

SKRIPSI

ANALISIS SENTIMEN PADA MEDIA SOSIAL *TWITTER* TERHADAP TOKOH GUS DUR MENGGUNAKAN METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE (SVM)*

Skripsi Ini Diajukan Sebagai Syarat Melaksanakan Kewajiban Studi Strata Satu
(S1) Program Studi Sistem Informasi



Disusun Oleh :

Nur Adinda Salsabila

11180930000022

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH
JAKARTA**

2022

PERNYATAAN

DENGAN INI SAYA MENYATAKAN BAHWA SKRIPSI INI BENAR-BENAR HASIL KARYA SENDIRI DAN BELUM PERNAH DIAJUKAN SEBAGAI SKRIPSI ATAU KARYA ILMIAH PADA PERGURUAN TINGGI ATAU LEMBAGA MANAPUN.

Ciputat, 8 Oktober 2022



NUR ADINDA SALSABILA

11180930000022





ABSTRAK

Nur Adinda Salsabila – 11180930000022, Analisis Sentimen pada Media Sosial *Twitter* Terhadap Tokoh Gus Dur Menggunakan Metode *Naïve Bayes* dan *Support Vector Machine (SVM)* di bawah bimbingan **Nur Aeni Hidayah** dan **Evy Nurmianti**

Twitter merupakan media sosial urutan ke 6 yang sering diakses di Indonesia. Masyarakat yang pengetahuannya kurang terhadap Gus Dur berani menyebarkan ujaran kebencian pada media sosial *twitter* atas kontroversinya. Penelitian ini merupakan penelitian kuantitatif tentang analisis sentimen yang mengekstrak sentimen tokoh Gus Dur pada *twitter* dengan tujuan menerapkan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)*. Alur penelitian menggunakan metode *SEMMA (Sample, Explore, Modify, Model, dan Assess)*. *Sample* berupa pengumpulan data set melalui *tweet* yang di-*crawling* selanjutnya menyeleksi atribut yang penting pada tahap *explore*. Selanjutnya *modify* berupa *text preprocessing* sehingga menjadi data terstruktur. *Model* untuk pelabelan menggunakan masing-masing metode yang akan diuji. Setelah pelabelan akan diklasifikasikan menggunakan kedua metode tersebut. *Assess* berupa hasil evaluasi menggunakan *confusion matrix* dan *k-fold cross validation*. Hasil akurasi *Naïve Bayes* adalah 78,36%, sedangkan *Support Vector Machine* menghasilkan nilai akurasi sebesar 84,27% sehingga *Support Vector Machine* lebih unggul 5,91% dari *Naive Bayes*. *Support Vector Machine* menganalisis 86 sentimen positif lebih banyak daripada sentimen negatif yang artinya Gus Dur dipandang positif pada media sosial *twitter*. Hasil akurasi dari *Naïve Bayes* dan *Support Vector Machine (SVM)* menghasilkan nilai yang baik untuk topik tokoh Gus Dur, sehingga algoritma tersebut dapat diterapkan untuk klasifikasi analisis sentimen pada data baru.

Kata Kunci: Analisis Sentimen, Gus Dur, *Naïve Bayes*, *Support Vector Machine*, *SEMMA*.

Bab I-V + xv + 72 Halaman + 7 Gambar + 7 Tabel + Daftar Pustaka + Lampiran Pustaka Acuan (51, 2018-2022)



KATA PENGANTAR

Puji syukur saya ucapkan kepada Allah SWT karena atas nikmat dan rahmat Nya, penulis dapat menyelesaikan skripsi ini sebagai syarat untuk mencapai gelar Sarjana Komputer Program Studi Sistem Informasi Fakultas Sains dan Teknologi Univeritas Islam Negeri Syarif Hidayatullah Jakarta. Selama proses penyusunan skripsi ini, tidak terlepas dari beberapa bantuan, saran, kritik dan dukungan. Dalam kesempatan ini, penulis berterima kasih kepada:

1. Bapak Ir. Nashrul Hakiem, S.Si., M.T., Ph.D selaku dekan Fakultas Sains dan Teknologi.
2. Bapak Aang Subiyakto, PhD. selaku ketua Program studi Sistem Informasi dan Bapak Nuryasin, S.Kom, M.Kom selaku sekretaris Program Studi Sistem Informasi.
3. Ibu Nur Aeni Hidyah, S.E.,M.M.S.I selaku dosen pembimbing I dan Ibu Evy Nurmiati, S.Kom.,MMSI selaku dosen pembimbing II yang telah meluangkan waktu, memberikan bantuan, bimbingan, dan motivasi.
4. Seluruh dosen dan staf karyawan Fakultas Sains dan Teknologi yang telah bekerjasama dari awal perkuliahan.
5. Wahid *Foundation* yang telah menyelenggarakan beasiswa kepenulisan skripsi tentang Gus Dur sehingga dapat menunjang proses penulisan skripsi ini.
6. Kedua orang tua peneliti, Bapak Nurkholik dan Ibu Atikah beserta keluarga, kerabat, dan sahabat yang selalu mendoakan serta mendukung saya dalam mengerjakan skripsi baik secara moril maupun material.
7. Seluruh sahabat dan teman seperjuangan sistem informasi angkatan 2018 terutama kelas A yang saling memberikan semangat dan partisipasinya selama proses menuju sidang.
8. Seluruh pihak yang secara langsung maupun tidak langsung telah mendukung dan mendoakan saya dalam pengerjaan skripsi ini.

Penulis menyadari bahwa masih ada kekurangan dalam skripsi ini. Oleh karena itu, penulis terbuka terhadap kritik dan saran yang membangun dari pembaca. Penulis berharap, skripsi ini dapat memberikan manfaat bagi pembaca.

Ciputat, 25 Oktober 2022

NUR ADINDA SALSABILA

11180930000022



DAFTAR ISI

HALAMAN JUDUL	ii
PENGESAHAN UJIAN.....	iv
ABSTRAK	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL.....	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah	5
1.3 Rumusan Masalah	5
1.4 Batasan Masalah.....	6
1.5 Tujuan Penelitian.....	6
1.6 Manfaat Penelitian.....	6
1.7 Metodologi Penelitian	7
1.7.1 Metode Pengumpulan Data.....	7
1.7.2 Metode Analisis Data.....	7
1.8 Sistematika Penulisan.....	8
BAB 2 LANDASAN TEORI	11
2.1 Analisis Sentimen.....	11
2.2 Media Sosial	12
2.3. <i>Twitter</i>	14
2.4 Tokoh Gus Dur	15
2.5 Metode <i>Data Mining</i>	16
2.5.1 Metode <i>KDD</i>	16
2.5.2 Metode <i>CRISP-DM</i>	16
2.5.3 Metode <i>SEMMA</i>	16

2.6	<i>Text Preprocessing</i>	17
2.7	Metode <i>Naïve Bayes</i>	18
2.8	Metode <i>Support Vector Machine (SVM)</i>	20
2.9	<i>Confusion Matrix</i>	20
2.10	<i>Rapidminer</i>	22
2.11	Penelitian Sejenis.....	23
BAB 3 METODOLOGI PENELITIAN		35
3.1	Pendekatan Penelitian.....	35
3.2	Waktu dan Tempat Penelitian	35
3.2.1	Waktu Penelitian	35
3.2.2	Tempat Penelitian.....	36
3.3	Populasi dan Sampel Penelitian.....	36
3.4	Pengembangan Model dan Hipotesis Penelitian	36
3.4.1	Metode <i>SEMMA</i>	36
3.4.2	Metode <i>Naïve Bayes</i>	37
3.4.3	Metode <i>Support Vector Machine (SVM)</i>	38
3.5	Metodologi Pengumpulan Data.....	39
3.5.1	Studi Literatur	39
3.5.2	<i>Crawling Data</i>	39
3.6	Analisis Data dan Interpretasi Hasil	40
3.7	Prosedur Penelitian.....	40
3.8	Perangkat Penelitian	41
3.9	Ranah Penelitian.....	42
BAB 4 HASIL DAN PEMBAHASAN		45
4.1	<i>Sample</i>	45
4.1.1	Penelitian Sejenis	45
4.1.2	<i>Crawling Data</i>	45
4.2	<i>Explore</i>	47
4.3	<i>Modify</i>	48
4.3.1	<i>Cleaning</i>	48
4.3.2	<i>Tokenize</i>	49

4.3.3	<i>Transform Case</i>	50
4.3.4	<i>Stopword Removal</i>	51
4.3.5	<i>Filtering</i>	52
4.3.6	<i>Wordcloud</i>	53
4.4	<i>Model</i>	53
4.4.1	<i>Pelabelan</i>	53
4.4.2	<i>Naïve Bayes</i>	54
4.4.3	<i>Support Vector Machine</i>	55
4.5	<i>Assess</i>	56
BAB 5 PENUTUP		64
5.1	<i>Kesimpulan</i>	64
5.2	<i>Saran</i>	64
DAFTAR PUSTAKA		66



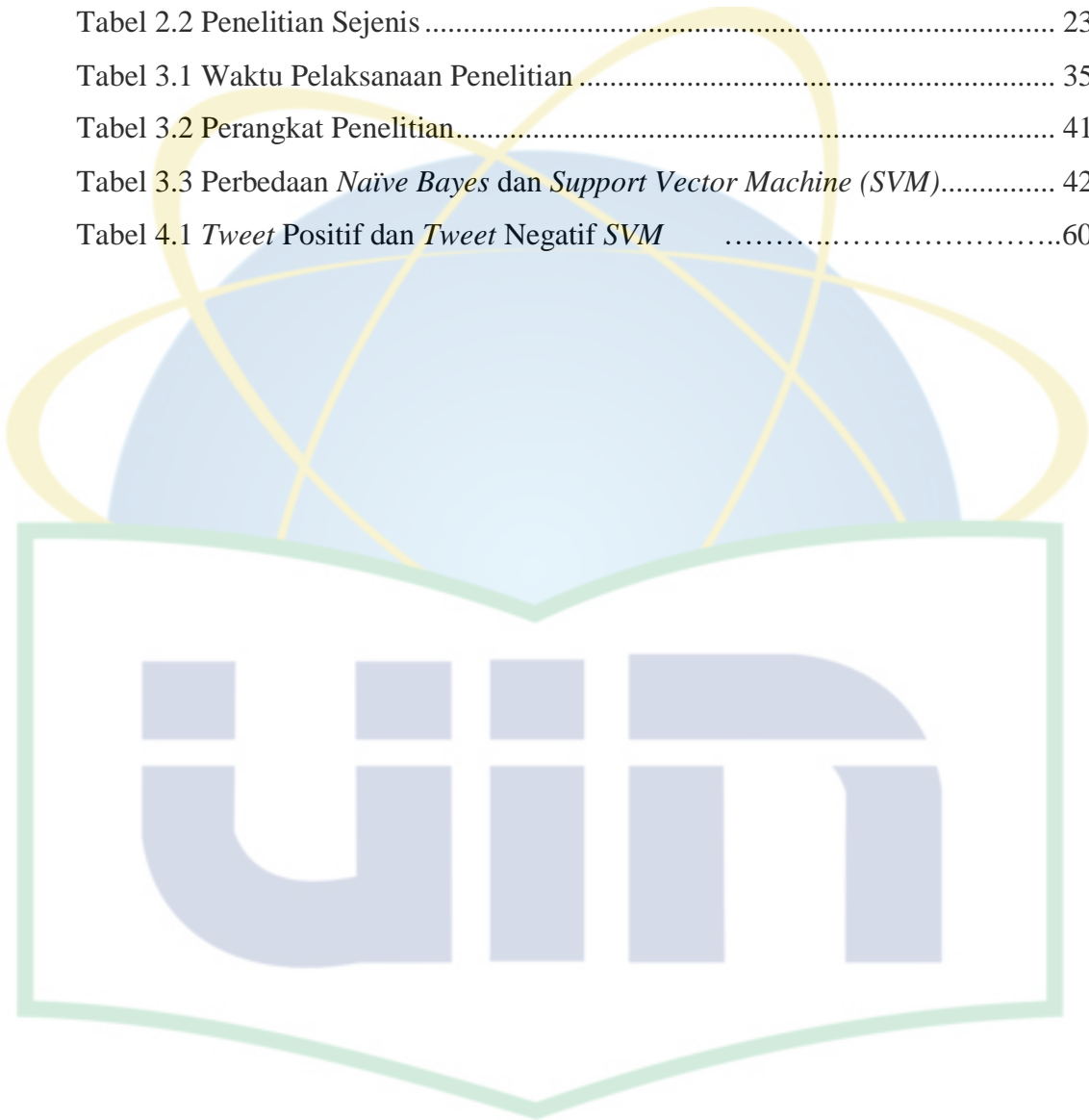
DAFTAR GAMBAR

Gambar 1.1 Data Presentase Pengguna Internet di Indonesia.....	2
Gambar 2.1 Perkembangan Pengguna Media Sosial di Indonesia.....	12
Gambar 2.2 Tren Pengguna Internet dan Media Sosial di Indonesia.....	12
Gambar 3.1 <i>Flowchart</i> Algoritma <i>Naïve Bayes</i>	38
Gambar 3.2 <i>Flowchart</i> Algoritma <i>Support Vector Machine (SVM)</i>	39
Gambar 3.3 Prosedur Penelitian (Pamungkas, 2018).....	41
Gambar 3.4 Penelitian Sejenis Metode <i>SEMMA</i>	43



DAFTAR TABEL

Tabel 1.1 Kelebihan dan Kekurangan Algoritma yang Digunakan	4
Tabel 2.1 <i>Confusion Matrix</i>	21
Tabel 2.2 Penelitian Sejenis	23
Tabel 3.1 Waktu Pelaksanaan Penelitian	35
Tabel 3.2 Perangkat Penelitian.....	41
Tabel 3.3 Perbedaan <i>Naïve Bayes</i> dan <i>Support Vector Machine (SVM)</i>	42
Tabel 4.1 <i>Tweet</i> Positif dan <i>Tweet</i> Negatif <i>SVM</i>	60





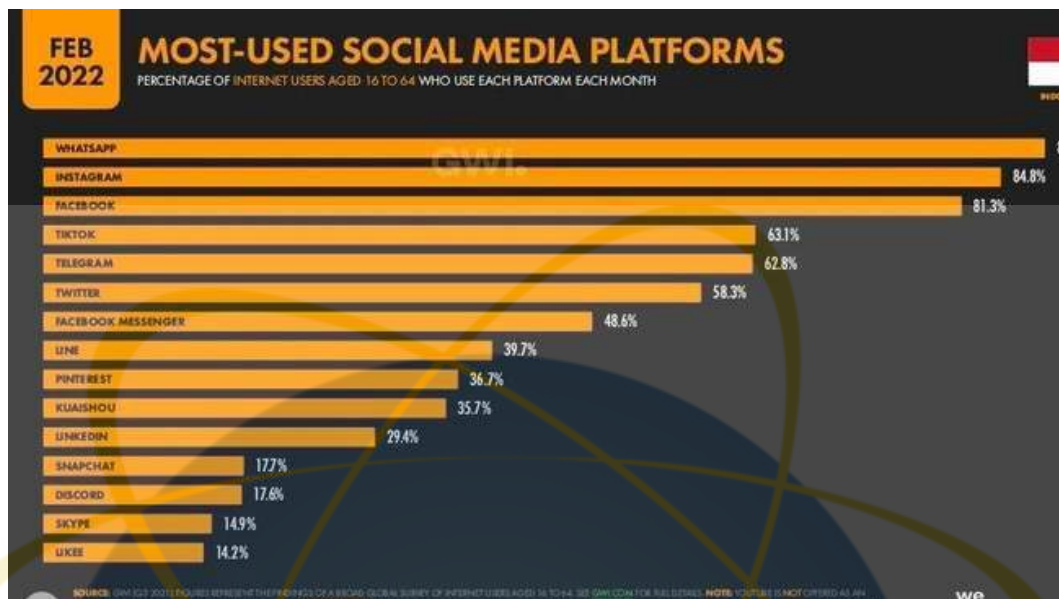
BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dunia digital semakin canggih sehingga perkembangan informasi lebih mudah didapatkan melalui portal atau media sosial. Informasi yang disajikan biasanya berupa teks pada media sosial berdasarkan pengelompokan isi informasi. Media sosial merupakan salah satu tempat untuk menyampaikan opini masyarakat. Opini masyarakat dapat berupa pujian atau ujaran kebencian dan *hoax* yang dapat menimbulkan perdebatan di media sosial. Salah satu media diskusi adalah dengan penggunaan *twitter* (Rusdaman & Rosiyadi, 2019).

Menurut hasil riset *We Are School* tahun 2022, penggunaan media sosial dapat mencapai 191,4 juta jiwa dari jumlah populasi sebesar 277,7 juta jiwa. *Twitter* mempunyai 383 juta *tweet* per hari dengan pengguna lebih dari 29,5 juta di Indonesia. *Twitter* juga merupakan media sosial urutan ke-6 yang sering diakses di Indonesia (Riyanto, 2022). Biasanya pengguna saling berbagi pesan melalui pesan singkat yang disebut *tweet*. *Tweet* inilah sebagai media menampung opini untuk mengetahui bagaimana sentimen yang muncul mengenai opini seseorang baik berupa opini positif ataupun negatif. *Twitter* digunakan sebagai analisis sentimen pada studi kasus berbagai bidang karena banyak dijadikan sebagai sumber data (I. Taufik dan S.A.Pamungkas, 2018). Selain itu, *twitter* menjadi pusat *trending* di Indonesia. *Twitter* merupakan media sosial yang mempunyai penyebaran informasi paling cepat dan tepat dari media sosial lainnya (Auliasari et al., 2021). *Twitter* mempunyai karakteristik yaitu dapat mengirim opini tanpa batas, mencari berita terkini, membagikan *tweet* pengguna lain dan memberi komentar. Adanya karakteristik *twitter* tersebut, para peneliti semakin mudah dalam menganalisis karena pengguna secara bebas dapat mengungkapkan pendapatnya, sehingga *twitter* menjadi sarana terbaik untuk melakukan analisis sentimen masyarakat (Ali et al., 2019). Berikut data presentase pengguna internet di Indonesia (Riyanto, 2022).



Gambar 1.1 Data Presentase Pengguna Internet di Indonesia

Gus Dur merupakan kyai kharismatik dan tokoh agama yang mempunyai pengaruh besar baik dalam negeri maupun luar negeri. Selama hidupnya, Gus Dur mempunyai banyak gelar dan penghargaan dari masyarakat, seperti bapak humanis dan pluralis. Menurut M. Yahya Al Mustaufi, dalam bukunya yang berjudul *Ajaran Gus Dur, Pemikiran Gus Dur dari Tasawuf hingga Demokrasi*, Gus Dur mampu memadukan nilai-nilai Islam yang selaras dengan perkembangan zaman. Namun, masih banyak masyarakat yang pengetahuannya kurang terhadap tokoh Gus Dur yang sebenarnya sehingga masyarakat berani menyebarkan ujaran kebencian terhadap Gus Dur. Gaya pemikiran Gus Dur baik sebelum dan setelah menjadi presiden pun masih menjadi perbincangan yang kontroversi di media sosial *twitter*. Akan tetapi dibalik sikap kontroversinya, Gus Dur sebagai sosok fenomenal dengan segala kecerdasan yang dimilikinya, tak heran jika kemudian sosoknya menjadi pusat penelitian dan topik diskusi (Mustaufi, 2019). Dengan menggunakan kata kunci Gus Dur, banyak opini masyarakat yang tertuang dalam *twitter*.

