

**ANALISIS SENTIMEN TERHADAP KASUS OMNICRON DI INDONESIA
DENGAN API TWITTER MENGGUNAKAN ALGORITMA NAIVE
BAYES CLASSIFIER**

*Sentiment Analysis of Omnicron in Indonesia using API Twitter with Naive Bayes
Classifier*

Untuk memenuhi sebagian persyaratan
Mencapai derajat sarjana S-1

SKRIPSI

Oleh:
MEISY ARINDA PUTRI
18311175



**PROGRAM STUDI S1 SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS TEKNOKRAT INDONESIA
BANDAR LAMPUNG
2022**

**HALAMAN PENGESAHAN
LAPORAN SKRIPSI S1 SISTEM INFORMASI**

**ANALISIS SENTIMEN TERHADAP KASUS OMNICON DI INDONESIA
DENGAN API TWITTER MENGGUNAKAN ALGORITMA NAIVE
BAYES CLASSIFIER**

Sentiment Analysis of Omnicron in Indonesia using API Twitter with Naive Bayes Classifier

Dipersiapkan dan disusun oleh

**MEISY ARINDA PUTRI
18311175**

Telah dipertahankan di depan Dewan Penguji
Pada tanggal 22 September 2022

Pembimbing,

Penguji,

Auliya Rahman Isnain, S.Kom., M.Cs.
NIK. 022 16 02 02

Zaenal Abidin, S.Si., S.Kom., M.T.
NIK. 021 13 02 04

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana
Tangga 16 November 2022

Fakultas Teknik dan Ilmu Komputer
Dekan,

Program Studi S1 Sistem Informasi
Ketua,

Dr. H. Mahathir Muhammad, S.E., M.M.
NIK. 023 05 00 09

Damayanti, S.Kom., M.Kom.
NIK. 022 09 03 04

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Meisy Arinda Putri
NPM : 18311175
Program Studi : S1 Sistem Informasi

Dengan ini menyatakan bahwa tugas akhir :

Judul : Analisis Sentimen Terhadap Kasus Omnicron di Indonesia
Dengan API Twitter Menggunakan Algoritma *Naive Bayes Classifier*
Pembimbing : Auliya Rahman Isnain, S.Kom., M.Cs.

Belum pernah diajukan untuk diuji sebagai persyaratan untuk memperoleh gelar akademik pada tingkatan di univertitas/ perguruan tinggi manapun. Tidak ada bagian dalam skripsi ini yang pernah dipublikasikan oleh pihak lain, kecuali bagian yang digunakan sebagai referensi, berdasarkan kaidah penulisan ilmiah yang benar.

Apabila dikemudian hari ternyata laporan tugas akhir yang saya tulis terbukti hasil saduran/plagiat, maka saya akan bersedia menanggung segala resiko yang akan saya terima.

Demikian pernyataan ini dibuat dengan sebenar – benarnya.

Bandar Lampung, 16 November 2022
Yang menyatakan,

Meisy Arinda Putri
NPM. 18311175

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPETNINGAN AKADEMIS

Sebagai sivitas akademik Fakultas Teknik dan Ilmu Komputer Universitas Tenokrat Indonesia, saya yang bertanda tangan di bawah ini :

Nama : Meisy Arinda Putri
NPM : 18311175
Program Studi : S1 Sistem Informasi
Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Fakultas Teknik dan Ilmu Komputer Universitas Tenokrat Indonesia **Hak Bebas Royalti Non eksklusif (*Non exclusive Royalty Free Right*)** atas karya ilmiah saya yang berjudul :

Analisis Sentimen Terhadap Kasus Omnicron di Indonesia Dengan API Twitter Menggunakan Algoritma *Naive Bayes Classifier* beserta perangkat yang ada (jika diperlukan). Dengan hak bebas royalti *non exclusive* ini Program Studi S1 Sistem Informasi Fakultas Teknik Dan Ilmu Komputer Universitas Teknokrat Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis, pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bandar Lampung
Pada tanggal : 16 November 2022

Yang menyatakan,

Meisy Arinda Putri
NPM. 18311175

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Skripsi dengan judul Analisis Sentimen Terhadap Kasus Omnicron di Indonesia Dengan API Twitter Menggunakan Algoritma *Naive Bayes Classifier*. Penulisan Skripsi ini dilakukan dalam rangka memenuhi salah satu syarat menyelesaikan pendidikan Strata 1 Program Studi SI Sistem Informasi Fakultas Teknik dan Ilmu Komputer Universitas Teknokrat Indonesia. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, sangatlah sulit bagi penulis untuk menyelesaikan laporan Skripsi ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Dr. H.M. Nasrullah Yusuf, S.E., M.B.A., selaku Rektor Universitas Teknokrat Indonesia.
2. Dr. H. Mahathir Muhammad, S.E., M.M., selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Teknokrat Indonesia.
3. Damayanti, S.Kom., M.Kom., selaku Ketua Program Studi S1 Sistem Informasi Fakultas Teknik dan Ilmu Komputer Universitas Teknokrat Indonesia.
4. Auliya Rahman Isnain, S.Kom., M.Cs., selaku Dosen Pembimbing yang telah meluangkan waktu untuk membimbing penulis menyelesaikan Proposal Skripsi ini.
5. Zaenal Abidin, S.Si., S.Kom., M.T., selaku Dosen Penguji.

Akhir kata, penulis berharap semoga Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu dan semoga Laporan Skripsi ini membawa manfaat.

Bandarlampung, 16 November 2022

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN LAPORAN SKRIPSI	ii
LEMBAR PERNYATAAN.....	iii
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	ix
ABSTRAK.....	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	4
BAB II LANDASAN TEORI.....	5
2.1 Tinjauan Pustaka	5
2.2 Keaslian Penelitian.....	10
2.3 Data Mining	10
2.3.1 Fungsi Data Mining	11
2.3.2 Tahapan Data Mining	12
2.4 Text Mining.....	14
2.4.1 Algoritma <i>TF-IDF</i>	17
2.4.2 Analisis Sentimen	18
2.5 <i>Naïve Bayes Classifier</i>	19
2.6 Covid-19.....	20
2.7 Twitter dan Twitter API.....	20
2.7.1 Pengertian Twitter	20
2.7.2 Struktur Data Twitter	21
2.7.3 Pengertian Twitter API.....	21
2.8 Rapid Miner	22
BAB III METODOLOGI PENELITIAN	24
3.1 Analisis Masalah	24
3.2 Kerangka Tahapan Penelitian	24
3.3 Proses Penarikan Data (<i>Crawling Data</i>).....	26
3.4 <i>Pre-processing Data</i>	32
3.5 <i>Weighting</i> (Algoritma <i>TF-IDF</i>).....	38

	Halaman
BAB IV IMPLEMENTASI DAN PEMBAHASAN	44
4.1 Implementasi	44
4.1.1 Implementasi <i>Naive Bayes Classifier</i> Menggunakan Rapid Miner	44
4.1.2 Hasil Implementasi Menggunakan Rapid Miner	52
4.1.3 Implementasi <i>Naive Bayes Classifier</i> Menggunakan Phyton ...	56
4.1.4 Hasil Implementasi Menggunakan Phyton	63
4.2 Pembahasan	66
 BAB V PENUTUP	 67
5.1 Kesimpulan	67
5.2 Saran	68
 DAFTAR PUSTAKA	 69

DAFTAR TABEL

	Halaman
Tabel 2. 1 Tinjauan Pustaka	5
Tabel 3. 1 Komentar Twitter	33
Tabel 3. 2 Memberikan Label Sentimen	33
Tabel 3. 3 Hasil <i>Cleansing</i>	34
Tabel 3. 4 Hasil <i>Case Folding</i>	35
Tabel 3. 5 Hasil <i>Tokenizing</i>	35
Tabel 3. 6 Hasil <i>Filtering</i>	36
Tabel 3. 7 Hasil <i>Stemming</i>	37
Tabel 3. 8 Hasil <i>Pre-Processing</i>	37
Tabel 3. 9 <i>Term Frequent</i> Seluruh Kata.....	38
Tabel 3. 10 Nilai IDF Seluruh Kata	40
Tabel 3. 11 Hasil Pembobotan TF-IDF.....	42
Tabel 4. 1 Pemberian Label Sentimen	44
Tabel 4. 2 Hasil Perbandingan	66

DAFTAR GAMBAR

	Halaman
Gambar 2. 1 Tahapan Data Mining	13
Gambar 2. 2 Proses <i>Text Mining</i>	15
Gambar 2. 3 Proses Folding	16
Gambar 2. 4 Proses <i>Tokenizing</i> dan <i>Filter</i>	16
Gambar 2. 5 Proses <i>Stemming</i>	17
Gambar 2. 6 Ilustrasi Algoritma <i>TF-IDF</i>	17
Gambar 2. 7 Tampilan Rapid Miner	23
Gambar 3. 1 Kerangka Tahapan Penelitian.....	25
Gambar 3. 2 <i>Search Twitter Operator</i>	26
Gambar 3. 3 Membuat Koneksi Baru	27
Gambar 3. 4 Memilih Tipe Koneksi	27
Gambar 3. 5 <i>Request Access Token</i>	28
Gambar 3. 6 Autentikasi	28
Gambar 3. 7 Otorisasi Akun Twitter.....	29
Gambar 3. 8 Token Akses	29
Gambar 3. 9 <i>Copy Code Akses Token</i>	30
Gambar 3. 10 Akses Token Terisi	30
Gambar 3. 11 Buat Proses.....	31
Gambar 3. 12 Hasil <i>Crawling Data</i>	31
Gambar 3. 13 Menyimpan Data Hasil <i>Crawling</i>	32
Gambar 3. 14 Data Sebelum <i>Pre-Processing</i>	32
Gambar 3. 15 Data Dalam Bentuk Dokumen	38
Gambar 4. 1 Operator <i>Read Excel</i>	46
Gambar 4. 2 <i>Import Data</i>	47
Gambar 4. 3 <i>Format Columns</i>	48
Gambar 4. 4 Menentukan Atribut Label Sentimen	49
Gambar 4. 5 Menerapkan Algoritma <i>TF-IDF</i>	49
Gambar 4. 6 Operator <i>Pre-Processing</i>	50
Gambar 4. 7 Menerapkan <i>Filter Stopword</i>	51
Gambar 4. 8 Menerapkan <i>Cross Validation</i>	51
Gambar 4. 9 Menerapkan Algoritma <i>Naive Bayes</i>	52
Gambar 4. 10 Tampilan <i>Process</i>	52
Gambar 4. 11 Hasil Pengujian Algoritma	53
Gambar 4. 12 <i>Source Code Crawling</i>	57
Gambar 4. 13 <i>Source Code TF-IDF</i> dan Pengujian.....	62
Gambar 4. 14 Hasil <i>Lexicon Based</i>	64

ABSTRAK

ANALISIS SENTIMEN TERHADAP KASUS OMNICON DI INDONESIA DENGAN API TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER

Oleh

MEISY ARINDA PUTRI
18311175

Varian terbaru dari virus Covid-19 yaitu *omicron* dinyatakan pertama kali masuk ke Indonesia pada 27 November 2021. Varian omicron dinilai paling banyak berasal dari WNI yang melakukan perjalanan dari luar negeri. Semakin meningkatnya kasus varian omicron di Indonesia hingga membuat pemerintah kembali memberlakukan PPKM di beberapa daerah. Meningkatnya kasus varian omicron tidak hanya berdampak pada kesehatan, tetapi berimbas ke berbagai bidang kegiatan masyarakat. Banyaknya masyarakat yang resah akan masuknya varian omicron di Indonesia membuat masyarakat mengungkapkan opini mereka. Opini-opini masyarakat tersebut ditujukan untuk berbagai pihak seperti pemerintah dan *public figure* melalui media sosial salah satunya melalui Twitter yang merupakan salah satu media sosial teratas yang digunakan untuk masyarakat menyuarakan opini-opini mereka.

Penelitian ini bertujuan untuk mengklasifikasikan opini-opini tersebut kedalam Sentimen Negatif, Sentimen Positif, dan Sentimen Netral menggunakan Algoritma *Naive Bayes Classifier*. Hasil pengujian menunjukkan bahwa nilai *Accuracy* mencapai nilai 50,22%, nilai *Precision* 48,72%, dan nilai *Recall* 50,70%. Hasil penerapan dan pengujian menggunakan Phyton lebih besar dibanding menggunakan *tools* Rapid Miner dimana nilai akurasi menggunakan Phyton mencapai 0.81 atau 81%, *precision* 0.90 atau 90%, dan *recall* 0.77 atau 77%.

Kata Kunci : *Omicron Indonesia, Analisis Sentimen, API Twitter, Naive Bayes Classifier, Text Mining*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Hingga saat ini dampak pandemi COVID-19 masih lah dirasakan oleh seluruh negara. Berbagai upaya telah dilakukan untuk menekan pandemi ini mulai dari diberlakukannya *lockdown*, *social distancing*, wajib menggunakan masker di tempat umum, hingga pemberian vaksin. Namun hingga saat ini COVID-19 masih terus mengalami mutasi mulai dari varian *alfa*, *beta*, dan yang paling terbaru adalah varian *omicron*. Varian terbaru dari virus ini yaitu *omicron* dinyatakan pertama kali masuk ke Indonesia pada 27 November 2021. Varian ini diduga berasal dari warga negara Indonesia (WNI) yang tiba dari Nigeria, (Rokom, 2021).

Varian omicron dinilai paling banyak berasal dari WNI yang melakukan perjalanan dari luar negeri. "Secara kumulatif kasus paling banyak berasal dari Turki dan Arab Saudi. Kemudian kebanyakan kasus konfirmasi Omicron adalah mereka yang sudah lengkap vaksinasi Covid-19," kata Kepala Biro Komunikasi dan Pelayanan Masyarakat Kemenkes Widyawati melalui keterangannya pada Sabtu, 8 Januari 2022, (Chairunnisa, 2022). Semakin meningkatnya kasus varian omicron di Indonesia hingga membuat pemerintah kembali memberlakukan PPKM di beberapa daerah. Meningkatnya kasus varian omicron tidak hanya berdampak pada kesehatan, tetapi berimbas ke berbagai bidang kegiatan masyarakat. Hal tersebut membuat masyarakat geram dan merasa dirugikan karena aktivitas pekerjaan yang telah mulai kembali dilakukan secara normal harus kembali dihentikan atau dilakukan dari rumah.

Banyaknya masyarakat yang resah akan masuknya varian omicron di Indonesia membuat masyarakat mengungkapkan opini mereka. Opini-opini

masyarakat tersebut ditujukan untuk berbagai pihak seperti pemerintah dan *public figure*. Dalam era dimana saat ini telah banyak masyarakat menggunakan sosial media untuk berkomunikasi hingga banyak yang menggunakan sosial media juga untuk mengutarakan opini mereka. Karena mengutarakan opini melalui sosial media dinilai lebih mudah dan lebih cepat tersampaikan kepada pihak yang diinginkan. Twitter merupakan salah satu media sosial teratas yang digunakan untuk masyarakat menyuarkan opini-opini mereka.

Apabila data opini masyarakat di Twitter diolah dan dianalisis dengan benar maka akan sangat penting dan bermanfaat bagi masyarakat atau organisasi-organisasi pemerintah yang dapat digunakan untuk negara Indonesia dalam memperbaiki sistem penanganan omicron saat ini. Data opini masyarakat di Twitter dapat ditarik atau diambil menggunakan API Twitter dengan menggunakan kata kunci atau *hashtag* sehingga data yang di dapatkan sesuai dengan topik yang akan dibahas. Namun banyaknya data opini yang didapatkan membuat data menjadi tidak terstruktur dan terkelompok.

Hal tersebut menunjukkan perlunya suatu teknik atau metode yang dapat digunakan untuk menganalisis data opini tersebut. Salah satu teknik yang dapat digunakan untuk menganalisis opini adalah *text mining*. *Text mining* merupakan salah satu teknik dari *data mining* yang biasa digunakan untuk melakukan analisis sentimen maka teknik ini dapat digunakan untuk melakukan analisis opini masyarakat terhadap kasus omicron di Indonesia.

Dengan adanya peluang sumber data yang sangat besar dapat dimanfaatkan untuk menghasilkan suatu *knowledge* yang bermanfaat. Pemanfaatan data yang bersumber dari media sosial merupakan terobosan baru untuk dijadikan sebagai

alternative sumber data. Analisis sentiment dapat digunakan untuk melakukan evaluasi suatu kinerja pelayanan dan sebagainya. Menurut Fink tahun 2011, jenis analisis sentiment yang sering digunakan dalam kalangan penelitian yaitu, analisis sentimen dalam bentuk dokumen atau kalimat. Terdapat beberapa metode Data Mining yang dapat digunakan untuk melakukan Teks Mining analisis sentiment, diantaranya adalah *Naïve Bayes Classifier*. *Naïve Bayes Classifier* merupakan metode klasifikasi yang menyatakan bahwa seluruh kegiatan memberikan kontribusi yang sama penting atau saling bebas pada pemilihan kelas tertentu, (Darwis, 2020).

Berdasarkan hal tersebut maka penulis akan melakukan analisis atau opini dari komentar masyarakat di twitter menggunakan metode *Naïve Bayes Classifier*. Dengan pemaparan diatas maka penulis memutuskan untuk melakukan penelitian dengan judul “**Analisis Sentimen Terhadap Kasus Omnicron di Indonesia dengan API Twitter Menggunakan Algoritma *Naive Bayes Classifier***”.

1.2 Rumusan Masalah

Berdasarkan latar belakang maka rumusan masalah pada penelitian ini adalah:

1. Bagaimana melakukan *crawling data* dari Twitter menggunakan API Twitter?
2. Bagaimanakah menerapkan metode *Naïve Bayes Classifier* untuk menentukan nilai positif, negative dan netral dalam analisis sentiment mengenai Omicron di Indonesia?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan yang ingin didapatkan antara lain:

1. Mendapatkan data opini menggunakan API Twitter.
2. Mengklasifikasi opini-opini masyarakat di Twitter.
3. Mengetahui nilai akurasi, presisi, dan *recall* dari hasil penerapan metode *Naïve Bayes Classifier*.

1.4 Batasan Masalah

Untuk melakukan penelitian yang terfokus maka dibuatlah batasan-batasan masalah dari penelitian ini sebagai berikut:

1. Data yang digunakan merupakan data opini yang diunggah ke Twitter dengan bahasa Indonesia dengan kata kunci atau *hashtag* yang berhubungan dengan varian omicron di Inonesia.
2. *Crawling data* (penarikan data) dilakukan dengan menggunakan API Twitter.
3. Dalam media yang digunakan untuk melakukan ujicoba pada data menggunakan aplikasi *Rapid Miner*.

1.5 Manfaat Penelitian

Dengan terselesaikannya penelitian ini maka penulis berharap dapat memberikan manfaat sebagai salah satu tinjauan penelitian terhadap penggunaan API Twitter untuk melakukan pengambilan data dari *platform* Twitter dan mengetahui kemampuan metode *Naïve Bayes Classifier* dalam menganalisis sentimen atau opini dari twitter.

