IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE (SVM)* UNTUK MENGETAHUI RESPON MASYARAKAT INDONESIA TERHADAP PEMBERIAN VAKSIN SINOVAC

Skripsi

Oleh :

Obey Al Farobi

11160910000076

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA

1442 H / 2021 M

IMPLEMENTASI METODE *SUPPORT VECTOR MACHINE (SVM)* UNTUK MENGETAHUI RESPON MASYARAKAT INDONESIA TERHADAP PEMBERIAN VAKSIN SINOVAC

Skripsi

Diajukan sebagai salah satu syarat untuk memperoleh gelar S.Kom.

Oleh :

Obey Al Farobi

11160910000076

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA

1442 H / 2021 M

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan hasil karya asli saya yang diajukan untuk memenuhi salah satu persyaratan memperoleh gelar strata 1 di UIN Syarif Hidayatullah Jakarta.
2. Semua sumber yang saya gunakan dalam penulisan ini telah saya cantumkan sesuai dengan ketentuan yang berlaku di UIN Syarif Hidayatullah Jakarta.
3. Apabila di kemudian hari terbukti bahwa karya ini bukan hasil karya asli saya atau merupakan hasil jiplakan dari karya orang lain, maka saya bersedia menerima sanksi yang berlaku di UIN Syarif Hidayatullah Jakarta.

Jakarta, 15 November 2021

Obey Al Farobi

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI

Sebagai sivitas akademik UIN Syarif Hidayatullah Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Obey Al Farobi

NIM : 11160910000076

Program Studi : Teknik Informatika

Fakultas : Sains dan Teknologi

Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Islam Negeri Syarif Hidayatullah Jakarta **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*)** atas karya ilmiah saya yang berjudul :

"IMPLEMENTASI METODE *SUPORT VECTOR MACHINE* (SVM) UNTUK MENGETAHUI RESPON MASYARAKAT INDONESIA TERHADAP PEMBERIAN VAKSIN SINOVAC"

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Islam Negeri Syarif Hidayatullah Jakarta berhak menyimpan, mengalihmedia/fomatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Jakarta

Pada tanggal: 15 November 2021

Yang menyatakan

(Obey Al Farobi)

Nama : **Obey Al Farobi**

Program Studi : **Teknik Informatika**

Judul : **Implementasi Metode *Suport Vector Machine* (SVM) Untuk Mengetahui Respon Masyarakat Indonesia Terhadap Pemberian Vaksin Sinovac**

ABSTRAK

Penelitian ini bertujuan melakukan implementasi *Support Vector Machine* dan Word Embedding dalam kasus sentiment analysis tentang tanggapan masyarakat tentang pemberian vaksin sinovac yangg diunggah di Twitter dan 3 class:positif, negative dan netral. Metode yang dipilih adalah metode klasifikasi *Support Vector Machine*. Sebelum melakukan klasifikasi, praprocessing pada penelitian ini meliputi tokenizasi, normalisasi, menghilangkan emoticon, Convert Negasi, *Stemming*, Stopword Removal serta Word embedding. Dataset yang digunakan berjumlah 30000 record. Program dirancang menggunakan Bahasa pemrograman python dengan beberapa library seperti keras, tensorflow dan pandas, nltk, Sckitlearn. Akurasi yang didapatkan pada pelatihan menggunakan *Support Vector Machine* sebesar 85%. Hasil pengujian adalah penelitian ini mampu melakukan sentiment analysis dengan kesalahan sebesar 15%.

Kata kunci : *Word embedding, Support Vector Machine, Sentiment Analysis*

Jumlah Pustaka : 20 Jurnal + 8 buku

Jumlah Halaman : Bab 1-6 = 81 halaman

KATA PENGANTAR

Puji syukur atas kehadirat Allah *Subhanahu wa Ta'ala* yang telah memberikan rahmat dan nikmat yang besar kepada penulis sehingga dapat menyelesaikan skripsi ini. Penulisan skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) Syarif Hidayatullah Jakarta.