Analisis sentimen adalah proses mengidentifikasi dan mengelompokan opini yang masih berbentuk teks ke dalam sentimen positif atau negatif. Dalam

hal ini, penerapan analisis sentimen dapat berupa menganalisa pendapat, sentimen, evaluasi, emosi, penilaian, atau sikap pada suatu produk, tokoh, organisasi, layanan, isu, atau peristiwa yang terjadi di masyarakatnya sendiri. Selain itu, analisis sentimen akan selalu berhubungan dengan masyarakat karena sumber informasi yang didapat dari media sosial dimana masyarakat sebagai penggunaanya (Rusdaman & Rosiyadi, 2019).

Penelitian ini menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine (SVM)* yang akan membandingkan akurasi kedua metode setelah melalui tahap *preprocessing*. Penerapan metode *Naïve Bayes* tergolong sangat mudah sehingga banyak digunakan dalam penelitian. Metode *Naïve Bayes* juga dapat diterapkan pada domain yang berbeda. Metode *Support Vector Machine (SVM)* adalah metode peningkatan *margin* yang berada dalam dua kelas yang berbeda. Menurut penelitian sebelumnya, metode *Support Vector Machine (SVM)* dan *Naïve Bayes* merupakan klasifikasi dengan akurasi paling tinggi (Budi & Mude, 2020).

Para peneliti banyak yang menggunakan kedua metode tersebut pada *twitter* untuk melakukan analisis sentimen. Penelitian dengan kedua metode tersebut pernah dilakukan oleh beberapa peneliti seperti, oleh Muhammad Dwison Alizah berjudul Sentimen Analisis Terkait *Lockdow* pada Sosial Media *Twitter* (Alizah et al., 2020). Penelitian oleh Andi Taufik dan Robi Sopandi yang berjudul Klasifikasi *Tweet Infulencer* NU dengan GNPF-Ulama menggunakan *Naïve Bayes* dan *Support Vector Machine* (Taufik et al., 2021). Penelitian oleh Deni Rusdaman dan Didi Rosiyadi yang berjudul Analisa Sentimen terhadap Tokoh Publik Menggunakan Metode *Naïve Bayes Classifier* dan *Support Vector Machine* (Rusdaman & Rosiyadi, 2019). Penelitian yang dilakukan oleh Rafiq Cahyani dengan skripsi yang berjudul Analisis Sentimen pada Media Sosial *Twitter* Terhadap Tokoh Publik Peserta Pilpres 2019 (Cahyani et al., 2019). Penelitian oleh Sitti Nurul Jannah Fitriyyah, Novi Safriadi, dan Enda Esyudha Pratama yang berjudul Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial *Twitter* Menggunakan Metode *Naive Bayes* (Nurul et al., 2019). Penelitian oleh

Nurul Khasanah Fitriyani dan Anggit Dwi Hartanto yang berjudul Analisis Sentimen Terhadap Tokoh Publik Menggunakan *Support Vector Machine* (Fitriyani, 2020). Penelitian oleh Taufik dan S.A.Pamungkas yang berjudul Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma *Support Vector Machine* (I. Taufik dan S.A.Pamungkas, 2018). Penelitian yang dilakukan oleh Digna Tata Lukmana, Sri Subanti, dan Yuliana Susanti yang berjudul Analisis Sentimen Terhadap Calon Presiden 2019 dengan *Support Vector Machine* di *Twitter* (Lukmana et al., 2019). Penelitian oleh Christ Memory Sitorus berjudul Prediksi Risiko Perjalanan Transportasi *Online* dari Data Telematik Menggunakan Algoritma *Support Vector Machine* (Sitorus et al., 2020).

Tabel 1.1 Kelebihan dan Kekurangan Algoritma yang Digunakan

Algoritma	Kelebihan	Kekurangan
<i>Naïve Bayes</i>	Algoritma yang sederhana, mudah dipahami, efisien, dan berakurasi tinggi (Rahayu & Farlina, 2021).	Tingkat akurasi yang tidak bisa diukur dengan satu probabilitas tetapi membutuhkan bukti-bukti yang lainnya juga (Irwansyah Saputra, 2022).
<i>Support Vector Machine</i>	Jumlah data yang sedikit mampu mengklasifikasikan dengan tingkat akurasi yang baik (Ichwan et al., 2019).	Sulit diaplikasikan untuk jumlah data yang besar (Ichwan et al., 2019).

Adanya urgensi tokoh Gus Dur yang mengundang banyak kontroversi di *twitter* baik positif maupun negatif menjadi suatu hal yang harus diketahui oleh publik. Dengan adanya penelitian ini dapat mengetahui pendapat publik di media sosial *twitter* mengenai tokoh Gus Dur apakah cenderung positif atau negatif. Oleh karena itu, penulis mengangkat tema penelitian dalam bentuk skripsi dengan judul “**Analisis Sentimen pada Media Sosial Twitter Terhadap Tokoh Gus Dur Menggunakan Metode Naive Bayes dan Support Vector Machine (SVM)**”.

1.2 Identifikasi Masalah

Berdasarkan latar belakang di atas, penulis telah mengidentifikasi permasalahan sebagai berikut:

- a. Gus Dur sebagai tokoh agamis dan nasionalis menjadi perbincangan masyarakat di *twitter*. Ada yang mendukung Gus Dur dan adapula yang tidak sependapat dengan pemikiran Gus Dur. Postingan *tweet* dapat memengaruhi orang lain dalam menceritakan sosok Gus Dur karena dapat tersebar dengan cepat.
- b. Gus Dur sebagai tokoh yang mempunyai pengaruh besar untuk perkembangan di Indonesia maupun luar negeri kurang diketahui oleh masyarakat sehingga mengakibatkan masyarakat menyebarkan ujaran kebencian terhadap Gus Dur. Dengan adanya analisis ini yang dilakukan pada media sosial *twitter* menggunakan kedua metode ini dapat melihat perbandingan kedua metode yang paling akurat mengenai bagaimana pandangan pengguna *twitter* terhadap tokoh Gus Dur.

1.3 Rumusan Masalah

Rumusan permasalahan berdasarkan identifikasi masalah pada penelitian ini adalah “Bagaimana analisis sentimen pada media sosial *twitter* terhadap tokoh Gus Dur menggunakan metode *Naives Bayes* dan *Support Vector Machine (SVM)*?”.

1.4 Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini sebagai berikut:

- a. Sumber data yang digunakan dari *twitter*.
- b. Penggunaan *tweet* berbahasa Indonesia.
- c. Pengambilan data dengan cara *crawling* menggunakan aplikasi *Rapidminer*.
- d. Hasil klasifikasi berupa sentimen positif dan negatif.

1.5 Tujuan Penelitian

Tujuan umum dari penelitian ini adalah:

- a. Mengetahui analisis sentimen pada media sosial *twitter* terhadap tokoh Gus Dur menggunakan metode *Naives Bayes* dan *Support Vector Machine (SVM)*.

Tujuan khusus dari penelitian ini adalah:

- a. Mengetahui sentimen positif atau sentimen negatif pada data set *twitter* mengenai tokoh Gus Dur.
- b. Mengetahui kata yang sering muncul pada data set.
- c. Mengetahui hasil akurasi, presisi, dan *recall*.

1.6 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

- a. Bagi penulis, penelitian ini merupakan sebuah eksplorasi teori-teori yang selama ini dipelajari, serta menambah wawasan, ilmu pengetahuan, dan pengalaman terhadap analisis sentimen *data mining*.

- b. Bagi Universitas, sebagai tolak ukur pengetahuan mahasiswa dalam menguasai ilmu sudah dipelajari dan sebagai referensi untuk penelitian selanjutnya.
- c. Bagi pembaca, memberikan informasi mengenai sentimen terhadap tokoh Gus Dur dan bermanfaat untuk referensi penelitian analisis sentimen di bidang sistem informasi.

1.7 Metodologi Penelitian

Metode yang digunakan pada penelitian ini adalah metode kuantitatif yang dibagi menjadi 2 tahap, yaitu metode pengumpulan data dan metode analisis data.

1.7.1 Metode Pengumpulan Data

1. Studi Literatur

Studi literatur dilakukan dengan mencari teori-teori dari kumpulan jurnal dan buku mengenai penelitian yang sejenis agar mendapatkan landasan teorir sesuai dengan masalah yang akan diteliti.

2. *Crawling Data*

Crawling data adalah metode pengumpulan data dengan cara mengambil data dari media sosial. Media sosial yang digunakan adalah twitter yang dihubungkan dengan aplikasi *rapidminer*.

1.7.2 Metode Analisis Data

Metodologi yang digunakan dalam penelitian ini mengacu pada *SEMMA Data Mining Process*, sebagai berikut (Alizah et al., 2020):

a. *Sample*

Ini merupakan tahap pengumpulan data yang dilakukan dengan mencari teori-teori terkait penelitian ini yang bersumber dari jurnal, buku, ataupun situs-situs yang berkaitan. Data didapatkan dengan mengamati secara langsung dan *crawling data* pada media sosial *twitter*.

b. *Explore*

Tahap ini menjelaskan deskripsi data dan visualisasi data. Deskripsi data merupakan penjelasan gambaran data informasi yang digunakan. Objek pada deskripsi data penelitian ini adalah opini pengguna *twitter* terhadap sosok Gus Dur. Data yang digunakan adalah *tweet* bahasa Indonesia dengan kata kunci Gus Dur.

c. *Modify*

Modifikasi data *tweet* yang belum diolah melalui beberapa proses, yaitu: *cleaning, tokenize, transform case, stopword removal* dan *filtering*.

d. *Model*

Klasifikasi data berdasarkan kelasnya untuk menentukan apakah termasuk opini positif atau negatif menggunakan model algoritma.

e. *Assess*

Evaluasi pemodelan yang ada dengan cara membandingkan hasil dari prediksi terhadap data uji dengan label sentimen yang sudah didapatkan sebelumnya. Hasilnya berupa nilai *confusion matix* yaitu akurasi, presisi, dan *recall*.

1.8 Sistematika Penulisan

Dalam penyusunan laporan skripsi terdiri dari lima bab, sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, identifikasi masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini membahas mengenai dasar teori yang mendukung penelitian mengenai analisis sentimen terhadap sosok Gus Dur menggunakan metode *Naïve Bayes* dan *Support Vector Machine (SVM)* pada media sosial *twitter*.

BAB 3 METODOLOGI PENELITIAN

Bab ini menjelaskan tentang metode yang digunakan, yaitu metode pengumpulan data dan metode untuk analisis data. Untuk analisis data menggunakan metode *Naïve Bayes* dan *Support Vector Machine (SVM)*.

BAB 4 HASIL DAN PEMBAHASAN

Bab ini menggambarkan profil singkat objek penelitian dan membahas hasil analisis sentimen terhadap sosok Gus Dur berdasarkan penelitian yang sudah dilakukan dengan metode yang telah ditentukan.

BAB 5 PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian berdasarkan rumusan masalah terkait analisis sentiment terhadap sosok Gus Dur serta berisi saran yang dapat digunakan untuk pengembangan penelitian selanjutnya.

DAFTAR PUSTAKA





BAB 2

LANDASAN TEORI

2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu contoh dari bidang *Natural Language Processing (NLP)* yang paling populer. *Natural Language Processing (NLP)* adalah bidang ilmiah yang membahas tentang bagaimana caranya agar komputer bisa bekerja dan berpikir seperti manusia. *Natural Language Processing (NLP)* merupakan bagian dari *Artificial Intelligence* atau kecerdasan buatan. Dalam perkembangan *data mining*, *Artificial Intelligence (AI)* merupakan salah satu dari empat cabang ilmu data mining, yaitu statistika, *database*, dan pencarian informasi. Dalam penerapannya, *Artificial Intelligence (AI)* juga memerlukan *machine learning* sebagai algoritma penyelesaian. Adanya *machine learning* digunakan untuk menggantikan manusia dalam mengambil keputusan. *Machine learning* tidak mempunyai perasaan seperti manusia sehingga keputusan yang diambil berdasarkan data yang sudah diolah (Irwansyah Saputra, 2022).

Analisis sentimen merupakan proses komputasi menggunakan teknik analisis teks dalam bentuk data tekstual dengan cara mengelola, memahami, dan mengelompokkan emosional baik positif atau negatif. Analisis sentimen banyak digunakan karena meningkatnya kebutuhan individu atau kelompok dalam mengetahui pendapat seseorang terhadap sesuatu. Analisis sentimen juga dipengaruhi oleh data set yang digunakan akan mengalami penanganan yang berbeda (Widayat, 2021). Analisis sentimen ditujukan bukan hanya untuk pribadi tapi juga untuk organisasi. Analisis sentimen juga disebut *opinion mining* yang merupakan proses menentukan emosional pengguna dengan cara menganalisis tulisan. Tak hanya tokoh, analisis sentimen juga dapat mengatasi masalah mengenai bisnis, program, produk, aplikasi, dan lain-lain yang dapat dikomentari publik (Herlinawati et al., 2020). Keunggulan analisis sentimen ini adalah menghemat waktu dan tenaga dalam melakukan penelitian dengan jumlah data yang besar. Berikut contoh penerapan analisis sentimen:

1. Bidang bisnis, misalnya mengetahui bagaimana reputasi merek produk baru di masyarakat sehingga dapat meningkatkan merek produk tersebut.
2. Bidang politik, misalnya mengetahui popularitas seorang tokoh sehingga dapat memberi pengetahuan mengenai tokoh tersebut.
3. Program, misalnya vaksinasi, *covid 19*, pilpres, dan lain-lain sehingga dapat memperbaiki kebijakan pemerintah tersebut.

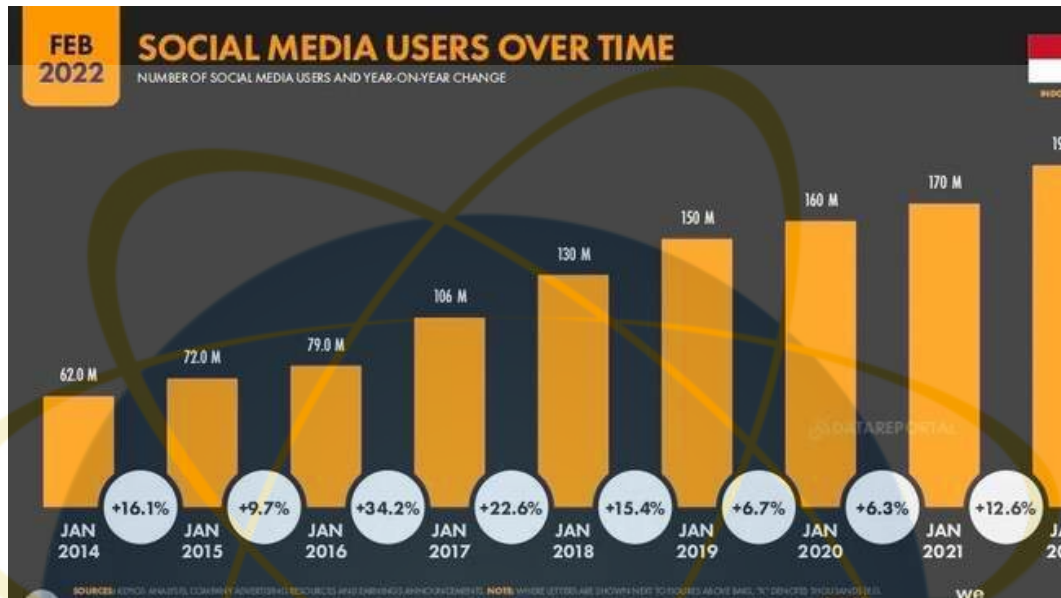
Secara umum, analisis sentimen terbagi menjadi lima langkah yaitu *crawling data*, *pre-processing*, *feature selection*, *classification*, dan *evaluation*. Analisis sentimen dapat mengubah data tidak beraturan menjadi data yang tersusun rapi. Manfaat adanya analisis sentimen yaitu sebagai evaluasi dan ide pada berbagai bidang. Analisis sentimen dapat menganalisis suatu kejadian, pernyataan, dan komentar yang kontroversi. Hasil dari analisis sentimen juga dapat menjadi sebuah gambaran bagi perusahaan, *public figure*, dan pemerintahan untuk menentukan langkah selanjutnya (Natasuwarna, 2020).

Terdapat beberapa jenis analisis sentimen yaitu *emotion detection*, *aspect-based sentiment analysis*, dan *fine grand sentiment analysis*. *Fine sentiment analysis* adalah jenis analisis yang memiliki penilaian spesifik dan biasa digunakan pada bidang *e-commerce*. *Emoticon detection* adalah jenis analisis yang bertujuan untuk mengetahui emosi yang ada pada pesan misalnya emosi bahagia, sedih, marah, dan lain-lain. *Aspect-based sentiment analysis* merupakan jenis analisis untuk mengetahui aspek yang berpengaruh dan penilaian dari pelanggan (Arviana, 2021).

2.2 Media Sosial

Di era digital media sosial semakin banyak digunakan di Indonesia bahkan di dunia. Media sosial merupakan media berbasis *online* yang digunakan untuk berinteraksi secara individu maupun kelompok dengan efektif dan efisien yang bersifat tak terbatas. Media sosial sebagai tempat mengajak yang dapat mengubah pandangan dan perilaku publik. Setiap tahunnya, media sosial selalu berkembang

pesat dan sudah familiar terutama di kalangan generasi Y dan Z (Ahmad, 2020). Berikut perkembangan pengguna media sosial di Indonesia (Riyanto, 2022).



Gambar 2.1 Perkembangan Pengguna Media Sosial di Indonesia

Berikut survei yang dilakukan oleh *We Are School* pada tahun 2022 (Riyanto, 2022).



Gambar 2.2 Tren Pengguna Internet dan Media Sosial di Indonesia

Berdasarkan gambar diatas, terdapat populasi jumlah penduduk di Indonesia sejumlah 277,7 juta, pengguna *mobile unique* 370,1 juta, pengguna internet 204,7 juta dan pengguna media sosial 191,4 juta. Media sosial juga dibagi menjadi tiga jenis yaitu (Hermawansyah & Pratama, 2021):

1. Platform pesan singkat, seperti aplikasi *Whatsapp*, *Facebook Messenger*, *Line*, dan *Telegram*.
2. Jejaring sosial, seperti *Facebook* dan *Twitter* yang paling banyak digunakan oleh pengguna.
3. Jaringan berbagi media, seperti *Youtube* dan *Instagram*.

