

**ANALISIS SENTIMEN PENGGUNA *TWITTER* MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* (SVM) PADA
KASUS KENAIKAN HARGA BBM**

RAHADI RAMLAN

NIM.H1091191037

SKRIPSI



**PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2023**

**ANALISIS SENTIMEN PENGGUNA *TWITTER* MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* (SVM) PADA
KASUS KENAIKAN HARGA BBM**

RAHADI RAMLAN

NIM. H1091191037

SKRIPSI

Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika



**PROGRAM STUDI STATISTIKA
JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK**

2023

ANALISIS SENTIMEN PENGGUNA *TWITTER* MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE (SVM)* PADA KASUS KENAIKAN HARGA BBM

Tanggung Jawab Yuridis Material Pada

Rahadi Ramlan
H1091191037

Disetujui Oleh

Pembimbing 1



Neva Satyahadewi, S.Si., M.Sc.,
CRA., CRP., CRMP.
NIP. 198212042005012001

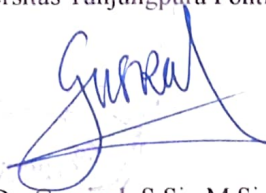
Pembimbing 2



Wirda Andani, M.Si
NIP. 199411152022032016

Disahkan
Oleh

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura Pontianak



Dr. Gusrizal, S.Si., M.Si.
NIP. 197108022000031001

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI
UNIVERSITAS TANJUNGPURA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
PONTIANAK**

TIM PENGUJI SKRIPSI

**ANALISIS SENTIMEN PENGGUNA *TWITTER* MENGGUNAKAN
METODE *SUPPORT VECTOR MACHINE* (SVM) PADA KASUS
KENAIKAN HARGA BBM**

**RAHADI RAMLAN
H1091191037**

NAMA/NIP	TIM PENGUJI	GOLONGAN/ JABATAN	TANDA TANGAN
Neva Satyahadewi, S.Si., M.Sc., CRA., CRP., CRMP. NIP. 198212042005012001	Pemimpin sidang merangkap anggota penguji	III/d Lektor	
Wirda Andani, M.Si. NIP. 199411152022032016	Sekretaris sidang merangkap anggota penguji	III/b Tenaga Pengajar	
Shantika Martha, M.Si. NIP. 198403082008122003	Ketua penguji	III/b Asisten Ahli	
Hendra Perdana, S.Si., M.Sc. NIP. 198810102019031020	Anggota penguji	III/b Tenaga Pengajar	

Berdasarkan Surat Keputusan

Dekan Fakultas Matematika dan Ilmu Pengatahuan Alam

Universitas Tanjungpura Pontianak

Nomor : 1996/UN22.8/TD.06/2023

Tanggal : 7 Juni 2023

Tanggal Lulus : 14 Juni 2023

HALAMAN PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi. Sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Pontianak, 14 Juni 2023



Rahadi Ramlan

ANALISIS SENTIMEN PENGGUNA *TWITTER* MENGGUNAKAN METODE *SUPPORT VECTOR MACHINE* (SVM) PADA KASUS KENAIKAN HARGA BBM

INTISARI

Twitter merupakan salah satu media sosial dengan pengguna aktif terbanyak yaitu sebesar 24 juta pengguna aktif. Informasi yang dipublikasikan pada *Twitter* dapat mengandung komentar dari pengguna terhadap suatu objek. Objek tersebut juga dapat berupa komentar positif atau negatif terhadap kejadian yang sedang banyak dibicarakan. Analisis sentimen digunakan untuk menentukan apakah data tersebut termasuk komentar negatif atau komentar positif karena komentar yang diambil pada *twitter* yaitu data tekstual. Metode yang bisa dipakai yaitu menggunakan metode *Support Vector Machine* (SVM) dalam analisis sentimen tentang komentar masyarakat terhadap kenaikan harga BBM pada *twitter*. Data komentar yang digunakan sebanyak 258 data *tweets* pada tanggal 4 September 2022 karena pada tanggal tersebut tepat sehari setelah kenaikan harga BBM. Sebelum melakukan analisis terlebih dahulu melakukan *preprocessing* untuk menghilangkan kata maupun informasi yang tidak diperlukan. Kemudian data dibagi menjadi data *training* dan data *testing* dengan perbandingan 80% dan 20%. Adapun hasil yang didapatkan dari 206 data *training* dan 52 data *testing* yaitu tingkat akurasi sebesar 82,69%, *sensitivity* sebesar 100%, dan *specificity* sebesar 79,07%. Kemudian dari hasil *testing* 52 data didapat hasil 43 komentar negatif dan 9 komentar positif sehingga dapat disimpulkan masyarakat lebih banyak yang tidak setuju dengan adanya kenaikan harga BBM.

