ANALISIS SENTIMEN PADA TWEET DENGAN TAGAR #BPJSRASARENTENIR MENGGUNAKAN METODE SUPPORT VECTORE MACHINE (SVM)

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat
Penyusunan Laporan Skripsi
Pada Fakultas Teknik
Universitas Islam Riau Pekanbaru

Oleh:

GADING TEGUH SANTOSO
143510514

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS ISLAM RIAU
PEKANBARU
2021

LEMBAR PENGESAHAN PEMBIMBING SKRIPSI

Nama	:	Gading Teguh Santoso
NPM	:	143510514
Fakultas	:	Teknik
Program Studi	:	Teknik Informatika
Jenjang Pendidikan	:	Strata Satu (S1)
Judul Skripsi	:	Analisis Sentiment Pada Tweet Dengan Tagar #BPJSRasarentenir Menggunakan Metode Support Vectore Machine (SVM)

Format sistematika dan pembahasan materi pada masing-masing bab dan sub bab dalam skripsi ini telah dipelajari dan dinilai relatif telah memenuhi ketentuan-ketentuan dan kriteria - kriteria dalam metode penulisan ilmiah. Oleh karena itu, skripsi ini dinilai layak dapat disetujui untuk disidangkan dalam ujian komprehensif.

Pekanbaru, 19 Maret 2021

Disahkan Oleh

Ketua Prodi Teknik Informatika	**Dosen Pembimbing**
Dr. ARBI HAZA NASUTION, B.IT(Hons)., M.IT	**Dr. ARBI HAZA NASUTION, B.IT(Hons)., M.IT**

LEMBAR PENGESAHAN
TIM PENGUJI UJIAN SKRIPSI

Nama	:	Gading Teguh Santoso
NPM	:	143510514
Fakultas	:	Teknik
Program Studi	:	Teknik Informatika
Jenjang Pendidikan	:	Strata Satu (S1)
Judul Skripsi	:	Analisis Sentiment Pada Tweet Dengan Tagar #BPJSRasarentenir Menggunakan Metode Support Vectore Machine (SVM)

Skripsi ini secara keseluruhan dinilai telah memenuhi ketentuan-ketentuan dan kaidah-kaidah dalam penulisan penelitian ilmiah serta telah diuji dan dapat dipertahankan dihadapan tim penguji. Oleh karena itu, Tim Penguji Ujian Skripsi Fakultas Teknik Universitas Islam Riau menyatakan bahwa mahasiswa yang bersangkutan dinyatakan **Telah Lulus Mengikuti Ujian Komprehensif Pada Tanggal 19 Maret 2021** dan disetujui serta diterima untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Strata Satu Bidang Ilmu **Teknik Informatika.**

Pekanbaru, 19 Maret 2021

Tim Penguji

1. **Ir.Des Suryani, M.Sc** — **Sebagai Tim Penguji I** (.....)
2. **Ause Labellapansa, ST., M.Cs., M.Kom** — **Sebagai Tim Penguji II** (.....)

Disahkan Oleh

Ketua Prodi Teknik Informatika	**Dosen Pembimbing**
Dr. ARBI HAZA NASUTION, B.IT(Hons)., M.IT	**Dr. ARBI HAZA NASUTION, B.IT(Hons)., M.IT**

LEMBAR PERNYATAAN BEBAS PLAGIARISME

Saya yang bertanda tangan dibawah ini:

Nama	: Gading Teguh Santoso
Tempat/Tgl Lahir	: Salimpaung, 14 April 1996
Alamat	: Perum Puri Air Dingin, Kec. Bukit Raya.

Adalah mahasiswa Universitas Islam Riau yang terdaftar pada:

Fakultas	: Teknik
Jurusan	: Teknik Informatika
Program Studi	: Teknik Informatika
Jenjang Pendidikan	: Strata-1 (S1)

Dengan ini menyatakan dengan sesungguhnya bahwa skripsi yang saya tulis adalah benar dan asli hasil dari penelitian yang telah saya lakukan dengan judul **"Analisis Sentiment Pada Tweet Dengan Tagar #BPJSRasarentenir Menggunakan Metode Support Vectore Machine (SVM)".**

Apabila dikemudian hari ada yang merasa dirugikan dan atau menuntut karena penelitian ini menggunakan sebagian hasil tulisan atau karya orang lain tanpa mencantumkan nama penulis yang bersangkutan, atau terbukti karya ilmiah ini **bukan** karya saya sendiri atau **plagiat** hasil karya orang lain, maka saya bersedia menerima sanksi sesuai dengan peraturan perundangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya untuk dapat digunakan sebagaimana mestinya.

Pekanbaru, 25 Maret 2021

Yang membuat pernyataan,

Gading Teguh Santoso

LEMBAR IDENTITAS PENULIS

Nama	: Gading Teguh Santoso
NPM	: 143510514
Tempat/Tanggal Lahir	: Salimpaung, 14 April 1996
Alamat Orang Tua	: Perawang
Nama Orang Tua	
Nama Ayah	: Nasirwan
Nama Ibu	: Harpineli
No.HP/Telp	: 082284896716
Fakultas	: Teknik
Program Studi	: Teknik Informatika
Masuk Th.Ajaran	: 2014
Keluar Th. Ajaran	: 2021
Judul Penelitian	: Analisis Sentiment Pada Tweet Dengan Tagar #BPJSRasarentenir Menggunakan Metode Support Vectore Machine (SVM)

Pekanbaru, 25 Maret 2021

Gading Teguh Santoso

HALAMAN
PERSEMBAHAN

Assalamu'alaikum Warahmatullahi Wabarakatu..

Alhamdulillah puji syukur penulis ucapkan kepada Allah SWT atas segala rahmat dan karunia-Nya yang telah diberikan kepada penulis sehingga dapat menyelesaikan tugas akhir dengan judul **"Analisis Sentiment Pada Tweet Dengan Tagar #BPJSRasarentenir Menggunakan Metode Support Vectore Machine (SVM)".**

Skripsi ini disusun untuk memenuhi persyaratan mencapai derajat strata-1 (S1) di program studi Teknik Informatika Fakultas Teknik Universitas Islam Riau. Penulis menyadari bahwa tanpa bantuan dari pihak-pihak lain, usaha yang penulis lakukan dalam menyelesaikan skripsi ini tidak akan membuahkan hasil yang berarti. Dalam kesempatan ini penulis ucapkan terima kasih kepada :

1. Allah SWT, karena hanya dengan izin dan karunia-Nya maka skripsi ini dapat selesai tepat pada waktunya. Segala puji bagi Allah yang maha mengabulkan segala doa.
2. Terkhusus orang tua tercinta yakni ayahanda dan ibunda tercinta beserta keluarga besar yang tak henti-hentinya selalau mensuport penulis dan membantu dalam segi materi dan moril serta do'a-do'anya sehingga penulis dapat menyelesaikan penulisan skripsi ini.
3. Bapak Arbi Haza Nasution, B.IT (Hons), M.IT selaku Dosen

Pembimbing atas bimbingan serta support dan motivasi yang diberikan.

4. Segenap Dosen Teknik Informatika, Universitas Islam Riau yang telah memberikan ilmu, pendidikan, dan pengetahuan kepada penulis selama duduk dibangku kuliah.
5. My Suport Farahdiva Assyfa Andrin dan Team Sentiment Analisis Univerisitas Islam Riau yang selalu memberikan semangat dan motivasi.

Akhir kata penulis mohon maaf atas kekeliruan dan kesalahan yang terdapat dalam skripsi ini dan berharap semoga skripsi ini dapat memberikan manfaat bagi pembaca.

Pekanbaru, 25 Maret 2021

Gading Teguh Santoso

KATA PENGANTAR

Alhamdulillahi Robbil 'Alamin. Segala puji syukur ke hadirat Allah SWT yang telah memberikan pertolongan pada hambanya yang lemah ini dalam setiap keadaan, termasuk dalam penelitian dan penulisan skripsi ini. Sholawat dan salam teruntuk Baginda Agung Nabi Muhammad SAW Sang *Insan Kamil* yang paling sempurna fisik dan akhlaknya yang sungguh kita rindukan perjumpaan dengan beliau di dunia terlebih di akhirat kelak.

Pelaksanaan penelitian dan penyusunan skripsi ini merupakan salah satu syarat untuk memperoleh gelar sarjana Teknik Informatika di Universitas Islam Riau.

Selanjutnya penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Ir. H. Abdul Kudus Zaini, MT selaku Dekan Fakultas Teknik dan selaku penasehat akademis.
2. Bapak Dr.Arbi Haza Nasution.,B.IT.,M.I.T selaku ketua Program Studi Teknik Informatika dan selaku dosen pembimbing yang telah senantiasa meluangkan waktu untuk memberikan arahan.
3. Bapak dan Ibu Dosen Teknik Informatika yang telah banyak memberikan ilmunya.
4. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan dukungan dan semangat sehingga penulis dapat menyelesaikan laporan skripsi ini.

Akhirnya, tiada gading yang tak retak. Penulis menyadari masih banyak kekurangan dan kelemahan dalam pelaksanaan dan penyusunan skripsi ini. Semoga penelitian ini dapat menjadi pengalaman yang berharga bagi penulis dalam mempersiapkan diri menghadapi persaingan di dunia kerja dan bermanfaat untuk masyarakat yang lebih luas.

Pekanbaru, Maret 2021

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ANALISIS SENTIMEN PADA TWEET DENGAN TAGAR #BPJSRASARENTENIR MENGGUNAKAN METODE SUPPORT VECTORE MACHINE (SVM)

ABSTRAK

Jejaring social membantu pengguna internet dalam berkomunikasi. Hal ini dikarenakan pengguna jejaring sosial dapat menyampaikan pesan dengan memanfaatkan fasilitas yang disiapkan oleh setiap media social.pesan-pesan para pengguna media sosial dapat dimanfaatkan dalam berbagai hal, seperti riview terhadap suatu produk atau riview terhadap suatu masalah pada politik atau masalah sosial sekarang ini. Hal ini bisa dilakukan dengan menganalisis sentiment para pengguna sosial media. Metode support vectore machine merupakan salah satu metode yang dapat digunakan untuk menganalisis sentiment. Dalam sentiment analisis mengunakan metode support vectore machine dilakukan dengan cara mengklasifikasikan sentiment kedalam kelas positif atau kelas negatif. Tingkat akurasi analisis sentiment terhadap #BPJSrasarentenir menggunakan metode support vectore machine adalah 94% dengan menggunakan 200 data kicauan.

Kata Kunci : Tweet, Analisis Sentiment, Support Vectore Machine (SVM)

SENTIMENT ANALYSIS ON TWEET WITH FARING #BPJSRASARENTENIR USING THE SUPPORT VECTORE MACHINE (SVM) METHOD

ABSTRACT

Social networking helps internet users communicate. This is because social network users can convey messages by utilizing the facilities prepared by each social media. Social media users' messages can be used in various ways, such as a review of a product or a review of a problem in politics or current social problems. This can be done by analyzing the sentiments of social media users. The support vectore machine method is one method that can be used to analyze sentiment. In sentiment analysis using the support vectore machine method, it is done by classifying sentiment into positive or negative classes. The accuracy rate of sentiment analysis for #BPJSrasarentenir using the support vectore machine method is 94% using 200 tweet data.

Keyword: Analisis Sentiment, Tweet, Support Vectore Machine (SVM)

BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut laporan terbaru We Are Social ditahun 2020, dikatakan bahwa ada sekitar 175 juta pengguna internet di Indonesia. Dari jumlah itu, sebanyak 160 juta adalah user yang menggunakan internet untuk mengakses media sosial. Beberapa media sosial yang banyak digunakan adalah berupa Facebook, Instagram, Youtube,Wechat,Twitter dan lain-lain.