2.3. Twitter

Salah satu media yang banyak digunakan adalah *twitter*. *Twitter* merupakan media sosial untuk bertukar pikiran dan pendapat. Pengguna *twitter* dapat mengirim dan menerima pesan *tweet* berupa teks, gambar, ataupun video. Perbedaan dengan media sosial lain, di *twitter* karakter untuk menulis pesan dibatasi sampai 280 karakter, sedangkan media sosial lainnya tidak dibatasi. *Twitter* bersifat publik sehingga status yang dibagikan dapat dilihat oleh orang lain meskipun bukan pengikutnya. Namun, pengiriman *tweet* juga dapat dibagikan hanya kepada temannya saja atau *followers*. *Twitter* mempunyai kelebihan yaitu jangkauan yang luas, dapat menjangkau publik figur, media promosi lebih luas, banyak jaringan, dan lebih mudah diukur kemampuannya. Berikut fitur yang ada pada *twitter* (I. Taufik dan S.A.Pamungkas, 2018):

1. *Tranding topic* adalah fitur yang menampilkan topik atau pembahasan teratas berupa *hashtag* yang banyak dibicarakan pengguna *twitter*.
2. *Hashtag* adalah fitur yang dapat mengelompokkan *tweet* atau pesan.
3. *Retweet* adalah fitur untuk membagikan *tweet* dari pengguna lain.
4. *Following* adalah fitur untuk menghubungkan antarpengguna atau sering disebut teman.

Data *twitter* dapat diambil menggunakan aplikasi yang dihubungkan dan *twitter*. Jika dibandingkan dengan media sosial lainnya, tidak mudah untuk mengumpulkan data secara terbuka. Media sosial lainnya tidak mengizinkan data akses karena kebijakan keamanan yang berbeda-beda. Selain itu, *twitter* juga mempunyai beberapa kecocokan dengan data mining, sebagai berikut (Wandani, 2021):

1. Format data *twitter* yang cocok dan nyaman bagi peneliti untuk dianalisis
2. Peraturan *twitter* untuk data relatif *fleksible* jika dibandingkan dengan API lainnya.
3. *Twitter* mempunyai desain yang *user friendly* atau mudah diakses bagi penggunanya.

2.4 Tokoh Gus Dur

Gus Dur atau yang mempunyai nama lengkap Abdurrahman Wachid ad Dakhil adalah tokoh nasional yang memperjuangkan nilai-nilai humanis di Indonesia. Gus Dur lahir di Jombang, Jawa Timur pada tanggal 4 Agustus 1940. Ayahnya bernama Wachid Hasyim putra dari KH Hasyim Asy'ari. Beliau mempunyai istri bernama Nyai Hj. Sinta Nuriyah. Mereka dikaruniai 4 orang anak yang bernama, Alissa Wahid, Yenny Wahid, Anita Wahid, dan Inayah Wahid. Beliau adalah kiayi dan tokoh Nahdlatul Ulama yang pernah menjabat sebagai ketua PBNU pada masanya juga merupakan presiden Republik Indonesia yang ke-4. Selain itu, beliau juga dikenal sebagai tokoh kontroversi yang banyak mengundang pro dan kontra atas kebijakannya selama menjadi presiden (Sa'diyah & Nurhayati, 2019).

Gus Dur juga sering mendapatkan penghargaan atas kontribusinya semasa hidup. Dilansir dari Jawa Pos TV, pada puncak Dies Natalis ke-61 Gus Dur mendapat penghargaan dari ITS karena telah berjasa dalam perkembangan Fakultas Teknologi Elektro dan Informatika Cerdas (FTEIC) yang diwakili oleh putri ke-4 nya yaitu Anisa Wahid. Penghargaan itu membuktikan bahwa Gus Dur berpengaruh terhadap perkembangan teknologi (Ginanjari, 2021). Gus Dur

mendapat penghargaan dari Universitas Malikkusaleh (Unimal) atas peran Gus Dur dalam penegrian Unimal sebagai bagian dari penyelesaian konflik Aceh pada tanggal 26 Februari 2021 (Bakri, 2022). Selain itu, setiap tahunnya pencinta Gus Dur atau sering disebut Gusdurian mengadakan Haul Gus Dur untuk mengenang semasa hidupnya. Adapula program di televisi yang bernama *A Tribute of Gus Dur* sebuah acara yang bertujuan untuk mengenang Gus Dur dan penghargaan terhadap Gus Dur.

2.5 Metode Data Mining

Metode dalam data mining adalah metode *Knowledge Discovery in Database (KDD)*, metode *Cross Industry Standard Process (CRISP-DM)* dan metode *SEMMA*. Setiap metode mempunyai kelebihanannya masing-masing yang dapat diterapkan sesuai dengan analisis masalah dan hasil yang diharapkan (Saputra, 2022).

2.5.1 Metode KDD

Metode *KDD* berfokus pada analisis pola yang diolah dari data. Metode *KDD* mempunyai tahapan yaitu *selection*, *preprocessing*, *transformation*, *data mining* dan *evaluation*.

2.5.2 Metode CRISP-DM

Metode *CRISP-DM* berfokus pada pemahaman kebutuhan bisnis pada data. Metode *CRISP-DM* mempunyai tahapan yaitu *bussines understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*..

2.5.3 Metode SEMMA

Metode *SEMMA* berfokus pada memodifikasi penambangan data dan pemodelan yang dirancang untuk membantu *user SAS Institute* (Saputra, 2022). Tahapan metode *SEMMA* sebagai berikut (Alizah et al., 2020):

a. *Sample*

Tahapan ini dilakukan dengan mencari teori-teori terkait penelitian ini yang bersumber dari jurnal, buku, ataupun situs-situs yang berkaitan. Data didapatkan dengan mengamati secara langsung dan *crawling* data pada media sosial *twitter*

b. *Explore*

Tahap ini menjelaskan deskripsi data dan visualisasi data. Deskripsi data merupakan penjelasan gambaran data informasi yang digunakan.

c. *Modify*

Pada tahap ini berupa modifikasi data dengan cara memilih, membuat, dan melakukan transformasi terhadap data yang diolah.

d. *Model*

Pada tahap ini melakukan penerapan model klasifikasi secara otomatis.

e. *Assess*

Pada tahap ini mengevaluasi dari pemodelan yang sudah dibuat.

2.6 *Text Preprocessing*

Data set harus melalui tahap *text preprocessing* terlebih dahulu karena data set tidak bisa digunakan tanpa melalui tahap pengelolaan data. *Text preprocessing* adalah suatu proses pengelolaan data set sebelum data tersebut diproses. Pada kenyataannya, masih banyak data set yang tidak bersih seperti kesalahan sistem saat pencatatan sehingga terjadinya data duplikat. Data yang belum diolah atau data tidak bersih kategorinya seperti format data yang tidak beraturan, adanya data kosong, tipe data yang berbeda-beda, adanya atribut yang tidak penting, dan lain sebagainya. Semakin bersih pra proses yang dilakukan, maka kemungkinan besar hasil data tersebut semakin akurat (Irwansyah Saputra, 2022). Terdapat tahapan *text preprocessing*, sebagai berikut (Rahman Isnain et al., 2021):

a. *Cleaning*

Cleaning adalah tahapan untuk membersihkan data pada set dan menyeleksi kata yang tidak diperlukan, tidak memiliki arti, atau arti yang memengaruhi sentiment seperti *html*, *link*, *mention*, dan *hashtag*.

b. *Tokenize*

Tokenisasi adalah pemecahan data set ke dalam bentuk token atau potongan kata agar mempermudah tahapan selanjutnya. Contohnya kalimat 'saya ingin makan' dipenggal menjadi ['saya', 'ingin', 'makan'] (Baihaqi et al., 2020).

c. *Transform case*

Transform case adalah tahapan merubah kalimat data teks menjadi teks yang seragam. Tahapan ini selalu ada dalam proses *text preprocessing* karena data yang ada tidak selalu terstruktur dalam penggunaan hurufnya. Dengan adanya tahapan ini dapat berperan dalam penyamarataan penggunaan huruf kapital. Sebagai contoh, kata "Data" dan "data" akan terbaca sebagai dua kata yang berbeda, sehingga melalui proses ini sistem dapat membaca secara efektif.

d. *Stopword Removal*

Stopword adalah tahapan untuk menghapus kata yang sering muncul tapi tidak memiliki arti penting dan maknanya tidak berpengaruh pada sistem, seperti 'oh', 'di', 'pada', dan sebagainya (Taufiqurrahman et al., 2021).

e. *Filter*

Filter adalah tahapan untuk menghapus kata-kata yang terlalu pendek dan terlalu panjang dengan minimal 3 huruf dan maksimal 25 huruf (Hajiati et al., 2021) .

2.7 Metode Naïve Bayes

Naïve Bayes merupakan salah satu algoritma dari *machine learning*. Dalam perkembangan *database*, *Naïve Bayes* termasuk *supervised learning* yaitu jenis *machine learning* yang membutuhkan sampel sebagai data latih yang memiliki label. *Supervised learning* dikelompokkan menjadi dua yaitu klasifikasi

dan regresi. Klasifikasi pada saat variabel menjadi kategori seperti merah atau kuning, penyakit atau tidak ada penyakit, dan sebagainya. Regresi pada saat variabel berupa nilai riil seperti berat, nilai uang, dan sebagainya. *Naïve Bayes* termasuk *supervised learning* klasifikasi seperti contoh lainnya yaitu *Support Vector Machine (SVM)*, *K-Nearest Neighbor (KNN)*, *Artificial Neural Network (ANN)*, *Trees Gradient Boosted (TGB)*, dan *Random Trees (RT)* sedangkan regresi seperti *Decision Tree*, *Logistic Regression*, dan *Kernel Regression* (Roihan et al., 2020).

Metode ini merupakan bagian dari metode *bayes* yang digunakan dalam klasifikasi teks didasarkan pada model penyederhanaan bahwa nilai atribut secara kondisional bersifat bebas jika nilai *output* diberikan. Metode *Naïve Bayes* banyak digunakan dalam teknik klasifikasi pada *twitter*. Metode ini juga digunakan dalam penambahan teks pada analisis sentimen dengan cara memprediksi probabilitas berdasarkan data sebelumnya. *Naïve Bayes* tidak bisa mendeteksi gambar, tetapi hanya bisa mendeteksi teks dan numerik. Untuk menghitung probabilitas dalam metode ini menggunakan pendekatan teorema *bayes* (Prabowo & Wiguna, 2021).

Teorema *bayes* ditemukan oleh Thomas Bayes yaitu seorang pendeta dari Inggris pada tahun 1763 dan disempurnakan oleh Laplace. Teorema *bayes* adalah pengenalan pola melalui pendekatan statistik yang fundamental. Teorema *bayes* dapat dideskripsikan seperti probabilitas terjadinya hubungan A dengan syarat hubungan B sudah terjadi, begitupun sebaliknya. Dalam bidang kedokteran modern, teorema *bayes* sering digunakan. Teorema *bayes* berperan dalam memperbaiki hitungan probabilitas dengan memanfaatkan data informasi tambahan (Retnosari, 2021). Berikut persamaan teori *bayes* (Prabowo & Wiguna, 2021):

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Persamaan di atas menunjukkan bahwa *Y* sebagai kelas spesifik, sedangkan *X* sebagai kelas yang belum diketahui. $(Y|X)$ merupakan probabilitas dari kelas berdasarkan hipotesa sebelumnya, sedangkan $P(X)$ merupakan probabilitas dari *Y*.

$P(Y/X)$ adalah hasil perkalian antara *likelihood* dan *prior* dibagi *evidence*. *Likelihood* adalah probabilitas atribut data X pada kelas Y , *prior* adalah probabilitas kelas Y dari total data set, dan *evidence* adalah probabilitas atribut data X dari total data set.

2.8 Metode Support Vector Machine (SVM)

Metode *Support Vector Machine (SVM)* termasuk jenis *machine learning supervised learning* yang membutuhkan *sample data*. *Support Vector Machine (SVM)* merupakan algoritma yang dikembangkan oleh Boser, Guyon, dan Vapnik pada tahun 1992. *Support Vector Machine (SVM)* memiliki konsep yang dikombinasi dari teori komputasi sebelumnya. Metode ini dapat mengubah data latih ke dimensi yang lebih tinggi menggunakan pola non linear (Hilda Kusumahadi et al., 2019). Metode ini juga sering digunakan dalam klasifikasi yang termasuk ke dalam *supervised learning* dan mempunyai konsep matematis yang jelas. Jika dibandingkan dengan metode lain, metode ini mempunyai tingkat akurasi yang lebih tinggi (Pratiwi et al., 2021). *Support Vector Machine (SVM)* dapat bekerja pada data set yang mempunyai dimensi tinggi menggunakan kernek trik (Pane et al., 2021). Pada pola hasil pelatihan, metode ini meliputi *machine learning* berdasarkan *Structural Risk Minimization (SRM)*. Pada metode ini dilakukan pencarian *hyperplan* terbaik atau batas keputusan untuk memisahkan dua kelas pada *input space*. Dalam hal ini berperan memisahkan antara *tweet* positif dari *tweet* negatif. Pencarian nilai *hyperplane* dilakukan menggunakan vektor bantu dan nilai *margin*. *Hyperplane* adalah bidang pemisah antara satu kelas dengan kelas lainnya. *Margin* adalah jarak antara *support vector machine* dengan *hyperplan*. (Santoso, 2021).

2.9 Confusion Matrix

Confusion matrix adalah matriks yang digunakan untuk melakukan evaluasi proses model klasifikasi berupa jumlah data uji yang benar dan salah. Dengan adanya matriks ini dapat mengetahui kualitas kinerja model klasifikasi (Normawati & Prayogi, 2021). Matriks ini berisi data target prediksi yang

dibandingkan dengan data target aktual. Data prediksi merupakan nilai yang didapatkan dari hasil pemodelan *machine learning*, sedangkan data aktual adalah nilai sebenarnya yang dimiliki. Adanya confusion matrix untuk mengetahui sejauh mana *machine learning* bekerja sesuai dengan yang diinginkan. *Confusion matrix* berisi berbagai performa yang dapat diukur seperti akurasi, presisi, *recall*, spesifikasi, dan *F1 Score* untuk mengetahui seberapa baik kinerja dari pemodelan yang telah dilakukan sebelumnya (Irwansyah Saputra, 2022). Contoh *confusion matrix* (Normawati & Prayogi, 2021).

Tabel 2.1 *Confusion Matrix*

Nilai Prediksi	Nilai Aktual	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive</i>	<i>False Negative</i>
<i>Negative</i>	<i>False Positive</i>	<i>True Negative</i>

Tabel di atas merupakan tabel *confusion matrix* dengan keterangan sebagai berikut:

- 1) *TP (True Positive)* = jumlah data nilai aktual kelas positif dan nilai prediksi kelas positif
- 2) *TN (True Negative)* = jumlah data nilai aktual negatif dan nilai prediksi negatif
- 3) *FP (False Positive)* = jumlah data nilai aktual positif dan nilai prediksi negatif
- 4) *FN (False Negative)* = jumlah data nilai aktual negatif dan nilai prediksi positif

Akurasi adalah nilai rasio data *tweet* yang sudah terdeteksi dalam pengujian. Nilai akurasi dapat menunjukkan kedekatan antara nilai prediksi sistem dan prediksi manusia (Forest & Bayes, 2021). Berikut rumusnya:

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Presisi adalah nilai ketepatan sistem mengenai informasi sistem untuk menunjukkan data positif dan data negatif yang benar. Nilai presisi ini dihasilkan dari nilai prediksi positif berbanding dengan jumlah nilai yang positif, dengan rumus sebagai berikut:

$$\text{Presisi} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall adalah nilai yang menunjukkan tingkat keberhasilan untuk mengetahui kembali informasi mengenai data positif dan negatif yang benar. *Recall* dihasilkan dari jumlah nilai *true* positif dibanding nilai aktual positif, sebagai berikut:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Spesifikasi adalah kebenaran memprediksi negatif dibandingkan dengan jumlah data negatif, dengan rumus sebagai berikut (Irwansyah Saputra, 2022):

$$\text{Spesifikasi} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

F1 Score adalah perbandingan rata-rata presisi dan recall yang dibobotkan, dengan rumus sebagai berikut (Irwansyah Saputra, 2022):

$$\text{F1 Score} = 2 \times \frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}}$$

2.10 *Rapidminer*

Aplikasi ini adalah *software* dari *Institute of Technology Blanchardstown* yang dibuat oleh Dr. Markus Hofmann. Aplikasi yang bersifat *open source* ini menggunakan bahasa pemrograman *Java*. *Rapidminer* dapat digunakan dalam

sistem operasi manapun. *Rapidminer* juga telah menyediakan fasilitas yang dibutuhkan untuk menganalisis data, sehingga pengguna tidak harus bisa *coding* (Srisulistiowati et al., 2021). *Rapidminer* sering kali digunakan untuk menganalisis jumlah data yang besar. Tampilan *rapidminer* yang bersifat *user friendly* menjadi salah satu kelebihan dari aplikasi ini. Algoritma yang digunakan disimpan dalam bentuk operator (Natasuwarna, 2019).

2.11 Penelitian Sejenis

Penelitian sejenis ini berisi penelitian sebelumnya yang berkaitan dengan tema penelitian yang diteliti. Adapun sumber penelitian sejenis berupa skripsi dan jurnal tentang analisis sentimen menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine*, yang disajikan pada Tabel 2.2 berikut.