BAB II LANDASAN TEORI

2.1 Tinjauan Pustaka

Tinjauan pustaka pada penelitian-penelitian yang dilakukan sebelumnya dalam mendukung penelitian yang sedang dilakukan. Berikut ini adalah penelitian yang telah dilakukan sebelumnya terkait dengan penelitian yang akan dilakukan oleh penulis :

Tabel 2. 1 Tinjauan Pustaka

No	Penulis, Tahun	Judul	Metode
1	Jeffri Prayitno Bangkit Saputra dan Racidon P. Bernarte, 2022	Algoritma Naïve Bayes dalam Memprediksi Penyebaran Omicron Varian Covid-19 di Indonesia: Implementasi dan Analisis	<i>Naïve Bayes Classifier</i>
2	Alvina Felicia Watratan, Arwini Puspita dan Dikwan Moeis, 2020	Implementasi Algoritma <i>Naïve Bayes</i> untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia	<i>Naïve Bayes Classifier</i>
3	Andi Akram Nur Risal, Nur Inayah Yusuf, Andi Baso Kaswar, dan Fathiah Adiba, 2021	Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naïve Bayes	<i>Naïve Bayes Classifier</i>
4	Dedi Darwis, Nery Siskawati, Zaenal Abidin, 2020	Penerapan Algoritma <i>Naive Bayes</i> Untuk Analisis Sentimen Review Data Twitter BMKG Nasional	<i>Naïve Bayes</i>
5	Debby Alita, Indah Sari, Auliya Rahman Isnain, Styawati 2021	Penerapan Naive Bayes Classifier untuk Pendukung Keputusan Penerima Beasiswa	<i>Naive Bayes Classifier</i>
6	Dedi Darwis, Eka Shintya Pratiwi, Ferico Octaviansyah Pasaribu 2020	Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia	<i>Support Vector Machine</i>

Tabel 2. 1 Tinjauan Pustaka (lanjutan)

No	Penulis, Tahun	Judul	Metode
7	Muhammad Ghudafa Taufik Akbar, Dwi Budi Srisullstiowati, 2021	Analisa Sentimen Efektifitas Vaksin terhadap Varian Covid-19 Omicron Berbasis Leksikon	<i>Support Vector Machine</i>
8	Auliya Rahman Isnain, Adam Indra Saktim Deby Alta, Nurman Satya Marga 2021	Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM	<i>Support Vector Machine</i>
9	Styawati, Nirwana Hendrastuty, Auliya Rahman Isnain, Ari Yanti Ramadhani 2021	Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja pada Twitter dengan Metode <i>Support Vector Machine</i>	<i>Support Vector Machine</i>
10	Fikri Aldi Nugraha, Nisa Hanum Harani, Roni Habibi dan Jl. Nuraini Siti Fatonah, 2020	Analisis Sentimen Social Distancing dan Physical Distancing Media Sosial Twiter Menggunakan Algoritma <i>Recurrent Neural Network (RNN)</i>	<i>Recurrent Neural Network (RNN)</i>

Literatur 1 (Saputra & Bernarte, 2022) dengan judul “Algoritma Naive Bayes dalam Memprediksi Penyebaran Omicron Varian Covid-19 di Indonesia Implementasi dan Analisis” dengan permasalahan ditemukan kesulitan yang terkait dengan tingkat penyebaran covid-19 yang terjadi khususnya di Indonesia, dengan isu-isu saat ini kemudian dinilai untuk memutuskan bagaimana mengatasinay dan menentukan luasnya masalah yang perlu diselidiki lebih lanjut. Proporsi instan yang diklasifikasikan dengan benar adalah 48,4848%, sedangkan persentase instance tang tidak diklasifikasikan adalah 51,5152%. Sebanyak 33 data kasus covid-19 per provinsi berhasil dikategorikan dengan tepat, dengan 16 data kasus covid-19 per Provinsi berhasil dikategorikan dengan 16 data terklasifikasikan dengan benar dan

mendapatkan nilai akurasi 46,4252% dan sebanyak 17 data kasus covid-19 per Provinsi salah diklasifikasikan dengan mendapatkan nilai akurasi sebesar 46,4252%.

Literatur 2 (Watratana, B, & Moeis, 2020) dengan judul “Implementasi Algoritma *Naïve Bayes* untuk Memprediksi Tingkat Penyebaran Covid-19 di Indonesia” dengan permasalahan cirus covid-19 dapat menyebabkan penyakit pada burung dan mamalia termasuk manusia. Pada manusia, corona virus menyebabkan infeksi saluran pernapasan yang umumnya ringan, seperti pilek, beberapa bentuk penyakit seperti: SARS, MERS dan Covid-19. Hasil penelitian menunjukkan bahwa 16 data dari 33 data yang di uji dalam kasus covid-19 per provinsi dengan keakuratan sebesar 48,4848%, dimana dari 33 data yang di uji dalam kasus Covid-19 per provinsi yang diuji terdapat 16 data yang berhasil diklasifikasikan dengan benar.

Literatur 3 (Risa, Yusuf, & Kaswar, 2021) dengan judul “Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma *Naïve Bayes*” dengan permasalahan terdapat lebih dari 100 kandidat vaksin covid-19 yang dikembangkan dan beberapa masih dalam tahap uji coba dan belum terdapat vaksin yang benar-benar efektif. Hasil penelitian menggunakan Algoritma *Naïve Bayes* mendapatkan akurasi sebesar 91% dimana dari 23 data latih, 2 diantaranya tidak diklasifikasikan secara tepat.

Literatur 4 (Darwis, Siskawati, et al., 2020) dengan judul “Penerapan Algoritma *Naïve Bayes* Untuk Analisis Sentimen Review Data Twitter BMKG Nasional” dimana Twittwe menjadi salah satu tempat yang dimanfaatkan untuk menyampaikan informasi yang diberikan BMKG Nasional. Hasil penelitian

menunjukkan Akurasi Algoritma *Naive Bayes* dalam melakukan klasifikasi sebesar 69,97%.

Literatur 5 (Alita et al., 2021) dengan judul “Penerapan Naive Bayes Classifier untuk Pendukung Keputusan Penerima Beasiswa” penelitian ini dilakukan untuk proses penentuan beasiswa sesuai dengan kriteria sehingga menjadi pembantu pengambil keputusan penerima beasiswa menggunakan metode *Naive Bayes Classifier*.

Literatur 6 dengan judul “ Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia” (Darwis, Pratiwi, et al., 2020) masyarakat memanfaatkan Twitter sebagai media penyalur opini terhadap kinerja KPK RI. Klasifikasi dilakukan dengan menggunakan metode *SVM* dengan hasil evaluasi akurasi sebesar 82%.

Literatur 7 dengan judul Analisa Sentimen Efektifitas Vaksin terhadap Varian Covid-19 Omicron Berbasis Leksikon (Akbar & Srisusilowati, 2021) dengan permasalahan peningkatan penyebaran Covid-19 tentang munculnya varian baru covid-19 yaitu omicron. Dengan menggunakan Metode *Support Vector Machine* penelitian ini adalah dataset yang diambil dengan cara *crawling* komentar youtube kanal berita berbahasa ingris kemudian dilakukan pelabelan teks berbasis leksikon menggunakan TextBlob dan Vader ke dalam komentar positif, negatif, dan netral. Terdapat perbedaan jumlah komentar positif, negatif, dan netral dengan kedua metode pelabelan text tersebut namun tidak terlalu jauh. Akurasi penerapan SVM terhadap dataset yang dilabeli dengan metode VADER mengungguli pelabelan komentar dengan metode TextBlob.

Literatur 8 (Isnain et al., 2021) dengan judul “ Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM” dengan permasalahan pendapat masyarakat tentang covid-19 di Indonesia dan dengan adanya virus corona ini banyak korban yang meninggal sehingga membuat masyarakat mengeluarkan opini mereka akan munculnya Virus Covid-19 dan Kebijakan Pemerintah dalam menanganinya. Penggunaan metode *Support Vector Machine* menunjukkan Akurasi sebesar 74%, Presisi sebesar 75%, *Recall* sebesar 92%, dan *F1 Score* sebesar 83%.

Literatur 9 (Styawati et al., 2021) dengan judul “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja pada Twitter dengan Metode *Support Vector Machine*” dengan permasalahan maraknya pembahasan mengenai Kartu Prakerja sebagai salah satu Program Pemerintah pada tahun 2020 dalam mengatasi pengangguran dan korban PHK. Hasil evaluasi yang dilakukan menunjukkan akurasi kernel linear sebesar 98.87%.

Literatur 10 dengan judul “Pengelompokkan Sentimen Pada Twitter Tentang Pendapat Masyarakat Karantina Selama Pandemi Covid-19 Menggunakan Metode DBSCAN, (Manis et al., 2021). dengan permasalahan dilakukan karantina untuk membatasi terjadinya penyebaran penyakit sehingga menimbulkan pro dan kontra dan membuat masyarakat mengungkapkan segala pendapat dan kritikan melalui twitter. Berdasarkan hasil pengujian dilakukan dengan data sebanyak 200 data mendapatkan nilai *silhouette coefficient* terbaik yaitu 0,32 pada nilai *epsilon* sebesar 20 dan *minPts* sebesar 15, sedangkan nilai *davies bouldin index* terbaik yaitu 0,10 pada nilai *epsilon* sebesar 15 dan *minPts* sebesar 15. Penelitian ini

mendapatkan hasil analisis lebih banyak pada sentimen netral yang berarti bahwa masyarakat berpendapat netral terhadap karantina selama pandemi Covid-19.

2.2 Keaslian Penelitian

Berdasarkan tinjauan pustaka yang telah dilakukan oleh penulis, maka perbedaan antara penelitian terdahulu dan penelitian yang dilakukan adalah sebagai berikut:

1. Data set yang digunakan merupakan *tweet* yang berisi komentar masyarakat Indonesia mengenai masuknya varian omicron ke Indonesia.
2. Data set yang digunakan diperoleh dengan cara *crawling* menggunakan Twitter API.
3. Penerapan dan visualisasi algoritma *Naïve Bayes Classifier* menggunakan *software* Rapid Miner.

2.3 Data Mining

Menurut *Gartner Group* data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2006).

Menurut (Santosa & Umam, 2018), data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar. sedangkan menurut (Khomarudin, 2016) data mining dapat diartikan sebagai proses mengekstrak atau menggali knowledge yang ada pada sekumpulan data, informasi dan knowledge yang didapat tersebut dapat digunakan pada banyak bidang seperti manajemen bisnis, pendidikan, kesehatan dan sebagainya.

2.3.1 Fungsi Data Mining

Menurut (Larose, 2005) dalam bukunya yang berjudul “*Discovering knowledge in data : an introduction to data mining*”, fungsi data mining dibagi menjadi:

1. Deskripsi

Terkadang peneliti dan analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numeric dari pada ke arah kategori. Model di bangun dengan *record* lengkap menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai variabel target di buat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa metode dan tehnik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat di pisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah.

5. Pengklasteran

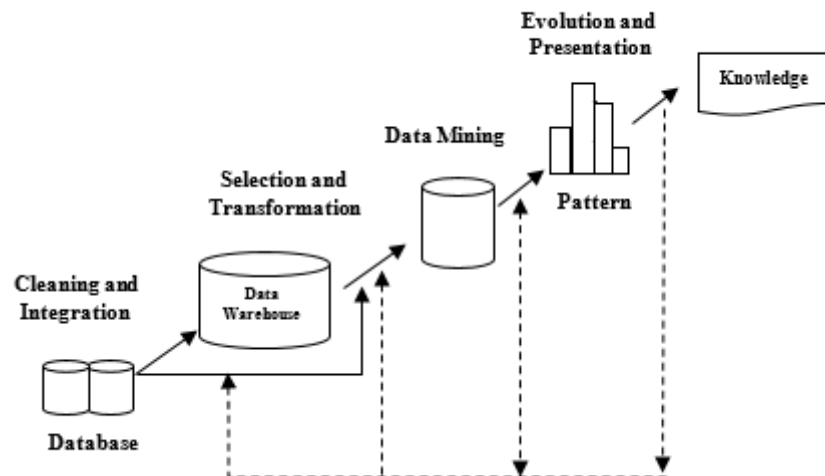
Pengklasteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Klaster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan *record-record* dalam klaster lain.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja (market basket analysis).

2.3.2 Tahapan Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada Gambar 2. berikut tahap-tahap yang ada pada data mining. Proses data mining terbagi dalam enam tahapan (Abdillah et al., 2016), yaitu:



Gambar 2. 1 Tahapan Data Mining (Abdillah et al., 2016)

Adapun penjelasan dari tahapan data mining pada ilustrasi gambar diatas adalah sebagai berikut:

1. Pembersihan Data (*Data Cleaning*)

Proses pembersihan data merupakan kegiatan membuang data yang sama (duplikasi), memeriksa data yang tidak konsisten, dan memperbaiki data yang salah.

2. Integrasi Data (*Data Integration*)

Proses integrasi data merupakan kegiatan menggabungkan data yang berasal dari berbagai sumber.

3. Seleksi Data (*Data Selection*)

Proses seleksi data ini akan dipilih data-data yang hendak digunakan dalam proses sesuai dengan kebutuhan analisis..

4. Transformasi Data (*Data Transformation*)

Pada tahap ini, data yang telah dipilih akan diubah ke dalam format yang telah ditentukan untuk kemudian di proses dan siap ditambah.

5. Proses Mining

Proses mining dilakukan dengan menerapkan metode data mining ke sistem sehingga diperoleh informasi yang diinginkan dari data-data tersebut.

6. Presentasi Pengetahuan (*Knowledge Presentation*)

Tahap ini merupakan tahap dimana hasil dari proses mining diterjemahkan kedalam bentuk informasi yang dapat dipahami oleh pihak yang membutuhkan.

2.4 Text Mining

Menurut (Aditya, 2015), *Text mining* adalah proses menambang data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen tersebut.

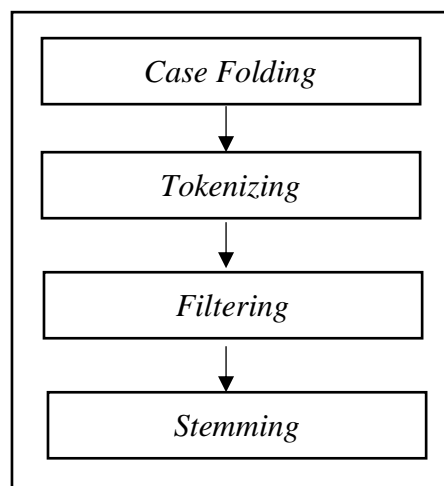
Text mining dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian dan menganalisa *unstructured text* dalam jumlah besar. Dalam memberikan solusi, *text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain seperti data mining, *information retrieval*, *statistic* dan matematika, *machine learning*, *linguistic*, *natural language processing*, dan *visualization*. Kegiatan riset untuk *text mining* antara lain ekstraksi dan penyimpanan teks, *processing* akan konten teks, pengumpulan data *statistic* dan *indexing* dan analisa konten.

Text mining memiliki permasalahan yang sama seperti data mining yaitu jumlah data yang besar, dimensi yang tinggi, data dan struktur yang terus berubah, data *noise*. Terdapat perbedaan data yang digunakan antara data mining dan *text*

mining. Data mining menggunakan data *structured data*, sedangkan *text mining* umumnya menggunakan data *unstructured data*, atau minimal *semistructured*.

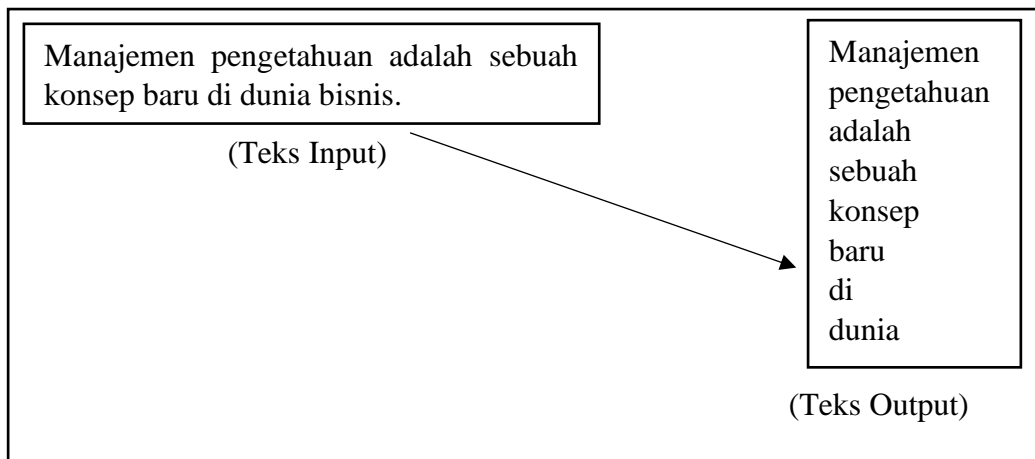
Pada umumnya teks mining memiliki beberapa karakteristik diantaranya yaitu dimensi yang tinggi, terhadap *noise* pada data, dan terdapat struktur teks yang tidak baik. Dalam mempelajari struktur data teks harus menentukan fitur-fitur yang mewakili setiap kata untuk setiap fitur yang ada pada dokumen, sebelum menentukan fitur-fitur yang mewakili, diperlukan tahap *preprocessing* yang dilakukan secara umum dalam teks mining pada dokumen, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming* (Raymond, 2006), seperti terlihat pada gambar

2.1



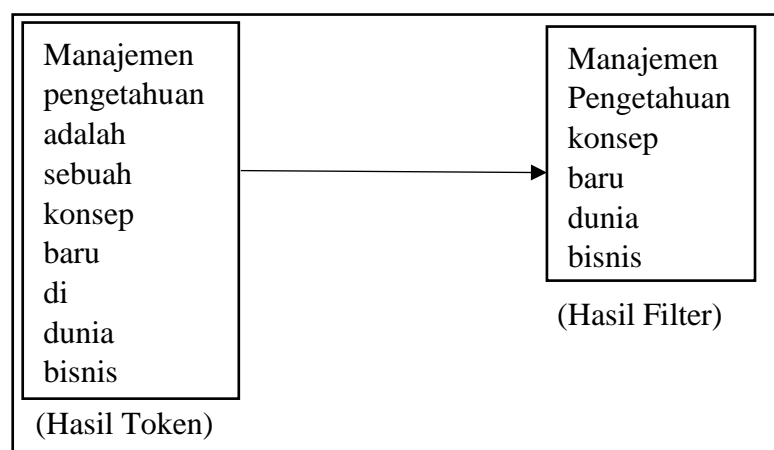
Gambar 2. 2 Proses Text Mining

Case folding yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf `a` sampai dengan `z` yang diterima. Karakter selain huruf dihilangkan dan dianggap deimeter. proses *folding* seperti gambar 2.2.



Gambar 2. 3 Proses Folding

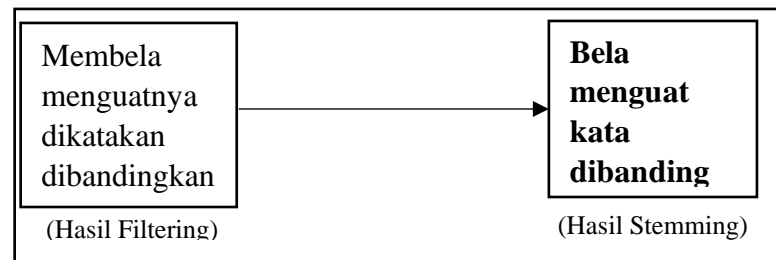
Tahap *Tokenizing* atau *parsing* adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya, sedangkan tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil *term*. Bisa menggunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-word*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari”, dan seterusnya. Proses *tokenizing* dan *filtering* seperti gambar 2.3.



Gambar 2. 4 Proses Tokenizing dan Filter

Tahap *stemming* adalah tahap mencari *root* kata dari tiap kata hasil *filtering*. Pada tahap *stemming* dilakukan proses pengambilan berbagai bentukan kata

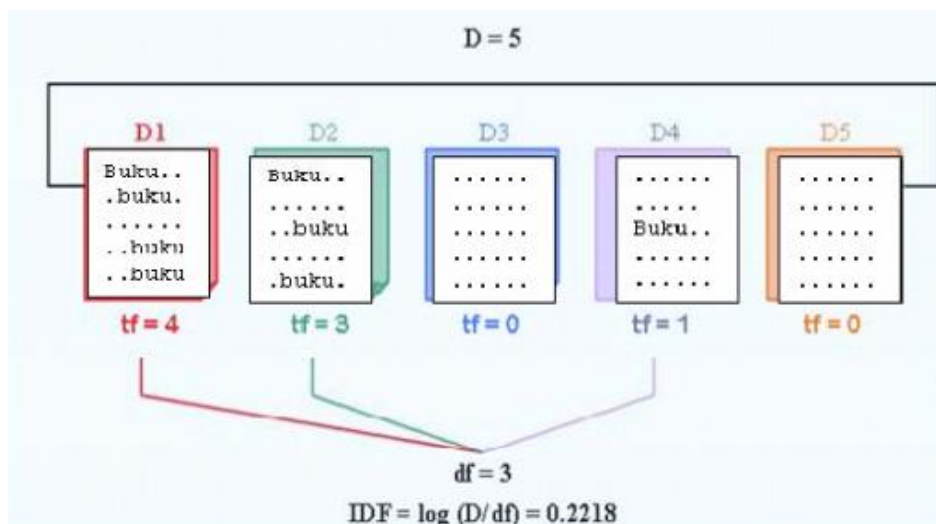
kedalam suatu representasi yang sama. Tahap ini kebanyakan dipakai untuk teks berbahasa inggris dan lebih sulit diterapkan pada teks berbahasa indonesia. Hal ini dikarenakan bahasa Indonesia tidak memiliki rumus bentuk baku yang permanen. Proses tahapan *stemming* pada teks berbahasa Indonesia seperti pada gambar



Gambar 2. 5 Proses *Stemming*

2.4.1 Algoritma *TF-IDF*

Algoritma *TF-IDF* merupakan algoritma pembobotan data berbentuk teks menjadi *numeric* atau *binary* sehingga data dapat diterapkan kedalam algoritma *data mining*. Algoritma ini merupakan jenis klasifikasi yang didasarkan oleh relevansi umpan balik dan diusulkan oleh (Rocchio, n.d.). Gambar 2.6 berikut ini merupakan ilustrasi dari Algoritma *TF-IDF*.