Media Social Twitter memiliki jumlah sekitar 19,5 juta user di Indonesia dari 500 juta user global (Kominfo,Indonesia). Jumlah yang cukup besar tersebut menimbulkan banyak ciutan dari para penggunanya tentang berbagai hal seperti: pendidikan, hiburan, pekerjaan, dan termasuk juga politik. Salah satu isu yang menjadi *trending topic* di Twitter pada tahun 2019 adalah tentang kenaikan iyuran BPJS.

BPJS adalah penyelenggara rencana jaminan sosial bidang kesehatan yang merupakan salah satu dari lima rencana Sistem Jaminan Sosial Nasional (SJSN), yaitu jaminan kesehatan, perlindungan kecelakaan kerja, perlindungan hari tua, perlindungan pensiun dan perlindungan kematian. UU No. 40 tahun 2004 berhubungan dengan sistem jaminan sosial nasional. Dengan bertambahnya donasi di tahun 2019, peningkatan ini memiliki banyak pro dan kontra, sehingga banyak netizen yang membuat tagar peningkatan bpjs untuk mengetahui apa yang terjadi.

Sentimen Analisis atau yang biasa disebut *opinion mining* adalah jenis *natural language*, yaitu pengolahan kata untuk melacak *mood* masyarakat suatu produk atau topik tertentu. Analisis sentimen, disebut opinion mining. (G.Vinodhini, M.Chandrasekaran 2012).

Penulis pada penelitian ini akan melakukan analisis sentimen para pengguna Twitter terhadap kenaikan iyuran pada BPJS tahun 2019. Dengan input berupa data *tweet* dalam Bahasa Indonesia, akan dilakukan klasifikasi dengan algoritma *SVM (SUPPORT VECTOR MACHINE)* untuk menentukan apakah *tweet* tersebut bersentimen positif atau negatif.

1.2 Identifikasi Masalah

Adapun identifikasi masalah yang bisa didapat dari latar belakang tersebut adalah "Dibutuhkannya analisis sentimen untuk menganalisa opini masyarakat mengenai topik yang sedang populer dibicarakan di media sosial twitter".

1.3 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan permasalahan yang akan diselesaikan dalam penelitian ini adalah bagaimana mengamati dan menganalisis opini pengguna Twitter mengenai kenaikan iyuran BPJS pada tahun 2019 dan kemudian mengklasifikasikan sentimen data *tweets* tersebut.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah tweet berbahasa Indonesia.

2. Jumlah data *tweets* yang digunakan adalah 300 data yang telah difilter dan diambil dari bulan Agustus 2019 sampai dengan Oktober 2019 dengan skala 100 data pertiga hari.
3. Metode yang digunakan untuk pengklasifikasian dalam penelitian ini adalah metode *SUPPORT VECTOR MACHINE (SVM).*
4. Hasil dari pengklasifikasian sentimen pada penelitian ini menggunakan metode SVM berupa sentimen positif dan negatif.
5. Menampilkan sentimen positif dan negatif berbentuk tabel.

1.5 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengklasifikasikan nilai sentimen dan menentukan kelas data ciutan (*tweets*) dari pengguna Twitter yang berkaitan dengan topik kenaikan iyuran BPJS menggunakan algoritma *SVM (SUPPORT VECTOR MACHINE).*

1.6 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik kepada penulis maupun pembaca tentang gambaran sentimen para pengguna Twitter terhadap kenaikan iyuran BPJS pada tahun 2019.

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Tinjauan Pustaka

Penelitian yang dilakukan sebelum menjadi rujukan yaitu penelitian oleh Taufik dan S.A. Pamungkas (2018) dimana penelitian yang berjudul "Analisis Sentimen Terhadap Tokoh Publik Menggunakan Algoritma *Support Vector Machine (SVM*)". Tujuan penelitian ini untuk mengimplementasikan Algoritma SVM dan menganalisa sentiment pada data tweet tentang tokoh public dengan keyword "Ahok" dan "@teman_ahok" agar dapat menentukan apakah tweet tersebut bernilai positif atau negatif.

Penelitian kedua dilakukan oleh Abdan Syakuro (2017) penelitian yang berjudul "Analisis Sentimen Masyarakat Terhadap *E-COMMERCE* Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier (NBC) Dengan Seleksi Fitur Information Gain (IG)". Penelitian ini bertujuan untuk membuktikan bahwa metode *Naïve Bayes Classifier* dengan seleksi fitur information Gain dapat digunakan dalam pengklasifikasian Analisis Sentimen E-Commerce.

Penelitian ketiga dilakukan oleh Taufik Kurniawan (2017) dengan judul "Implentasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Media Mainstream Menggunakan *Niaves Bayes Classifier* Dan *Support Vectore Machine (SVM)".* Penelitian ini bertujuan untuk mendapatkan hasil ketepatan klasifikasi sentiment pengguna terhadap media televisi dan menetukan sentimen masyarakat apakah itu positif atau megatif.

Terakhir penelitian yang dilakukan oleh (Varian Habbie Bukhari, 2015) dengan judul *Sentiment Analysis* Menggunakan *K-Nearest Neighbor* Dengan Perbandingan Fungsi Jarak (Studi Kasus : Twitter Indosat Dan Telkomsel). Tujuan dari penelitian ini adalah melakukan pengklasifikasian data uji dengan menggunakan metode *K-Nearest Neighbor* dan menguji serta membandingkan akurasi metode *K-Nearest Neighbor* dengan dua metode perhitungan jarak : *Minkowski Distance* dan *City Block Distance*. Dalam penelitian ini, dihasilkan kesimpulan bahwa metode pengukuran jarak dengan *Minkowski Distance* memiliki akurasi yang lebih baik dari pada *City Block Distance*.

2.2 Dasar Teori

2.2.1 Sentiment Analysis

Analisis sentimen atau biasa disebut dengan opinion mining merupakan cabang dari data mining yang bertujuan untuk menganalisis, memahami, mengolah, dan mengekstrak data tekstual berupa opini tentang entitas seperti produk, jasa, organisasi, orang, dan topik tertentu. Analisis ini digunakan untuk mengekstrak beberapa informasi dari dataset yang ada.

Menurut (Bo Pang dan Lillian, 2008), analisis sentimen merupakan cabang penelitian dalam domain Text Mining. Secara umum, analisis sentimen berkaitan dengan studi komputasi opini, perasaan, dan emosi yang diungkapkan secara tekstual. Analisis sentimen bertujuan untuk mengekstrak atribut dari suatu komentar (opini, perasaan, dan emosi) yang diungkapkan secara tekstual. Setelah itu akan dilakukan evaluasi terhadap komentar tersebut, yaitu positif atau negatif.

Menurut (Liu, 2012), analisis sentimen adalah bidang ilmu yang menganalisis pendapat, sikap, evaluasi dan evaluasi peristiwa, topik, organisasi dan orang.

2.2.2 Twitter

Twitter adalah situs mikroblog media sosial yang memungkinkan pengguna memposting tweet sepanjang 140 karakter yang disebut tweet. Twitter dioperasikan oleh Twitter, Inc., yang didirikan pada Maret 2006 oleh Jack Dorsey. Twitter berkembang pesat dan dengan cepat mendapatkan popularitas di seluruh dunia, per Januari 2013. Ada lebih dari 500 juta pengguna terdaftar, di mana 200 juta di antaranya adalah pengguna aktif.Lonjakan penggunaan Twitter biasanya terjadi di acara populer. Diawal 2013, pengguna Twitter telah mengirim lebih dari 340 juta tweet setiap hari, dan Twitter telah menangani lebih dari 1,6 miliar permintaan pencarian setiap hari. Keunggulan Twitter dibandingkan media sosial lainnya adalah sebagai berikut:

1. Tweet 140 karakter yang sederhana memudahkan pengiriman melalui SMS atau email.
2. Fleksibel, Anda dapat menge-tweet menggunakan berbagai program dan perangkat.
3. Hot News Tweet langsung diumumkan ke seluruh pemakai twitter.
4. Tanpa Batas Anda bisa melihat semua tweet orang tanpa harus menjadi teman dahulu.

5. Privasi Tweet menyediakan layanan komunikasi langsung ke pengguna Twitter lain tanpa dibaca oleh pengguna lain.Ini cocok untuk komunikasi bersifat prifat.

2.2.3 *Klasifikasi*

Klasifikasi adalah proses menemukan sekumpulan model atau fungsi yang mendeskripsikan dan membedakan kelas data. Tujuan dari klasifikasi adalah untuk memprediksi kelas dari suatu objek yang kelasnya tidak diketahui. Klasifikasi memiliki dua proses yaitu membangun model klasifikasi dari sekumpulan kelas data yang ada, didefinisikan sebelumnya (training data set) dan menggunakan model tersebut untuk klasifikasi tes data *(prediction*) serta mengukur akurasi dari model. Klasifikasi dapat dimanfaatkan dalam berbagai aplikasi seperti diagnosa seperti medis, selective marketing, pengajuan kredit perbankan, email dan analisis sentimen. Klasifikasi dapat disajikan dalam berbagai macam model klasifikasi seperti *decision trees, naïve bayes classifier, k-nearest-neighbourhood classifier, neural network,Support vectore machine*.

2.2.4 *Pre-Processing*

Preprocessing adalah langkah awal dari penambangan teks untuk mengubah data ke format yang diperlukan. Proses ini dilakukan dalam rangka mencari, mengolah dan mengatur informasi serta menganalisis hubungan tekstual berdasarkan data terstruktur dan tidak terstruktur (Nugroho, 2016).

Gambar 2.1 Alur *Pre-Processing*

1. *Case folding*

   Adalah proses perubahan semua huruf pada dokumen tweet menjadi huruf kecil. Hanya huruf a sampai z yang akan diproses. Karakter selain huruf akan diabaikan.

2. *Tokenizing*

   Tokenisasi adalah proses untuk memotong dokumen menjadi pecahan kecil yang dapat berupa bab, sub-bab, paragraph, kalimat, dan kata (token).

3. *Filtering*

   Adalah tahap mengambil kata-kata penting dari hasil token.

4. *Stop word*

   Membuang kata-kata yang tidak terlalu penting sperti kata: di,dan,atau,ke dll.

5. *Stemming*

   Stemming merupakan proses untuk mencari stem (kata dasar) dari kata hasil stopword removal (filtering). Terdapat dua aturan dalam melakukan stemming yaitu dengan pendekatan kamus dan pendekatan aturan (Utomo, 2013).

2.2.*5 Text Mining*

Text Mining suatu proses untuk menemukan informasi dalam koleksi teks besar, dan secara otomatis mengidentifikasi pola dan hubungan yang menarik dalam data tekstual. Text Mining berupa penelitian yang sangat interdisipliner, mulai dari bidang data, pengolahan bahasa alami, pembelajaran mesin, dan pengambilan informasi (Feldman, Ronen, dan Sanger, James. 2007). Penggunaan dari text mining dilakukan untuk klasterisasi, klasifiasi, information retrieval, dan information extraction (Berry & Kogan, 2010).