Kata Kunci: *Machine learning, Twitter, Preprocessing, Sentimen*

SENTIMENT ANALYSIS OF TWITTER USERS USING THE SUPPORT VECTOR MACHINE (SVM) METHOD IN THE CASE OF A FUEL PRICE INCREASE

ABSTRACT

Twitter is one of the most active social media platforms, with a total of 24 million active users. Information shared on Twitter can include user comments on a particular subject. These comments can be either positive or negative, reflecting the ongoing discussions on various topics. Sentiment analysis is utilized to determine whether the data consists of negative or positive comments since the extracted comments from Twitter are in textual form. One method that can be employed is using Support Vector Machine (SVM) for sentiment analysis regarding public comments on the increase in fuel prices on Twitter. A total of 258 tweet data were used for analysis on September 4, 2022, which is one day after the fuel price hike. Prior to analysis, preprocessing was conducted to remove unnecessary words and information. Subsequently, the data was divided into training data and testing data, with an 80% and 20% ratio, respectively. The results obtained from 206 training data and 52 testing data showed an accuracy rate of 82.69%, a sensitivity rate of 100%, and a specificity rate of 79.07%. Furthermore, out of the 52 testing data, 43 were classified as negative comments and 9 as positive comments, indicating that the majority of people disagreed with the fuel price increase.

Keywords: Machine learning, Twitter, Preprocessing, Sentiment

PRAKATA

Puji dan syukur saya panjatkan kepada Allah Swt. atas ridhonya saya dapat menyelesaikan penyusunan skripsi ini. Adapun judul skripsi yang saya ajukan adalah “**Analisis Sentimen Pengguna Twitter Menggunakan Metode Support Vector Machine (SVM) Pada Kasus Kenaikan Harga BBM**”.

Skripsi ini diajukan untuk memenuhi syarat kelulusan mata kuliah Tugas Akhir di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura. Dalam penyelesaian tugas akhir ini tidak terlepas dari bimbingan, dukungan, motivasi serta doa dari berbagai pihak. Penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada

1. Allah *Subhanahu Wa Ta'ala* yang telah memberikan segala rahmat, hidayah, karunia, pertolongan, petunjuk, kemudahan dan kekuatan pada penyelesaian tugas akhir ini.
2. Kedua orang tua penulis, Umak Ramini & Ayah Asrul yang telah mendidik, memberi dukungan besar serta memotivasi penulis.
3. Saudara saya tercinta Nurfitiana dan Fikri Akmal yang selalu mendukung penulis saat senang maupun sedih.
4. Keluarga Besar penulis yang selalu memotivasi dalam kemauan untuk belajar dan menggapai cita-cita.
5. Ibu Neva Satyahadewi, M.Sc., CRA, CRP, CRMP selaku Dosen pembimbing pertama yang telah memberikan bimbingan dan arahan kepada penulis.
6. Ibu Wirda Andani, S.Si, M.Si selaku Dosen pembimbing kedua yang telah memberikan bimbingan dan arahan kepada penulis.
7. Ibu Shantika Martha, M.Si selaku dosen penguji Pertama yang telah memberikan masukan dan saran kepada penulis.
8. Bapak Hendra Perdana, M.Sc selaku dosen penguji Kedua yang telah memberikan masukan dan saran kepada penulis.
9. Teman-teman Statistika angkatan 2019 yang kebersamaan penulis selama masa perkuliahan.

10. Comdev & Outreaching serta Ditjen Belmawa Kemenristetdikti yang telah memberikan Beasiswa Bidikmisi.
11. Semua pihak yang telah membantu dan tidak dapat disebutkan satu persatu.

Semoga segala kebaikan dan pertolongan semuanya mendapat berkah dari Allah Swt. dan akhirnya saya menyadari bahwa skripsi ini masih jauh dari kata sempurna, karena keterbatasan ilmu yang saya miliki. Untuk itu saya dengan kerendahan hati mengharapkan saran dan kritik yang sifatnya membangun dari semua pihak demi membangun laporan penelitian ini.