Penulis sangat menyadari bahwa skripsi ini tidak akan bisa terselesaikan tanpa bantuan dan dukungan dari berbagai pihak. Maka pada kesempatan ini penulis bermaksud mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak, diantaranya :

1. Allah SWT, Tuhan yang Maha Pengasih lagi Maha Penyayang yang telah memberikan nikmat yang tak kunjung henti kepada penulis serta diberikan kelancaran dalam menyelesaikan skripsi.
2. Nashrul Hakiem, S.Si., M.T., Ph.D selaku Dekan Fakultas Sains dan Teknologi.
3. Dr. Imam Marzuki Shofi, M.T selaku Ketua Program Studi Teknik Informatika Fakultas Sains dan Teknologi.
4. Khodijah Hulliyah, M.Si., Ph.D dan Siti Ummi Masruroh, M.Sc selaku Dosen Pembimbing I dan II yang senantiasa membimbing, mengarahkan, dan memotivasi penulis selama penyusunan skripsi.
5. Orang tua dan keluarga Teteh Tetehku Kakang Kakangku Terbaik yang selalu mendoakan dan mendukung Serta Memotivasi penulis.
6. Sahabat sekaligus teman seperjuangan, Muhammad Nur, Muhammad Farhan Suralaga, Abdul Halim Arraisuli, dan I'im Umamil Khoiri yang selalu memberi dukungan serta semangat satu sama lain hingga saat ini.
7. Teman – Teman Crew GandoLL Base Channel yang senantiasa memberi dukungan dan semangat kepada penulis.
8. Teman – teman seperjuangan kelas TI C Angkatan 2016 yang senantiasa memberikan dukungan semangat satu sama lain.
9. Serta semua pihak yang telah membantu penyelesaian skripsi ini yang tidak dapat disebut namanya satu per satu.

Tidak menutup kemungkinan bahwa skripsi ini tidak luput dari kekurangan dan kesalahan. Oleh karena itu, penulis memohon maaf atas segala kekurangan ataupun kesalahan baik dari segi keilmuan atau penulisan. Bila ada kritik dan saran yang

tentunya dibutuhkan untuk penelitian ini, dapat dikirimkan ke email obey.al16@mhs.uinjkt.ac.id. Akhir kata, semoga skripsi ini dapat bermanfaat bagi para pembaca pada umumnya dan penulis pada khususnya. Amin.

Jakarta, 15 November 2021

Obey Al Farobi

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1.Latar Belakang

Berdasarkan data dari *we are social dan hootsuite*, jumlah pengguna sosial media aktif di Indonesia mencapai 160 juta pengguna. Jumlah tersebut sebesar 59% dari total penduduk di Indonesia. Lebih lanjut dari data tersebut, pertambahan yang terjadi dari tahun 2019 untuk pengguna aktif adalah sebesar 8,1 % atau 12 juta pengguna. Dari semua data pengguna aktif sosial media, hampir seluruhnya atau sebesar 99% melakukan akses melalui mobile platform. Dari data yang ada, dapat diambil kesimpulan bahwa memang pasar untuk pengguna sosial media sangat besar, banyak aktivitas yang dihabiskan penduduk Indonesia dalam mengakses sosial media. (Hootsuite & We Are Social, 2020)

Peluang yang dapat ditangkap dari fakta dan data tersebut adalah ada informasi lebih lanjut yang dapat digali dari aktivitas yang terjadi di sosial media. Proses penggalian informasi yang dapat dilakukan ada banyak cara, seperti *crawling* ataupun *scrapping*. Proses *crawling* dapat dilakukan untuk mengambil data atau konten yang ada di sosial media. Data yang diambil masih data mentah dan kotor yang perlu dilakukan tahapan *pre processing* dan melalui banyak tahapan sehingga dapat dihasilkan informasi yang baru dan bermanfaat. Hal tersebut sering disebut sebagai analisis sentimens (Ruan et al., 2018). Sentimen merupakan proses komputasional dalam mengidentifikasi dan mengategori opini-opini dalam bentuk potongan teks, khususnya untuk mengukur maksud si pembuat potongan teks terhadap topik tertentu, dapat bernada positif, negatif, atau netral. Dalam konteks respon masyarakat terhadap kebijakan pemberian vaksin, sentimen yang sering muncul biasanya adalah sentimen yang bernilai positif

Dalam bentuk pujian dan apresiasi maupun sentiment bernilai negatif dalam bentuk kritik.