2.2.6 Pembobotan TF-IDF

Metode TF-IDF merupakan metode penghitungan bobot setiap kata yang paling umum digunakan dalam temu kembali informasi. Cara ini juga dikenal efektif, mudah, dan memberikan hasil yang akurat. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) untuk setiap token (word) pada setiap dokumen dalam korpus. Cara ini akan menghitung bobot setiap token pada dokumen d dengan rumus:

$$tf = 0.5 + 0.5 * (ft,d)/max(ft,d) \quad \text{……[2.1]}$$

TermFrekuensi istilah menentukan frekuensi (tingkat frekuensi) istilah dalam dokumen. Sedangkan frekuensi dokumen adalah banyaknya dokumen yang terdapat istilah. Seperti yang anda lihat pada rumus di atas, yang digunakan dalam perhitungan adalah nilai *term frequency* dari *document frequency*. N disebut dengan jumlah seluruh dokumen dan df menyatakan nilai *document frequency* dari term yang akan dicari inversnya. Nilai *invers document frequency* (idf) diperoleh dari perhitungan berikut :

$$Idf = \log\left(\frac{N}{df}\right) \quad \text{…..…….. [2.2]}$$

$$W = tf * idf \quad \text{……[2.3]}$$

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses sortir, dimana semakin tinggi nilai W maka semakin besar kesamaan dokumen dengan kata kunci dan sebaliknya.

Penjelasan :

d : dokumen ke-d

t : kata ke-t dari kata kunci

W : bobot dokumen ke-d terhadap kata ke-t

tf : jumlah kata dalam dokumen yang dicari

IDF : Inversed Document Frequency

ft,d : frekuensi kata pada d

df : banyak dokumen yang berisi kata pencarian

2.2.7 Ketidak Seimbangan Data (*Imbalanced Dataset)*

(Bunkhumpornpat et al. (2012). Data dikatakan tidak seimbang ketika suatu kelas data memiliki jumlah amatan yang jauh lebih sedikit dibandingkan kelas lainnya. Data tidak seimbang merupakan suatu kondisi data yang penyebarannya berbeda. Sehingga terdapat kelas yang mendominasi (kelas mayoritas) dibandingkan dengan kelas lain (kelas minoritas). Ketidak seimbangan data tersebut dapat menyebabkan ketepatan klasifikasi dan akurasi menjadi berkurang. Metode klasifikasi *support vector machine (SVM)* merupakan salah satu metode yang memiliki kemampuan klasifikasi yang cukup baik.

2.2.8 *Support Vector Machine*

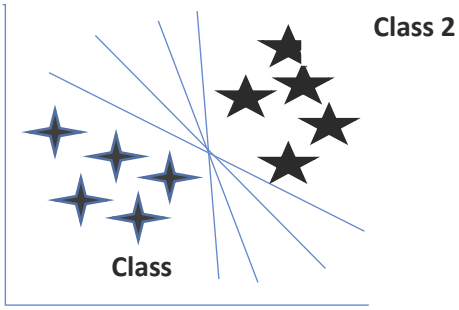
Support Vector Machine Classification (Christianini et al., 2000) adalah sistem pembelajaran mesin yang menggunakan ruang hipotetis berupa fungsi linier dalam ruang fitur berdimensi besar, dilatih dengan algoritma pembelajaran berdasarkan teori optimasi dengan mengimplementasikan deviasi pembelajaran yang dihasilkan dari teori pembelajaran statistik keluar.

Menurut (Santos, 2015), Support Vector Machine (SVM) adalah teknik prediksi yang relatif baru untuk klasifikasi dan regresi. Mesin Vektor Dukungan

berada dalam kelas pembelajaran yang diawasi di mana perlu memiliki fase pelatihan menggunakan pelatihan SVM berurutan yang diikuti dengan fase pengujian.

Support Vector Machine (SVM) adalah metode klasifikasi menggunakan Machine Learning (supervised learning) yang memprediksi kelas dari model atau pola berdasarkan hasil proses pelatihan. Grading dilakukan dengan mencari hyperplane atau batas keputusan yang memisahkan satu kelas dengan yang lain, yang dalam hal ini berperan dalam memisahkan tweet sentimen positif (bertanda +1) dari tweet sentimen negatif (bertanda -1). SVM mencari nilai hyperplane menggunakan vektor bantu dan nilai margin. Pada penelitian ini data masukan yang memiliki representasi vektor diperoleh dari proses penimbangan. Dengan melakukan pelatihan dalam klasifikasi SVM,itu kemudian akan menghasilkan nilai atau pola yang akan digunakan dalam proses pengujian SVM untuk menandai sentimen di tweet.

Proses pembelajaran dalam masalah klasifikasi diterjemahkan sebagai upaya untuk temukan garis (bidang-hiper) yang memisahkan kedua kelompok. Gambar tahapan pembelajaran di SVM adalah:



Gambar 2.2 *hyperplane* kelas positif dan negatif

Hyperplane pemisah terbaik antara dua kelas dapat ditemukan dengan mengukur margin hyperplane dan menemukan titik maksimumnya. Margin adalah jarak antara hyperplane dan data terdekat dari setiap kelas. Subset terdekat dari kumpulan data pelatihan disebut vektor pembawa. Garis penuh pada Gambar 2.2 menunjukkan bidang-hiper terbaik yang berada tepat di tengah-tengah kedua kelas. Upaya menemukan lokasi hyperplane yang optimal menjadi dasar dari proses pembelajaran di SVM (Nugroho, 2008).

Data yang tersedia dilambangkan dengan x ∈ R d, sedangkan label terkait dilambangkan dengan yi ∈ {-1 + 1} untuk i = 1,2, ..., 1, di mana l adalah jumlah data. Diasumsikan bahwa kedua kelas -1 dan +1 dapat dipisahkan sepenuhnya oleh bidang-hiper dengan dimensi d yang ditentukan. Diasumsikan bahwa kedua kelas -1 dan +1 dapat dipisahkan sepenuhnya oleh hyperplane berdimensi d, yang didefinisikan: w.x + b = 0 ……. [2.4]

Pola xi milik kelas -1 (sampel negatif) dapat dirumuskan sebagai pola memenuhi pertidaksamaan: w.x + b = -1 ……. [2.5]

sedangkan pola yang termasuk dalam kelas +1 (sampel positif):

w.x + b = +1 …… .. [2.6]

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara hyperplane dan titik terdekatnya, yaitu 1 / || di ||. Ini dapat dirumuskan sebagai masalah pemrograman kuadratik (QP), yaitu mencari titik minimum persamaan dengan batasan persamaan:

$\min \tau(w) = \frac{1}{2} ||w||^2$ ……...…….[2.7]

$yi(xi.w+b) -1 \geq 0$ ……[2.8]

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, diantaranya Lagrange Multiplier.

1. Mencari Lagrange Multipliers (αi)

$$\tilde{L}(a) = \sum_{i=1}^{n} ai - \frac{1}{2}\sum_{i,j} aiajyiyjxi^{T}xj \quad \text{……[2.9]}$$

Dikenakan (untuk setiap i=1,… .,n)

Keterangan:

yi: kelas data latih (+1/-1)

yj : kelas data latih (+1/-1)

xi : vektor bobot kalimat komentar

xj : vektor bobot kalimat komentar

2. Mencari Nilai Bobot (w)

$$W = \sum_{i=1}^{n}(\propto i\ yixi) \quad \text{……..……[2.10]}$$

Keterangan :

w : vektor bobot.

yi : kelas data latih (+1/-1).

xi : vektor bobot kalimat komentar yang menjadi vektor pendukung.

3. Mencari Nilai Bias (b)

$$b = \frac{1}{NSV}\sum_{i=1}^{NSV}(w\ .\ xi - yi) \quad \text{……...[2.11]}$$

Keterangan :

NSV : jumlah vektor pendukung.

w : vektor bobot

yi: kelas data latih (+1/-1)

xi : vektor bobot kalimat komentar yang menjadi vektor pendukung.

Proses pengklasifikasian (pengujian) dalam SVM menggunakan persamaan:

$$f(\vec{t}) = sgn\left(\sum_{i=1,.xi\in SV}^{n} \alpha iyi < \vec{t}\ .\vec{x}i > +b\right) \text{……[2.12]}$$

Keterangan :

t : vektor bobot data uji

xi: vektor pendukung

b : nilai bias

yi: kelas atau label dari vektor pendukung (+1/-1)

αi adalah Lagrange multipliers, yang bernilai nol atau positif (αi 0). Nilai optimal dari persamaan dapat dihitung dengan meminimalkan L terhadap w dan b, dan memaksimalkan L terhadap αi. Dengan memperhatikan sifat bahwa pada titik optimal gradient L=0, persamaan langkah dapat dimodifikasi sebagai maksimalisasi problem yang hanya mengandung αi saja, sebagaimana persamaan *Maximize* :

$$\sum_{i=1} ai - \frac{1}{2}\sum_{i,j=1} ai\ aj\ yi.yj, \vec{x}_i, \vec{x}_j \quad \text{.....[2.13]}$$

2.2.9 *Confusion Matrix*

Confusion matrix merupakan dataset hanya memiliki dua kelas, kelas yang satu sebagai positif dan kelas yang lain sebagai negatif . Metode ini menggunakan tabel matrix seperti pada tabel 2.1.

Tabel 2.1 *Confusion Matrix*

		Prediksi		Total
		(negatif)	(Positif)	
Example	(Negatif)	p	q	p+q
	(Positif)	u	v	u+v
	tal	p+u	q+v	m

Keterangan:

a. p adalah jumlah prediksi yang tepat bahwa instance bersifat negatif (tn).

b. q adalah jumlah prediksi yang salah bahwa instance bersifat positif (fn).

c. u adalah jumlah prediksi yang salah bahwa instance bersifat negatif (fp).

d. v adalah jumlah prediksi yang tepat bahwa instance bersifat positif (tp).

Berikut adalah persamaan model confusion matrix:

a. Nilai akurasi (acc) adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$\text{Acc} = \frac{TP+TN}{(TP+FP+FN+TN)} \qquad \text{.....[2.17]}$$

b. Tingkat negatif benar (tn) didefinisikan sebagai proporsi kasus negatif yang diklasifikasikan dengan benar, yang dihitung dengan menggunakan persamaan:

$$\text{tn} = \frac{p}{p+q} \qquad \text{.....[2.18]}$$

c. Tingkat negatif palsu (fn) adalah proposi kasus positif yang salah diklasifikasikan sebagai negatif, yang dihitng dengan menggunakan persamaan:

$$\text{fn} = \frac{u}{u+v} \qquad \text{.....[2.19]}$$

d. Tingkat negatif palsu (fp) adalah proporsi kasus negatif yang salah diklasifikasikan sebagai positif, yang dihitung dengan menggunakan persamaan:

$$\text{fp} = \frac{q}{p+q} \qquad \text{.....[2.20]}$$

e. Penarikan kembali (recall) atau tingkat positif benar (tp) adalah proporsi kasus positif yang diklasifikasikan dengan benar, yang dihitung dengan menggunakan persamaan:

$$\text{tp} = \frac{TP}{TP+FN} \qquad \text{.....[2.21]}$$

f. Presisi (p) adalah proporsi prediksi kasus positif yang benar, yang dihitung dengan menggunakan persamaan:

$$prc = \frac{TP}{TP+FP} \quad \text{.....[2.22]}$$

2.2.10 *Evaluation Measure*

Evaluatio Measure adalah sistem pencarian informasi digunakan untuk menilai seberapa baik hasil pencarian, memuaskan maksud permintaan pengguna. Metrik seperti itu sering dibagi menjadi dua yaitu : metrik online untuk melihat interaksi pengguna dengan sistem pencarian, sementara metrik offline untuk mengukur relevansi, dengan kata lain seberapa besar kemungkinan setiap hasil, atau halaman hasil mesin pencari (SERP) secara keseluruhan, dan bertemu kebutuhan informasi pengguna. Evaluasi yang digunakan adalah: precision, recal,akurasi dan f1-score (Powers, David M W. 2011).