Pontianak, Juni 2023

Rahadi Ramlan

DAFTAR ISI

INTISARI.....	i
ABSTRACT	ii
PRAKATA	iii
DAFTAR ISI	v
DAFTAR GAMBAR	vii
DAFTAR TABEL	viii
DAFTAR LAMPIRAN	ix
DAFTAR SIMBOL	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	4
1.5 Tinjauan Pustaka.....	4
1.6 Metodologi Penelitian.....	5
BAB II LANDASAN TEORI.....	8
2.1 <i>Twitter</i>	8
2.2 Analisis Sentimen	8
2.3 Text Mining	9
2.4 <i>Machine Learning</i>	9
2.5 Klasifikasi	10
2.6 Fungsi Kernel.....	10
2.7 Matriks <i>Hessian</i>	11
2.8 <i>Dot Product</i>	11
BAB III PENGAMBILAN DATA DENGAN WEB SCRAPING DAN PENGKLASIFIKASIAN MENGGUNAKAN SUPPORT VECTOR MACHINE	13
3.1 <i>Web Scrapping</i>	13
3.2 <i>Text Preprocessing</i>	13
3.3 Pembobotan Kata TF-IDF (<i>Term Frequency-Inverse Document Frequency</i>)	14
3.4 <i>Support Vector Machine</i>	15
3.5 Evaluasi Hasil Klasifikasi	18
BAB IV HASIL DAN PEMBAHASAN	20

4.1 Sumber Data	20
4.2 <i>Data Selection</i>	20
4.3 <i>Preprocessing Data</i>	21
4.3.1 <i>Cleaning Data</i>	21
4.3.2 <i>Case Folding</i>	22
4.3.3 <i>Tokenizing</i>	22
4.3.4 <i>Normalization</i>	23
4.3.5 <i>Filtering</i>	24
4.4 Pembobotan Kata TF-IDF	25
4.5 Klasifikasi <i>Support Vector Machine</i> (SVM).....	27
4.6 Evaluasi Hasil Klasifikasi	31
BAB V KESIMPULAN DAN SARAN	33
5.1 Kesimpulan	33
5.2 Saran	33
DAFTAR PUSTAKA	34
LAMPIRAN	37

DAFTAR GAMBAR

Gambar 1. 1 <i>Flowchart Support Vector Machine (SVM)</i>	7
Gambar 3. 1 <i>Hyperplane SVM</i>	16

DAFTAR TABEL

Tabel 3. 1 <i>Confusion Matrix</i>	19
Tabel 4. 1 Sampel Data <i>Twitter</i>	20
Tabel 4. 2 Proses <i>Cleaning Data</i>	21
Tabel 4. 3 Proses <i>Case Folding</i>	22
Tabel 4. 4 Proses <i>Tokenizing</i>	23
Tabel 4. 5 Proses <i>Normalization</i>	24
Tabel 4. 6 Proses <i>Filtering</i>	25
Tabel 4. 7 Data contoh perhitungan <i>TF-IDF</i>	25
Tabel 4. 8 Perhitungan <i>TF-IDF</i>	26
Tabel 4. 9 Fungsi Kernel	27
Tabel 4. 10 Hasil Perhitungan Nilai Fungsi Kernel	28
Tabel 4. 11 Nilai Matriks <i>Hessian</i>	28
Tabel 4. 12 Hasil Perhitungan Nilai <i>Error</i>	29
Tabel 4. 13 Hasil Perhitungan Nilai <i>Delta Alpha</i>	29
Tabel 4. 14 Hasil Perhitungan Nilai <i>Alpha Baru</i>	30
Tabel 4. 15 <i>TF-IDF</i> Sampel Data Uji.....	30
Tabel 4. 16 Hasil Perhitungan <i>Dot Product</i> Data Uji Dan Data Latih.....	31
Tabel 4. 17 Hasil <i>Confusion Matrix</i>	32

DAFTAR LAMPIRAN

Lampiran 1. Syntax <i>web scrapping</i>	37
Lampiran 2. Syntax <i>Preprocessing</i> Data.....	37
Lampiran 3. Syntax TF-IDF	41
Lampiran 4. Syntax Klasifikasi SVM	43