Tabel 2.2 Penelitian Sejenis

No	Judul	Penulis	Tujuan	Metode dan Alat	Data Set	Hasil	Kekurangan dan Kelebihan
1.	Klasifikasi <i>Tweet Influencer</i> NU dengan GNPF-Ulama Menggunakan <i>Naive Bayes</i> dan <i>Support Vector Machine</i>	(Taufik et al., 2021)	Mengukur akurasi <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> pada <i>Influencer</i> NU dan GNPF-Ulama	- <i>Naïve Bayes</i> - <i>Support Vector Machine</i> - <i>Software Rapid Miner</i>	- Data set sejumlah 400 - Sumber data dari <i>Library Tweepy</i> pada data <i>twitter</i> - Label	Metode <i>Support Vector Machine</i> menghasilkan nilai akurasi dan <i>AUC</i> yang lebih baik dari <i>Naive Bayes</i> yakni sebesar	- Kekurangannya data set masih sedikit - Kelebihannya akurasi dari kedua metode dinilai cukup baik.

					data set dilakukan secara manual	77.28% dan AUC sebesar 0.863.	
2.	Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode <i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i>	(Rusdian & Rosiyadi, 2019)	Membandingkan tingkat akurasi metode <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> pada data <i>twitter</i> tokoh publik.	- <i>Naïve Bayes Classifier</i> - <i>Support Vector Machine</i> - <i>Software twiny</i>	- Data set sejumlah 200 - Sumber data dari data <i>twitter</i>	Algoritma <i>Support Vector Machine</i> memiliki tingkat keakurasiannya sebesar 73.96% sedangkan untuk algoritma <i>Naïve Bayes Classifier</i> memiliki tingkat keakurasiannya sebesar 71.94%	- Kekurangan data <i>tweet</i> masih sedikit - Kelebihan tingkat akurasi kedua metode cukup baik
3.	Analisis Sentimen Program Acara di	(Ramadhan et al., 2019)	Mengetahui akurasi dari	- <i>Naïve Bayes</i> - <i>Support</i>	- Sumber data dari <i>twitter</i>	Seluruh program acara didapatkan	- Kekurangan tidak menghasilkan secara spesifik

	Sctv Pada <i>Twitter</i> Menggunakan Metode <i>Naive Bayes</i> dan <i>Support Vector Machine</i>		metode <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> program SCTV pada <i>twitter</i> .	<i>Vector Machine</i> - <i>N-gram</i>	tahun 2019 dengan jumlah 4198 <i>tweet</i>	hasil akurasi 88,57%	tingkat akurasi kedua metode - Kelebihan memiliki banyak data <i>tweet</i>
4.	Analisa Sentimen <i>Twitter</i> Pada Pilpres 2019 Menggunakan Algoritma <i>Naive Bayes</i>	(Mahbub ah et al., 2019)	Mengeta hui tingkat akurasi metode <i>Naïve Bayes</i> di <i>twitter</i> pada Pilpres 2019	<i>Naïve Bayes</i>	- Sumber data dari <i>twitter</i> dengan jumlah 300 <i>tweet</i>	Klasifikasi data <i>tweet</i> mengguna kan algoritma <i>naive bayes classifier</i> memberik an akurasi sebesar 73%.	- Kekurangan data set yang masih sedikit - Akurasi metode cukup baik 73%

						<i>Precision</i> kelas negatif sebesar 78% dan <i>precision</i> kelas positif sebesar 66%	
5.	Analisis Sentimen Terhadap Calon Presiden 2019 dengan <i>Support Vector Machine</i> di <i>Twitter</i>	(Lukman a et al., 2019)	Mengeta hui opini masyarakat at pada calon presiden 2019 di <i>twitter</i> dengan metode <i>Support Vector Machine</i>	- <i>Support Vector Machine</i> - <i>Naïve Bayes Classifier</i>	- Sumber data diambil dari <i>twitter</i> berjumlah 20.000 <i>tweet</i> dengan kata kunci “Jokowi” dan “Prabowo”	Klasifikasi <i>SVM</i> dengan fungsi Kernel dapat melakukan klasifikasi dengan akurasi sebesar 86,82% untuk <i>tweet</i> dengan kata kunci “Jokowi” dan 86,27% untuk	- Kekurangan hanya menggunakan kernel linier tanpa mempertimban gkan kernel lain. - Kelebihan akurasi dari metode tersebut lebih dari 80%.

						untuk <i>tweet</i> dengan kata kunci “Prabowo ”	
6.	Analisis Sentimen Opini Terhadap Vaksin <i>Covid-19</i> pada Media Sosial <i>Twitter</i> Menggunakan <i>Support Vector Machine</i> dan <i>Naive Bayes</i>	(Fitriana et al., 2021)	Mengana lisis opini vaksin <i>Covid-19</i> pada <i>twitter</i> dengan <i>support vector machine</i> dan <i>naïve bayes</i> .	- <i>Support Vector Machine</i> - <i>Naïve Bayes</i>	- Sumber data dari <i>API twitter</i>	Algoritma <i>SVM</i> memiliki performa lebih baik pada bagian akurasi, presisi dan <i>recall</i> dengan nilai 90,47%, 90,23%, 90,78% dan algoritma <i>Naïve Bayes</i> adalah 88,64%, 87,32%, 88,13%, dengan	- Kekurangan data set sedikit - Kelebihan memiliki akurasi yang tinggi

						selisih akurasi 1,83%, presisi 2.91% dan <i>recall</i> 2.65%.	
7.	Analisis Sentimen <i>Zoom Cloud Meetings</i> di <i>Play Store</i> Menggunakan <i>Naïve Bayes</i> dan <i>Support Vector Machine</i>	(Herlinawati et al., 2020)	Mengetahui kinerja dan akurasi <i>Zoom Cloud Meeting</i> melalui analisis sentimen	- <i>Naïve Bayes</i> - <i>Support Vector Machine</i>	- Sumber data diambil dari 1000 <i>data set</i>	Tingkat akurasi yang didapatkan algoritma <i>Support Vector Machine (SVM)</i> 81,22% dibandingkan algoritma <i>Naïve Bayes (NB)</i> yaitu 74,37%.	- Kelebihannya jumlah <i>data set</i> hampir seimbang (<i>balanced</i>) - Akurasi metode baik diatas 80%
8.	Analisis Sentimen Terhadap Tokoh Publik	(I. Taufik dan S.A.Pamungkas,	Mengetahui sentimen tokoh publik	- <i>Support Vector Machine (SVM)</i>	- Data berasal dari <i>twitter</i> dengan	Kernel Sigmoid memiliki tingkat akurasi	- Kekurangannya data <i>tweet</i> masih sedikit. - Kelebihannya

	Menggunakan Algoritma <i>Support Vector Machine (SVM)</i>	2018)	dengan kata kunci “ahok”		jumlah 630 data <i>tweet</i> .	paling baik sekitar 81% dilanjutkan dengan kernel linear sekitar 77%. Untuk kernel Gaussian dan Polinom nilai akurasi sekitar 50%.	menggunakan kernel lengkap seperti kernel <i>linier, sigmoid, gaussian</i> dan <i>polinom</i> .
9.	Analisis Sentimen <i>Twitter</i> Kuliah Online Pasca <i>Covid-19</i> Menggunakan Algoritma <i>Support</i>	(Setiawan & Utami, 2021)	Mengetahui sentimen terhadap kebijakan kuliah online apakah cenderung positif, negatif,	- <i>Naïve Bayes</i> - <i>Support Vector Machine</i>	- Data diambil melalui <i>API Twitter</i> dengan kata kunci kuliah online didapat	<i>Naïve bayes</i> mendapatkan hasil akurasi 81,20% dan <i>SVM</i> dengan hasil akurasi optimal	- Kekurangannya data yang diperoleh masih sedikit - Kelebihannya tingkat akurasi kedua metode sangat baik yaitu di atas

	<i>Vector Machine</i> dan <i>Naive Bayes</i>		atau netral.		350 kata	yaitu 85%,	80%
10.	Sentimen Analisis Terkait <i>Lockdown</i> pada Sosial Media <i>Twitter</i>	(Alizah et al., 2020)	Mengetahui sentimen masyarakat terhadap kebijakan <i>lockdown</i>	- <i>Naive Bayes</i> - <i>Support Vector Machine</i>	Data diambil melalui <i>API Twitter</i> dengan jumlah 15.494 <i>tweet</i> .	<i>Naive Bayes</i> dengan akurasi 81% dan <i>Support Vector Machine</i> dengan akurasi 87%	- Kelebihannya data <i>tweet</i> sudah banyak - Kelebihannya tingkat akurasi kedua metode sangat baik yaitu di atas 80%
11.	Komparasi Algoritma <i>Naive Bayes</i> dan <i>Support Vector Machine</i> untuk Memprediksi Pengaruh Pembelajaran <i>Daring</i> Terhadap Motivasi Belajar	(Indrayuni & Informatika, 2018)	Mengetahui penilaian <i>review</i> positif atau negatif dari penonton mengenai film	- <i>Naive Bayes</i> - <i>Support Vector Machine</i>	Data diambil dari www.cinematic.com dengan 200 data <i>review</i> .	Akurasi algoritma <i>Naive Bayes</i> 84,50% dan <i>Support Vector Machine</i> 90%.	- Kekurangan data <i>tweet</i> masih sedikit - Kelebihannya tingkat akurasi kedua metode sangat baik yaitu di atas 80%

12.	Algoritma Klasifikasi <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> Dalam Layanan Komplain Mahasiswa	(Herman to et al., 2020)	Mengetahui perbandingan kedua metode dalam menangani data komplain mahasiswa	- <i>Naïve Bayes</i> - <i>Support Vector Machine</i>	Diambil dari database sisfo akademik (student.sbsi.ac.id).	Metode <i>Support Vector Machine</i> sebesar 84,45% dan <i>Naïve Bayes</i> sebesar 69,75%.	- Kekurangan jumlah data tidak disebutkan. - Kelebihan akurasi algoritma <i>SVM</i> lebih dari 80%.
13.	Analisis Sentimen <i>Gofood</i> Berdasarkan <i>Twitter</i> Menggunakan Metode <i>Naïve Bayes</i> dan <i>Support Vector Machine</i>	(Petiwi et al., 2022)	Menganalisa opini masyarakat terhadap kinerja <i>Gojek</i> (<i>Gofood</i>) di Indonesia	- <i>Naïve Bayes</i> - <i>Support Vector Machine</i>	Diambil dari komentar pengguna <i>Gofood</i> di <i>twitter</i> dengan <i>library tweepy</i> dan <i>Twitter API</i>	Nilai akurasi <i>Support Vector Machine</i> 98,5% dan <i>Naïve Bayes</i> 91,5%	- Kekurangan tidak disebutkan jumlah data yang diambil - Kelebihannya kedua metode nilai akurasi di atas 80%
14.	Analisis Sentimen	(Ramani zar et al.,	Membandingkan	- <i>Naïve Bayes</i>	Diambil dari	<i>Algoritma Support</i>	- Kekurangan jumlah data

	Pengguna <i>Twitter</i> terhadap Konflik antara Palestina dan Israel Menggunakan Metode <i>Naïve Bayes Classification</i> dan <i>Support Vector Machine</i>	2021)	penggun aan metode klasifikasi dalam mengide ntifikasi sentimen publik	- <i>Support Vector Machine</i>	<i>twitter</i> berbahasa indonesi a sejumlah 316 data.	<i>Vector Machinr (SVM)</i> lebih baik daripada <i>Naïve Bayes</i> dengan akurasi 74% dan 80%	hanya sedikit - Kelebihannya kedua metode ini cukup baik dalam menentukan nilai akurasi
15.	Prediksi Risiko Perjalanan Transportasi <i>Online</i> Dari Data Telematik Menggunakan Algoritma <i>Support Vector Machine</i>	(Sitorus et al., 2020)	Impleme ntasi data g untuk pembuatan model menggunakan algoritma <i>SVM</i>	- <i>Support Vector Machine (SVM)</i>	Data telemati k yang berasal dari <i>smartph one</i>	Akurasi terbaik yaitu dengan parameter $C = 100$, pada fold 1 dengan akurasi 65,63%.	- Kekurangannya data yang diambil masih terbatas pada <i>smartphone</i> -Kelebihannya mempunyai nilai parameter yang terbaik
16.	Prediksi	(Komput	mempred	- <i>Naïve</i>	Dari	<i>Naïve</i>	- Kelebihannya

	<i>Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan ID3</i>	er et al., 2022)	iksi nasabah mana yang akan melakukan churn dilihat dari atribut yang telah disesuaikan	<i>Bayes - ID3 - RFM (Recency, Frequency, Monetary)</i> - <i>Rapidminer</i>	<i>website Kaggle</i> yang merupakan data nasabah milik bank berjumlah 10.000 data	<i>Bayes</i> , hasil dari akurasinya mendapatkan 85.17% dan menggunakan <i>ID3</i> , hasil dari akurasinya mendapatkan 79.17%	mempunyai data <i>tweet</i> yang banyak dan nilai akurasinya tinggi
17.	Implementasi <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> Dengan <i>Lexicon Based</i> Untuk Analisis Sentimen Pada <i>Twitter</i>	(Arindini Nuri, 2022)	menerapkan algoritma <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> .	- <i>Naïve Bayes</i> - <i>Support Vector Machine (SVM)</i> - <i>Lexicon Based</i>	Data diambil dari <i>twitter</i> menggunakan <i>API Twitter</i>	Hasil akurasi dari metode <i>Naïve Bayes</i> adalah 74% dan metode <i>Support Vector Machine</i> dengan akurasi 85%.	- Kelebihannya mempunyai data cukup banyak dengan hasil yang sangat baik dari <i>SVM</i> .



BAB 3

METODOLOGI PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini menentukan tingkat akurasi yang paling tinggi dan hasil sentimen pengguna twitter apakah positif atau negatif yang dapat memengaruhi pandangan masyarakat terhadap tokoh Gus Dur dianalisis menggunakan pendekatan kuantitatif. *Microsoft Word 2010* dan *rapidminer* adalah aplikasi yang membantu dalam persiapan laporan dan pemrosesan data dari *twitter*. *Mendeley Dekstop* adalah aplikasi yang membantu memasukan referensi laporan penelitian ini.

3.2 Waktu dan Tempat Penelitian

3.2.1 Waktu Penelitian

Tabel 3.1 Waktu Pelaksanaan Penelitian

No.	Tahapan	Juni 2022	Juli 2022	Agustus 2022	September 2022	Oktober 2022
1.	Landasan teori					
2.	Pengumpulan data					
3.	Analisis data					
4.	Interpretasi					
5.	Pembuatan laporan					

3.2.2 Tempat Penelitian

Penelitian ini dilakukan pada media sosial *twitter* menggunakan aplikasi *rapidminer*.

3.3 Populasi dan Sampel Penelitian

Penelitian ini menggunakan populasi dan *sample* berupa *tweet* pada media sosial *twitter* dengan kata kunci Gus Dur.

3.4 Pengembangan Model dan Hipotesis Penelitian

3.4.1 Metode SEMMA

Metodologi yang digunakan dalam penelitian ini mengacu pada *SEMMA Data Mining Process* (Alizah et al., 2020). Metode *SEMMA* fokus pada modifikasi, penambahan data, dan pemodelan yang dirancang untuk membantu pengguna *software SAS enterprise miner* (Akuntansi & Primakara, 2022). Jika dibandingkan dengan metode proses data mining yang lain seperti *CRISP-DM* dan *KDD*, metode *SEMMA* memiliki tahapan yang sederhana sehingga memberikan keleluasaan pada penerapannya (Sitorus et al., 2020). Kelebihan dari metode ini tahapannya mudah dipahami dan dapat dijadikan pengembangan serta pemeliharaan proyek *data mining* yang terstruktur (Komputer et al., 2022). Berikut tahapan metode *SEMMA* (Alizah et al., 2020):

a. Sample

Berikut adalah alur mengambil data dari *twitter*:

- 1) Memiliki akun *twitter*
- 2) Menghubungkan *rapidminer* dengan akun *twitter* yang digunakan
- 3) *Crawling* data *tweet* pada *rapidminer* dengan kata kunci “gus dur” dengan batas maksimal 10000 *record* dalam sekali pengambilan dan hanya *tweet* yang berbahasa Indonesia saja.

b. *Explore*

Objek pada deskripsi data penelitian ini adalah opini pengguna *twitter* terhadap sosok Gus Dur. *Tweet* yang diambil berupa *tweet* berbahasa Indonesia dengan kata kunci Gus Dur.

c. *Modify*

Modifikasi data *tweet* yang belum diolah melalui beberapa proses, yaitu: *cleaning*, tokenisasi, *transform cases*, *stopword*, dan *filtering*. Selanjutnya melakukan visualisasi kata yang sering muncul dalam bentuk *wordcloud*.

d. *Model*

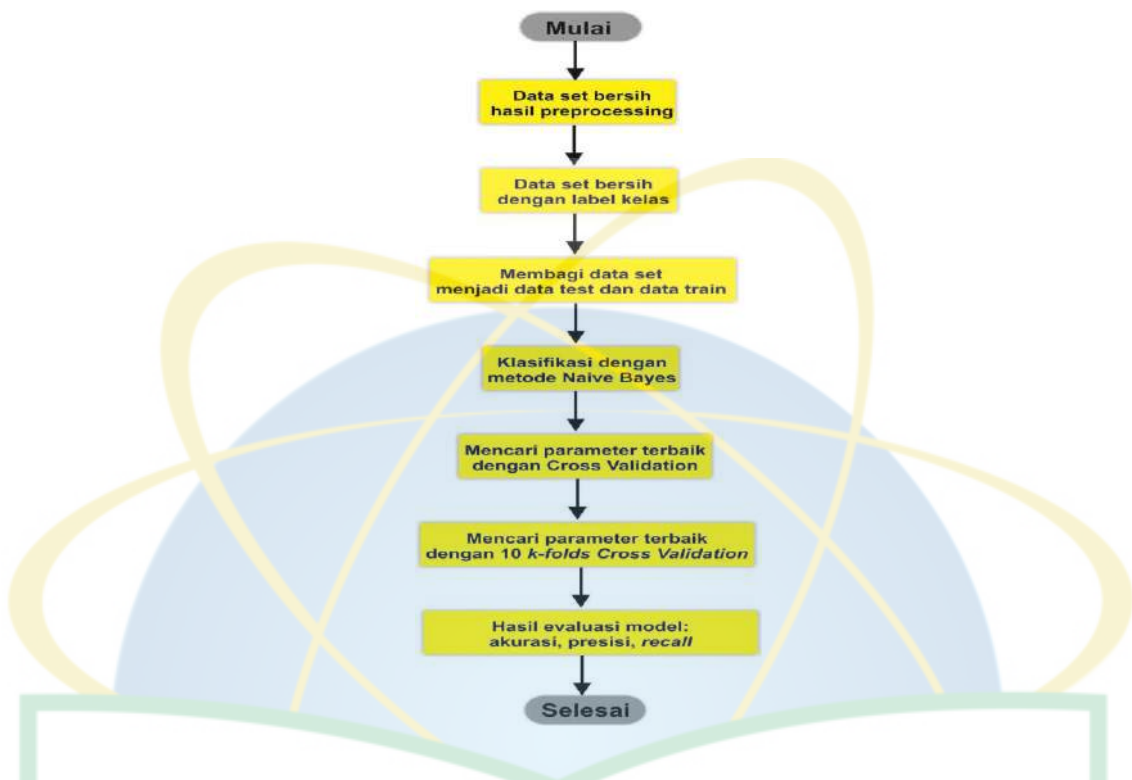
Pada tahap ini dilakukan pelabelan data berdasarkan kelasnya untuk menentukan opini positif atau negatif menggunakan pelabelan manual sebagian dan otomatis mengacu pada *dictionary*. Setelah itu, data set berupa *data train* dan *data test* diolah berdasarkan kedua metode.

e. *Assess*

Pada tahap ini membandingkan hasil dari prediksi terhadap data uji dengan label sentimen pada data uji yang sebelumnya sudah didapatkan. Hasil evaluasi berdasarkan nilai *confusion matix* yaitu akurasi, presisi, dan *recall*.

3.4.2 Metode *Naïve Bayes*

Dalam proses pengklasifikasian, keuntungan dari metode *Naïve Bayes* adalah data pelatihan yang dibutuhkan dalam jumlah kecil untuk menentukan estimasi parameter. Perbedaan yang paling mendasar antara teorema *Bayes* dengan metode lainnya adalah parameter *Bayes* dianggap menjadi variabel random, sedangkan dalam statistik klasik, parameter tidak dapat diketahui. Hubungan antara peluang bersyarat dari kejadian H dan X disebut teorema *Bayes*. Nilai akurasi yang dihasilkan lebih tinggi dibandingkan metode lainnya (Sains, 2019). Gambar 3.1 berikut adalah *flowchart* dari metode *Naïve Bayes*.