2.2.11 *Supervised Learning*

Supervised learning adalah suatu teknik machine learning yang terawasi dan diberi label,dimana hasil akhirnya sudah diketahui dengan membuat suatu fungsi dari data latihan. Data latihan terdiri dari pasangan input dan output yang diharapkan dari input yang bersangkutan. Metode ini bekerja dengan memberikan data input dan menggambarkan output secara langsung.. Tujuh Metode machine Learning yaitu, *Decision Table, Random Forest (RF), Naïve Bayes (NB), Support Vector Machine (SVM), Neural Networks (Perceptron), Jrip and Decision Tree (J48) using Waikato Environment for Knowledge Analysis (WEKA) machine learning tool (*Mark Ryan M. Talabis. D. Kaye, in 2015*).*

2.2.12 *RapidMiner*

Rapid Miner adalah analitik penambangan data, penambangan teks, dan perangkat lunak analitik prediktif. Menggunakan berbagai teknik deskriptif dan prediktif dalam memberikan wawasan kepada pengguna sehingga mereka dapat membuat keputusan terbaik dan memiliki sekitar 500 operator data mining termasuk operator untuk input, output, data preprocessing dan visualisasi.

2.2.13 *Jupyter Notebook*

Jupyter Notebooks (sebelumnya IPython Notebooks) adalah lingkungan komputasi interaktif berbasis web untuk membuat dokumen Notebook Jupyter. Istilah "notebook" secara sehari-hari dapat membuat referensi ke banyak entitas yang berbeda, terutama aplikasi web Jupyter, server web Jupyter Python, atau format dokumen Jupyter tergantung pada konteksnya.Dokumen Notebook Jupyter adalah dokumen JSON menurut skema pembuatan

berisi daftar sel input / output yang diurutkan yang dapat berisi kode, teks (menggunakan Markdown), matematika, plot, dan media kaya, biasanya diakhiri dengan ".ipynb" perpanjangan. Notebook Jupyter dapat dikonversi ke sejumlah format output standar terbuka (HTML, slide presentasi, LaTeX, PDF, ReStructuredText, Markdown, Python) melalui "Unduh Sebagai" di antarmuka web, melalui perpustakaan nbconvert atau "jupyter nbconvert" perintah antarmuka baris dalam shell. Untuk menyederhanakan visualisasi dokumen notebook Jupyter di web, perpustakaan nbconvert disediakan sebagai layanan melalui NbViewer yang dapat mengambil URL ke dokumen notebook yang tersedia untuk umum,

mengonversinya menjadi HTML dengan cepat dan menampilkannya kepada pengguna.Antar muka.

2.2.14 *Python*

Menurut pengertian dari Python Software Foundation (2016), Python adalah bahasa pemrograman interpreatatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. Bahasa pemograman Python disebut sebagai bahasa yang kemampuan, menggabungkan kapabilitas, dan sintaksis kode yang sangat jelas, dan juga dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif danpython bahasa pemrograman tingkat tinggi yang ditafsirkan, interaktif dan berorientasi objek. Python dengan desain yang sangat mudah dibaca dan dipahami, karena sama seperti bahasa pemograman yang lainnya yaitu dengan menggunakan kata bahasa inggris. Selain itu, penggunaan rumus atau sintaksis lebih sedikit.

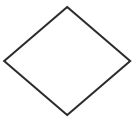
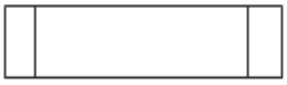
Python terutama mendukung banyak paradigma pemrograman; tapi tidak terbatas; pada pemrograman berorientasi objek, pemrograman imperatif, dan pemrograman fungsional. Salah satu fitur yang tersedia pada Python adalah bahasa pemrogramannya yang dilengkapi dengan manajemen memori otomatis. Seperti halnya bahasa pemrograman dinamis lainnya, python umumnya digunakan sebagai bahasa scripting, walaupun pada praktiknya penggunaan bahasa ini lebih luas dan mencakup penggunaan konteks yang biasanya tidak dilakukan dengan menggunakan bahasa scripting. Python dapat digunakan untuk berbagai tujuan pengembangan perangkat lunak dan dapat dijalankan di berbagai platform sistem operasi.

Seperti bahasa pemrograman dinamis, python sering digunakan sebagai bahasa scripting dengan interpreter terintegrasi ke dalam sistem operasi. Kode Python dapat dijalankan pada sistem berbasis: Linux / Unix, Windows, Mac OSX, OS / 2, Amiga, Palm OS, Symbian (untuk produk Nokia).

2.2.15 *Flowchart*

Flowchart adalah sebuah bagan-bagan yang memiliki arus yang menggambarkan langkah-langkah penyelesaian suatu masalah. *Flowchart* merupakan cara penyajian dari suatu algoritma (Ladjamudin, 2006:265). Simbol *flowchart* dan fungsinya dapat dilihat pada tabel 2.2.

Tabel 2.2 Simbol dan Fungsi *Flowchart*

No	Simbol	Keterangan	Fungsi
1		Terminator	Awal / akhir program
2		Flow Line	Arah aliran program
3		Preparation	Proses inisialisasi / pemberian nilai awal
4		Proses	Proses pengolahan data
5		Decision	Perbandingan pernyataan, penyeleksian data yang memberikan pilihan untuk langkah selanjutnya
6		Predefined process	Permulaan sub program / proses menjalankan sub program
7		Input / Output Data	Proses input / output data, parameter, informasi

8		On Page Connector	Penghubung bagian-bagian flowchart yang berada pada suatu halaman
9		Off Page Connector	Penghubung bagian-bagian flowchart yang berada pada halaman berbeda

BAB III

METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

3.1.1 Metode Penelitian

Beberapa metode untuk mendapatkan data atau informasi dalam pemecahan masalah. Metode yang dilakukan tersebut antara lain:

1. Studi Kepustakaan

   Metode pengumpulan data dengan metode kepustakaan dilakukan dengan pengumpulan jurnal, literatur, paper, makalah, buku, maupun situs internet sebagai sumber pustaka yang berkaitan dengan materi penulisan khususnya analisis sentimen menggunakan metode *Support Vectore Machine (SVM).*

2. Pengumpulan Data Tweet

   Data yang diperoleh merupakan sumber yang diambil secara langsung dari Twitter dan menggunakan aplikasi RapidMiner dengan *keyword* pencarian #BPJSRasaRentenir .

3. Perancangan

   Dalam proses ini ada beberapa hal yang dilakukan seperti:Desain alur pengambilan data, desain alur sistem dan metode, desain pemrograman.

4. Pengujian dan Evaluasi

   Pada tahap ini dilakukan pengujian sistem untuk mnegetahui dimana letak kesalahan yang mungkin terjadi dan dapat ditanggulangi.

5. Penyusunan Laporan

Pada tahap ini penyusunan laporan dibuat sebagai dokumentasi yang berfungsi untuk dapat mempermudah dipelajari dan dikembangkan oleh orang lain.

3.1.2 Spesifikasi Perangkat Keras (*Hardware*)

Perangkat keras (*hardware*) yang digunakan dalam penelitian ini yaitu menggunakan 1 buah laptop. Spesifikasi laptop sebagai berikut :

1. *Processor* : AMD A9-9425 RADEON R5, 5 COMPUTE CORES 2C+3G 3.10GHz
2. *RAM* : 4GB
3. *System Type* : 64bit *operating system*

3.1.3 Spesifikasi Perangkat Lunak (*Software)*

Perangkat lunak (*software*) yang digunakan dalam penelitian ini sebagai berikut :

1. Windows 10 64bit

   Merupakan system operasi yang digunakan pada laptop Asus VivoBook Max X441B.

2. Google Chroome & Mozilla Firefox

   Mesin pencari atau disebut sebuah peramban web sumber terbuka.

3. RapidMiner

   Perangkat yang digunakan dalam pengambilan data pada twitter berupa komentar atau tanggapan sebuah user terhadap tagar (#) topik yang sedang trending di twitter.

4. Microsoft Excel

   Digunakan untuk memfilter dan memilih data *tweets* yang telah didapatkan dari RapidMiner.

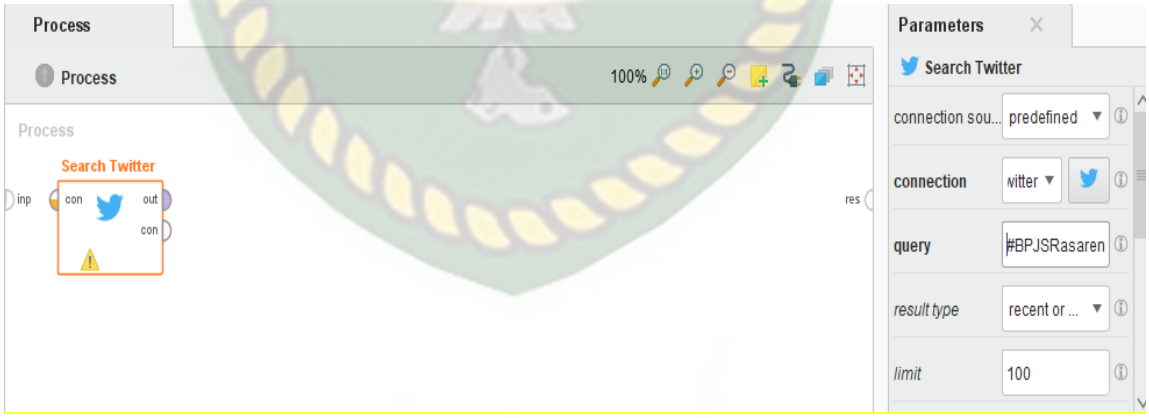
5. Python 3

   Adalah perangkat atau aplikasi yang digunakan dalam membuat dan meracang program.

3.2 Perancangan Sistem

3.2.1 Pengumpulan Data

Dalam penelitian ini, peneliti menggunakan sebanyak 300 data *tweet.* Sebanyak 200 data akan dijadikan sebagai data latih dan 100 data lagi akan dijadikan sebagai data uji. Dalam pengumpulan data *tweet* menggunakan Bahasa Indonesia yang bertagarkan BPJSRasarentenir. Berikut langkah-langkah pengambilan data *tweet* menggunakan RapidMiner pada gambar 3.1.



Gambar 3.1 Proses Crawling Data *Tweet*

Tahapan pertama pada pengambilan data pada RapidMiner Studio adalah dengan memilih operator yang akan digunakan, dimana pada penelitian ini menggunakan

tweet sehingga operator yang digunakan adalah *Search Twitter*. Pada operator *tweet* sendiri terdapat parameter untuk pencarian data yaitu:

Connection : Berfungsi sebagai penghubung untuk pengambilan data.

Query : Berfungsi untuk mencari judul tagar yang ingin digunakan.

Result type : Jenis pencarian data terdapat dua pilihan terbaru atau populer.

Limit : Jumlah batasan data yang akan diambil.