DAFTAR SIMBOL

W	=	Bobot <i>TF-IDF</i> kata pada dokumen tertentu
idf	=	Bobot <i>inverse document frequency</i> kata
tf	=	Jumlah frekuensi kata pada dokumen tertentu
N	=	Jumlah dokumen
df	=	Jumlah frekuensi kata pada seluruh dokumen
w	=	Bobot <i>vector</i>
x	=	Data ke
b	=	Nilai Bias
D	=	Matriks <i>Hessian</i>
K	=	Fungsi Kernel
E	=	Nilai <i>error</i>
δa	=	Nilai <i>delta alpha</i>

BAB I

PENDAHULUAN

1.1 Latar Belakang

Bahan Bakar Minyak (BBM) merupakan komoditas yang memegang peranan penting dalam kegiatan ekonomi suatu negara. BBM memiliki peranan sebagai bahan pendukung proses produksi bagi industri. Kenaikan harga BBM menyebabkan biaya produksi semakin meningkat, sehingga keuntungan atau laba perusahaan akan mengalami penurunan. Oleh karena itu, kenaikan harga BBM tidak hanya berdampak pada perusahaan pada khususnya, tetapi juga pada masyarakat kecil pada umumnya (Ardani, 2018).

Berdasarkan informasi yang diperoleh dari (Sihombing, 2022), BBM mengalami kenaikan pada tanggal 3 September 2022 jam 14.00. Perubahan harga BBM tersebut meliputi harga Pertalite dari Rp7.650/liter menjadi Rp10.000/liter, harga Solar subsidi dari Rp5.150/liter menjadi Rp6.800/liter, dan harga Pertamina dari Rp12.500/liter menjadi Rp14.500/liter.

Kenaikan harga BBM menimbulkan reaksi yang berbeda di kalangan masyarakat. Kenaikan harga BBM menyebabkan ketidakpuasan terutama pada kelompok menengah ke bawah. Dampak dari kenaikan harga BBM yang dirasakan masyarakat termasuk meningkatnya biaya rumah tangga, transportasi, dan bahan pokok. Kenaikan harga BBM ini dikhawatirkan akan memicu inflasi yang lebih tinggi dan berdampak pada harga kebutuhan pokok masyarakat. Selain itu, daya beli masyarakat di pasar domestik juga dapat melemah karena pendapatan mereka tidak meningkat kecuali bagi PNS. Dari segi sosial, kenaikan harga BBM juga dapat berdampak pada kemiskinan yang semakin tinggi (Sandy, Manongga, & Iriani, 2015)

Kenaikan harga Bahan Bakar Minyak yang terjadi di Indonesia diakibatkan karena pemerintah berusaha menyeimbangkan harga minyak dunia yang semakin tinggi dan membengkaknya anggaran subsidi terhadap BBM. Kenaikan BBM disebabkan oleh meningkatnya harga minyak mentah di pasar perdagangan dunia

sampai level harga \$94/barel di tanggal 12 September 2022. Untuk tahun 2022, pemerintah telah menetapkan subsidi BBM sebesar 502 triliun rupiah (Sihombing, 2022). Kenaikan harga BBM selalu mendapat tanggapan pro dan kontra dari masyarakat. Tanggapan tersebut lebih banyak dilakukan pada media sosial, salah satu media yang digunakan adalah *Twitter*.

Twitter yang memiliki jumlah pengguna aktif sebanyak 24 juta orang pada awal Maret 2023 dan pada tahun 2022 sebesar 98,02% pengguna internet di Indonesia menyebutkan bahwa alasan utama mereka menggunakan adalah untuk media sosial (Jemadu, 2022). *Twitter* merupakan salah satu media sosial yang digunakan untuk memperoleh informasi tentang bisnis, hiburan, ekonomi, politik, dan topik lainnya. Seiring dengan tingginya frekuensi publikasi *tweet* melalui *twitter*, pengguna juga dapat menyampaikan komentar atau pandangan mereka terhadap suatu objek, seperti produk atau layanan yang terkait dengan kejadian yang terjadi di sekitar masyarakat (Tineges, Triayudi, & Sholihati, 2020).

Pada *twitter* memiliki banyak data komentar, jadi menyalin data secara manual akan membutuhkan waktu yang lama. Salah satu teknik yang digunakan untuk mengambil data atau informasi berupa data tekstual dari *website* adalah penggunaan *web scraping*. *Web Scraping* akan melakukan ekstraksi data tertentu saja dari *website* yang dituju. Data hasil *Web Scraping* tersebut akan digunakan untuk analisis. Analisis hasil dari *web scrapping* biasa digunakan dalam klasifikasi. Klasifikasi merupakan proses menemukan sebuah model atau fungsi yang menjelaskan dan membedakan data menjadi kelompok-kelompok tertentu. Dalam proses ini, dilakukan pengecekan terhadap karakteristik objek dan kemudian menempatkannya ke dalam salah satu kelompok yang telah ditetapkan sebelumnya. Dalam klasifikasi salah satu analisis yang umum dipakai dalam analisis data teks yaitu *sentiment analysis* (Yani, Pratiwi, & Muhandi, 2019).