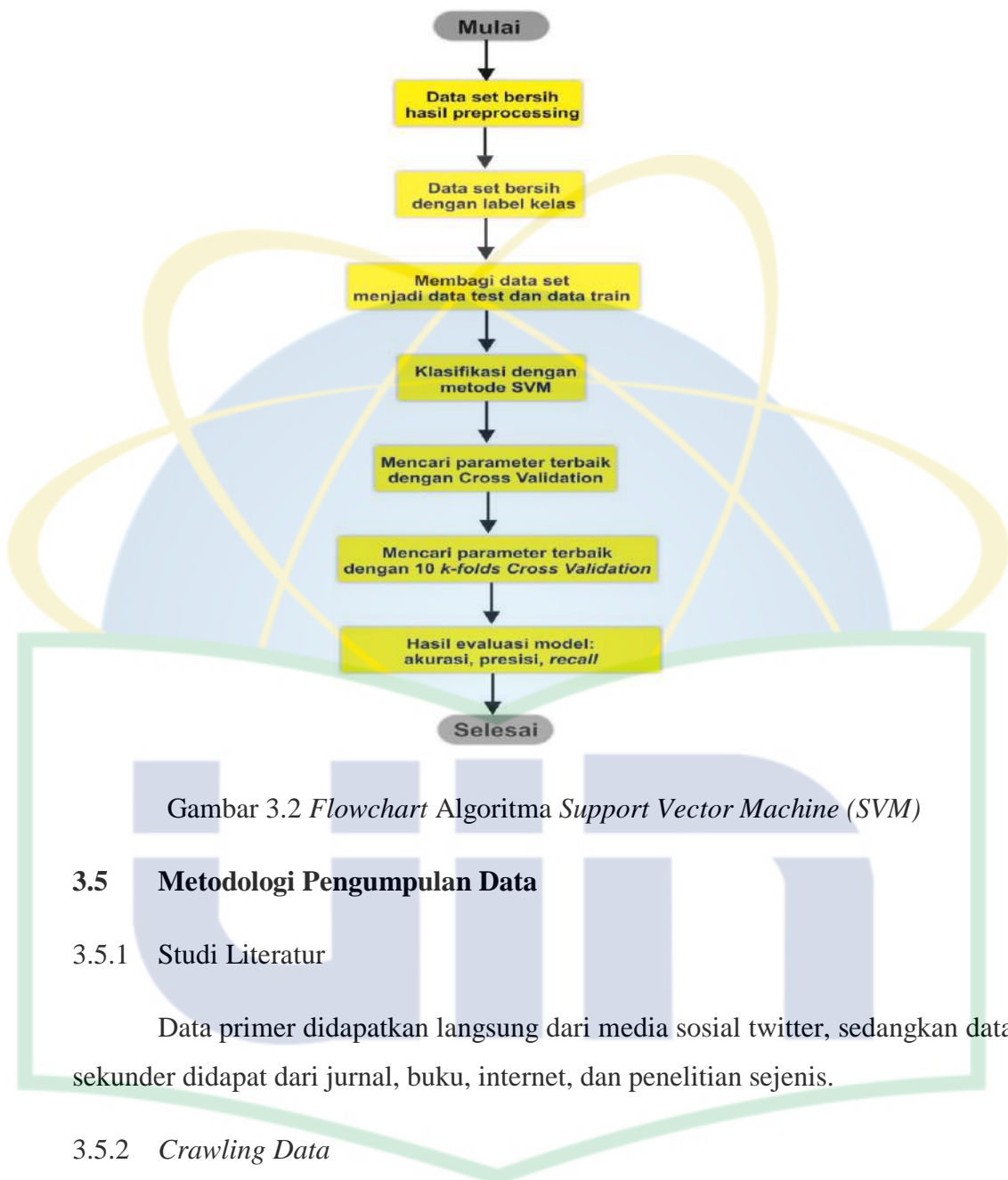


Gambar 3.1 Flowchart Algoritma Naïve Bayes

3.4.3 Metode Support Vector Machine (SVM)

Kelebihan menggunakan metode ini adalah dapat menghasilkan model klasifikasi yang baik dengan formulasi yang jelas dan hanya sedikit parameter yang diatur. Selain itu, metode ini mudah dalam penerapannya karena penentuan SVM dapat ditentukan menggunakan *QR (Quadratic Programming)* dan memiliki kemampuan generalisasi yang tinggi. Berikut alur metode *Support Vector Machine* (Ailiyya, 2020):

1. Melakukan transformasi data ke dalam ruang
2. Pengukuran pada margin antarkelas
3. Memahami parameter dalam kernel
4. Menemukan *hyperplane* terbaik



3.5 Metodologi Pengumpulan Data

3.5.1 Studi Literatur

Data primer didapatkan langsung dari media sosial twitter, sedangkan data sekunder didapat dari jurnal, buku, internet, dan penelitian sejenis.

3.5.2 Crawling Data

Crawling data adalah metode pengumpulan data dengan cara mengambil data dari media sosial. Pada penelitian ini menggunakan *twitter* yang dihubungkan dengan aplikasi *rapidminer*.

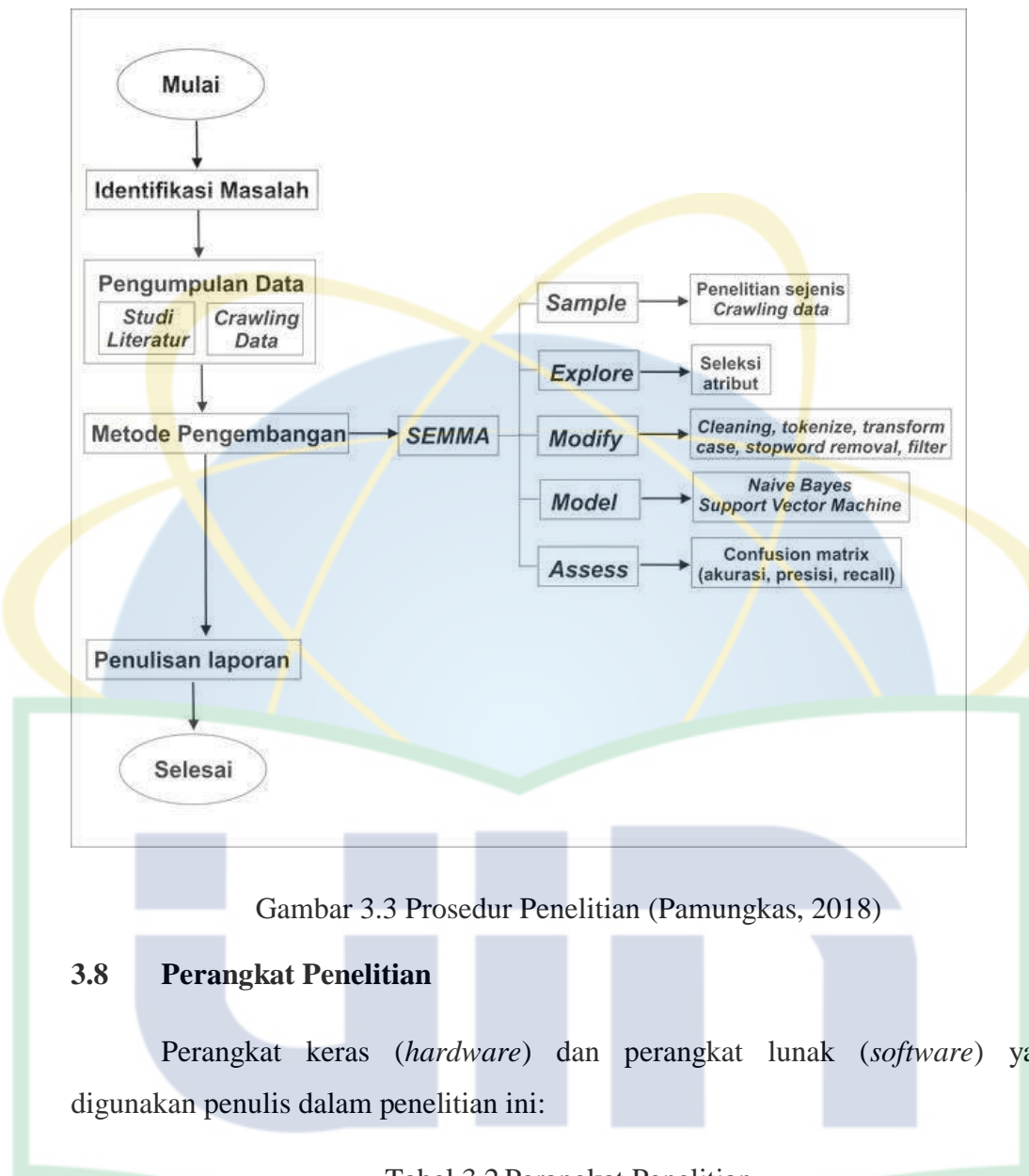
3.6 Analisis Data dan Interpretasi Hasil

Tahapan analisis data yang digunakan dalam penelitian ini adalah metode *SEMMA* (*Sample, Explore, Modify, Model, Assess*). Pada tahap *sample* dilakukan pengumpulan data berupa penelitian sejenis dan *crawling data* dari *twitter*. Tahap *explore* berupa menyeleksi atribut yang tidak digunakan. Tahap *modify* dilakukan pada data tidak terstruktur yang terdiri dari *cleaning, transform cases, tokenize, stopword removal, dan filtering* sehingga menjadi data terstruktur. Tahap *model* dilakukan pengolahan data set terstruktur menggunakan metode *Naïve Bayes* dan *Support Vector Machine (SVM)*. Tahap *assess* mengevaluasi penilaian terhadap pemodelan berupa akurasi, presisi, dan *recall*. Setelah itu, tahapan terakhir adalah kesimpulan dan saran penelitian.

Peneliti juga membandingkan penelitian ini dengan penelitian sebelumnya yang masih relevan. Hasil analisis dan interpretasi data akan disajikan secara rinci pada Bab 4.

3.7 Prosedur Penelitian

Prosedur penelitian digunakan sebagai kerangka pemikiran, bimbingan, dan penelitian dasar. Prosedur yang digunakan dijelaskan dalam Gambar 3.3 di bawah ini.



Gambar 3.3 Prosedur Penelitian (Pamungkas, 2018)

3.8 Perangkat Penelitian

Perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan penulis dalam penelitian ini:

Tabel 3.2 Perangkat Penelitian

Hardware	Laptop ASUS	Intel Celeron
		4GB RAM
		HDD
		Monitor
Software	Sistem Operasi	Windows 10

	<i>Tools</i>	<i>Rapidminer</i>
--	--------------	-------------------

3.9 Ranah Penelitian

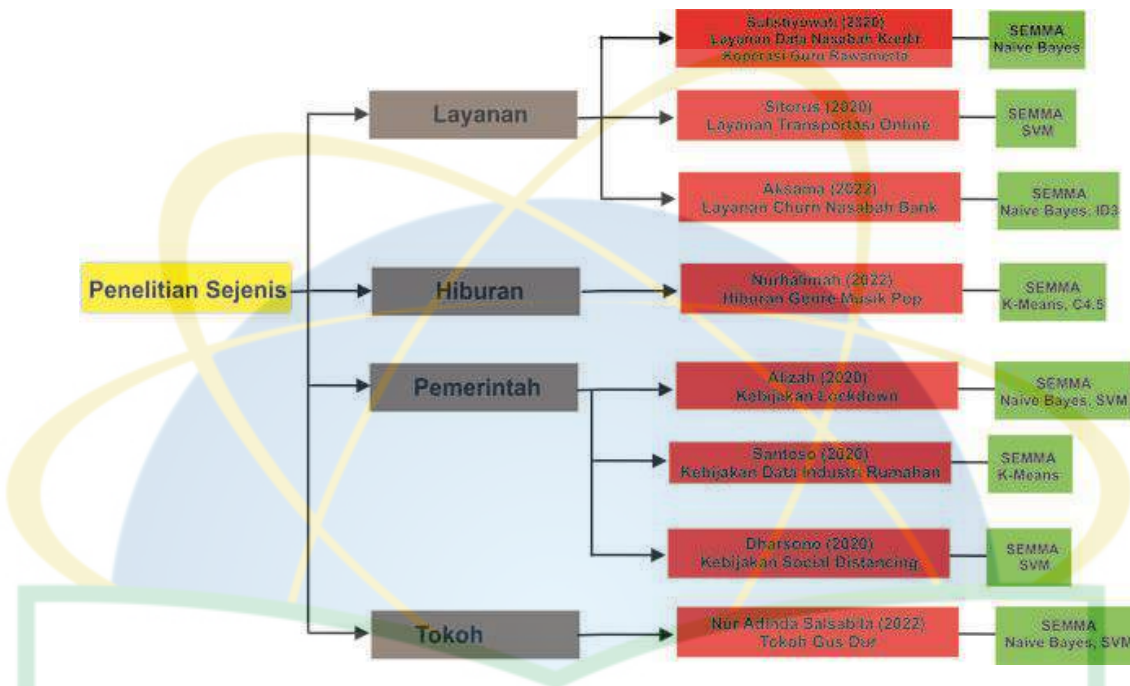
Berdasarkan penelitian sejenis, penelitian oleh (Alizah et al., 2020) merupakan referensi utama dalam penelitian ini. Kedua penelitian ini menggunakan objek penelitian yang sejenis, metode yang sama yaitu metode *SEMMA*, algoritma yang sama yaitu *Naïve Bayes* dan *Support Vector Machine*, sumber data yang sama, pelabelan yang sama dan hasil akurasi tertinggi dengan algoritma yang sama. Penulis menggunakan pelabelan manual dan pelabelan otomatis menggunakan masing-masing metode.

Adanya perbandingan kedua metode, hasil penelitian ini menjadi lebih akurat karena yang dipilih adalah metode dengan tingkat akurasi paling tinggi. Selain itu, hasil perbandingan dua metode dapat menjadi referensi untuk penelitian selanjutnya dalam memilih algoritma. Berikut Tabel 3.3 mengenai perbedaan metode *Naïve Bayes* dan *Support Vector Machine (SVM)* (Dhany, 2019):

Tabel 3.3 Perbedaan *Naïve Bayes* dan *Support Vector Machine (SVM)*

<i>Naïve Bayes</i>	<i>Support Vector Machine (SVM)</i>
Berorientasi pada probabilitas berdasarkan data sebelumnya.	Berorientasi pada linier classifier agar dapat bekerja pada problem non linier
Parameter dianggap variabel random.	Parameter sudah ditentukan dan bersifat tetap.
Menggunakan distribusi posterior yang merupakan perkalian antara fungsi <i>likelihood</i> dan prior.	Melakukan inferensia hanya berdasarkan fungsi <i>likelihood</i> .
Hanya digunakan untuk klasifikasi	Bisa digunakan untuk klasifikasi dan regresi

Penelitian ini juga menjelaskan secara rinci proses pelabelan yang belum dijelaskan pada penelitian sebelumnya. Berikut Gambar 3.4 merupakan penelitian sebelumnya menggunakan metode *SEMMA*.



Gambar 3.4 Penelitian Sejenis Metode *SEMMA*

Penerapan penelitian analisis sentimen menggunakan metode proses data mining *SEMMA* ada pada layanan, hiburan dan kebijakan pemerintahan. Oleh karena itu, penelitian analisis sentimen terhadap tokoh ini dilakukan sebagai kontribusi penelitian untuk menjadi referensi dalam proses data mining metode *SEMMA* pada bidang sistem informasi.



BAB 4

HASIL DAN PEMBAHASAN

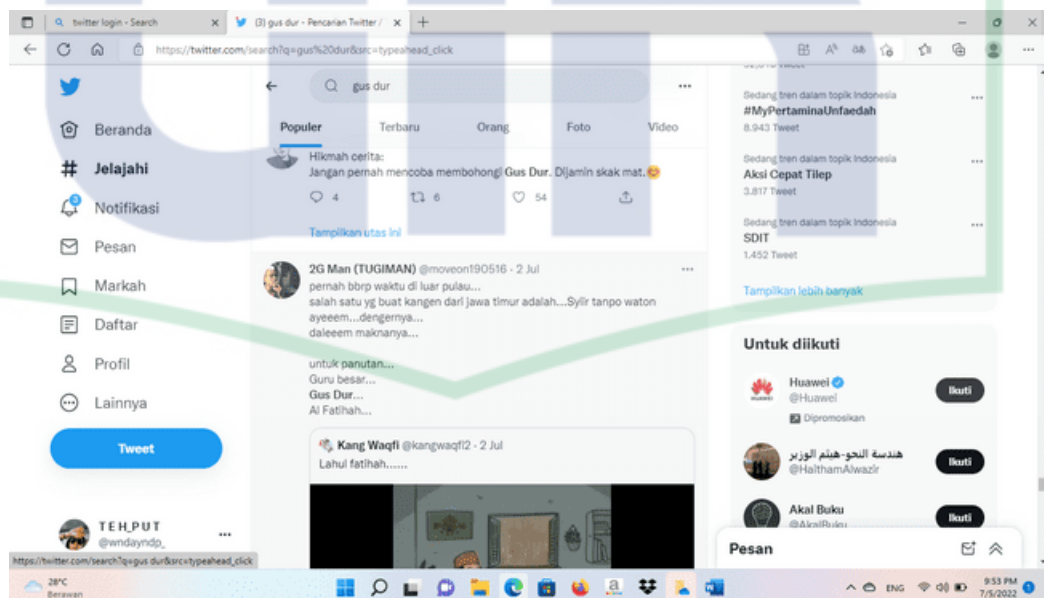
4.1 Sample

4.1.1 Penelitian Sejenis

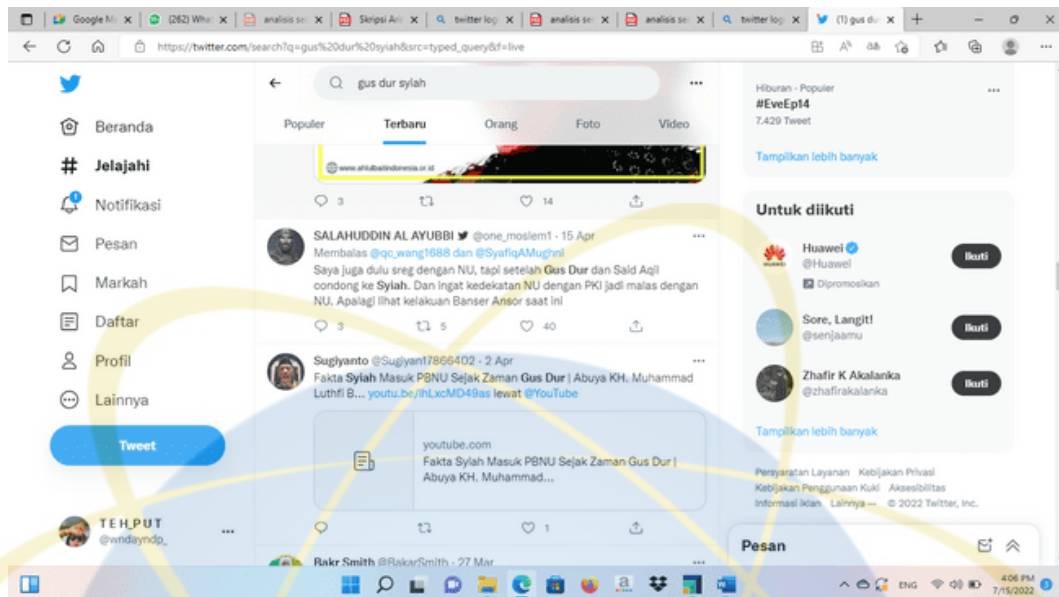
Penelitian sejenis yang dilakukan terkait materi analisis sentimen, metode *Naïve Bayes*, metode *Support Vector Machine*, dan pelabelan dari jurnal yang telah dijelaskan.

4.1.2 Crawling Data

Crawling data yang diambil dari *twitter* dengan kata kunci ‘gus dur’. *Crawling* data yang diambil dalam bentuk teks yang merupakan *tweet* dari pengguna *twitter*. Data yang diambil pada tanggal 5-25 Agustus 2022 berjumlah 743 *tweet* setelah melalui tahap *text preprocessing* berjumlah 102 *tweet* berupa opini positif dan negatif. Berikut contoh *tweet* positif dan negatif dari pengguna *twitter*.