To-User	To-User-Id	Language	Source	Text	Geo-Locatio...	Geo-Locatio...	Retweet-Co...
?	-1	en	<a href="http...	Congratulatio...	?	?	11536
?	-1	en	<a href="http...	$TSLA adds ...	?	?	2626
?	-1	en	<a href="http...	Tesla $TSLA ...	?	?	2067
?	-1	en	<a href="http:/...	$HOT .0011	?	?	0
?	-1	en		RT @daviden...	?	?	1
?	-1	es	<a href="http...	tengo la plen...	?	?	0
?	-1	en	<a href="http...	Go figure. @e...	?	?	0
freshjiva	129235915	en	<a href="http...	@freshjiva It'...	?	?	0
?	-1	en	<a href="http:/...	RT @nichola...	?	?	1
?	-1	en	<a href="http:/...	2021 bull run ...	?	?	1
?	-1	en	<a href="http...	RT @Cryptos...	?	?	1
?	-1	en	<a href="http:/...	RT @DeCrypt...	?	?	1
?	-1	en	<a href="http:/	RT @itsALLri	?	?	84

Gambar 3.2 Hasil Pencarian Data *Tweet*

Hasil pengambilan data dari *tweet* ini akan dilanjutkan dengan mengambil bagian Text saja dan yang lain diabaikan karena data yang akan diolah hanya bagian text, lalu dipindahkan kedalam MS.Excel , berikut tampilan data yang sudah dipindahkan kedalam excel pada gambar 3.3.

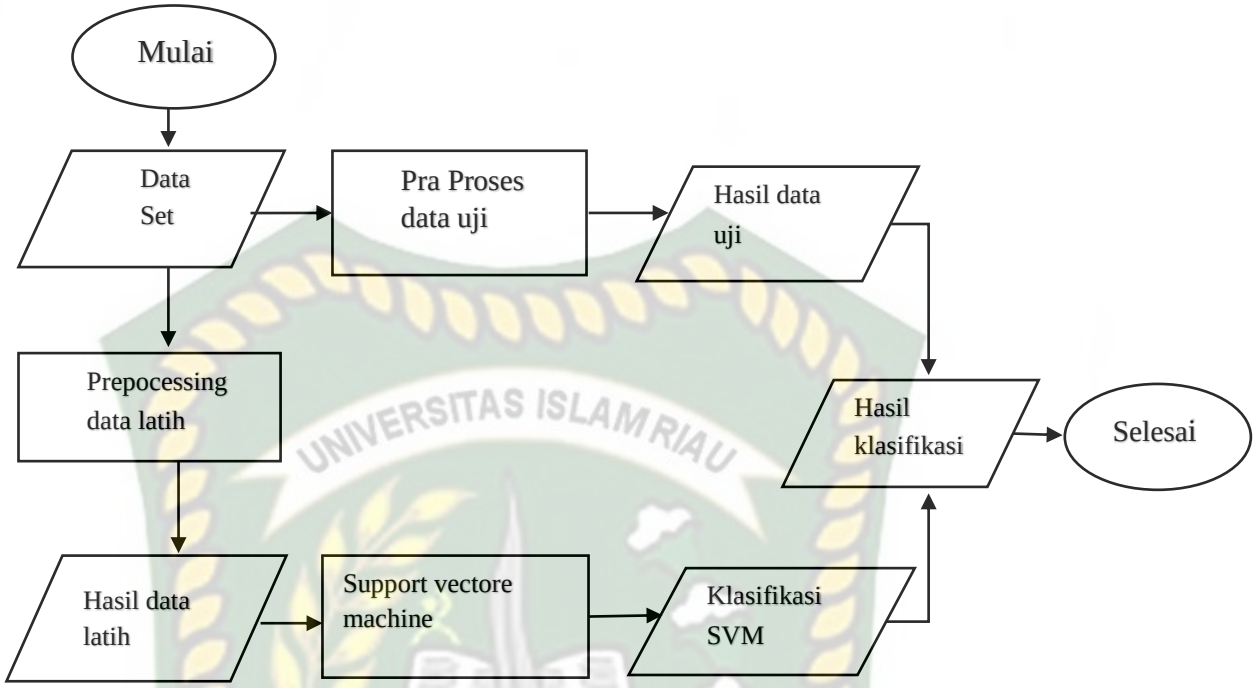
1	Yang ditakuti rakyat adalah kenaikan harga2 barang yang meroketBukan #radikalismehttps://t.co/HhJgKrZQEeTak Cuma Iura
1	"Kami tidak mau kalau hanya naik saja untuk menutupi kekurangan tapi tidak ada kenaikan dalam hal pelayanan," kata Wak
1	RT @_pena_alfath: Seharusnya negara yang bertanggung jawab menjamin kesehatan rakyatnya Bukan rakyat yang disuruh i
1	gw nanya sebagai awam, ye. Dari uang yang masuk ke kas #BPJS, antara yang dibelanjain buat ngobatin peserta dan buat ng
0	RT @J1n_Official: Sebenarnya dari Dulu udah Bnyk yg Sadar, Telah Menjadi Bodoh Cuma JAIM..Melindungi harga diri#BPJSRa
0	RT @katamereka7: ada yang teriak2 #BPJSRasaRentenir ?pas ditanya ikut BPJS gak.?gak ikut.! ??#saveindonesia#IndonesiaS
0	RT @J1n_Official: Sebenarnya dari Dulu udah Bnyk yg Sadar, Telah Menjadi BodohCuma JAIM..Melindungi harga diri#BPJSRa
0	RT @moslimcommunity: Dulu Sudah Diingatkan BPJS Haram, Tapi Ngeyal, Sekarang Rasakan Akibatnya! #BPJSRasaRentenir
0	#BPJSRasaRentenir What.... ? Hidup mu terus mengeluh... Gak suka... Gk usa kau paKE BPJS... Kau kan ahli surga... Doakan
0	Saya sih setuju saja kalau iuran BPJS dinaikkan, asalkan gaji direktur & karyawannya ditinjau kembali. Sudah keterlaluan yg
0	#BPJSRasaRentenirKenaikan iyuran BPJS sangat memberatkan.,!!., Ingat tdk satupun didunia ini yg ingin sakit., hanya org bo
0	ada yang teriak2 #BPJSRasaRentenir ?pas ditanya ikut BPJS gak.?gak ' ' ??#saveindonesia#IndonesiaSatu https://t.co/K3
0	RT @moslimcommunity: Dulu Sudah Diingatkan BPJS Haram, Tapi Ng Sekarang Rasakan Akibatnya! #BPJSRasaRentenir

Gambar 3.3 Data Dalam Bentuk Excel

Setelah memindahakan data kedalam excel tahapan selanjutnya adalah mengelompokan data tersebut menjadi dua kelas, yaitu kelas "1" untuk data positif dan kelas "0" untuk negatif. Selanjutnya data excel yang telah dibuat dikonversikan kedalam bentuk data *CSV* (*Comma Separated Values*) yang bertujuan untuk pemanggilan agar dapat dibaca dan diproses oleh sistem yang digunakan.

3.2.2 Alur Proses Sistem

Tahapan dalam melakukan analisis sentimen dengan metode *klasifikasi SVM* dimulai dengan input data yang berupa data latih dan data uji yang kemudian diproses pada tahapan *pre-processing* hingga proses klasfikasi.



Gambar 3.4 Deskripsi Umum Alur Sistem

Gambar 3.4 menunjukan tahapan-tahapan yang dijalankan pada penelitian ini. Berikut ini adalaha penjabaran dari tiap-tiap proses yang diakukan:

1. Input Data

Pengimputan berdasarkan data yang diambil dari twitter menggunakan rapidminer dan telah diberikan label pada setiap data ketika didalam excel.

2. Preproses program dan pembobotan kata

Data latih dan data uji yang telah diinputkan kemudian diproses oleh sistem untuk dilakukannya proses stopwords atau bisa disebut penghapusan kata tidak penting, contoh kata yang akan diproses oleh *stopword* seperti kata: yang,di,dan,kok,ke dll.

3. Penerapan Metode *Support Vectore Machine*

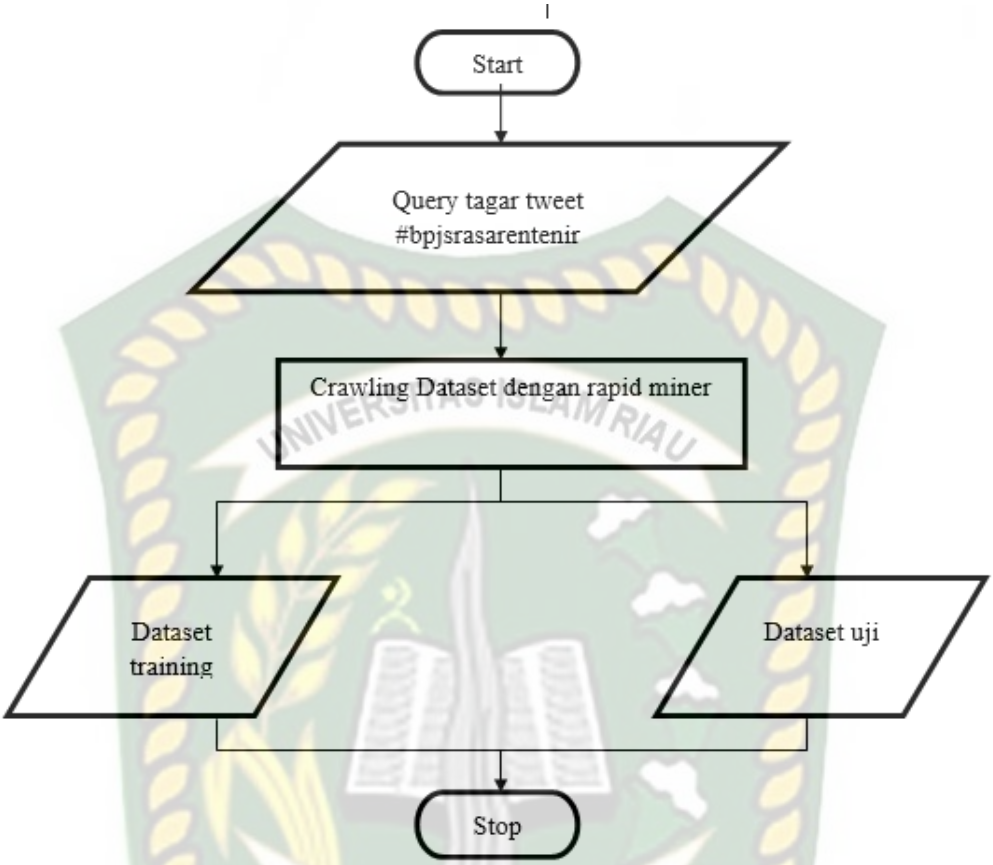
Setelah penerapan preprocessing langkah selanjutnya masuk ke tahap metode *Support Vectore Machine* yang akan melakukan pembobotan pada data yang telah diproses dan menentukan akurasi data latih yang telah di inputkan.

4. Proses pengimputan dan pengujian data uji

Masuk ketahap berikutnya, setelah proses pengujian data latih selanjutnya sistem yang dibuat akan melakukan pengujian menggunaka data uji yang di inputkan kedalam sistem untuk mendapatkan hasil sentimen dikelas positif atau negatif yang diberi label (0) untuk negatif dan (1) positif.

3.2.3 Proses *Dataset*

Dataset berupa teks berbahasa Indonesia yang diambil dari twitter. Data yang diambil untuk penelitian ini mengguanakn *query* ‘bpjsrasarentenir’. *Query* tersebut merupakan tagar yang sedang *trending topic* di twitter. Alur proses pengambilan *dataset* seperti pada gambar 3.5.



Gambar 3.5 Alur *Dataset*

Dataset dari hasil *crawling* ini akan dibagi menjadi dua bagian yaitu data *training* (data latih) dan data *testing* (data uji) yang dipresentasikan pada gambar 3.5. Data latih diklasifikasikan menggunakan *Support vectore machine* dengan label sentimen negatif dan positif. Sedangkan data uji diklasifikasikan secara manual dengan label setimen negatif dan positif. Data uji nanti nya akan digunakan pada saat *evaluasi* untuk menentukan keakuratan data pada sistem. Pengambilan data *training* dan data *testing* dilakukan dengan waktu pengambilan yang berbeda.