Dikutip dari detik, kebijakan pemerintah untuk melakukan vaksinasi terhadap masyarakat menuai pro dan kontra. Banyak yang setuju dengan langkah yang diambil pemerintah, tetapi banyak juga yang masih meragukan kualitas dan keamanan dari vaksin yang diberikan, yakni vaksin sinovac. PT Bio Farma (Persero) menyebut ada 4 juta dosis vaksin COVID-19 yang siap disuntikkan pada Februari. Vaksin yang siap disuntikkan ini berasal dari bahan baku yang sudah diterima Bio Farma sejak Oktober 2020 yaitu vaksin sinovac. Bio Farma sudah menerima setidaknya 3 juta bahan Baku vaksin dari Sinovac. Di mana, pendistribusian dilakukan menjadi dua tahap yaitu sebanyak 1,2 juta dosis dan sebanyak 1,8 juta bahan baku vaksin pada Desember 2020. (Finance, 2021)

Hal tersebut mengindikasikan bahwa proses vaksinasi kepada masyarakat Akan dilakukan dengan kuantitas yang lebih banyak dan jangkauan yang lebih luas lagi. Untuk dapat menangkap respon yang terjadi di masyarakat salah satunya dilakukan analisis sentiment, jadi data yang didapatkan dari sosial media yang mobilitasnya tinggi dan menghasilkan banyak data setiap harinya Akan dilakukan pelabelan sentiment, dalam kategori positif, negatif, dan netral. Ada banyak metode yang dapat mengklasifikasikan pelabelan dalam sentiment.

Banyak metode atau algoritma yang dapat diterapkan dalam analisis sentimen, beberapa jenis metode yang dapat digunakan dalam melakukan analisis sentiment adalah algoritma atau metode klasifikasi yang Akan berguna dalam mengelompokan sentimen. Beberapa jenis metode tersebut seperti K-*Nearest Neighbour, Naïve Bayes, C.45, Support Vector Machine* dan beberapa metode klasifikasi lainnya.

Penelitian rujukan pertama adalah penelitian yang dilakukan oleh Oki Arifin, dkk Analisa Perbandingan Tingkat Performansi Metode *Support*

Vector Machine Dan *Naïve Bayes Classifier* Untuk Klasifikasi Jalur Minat SMA. Dari penelitian ini menyimpulkan bahwa hasil akurasi SVM lebih tinggi dengan akurasi sebesar 96,88% dan Naïve Bayes dengan akurasi sebesar 89,63%. (Arifin & Sasongko, 2018). Selanjutnya, penelitian dari Anisa Eka Puridewi, Jaka Nugraha. Hasil dari penelitian menyimpulkan bahwa Algoritma *Naïve Bayes* dan *Support Vector Machine* memiliki akurasi yang sama sebesar 0,6758 atau prosentase sebesar 67%. (Anisa Eka Puridewi, 2018)

Dari kedua penelitian tersebut, akurasi yang dihasilkan oleh metode *Support Vector Machine (SVM)* cukup baik dengan menghasilkan akurasi yang cukup tinggi. Sehingga peneliti Akan menggunakan metode *Support Vector Machine (SVM)* dalam penelitian ini untuk mendeteksi tanggapan tentang vaksin Sinovac. *Focus* pada penelitian ini hanya vaksin Sinovac dikarenakan data yang diambil dari kaggle pada bulan Oktober 2020-Januari 2021 baru muncul vaksin sinovac.

1.2. Rumusan Masalah

Berdasarkan latar belakang diatas, maka peneliti mengambil rumusan masalah sebagai berikut:

a. Bagaimana menerapkan algoritma *Support Vector Machine (SVM)* dalam melakukan klasifikasi sentiment di sosial media terhadap pemberikan vasksin sinovac?
b. Berapa akurasi yang dihasilkan oleh metode *Support Vector Machine (SVM)* dalam melakukan klasifikasi sentiment?
c. Bagaimana melakukan deteksi otomatis menggunakan SVM untuk membedakan sentiment positif, *negative* dan netral?