Gambar 4.1 Contoh *tweet* positif



Gambar 4.2 Contoh *tweet negatif*

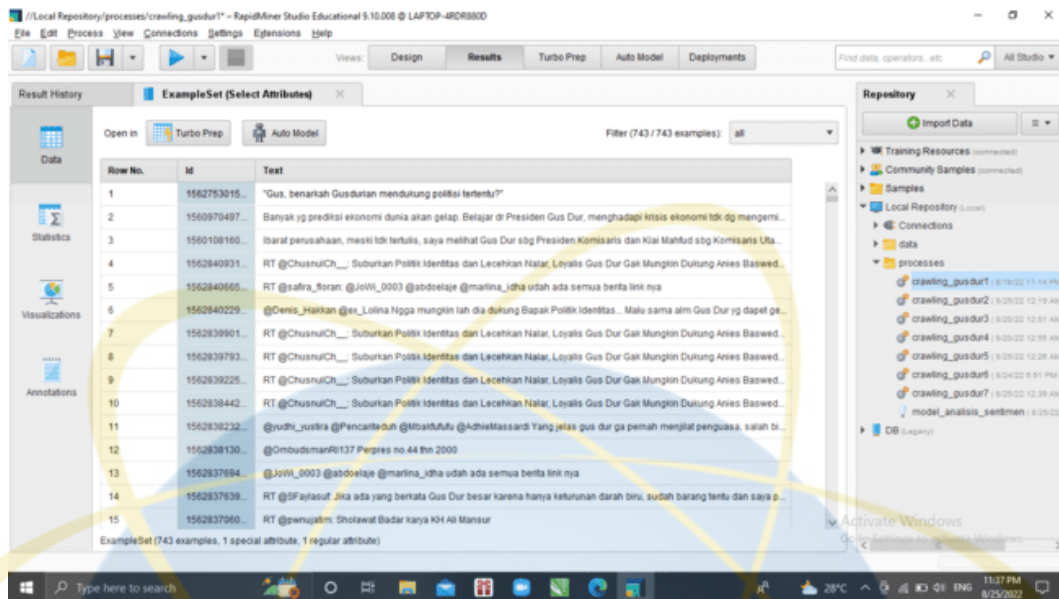
Dalam *crawling* data, penelitian ini menggunakan aplikasi *Rapidminer* yang disimpan dalam bentuk *file csv*. Peneliti menggunakan operator '*Search Twitter*' yang disambungkan pada akun *twitter* tersebut melalui koneksi *twitter* di *rapidminer*. Berikut proses *crawling* data dan hasilnya.

Row No.	Id	Created At	From User	From User Id	To User	To User Id	Language	Source	Text
1	1562753015...	Aug 25, 2022	Jaringan GU...	200604596	?	-1	in	<a href="http...	"Gus
2	1560970487...	Aug 20, 2022	Luqman Haki...	1185472145...	?	-1	in	<a href="http...	Ban
3	1560108160...	Aug 18, 2022	Luqman Haki...	1185472145...	LuqmanBee...	1185472145...	in	<a href="http...	Bar
4	1562839901...	Aug 25, 2022	Zi Khincaiko	931278234	?	-1	in	<a href="http...	RT @
5	1562839793...	Aug 25, 2022	KaptenKilik	44301509	?	-1	in	<a href="http...	RT @
6	1562839225...	Aug 25, 2022	Cempuloko	390604768	?	-1	in	<a href="http...	RT @
7	1562838442...	Aug 25, 2022	Nanda Hadin	1329226745...	?	-1	in	<a href="http...	RT @
8	1562838232...	Aug 25, 2022	Beloro Guna	1514283660	yudhi_yusira	106770398	in	<a href="http...	@yu
9	1562838130...	Aug 25, 2022	palatol	2309937103	Ombudsman...	125923625	in	<a href="http...	@Jo
10	1562837694...	Aug 25, 2022	oBafzad	1186531317...	JWV_0003	1263109250...	in	<a href="http...	@Jo
11	1562837638...	Aug 25, 2022	Banyu_Biru	367298599	?	-1	in	<a href="http...	RT @
12	1562837060...	Aug 25, 2022	Yudi Aef	1851367612	?	-1	in	<a href="http...	RT @
13	1562834217...	Aug 25, 2022	Ismi any kati...	1047853514...	?	-1	in	<a href="http...	RT @
14	1562833499...	Aug 25, 2022	Saiman Anif F...	1399803803...	?	-1	in	<a href="http...	Jka

Gambar 4.3 Hasil *Crawling Data Twitter*

4.2 Explore

Hasil *crawling* terdiri dari 12 kolom yang berisi id merupakan nomor *id twitter*, *created_at* merupakan waktu *tweet* diunggah oleh pengguna, *from_user* merupakan nama pengguna *twitter*, *from_user_id* merupakan *id* pengguna *twitter*, *to-user* merupakan *tweet* untuk *user* yang lain, *to-user-id* merupakan nomor *id user* yang di-*tweet*, *language* merupakan bahasa yang digunakan, *source* merupakan *link*, *text* merupakan isi *tweet*, *geo-location* merupakan lokasi pengguna *twitter*, dan *retweet-count* merupakan jumlah *retweet*. Agar mempermudah dalam menganalisis data, maka hanya atribut *text* saja yang ditampilkan. Gambar 4.4 berikut adalah hasil dari *filter Select Attributes*.



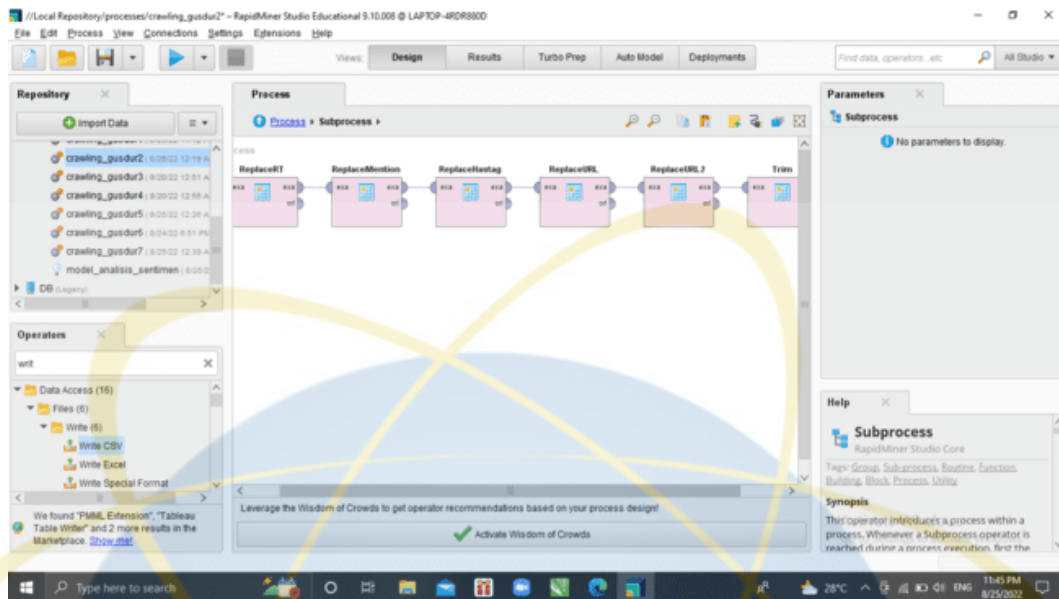
Gambar 4.4 Hasil *Filter Select Attributes*

4.3 Modify

Modifikasi dilakukan pada data set berupa *text preprocessing*. Tahapan ini bertujuan agar sistem komputer lebih mengenali bentuk data set. Selain itu, tahapan ini dapat mengubah data yang tidak tersusun menjadi data yang tersusun rapi. Dalam *text preprocessing* terdapat 5 proses yaitu *cleaning*, *tokenize*, *transform cases*, dan *filtering*.

4.3.1 Cleaning

Pada proses ini menghapus atribut yang tidak dibutuhkan dan tidak memiliki makna, seperti *hashtag*, *mention*, *retweet*, *whitespace*, dll. Berikut gambar 4.6 adalah contoh *cleaning* pada salah satu data set.



Gambar 4.5 Proses *Cleaning*

Row No.	Text
1	Gusmin tegaskan blw Jaringan GUSOURian TIDAK pernah mendukung calon tertentu dim kontestasi politik di level mana pun.
2	Gus Dur memilih memperkuat ekonomi desa dan UMKMUMKM.
3	tinggal cari ibu itu, sis.??
4	Ngga mungkin lah dia dukung Bapak Politik identitas... Malu sama alm Gus Dur yg dapet gelar Bapak Pluralisme... Bu mah kerajaan butter Anie...
5	Yang jelas gus dur ga pernah menjilat penguasa, salah bilang salah, ga kays ente, jokowi maha suci wkwkwkwk
6	k.h. abdurrachman wahid
7	public service hero
8	(GUS DUR)
9	.
10	—Maljum Jangan Lupa ?

Gambar 4.6 Hasil *Cleaning*

4.3.2 Tokenize

Selanjutnya proses tokenisasi adalah proses memecahkan kalimat menjadi potongan kata atau token untuk mengetahui asal munculnya kata. Berikut ini gambar 4.8 hasil tokenisasi pada data set.

Row No.	Sentimen	text	ABRI	AD	APA	AS	Abraham	Acara	Acco
1	positif	Gusmin tega...	0	0	0	0	0	0	0
2	positif	Gus Dur me...	0	0	0	0	0	0	0
3	positif	Ngga mungki...	0	0	0	0	0	0	0
4	positif	Yang jelas gu...	0	0	0	0	0	0	0
5	positif	Kim nggak ke...	0	0	0	0	0	0	0
6	positif	Aim KH Gus...	0	0	0	0	0	0	0
7	negatif	Emang loyal...	0	0	0	0	0	0	0
8	positif	Dulu Preside...	0	0	0	0	0	0	0
9	positif	Pada masa P...	0	0	0	0	0	0	0
10	positif	Kita patut bert...	0	0	0	0	0	0	0
11	positif	Saya salah s...	0	0	0	0	0	0	0
12	negatif	Nenek lagi b...	0	0	0	0	0	0	0
13	positif	Kata aim GU...	0	0	0	0	0	0	0
14	positif	saya sejak ik...	0	0	0	0	0	0	0

Gambar 4.7 Hasil *tokenize*

4.3.3 *Transform Case*

Transform case merupakan proses pemerataan huruf dari huruf kapital menjadi huruf kecil atau sebaliknya. Pada penelitian ini data set diubah menjadi huruf kecil semua karena *tweet* yang sudah diambil sebagian besar menggunakan huruf kecil. Berikut gambar 4.5 contoh hasil *transform case* data set.

Row No.	Sentimen	text	aamin	abraham	abri	absolut	acara	accords	ad
1	positif	gusmin lega...	0	0	0	0	0	0	0
2	positif	gus dur mem...	0	0	0	0	0	0	0
3	positif	ngga mungl...	0	0	0	0	0	0	0
4	positif	yang jelas gu...	0	0	0	0	0	0	0
5	positif	km nggak ke...	0	0	0	0	0	0	0
6	positif	alm kh gus d...	0	0	0	0	0	0	0
7	negatif	emang loyal...	0	0	0	0	0	0	0
8	positif	dulu preside...	0	0	0	0	0	0	0
9	positif	pada masa p...	0	0	0	0	0	0	0
10	positif	kita palut bert...	0	0	0	0	0	0	0
11	positif	kaya salah s...	0.330	0	0	0	0	0	0
12	negatif	nenek lagi be...	0	0	0	0	0	0	0
13	positif	kata alm gus...	0	0	0	0	0	0	0

Gambar 4.8 Hasil *Transform Case*

4.3.4 *Stopword Removal*

Proses ini dilakukan untuk menghapus kata-kata yang banyak penggunaannya tapi tidak mempengaruhi sentimen suatu kalimat. *Stopword removal* pada penelitian ini menggunakan *dictionary* yang merupakan *corpus stopwords* berbahasa Indonesia. Pada gambar 4.10 adalah contoh hasil *stopword removal* pada data set.

Row No.	Sentimen	text	
1	positif	gusmin tegaskan bhw jaringan gusdurian mendukung calon kontestasi politik level	0
2	positif	gus dur memilih memperluas ekonomi desa ukm ukm	0
3	positif	ngga dukung politik identitas malu alm gus dur yg dapat gelar pluralisme mah kerajaan buzzer anies baswedan t co ke ayu	0
4	positif	gus dur ga menjilat penguasa salah bilang salah ga kaya ente jokowi maha suci	0
5	positif	km nggakt hua orang yg gus dur waras hatinya tdk dikanglangi hasad	0
6	positif	alm kh gus dur udah ngasi julukan klo dpr isinya anak ti ga berubah ampe lebaran kuda	0
7	negatif	emang loyalis gus dur ekor sih ngaruh gitu	0
8	positif	presiden gus dur mmg berencana menaikkan gaji pns tpi poli scr signifikan beliau birokrasi pendapatannya angka korupsi	0
9	positif	presiden gus dur dipercaya pns perhatian segi kemakmuran presiden gus dur berjasa pns kakak	0
10	positif	patut berterimakasih alm presiden gus dur beliau menikmah kemertahan imlek	0
11	positif	salah satunya fans gus dur smoga ganjar presiden ri jokowi aamin	0.400
12	negatif	nenek behalusinasi kayaknya ngarep dukungan keluarga gus dur ngimplimu keduwuren	0
13	positif	alm gus dur benarnya emang gaduhnya nuangan tr	0

Gambar 4.9 Hasil *Stopword Removal*

4.3.5 *Filtering*

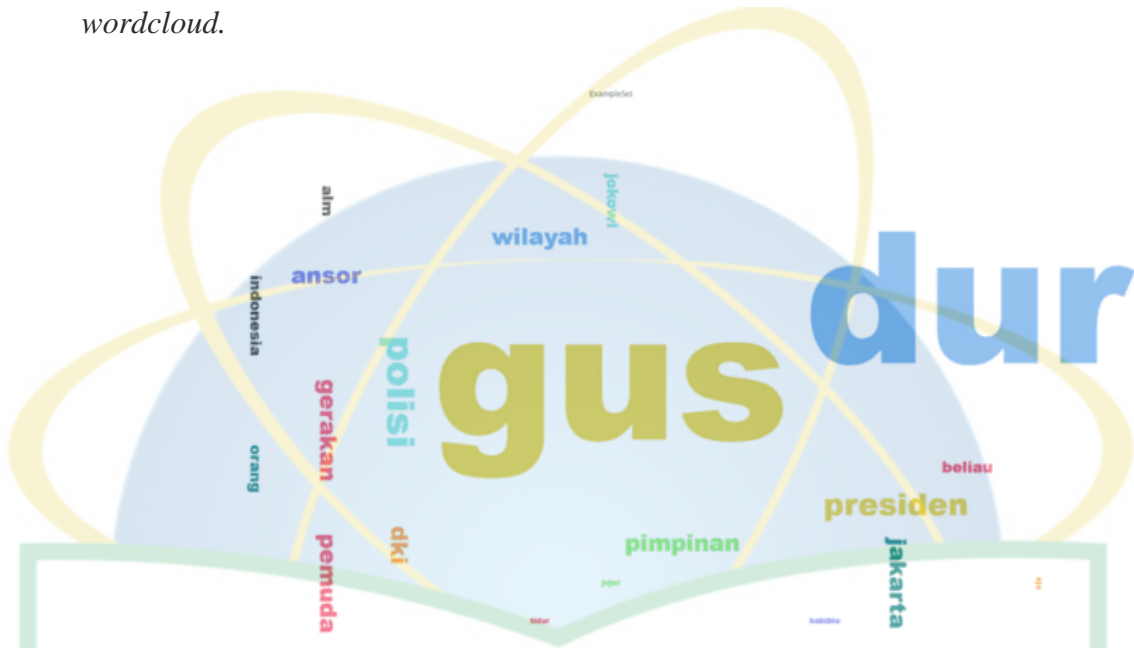
Dalam proses ini dilakukan penghapusan kata yang memiliki huruf terlalu pendek atau yang disingkat dan huruf yang terlalu panjang. Berikut adalah contoh *filtering* pada data set.

Row No.	Sentimen	text	
1	positif	gusmin tegaskan bhw jaringan gusdurian mendukung calon kontestasi politik level	0
2	positif	gus dur memilih memperluas ekonomi desa ukm ukm	0
3	positif	ngga dukung politik identitas malu alm gus dur yg dapat gelar pluralisme mah kerajaan buzzer anies baswedan ayubgm	0
4	positif	gus dur menjilat penguasa salah bilang salah kaya ente jokowi maha suci	0
5	positif	km nggakt hua orang gus dur waras hatinya tdk dikanglangi hasad	0
6	positif	alm gus dur udah ngasi julukan klo dpr isinya anak berubah ampe lebaran kuda	0
7	negatif	emang loyalis gus dur ekor sih ngaruh gitu	0
8	positif	presiden gus dur mmg berencana menaikkan gaji pns tpi poli scr signifikan beliau birokrasi pendapatannya angka korupsi lun...	0
9	positif	presiden gus dur dipercaya pns perhatian segi kemakmuran presiden gus dur berjasa pns kakak	0
10	positif	patut berterimakasih alm presiden gus dur beliau menikmah kemertahan imlek	0
11	positif	salah satunya fans gus dur smoga ganjar presiden jokowi aamin	0.400
12	negatif	nenek behalusinasi kayaknya ngarep dukungan keluarga gus dur ngimplimu keduwuren	0
13	positif	alm gus dur benarnya emang gaduhnya nuangan	0

Gambar 4.10 Hasil *Filtering*

4.3.6 Wordcloud

Wordcloud merupakan visualisasi data set yang diambil dari kata yang sering muncul pada *tweet*. Gambar 4.12 adalah contoh visualisasi data dengan *wordcloud*.



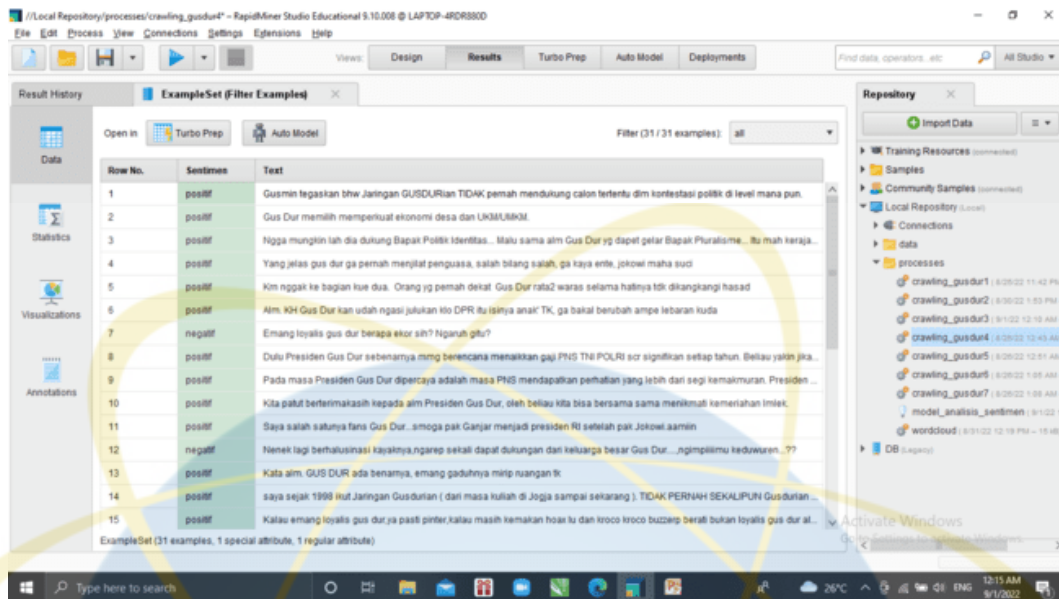
Gambar 4.11 Contoh *Wordcloud*

4.4 Model

Tahap pemodelan dalam penelitian ini meliputi pelabelan secara manual dan dua pemodelan klasifikasi dengan metode *Naïve Bayes* dan *Support Vector Machine (SVM)*.