Gambar 2. 6 Ilustrasi Algoritma *TF-IDF*

Keterangan:

D1, D2, D3, D4, D5 = Dokumen

D = Total Dokumen

Tf = Banyaknya kata yang dicari pada sebuah dokumen

Df = Banyak dokumen yang mengandung kata yang dicari

Terdapat 3 tahap dalam menggunakan algoritma ini, dimana tahapannya adalah sebagai berikut:

1) *Term Frequent*

Tf adalah frekuensi dari suatu kata yang muncul dalam satu dokumen.

2) *Inverse Document Frequency*

IDF adalah nilai *invers* dari dokumen yang mengandung kata tersebut.

Berikut ini merupakan persamaan untuk menghitung nilai IDF :

$$IDF(t) = \log\left(\frac{n}{df(t)}\right) \quad (2.1)$$

Dimana: n = total dokumen

$df(t)$ = jumlah dokumen yang mengandung kata t

3) *Weighting TF-IDF* (Pembobotan)

Untuk menemukan nilai bobot dari data dilakukan dengan cara mengalikan nilai TF dan IDF . Berikut ini merupakan persamaan dari pembobotan $TF-IDF$:

$$W(dt) = TF(d, t) \times IDF(t) \quad (2.2)$$

2.4.2 Analisis Sentimen

Menurut (Mujilahwati, 2016), Analisis Sentimen atau *opinion mining* adalah kajian tentang cara untuk memecahkan masalah dari opini masyarakat, sikap dan emosi suatu entitas, dimana entitas, dimana entitas tersebut dapat mewakili

individu. Menurut (I et al., 2012), analisis Sentimen atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif. Covid-19 adalah penyakit jenis baru yang.

2.5 *Naïve Bayes Classifier*

Algoritma naive bayes classifier merupakan algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasi data uji pada kategori yang paling tepat (Feldman & Sanger, 2007). *Naive Bayes Classifier* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas, (Patil & Sherekar, 2013). Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Persamaan Algoritma Naive Bayes Classifier adalah, (Bustami, 2013)

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (2.4)$$

Di mana :

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

Untuk mengetahui nilai dari data yang akan diprediksi akan dilakukan penghitungan pemetaan data menggunakan $Vmap$, perhitungan nilai $Vmap$ dilakukan dengan menggunakan persamaan sebagai berikut:

$$Vmap = V_j\{Positif, Negatif\}P(V_j) \prod P(X_i | V_j) \quad (2.5)$$

2.6 Covid-19

Corona virus merupakan virus yang dapat menyebabkan gejala ringan sampai berat. setidaknya ada 2 jenis corona virus yang diketahui dapat menyebabkan penyakit dengan gejala berat seperti *Middle East Respiratory Syndrome* (MERS) dan *Severe Acute Respiratory Syndrome* (SARS). Covid-19 adalah penyakit yang belum pernah diidentifikasi sebelumnya, virus penyebab Covid-19 ini disebut dengan Sars-Cov-2. SARS dibawa atau ditransmisikan oleh luwak ke manusia dan MERS dari unta ke manusia, (Organization, 2020).

2.7 Twitter dan Twitter API

2.7.1 Pengertian Twitter

Menurut (Ahuja & Dubey, 2017), Twitter merupakan *platform* media sosial yang digunakan orang-orang diseluruh dunia sebagai tempat yang tepat untuk menyampaikan pendapat mereka. Menurut (Hasan, 2017), Twitter adalah suatu jejaring sosial yang membatasi penggunaannya untuk mengirim sebuah tweet dengan 140 karakter tidak lebih. Twitter merupakan layanan jejaring sosial yang berguna untuk menghubungkan antara pengguna satu dengan pengguna lainnya.

2.7.2 Struktur Data Twitter

Pesan *twitter* memiliki banyak ragam struktur data.(A. Go & L. Huang, 2009), menjabarkan karakteristik *twitter* sebagai berikut :

1. Pada pesan *twitter*, setiap *tweet* hanya berisi panjang maksimal 140 karakter, dengan nilai rata-rata panjang setiap *tweet* dimana diketahui rata-rata *tweet* adalah 14 kata atau 78 karakter.
2. Data *twitter* dapat bersumber dari beberapa tempat. Dengan *twitter* API data dengan mudah didapat.
3. Pengguna *twitter* dapat dengan mudah menggunakan media apapun untuk menulis dan mengirimkan pesan mereka, termasuk penggunaan media ponsel. Kemunculan kesalahan penulisan ataupun penggunaan bahasa asing jauh lebih tinggi. Terdapat ragam topik didalamnya.

Menurut (Davido & Tsur, 2010), menyimpulkan bahwa sebuah *tweet* biasanya mengandung alamat URL, alamat pengguna twitter yang disebut *username* (@+*username*), atau konten *tag* yang disebut *hashtag* (#), dan *emoticon*. Penggunaan *hashtag* dan *emoticon* dianggap juga dapat tidak mewakili dari sentimen dalam sebuah *tweet* (A. Go & L. Huang, 2009)

2.7.3 Pengertian Twitter API

Twitter Application Programming Interface (API) merupakan fungsi-fungsi untuk menggantikan bahasa yang digunakan dalam *system calls* dengan bahasa yang lebih terstruktur dan mudah dimengerti oleh programmer. Fungsi yang dibuat dengan menggunakan API kemudian akan memanggil *system calls* sesuai dengan sistem operasinya. Tidak menutup kemungkinan nama dari *system calls* sama

dengan nama di API (Arifin & Suryatul). API twitter terdiri dari 3 (tiga) bagian yaitu:

1. *Search API*

Search API dirancang untuk memudahkan *user* dalam mengelola *query search* di konten twitter. *User* dapat menggunakannya untuk mencari *tweet* berdasarkan *keyword* khusus atau mencari *tweet* lebih spesifik berdasarkan *username twitter*. *Search API* menyediakan akses pada *Data Trending Topic*.

2. *Representational State Transfer (REST) API*

Representational State Transfer (REST) API memperbolehkan *Developer* untuk mengakses inti dari twitter seperti *timeline*, *status update*, dan informasi pengguna. *Representational State Transfer (REST) API* digunakan dalam membangun sebuah aplikasi twitter yang kompleks yang memerlukan inti dari *twitter*.

3. *Streaming API*

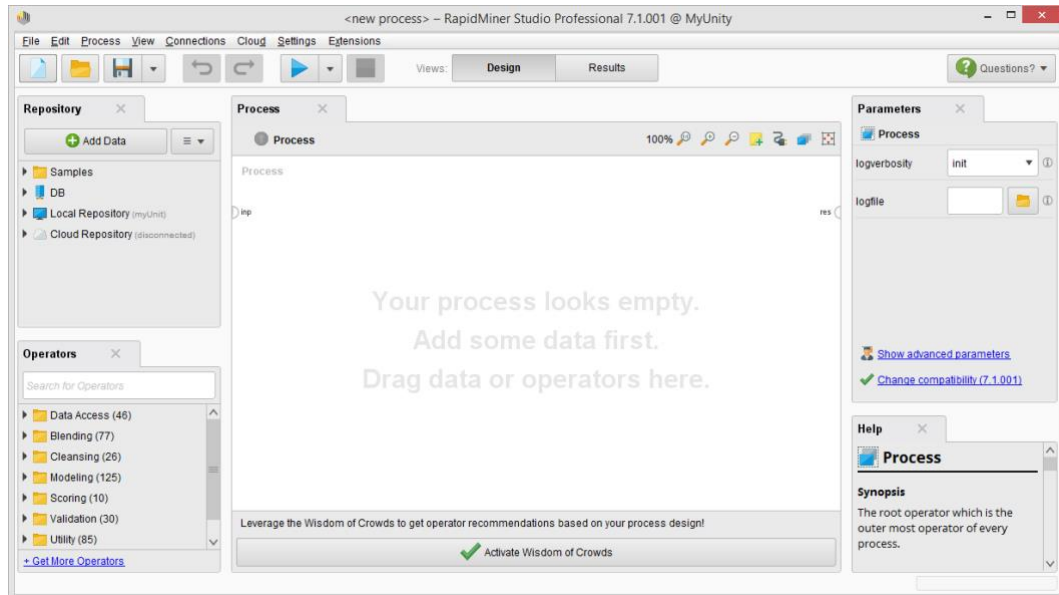
Streaming API memperbolehkan *developer* untuk kebutuhan yang lebih intensif seperti melakukan penelitian dan analisis data. *Streaming API* menghasilkan aplikasi yang dapat mengetahui statistik *status update*, *follower*, dan lain sebagainya.

2.8 Rapid Miner

Penelitian ini menggunakan *tools Rapid Miner 7.1* untuk menerapkan *association rule* terhadap data penelitian yang digunakan. Rapid Miner adalah *software* yang bersifat *open source* untuk melakukan analisis terhadap *data mining*, *text mining*, dan analisis prediksi, (L. Elvitaria & M. Havenda, 2017). Rapid Miner merupakan sebuah solusi untuk melakukan analisis terhadap data mining dengan

menggunakan teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik, (C, 2013) .

Tampilan dari Rapid Miner 7.1 dapat dilihat pada gambar 2.6.



Gambar 2. 7 Tampilan Rapid Miner

BAB III METODOLOGI PENELITIAN

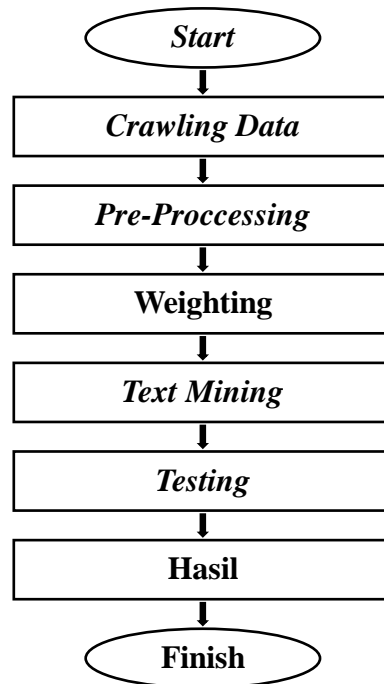
3.1 Analisis Masalah

Data opini masyarakat di Twitter jika diolah dan dianalisis dengan benar maka akan sangat penting dan bermanfaat bagi masyarakat atau organisasi-organisasi pemerintah yang dapat digunakan untuk negara Indonesia dalam memperbaiki sistem penanganan omicron saat ini. Data opini masyarakat di Twitter dapat ditarik atau diambil menggunakan API Twitter dengan menggunakan kata kunci atau *hashtag* sehingga data yang di dapatkan sesuai dengan topik yang akan dibahas. Namun banyaknya data opini yang didapatkan membuat data menjadi tidak terstruktur dan terkelompok.

Hal tersebut menunjukkan perlunya suatu teknik atau metode yang dapat digunakan untuk menganalisis data opini tersebut. Salah satu teknik yang dapat digunakan untuk menganalisis opini adalah *text mining*. *Text mining* merupakan salah satu teknik dari *data mining* yang biasa digunakan untuk melakukan analisis sentimen maka teknik ini dapat digunakan untuk melakukan analisis opini masyarakat terhadap kasus omicron di Indonesia.

3.2 Kerangka Tahapan Penelitian

Kerangka tahapan penelitian ini dibuat oleh penulis sebagai acuan terstruktur pelaksanaan penelitian. Kerangka Tahapan Penelitian ini disesuaikan dengan tahapan *Text Mining*. Berikut ini merupakan tahapan penelitian yang akan penulis laksanakan tertera pada gambar 3.1.



Gambar 3. 1 Kerangka Tahapan Penelitian

Berikut ini merupakan penjelasan tahapan-tahapan dari kerangka penelitian diatas:

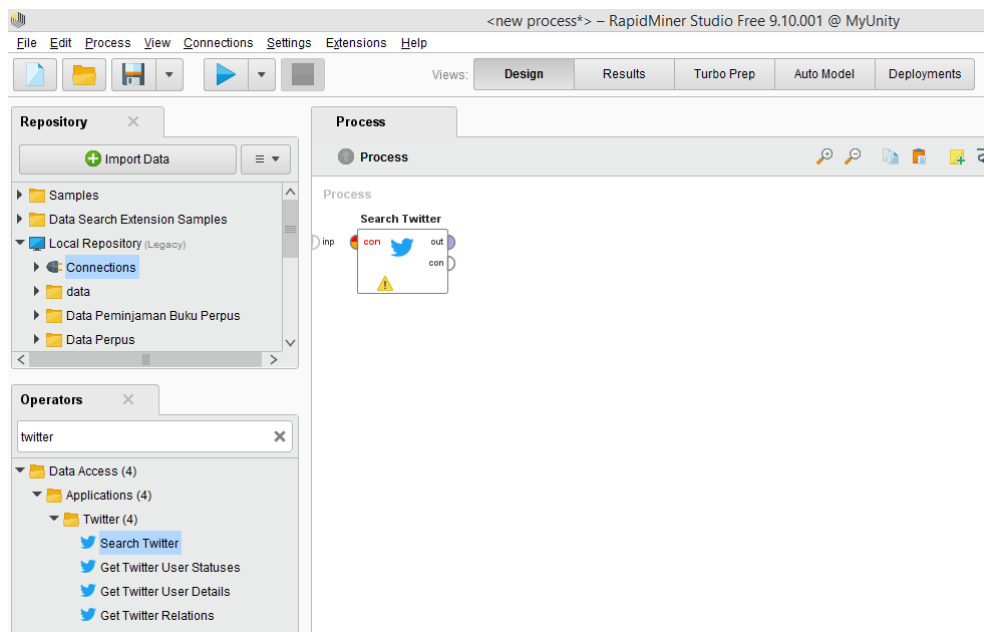
1. *Crawling Data*, yaitu mencari data *twitter* berisi opini masyarakat yang menggunakan *software* Rapid Miner.
2. *Pre-Processing*, melakukan *Cleansing*, *Case Folding*, *Tokenizing*, dan *Stemming* untuk menghasilkan data bersih.
3. *Weighting*, melakukan pembobotan pada data yang berbentuk teks.
4. *Text Mining*, melakukan penggalian data menggunakan Algoritma *Naive Bayes Classifier*.
5. Pengujian, melakukan pengukuran performa algoritma dengan menghitung *Accuracy*, *Recall*, *Precision*.
6. Mendapatkan hasil Klasifikasi dari analisis sentimen.

3.3 Proses Penarikan Data (*Crawling Data*)

Untuk mendapatkan data komentar dari twitter perlu meminta akses ke API Twitter dimana sebelum itu harus memiliki akun twitter, selain itu pastikan perangkat terhubung dengan akses internet. Proses penarikan atau pengambilan data komentar dari twitter (*crawling data*) dilakukan dengan menggunakan *software* Rapid Miner 9.1. Berikut ini merupakan tahapan *crawling data*:

1. Gunakan Operator Search Twitter

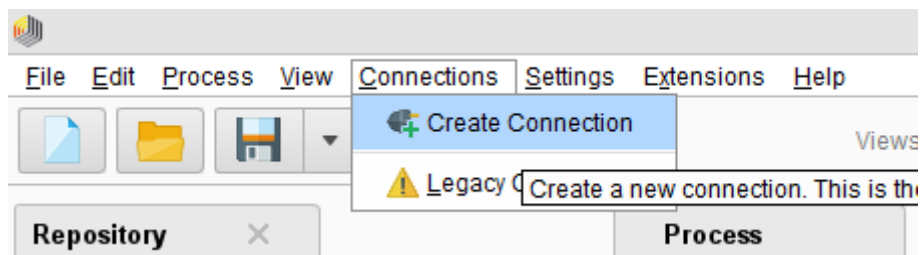
Lakukan pencarian operator pada *operator search bar* untuk mengidentifikasi pengambilan data merupakan data dari twitter lalu drag ke halaman proses.



Gambar 3. 2 Search Twitter Operator

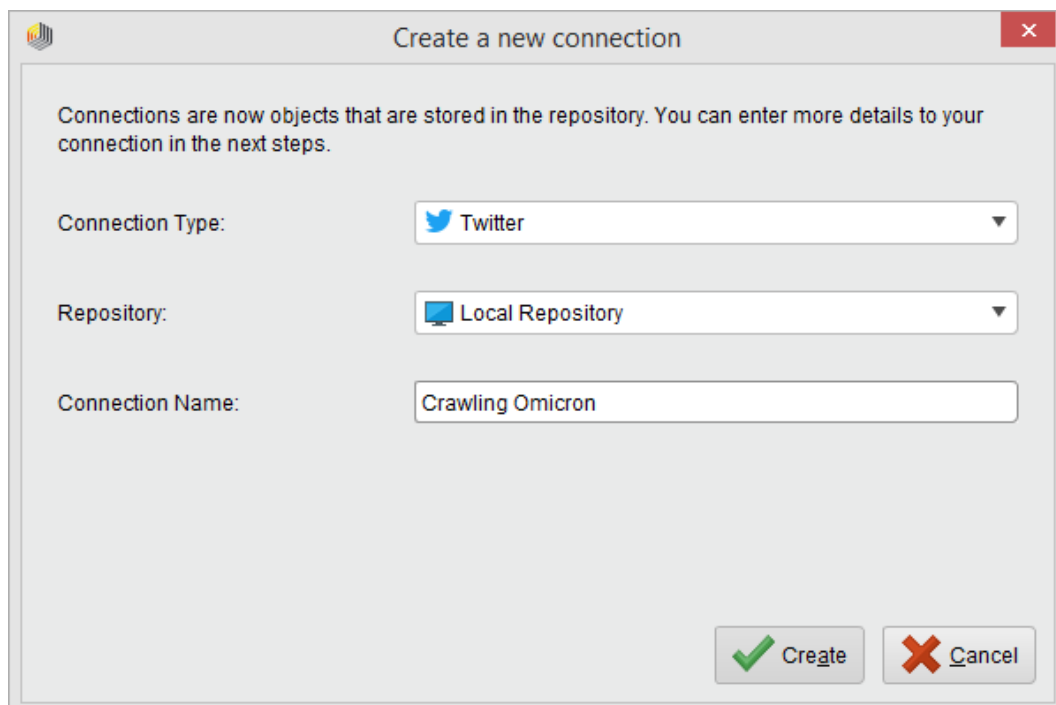
2. Buat Koneksi

Untuk melakukan penarikan data perlu membuat koneksi dari Rapid Miner ke API Twitter dengan membuat koneksi baru pada tab *Connection*.



Gambar 3. 3 Membuat Koneksi Baru

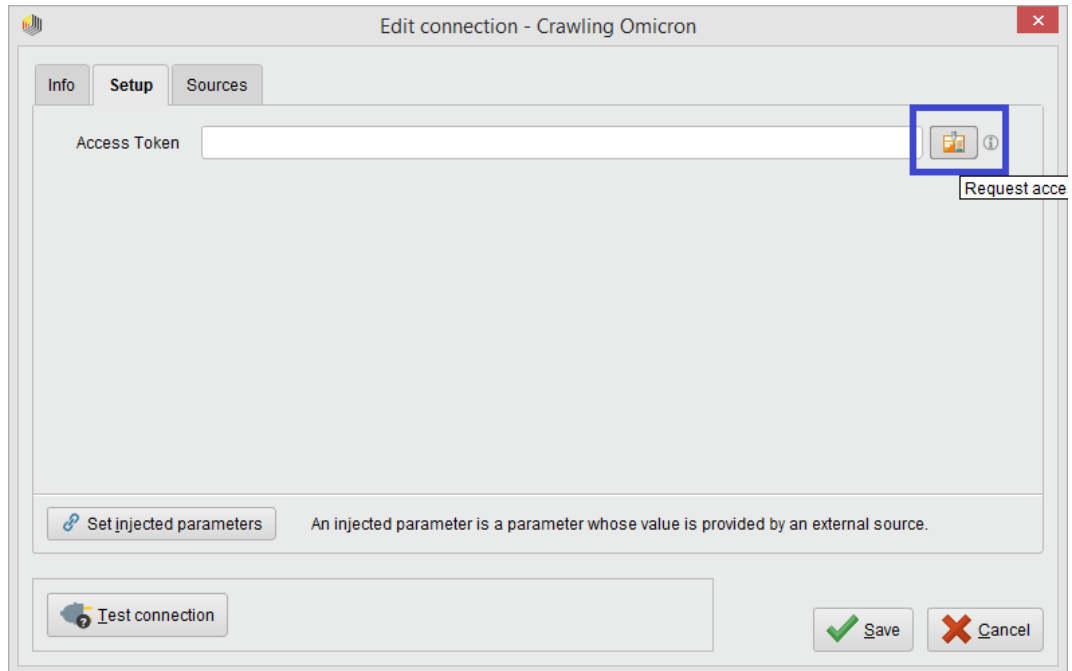
Selanjutnya pilih tipe koneksi pada *Connection Type* menjadi Twitter, pilih *Repository* pada *Local Repository* dan berikan nama koneksi tersebut lalu klik *create*.