1.3. Batasan Masalah

Sentimen analisis ini memiliki cakupan yang luas, untuk itu, agar penelitian lebih fokus, maka peneliti membuat batasan masalah yaitu :

a. Menggunakan algoritma *Support Vector Machine (SVM)* untuk klasifikasi sentiment.

b. Data yang digunakan dalam penelitian diambil dari Kaggle berupa text berbahasa Indonesia.
c. Klasifikasi berupa positif, negatif dan netral
d. Periode pengambilan dataset dimulai dari 11 Oktober 2020 – 29 Januari 2021
e. Teks yang diambil hanya berbahasa Indonesia

1.4.Tujuan Penelitian

Mengetahui hasil dari penerapan algoritma *Support Vector Machine (SVM)* dalam melakukan klasifikasi sentiment mengenai vaksin sinovac dan mengetahui akurasi dari algoritma *Support Vector Machine (SVM)* dalam melakukan klasifikasi.

1.5.Manfaat Penelitian

1. Bagi Penulis

   Menerapkan ilmu yang didapatkan, dalam hal ini ilmu data mining atau data science untuk melakukan analisis sentiment dengan topik vaksinasi sinovac dan mengklasifikasinnya menggunakan algoritma *Support Vector Machine (SVM)*.

2. Bagi Pembaca

   Memberikan wawasan dan referensi tentang ilmu data mining dan mendapatkan gambaran langkah proses sentiment analisis menggunakan algoritma *Support Vector Machine (SVM)*.

3. Bagi Universitas

   Dapat mengukur tingkat kemampuan mahasiswa selama belajar dalam perkuliahan dan dapat menambah referensi literatur terhadap algoritma *Support Vector Machine (SVM)*.

1.6. Metode Penelitian

1.6.1. Metode Pengumpulan Data

Dalam melakukan pengumpulan data, penulis menggunakan metode berikut ini :

1. Studi Literatur

   Penulis mengumpulkan data dari jurnal atau karya tulis yang relevan, hal tersebut dapat membantu penulis untuk menambah referansi sesuai dengan permasalahan.

1.6.2. Metode Implementasi

Dalam mengimplementasikan model yang penulis ajukan, maka penulis melakukan dengan beberapa tahapan :

1. Pengumpulan Data

   Pengumpulan data dilakukan dengan mencari sumber data dari sosial media untuk dilakukan mendapatkan data mentah yang selanjutnya Akan dilakukan proses pengolahan dan interpetasi data.

2. *Pre-Processing*

   Pre-Processing merupakan tahapan dimana pengolahan data awal setelah proses pengumpulan data. Pengolahan data pada tahap *pre-processing* yang dilakukan dengan mengolah data mentah dengan mulai memisahkan data yang memang tidak bisa diolah dengan data yang bisa diolah dan dilanjutkan pada proses selanjutnya.

3. Analisis

   Analisis dilakukan setelah data berhasil dilakukan *pre-processing*. Data Akan dianalisis apakah sudah layak dilakukan

implementasi dalam proses klasifikasi atau perlu melakukan pengumpulan data kembali.

4. Implementasi

   Implementasi yang dilakukan dengan melakukan label sentimen ke dalam data *training* dan *testing*.

5. Pengujian

   Pengujian dilakukan untuk mendapatkan hasil akurasi dari algoritma dalam melakukan klasifikasi.

1.7. Sistematika Penulisan

Sistematika penulisan yang dilakukan dalam penelitian ini terdiri dari beberapa berapa bagian :

BAB I PENDAHULUAN

Bab ini membahas tentang hal umum dalam penelitian, seperti latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang pengertian dan teori-teori yang dibutuhkan dalam melaksanakan penelitian ini.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan uraian secara rinci mengenai metode yang digunakan pada saat penelitian yaitu metode pengumpulan data, metode pengimplementasian dan lain sebagainya.

BAB IV IMPLEMENTASI

Bab ini membahas tentang analisis kebutuhan, perancangan danimplementasi sistem sesuai dengan metode yang digunakan dalampenelitian.

BAB V HASIL DAN PEMBAHASAN

Bab ini membahas tentang output dari analisis kebutuhan, perancangan dan implementasi yang dilakukan oleh peneliti sesuai dengan algoritma yang dipakai.