4.4.1 Pelabelan

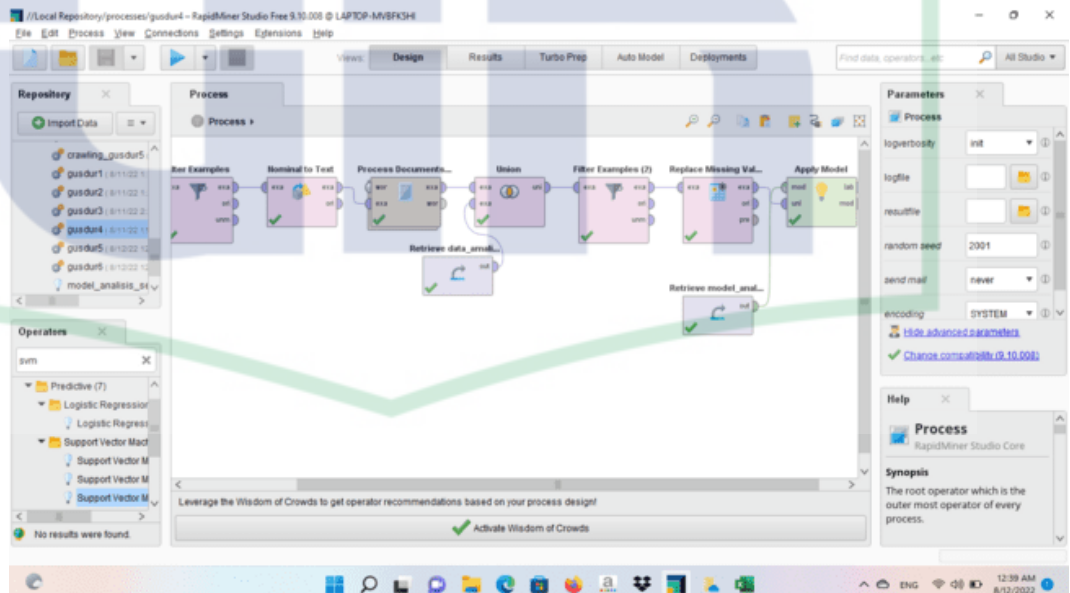
Data yang sudah melalui *preprocessing* kemudian melakukan pelabelan secara manual dan otomatis. Pelabelan secara manual dilakukan pada sebagian *tweet*, sedangkan pelabelan otomatis mengacu pada *dictionary* opini positif dan negatif. Berikut merupakan hasil pelabelan baik secara manual.



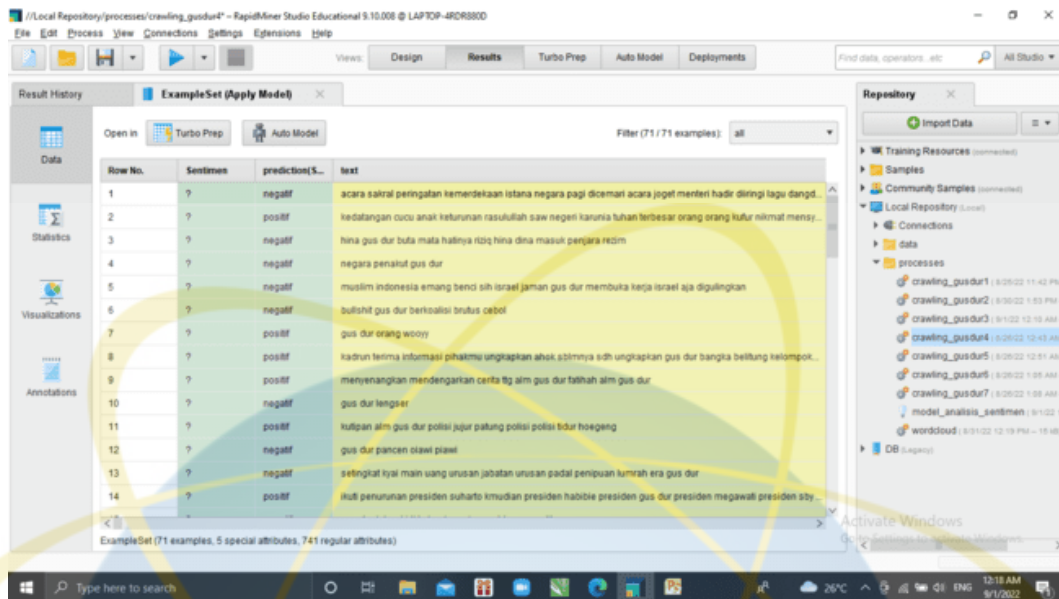
Gambar 4.12 Contoh Pelabelan Manual (Data Latih)

4.4.2 Naïve Bayes

Pada penelitian ini data set dibagi dua bagian yaitu data uji dan data latih dengan rasio *test* dan *train* yaitu 30% data latih dan 70% data uji.



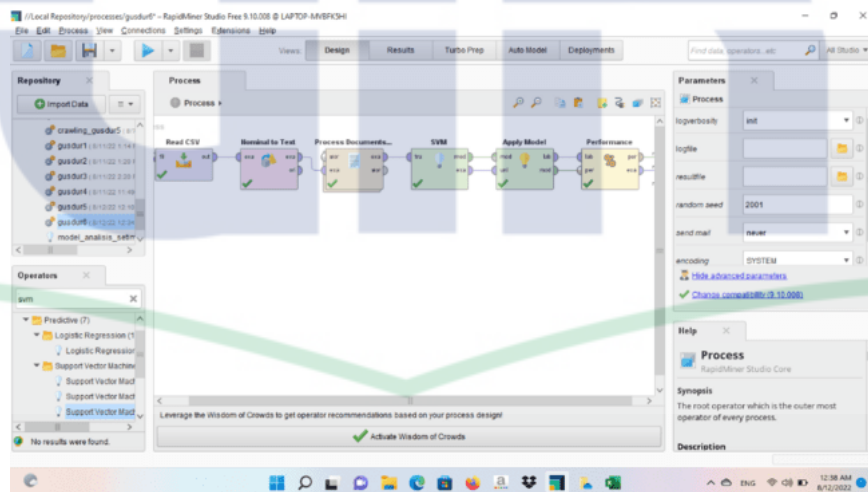
Gambar 4.13 Proses Pelabelan Naïve Bayes (Data Uji)



Gambar 4.14 Hasil Pelabelan *Naïve Bayes* (Data Uji)

4.4.3 *Support Vector Machine*

Seperti halnya metode *Naïve Bayes*, tahap ini membagi data set menjadi dua bagian yaitu data latih dan data uji dengan rasio *test* dan *train* yaitu 30% data latih dan 70% data uji.



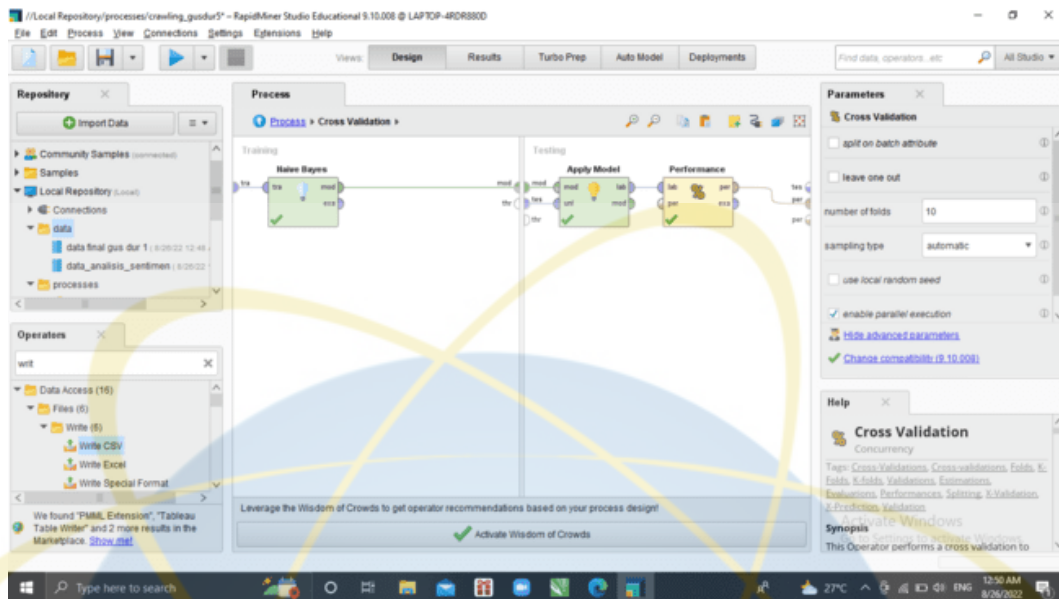
Gambar 4.15 Proses Pelabelan *SVM* (Data Uji)

Row No.	Sentences	prediction(S...	confidence(...	confidence(...	text	aamin	abraham	atri	absc
32	?	posItf	0.734	0.266	acara sakral ...	0	0	0	0
33	?	posItf	0.733	0.267	kedatangan c...	0	0	0	0
34	?	posItf	0.733	0.267	hina gus dur...	0	0	0	0
35	?	posItf	0.733	0.267	negara pena...	0	0	0	0
36	?	posItf	0.733	0.267	muslim indo...	0	0	0	0
37	?	posItf	0.733	0.267	bulshat gus d...	0	0	0	0
38	?	posItf	0.733	0.267	gus dur gran...	0	0	0	0
39	?	posItf	0.733	0.267	kadrun larim...	0	0	0	0
40	?	posItf	0.733	0.267	menyenangt...	0	0	0	0
41	?	posItf	0.733	0.267	gus dur leng...	0	0	0	0
42	?	posItf	0.734	0.266	kutipan alm g...	0	0	0	0
43	?	posItf	0.733	0.267	gus dur panc...	0	0	0	0
44	?	posItf	0.733	0.267	seingat kya...	0	0	0	0

Gambar 4.16 Hasil Pelabelan SVM (Data Uji)

4.5 Assess

Assess merupakan tahap untuk mengevaluasi kedua metode. Hasil evaluasi yang digunakan pada kedua metode ini berupa *nilai confusion matrix* berisi nilai akurasi, presisi, dan *recall* yang diambil dari data *test*. Dalam menentukan nilai *confusion matrix*, penelitian ini menggunakan *k-fold cross validation* dengan nilai $k=10$ agar nilai yang dihasilkan maksimal.



Gambar 4.17 Tahap Cross Validation Metode Naïve Bayes

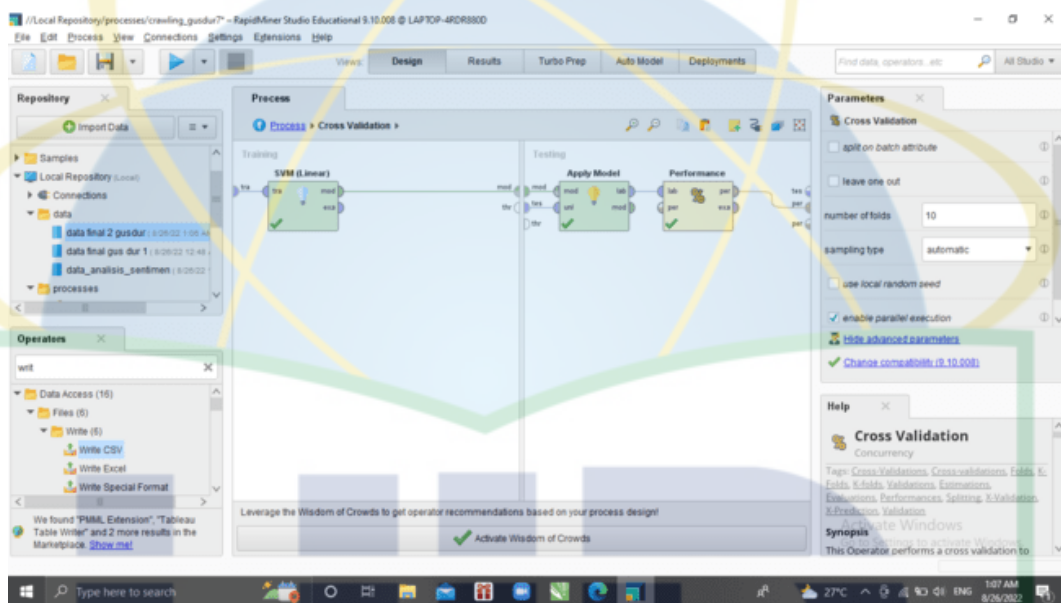
	true positif	true negatif	class precision
pred positif	61	12	83.56%
pred negatif	10	19	65.52%
class recall	85.92%	61.29%	

accuracy: 78.36% +/- 5.16% (micro average: 78.43%)

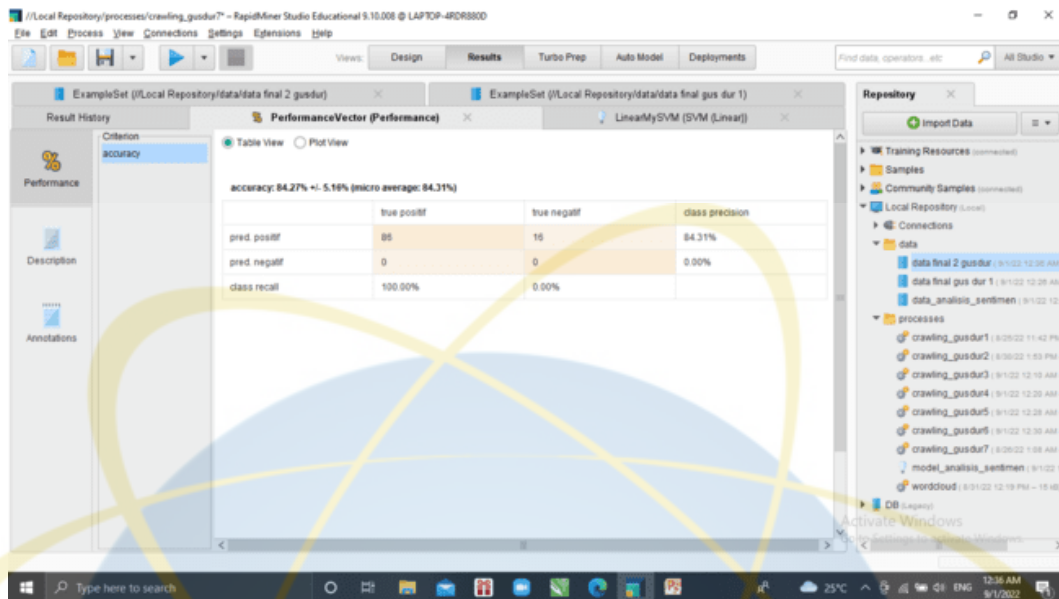
Gambar 4.18 Confusion Matrix Algoritma Naïve Bayes

Nilai akurasi yang didapat algoritma *Naïve Bayes* adalah 78,36% artinya sejumlah 78,36% model *Naïve Bayes* dapat mengklasifikasikan data yang benar. *Margin error* artinya jumlah kesalahan dalam pengambilan *sample* sebesar +- 5%

9,10 %. Terdapat 71 sentimen positif dan 31 sentimen negatif. Presisi merupakan perbandingan antara hasil prediksi dengan data yang diminta yaitu sejumlah 85,92% pada prediksi positif dan 61,29% pada prediksi negatif. Nilai *recall* menggambarkan keberhasilan model dalam menemukan kembali informasi dengan perbandingan antara rasio prediksi benar positif dan keseluruhan data prediksi positif. Didapatkan hasil pada data positif 83,56% dan pada data negatif 65,52%.



Gambar 4.19 Proses *Cross Validation* Algoritma *Support Vector Machine*



Gambar 4.20 Nilai Akurasi Algoritma *Support Vector Machine*

Nilai akurasi yang didapat algoritma *Support Vector Machine* adalah 84,27% artinya sejumlah 84,27% model ini dapat mengklasifikasikan data dengan benar. Sedangkan *margin error* atau kesalahan dalam pengambilan *sample* sejumlah +/- 5,16%. Terdapat 86 sentimen positif dan 16 sentimen negatif. Presisi merupakan perbandingan antara hasil prediksi dengan data yang diminta yaitu 100% pada prediksi positif. Nilai *recall* menggambarkan keberhasilan model dalam menemukan kembali informasi dengan perbandingan antara rasio prediksi benar positif dan keseluruhan data prediksi positif yaitu 84,31% pada data positif.

Metode klasifikasi *Naive Bayes* mendapatkan akurasi sebesar 78,36%. Sedangkan metode klasifikasi *support vector machine* mendapatkan akurasi sebesar 84,27% dengan tingkat eror lebih rendah daripada metode *Naive Bayes* sehingga metode *Support Vector Machine* memiliki kinerja yang lebih baik pada penelitian ini dibandingkan dengan metode *Naive Bayes*. Seperti halnya dalam penelitian yang dilakukan oleh Andi Taufik (Taufik et al., 2021) dengan hasil akurasi *Support Vector Machine (SVM)* lebih baik daripada *Naive Bayes* yaitu sebesar 77,28%. Selain itu, penelitian Hermanto (Hermanto et al., 2020) dengan hasil akurasi 84,45% untuk *Support Vector Machine (SVM)* yang lebih baik

daripada *Naïve Bayes*. Penelitian yang dilakukan oleh Deri (Ramadhan et al., 2019) menghasilkan akurasi 88,57% untuk *Support Vector Machine (SVM)* lebih baik dari *Naïve Bayes*. Penelitian yang dilakukan oleh Ariansyah (Ariansyah & Kusmira, 2021) menghasilkan bahwa *Support Vector Machine (SVM)* mendapatkan nilai akurasi lebih baik daripada *Naïve Bayes* yaitu sebesar 97,22%. Penelitian yang dilakukan oleh (Rusdianan & Rosiyadi, 2019) menghasilkan *SVM* sebesar 73,96% dan *Naïve Bayes* sebesar 71,94%. Penelitian yang dilakukan oleh (Fitriana et al., 2021) menghasilkan akurasi *SVM* sebesar 90,47% dan *Naïve Bayes* sebesar 88,64%. Penelitian yang dilakukan oleh (Ramanizar et al., 2021) menghasilkan nilai akurasi untuk *SVM* sebesar 80% dan *Naïve Bayes* sebesar 74%. Penelitian yang dilakukan oleh (Setiawan & Utami, 2021) menghasilkan nilai akurasi *SVM* sebesar 85% dan *Naïve Bayes* sebesar 81,20%. Penelitian yang dilakukan oleh (Indrayuni & Informatika, 2018) dengan nilai akurasi *Naïve Bayes* 84,50% dan *SVM* 90%. Penelitian yang dilakukan oleh (Herlinawati et al., 2020) menghasilkan nilai akurasi *SVM* sebesar 81,22% dan *Naïve Bayes* sebesar 74,37%.

