung Verawardina, Firman Edi, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," J. Media Inform. Budidarma, vol. 5, no. 1, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [6] M. R. Fauzi, R. A. Pratama, P. Laksono, and P. Eosina, "Analisis Big Data Sentimen Konsumen UMKM Sektor Kuliner Menggunakan Multi-Label K-Nearest Neighbor," J. Tek. Inform., vol. 9, no. 1, pp. 9–20, 2021, doi: 10.32832/kreatif.v9i1.3587.
- [7] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," Smatika J., vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [8] F. A. D. Aji Prasetya Wibawa, Muhammad Guntur Aji Purnama, Muhammad Fathony Akbar, "Metode-metode Klasifikasi," Pros. Semin. Ilmu Komput. dan Teknol. Inf., vol. 3, no. 1, pp. 134–138, 2018.
- [9] A. Muzaki and A. Witanti, "Sentiment Analysis of the Community in the Twitter To the 2020 Election in Pandemic Covid-19 By Method Naive Bayes Classifier Sentimen Analisis Masyarakat Di Twitter Terhadap Pilkada 2020 Ditengah Pandemic Covid-19 Dengan Metode Naïve Bayes Classifier," J. Tek. Inform., vol. 2, no. 2, pp. 101–107, 2021.

- [10] E. Manalu, F. A. Sianturi, and M. R. Manalu, "Penerapan Algoritma Naive Bayes Untuk Memprediksi Jumlah Produksi Barang Berdasarkan Data Persediaan dan Jumlah Pemesanan Pada CV. Papadan Mama Pastries," *J. Mantik Penusa*, vol. 1, no. 2, pp. 16–21, 2017.
- [11] N. Ayunandita and S. Dadi Riskiono, "Permodelan Sistem Informasi Akademik Menggunakan Extreme Programming Pada Madrasah Aliyah (Ma) Mambaul Ulum Tanggamus," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 2, no. 2, pp. 196–204, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/informatika>
- [12] S. N. J. Fitriyyah, N. Safriadi, and E. E. Pratama, "Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 279–285, 2019, doi: 10.26418/jp.v5i3.34368.
- [13] R. Y. Hayuningtyas, "Penerapan Algoritma Naive Bayes untuk Rekomendasi Pakaian Wanita," *J. Inform.*, vol. 6, no. 1, pp. 18–22, 2019, doi: 10.31294/ji.v6i1.4685.
- [14] R. H. Bagus Muhammad Akbar, Ahmad Taufiq Akbar, "Analisis Sentimen dan Emosi Vaksin SINOVAC Pada TWITTER Menggunakan NAIVE BAYES dan VALENCE SHIFTER," *J. Teknol. Terpadu*, vol. 7, no. 2, pp. 83–92, 2021.