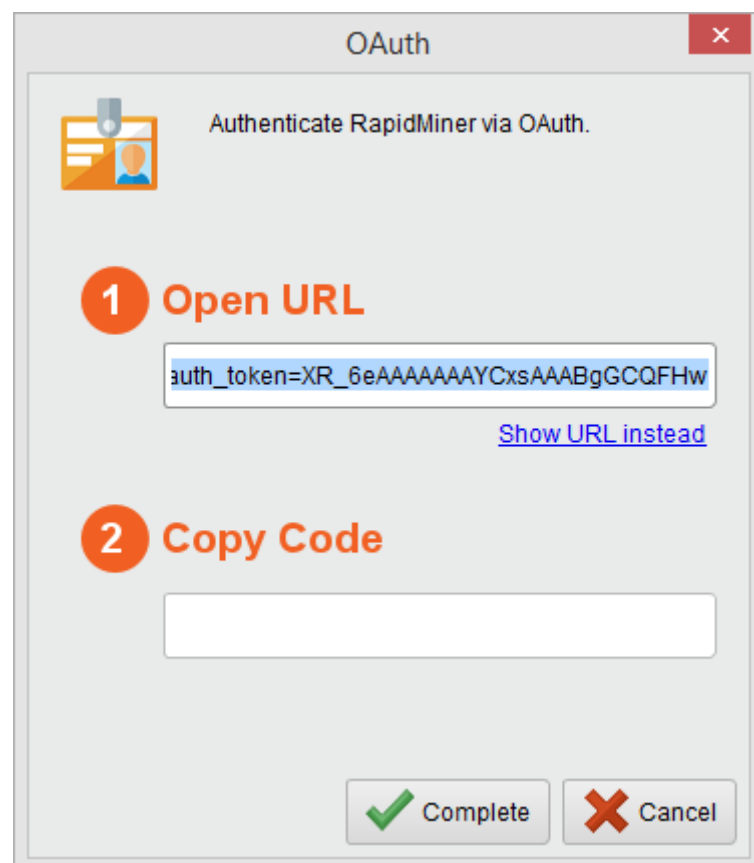
Gambar 3. 4 Memilih Tipe Koneksi

3. Buat Permintaan Akses Token

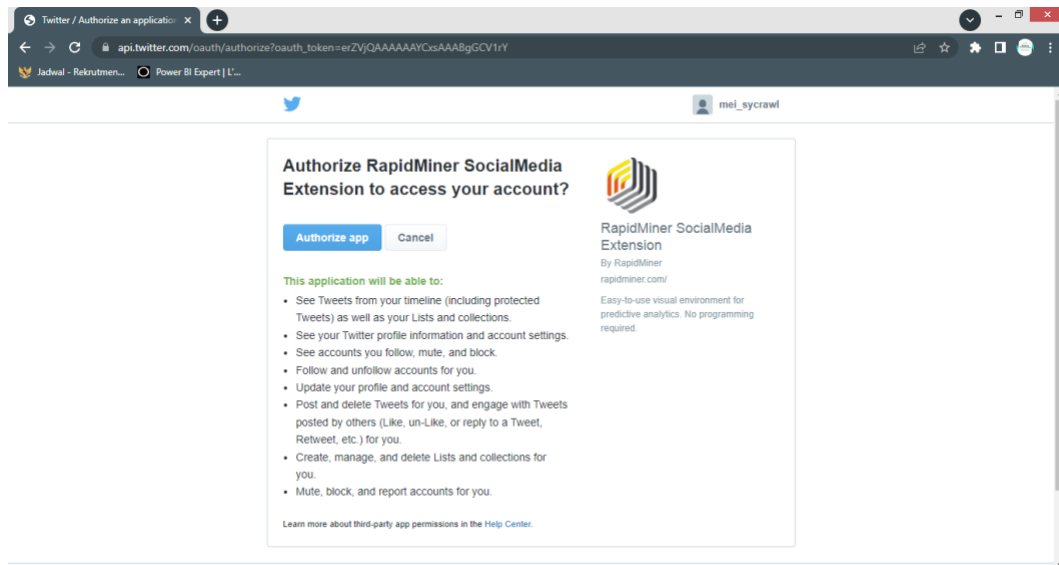
Setelah berhasil membuat koneksi maka selanjutnya adalah meminta akses token ke API Twitter. Pada tahap ini *software* rapid miner akan meminta akses kepada API Twitter untuk dapat mengambil data. dan API Twitter akan memberikan Token Akses kepada Rapid Miner.



Gambar 3.5 Request Access Token

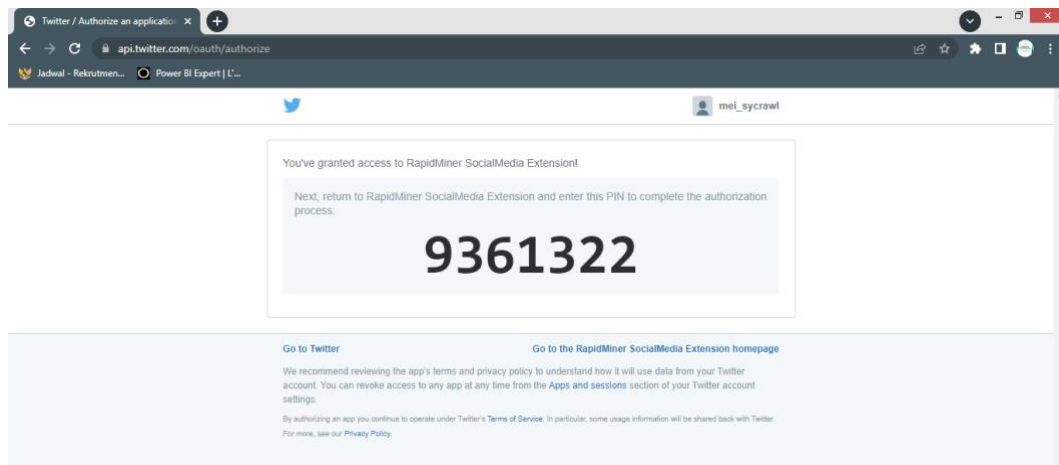


Gambar 3.6 Autentikasi



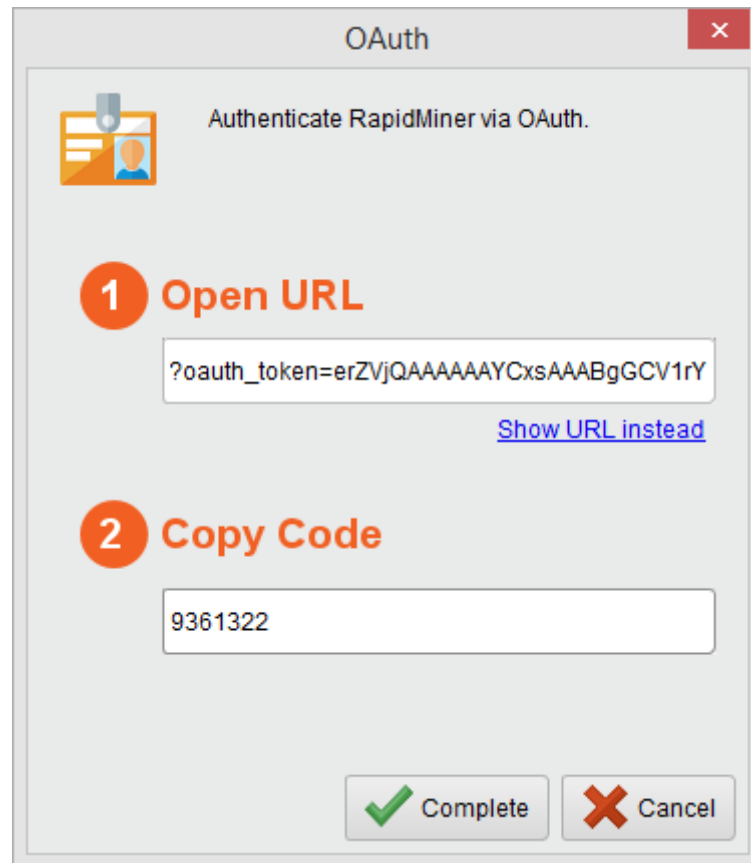
Gambar 3. 7 Otorisasi Akun Twitter

Setelah mendapatkan URL untuk meminta akses token maka *copy* URL tersebut ke *browser* untuk otorisasi akun twitter yang akan digunakan untuk pengambilan data. Jika sudah melakukan otorisasi maka akan ditampilkan token akses seperti dibawah ini.

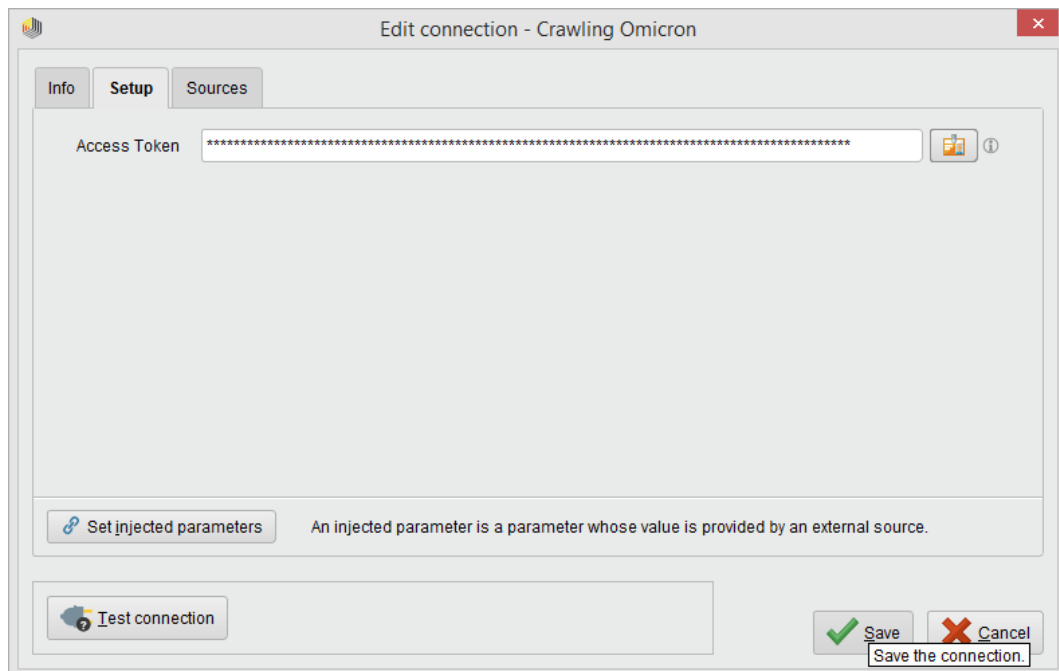


Gambar 3. 8 Token Akses

Selanjutnya salin token akses tersebut dan *paste* kembali ke Autentikasi Rapid Miner dan klik *complete*. Maka akan tampil *Access Token* akan terisi, pada tahap ini koneksi bisa disimpan dengan klik tombol *save*.



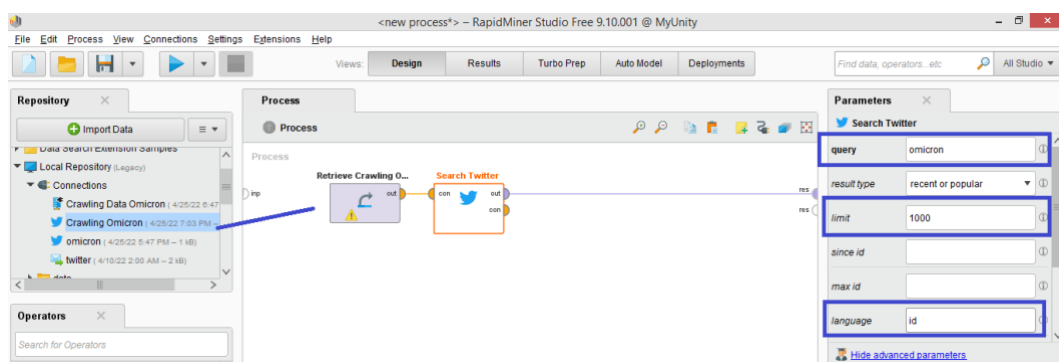
Gambar 3. 9 Copy Code Akses Token



Gambar 3. 10 Akses Token Terisi

4. Buat Proses Penarikan Data

Selanjutnya gunakan koneksi yang telah dibuat dan hubungkan dengan operator *Search Twitter* untuk memproses pengambilan data. Pada *Parameters* isi *query* dengan kata kunci yang akan digunakan yaitu “omicon”, *limit* maksimal pada penarikan data menggunakan Rapid Miner adalah 1000, dan pada *language* isi dengan menggunakan kode bahasa indonesia yaitu “id” sehingga data komentar yang diambil hanyalah yang berbahasa Indonesia.



Gambar 3. 11 Buat Proses

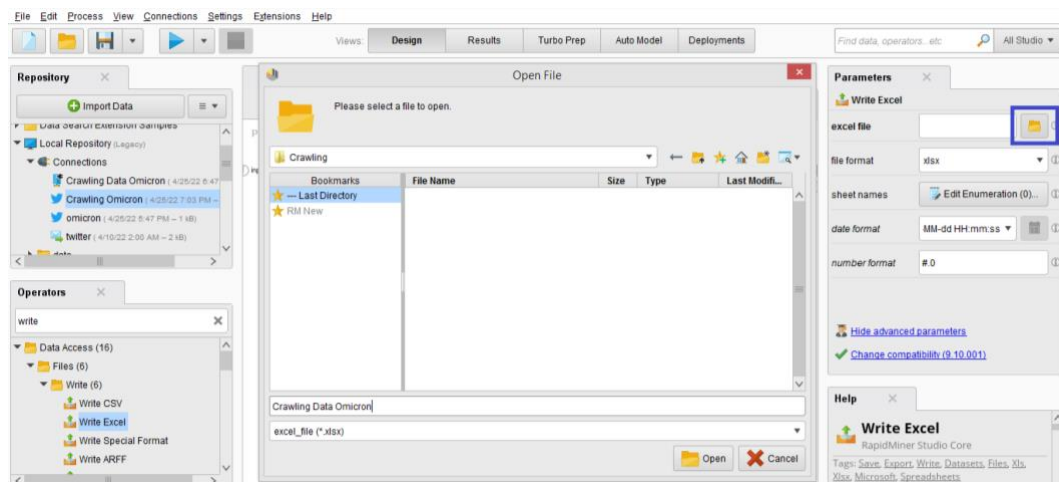
Setelah proses operator terskoneksi lakukan *running process* maka akan muncul hasil penarikan data seperti pada gambar berikut.

Created-At	From-User	From-User-Id	To-User	To-User-Id	Language	Source	Text	Geo-Locatio...	Geo-Locatio...	Retwe
18346340...	Apr 25, 2022	Noor Hisham...	1595149746	?	-1	in	<a href="http://...	Kini kita berada di fasa peralih...	?	111
15629770...	Apr 17, 2022	Noor Hisham...	1595149746	?	-1	in	<a href="http://...	Varian omicon adalah kurang ...	?	455
16634021...	Apr 20, 2022	Jabatan Pen...	114914506	?	-1	in	<a href="http://...	Vaksinasi dan dos penggalak L...	?	70
18558668...	Apr 25, 2022	baso amir	54161214	?	-1	in	<a href="http://...	Memang Omicon? :) https://t.c...	?	0
18558076...	Apr 25, 2022	Cream Beast	1477894954	pawseidonn	8742271605...	in	<a href="http://...	@pawseidonn @Ttmibb ltu o...	?	0
18557595...	Apr 25, 2022	David Sidharta	1312522381...	detikcom	69183155	in	<a href="http://...	@detikcom Karena china terfal...	?	0
18556056...	Apr 25, 2022	Rafik Alhaqi	1515501450...	?	-1	in	<a href="http://...	dulu corona	?	0
18554154...	Apr 25, 2022	Bagus	8171949072...	bengxzi	1135778225...	in	<a href="http://...	@bengxzi @haylortan Amaan k...	?	0
18553888...	Apr 25, 2022	Mik Jeger	7563674028...	?	-1	in	<a href="http://...	RT @tvOneNews: Subvarian B...	?	11
18552985...	Apr 25, 2022	Lyra	1307827626...	?	-1	in	<a href="http://...	RT @DGHisham: Kini kita ber...	?	111
18552966...	Apr 25, 2022	Tras el valle	9626634991...	?	-1	in	<a href="http://...	RT @jaranaz94: "Casi 300"	?	28
18552959...	Apr 25, 2022	Tras el valle	9626634991...	?	-1	in	<a href="http://...	RT @TrasElValle_OBC: "Casi ...	?	27
18551591...	Apr 25, 2022	Farah F Khu...	233888000	?	-1	in	<a href="http://...	little did i know abis nge tweet L...	?	0
18551362...	Apr 25, 2022	Rhea Para	1316995278...	?	-1	in	<a href="http://...	RT @DGHisham: Kini kita ber...	?	111

Gambar 3. 12 Hasil Crawling Data

5. Simpan Data

Setelah data berhasil didapat maka data dapat disimpan ke dalam file CSV dengan menggunakan operator Writer CSV lalu dihubungkan dengan koneksi *Search Twitter*. Pilih direktori untuk menyimpan file dengan format CSV.



Gambar 3. 13 Menyimpan Data Hasil Crawling

3.4 Pre-processing Data

Pre-processing Data merupakan tahap pembersihan data dari karakter-karakter dan kata yang tidak memiliki makna. Berikut ini merupakan tampilan data yang perlu dilakukan pembersihan sehingga mendapatkan kalimat komentar yang bermakna dan bersih dari karakter-karakter yang tidak diperlukan.