Berdasarkan penelitian yang telah dilakukan, interpretasi hasil penelitian ini mengambil metode *Support Vector Machine (SVM)* dengan tingkat akurasi 84,27%, 86 sentimen positif, dan 16 sentimen negatif. *Tweet* positif berisi perkataan Gus Dur semasa hidupnya yang terbukti sampai sekarang, sedangkan *tweet* negatif berisi ujaran kebencian dari yang tidak mengetahui sosok Gus Dur sebenarnya. Berikut tabel 4.1 berisi *tweet* positif dan negatif.

Tabel 4.1 *Tweet* Positif dan *Tweet* Negatif SVM

<i>Tweet</i> Positif	<i>Tweet</i> negatif
ibaratnya, sudah jatuh tertimpa tangga, beliau gus dur sudah diturunkan jadi presiden, jabatan beliau pkb juga turunkan oleh muhaimin ini buat hati	apakah gus dur benci islam padahal gus dur sendiri islam dan tokoh agama

nyesek	
gua dukung prabowo aja dah kasihan beliau udah tua, udah beberapa kali nyapres sekalian biar ramalan almarhum gus dur terwujud	sejak rezim gus dur lembaga kepresidenan negara kurang berwibawa alasan keterbukaan egaliter gak segitunya kali tetap ada yang harus dijaga marwah dan kehormatannya
salah satu humor yang masih melekat hingga detik ini yaitu humor gus dur, presiden yang mengatakan ada tiga polisi jujur indonesia yaitu polisi tidur patung polisi dan jenderal hoegeng	mungkin mau balik arah aja dari ngelanturnya gus dur
tidak melupakan megawati juga bukti nya waktu dia mencalonkan presiden aku nggak memilih dia dan partainya kesalahan terbesar ada amin rais mega soalnya mengompori dia dan omongan gus dur terbukti mega gak terpilih lagi jadi presiden dan amien rais nggak jadi apa2	gus dur naikin gaji pegawai sampai 40% korupsi birokrasi makin marak
bener plesetan gus dur lembaga wakil rakyat seperti sekolah taman kanak kanak	makanya saya heran kok anaknya gus dur dukung gabener ngga salah mikir mba
cuma gus dur dan jokowi bisa bikin lembaga kepresidenan egaliter	kerusakan lembaga kepolisian penegak hukum sudah lama diteridentifikasi sejak era gus dur tanda kerusakan lembaga kepolisian lahirnya kpk
karena negara ini memang penakut	prabowo ngarep omongan gus dur jadi

tegas gus dur	kenyataan jadi presiden usia tua
berkat eyang guru tercinta bapa gus dur kami kita semua jadi tau arti toleransi saling menghargai dan menghormati perbedaan	semakin terbuka akhirnya kebobrokan busuk tubuh instuisi negara slogan melindungi mengayomi rakyat





BAB 5

PENUTUP

5.1 Kesimpulan

Kesimpulan yang dapat diambil adalah:

1. Hasil klasifikasi metode *Naïve Bayes* dengan jumlah 71 sentimen positif dan 31 sentimen negatif. Nilai akurasi 78.36%, *margin error* 9.10%, presisi prediksi positif 85,92%, presisi prediksi negatif 61,292%, *recall* data positif 83,56%, dan *recall* data negatif 65,52%.
2. Hasil klasifikasi metode *Support Vector Machine (SVM)* dengan jumlah 86 sentimen positif dan 16 sentimen negatif. Nilai akurasi 84.27%, *margin error* 5.16%, presisi prediksi positif 100%, dan *recall* data positif 84,31%.
3. Berdasarkan nilai akurasi, presisi, dan *recall* dari kedua metode, metode *Support Vector Machine (SVM)* mempunyai nilai yang lebih tinggi dibandingkan dengan metode *Naïve Bayes* sehingga yang kita ambil adalah metode *Support Vector Machine (SVM)*.
4. Berdasarkan hasil sentimen dari *Support Vector Machine (SVM)*, sentimen positif berjumlah 86 dan sentimen negatif berjumlah 16. Jadi, sentimen positif lebih banyak daripada sentimen negatif yang artinya pengguna *twitter* lebih banyak memberi opini positif terhadap Gus Dur.

5.2 Saran

Adapun saran dari penulis untuk penelitian selanjutnya:

1. Penelitian selanjutnya dapat menggunakan media sosial lain seperti *facebook* atau *instagram*.
2. Penelitian selanjutnya dapat menggunakan aplikasi lain dengan bahasa pemrograman seperti *python* atau bahasa *r*.
3. Penelitian selanjutnya dapat menambahkan sentimen netral.
4. Penelitian selanjutnya dapat menggunakan selain bahasa Indonesia



DAFTAR PUSTAKA

- Ahmad, A. (2020). *Media Sosial dan Tantangan Masa Depan Generasi Milenial*. 08(02), 134–148.
- Akuntansi, S. I., & Primakara, S. (2022). *Analisis Transaksi Penjualan Barang Menggunakan Metode Apriori pada UD . Ayu Tirta Manis. 1*, 11–20.
- Ali, W. P., Sibaroni, Y., & Si, S. (2019). *Analisis Sentimen Masyarakat Terhadap Kinerja Presiden Indonesia Dalam Aspek Ekonomi , Kesehatan , dan Pembangunan Berdasarkan Opini dari Twitter Program Studi Sarjana Informatika Fakultas Informatika Universitas Telkom Bandung Analisis Sentimen Masyarakat*. 6(2), 8637–8649.
- Alizah, M. D., Nugroho, A., Radiyah, U., & Gata, W. (2020). *Sentimen Analisis Terkait Lockdown pada Sosial Media Twitter*. *Indonesian Journal on Software Engineering (IJSE)*, 6(2), 223–229.
<https://doi.org/10.31294/ijse.v6i2.8991>
- Ariansyah, A., & Kusmira, M. (2021). *Analisis Sentimen Pengaruh Pembelajaran Daring Terhadap Motivasi Belajar Di Masa Pandemi Menggunakan Naive Bayes Dan Svm*. *Faktor Exacta*, 14(3), 100.
<https://doi.org/10.30998/faktorexacta.v14i3.10325>
- Arindini Nuri. (2022). *IMPLEMENTASI NAÏVE BAYES DAN SUPPORT VECTOR MACHINE DENGAN LEXICON BASED UNTUK ANALISIS SENTIMEN PADA TWITTER*. 90.
- Arviana, G. N. (2021). *Sentiment Analysis, Teknik untuk Pahami Maksud di Balik Opini Pelanggan*. 1 Februari.
- Auliasari, K., Prasetya, R. P., & Industri, F. T. (2021). *ANALISIS SENTIMENT CYBERBULLYING PADA SOSIAL MEDIA TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE*. 5(2), 548–556.

Baihaqi, W. M., Pinilih, M., Rohmah, M., Informasi, T., Purwokerto, U. A., Informasi, S., Purwokerto, U. A., & Korespondensi, P. (2020). Kombinasi K-Means Dan Support Vector Machine (Svm) Untuk K-Means and Support Vector Machine (Svm) Combination To Predict Sara Elements on Tweet. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(3), 501–510. <https://doi.org/10.25126/jtiik.202072126>

Bakri. (2022). *Unimal Beri Penghargaan Untuk Keluarga Gus Dur*. Serambinews.Com.

Budi, L., & Mude, A. (2020). *Perbandingan Metode Klasifikasi Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store*. 12(2), 154–161.

Fitriana, F., Utami, E., & Fatta, H. Al. (2021). *Analisis Sentimen Opini Terhadap Vaksin Covid-19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naïve Bayes*. 5(1), 19–25.

Forest, R., & Bayes, N. (2021). *Perbandingan Akurasi , Recall , dan Presisi Klasifikasi pada Algoritma*. 5(April), 640–651. <https://doi.org/10.30865/mib.v5i2.2937>

Ginanjari, D. (2021). *ITS Beri Penghargaan untuk Gus Dur*. JawaPos.Com.

Hajiati, S. R. I., Informatika, P. S., Komunikasi, F., Informatika, D. A. N., & Surakarta, U. M. (2021). *PERBANDINGAN TINGKAT AKURASI ALGORITMA RESTRICTED BOLTZMANN MACHINE DAN BACKPROPAGATION DALAM*.

Herlinawati, N., Yuliani, Y., Faizah, S., Gata, W., & Samudi, S. (2020). Analisis Sentimen Zoom Cloud Meetings di Play Store Menggunakan Naïve Bayes dan Support Vector Machine. *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 293. <https://doi.org/10.24114/cess.v5i2.18186>

Hermanto, Mustopa, A., & Kuntoro, A. Y. (2020). Algoritma Klasifikasi Naive

Bayes Dan Support Vector. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, 5(2), 211–220.

Hermawansyah, A., & Pratama, A. R. (2021). Analisis Profil dan Karakteristik Pengguna Media Sosial di Indonesia Dengan Metode EFA dan MCA. *Techno.Com*, 20(1), 69–82. <https://doi.org/10.33633/tc.v20i1.4289>

Hilda Kusumahadi, S., Junaedi, H., & Santoso, J. (2019). Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 54–60. <https://doi.org/10.30591/jpit.v4i1.1125>

I. Taufik dan S.A.Pamungkas. (2018). Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal "LOG!K@,"* 8(1), 71–79.

Ichwan, M., Dewi, I. A., & S, Z. M. (2019). Klasifikasi Support Vector Machine (SVM) Untuk Menentukan TingkatKemanisan Manga Berdasarkan Fitur Warna. *MIND Journal*, 3(2), 16–23. <https://doi.org/10.26760/mindjournal.v3i2.16-23>

Indrayuni, E., & Informatika, M. (2018). *KOMPARASI ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE UNTUK ANALISA SENTIMEN REVIEW FILM*. 14(2), 175–180.

Irwansyah Saputra, D. A. K. (2022). *Machine Learning untuk Pemula*. Informatika Bandung.

Komputer, S., Aksama, M. C., & Wahyuniati, A. (2022). *Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Naïve Bayes dan*. 17(1), 9–18.

Lukmana, D. T., Subanti, S., & Susanti, Y. (2019). Analisis Sentimen Terhadap Calon Presiden 2019 Dengan Support Vector Machine Di Twitter. *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT, 2002*, 154–160.

- Mahbubah, L. D., Zuliarso, E., Studi, P., Informasi, S., Informasi, F. T., Stikubank, U., Sentimen, A., Subianto, P., & Trees, D. (2019). Analisa Sentimen Twitter Pada Pilpres 2019 Menggunakan. *Sintak*, 194–195.
- Mustaufi, M. Y. Al. (2019). *Ajaran Gus Dur Pemikiran Gus Dur dari Tasawuf hingga Demokrasi* (F. Karyadi (ed.)).
- Natasuwarna, A. P. (2019). Tantangan menghadapi era revolusi 4.0 - Big data and data mining. *Seminar Nasional Hasil Pengabdian Kepada Masyarakat*, 23–27.
- Natasuwarna, A. P. (2020). Seleksi Fitur Support Vector Machine pada Analisis Sentimen Keberlanjutan Pembelajaran Daring. *Techno.Com*, 19(4), 437–448. <https://doi.org/10.33633/tc.v19i4.4044>
- No, V. (2020). *Analisis Sentimen Terhadap Tokoh Publik Menggunakan Support Vector Machine*. 5(1), 8–12.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika (J-Sakti)*, 5(2), 697–711.
- Nurul, S., Fitriyyah, J., Safriadi, N., & Pratama, E. E. (2019). *Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes*. 5(3), 279–285.
- Pamungkas, I. T. S. A. (2018). *ANALISIS SENTIMEN TERHADAP TOKOH PUBLIK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM). 1*, 69–79.
- Pane, S. F., Owen, A., & Prianto, C. (2021). Analisis Sentimen UU Omnibus Law pada Twitter Menggunakan Metode Support Vector Machine. *Jurnal Telekomunikasi Dan Komputer*, 11(2), 130. <https://doi.org/10.22441/incomtech.v11i2.10874>

Petiwi, M. I., Triayudi, A., & Sholihati, I. D. (2022). Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine. *Jurnal Media Informatika Budidarma*, 6(1), 542.
<https://doi.org/10.30865/mib.v6i1.3530>

Prabowo, W. A., & Wiguna, C. (2021). Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM. *Jurnal Media Informatika Budidarma*, 5(1), 149. <https://doi.org/10.30865/mib.v5i1.2604>

Pratiwi, R. W., H, S. F., Dairoh, D., Af'idah, D. I., A, Q. R., & F, A. G. (2021). Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM). *INISTA: Journal of Informatics, Information System, Software Engineering and Applications*, 4(1), 40–46.

Rahayu, monica putri, & Farlina, Y. (2021). penerapan metode naive bayes dalam prediksi penyebab kecelakaan kerja CV. Deka Utama. *Jurnal Larik*, 1(1), 21–27.

Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm. *Jdmsi*, 2(1), 31–37.

Ramadhan, D. A., Budi, E., & Si, S. S. (2019). ANALISIS SENTIMEN PROGRAM ACARA DI SCTV PADA TWITTER MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE. 6(2), 9736–9743.

Ramanizar, H., Fajri, A., Binsar Sinaga, R., Mubarok, H., Pangestu, A. D., & Prasvita, D. S. (2021). Analisis Sentimen Pengguna Twitter terhadap Konflik antara Palestina dan Israel Menggunakan Metode Naïve Bayesian Classification dan Support Vector Machine. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA) Jakarta-Indonesia, September*, 166–175.

Retnosari, R. (2021). PADA PERBANKAN DENGAN METODE NAIVE BAYES.

8(1).

Riyanto, A. D. (2022). *Hootsuite (We are Social): Indonesian Digital Report 2021*.

Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). *Pemanfaatan Machine Learning dalam Berbagai Bidang : Review paper*. 5(April), 75–82.

Rusdaman, D., & Rosiyadi, D. (2019). Analisa Sentimen Terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine. *CESS (Journal of Computer Engineering System and Science)*, 4(2), 230–235.

Sa'diyah, H., & Nurhayati, S. (2019). TADRIS : JURNAL PENDIDIKAN ISLAM Pendidikan Perdamaian Perspektif Gus Dur : *Tadris : Jurnal Pendidikan Islam*, Volume14(2), 175–188. <https://doi.org/10.19105/tjpi>.

Sains, F. (2019). *ANALISIS ALGORITHMS SUPPORT VECTOR MACHINE DENGAN NAIVE BAYES KERNEL PADA KLASIFIKASI DATA*. 6.

Santoso, G. T. (2021). *Analisis sentimen pada tweet dengan tagar #bpjsrasarentenir menggunakan metode support vectore machine (svm) skripsi*.

Setiawan, H., & Utami, E. (2021). *Analisis Sentimen Twitter Kuliah Online Pasca Covid-19 Menggunakan Algoritma Support Vector Machine dan Naive Bayes*. 5(1), 43–51.

Sitorus, C. M., Rizal, A., & Jajuli, M. (2020). *Prediksi Risiko Perjalanan Transportasi Online Dari Data Telematik Menggunakan Algoritma Support Vector Machine*. 6, 254–265.

Srisulistiowati, D. B., Khaerudin, M., & ... (2021). Sistem Informasi Prediksi Penjualan Alat Tulis Kantor Dengan Metode Fp-Growth (Studi Kasus Toko Koperasi Sekolah Bina Mulia). *JSI (Jurnal ...)*, 243–256.

Studi, P., Informasi, S., Sains, F., Teknologi, D. A. N., Islam, U., & Sunan, N. (2019). *Analisis sentimen pada media sosial twitter terhadap tokoh publik peserta pilpres 2019*.

Taufik, A., Sopandi, R., Studi, P., Informasi, S., Mandiri, U. N., Studi, P., Informatika, T., & Mandiri, U. N. (2021). *KLASIFIKASI TWEET INFLUENCER NU DENGAN GNPf-ULAMA MENGGUNAKAN NAIVE BAYES DAN SUPPORT VECTOR MACHINE CLASSIFICATION OF NU INFLUENCER TWEET WITH GNPf-ULAMA* Populisme Islam belakangan ini merupakan salah satu topik sentral dalam kajian politik di Indonesia . 5, 258–276.

Taufiqurrahman, F., Faraby, S. Al, & Purbolaksono, M. D. (2021). Klasifikasi Teks Multi Label pada Hadis Terjemahan Bahasa Indonesia Menggunakan Chi Square dan SVM. *E-Proceeding of Engineering*, 8(5), 10650–10659.

Wandani, A. (2021). *Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K-NN , Random Forest , dan Naive Bayes*. 5(September), 651–665.

Widayat, W. (2021). Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning. *Jurnal Media Informatika Budidarma*, 5(3), 1018. <https://doi.org/10.30865/mib.v5i3.3111>





**KEMENTERIAN AGAMA
UIN SYARIF HIDAYATULLAH JAKARTA
FAKULTAS SAINS DAN TEKNOLOGI**

Jl. Ir. H. Juanda No. 95 Ciputat 15412 Indonesia
Telp. (62-21) 7493806, 7493547 Fax. (62-21) 7493315

Website : fst.uinjkt.ac.id
Email : fst@uinjkt.ac.id

Nomor : B - 1563E/F9/ KM.01 /03/2022
Lampiran : -
Perihal : Pembimbing Skripsi

Jakarta, 23 Maret 2022

Kepada Yth.

1. Nur Aeni Hidayah S.E., M.M.S.I.
2. Evy Nurmiati S.Kom, M.MSI

Assalamualaikum, Wr Wb

Dengan ini diharapkan kesediaan Saudara untuk menjadi pembimbing I/II/ (Materi/Teknis)* penulisan skripsi mahasiswa:

Nama : NUR ADINDA SALSABILA

NIM : 11180930000022

Program Studi : Sistem Informasi

Judul Skripsi : ANALISIS SENTIMEN PADA MEDIA SOSIAL
TWITTER TERHADAP TOKOH GUS DUR
MENGUNAKAN METODE NAÏVE BAYES DAN
SUPPORT VECTOR MACHINE (SVM)

Judul tersebut telah disetujui oleh Program Studi bersangkutan pada tanggal dengan outline, abstraksi dan daftar pustaka terlampir. Bimbingan skripsi ini diharapkan selesai dalam waktu 6 (enam) bulan setelah ditandatanganinya surat penunjukan pembimbing skripsi

Apabila terjadi perubahan terkait dengan skripsi tersebut selama proses pembimbingan, harap segera melaporkan kepada Program Studi bersangkutan.

Demikian atas kesediaan Saudara, kami ucapkan terima kasih.

Wassalamu'alaikum Wr.Wb



Jakarta, 23 Maret 2022

a.n Dekan

Wakil Dekan Bid. Akademik



Dr. Ir. Siti Rochani, M.Si.

NIP. 196203081989032001

Cek Keaslian Surat di <https://e-letter.fst.uinjkt.ac.id/scan>