BAB VI PENUTUP

Bab ini berisikan kesimpulan dari apa yang telah dilakukan oleh peneliti dan saran dari pembaca untuk pengembangan penelitian menjadi lebih baik

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1.TINJAUAN PUSTAKA

Beberapa rujukan dari literatur berikut ini digunakan sebagai acuan dalam proses penelitian yang Akan dilakukan. Studi literature Akan bermanfaat untuk menangkap permasalahan yang mungkin muncul dan dari penyelesaian masalah dalam penelitian yang sudah pernah dilakukan sebelumnya.

Penelitian yang dilakukan oleh Oki Arifin, dkk Analisa Perbandingan Tingkat Performansi Metode *Support Vector Machine* Dan Naïve Bayes Classifier Untuk Klasifikasi Jalur Minat SMA. Dari penelitian ini menyimpulkan bahwa hasil akurasi SVM lebih tinggi dengan akurasi sebesar 96,88% dan Naïve Bayes dengan akurasi sebesar 89,63%. (Arifin & Sasongko, 2018)

Selanjutnya, penelitian dari Anisa Eka Puridewi, Jaka Nugraha. Hasil dari penelitian menyimpulkan bahwa Algoritma Naïve Bayes dan *Support Vector Machine* memiliki akurasi yang sama sebesar 0,6758 atau prosentase sebesar 67%. (Anisa Eka Puridewi, 2018)

Umi Rofiqoh dkk juga pernah menuliskan tentang SVM dan Lexicon Based Features pada klasifikasi sentiment layanan telekomunikasi seluler di media sosial twitter. Penelitian ini menghasilkan akurasi terbesar dengan skema tanpa lexicon based features dengan hasil sebesar 84%, precision 76%, Recall 86% dan F-Measure 80%. (Rofiqoh et al., 2017)

Rujukan yang keempat adalah dari Rian Tineges dkk yang menerapkan metode *Support Vector Machine*. Pengambilan data dilakukan pada media sosial twitter menggunakan API twitter pada bagasa pemrograman python. Hasil pengujian di penelitian ini adalah akurasu sebesar 87%, precision sebesar 86%, recall 95%, error rate 13%, dan f1-score 90%. (Tineges et al., 2020)

Penerapan metode SVM dalam melakukan analisis sentiment yang dituliskan Fadholi Fatharanto, mengambil data dari sosial media twitter. Data diambil pada 2 pelanggan layanan internet rumah, dimana menghasilkan Algoritma *Support Vector Machine* mendapatkan hasil persentase nilai positif sebesar 41,2% dan negatif sebesar 58,8% untuk Telkom dan persentase nilai positif sebesar 35,2% dan negatif sebesar 64,8% untuk Biznet. (Sari & Haranto, 2019)

Rujukan berikutnya mengambil data dari hasil kuisioner evaluasi dosen dimana penerapan algoritma *Support Vector Machine* digunakan untuk melakukan klasifikasi sentiment yang ditimbulkan. Akurasi tertinggi SVM pada sistem ini yaitu 67,83%. Akurasi tertinggi dicapai pada sistem yang tidak menerapkan perubahan bobot pada dokumen uji dan menggunakan range > 0 untuk kelas sentimen positif, < 0 untuk kelas sentimen negatif, dan 0 untuk kelas sentimen netral, serta pelatihan menggunakan set datalatih ke 6.(Santoso et al., 2017)

Penelitian referensi ke 7 dari Sri Widaningsih yang membuat perbandingan prediksi waktu kelulusan menggunakan metode naïve bayes, c.45, svm, dan knn. Data yang digunakan adalah nilai indeks prestasi semester 3-6 (IPS 3-6). Hasilnya adalah Algoritma Naïve Bayes memiliki akurasi tertinggi sebesar 76,79% diikuti SVM, C4.5 dan KNN. (Widaningsih, 2019)

Jurnal rujukan berikutnya dari Arsya Monica Pravina yang melakukan analisis sentiment mengenai opini maskapai penerbangan di media sosial twitter dengan menerapkan algoritma SVM. Hasil akurasi yang dihasilkan sebesar 40%, precision 40%, 100% recall, dan f-measure sebesar 57,14%. (Arsya Monica Pravina, Imam Cholissodin, 2019)