Created-From-Use-From-Use-To-User	To-User-It-Language	Source	Text
##### Noor Hish 1.6E+09	-1 in	<a href="Kini kita berada di fasa peralihan endemik iaitu fasa mitigasi dalam kesihatan awam di mana virus covid-19 iaitu varian omicron berada di seki	
##### tvOneNews 55507370	-1 in	<a href="Subvarian Baru Omicron Ditemukan di China, Lebih Cepat Menular https://t.co/Sm0sESKIMA	
##### Noor Hish 1.6E+09	-1 in	<a href="Varian omicron adalah kurang ganas jika dibandingkan dgn varian delta tetapi masih berbahaya terutamanya kepada mereka yg tidak mendap	
##### Basmi Ros 2.92E+08	-1 in	<a href="WHO melaporkan tren kasus global terus menurun signifikan sejak akhir Maret. Namun, China masih 'diamuk' Omicron. Pakar	
##### baso amir 54161214	-1 in	<a href="Memang Omicron?): https://t.co/FIODt6Mryj	
##### Cream Be 1.48E+18 pawseido 8.74E+17 in	<a href="pawseidonn @Ttxtmlbb itu omicron kak. Ada tanduknya		
##### David Sidi 1.31E+18 detikcom 69183155 in	<a href="detikcom Karena china terlalu pingin bersihBiasa nya beberapa orang kena lockdown satu kota.....padahal biarin aja kena covid apalagi o		
##### Rafi Alhaq 1.52E+18	-1 in	<a href="dulu corona	
##### Bagus 8.17E+17 bengzxi 1.14E+18 in	<a href=" @bengzxi @haylorton Amaan koo pas yg kasus omicron sama delta naik aku sering renang		
##### Mik Jeger 7.56E+17	-1 in	<a href="RT @tvOneNews: Subvarian Baru Omicron Ditemukan di China, Lebih Cepat Menular https://t.co/Sm0sESKIMA	
##### Lyra 1.31E+18	-1 in	<a href="RT @DGHisham: Kini kita berada di fasa peralihan endemik iaitu fasa mitigasi dalam kesihatan awam di mana virus covid-19 iaitu varian omicr	
##### Tras el val 9.63E+17	-1 in	<a href="RT @jaranaz94: 'Casi 300'	
##### Tras el val 9.63E+17	-1 in	<a href="RT @TrasElValle_OBC: 'Casi 300'	
##### Farah F Kh 2.34E+08	-1 in	<a href="Little did i know abis nge tweet ini besoknya saya omicron ? https://t.co/x6HdZTD0zG	
##### Rhea Para 1.32E+18	-1 in	<a href="RT @DGHisham: Kini kita berada di fasa peralihan endemik iaitu fasa mitigasi dalam kesihatan awam di mana virus covid-19 iaitu varian omicr	
##### Virza Sani 1.69E+08	-1 in	<a href="RT @detikcom: WHO melaporkan tren kasus global terus menurun signifikan sejak akhir Maret. Namun, China masih 'diamuk' Omicron. Pakar	
##### SUJANIN R 7.61E+17 AgesAgu 1.02E+18 in	<a href=" @AgesAgus Itulah tujuan mereka untuk menguatkan isu yg mereka bangun tentang islam phobia, toh yg mereka tunggu adalah 'arangan n		
##### Kimberley 1.33E+18 Embunpat 1.41E+18 in	<a href=" @Embunpat01390476 Disana lagi di lock down jadi ga berani pulang... mau besaran omicron di sini.		
##### detikHeal 1.04E+08	-1 in	<a href="Kasus kematian COVID-19 dan infeksi pasien bergejala makin meningkat di Shanghai. Banyak warga berteriak frustrasi akibat lockdown ketat. f	
##### ROMO 1.19E+08	-1 in	<a href="RT @sunnyhujan: @jowohitpost ternyata orang yang sama, konser waktu kasus omicron lagi naik https://t.co/RzuLinkOm	
##### Just Fika 69082638	-1 in	<a href="RT @detikcom: WHO melaporkan tren kasus global terus menurun signifikan sejak akhir Maret. Namun, China masih 'diamuk' Omicron. Pakar	
##### detikcom 69183155	-1 in	<a href="WHO melaporkan tren kasus global terus menurun signifikan sejak akhir Maret. Namun, China masih 'diamuk' Omicron. Pakar mengungkap bl	

Gambar 3. 14 Data Sebelum Pre-Processing

Tahap *pre-processing* dilakukan berfokus pada isi komentar twitter. Untuk memperjelas penjelasan pada proses ini maka penulis akan mengambil contoh 7 komentar untuk dijelaskan tahap *pre-processing* nya. Berikut ini merupakan tabel berisi teks komentar twitter.

Tabel 3. 1 Komentar Twitter

No.	Text
1	Subvarian Baru Omicron Ditemukan di Cina, Lebih Cepat Menular https://t.co/Sm0sESKiMA
2	Memang Omicron? :) https://t.co/FlODt6Mryj
3	@pawseidonn @Txtmlbb Itu omicron kak. Ada tanduknya
4	@detikcom Karena china terlalu pingin bersihBiasa nya beberapa orang kena lockdown satu kota.....padahal biarin aja kena covid apalagi omicron yg cuma gejala ringanitung2 vaksin ulang
5	dulu corona sekarang omicron
6	@bengxzi @haylortan Amaan koo pas yg kasus omicron sama delta naik aku sering renang
7	little did i know abis nge tweet ini besoknya saya omicron ? https://t.co/x6Hd2TdOzG

Sebelum dilakukan tahap *pre-processing* isi komentar-komentar tersebut diberikan label sentimen secara manual dan mengubah bahasa kedalam bahasa Indonesia yang baku. Label sentimen yang digunakan oleh penulis pada penelitian ini merupakan Sentimen Positif, Sentimen Netral, dan Sentimen Negatif.

Tabel 3. 2 Memberikan Label Sentimen

No.	Text	Sentimen
1	Subvarian Baru Omicron Ditemukan di Cina, Lebih Cepat Menular https://t.co/Sm0sESKiMA	Negatif
2	Memang Omicron? :) https://t.co/FlODt6Mryj	Netral
3	@pawseidonn @Txtmlbb Itu omicron kak. Ada tanduknya	Negatif
4	@detikcom Karena china terlalu pingin bersihBiasa nya beberapa orang kena lockdown satu kota.....padahal biarin aja kena covid apalagi omicron yg cuma gejala ringanitung2 vaksin ulang	Positif
5	dulu corona sekarang omicron	Negatif
6	@bengxzi @haylortan Amaan koo pas yg kasus omicron sama delta naik saya sering renang	Positif
7	little did i know abis nge tweet ini besoknya saya omicron ? https://t.co/x6Hd2TdOzG	Negatif

Setelah data diberikan label sentimen maka dilanjutkan ke tahap *Pre-processing*, dimana terdiri dari tahap-tahap berikut:

1. *Cleansing*

Cleansing merupakan tahap yang bertujuan untuk menghilangkan karakter atau simbol, *link url*, *username*, *mention*, *hashtag*, *retweet*, dan *emoticon*.

Tabel 3. 3 Hasil *Cleansing*

No.	Text	Hasil Cleansing
1	Subvarian Baru Omicron Ditemukan di Cina, Lebih Cepat Menular https://t.co/Sm0sESKiMA	Subvarian Baru Omicron Ditemukan di Cina, Lebih Cepat Menular
2	Memang Omicron? :) https://t.co/FIODt6Mryj	Memang Omicron?
3	@pawseidonn @Txmlbb Itu omicron kak. Ada tanduknya	Itu omicron kak. Ada tanduknya
4	@detik.com Karena cina terlalu ingin bersih, beberapa orang yang terjangkit corona akan di lockdown satu kota. sebenarnya dibiarkan saja jika hanya gejala ringan, anggap saja vaksin ulang	Karena cina terlalu ingin bersih, beberapa orang yang terjangkit corona akan di lockdown satu kota. sebenarnya dibiarkan saja jika hanya gejala ringan, anggap saja vaksin ulang
5	dulu corona sekarang omicron	dulu corona sekarang omicron
6	@bengxzi @haylortan Aman kok diwaktu kasus omicron dan delta naik saya sering renang	Aman kok diwaktu kasus omicron dan delta naik saya sering renang
7	siapa yang akan tahu sehabis membuat tweet ini besoknya saya omicron ? https://t.co/x6Hd2TdOzG	siapa yang akan tahu sehabis membuat tweet ini besoknya saya omicron ?

2. *Case Folding*

Case Folding merupakan tahapan dimana seluruh *case* didalam data sampel yang telah di *cleansing* akan diganti menjadi bentuk standar. Bentuk standar yang dimaksud adalah dengan mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya huruf “a” sampai dengan “z” yang diterima. Sedangkan karakter lainnya dianggap sebagai delimiter atau pembatas.

Tabel 3. 4 Hasil *Case Folding*

No.	Text Cleansing	Hasil Case Folding
1	Subvarian Baru Omicron Ditemukan di Cina, Lebih Cepat Menular	subvarian baru omicron ditemukan di cina, lebih cepat menular
2	Memang Omicron?	memang omicron?
3	Itu omicron kak. Ada tanduknya	itu omicron kak. ada tanduknya
4	Karena cina terlalu ingin bersih, beberapa orang yang terjangkit corona akan di lockdown satu kota. sebenarnya dibiarkan saja jika hanya gejala ringan, anggap saja vaksin ulang	karena cina terlalu ingin bersih, beberapa orang yang terjangkit corona akan di lockdown satu kota. sebenarnya dibiarkan saja jika hanya gejala ringan, anggap saja vaksin ulang
5	dulu corona sekarang omicron	dulu corona sekarang omicron
6	Aman kok diwaktu kasus omicron dan delta naik saya sering renang	aman kok diwaktu kasus omicron dan delta naik saya sering renang
7	siapa yang akan tahu sehabis membuat tweet ini besoknya saya omicron ?	siapa yang akan tahu sehabis membuat tweet ini besoknya saya omicron ?

3. *Tokenizing*

Tahap *tokenizing* atau bisa disebut *parsing* merupakan tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Tanda baca akan dihilangkan sehingga menghasilkan sebuah kalimat atau kata yang berdiri sendiri. Tabel dibawah ini merupakan hasil dari proses *tokenizing*.

Tabel 3. 5 Hasil *Tokenizing*

No.	Text Case Folding	Hasil Tokenizing
1	subvarian baru omicron ditemukan di cina, lebih cepat menular	subvarian baru omicron ditemukan di cina lebih cepat menular
2	memang omicron?	memang omicron
3	itu omicron kak. ada tanduknya	itu omicron kak ada tanduknya
4	karena cina terlalu ingin bersih, beberapa orang yang terjangkit corona akan di lockdown satu kota. sebenarnya dibiarkan saja jika hanya gejala ringan, anggap saja vaksin ulang	karena cina terlalu ingin bersih beberapa orang yang terjangkit corona akan di lockdown satu kota sebenarnya dibiarkan saja jika hanya gejala ringan anggap saja vaksin ulang

Tabel 3. 4 Hasil *Tokenizing* (lanjutan)

No.	Text Case Folding	Hasil Tokenizing
5	dulu corona sekarang omicron	dulu corona sekarang omicron
6	aman kok diwaktu kasus omicron dan delta naik saya sering renang	aman kok diwaktu kasus omicron dan delta naik saya sering renang
7	siapa yang akan tahu sehabis membuat tweet ini besoknya saya omicron ?	siapa yang akan tahu sehabis membuat tweet ini besoknya saya omicron

4. *Filtering*

Dalam tahap ini akan dilakukan *filtering* untuk mengambil kata penting. Digunakan algoritma *stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata yang penting). Kata-kata yang dapat dihilangkan adalah kata-kata yang tidak deskriptif dalam pendekatan *bag-of-word*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari”, dan sebagainya.

Tabel 3. 6 Hasil *Filtering*

No.	Text Tokenizing	Hasil Filtering
1	subvarian baru omicron ditemukan di cina lebih cepat menular	subvarian baru omicron ditemukan cina lebih cepat menular
2	memang omicron	memang omicron
3	itu omicron kak ada tanduknya	omicron tanduknya
4	karena cina terlalu ingin bersih beberapa orang yang terjangkit corona akan di lockdown satu kota sebenarnya dibiarkan saja jika hanya gejala ringan anggap saja vaksin ulang	cina bersih orang terjangkit corona lockdown kota sebenarnya dibiarkan gejala ringan anggap vaksin ulang
5	dulu corona sekarang omicron	dulu corona sekarang omicron
6	aman kok diwaktu kasus omicron dan delta naik saya sering renang	aman diwaktu kasus omicron delta naik saya renang
7	siapa yang akan tahu sehabis membuat tweet ini besoknya saya omicron	sehabis tweet besoknya saya omicron

5. *Stemming*

Tahap *stemming* adalah tahap mencari akar atau dasar kata dari tiap kata hasil *filtering*. Pada tahap ini seluruh kata yang memiliki imbuhan akan diubah menjadi kata dasar.

Tabel 3. 7 Hasil *Stemming*

No.	Text Filtering	Hasil <i>Stemming</i>
1	subvarian baru omicron ditemukan cina lebih cepat menular	subvarian baru omicron temu cina lebih cepat menular
2	memang omicron	memang omicron
3	omicron tanduknya	omicron tanduknya
4	cina bersih orang terjangkit corona lockdown kota sebenarnya dibiarkan gejala ringan anggap vaksin ulang	cina bersih orang jangkit corona lockdown kota benar biar gejala ringan anggap vaksin ulang
5	dulu corona sekarang omicron	dulu corona sekarang omicron
6	aman diwaktu kasus omicron delta naik saya renang	aman waktu kasus omicron delta naik saya renang
7	sehabis tweet besoknya saya omicron	sehabis tweet besok saya omicron

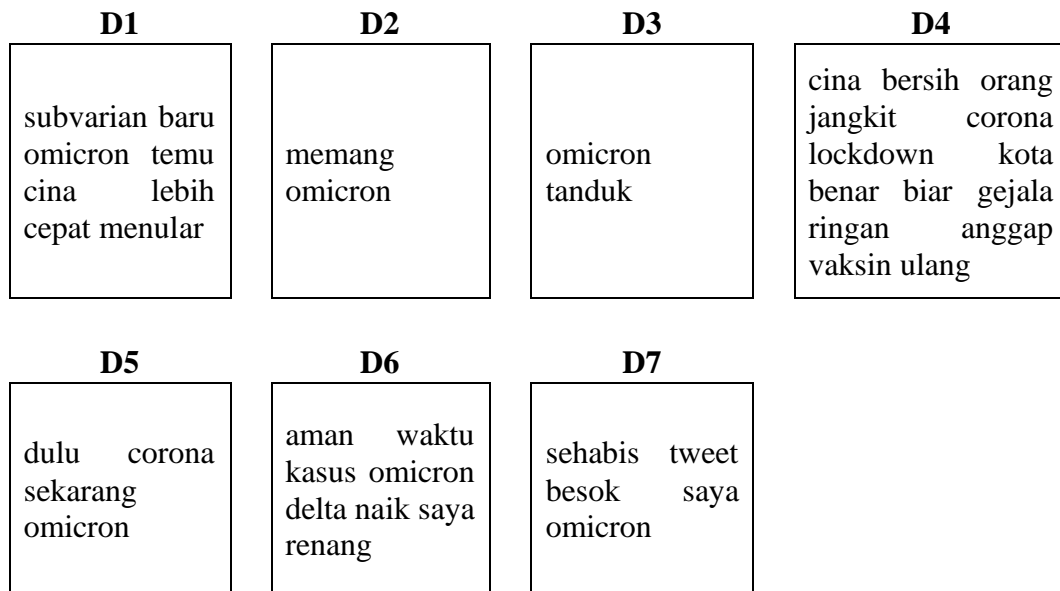
Setelah dilakukan tahap *pre-processing* maka didapatkan data yang akan digunakan untuk proses selanjutnya. Berikut ini merupakan data yang telah bersih setelah dilakukan *pre-processing*.

Tabel 3. 8 Hasil *Pre-Processing*

No.	Hasil <i>Pre-Processing</i>
1	subvarian baru omicron temu cina lebih cepat menular
2	memang omicron
3	omicron tanduknya
4	cina bersih orang jangkit corona lockdown kota benar biar gejala ringan anggap vaksin ulang
5	dulu corona sekarang omicron
6	aman waktu kasus omicron delta naik saya renang
7	sehabis tweet besok saya omicron

3.5 Weighting (Algoritma TF-IDF)

Dalam proses pembobotan ini setiap *record* atau baris kalimat akan disebut sebagai Dokumen (D). Berdasarkan hasil *pre-processing* maka berikut ini merupakan data yang akan digunakan untuk proses pembobotan (*weighting*).



Gambar 3. 15 Data Dalam Bentuk Dokumen

Mencari nilai bobot dari seluruh dokumen dilakukan dengan langkah-langkah Algoritma TF-IDF sebagai berikut:

1. *Term Frequent*

Pada tahap ini menghitung frekuensi kemunculan kata pada masing-masing dokumen. berikut ini merupakan frekuensi kemunculan dari setiap kata yang ada pada seluruh dokumen.

Tabel 3. 9 Term Frequent Seluruh Kata

	TF						
	D1	D2	D3	D4	D5	D6	D7
subvarian	1	0	0	0	0	0	0
baru	1	0	0	0	0	0	0
omicron	1	1	1	0	1	1	1

Tabel 3. 9 *Term Frequent* Seluruh Kata (lanjutan)

	TF						
	D1	D2	D3	D4	D5	D6	D7
temu	1	0	0	0	0	0	0
cina	1	0	0	1	0	0	0
lebih	1	0	0	0	0	0	0
cepat	1	0	0	0	0	0	0
menular	1	0	0	0	0	0	0
memang	0	1	0	0	0	0	0
tanduk	0	0	1	0	0	0	0
bersih	0	0	0	1	0	0	0
orang	0	0	0	1	0	0	0
jangkit	0	0	0	1	0	0	0
corona	0	0	0	1	1	0	0
lockdown	0	0	0	1	0	0	0
kota	0	0	0	1	0	0	0
benar	0	0	0	1	0	0	0
biar	0	0	0	1	0	0	0
gejala	0	0	0	1	0	0	0
ringan	0	0	0	1	0	0	0
anggap	0	0	0	1	0	0	0
vaksin	0	0	0	1	0	0	0
ulang	0	0	0	1	0	0	0
dulu	0	0	0	0	1	0	0
sekarang	0	0	0	0	1	0	0
aman	0	0	0	0	0	1	0
waktu	0	0	0	0	0	1	0
kasus	0	0	0	0	0	1	0
delta	0	0	0	0	0	1	0
naik	0	0	0	0	0	1	0
saya	0	0	0	0	0	1	1
renang	0	0	0	0	0	1	0
sehabis	0	0	0	0	0	0	1
tweet	0	0	0	0	0	0	1
besok	0	0	0	0	0	0	1

2. *Inverse Document Frequency (IDF)*

Selanjutnya mencari nilai *IDF* dengan menggunakan persamaan (2.1)

sebagai berikut):

$$IDF(t) = \log\left(\frac{n}{df(t)}\right) \quad (2.1)$$

Dimana: n = total dokumen

$df(t)$ = jumlah dokumen yang mengandung kata t

Berikut ini merupakan contoh perhitungan nilai IDF dari kata “subvarian”:

Diketahui: $n = 7$

$df(t) = 1$

Maka : $IDF(t) = \log\left(\frac{7}{1}\right) = 0.8451$

Tabel 3.10 dibawah ini merupakan tabel hasil perhitungan nilai IDF dari seluruh kata.

Tabel 3. 10 Nilai IDF Seluruh Kata

	TF							DF	IDF
	D1	D2	D3	D4	D5	D6	D7		
subvarian	1	0	0	0	0	0	0	1	0.8451
Baru	1	0	0	0	0	0	0	1	0.8451
Omicron	1	1	1	0	1	1	1	6	0.0669
Temu	1	0	0	0	0	0	0	1	0.8451
Cina	1	0	0	1	0	0	0	2	0.5441
Lebih	1	0	0	0	0	0	0	1	0.8451
Cepat	1	0	0	0	0	0	0	1	0.8451
menular	1	0	0	0	0	0	0	1	0.8451
memang	0	1	0	0	0	0	0	1	0.8451
tanduk	0	0	1	0	0	0	0	1	0.8451
bersih	0	0	0	1	0	0	0	1	0.8451
orang	0	0	0	1	0	0	0	1	0.8451
jangkit	0	0	0	1	0	0	0	1	0.8451
corona	0	0	0	1	1	0	0	2	0.5441
lockdown	0	0	0	1	0	0	0	1	0.8451
kota	0	0	0	1	0	0	0	1	0.8451
benar	0	0	0	1	0	0	0	1	0.8451
biar	0	0	0	1	0	0	0	1	0.8451
gejala	0	0	0	1	0	0	0	1	0.8451
ringan	0	0	0	1	0	0	0	1	0.8451
anggap	0	0	0	1	0	0	0	1	0.8451
vaksin	0	0	0	1	0	0	0	1	0.8451
ulang	0	0	0	1	0	0	0	1	0.8451

Tabel 3. 10 Nilai IDF Seluruh Kata (lanjutan)

	TF							DF	IDF
	D1	D2	D3	D4	D5	D6	D7		
dulu	0	0	0	0	1	0	0	1	0.8451
sekarang	0	0	0	0	1	0	0	1	0.8451
aman	0	0	0	0	0	1	0	1	0.8451
waktu	0	0	0	0	0	1	0	1	0.8451
kasus	0	0	0	0	0	1	0	1	0.8451
delta	0	0	0	0	0	1	0	1	0.8451
naik	0	0	0	0	0	1	0	1	0.8451
saya	0	0	0	0	0	1	1	2	0.5441
renang	0	0	0	0	0	1	0	1	0.8451
sehabis	0	0	0	0	0	0	1	1	0.8451
tweet	0	0	0	0	0	0	1	1	0.8451
besok	0	0	0	0	0	0	1	1	0.8451

3. Weighting (TF-IDF)

Setelah nilai Tf dan IDF dari setiap kata diketahui maka dapat dilakukan pembobotan untuk setiap kata dalam dokumen dengan menggunakan persamaan (2.3) sebagai berikut:

$$W(dt) = TF(d,t) \times IDF(t) \quad (2.2)$$

Berikut ini salah satu contoh perhitungan pembobotan (W) pada D1 yang terdiri dari kata:

$$\text{“subvarian”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.8451 = 0.8451$$

$$\text{“baru”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.8451 = 0.8451$$

$$\text{“omicron”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.0669 = 0.8451$$

$$\text{“temu”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.8451 = 0.8451$$

$$\text{“cina”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.5441 = 0.8451$$

$$\text{“lebih”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.8451 = 0.8451$$

$$\text{“cepat”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.8451 = 0.8451$$

$$\text{“menular”} = W(dt) = TF(d,t) \times IDF(t) = 1 \times 0.8451 = 0.8451$$

Maka total bobot untuk D1 adalah sebagai berikut

$$D1 = 0.8451 + 0.8451 + 0.0669 + 0.8451 + 0.5441 + 0.8451 + 0.8451 + 0.8451$$

$$D1 = 5.6816$$

Tabel 3.11 dibawah ini merupakan tabel hasil pembobotan TF-IDF dari seluruh kata.