Dalam pengklasifikasian pada *sentiment analysis* ada beberapa metode yang bisa dipakai, yaitu *Naïve Bayes Classifier* (NBC), *Maximum Entropy*, dan *Support Vector Machine* (SVM). Metode NBC merupakan metode yang membuat asumsi yang sangat kuat (naif) akan independensi dari masing-masing kelas kejadian yang diberikan label. Metode *Maximum Entropy* merupakan metode yang berbasiskan

probabilitas yang termasuk dalam kelas model eksponensial. Prinsip dari *maximum entropy* mencari distribusi $p(a|b)$ yang akan memberikan nilai *entropy maksimum*. Metode SVM menggunakan *machine learning (supervised learning)* yang memprediksi kelas berdasarkan model atau pola dari hasil proses *training*. Nilai atau pola yang dihasilkan dari metode SVM sebenarnya adalah sebuah garis pemisah yang disebut dengan *hyperplane* (Cindo, Rini, & Ermatita, 2019).

Kemampuannya untuk mengolah data berdimensi besar menjadi keunggulan metode SVM dibanding dengan *classifier* lain. Pada data teks, keunggulan metode SVM untuk mengolah data berdimensi besar dapat dimanfaatkan, karena sifat data teks yang biasanya berdimensi besar (Purnamawan, 2015).

Pada penelitian ini menggunakan metode SVM dengan data hasil *tweet* di *twitter* dengan kata kunci BBM Naik. Data yang digunakan sebagai data latih dan data uji. Tujuan dari penelitian ini untuk menentukan tingkat akurasi metode SVM dalam mengklasifikasikan *tweet* dan menentukan tingkat kepuasan masyarakat terhadap kenaikan harga BBM berdasarkan media sosial *twitter*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang dikemukakan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana menentukan tingkat akurasi metode SVM dalam mengklasifikasikan *tweet*.
2. Bagaimana menentukan tingkat kepuasan masyarakat terhadap kenaikan harga BBM berdasarkan media sosial *twitter*.

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dikemukakan, tujuan penelitian dalam penelitian ini adalah:

1. Menentukan tingkat akurasi metode SVM dalam mengklasifikasikan *tweet*.
2. Menentukan tingkat kepuasan masyarakat terhadap kenaikan harga BBM berdasarkan media sosial *twitter*.

1.4 Batasan Masalah

Agar penelitian ini lebih terarah, diperlukan batasan masalah. Batasan masalah dalam penelitian ini antara lain:

1. *Tweet* dari Media Sosial *Twitter* dengan kata kunci BBM Naik yang diambil pada tanggal 4 September 2022 dengan *tweet* yang digunakan adalah *tweet* berbahasa Indonesia dan jumlah *tweet* sebanyak 1001 *tweet*.
2. Metode yang digunakan adalah SVM dengan teknik *Web Scrapping*. Analisis pada penelitian ini menggunakan bantuan *software Python* dengan web *Jupyter Notebook*. Penelitian ini menggunakan salah satu fungsi kernel yaitu kernel linear.
3. Penentuan yang termasuk ke dalam komentar atau bukan komentar dan pemisahan antara komentar positif dan komentar negatif itu dilakukan subjektif dari peneliti. Komentar positif yaitu pendapat dari masyarakat di *twitter* yang secara keseluruhan komentar tersebut mendukung atau setuju terhadap kenaikan BBM, begitupula sebaliknya komentar negatif yaitu pendapat dari masyarakat di *twitter* yang secara keseluruhan komentar tersebut menolak atau tidak setuju terhadap kenaikan BBM.
4. Pemberian nilai awal (inisialisasi) terhadap parameter yang diaplikasikan yaitu λ (*Lambda*) bernilai 0,5 ; α_1 (*Alpha*) bernilai 0,5 ; γ (*Learning rate*) bernilai 0,5 ; C (*variable slack*) bernilai 1 ; ε (*Epsilon*) bernilai 0,001.