3.2.4 *Preprocessing*

Preprocessing adalah tahapan untuk mengolah kata mentah yang baru didapat. Langkah awal setelah data didapat adalah preprocessing. Dalam *preprocessing* sendiri terdapat beberapa tahap, berikut adalah tahapan-tahapan tersebut:

Tabel 3.1 Tahap *Prepocessing*

No	Sebelum	Sesudah
1.	Ternyata solusi menku yang dapat penghargaan dunia cuma sebatas menaikan iuran tagihan	ternyata solusi menku yang dapat penghargaan dunia cuma sebatas menaikan iuran tagihan
2.	Ternyata solusi menku yang dapat penghargaan dunia cuma sebatas menaikan iuran tagihan	|Ternyata| |solusi| |menku| |yang| |dapat| |penghargaan| |dunia| |cuma| |sebatas| |menaikan| |iuran| |tagihan|
3.	Ternyata solusi menku yang dapat penghargaan dunia cuma sebatas menaikan iuran tagihan	Ternyata solusi menku dapat penghargaan sebatas menaikan iuran tagihan

Penjelasan tahapan prepocessing pada tabel 3.1 :

a. Tahapan awal pada *prepocessing* ini melakukan *Case Folding* yaitu merubah huruf besar menjadi huruf kecil,hanya huruf A sampai Z yang akan diproses.

b. Dari proses diatas dilakukan proses *Tokenizing* yaitu proses untuk memotong kata per kata hingga menjadi beberapa bagian.

c. Proses selanjutnya yaitu *Filtering dan Stopword* yaitu proses menghapus kata yang tidak penting dan menampilkan kata penting.

3.2.5 *Term Frequency (TF)*

Pada tahapan ini berfugsi untuk menghitung jumlah kemunculan kata pada dokumen. Apabila jumlah sebuah *term*/kata probabilitas kemunculannya pada sebuah dokumen maka *term*/kata tersebut perlu mendapat *query*. Menggunakan rumus $\text{tf} = 0.5 + 0.5 * \frac{ft,d}{\max(ft,d)}$

3.2.6 *Inverse Document Frequescy* (*IDF*)

Merupakan jumlah dokumen yang berisikan *term* yang dicari dalam dokumen *data set* yang fungsinya adalah mendefinisikan kontribusi dari *term* kedalam dokumen dengan menggunakan rumus $Idf = \log\left(\frac{N}{df}\right)$

3.2.7 *Term Frequency-Inverse Document Frequency* (*TF-IDF*)

Metode pembobotan yang diintegrasikan dari *term frequency* dan *inverse document frequency* dengan menggunakan rumus W = tf * idf metode ini berfungsi untuk mencari representasi nilai dari kumpulan data. Berikut ini contoh data yang akan dihitung menggunakan *TF-IDF* sebagai berikut:

Tabel 3.2 data/document

Data A	Kenaikan BPJS tidak sebanding dengan manfaatnya
Data B	Kenaikan jadi kado terburuk awal pemerintahan Jokowi
Data C	Jokowi teken perpres iuran BPJS
Data D	Tujuan kenaikan iuran BPJS untuk meningkatkan kualitas pelayanan

Pada table 3.2 menggunakan contoh sebanyak 4 data yang akan dproses, dimana data tersebut akan dipisah menjadi bagian kecil untuk menghitung bobot pada setiap *term*/kata.

Selanjutnya adalah tahapan pembobotan pada setiap kata dimana menghitung jumlah kemunculan atau disebut dengan *term frequency* pada setiap data dapat dilihat pada tabel 3.3 sebagai berikut.

Tabel 3.3 Tf-Idf

Term	**TF**				**df**	**idf = log $(\frac{N}{df})$**				**Bobot (w) = tf x idf**			
	A	**B**	**C**	**D**		**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Kenaikan	1	1	0	1	**3**	0.1249	0.1249	0	0.1249	0.1249	0.1249	0	0.1249
BPJS	1	0	1	1	**3**	0.1249	0	0.1249	0.1249	0.1249	0	0.1249	0.1249
Tidak	1	0	0	0	**1**	0.6020	0	0	0	0.6020	0	0	0
sebanding	1	0	0	0	**1**	0.6020	0	0	0	0.6020	0	0	0
Dengan	1	0	0	0	**1**	0.6020	0	0	0	0.6020	0	0	0
manfaatnya	1	0	0	0	**1**	0.6020	0	0	0	0.6020	0	0	0
jadi	0	1	0	0	**1**	0	0.6020	0	0	0	0.6020	0	0
kado	0	1	0	0	**1**	0	0.6020	0	0	0	0.6020	0	0
terburuk	0	1	0	0	**1**	0	0.6020	0	0	0	0.6020	0	0
awal	0	1	0	0	**1**	0	0.6020	0	0	0	0.6020	0	0
pemerintahan	0	1	0	0	**1**	0	0.6020	0	0	0	0.6020	0	0
Jokowi	0	1	1	0	**2**	0	0.3010	0.3010	0	0	0.3010	0.3010	0
teken	0	0	1	0	**1**	0	0	0.6020	0	0	0	0.6020	0
perpres	0	0	1	0	**1**	0	0	0.6020	0	0	0	0.6020	0
iuran	0	0	1	1	**2**	0	0	0.3010	0.3010	0	0	0.3010	0.3010
tujuan	0	0	0	1	**1**	0	0	0	0.6020	0	0	0	0.6020
untuk	0	0	0	1	**1**	0	0	0	0.6020	0	0	0	0.6020
meningkatkan	0	0	0	1	**1**	0	0	0	0.6020	0	0	0	0.6020
kualitas	0	0	0	1	**1**	0	0	0	0.6020	0	0	0	0.6020
pelayanan	0	0	0	1	**1**	0	0	0	0.6020	0	0	0	0.6020

3.2.8 Klasifikasi *Support Vectore Machine (SVM)*

Data berupa kata-kata atau kalimat yang kita inputkan telah melalui tahap *filtering* dan *stopword* selanjutnya akan dilakukan pembobotan pada kata-kata tersebut setelah pembobotan diklasifikasi apabila nilai *term* pada kalimat bernilai >0 maka akan diletakan pada komentar positif dan apabila *term* kalimat bernilai < 0 maka akan diletakan pada komentar negatif.

Dalam klasifikasi *system* hanya melihat pada titik dan ruang pada dokumen tersebut untuk tujuan pemodelan ruang *vector* yang digunakan kemudian memberikan setiap kata dalam dokumen yang akan diproses. *Svm* dalam proses klasifikasi ini bertujuan untuk menemukan garis terbaik untuk membagi kedalam dua kelas kemudian diklasifikasikan dokumen uji berdasarkan pada sisi mana garis itu muncul. Apabila data yang dilatih terpisah secara linear maka bisa dilakukan pemilihan untuk dua *hyperplane* dari margin dengan cara menghilangkan atau tidak ada nilai antara *hyperplane* dan *margin* kemudian dimaksimalkan jaraknya dengan menggunakan persamaan $||\frac{2}{w}||$. Cara ini untuk meminimalkan ||w||. Untuk menghindari titik data (i) jatuh kedalam margin maka dirumuskan sebagai berikut:

w.xi – b ≥ 1 untuk xi kelas pertama atau w.xi – b ≤ -1 untuk xi kelas kedua yang dijelaskan dengan persamaan yi(w.xi – b) ≥1, untuk semua 1 ≤ i ≤ n sehingga untuk mendapatkan optimasi maka minimalkan nilai w dan b ||w|| untuk setiap I = 1,….n yang dirumuskan yi(w.xi – b) ≥ 1 apabila kelompok *hyperplane* ditemukan dengan membagi nilai maka yi(w.xi – b) ≥ 0 untuk itu perlu

dikirimkan semua nilai minimal dengan cara mengirimkan semua xi untuk semua +∞ sehingga nilai minimum ini akan dicapai.

Pencarian bobot w dan factor bias b untuk bidang pembatas sebagai berikut: $yi(wxi + b) \geq 1$,

D1=0.1249 w1 + 0.1249w2 + 0.6020w3 + 0.6020w4 + 0.6020w5 + 0.6020w6 + b≥1

D2=0.1249w1+0.6020w7+0.6020w8+0.6020w9+0.6020w10+0.6020w11+0.3010w12+ b≥1

D3=0.3010w12+0.6020w13+0.6020w14+0.3010w15+0.3010w2+b≥1

D4=0.6020w16+0.1249w1+0.3010w15+0.1249w2+0.6020w17+0.6020w18+0.6020w19 +0.6020w20+b≥1

Maka matrik K yang terbentuk adalah sebagai berikut :

Tabel 3.4 *Matrik* $xi\ xj^T$

	1	2	3	4	
1	1.4808	0.0156	0.0156	0.0312	-
2	0.0156	1.9182	0.0906	0.0156	-
3	0.0156	0.0906	0.9216	0.1062	+
4	0.0312	0.0156	0.1062	1,9338	+
	-	-	+	+	

Pada tabel 3.4 adalah tampilan dari tabel bobot hasil yang diperoleh pada setiap data dan dimasukan kedalam tabel *Matrix.*

$$max \sum_{i=1}^{e} \alpha i - \frac{1}{2} \sum_{i,j=1}^{e} yiyj\alpha i\alpha jxi^{T}\, xj$$

$$Subject\ to = \sum_{i=1}^{e} \alpha iyi = 0$$

$$0 \leq \alpha i, i = 1, \ldots . e,$$

Max [(a1+a2+a3+a4) – (1,4808 a1a1 + 0.0156 a1a2 + 0.0156 a1a3 + 0.0321 a1a4 + 0.0156 a2a1 + 1.9182 a2a2 + 0.0906 a2a3 + 0.0156 a2a4 + 0.0156 a3a1 + 0.0906 a3a2 + 0.9216 a3a3 + 0.106 a3a4 + 0.0312 a4a1 + 0.0156 a4a2 + 0.1062 a4a3 + 1.9338 a4a4)/2]

a1 = 1

a2 = 2

a3 = 1

a4 = 2

$$(x) = sign(\sum_{i=1}^{N} a{:}\, y.K(x, xi) + b)$$

fungsi *sign*() adalah semacam fungsi *normalisasi*, jika nilai x di dalam fungsi *sign*(x) adalah > 0 maka fungsi tersebut memberikan nilai 1, jika nilai x dalam fungsi adalah < 0 maka fungsi akan memberikan nilai -1, dan jika nilai x sama dengan 0 maka fungsi akan memberikan nilai 0.

Tabel 3.5 Bobot d1 sampai d4

	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
D1	0.1249	0.1249	0.6020	0.6020	0.6020	0.6020	0	0	0	0
D2	0.1249	0	0	0	0	0	0.6020	0.6020	0.6020	0.6020
D3	0	0.1249	0	0	0	0	0	0	0	
D4	0.1249	0.1249	0	0	0	0	0	0	0	

	t11	t12	t13	t14	t15	t16	t17	t18	t19	t20
D1	0	0	0	0	0	0	0	0	0	0
D2	0.1249	0.3010	0	0	0	0	0	0	0	0
D3	0	0.3010	0.6020	0.6020	0.3010	0	0	0	0	0
D4	0.1249	0.1249	0	0	0.3010	0.6020	0.6020	0.6020	0.6020	0.6020

Pada tabel 3.5 ini adalah tampilan bobot setiap kata pada data yang digunakan, dimana *term*/kata yang diperoleh dari 4 data diatas adalah sebanyak 20 *term*/kata yang disingkat menjadi t1-t20.