Tabel 3. 11 Hasil Pembobotan TF-IDF

	<i>Weight (TF-IDF)</i>						
	D1	D2	D3	D4	D5	D6	D7
Subvarian	0.8451	0	0	0	0	0	0
Baru	0.8451	0	0	0	0	0	0
omicron	0.0669	0.0669	0.0669	0	0.0669	0.0669	0.0669
Temu	0.8451	0	0	0	0	0	0
Cina	0.5441	0	0	0.5441	0	0	0
Lebih	0.8451	0	0	0	0	0	0
Cepat	0.8451	0	0	0	0	0	0
menular	0.8451	0	0	0	0	0	0
memang	0	0.8451	0	0	0	0	0
tanduk	0	0	0.8451	0	0	0	0
Bersih	0	0	0	0.8451	0	0	0
Orang	0	0	0	0.8451	0	0	0
jangkit	0	0	0	0.8451	0	0	0
corona	0	0	0	0.5441	0.5441	0	0
lockdown	0	0	0	0.8451	0	0	0
Kota	0	0	0	0.8451	0	0	0
Benar	0	0	0	0.8451	0	0	0
Biar	0	0	0	0.8451	0	0	0
Gejala	0	0	0	0.8451	0	0	0
Ringan	0	0	0	0.8451	0	0	0
anggap	0	0	0	0.8451	0	0	0
Vaksin	0	0	0	0.8451	0	0	0
Ulang	0	0	0	0.8451	0	0	0
Dulu	0	0	0	0	0.8451	0	0
sekarang	0	0	0	0	0.8451	0	0
Aman	0	0	0	0	0	0.8451	0
Waktu	0	0	0	0	0	0.8451	0
Kasus	0	0	0	0	0	0.8451	0
Delta	0	0	0	0	0	0.8451	0

Tab 3. 11 Hasil Pembobotan TF-IDF (lanjutan)

	<i>Weight (TF-IDF)</i>						
	D1	D2	D3	D4	D5	D6	D7
Naik	0	0	0	0	0	0.8451	0
Saya	0	0	0	0	0	0.5441	0.5441
Renang	0	0	0	0	0	0.8451	0
sehabis	0	0	0	0	0	0	0.8451
Tweet	0	0	0	0	0	0	0.8451
Besok	0	0	0	0	0	0	0.8451
Total	5.6816	0.9120	0.9120	11.2293	2.3012	5.6816	3.1463

Berdasarkan pembobotan diatas maka didapat hasil *TF-IDF* dari setiap dokumen adalah sebagai berikut:

$$D1 = 5.6816$$

$$D5 = 2.3012$$

$$D2 = 0.9120$$

$$D6 = 5.6816$$

$$D3 = 0.9120$$

$$D7 = 3.1463$$

$$D4 = 11.2293$$

BAB IV IMPLEMENTASI DAN PEMBAHASAN

4.1 Implementasi

4.1.1 Implementasi *Naive Bayes Classifier* Menggunakan Rapid Miner

Analisis sentimen yang dilakukan menggunakan total 3.003 komentar twitter berupa tanggapan terhadap masuknya varian baru dari Covid-19 yaitu Omicron. Sebelum dilakukan proses analisis sentimen, seluruh data tersebut diberikan Label Sentimen terlebih dahulu. Pemberian label sentimen ini dilakukan oleh salah satu **Tenaga Kesehatan** yaitu **Saudari Aminah Ghina, A.Md. Keb.** Berikut ini merupakan format pelabelan yang dilakukan.

Tabel 4. 1 Pemberian Label Sentimen

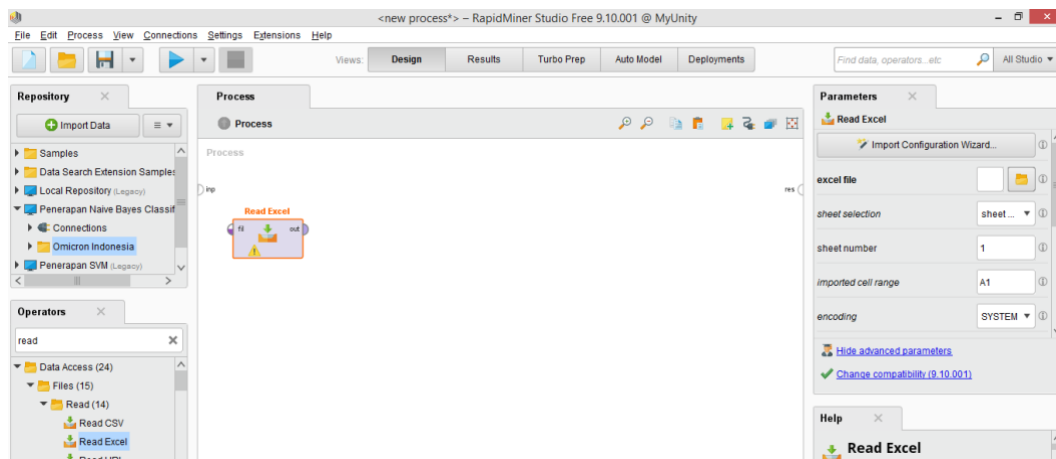
Teks Komentar Twitter	Label
RT @NarasiNewsroom: Varian baru COVID-19, Omicron BA.4 dan BA.5 telah terdeteksi di Indonesia dan saat ini sudah ditemukan sebanyak delapan.	Negatif
Omicron BA ?! Subvarian baru rasa keju susu.	Negatif
RT @CNNIndonesia: Mutasi varian Omicron dengan subvarian BA.4 dan BA.5 disebut sebagai penyebab angka kasus Covid-19 meningkat.	Negatif
Alert! Taiwan Laporkan Kasus Pertama Sub-varian Omicron Baru	Negatif
Mutasi varian Omicron dengan subvarian BA.4 dan BA.5 disebut sebagai penyebab angka kasus Covid-19 meningkat. Singapura sudah memprediksi gelombang baru dari subvarian Omicron akan terjadi pada Juli 2022.	Negatif
RT @cnbcindonesia: Varian Omicron Baru Masuk Ibu Kota, Jakarta Bersiap!	Negatif
Ancaman reinfeksi COVID-19 masih ada, seiring dengan munculnya varian baru Omicron BA.4 dan BA.5.	Negatif
tiba-tiba rame berita omicron varian baru.	Negatif
Waspada!! Ada 8 Orang Terpapar Varian Baru Covid-19 BA 4 dan BA 5 Fokus Sebaran varian baru virus omicron BA.4 dan BA.5 dan naiknya suspek Covid-19 dalam beberapa hari terakhir terdeteksi.	Negatif
RT @agung22reborn: @nuicemedia Sudah menyusut kah harta beliau sampai-sampai membuat statement terbaru mengenai covid yang sudah hampir kita lupakan	Negatif

Implementasi Algoritma *Naive Bayes Classifier* pada penelitian ini menggunakan tools Rapid Miner. Dalam proses penarikan data atau *crawling data* juga menggunakan tools Rapid Miner. Dimana pada proses penarikan data dilakukan dengan menggunakan kata kunci “Omicron”, “Omicron Indonesia”, “Varian Omicron”, “Covid-19 Omicron”. Untuk proses penarikan data dapat dilihat pada sub bab pembahasan sebelumnya yaitu pada sub bab **3.3 Proses Penarikan Data (*Crawling Data*)**.

Pada bab ini akan membahas mengenai bagaimana menggunakan tools Rapid Miner untuk melakukan Analisis Sentimen terhadap data komentar Twitter yang telah didapatkan. Berikut ini merupakan langkah-langkah penerapan Analisis Sentimen menggunakan *Naive Bayes Classifier* dengan tools Rapid Miner.

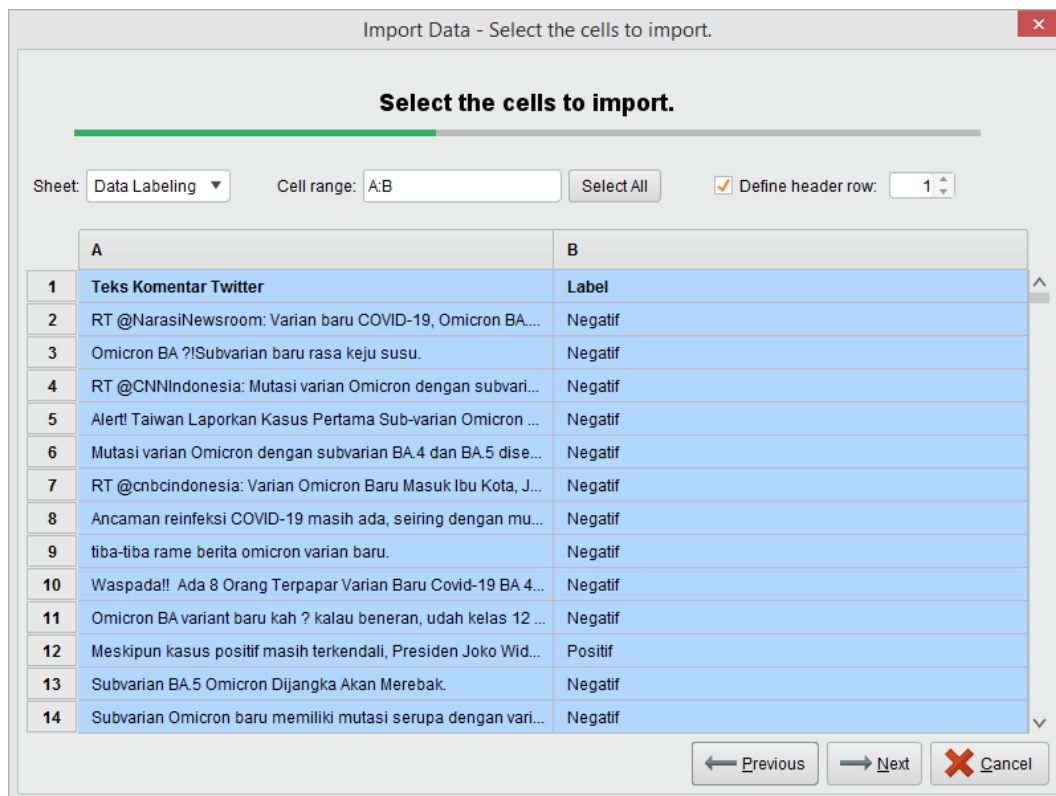
1) Memasukan Data

Terdapat banyak pilihan operator yang dapat digunakan untuk memasukan data kedalam lembar proses pada Rapid Miner diantaranya adalah *Read Excel*, *Read CSV*, *Read URL*, *Read SPSS*, dan masih banyak lagi pilihan lainnya. Salah satu kelebihan menggunakan tools Rapid Miner adalah dapat langsung membaca file dengan format *.xlsx atau *.xls sehingga tidak perlu mengubah file excel menjadi file csv. Pada penelitian ini data yang digunakan terdapat dalam sebuah file excel dengan format *.xlsx maka operator yang digunakan oleh penulis adalah operator *Read Excel*. Untuk memasukan data kedalam Rapid Miner dapat dilakukan dengan memilih operator *Read Excel* dan memasukannya ke lembar proses seperti pada gambar dibawah ini.



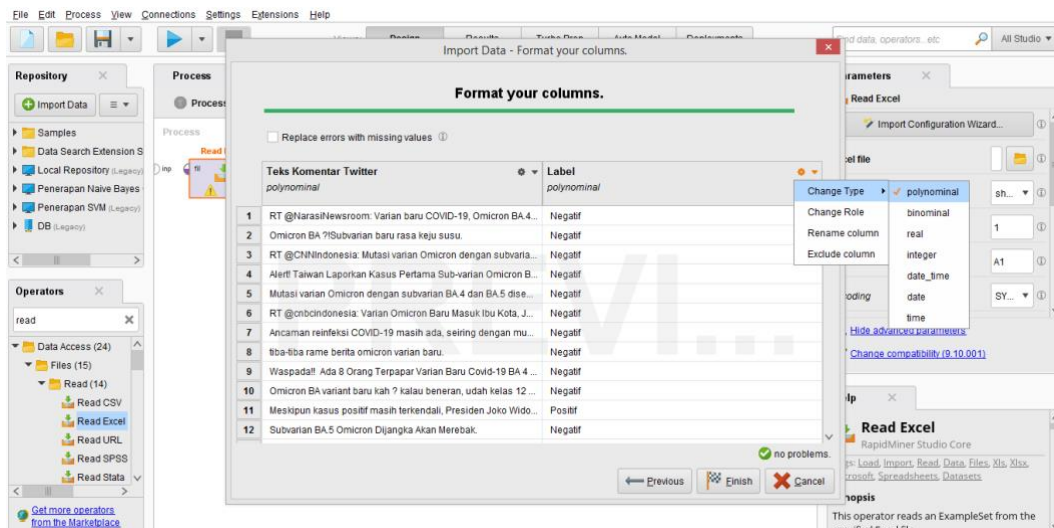
Gambar 4.1 Operator *Read Excel*

Selanjutnya pada tab parameter klik *Import Configuration Wizard* untuk memasukan data kedalam operator *Read Excel*. Lalu cari lokasi dimana tersimpannya data dan kemudian pilih data tersebut. Setelah memilih data maka akan diarahkan untuk melakukan pilihan *cell* untuk di *import*. Pada tahap ini penulis memilih *Sheet* Data Labeling dan *cell range* A:B karena kolom komentar Twitter dan Labeling Sentimen terdapat pada kolom tersebut. Berikut ini merupakan tampilan dari *Import Data*



Gambar 4. 2 Import Data

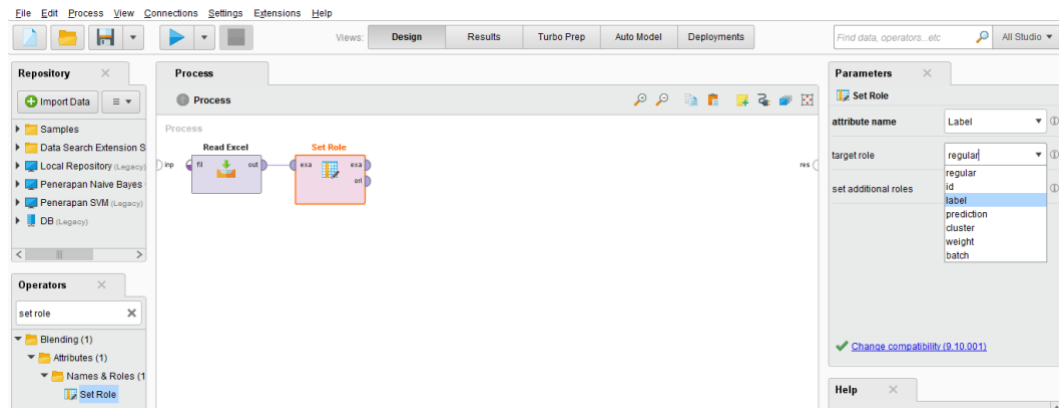
Setelah data di *Import*, maka akan muncul tampilan *Format Column* seperti pada Gambar 4.3. Didalam Rapid Miner terdapat beberapa pilihan tipe data yang dapat digunakan yaitu *polynomial*, *binominal*, *real*, *integer*, *date_time*, *date*, dan *time*. Tipe data yang akan digunakan oleh penulis pada Kolom Label adalah sebagai *Polynomial*. Penulis tidak menggunakan tipe data *binominal* karena tipe data ini digunakan untuk data yang hanya memiliki 2 pilihan/kelas, sedangkan data yang akan digunakan oleh penulis merupakan data yang memiliki 3 pilihan/kelas. karena Label Sentimen yang digunakan adalah 3 Sentimen yaitu Sentimen Positif, Sentimen Netral, dan Sentimen Negatif. Selanjutnya apabila sudah tidak terdapat kesalahan pada data atau *error* maka klik *Finish* maka data telah berhasil masuk ke dalam *operator Read Excel*. Untuk lebih jelasnya ditampilkan dalam gambar 4.3 dibawah ini.



Gambar 4.3 *Format Columns*

2) Membentuk *Set Role*

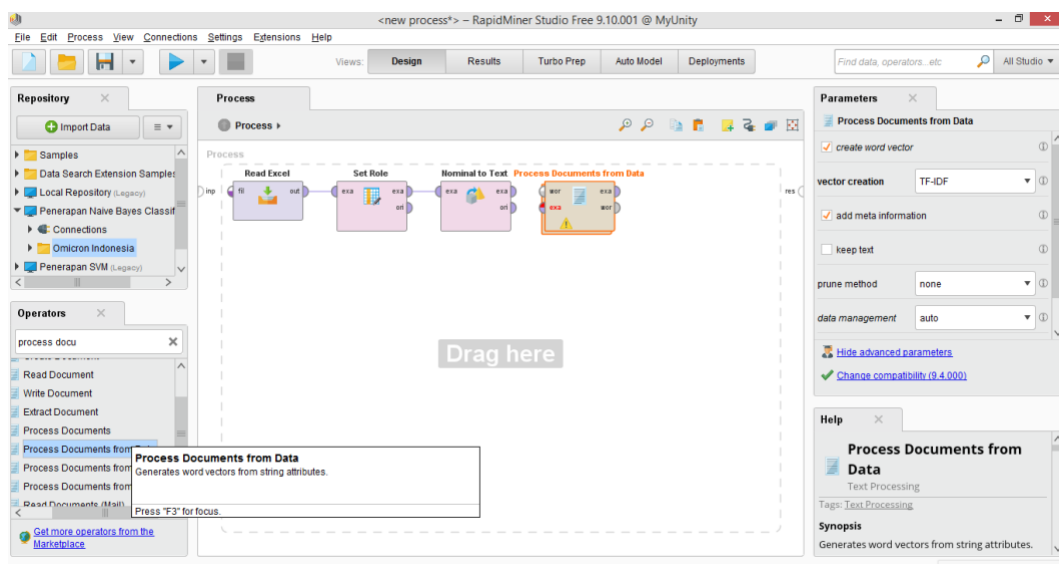
Operator Set Role digunakan untuk memberikan peran terhadap masing-masing atribut yang ada pada data. Dalam Operator ini diberikan beberapa pilihan peran untuk atribut yang digunakan yaitu *regular*, *id*, *label*, *prediction*, *cluster*, *weight*, dan *batch*. Data yang digunakan oleh penulis terdapat 2 atribut yaitu Teks Komentar Twitter dan Label. Dimana untuk melakukan klasifikasi menggunakan Algoritma *Naive Bayes* diperlukan atribut yang berperan sebagai *label*. Maka pada penelitian ini atribut yang berperan sebagai *label* adalah atribut/kolom Label. Pada tab parameter pilih attribute name menjadi Kolom “Label” dan pada *target role* pilih sebagai *label*. B



Gambar 4. 4 Menentukan Atribut Label Sentimen

3) Proses TF-IDF

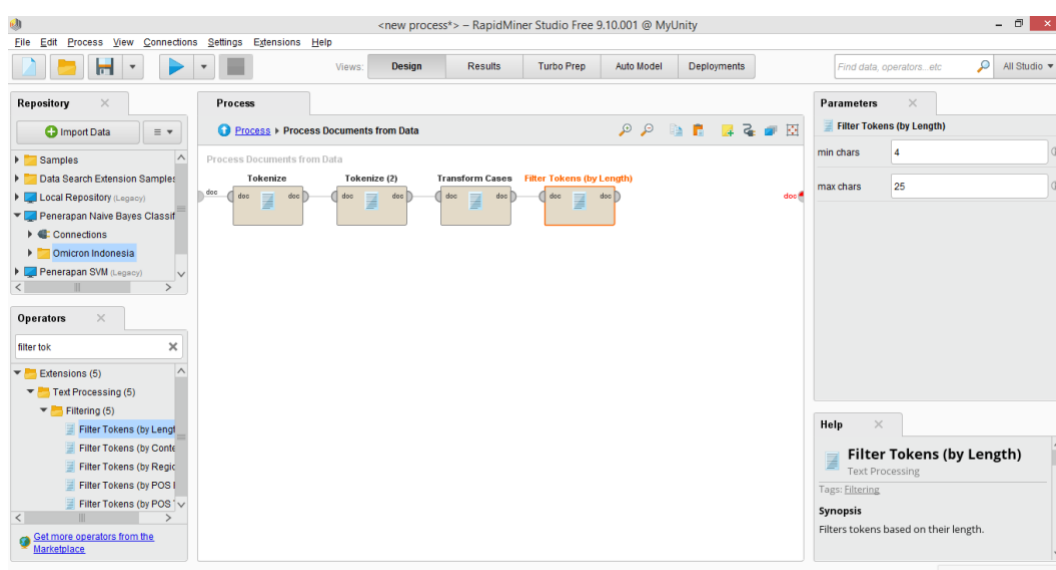
Untuk menerapkan proses TF-IDF digunakan *operator Process Documents from Data*. Sebelum menerapkan operator tersebut pastikan telah diterapkan operator *Nominal to Text* untuk membuat seluruh data terbaca sebagai teks karena analisis yang dilakukan adalah Analisis Sentimen yang merupakan menganalisis teks. Pada tab Parameter gunakan *vector creation* sebagai TF-IDF.