1.5 Tinjauan Pustaka

(Rofiqoh, Perdana, & Fauzi, 2017) melakukan penelitian untuk mencari tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler Indonesia pada *twitter*. Penelitian tersebut menerapkan metode Metode *Support Vector Machine* dan *Lexicon Based Features*. Hasil dari penelitian tersebut adalah tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler pada *Twitter* dengan metode *Support Vector Machine* dan *Lexicon Based Features* menghasilkan nilai

accuracy sebesar 79%, *precision* sebesar 65%, *recall* sebesar 97%, dan *f-measure* sebesar 78%. Nilai *recall* yang sangat besar tersebut dipengaruhi oleh banyaknya data uji yang sebenarnya positif dan terdeteksi sebagai positif oleh sistem. Hal ini membuktikan bahwa analisis sentimen dengan metode *Support Vector Machine* dan *Lexicon Based Features* baik digunakan untuk klasifikasi analisis sentimen.

(Tineges, Triayudi, & Sholihati, 2020) menerapkan metode klasifikasi SVM dalam menentukan klasifikasi sentimen terhadap layanan indihome berdasarkan *twitter*. Digunakan metode klasifikasi SVM dalam menentukan klasifikasi sentimen terhadap layanan indihome berdasarkan *twitter* dengan menggunakan data *tweet* yang diperoleh dari *twitter*. Hasil evaluasi menggunakan *Confussion Matrix*, didapat akurasi sebesar 87% dengan ketepatan antara hasil prediksi dengan data sebenarnya (*precision*) sebesar 86%, tingkat keberhasilan sistem dalam memprediksi sebuah data (*recall*) sebesar 95%, tingkat kesalahan semua data yang diprediksi (*error rate*) sebesar 13%, sedangkan untuk nilai perbandingan rata-rata *precision* dan *recall* (*f1-score*) adalah sebesar 90%.

(Darwis, Pratiwi, & Pasaribu, 2020) melakukan penelitian tentang analisis sentimen pada data *twitter* terhadap Komisi Pemberantasan Korupsi Republik Indonesia menggunakan metode SVM. Dari penelitian tersebut diperoleh hasil pengujian dan evaluasi yang dilakukan pada nilai akurasi, *precession*, *recall*, dan *F1-Score*, dapat disimpulkan bahwa sentimen masyarakat pengguna *twitter* mengenai kinerja KPK dengan persentase sangat kurang baik yaitu ditunjukan dengan munculnya kecondongan sentimen negatif sebesar 77% dengan keakuratan hasil pengujian akurasi sebesar 82% dan pengujian *precision* sebesar 90%, serta *recall* sebesar 88% dan *f1-score* sebesar 89%.