Misal ada kalimat baru = “jokowi meningkatkan pelayanan BPJS”.

Setelah dipecah kalimat baru ini mengandung ‘jokowi’, ’meningkatkan’, ’pelayanan’, ’BPJS’ maka ketika kalimat ini dimasukan kedalam *vector* kata kunci untuk proses klasifikasi menjadi.

Tabel 3.6 kemunculan *term* pada kalimat baru

t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14	t15	t16	t17	t18	t19	t20
0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1

Pada tabel 3.6 menerapkan kemunculan kata baru pada *term*/kata lama sedangkan tabel 3.7 adalah bobot dari kata baru pada setiap data.

Tabel 3.7 bobot *term*/kata baru pada data

	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10
D1	0	0.1249	0	0	0	0	0	0	0	0
D2	0	0	0	0	0	0	0	0	0	0
D3	0	0.1249	0	0	0	0	0	0	0	0
D4	0	0.1249	0	0	0	0	0	0	0	0

	t11	t12	t13	t14	t15	t16	t17	t18	t19	t20
D1	0	0	0	0	0	0	0	0	0	0
D2	0	0.3010	0	0	0	0	0	0	0	0
D3	0	0.3010	0	0	0	0	0	0	0	0
D4	0	0	0	0	0	0	0	0.6020	0	0.6020

K (x,xi) = fungsi ini menggunakan 2 buah matrik.

K (x,x1) = mengartikan perkalian matrik x dengan matrix d 1.

Tabel 3.8 Matrik k

	x1	x2	x3	x4
K(x,x1)	0.1249	0.3010	0.4259	1.3289

Hasil bobot yang diperoleh data baru pada tabel 3.8

Tabel 3.9 Matrik y

y			
-1	-1	1	1

Kelas data yang dipakai dengan menggunakan matrix y pada tabel 3.9

Tabel 3.10 Matrik x,xi

-0,1249	-0,3010	0,4259	1,3289	-0,1249+-0,6020+0,4259+2,6578=1,2327
1	2	1	2	

Pada tabel 3.10 dimana setiap hasil bobot kata baru dikalikan dengan bobot data lama yang telah dibulatkan

$$a1y1k(x,x1) + a2y2k(x,x2) + a3y3(x,x3) + a4y4k(x,x4) + b$$

$$(1 * (-1) * 0.1249) + (2 * (-1) * 0.3010) + (1 * (1) * 0.4259)$$

$$+ (2 * (1) * 1.3289)$$

+(-1) = 1,2327

f(x) = sign(1,2327) = 1

maka kalimat baru “Jokowi meningkatkan pelayanan BPJS” masuk kedalam kelas (1)

BAB IV

HASIL DAN PEMBAHASAN

4.1 Model Analisis Sentimen

Untuk penelitian, penulis memakai *jupyter python* sebagai aplikasi dalam pembuatan analisis sentiment karna dalam pemakaian jupyter python yang simple dan mudah dipahami. Berikut tahapan pada analisis sentiment.

4.1.1 Pelatihan

Pada proses pelatihan ini bertujuan pembentuk model prediksi data yang telah diambil. Untuk penelitian ini penulis menggunakan data latih berjumlah 200 data yang diambil dari twitter dan dibagi menjadi dua model kelas yang pertama kelas "Positif" dan yang kedua kelas "Negatif". Berikut tampilan dari data latih yang ditampilkan oleh sistem pada gambar 4.1.

Out[18]:

	liked	txt
0	1	RT @muit1924: Harusnya tetap dukung meski kebi...
1	1	Lagi ramai #BPJSRasaRentenir ya? Dulu, saya pe...
2	1	Kadang berfikir untuk tidak punya fikiran dala...
3	1	RT @J1n_Official: Sebenarnya dari Dulu udah Bn...
4	1	Alasan yang menurut gua mutlak dari pencabutan...
...	...	...
195	0	RT @bayangan_semar: #BPJSRasaRentenir\nKetika ...
196	0	RT @BinSukamdo_: @hacknet__71 Tengyu kamsiah p...
197	0	#BPJSRasaRentenir Jadi Trending Topic Gegara M...
198	0	Yang ditakuti rakyat adalah kenaikan harga2 ba...
199	0	"Kami tidak mau kalau hanya naik saja untuk me...

200 rows × 2 columns

Gambar 4.1 Tampilan Data Latih

Pada gambar 4.1 terdapat tampilan text atau kalimat yang telah diambil dari twitter,dimana data *tweet* tersebut disimpan kedalam excel lalu di ubah menjadi format *csv*. Pada data *tweet* terdapat dua buah kelas kalimat dimana 1 ditandai sebagai kelas positif dan 0 ditandai sebagai kelas negatif.

Tahapan selanjutnya yaitu menguji akurasi data latih yang telah kita inputkan sebanyak 200 data,dimana akurasi yang didapat adalah 0.94 dengan menggunakan metode *SVM (Support Vectore Machine).*

4.1.2 Pengujian Data uji

Pengujian yang dilakukan dengan data uji yang telah disiapkan adalah proses untuk mengetahui tingkat keakuratan pada data yang akan kita uji nanti, dimana pengguna menginputkan data dan menghasilkan sebuah kelas yang telah dibentuk berupa biner yaitu “1”untuk kelas positif dan “0” untuk kelas negatif. Hasil ini diperoleh sistem melalui pembelajaran dara data latih yang telah di inputkan, berikut tampilan *form* data uji.

	txt
0	#BPJSRasaRentenir Jadi Trending Topic Gegara M...
1	RT @WarisLubis: Beban yang ditanggung begitu b...
2	RT @WarisLubis: Beban yang ditanggung begitu b...
3	RT @BamsBulaksumur: Mahasiswa : "Pak kok ada y...
4	Pagi"gw misuh" k bpjs pemerintah ma pengguna b...
...	...
95	RT @muit1924: @ik4mawar3 @rasmanduri @Lizahra_...
96	RT @muit1924: @ik4mawar3 @rasmanduri @Lizahra_...
97	RT @muit1924: @ik4mawar3 @rasmanduri @Lizahra_...
98	RT @Anggraini_4yu: Tiap Bulan WAJIB BAYAR\nGa ...
99	RT @muit1924: @ik4mawar3 @rasmanduri @Lizahra_...

100 rows × 1 columns

Gambar 4.2 Tampilan data uji

Hasil dari data latih akan di uji dengan data yang telah diinputkan oleh *user* agar mendapatkan ketepatan prediksi. Dimana penulis menggunakan 100 contoh data yang dibagi menjadi dua bagian ,yang pertama dikelas negatif dan kedua dikelas positif. Pengujian ini menggunakan *whitebox* yaitu salah satu metode yang berfokus pada sisi *source code* di sistem dan uji manual.

Dari hasil *source code* tersebut maka akan ditampilkan sebuah table ketepatan prediksi dari kata yang telah di inputkan, hasil dari *source code* yang telah diinputkan akan menghasilkan berupa tabel ketepatan prediksi oleh sistem seperti gambar dibawah ini.

	Tweet	Sentiment	Uji Manual	Ketepatan Prediksi
0	#BPJSRasaRentenir Jadi Trending Topic Gegara M...	0	0	Tepat
1	Perangkat sosial negara dipakai sebagai tukang...	0	0	Tepat
2	TARING BPJS putus ditangan mas jae Defisit men...	0	0	Tepat
3	RT @KhaulaAlAzhar: Nampaknya rezim mau menikma...	0	0	Tepat
4	RT @rasmanduri: Berita Tahun 2015)Hizbut Tahri...	0	0	Tepat
5	Beban yang ditanggung begitu berat bagi rakyat...	1	0	Tidak Tepat
6	RT @rasmanduri: Berita Tahun 2015)Hizbut Tahri...	0	0	Tepat
7	RT @dzoemient12: Dipaksa dan terpaksa masuk bp...	0	0	Tepat
8	RT @Muji_TabloidMU: @rmol_id Kayaknya baru pua...	0	0	Tepat
9	RT @Gambio7: Yg mewajibkan BPJS pemerintah, ra...	1	0	Tidak Tepat
10	RT @arjw15: katanya untuk menolong rakyat eh k...	0	0	Tepat
11	#BPJSRasaRentenirMau berobat susah bener. Kala...	0	0	Tepat
12	RT @BambangBAlfary: Yg di naikan itu pelayanan...	0	0	Tepat
13	RT @Anggraini_4yu: Tiap Bulan WAJIB BAYARGa Ba...	0	0	Tepat
14	Kejutan di bulan ini adalah tarif BPJS yang na...	1	0	Tidak Tepat
15	RT @opposite6890: Naaah.. ini dia Wirr.. @wira...	0	0	Tepat
16	@Rusydi__riau40 @LisaAmartatara3 @K1ngPurwa @P...	0	0	Tepat
17	@RstyCayah @jokowi Wong edan lagi ngoceh koq d...	1	0	Tidak Tepat
18	Percayalah hari gni semuanya UUD (ujung2nya du...	0	0	Tepat
19	RT @opposite6890: Naaah.. ini dia Wirr.. @wira...	0	0	Tepat
20	@geloraco Tidak memberatksn, hanya mencekik#BP...	0	0	Tepat
21	sunggu nasip PT LECES BUMN sdh lama berkotat ...	0	0	Tepat

Gambar 4.3 Tampilan Ketepatan Prediksi

Sentimen adalah hasil uji dari sistem, sementara uji manual merupakan hasil dari pengujian pengguna. Jika hasil uji dari sistem sama dengan hasil uji dari pengguna maka prediksi nya adalah tepat, apabila jika hasil uji tidak sama maka prediksi yang dihasilakan adalah tidak tepat.

4.2 *Evaluation Measure*

Evaluasi *measure* adalah pengujian hasil klasifikasi dengan mengukur nilai kebenaran dari sistem. Dimana dataset memiliki dua kelas,yaitu terbagi menjadi kelas positif dan kelas negatif. Pada kasus ini ada dua entri baris dan kolom *confusion matrix* yaitu sebagai *true and false positives* dan *true and false negatives,* seperti pada tabel 4.1 di bawah ini:

4.1 Tabel *Confusion Matrix*

		Predicted	
		Positive	Negative
Positive		TP	FN
Negative		FP	TN

True Positif = Suatu keadaan dimana sistem mendeteksi hasil positif dan hasil dari uji manual menunjukan hasil yang positif pula.

True Negatif = Suatu keadaan dimana sistem mendeteksi hasil negative dan hasil dari uji manual menunjukan hasil negatif pula.

False Positif = Suatu keadaan dimana sistem mendeteksi hasil negatif dan hasil dari uji manual menunjukan hasil positif.

False Negatif = Suatu keadaan dimana sistem mendeteksi hasil positif dan hasil dari uji manual menunjukan hasil negatif.

4.2.1 *Accuracy*

Accuracy adalah fraksi prediksi yang benar, data yang berhasil diklasifikasikan dengan benar oleh sistem. *Accuracy* didapatkan dari hasil perhitungan yang ditunjukkan pada persamaan di bawah ini.

$$Accuracy = \frac{TP+TN}{(TP+FP+FN+TN)}$$

Dari 100 sample data *testing* yang sudah diproses ketepatan prediksi yang sudah didapatkan sebagai berikut:

True Positif : 34

True Negatif : 57

False Positif : 3

False Negatif : 6

Cara menghitung akurasinya adalah sebagai berikut:

$$Accuracy = \frac{34+57}{(34+3+6+57)} = \frac{91}{(100)} = 0.91$$

Pada penelitian ini menggunakan bilangan bulat karna pada perhitungan ini menentukan *range* dengan nilai 1-100 maka dikalikan dengan 100, sehingga hasil yang didapat adalah 91. Sehingga hasil yang diklasifikasikan dengan tepat oleh sistem memiliki akurasi yang baik.