Gambar 4. 5 Menerapkan Algoritma TF-IDF

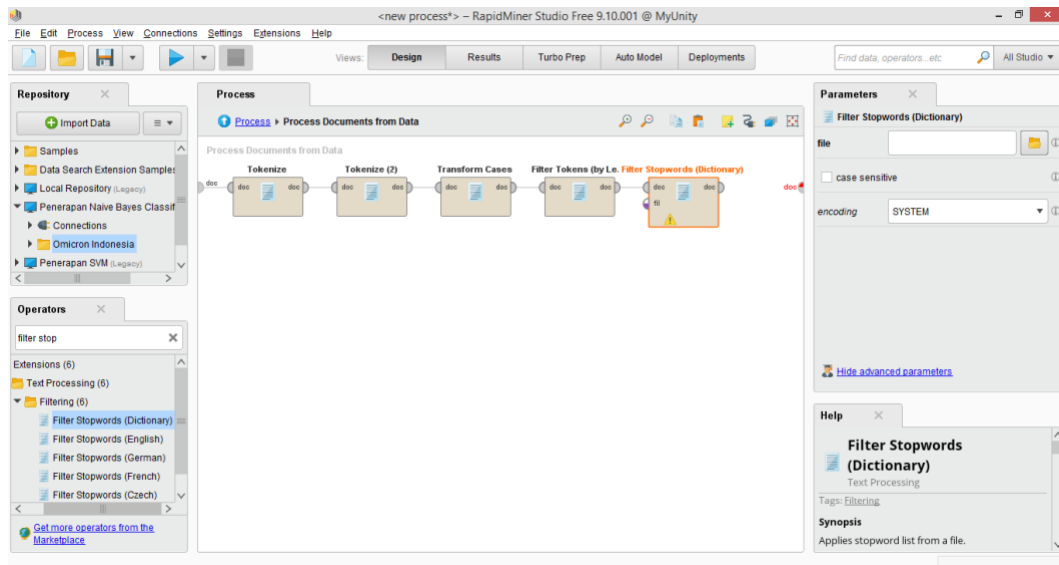
Selanjutnya klik dua kali pada operator *Process Documents From Data* untuk menerapkan tahap *pre-processing* menggunakan operator *Tokenize* untuk

menghilangkan karakter-karakter tulisan seperti tanda baca dan hashtag. *Tokenize* yang kedua digunakan untuk menghilangkan karakter-karakter yang tidak terlihat seperti spasi atau enter berlebih. *Transform Case* digunakan untuk mengubah seluruh kata menjadi huruf kecil atau *lower case*. *Filter Token* digunakan sebagai pembacaan kata dimulai dengan kata yang memiliki minimal 4 huruf dan maksimal 25 huruf. Maka penerapan *operator* untuk tahap *pre-processing* dapat dilihat seperti gambar dibawah ini.



Gambar 4. 6 Operator Pre-Processing

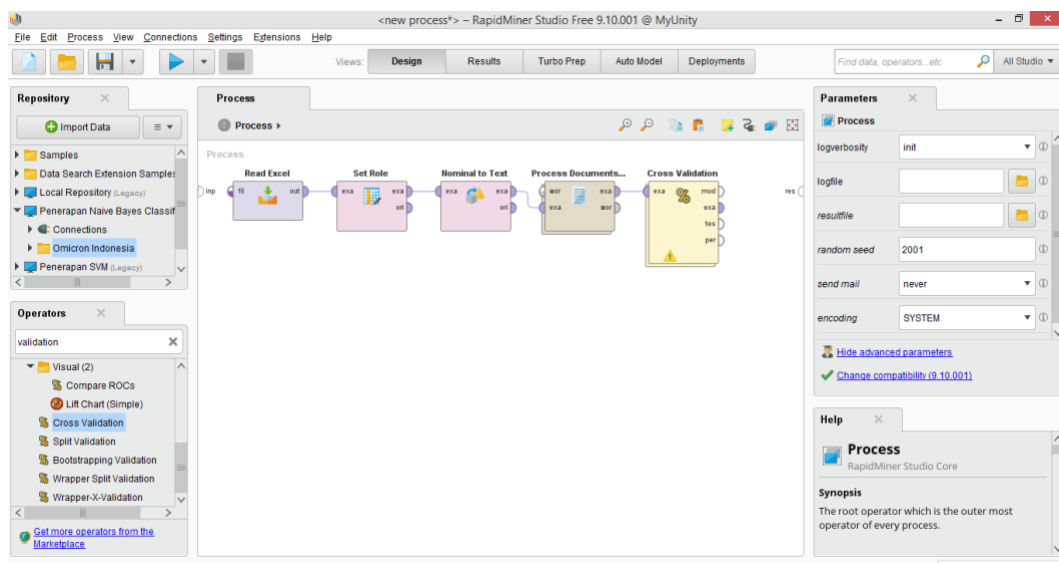
Pada tahap *pre-processing* yang terakhir adalah gunakan operator *stopwords by dictionary* untuk mengenali kata-kata penghubung atau kata-kata tanpa makna dalam bahasa Indonesia. pada tab parameter silahkan pilih file *stopword* dalam bahasa Indonesia yang telah disimpan dalam format *.txt.



Gambar 4. 7 Menerapkan *Filter Stopword*

4) Menerapkan Algoritma *Naive Bayes Classifier*

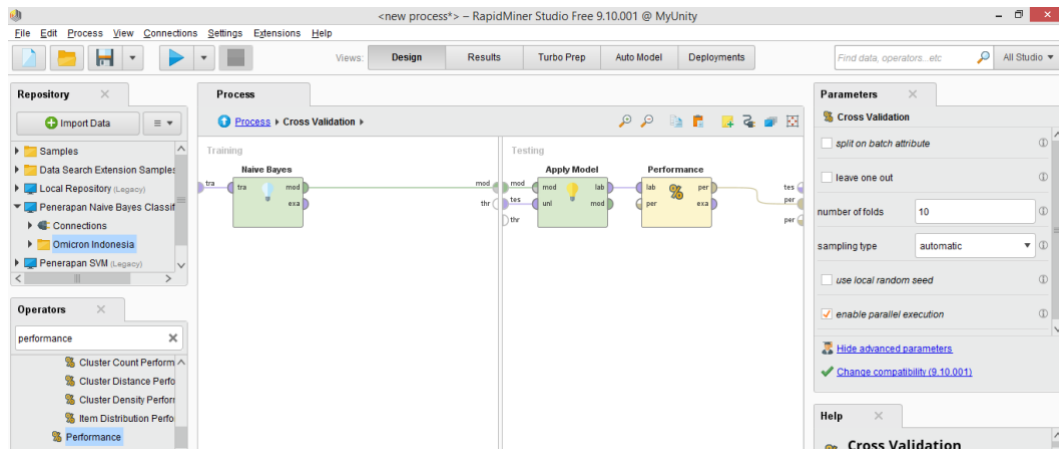
Untuk menerapkan algoritma *Naive Bayes Classifier* langsung gunakan *operator* pengujian yaitu *Cross Validation* seperti pada gambar dibawah ini.



Gambar 4. 8 Menerapkan *Cross Validation*

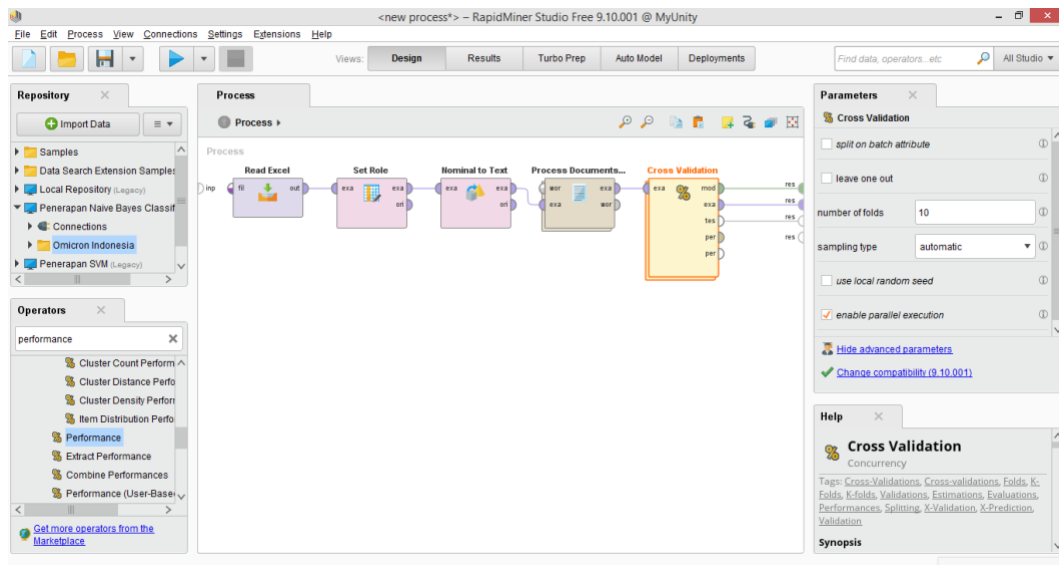
Selanjutnya klik dua kali pada *operator* tersebut untuk menerapkan Algoritma *Naive Bayes Classifier* seperti pada gambar dibawah ini. Pada tab *Training*

masukan *operator Naive Bayes* dan pada tab *Testing* masukan *operator Apply Model dan Performance*.



Gambar 4. 9 Menerapkan Algoritma *Naive Bayes*

Setelah menerapkan algoritma maka hubungkan seluruh *connector* seperti pada Gambar 4.10 dan lakukan *running process*.



Gambar 4. 10 Tampilan *Process*

4.1.2 Hasil Implementasi Menggunakan Rapid Miner

Dalam penelitian ini pengujian data dilakukan dengan menggunakan metode *Cross Validation*, dimana dengan menggunakan metode ini seluruh data latih (*data training*) yang digunakan akan digunakan juga sebagai data uji (*data*

testing). Dalam metode pengujian seluruh data berperan sebagai data latih dan data data uji. Penggunaan *Cross Validation* dinilai lebih memberikan hasil pengujian yang paling akurat dan efektif.

Setelah melakukan *modeling* pada Rapid Miner dan melakukan *running process* maka akan muncul dari hasil pengujian yaitu *confusion matrix* nya seperti Gambar 4.11.

Screenshot of Rapid Miner Studio showing a confusion matrix for a Naive Bayes model. The interface includes a menu bar, toolbars, and a sidebar with 'Performance', 'Description', and 'Annotations' sections. The main area displays a table with columns for true Negatif, true Positif, true Netral, and class precision, and rows for predicted Negatif, Positif, and Netral, along with a class recall row.

	true Negatif	true Positif	true Netral	class precision
pred. Negatif	754	219	187	65.00%
pred. Positif	377	463	122	48.13%
pred. Netral	439	151	291	33.03%
class recall	48.03%	55.58%	48.50%	

Gambar 4. 11 Hasil Pengujian Algoritma

Berdasarkan *confusion matrix* diatas dapat terlihat bahwa nilai akurasi penerapan dan pengujian data adalah 50,22%. Untuk nilai *precision* dan *recall* yang ditunjukkan adalah per masing-masing kelas klasifikasi. Maka untuk mengetahui nilai *precision* dan *recall* keseluruhan dilakukan perhitungan menggunakan persamaan-persamaan sebagai berikut:

1) *Precision*

Untuk mendapatkan nilai *precision* keseluruhan dapat dilakukan dengan menjumlahkan seluruh nilai *precision* pada masing-masing kelas lalu dibagi jumlah kelas klasifikasi dimana jumlah kelas klasifikasi pada penelitian ini ada 3. Berikut ini perhitungan nilai *precision*:

Diketahui	: <i>Precision</i> Negatif	: 65,00%
	<i>Precision</i> Positif	: 48,13%
	<i>Precision</i> Netral	: 33,03%
	Jumlah Kelas	: 3

$$\begin{aligned} \text{Maka, Nilai } Precision &= \frac{Precision\ Negatif + Precision\ Positif + Precision\ Netral}{Jumlah\ Kelas} \\ &= \frac{65,00\% + 48,13\% + 33,03\%}{3} = 48,72\% \end{aligned}$$

2) *Recall*

Untuk mendapatkan nilai *recall* keseluruhan dapat dilakukan dengan menjumlahkan seluruh nilai *recall* pada masing-masing kelas lalu dibagi jumlah kelas klasifikasi dimana jumlah kelas klasifikasi pada penelitian ini ada 3. Berikut ini perhitungan nilai *recall*:

Diketahui	: <i>Recall</i> Negatif	: 48,03%
	<i>Recall</i> Positif	: 55,58%
	<i>Recall</i> Netral	: 48,50%
	Jumlah Kelas	: 3

$$\begin{aligned} \text{Maka, Nilai } Recall &= \frac{Recall\ Negatif + Recall\ Positif + Recall\ Netral}{Jumlah\ Kelas} \\ &= \frac{48,03\% + 55,58\% + 48,50\%}{3} = 50,70\% \end{aligned}$$

Dari pembahasan yang telah dilakukan maka hasil yang didapatkan penulis pada penerapan Algoritma *Naive Bayes Classifier* untuk melakukan analisis Sentimen Komentar Twitter adalah sebagai berikut:

- a. Dari pengujian 3.003 data yang berhasil diuji sebagai Klasifikasi Sentimen Negatif berjumlah 1.570 data, data yang berhasil diuji sebagai Klasifikasi

Sentimen Positif berjumlah 833 data, dan data yang berhasil diuji sebagai Klasifikasi Sentimen Netral berjumlah 600 data.

- b. *Accuracy*, atau tingkat keakuratan algoritma *Naive Bayes Classifier* dalam melakukan analisis sentimen menggunakan data tanggapan Varian Omicron di Indonesia berhasil mengklasifikasikan kedalam Sentimen Positif, Sentimen Negatif, dan Sentimen Netral dengan nilai *accuracy* mencapai 50,22%.
- c. Nilai ketepatan antara informasi yang dihasilkan oleh algoritma *Naive Bayes Classifier* dengan hasil yang diharapkan ditunjukkan dengan hasil dari nilai *Precision* yaitu mencapai 48,72% dengan nilai *precision* pada masing-masing Sentimen yaitu Sentimen Negatif 65%, Sentimen Positif 48,13% dan Sentimen Netral 33.03%.
- d. Sedangkan tingkat keberhasilan algoritma untuk menemukan informasi dari analisis sentimen ditunjukkan oleh nilai *Recall* yaitu mencapai 50,70% dengan nilai *recall* pada masing-masing sentimen yaitu Sentimen Negatif 48,03%, Sentimen Positif 55,58% dan Sentimen Netral 48,50%.

Berdasarkan hasil tersebut didapat bahwa lebih banyak orang yang memberikan sentimen negatif terhadap masuknya varian Omicron ke Indonesia. Hasil analisis sentimen yang didapat menunjukkan bahwa kemampuan Algoritma *Naive Bayes Classifier* untuk melakukan klasifikasi terhadap 3.003 data yang ada memiliki akurasi mencapai 50,22%. Dengan tingkat keberhasilan untuk menemukan informasi yang dibutuhkan mencapai nilai *recall* sebesar 50,70% dan memiliki ketepatan informasi mencapai nilai *precision* sebesar 48,72%.

4.1.3 Implementasi *Naive Bayes Classifier* Menggunakan Phyton

Implementasi algoritma *Naive Bayes Classifier* menggunakan bahasa Phyton dengan *tools Google Colab* . Berikut ini merupakan *source code* dari proses *crawling* hingga penerapan Algoritma.

1) *Crawling Data & Labeling*

Crawling data adalah proses penarikan data komentar twitter menggunakan Phyton ke platform Twitter. Sebelum melakukan penarikan data komentar hal yang harus dilakukan yaitu memberikan/mengatur variable jumlah untuk data komentar twitter yang akan kita akan ambil dengan isi kata kunci sesuai dengan yang akan dicari. Dalam penarikan data komentar twitter terdapat label data yang diberikan yaitu label Positif, Neutral dan Negatif. Berikut ini merupakan source code yang digunakan dalam melakukan penarikan data sesuai dengan jumlah yang telah ditentukan.

```

maxTweets = 5000
tweet_list2 = []
for i, tweet in enumerate (sntwitter.TwitterSearchScrapper('Omicron di
Indonesia since: 2022-01-01 until:2022-07-20').get_items()):
    If i>maxTweets :
        Break
    Tweets_list2.append ([tweet.date, tweet.id,
tweet.content,tweet.username])
tweets_df2 = pd.DataFrame (tweets_list2, Columns = ['Datetime', '
Tweet
Id', ' Text', 'Username'])
Tweets_df2.to_csv("Omicron_Crawling.csv",
index = False)

[] data = pd.read_csv('Omicron_Crawling.csv')

```

```

> Install dan Import Library yang dibutuhkan
[ ] ↳ 2 sel tersembunyi

Crawling data dari Twitter

#Mengatur variable jumlah untuk data tweet
maxTweets = 5000

#Membuat list data twitter
tweets_list2 = []
#Menggunakan TwitterSearchScrapecraper untuk scrape data dan menyimpan ke list
for i,tweet in enumerate(sntwitter.TwitterSearchScrapecraper('Omicron di indonesia since:2022-01-01 until:2022-07-20').get_items()):
    if i>maxTweets:
        break
    tweets_list2.append([tweet.date, tweet.id, tweet.content, tweet.username])

#Membuat Dataframe dari list Twitter diatas
tweets_df2 = pd.DataFrame(tweets_list2, columns=['Datetime', 'Tweet Id', 'Text', 'Username'])
tweets_df2.to_csv("Omicron_Crawling.csv", index = False)

data = pd.read_csv('Omicron_Crawling.csv')

```

Gambar 4. 12 Source Code Crawling

Berikut ini merupakan hasil penarikan data dan pemberian label menggunakan Phyton pada *google colab* yaitu Label Positif, Label Negatif, Label Netral.

Dataset.head (5)

```
[] scores = [analyser.polarity_scores (x) for x in dataset ['Tweet']]
```

```
[] dataset['Score'] = [x['compound'] for x in scores]
```

Table 4.1 Hasil Crawling dan Labeling

	Tweet	Klasifikasi
0	Entering the Subvariant omicron b Indonesia t	Positif
1	a new variant of covid omicron b Indonesia app	Positif
2	entered subvariant omicron b Indonesia the val	Positif
3	member of commission ix of the dpr pdi faction	Negatif
4	Prediction of the distribution pattern of the	Netral

2) Import Data

Proses ini dilakukan proses memasukkan data yang sudah di crawling dan di labeling dengan menggunakan penyimpanan menggunakan format csv. Berikut ini *source code* yang digunakan.

```

[] #read_Data
Import pandas as pd
def load_data():
    data= pd.read_csv('Omicron_Crawling.csv')
    Return data
[] tweet_df = load_data()
[] tweet_df = pd.DataFrame (tweet_df [['Username','Text']])
Tweet_df.head(3)

```

Table 4.2 Import Data ke Phyton

	Username	Text
0	wow_keren	Masuknya Subvarian Omicron BA 2 75 di Indonesia
1	newsmerahputih	Munculnya subvarian terbaru dari COVID-19 Omic
2	Beritacovid	#Masuknya Subvarian Omicron BA 2 75 di Indones

3) *Pre-Processing*

Data yang telah di impot selanjutnya akan dilakukan tahap *Pre-Processing* yaitu *Case Folding*, *Tokenizing*, *Filtering*, *Stemming*. Berikut ini *source code* dan hasil dari prosesnya.