1.6 Metodologi Penelitian

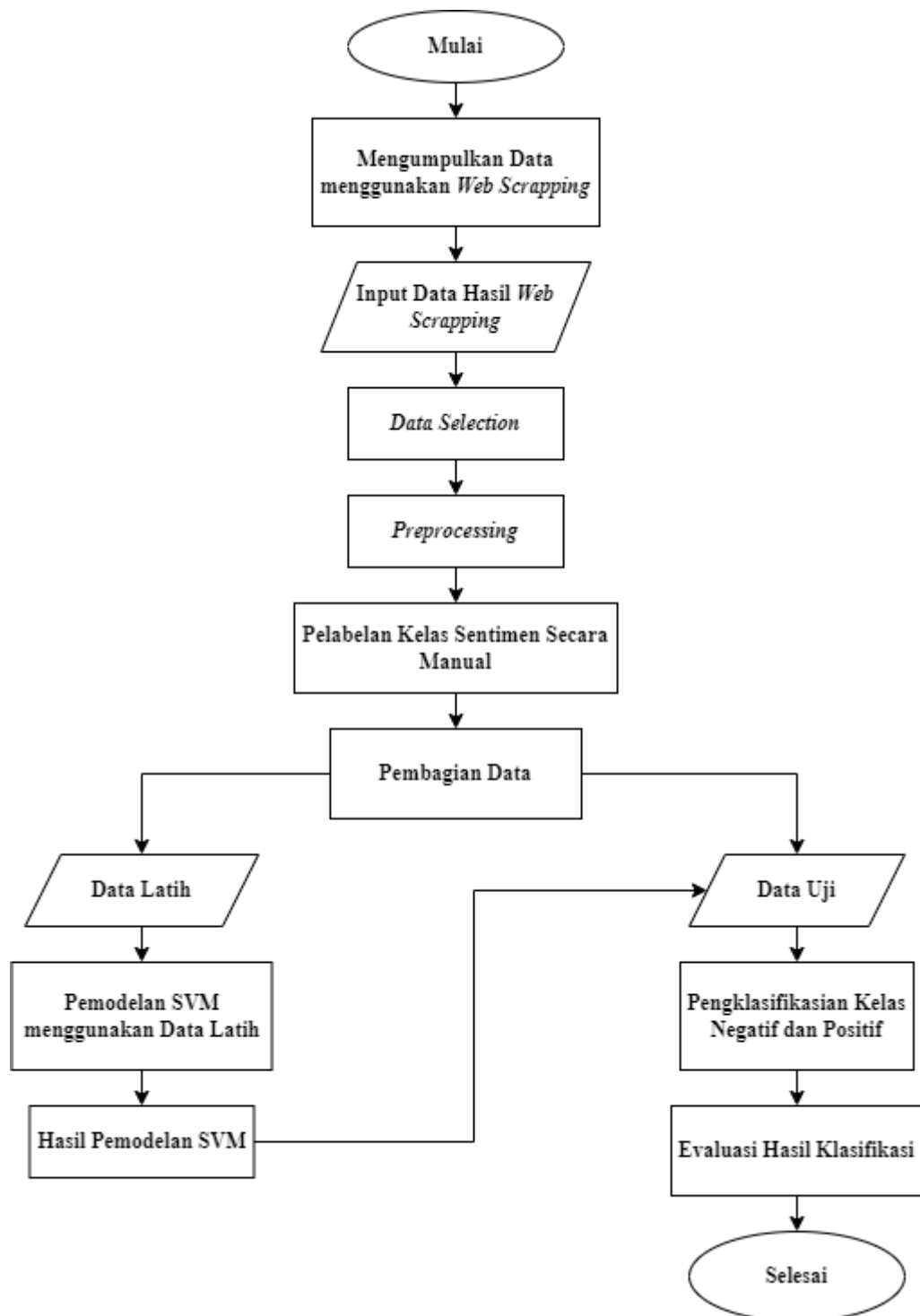
Penelitian ini termasuk kedalam jenis penelitian studi literatur dimana dalam penulisan penelitian ini dilakukan dengan mencari referensi teori yang relevan dengan permasalahan yang akan diteliti. Data yang digunakan dalam penelitian ini adalah data sekunder yaitu data yang diperoleh dari media sosial *twitter* pada

tanggal 4 September 2022. Adapun pengambilan data ini menggunakan Teknik *Web Scrapping* dengan memakai kata kunci “BBM Naik”.

Langkah-langkah dalam melakukan analisis *Support Vector Machine* adalah sebagai berikut:

1. Pengambilan data *tweet* dengan menggunakan teknik *web scrapping* dengan menggunakan kata kunci “BBM Naik”.
2. Data *selection*, melakukan seleksi terhadap komentar yang termasuk ke dalam komentar dan bukan komentar. Kemudian dari hasil seleksi dilakukan pemisahan antara komentar positif dan komentar negatif.
3. *Text Preprocessing* bertujuan untuk memperbaiki informasi dari semua data ke dalam bentuk yang lebih baku. Tahapan *preprocessing* terdiri dari *cleaning data*, *case folding*, *normalization*, *tokenizing*, *filtering*.
4. Pembobotan kata dengan TF-IDF.
5. Pelabelan kelas sentimen secara manual untuk didapat kelas positif atau negatif.
6. Membagi data *training* dan data *testing* dengan perbandingan 80% data *training* dan 20% data *testing*.
7. Data latih (*training*) sebanyak n akan digunakan untuk proses pelatihan menggunakan Metode SVM dengan menghitung nilai optimal *hyperplane* untuk mendapatkan parameter b dan α nilai yang akan digunakan dalam proses pengujian.
8. Menampilkan hasil akhir klasifikasi. Output yang dihasilkan proses pengujian terhadap suatu data uji berupa hasil klasifikasi data uji apakah data tersebut masuk ke dalam kelas negatif atau positif.
9. Menentukan evaluasi hasil klasifikasi.

Alur dari penelitian ini dapat dilihat dari *flowchart* berikut:



Gambar 1. 1 Flowchart Support Vector Machine (SVM)