Tampilan dari *source code* pada program adalah sebagai berikut:

```
def compute_akurasi(tp, tn, fn, fp) :
    '''
    akurasi = TP + TN / TP + TN + FP + FN
    '''
    return ((tp + tn) *100)/ float(tp + tn +fn + fp)
```

```
print ('akurasi :', compute_akurasi(tp, tn, fn, fp))
```

```
akurasi : 91.0
```

Gambar 4.4 Perhitungan akurasi

Bisa disimpulkan dalam pengujian ini bahwa perhitungan akurasi sistem dan akurasi manual dinyatakan sama, maka pengujian ini dianggap sesuai denga apa yang diharapkan.

Untuk evaluasi yang telah dilakukan dapat menghasilkan nilai tinggi karna data *testing* yang diinputkan memiliki kosakata yang terdapat pada data *training*, sehingga sistem dapat mempelajari dari data *training* dan memproses data *testing* dengan baik.

4.3 Antar Muka Pada Sentiment Analisis

Pada penelitian ini setiap *user* bisa mengunakan sebuah *form* yang telah dibuat untuk dapat melihat hasil dari data *tweet*. Dalam antar muka ini setiap user bisa menginputkan sebuah data berupa text yang nantinya akan diproses oleh sistem dan langsung dieksekusi sehingga menghasilkan output berupa prediksi, ada pun prediksi yang dihasilkan berupa prediksi positif atau negatif. Berikut ini tampilan dari antar muka bisa dilihat pada gambar 4.5 berupa tampilan form input dan pada gambar 4.6 berupa tampilan *form* output hasil keluaran.

Gambar 4.5 *Form Input User*

Gambar 4.6 *Form Hasil Output User*

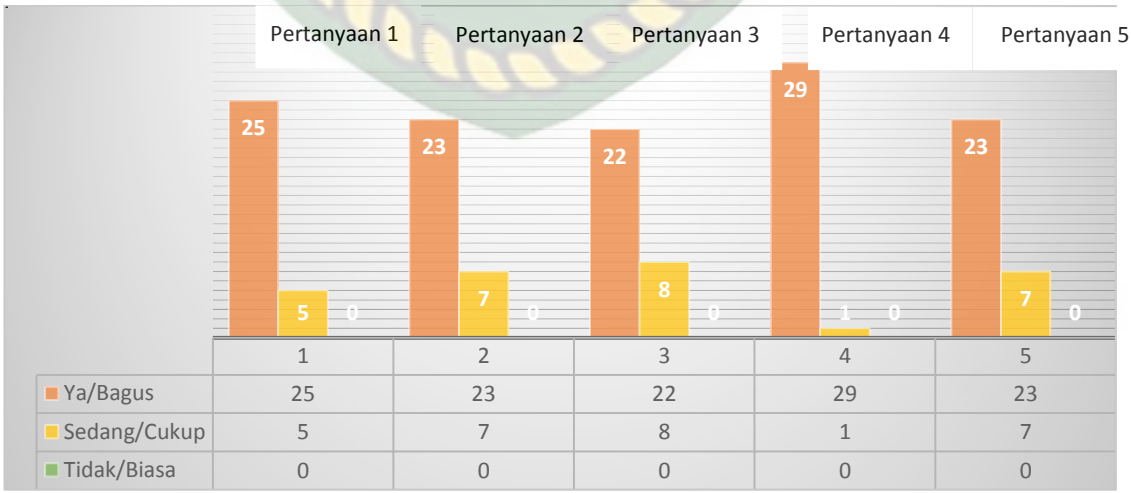
4.4 Pengujian Kepada User

4.4.1 Implementasi User

Implementasi sistem yang dipakai adalah membuat kuisioner dengan 5 pertanyaan dan 30 koresponden yang mana ditujukan kepada user yang ingin memakai sistem ini. kepada 30 koresponden diajukan pertanyaan yang terkait dengan kinerja atau *performance* dari sistem. Adapun kelima pertanyaan yang dimaksud adalah sebagai berikut :

1. Apakah informasi yang ditampilkan mudah dimengerti oleh user?
2. Apakah bahasa yang digunakan pada analisis sentimen mudah dimengerti?
3. Bagaimana pendapat anda tentang tampilan pada analisis sentimen ini?
4. Apakah analisis sentimen ini cukup mudah digunakan (dioperasikan)?
5. Menurut anda, apakah analisis sentimen ini sudah layak dipublikasikan?

Dari pertanyaan-pertanyaan diatas, maka hasil jawaban atau tanggapan dari koresponden terhadap kinerja atau *performance* dari sistem berdasarkan pertanyaan yang diajukan adalah sebagai berikut:



	1	2	3	4	5
Ya/Bagus	25	23	22	29	23
Sedang/Cukup	5	7	8	1	7
Tidak/Biasa	0	0	0	0	0

Gambar 4.7 Hasil Kuisioner

Keterangan:

1. Apakah informasi yaditampilkan mudah dimengerti oleh user memiiki nilai YA : 25 koresponden, SEDANG : 5 koresponden, TIDAK : 0 koresponden.
2. Apakah bahasa yang digunakan pada analisis sentimen mudah dimengerti memiliki nilai YA : 23 koresponden, SEDANG : 7 koresponden, TIDAK : 0 koresponden.
3. Bagaimana pendapat anda mengenai tampilan pada analisis sentimen ini memiliki nilai BAGUS : 22 koresponden, CUKUP : 8 koresponden, BIASA : 0 koresponden.
4. Apakah analisis sentimen ini cukup mdah digunakan (dioperasikan) memiliki nilai YA : 29 koresponden, SEDANG : 1 koresponden, TIDAK : 0 koresponden.
5. Menurut anda, apakah analisis sentimen ini sudah layak dipublikasikan memiliki nilai YA : 23 koresponden, CUKUP : 7 koresponden, TIDAK : 0 koresponden.

4.4.2 Kesimpulan Implementasi Sistem

Berdasarkan hasil kuisioner diatas maka bisa disimpulkan bahwa aplikasi manajemen jurnal online ini memiliki persentase seperti yang dapat dilihat pada tabel 4.2.

Tabel 4. 2 Hasil Nilai Persentase Tiap Pertanyaan Kuisioner

No	Pertanyaan	Jumlah Persentase Koresponden		
		Ya / Bagus	Sedang / Cukup	Tidak / Biasa
1	Apakah informasi yaditampilkan mudah dimengerti oleh user?	83%	17%	0%
2	Apakah bahasa yang digunakan pada analisis sentimen mudah dimengerti?	77%	23%	0%
3	Bagaimana pendapat anda mengenai tampilan pada analisis sentimen ini?	73%	27%	0%
4	Apakah analisis sentimen ini cukup mdah digunakan (dioperasikan)	97%	3%	0%
5	Menurut anda, apakah analisis sentimen ini sudah layak dipublikasikan	77%	23%	0%

Dari hasil persentase tabel 4.2 nilai presentase tiap pertanyaan kuisioner, analisis sentimen pada twitter dengan tagar #BPJSRasarentenir memiliki *performance* baik dengan nilai persentase rata-rata sebesar 81.4%, sehingga aplikasi ini dapat diimplementasikan ke publik.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan bahwa penelitian tentang analisis sentimen pengguna Twitter pada topik #BPJSrasarentenir dengan 200 sample data menggunakan metode *Support Vectore Machine* telah berhasil dilakukan, denga hasil akurasi 94%. Penelitian ini menghasilkan model yang dapat melakukan klasifikasi dan prediksi tentang tanggapan masyarakat terhadap BPJS apakah itu tanggapan positif maupun negatif. Dan pada penelitian ini analisis sentiment masyarakat tentang #BPJSrasarentenir lebih beranggapan negatif.

5.2 Saran

Riset yang dilakukan ini tidak terlepas dari kekurangan dan kelemahan. Oleh karena itu, peneliti menyarankan adanya perbaikan-perbaikan :

1. Penelitian selanjutnya diharapkan menggunakan dua atau lebih metode atau algoritma untuk meningkatkan keakuratan sistem.
2. Sistem belum dapat mengidentifikasi *slang word* / kata-kata yang disingkat yang berpengaruh pada keakuratan perhitungan sehingga diperlukan adanya fitur untuk mengidentifikasi masalah tersebut.

Sistem yang dibangun masih terpisah antara modul untuk mengambil data dari Twitter dan modul untuk perhitungan sentimen analisis. Pada penelitian selanjutnya diharapkan dapat dibuat sistem yang lebih terintegrasi

DAFTAR PUSTAKA

Akbari, M. I. H. A. D., Astri Novianty S.T., M. & Casi Setianingsih S.T., M. (2012). *Analisis Sentimen Menggunakan Metode Learning Vector Quantization.* Telkom University.

B.Liu. (2012). *Sentimen Analysis and opinion mining.* Morgan & Claypool Publisher.

Berry, M.W. & Kogan, J. (2010). *Text Mining Aplication and theory.* WILEY : United Kingdom.

Bramer, Max. (2007). *Principles of Data Mining.* London: Springer. ISBN-10: 1-84628-765-0,ISBN-13: 978-1-84628-765-7.

Bukhari, Varian Habbie. (2015). *Sentiment Analysis Menggunakan K-Nearest Neighbor dengan Perbandingan Fungsi Jarak (Studi Kasus : Twitter Indosat dan Telkomsel).* Bandung : Universitas Widyatama.

Bunkhumpornpat, C., K. Sinapiromsaran, C. Lursinsap. (2012). *density-based synthetic minority over-sampling technique.* Application Intelligence 36, 664–684.

Christianini, N.,& Shawe-Taylor, J. (2000). *Support vectore machines.* Cambridge, UK:Cambridge University,Press, 93(443), 935948.

Feldman, Ronen, dan Sanger, James. (2007). *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data.* New York: Cambridge University Press.

Manning, C., Raghavan, P. & Schütze, H,. (2009). *An Introduction to Information Retrieval.* Cambridge: Cambridge University.

Mark Ryan M. Talabis. D. Kaye, . (2015). *Supervised learning.* Nugroho, A. S. (2008). *Support Vector Machine: Paradigma Baru Dalam Softcomputing.* Konferensi Nasional Sistem Dan Informatika , 92-99.

Nugroho, G. A. P,. (2016). *Analisis Sentimen Data Twitter Menggunakan K-Means Clustering.*

Pang, B., & Lee,L. (2008). *Opinion Mining and sentiment analysis.* Foundations and trends in information retrieval, 2(1-2), 1-135.

Powers, David M W. (2011). *Evaluation From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation.* South Australia, Flinders University.

Santosa, B,. (2015). *Tutorial Support Vector Machine.*

Simon Haykin. (1999). *Neural Network: A Comprehensive Foundation.* Prentice Hall, New Jersey: A Comprehensive Foundation.

Utomo, M. S,. (2013). *Implementasi Stemmer Tala pada Aplikasi Berbasis Web.* Jurnal Teknologi Informasi DINAMIK, Volume 18, pp. 41-45.

Dian Indriani. (2019). *Analisis sentiment pada tweet dengan tagar #kpujangancurang menggunakan metode Naïve Bayes. Riau:Universitas Islam Riau.*