Case Folding

Proses ini digunakan untuk mengubah seluruh karakter huruf dari bentuk kapital menjadi huruf kecil semua. Berikut ini merupakan *source code* dan hasil dari proses *Case Folding*.

```

Def remove_punct(text):
    Text = “.join([char for char in text if char not in string.punctuation])
    Return text
Tweet_df ['clean'] = tweet_df["Tweet_Clean"].apply(lambda
x:remove_punct(x))
tweet_df.head(3).

```

Table 4.3 Hasil *Pre-Processing*

	Username	Text	clean_tweet	remove_link
0	wow_keren	Masuknya Subvarian Omicron BA 2 75 di Indonesia	Masuknya Subvarian Omicron BA 2 75 di Indonesia	Masuknya Subvarian Omicron BA 2 75 di Indonesia
1	newsmerahputih	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic
2	beritacovid	#Masuknya Subvarian Omicron BA 2 75 di Indones	#Masuknya Subvarian Omicron BA 2 75 di Indones	#Masuknya Subvarian Omicron BA 2 75 di Indones
3	tribunnewswiki	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad
4	ema_axa	Pola Penyebaran SubvarianOmicron BA 2 75 di I	Pola Penyebaran SubvarianOmicron BA 2 75 di I	Pola Penyebaran Subvarian Omicron BA 2 75 di I

Tokenizing

Proses *tokenizing* atau bisa disebut *parsing* merupakan tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Berikut ini *source code* dan hasil *Tokenizing*.

```
[ ] #hasil tokenize
    Tweet_df.head(5)
```

Table 4.4 Hasil *Tokenizing*

	Username	Text	clean_tweet	remove_link	remove_hastag	Tweet_Clean
0	wow_keren	Masuknya Subvarian OmicronBA 275 di Indonesia	Masuknya Subvarian Omicron BA 275 di Indonesia	Masuknya Subvarian Omicron BA 275 di Indonesia	Masuknya Subvarian Omicron BA 275 di Indonesia	[Masuk, subvarian, omicron, ba indonesia, nil
1	newsmerahputih	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic	[muncul, subvarian, baru covid, omicron, ba
2	Beritacovid	#Masuknya Subvarian OmicronBA 275 di Indones	#Masuknya Subvarian Omicron BA 275 di Indones	#Masuknya Subvarian Omicron BA 275 di Indones	Masuknya Subvarian Omicron BA 275 di Indonesia	[Masuk, subvarian, omicron, ba indonesia, nil
3	Tribunnews wiki	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad	[anggota, komisi, ix, dpr, fraksi, pdi, p, rah
4	ema_axa	Pola Penyebaran Subvarian OmicronBA 275 di l	Pola Penyebaran Subvarian Omicron BA 275 di l	Pola Penyebaran Subvarian Omicron BA 275 di l	Pola Penyebaran Subvarian Omicron BA 275 di Indo	[pola, sebar, subvarian, omicron, ba, indonesi

Filtering

Tahap *filtering* dilakukan untuk menghapus kolom, menyisakan kolom tweet dengan mengambil kata penting yang terdapat pada komentar. Digunakan algoritma stoplist (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata yang penting). Tahap *filtering* digunakan untuk menyimpan hasil *processing* dan *filtering* ke dalam format csv.

```
Tweet_df.drop(tweet_df.columns[[0,1,2,3,4,5]], axis = 1, inplace = True)
Tweet_df.to_csv('HasilProcessingOmicron.csv', encoding='utf8', index=False)
tweet_df.head(3)
```

Table 4.5 Hasil *Filtering*

	Clean
0	masuk subvarian omicron ba indonesia nilai tan
1	muncul subc\varian baru covid omicron ba indones
2	masuk subvarian omicron ba indonesia nilai tan

Stemming

Tahap *stemming* adalah tahap mencari akar atau dasar kata dari tiap kata hasil *filtering*. Pada tahap ini seluruh kata yang memiliki imbuhan akan diubah menjadi kata dasar.

4) *Weighting*, Penerapan Algoritma dan Pengujian

Pada tahap ini dilakukan pembobotan menggunakan Algoritma *TF-IDF* dan menerapkan Algoritma *Naive Bayes*.

```
[ ] !pip install -Uscikit-learn
[ ] import pandas as pd
    Import numpy as np
    df = pd.read_csv('HasilProcessingOmicrons.csv', sep=' ', ')
    df.head(5)
[ ] tweet_data = pd.read_csv("HasilProcessingOmicrons.csv",
    usecols=['Tweet_Clean'])
    tweet_data.columns = ["Tweet_Clean"]
    tfidf_mat = normalized_counts.multiply(IDF_vector).toarray()
    dfbtf = pd.DataFrame (data=tfidf_mat, columns=[a])
    dfbtf
[ ] dfbtf.to_csv"Hasil_vectorspaceMatrix.csv")
```

Table 4.6 Source Code TF-IDF dan Pengujian

	Username	Text	clean_tweet	remove_link	remove_hashtag	Tweet_Clean
0	wow_keren	Masuknya Subvarian Omicron BA 2 75 diIndonesia	Masuknya Subvarian Omicron BA 2 75 di Indonesia	Masuknya Subvarian Omicron BA 2 75 di Indonesia	Masuknya Subvarian OmicronBA 2 75 di Indonesia D	["Masuk","subvarian","omicron","ba","indon
1	newsm erahputih	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic	Munculnya subvarian terbaru dari COVID-19 Omic	["muncul","subvarian","baru","covid","omic"
2	beritacovid	#Masuknya Subvarian OmicronBA 2 75 di Indones	#Masuknya Subvarian Omicron BA 2 75 di Indones	#Masuknya Subvarian Omicron BA 2 75 di Indones	Masuknya Subvarian OmicronBA 2 75 di Indonesia D	["Masuk","subvarian","omicron","ba","indon"
3	tribunn ewswiki	Anggota Komisi IX DPR dari Fraksi PDI-P Rahmad	Anggota KomisiIX DPR dari Fraksi PDI-P Rahmad	Anggota KomisiIX DPR dari Fraksi PDI-P Rahmad	Anggota Komisi IXDPR dari Fraksi PDI-P Rahmad	["anggota","komisi","ix","dpr","fraksi","..
4	ema_axa	Pola Penyebaran Subvarian OmicronBA 2 75 di l	Pola Penyebaran Subvarian Omicron BA 2 75 di l	Pola Penyebaran Subvarian Omicron BA 2 75 di l	Pola Penyebaran Subvarian OmicronBA 2 75 di Indo	["pola","sebar","subvarian","omicron","ba"...

	akibat	alami	amp	anak	angka	antisipasi	ba	ba	baba	ba	...	vaksin	vaksinasi	varian	varian	virus	warga	waspada	ya	yg		
0	0.0	0.0	0.0	0.0	0.0	0.0	0,246119	0.0	0.0	0.0	...	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000	
2	0.0	0.0	0.0	0.0	0.0	0.0	0,169207	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	0.000
3	0.0	0.0	0.0	0.0	0.0	0.0	0,208254	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	0.000
4	0.0	0.0	0.0	0.0	0.0	0.0	0,338413	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	0.000
5	0.0	0.0	0.0	0.0	0.0	0.0	0,270731	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	0.000
....
4103	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	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	0.000
4104	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	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	0.000
4105	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	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	0.000
4106	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	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	0.375%
4107	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	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	0.000

4108 rows x 100 columns

Gambar 4. 13 Source Code TF-IDF dan Pengujian

Translate Data Omicorn

Proses ini merupakan tahap penerjemahan data hasil *processing* ke bahasa Inggris. Selanjutnya, untuk proses *Labeling* digunakan *lexicon based*. Dimana *Lexicon Based* merupakan proses mengidentifikasi suatu kalimat mengandung opini atau tidak. Berikut ini merupakan *source code* yang digunakan.

```

- Translate Data Omicorn

#Install googletrans menggunakan pip
!pip3 install googletrans==3.1.0a0

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting googletrans==3.1.0a0
  Downloading googletrans-3.1.0a0.tar.gz (19 kB)
Collecting httpx==0.13.3
  Downloading httpx-0.13.3-py3-none-any.whl (55 kB)
    |#####| 55 kB 3.2 MB/s
Collecting rfc3986<2,>=1.3
  Downloading rfc3986-1.5.0-py2.py3-none-any.whl (31 kB)
Requirement already satisfied: idna==2.* in /usr/local/lib/python3.7/dist-packages (from httpx==0.13.3->googletrans==3.1.0a0) (2.10)
Collecting sniffio
  Downloading sniffio-1.3.0-py3-none-any.whl (10 kB)
Requirement already satisfied: certifi in /usr/local/lib/python3.7/dist-packages (from httpx==0.13.3->googletrans==3.1.0a0) (2022.9.24)
Requirement already satisfied: chardet==3.* in /usr/local/lib/python3.7/dist-packages (from httpx==0.13.3->googletrans==3.1.0a0) (3.0.4)
Collecting hstspreload
  Downloading hstspreload-2022.10.1-py3-none-any.whl (1.4 MB)
    |#####| 1.4 MB 28.4 MB/s
Collecting httpcore==0.9.*
  Downloading httpcore-0.9.1-py3-none-any.whl (42 kB)
    |#####| 42 kB 1.3 MB/s
Collecting h11<0.10,>=0.8
  Downloading h11-0.9.0-py2.py3-none-any.whl (53 kB)
    |#####| 53 kB 2.3 MB/s
Collecting h2==3.*
  Downloading h2-3.2.0-py2.py3-none-any.whl (65 kB)
    |#####| 65 kB 3.8 MB/s
Collecting hpack4,>=3.0

```

Gambar 4. 20 Translate Data Omicorn

Lexicon Based Naïve Buys

Pada penelitian ini, dilakukan pengujian ekstraksi fitur menggunakan metode *lexicon-based*. Tahapan *Lexicon Based* merupakan salah *satu* metode ekstraksi fitur yang dapat meningkatkan performansi sistem karena mampu mengekstrak kalimat opini dengan nilai presisi yang cukup tinggi. Pada tahap ini, setiap tweet yang ada pada dataset akan dianalisa satu per satu. Tweet yang mengandung kata yang terdapat pada kamus kata opini akan dilabeli sebagai tweet opini. Berikut ini merupakan *source code* metode *lexicon based*:

```
#Lexicon Based
```

```
!pip install VaderSentiment
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
```

```
Collecting VaderSentiment
```


4.1.4 Hasil Implementasi Menggunakan Python

Pada gambar dibawah ini merupakan hasil dari penerapan algoritma dan pengujian data. Dimana nilai akurasi menggunakan Python mencapai 0.81 atau 81%, *precision* 0.90 atau 90%, dan *recall* 0.77 atau 77%.

```
dataset.tail(3)
from sklearn import metrics
print(metrics.classification_report(dataset[""Klasifikasi"],dataset['Sentiments]))
```

Table 4.7 Confusion Matrix

	Tweet	Klasifikasi	Score	Sentiment
4105	Luhut Terawang, the peak of Omicron Indonesia	Netral	0,0000	Netral
4106	detection of omicron in one area, based on dat	Positif	0,4404	Positif
4107	Yes, it's already Ramadan, the anti-islam Regi	Negatif	0,0956	Negatif
	Precision	Recall	f1- score	support
Negatif	0,75	0,76	0,76	870
Netral	0,73	0,89	0,8	1226
Positif	0,9	0,77	0,83	2012
Accuracy			0,81	4108
macro avg	0,79	0,81	0,80	4108
weighted avg	0,82	0,81	0,81	4108

4.2 Pembahasan

Berdasarkan pembahasan diatas maka didapatkan tabel perbandingan dari hasil penerapan Algoritma *Naive Bayes Classifier* menggunakan Rapid Miner dan menggunakan Phyton.

Tabel 4. 2 Hasil Perbandingan

Hasil Pengujian	Rapid Miner	Phyton
Akurasi	50,22%	81%
<i>Precision</i>	48,72%	90%
<i>Recall</i>	50,70%	77%

Berdasarkan hasil diatas dapat dikatakan bahwa hasil penerapan dan pengujian menggunakan Phyton lebih besar dibanding menggunakan *tools* Rapid Miner.

BAB V PENUTUP

5.1 Kesimpulan

Pembahasan yang telah dipaparkan pada bab-bab sebelumnya telah dapat ditarik kesimpulan penelitian. Kesimpulan yang dapat ditarik dari penelitian ini antara lain adalah sebagai berikut:

1. Data opini Twitter yang dapat ditarik menggunakan API Twitter melalui *tools* Rapid Miner berjumlah 3.003 data.
2. Dari pengujian 3.003 data yang berhasil diuji sebagai Klasifikasi Sentimen Negatif berjumlah 1.570 data, data yang berhasil diuji sebagai Klasifikasi Sentimen Positif berjumlah 833 data, dan data yang berhasil diuji sebagai Klasifikasi Sentimen Netral berjumlah 600 data. Situasi ini menunjukkan bahwa lebih banyak respon Negatif yang diberikan oleh para pengguna Twitter terhadap masuknya varian Covid-19 ke Indonesia.
3. *Accuracy*, atau tingkat keakuratan algoritma *Naive Bayes Classifier* dalam melakukan analisis sentimen menggunakan data tanggapan Varian Omicron di Indonesia berhasil mengklasifikasikan kedalam Sentimen Positif, Sentimen Negatif, dan Sentimen Netral dengan nilai *accuracy* mencapai 50,22%.
4. Nilai ketepatan antara informasi yang dihasilkan oleh algoritma *Naive Bayes Classifier* dengan hasil yang diharapkan ditunjukkan dengan hasil dari nilai *Precision* yaitu mencapai 48,72%.

5. Sedangkan tingkat keberhasilan algoritma untuk menemukan informasi dari analisis sentimen ditunjukkan oleh nilai *Recall* yaitu mencapai 50,70%.
6. Hasil penerapan dan pengujian menggunakan Phyton lebih besar dibanding menggunakan *tools* Rapid Miner dimana nilai akurasi menggunakan Phyton mencapai 0.81 atau 81%, *precision* 0.90 atau 90%, dan *recall* 0.77 atau 77%.

5.2 Saran

Saran yang dapat penulis berikan ditujukan kepada penulis selanjutnya adalah untuk menggunakan Algoritma lain dalam melakukan Analisis Sentimen terhadap Komentar Twitter, atau dapat melakukan perbandingan menggunakan Algoritma Klasifikasi yang lain.

DAFTAR PUSTAKA

- A. Go, R. B., & L. Huang. (2009). *Twitter Sentiment Clasification Using Distant Supervision*.
- Abdillah, G., Putra, F. A., & Renaldi, F. (2016). Penerapan data Mining Pemakaian Air Pelanggan untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru di PDAM Tirta Raharja Menggunakan Algoritma K-Means. *Seminar Nasional Teknologi Informasi Dan Komunikasi (SENTIKA)*, pp 498-506.
- Aditya, B. . (2015). Penggunaan Web Crawler untuk Menghimpun Tweet dengan Metode Pre-Processing Text Mining. *Jurnal Infotel*, 7 No. 2.
- Ahuja, S., & Dubey, G. (2017). Clustering and Sentimen Analysis on twitter Data. *International Conference on Telecommunication and Networks (TEL-NET)*.
- Akbar, M. G. T., & Srisusilowati, D. B. (2021). Analisa Sentimen Efektifitas Vaksin terhadap Varian COVID 19 Omicron Berbasis Leksikon. *Journal of Information and Information Security*, Vol. 2(No. 2).
- Alita, D., Sari, I., Isnain, A. R., & Styawati. (2021). Penerapan Naive Bayes Classifier untuk Pendukung Keputusan Penerima Beasiswa. *Jurnal Data Mining Sistem Informasi*, Vol. 2(No. 1), Hal. 17-23.
- Arhami, S. M., & Nasir, S. M. (2020). *Data Mining : Algoritma dan Implementasi*. Penerbit Andi.
- Bustami. (2013). Penerapan Algoritma Naive Bayes untuk Mengklasifikasikan Data Nasabah Asuransi. *TECHSI: Jurnal Penelitian Teknik Informatika*, Voll. 3 No, Hal. 127-146.
- C, D. A. et al. (2013). *Belajar Data Mining dengan Rapid Miner*.
- Chairunnisa, N. (2022). *Kasus Omicron Masih Didominasi WNI dari Luar Negeri, Terbanyak dari Turki*. Tempo.Co. <https://tinyurl.com/256zw9hw>
- Darwis, D., Pratiwi, E. S., & Pasaribu, F. O. (2020). Penerapan Algoritma SVM untuk Analisis Sentimen pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Jurnal Ilmiah Edutic*, Vol. 7(No. 1), Hal. 1-11.
- Darwis, D., Siskawati, N., & Abidin, Z. (2020). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter BMKG Nasional. *Jurnal TEKNO KOMPAK*, Vol. 15(No. 1), Hal. 131-145.
- Davido, D., & Tsur, O. (2010). *Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon*.

- Feldman, R., & Sanger, J. (2007). *Advanced Approache in Analyzing Unstructured Data, The Text Mining Handbook*.
- Hasan, B. (2017). *Remaja Berkualitas Problematika Remaja dan Solusinya*. Pustaka Pelajar Offset`.
- I, R., Pranomo, S. ., & Dahlan, E. A. (2012). *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Pubik pada Perguruan Tinggi*. *Electrics, Electronics, Communicatioons, Controls, Information, System (EECCIS)*.
- Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM. *Jurnal Data Mining Sistem Informasi, Vol. 2*(No. 1), Hal. 31-37.
- Khomarudin, A. N. (2016). *Teknik Data Mining: Algoritma K-Means Clustering*. *IlmuKomputer.Com*, 1-12.
- L. Elvitaria, & M. Havenda. (2017). Memprediksi Tingkat Peminat Ekstrakurikuler Pada Siswa SMK Analisis Kesehatan Abdurrab Menggunakan Algoritmaa C4.5 (`Studi Kasus: SMK Analisis Kesehatan Abdurrab). *RABIT (Jurnal Teknologi Dan Sistem Informasi Univrab), VOL. 2 No.*
- Larose, D. T. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. John Willey& Sons, Inc.
- Larose, D. T. (2006). *Data Mining Methods and Models*. Hoboken New Jersey: Jhon Wiley & Sons, Inc.
- Manis, N. D., Sari, Y. A., & Cholissodin, I. (2021). Pengelompokan Sentimen Pada Twitter Tentang Pendapat Masyarakat Terhadap Karantina Selama Pandemi COVID-19 Menggunakan Metode DBSCAN. *Jurnal Pengmbanan Teknologi Informasi Dan Ilmu Komputer, Vol. 5*(No. 2).
- Mujilahwati, S. (2016). *Pre-processing Text Mining Pada Twitter*. Seminar Teknologi Informasi Nasional dan Komunikasi (SENTIKA).
- Organization, W. H. (2020). Novel Coronavirus (2019-nCoV) SITUATION REPORT - 1. *Materials and Methods, Vol. 10*.
- Patil, T. ., & Shrekar, M. . (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal of Computer Science and Applications, Vol. 6, No*, Hal. 256-261.
- Raymond, J. M. (2006). *CS 391L: Machine Learning Text Categorization*. University Of Texas at Austin.
- Rocchio, J. (n. d. (n.d.)). *The Smart Retrieval System-Experiments in Automatic Document Processing*.

- Rokom. (2021). *Kasus Pertama Omicron di Indonesia Diduga dari WNI yang Datang dari Nigeria*. Sehat Negeriku Kemenkes RI. <https://tinyurl.com/52xceudd>
- Santosa, B., & Umam, A. (2018). *Data Mining dan Big Data Analytics*. Penebar Media Pustaka.
- Saputra, J. P. B., & Bernarte, R. P. (2022). Algoritma Naive Bayes dalam Memprediksi Penyebaran Omicron Varian Covid-19 di Indonesia: Implementasi dan Analisis. *Jurnal Internasional Informatika Dan Sistem Informasi, Vol. 5, No.*
- Styawati, Hendrastuty, N., Isnain, A. R., & Ramadhani, A. Y. (2021). Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja pada Twitter dengan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT), Vol. 6(No. 3), Hal. 150-155.*