Tulisan dari Mohammad Rezwanul Huq mengenai perbandingan KNN dan SVM yang menggunakan data dari sosial media twitter menghasilkan kesimpulan hasil perbandingan algoritma pengklasifikasi (SCA) berkinerja lebih baik daripada SVM. (Rezwanul et al., 2017)

Rujukan dari tulisan munir Ahmad dkk yang menggunakan 2 dataset yaitu Twitter dataset for self-driving cars and Twitter dataset for Apple products. Metode

yang diterapkan dalam klasifikasi ini adalah metode SVM dimana hasilnya untuk dataset pertama file presisi rata-rata, recall dan f-measure 55,8%, 59,9% dan 57,2%. Untuk dataset kedua, Presisi rata-rata, Recall dan F-Measure adalah 70,2%, 71,2% dan 69,9%. (Ahmad & Ali, 2017)

Rujukan yang berikutnya adalah perbandingan dari metode naïve bayes dan *Support Vector Machine (SVM)* yang menghasilkan kesimpulan bahwa Persentase akurasi tertinggi diperoleh SVM Polynomial, sedangkan Naive Bayes menghasilkan persentase akurasi terendah (Fibrianda & Bhawiyuga, 2018).

Komparasi serupa juga pernah dilakukan oleh Imam riadi dkk yang melakukan perbandingan antara *Naïve Bayes* dan *Support Vector Machine (SVM)* dimana penelitian ini menghasilkan kesimpulan *Support Vector Machine* memiliki probabilitas 0,8 dimana nilainya lebih tinggi dari *Naïve Bayes* sebesar 0,1 (Riadi et al., 2019).

Sementara penelitian penerapan metode *Support Vector Machine (SVM)* pernah dituliskan oleh Arsya Monica Pravina dkk yang dapat menarik kesimpulan di akhir penelitian yaitu dengandigunakan Lexicon Based Features dengan iterasi sebanyak 50 kali memberikan hasil accuracy sebesar 40% (Arsya Monica Pravina, Imam Cholissodin, 2019).

Dimas Joko Hardyanto pernah juga menerapkan *Support Vector Machine (SVM)* pada kasus review barang dalam bahasa Indonesia, dimana kesimpulan yang dapat diambil adalah berdasarkan hasil pengujian,diperoleh akurasi metode *Support Vector Machine* dan Query Expansion sebesar 96,25% dan akurasi menggunakan metode *Support Vector Machine* tanpa Query Expansion sebesar 94,75% (Haryanto et al., 2018).

Analisis sentimen menggunakan *Support Vector Machine (SVM)* yang dilakukan oleh Wanda Athira Luqyana dkk menghasilkan kesimpulan akhir didapatkan parameter terbaik pada metode SVM yaitu dengan nilai degree kernel polynomial sebesar 2 nilai learning rate sebesar 0,0001, dan jumlah iterasi maksimum yang digunakan adalah 200 kali (Luqyana et al., 2018).

Masih dalam penerapan metode *Support Vector Machine (SVM)* yang dilakukan oleh Muhammad Rangga Aziz Nasution dkk pada tahun 2019 yang menyatakan bahwa *Support Vector Machine* lebih unggul dengan nilai 89,70% tanpa K-Fold Cross Validation dan 88,76% dengan K-Fold Cross Validation (Nasution & Hayaty, 2019).

Selanjutnya adalah komparasi antara Naïve Bayes dan *Support Vector Machine (SVM)* yang dituliskan oleh Hennie Thuteru pada Jurnal Pengembangan IT yang menunjukan rata-rata tingkat akurasi metode klasifikasi SVM yang lebih baik dari pada metode NBC, yaitu sebesar 76.42% (Tuhuteru & Iriani, 2018).

Support Vector Machine (SVM) yang digunakan oleh Ferdi Alvianda dkk pada analisis sentiment konten radikal menggunakan media sosial twitter menghasilkan sebuah kesimpulan akhir berupa tingkat akurasi tertinggi yang dihasilkan dari penelitian ini adalah 70% (Alvianda & Adikara, 2019).

Pada media sosial twitter juga dijadikan sumber data bagi penelitian Ahmad Choirun Najib dkk dengan menerapkan metode *Support Vector Machine (SVM)* menyimpulkan bahwa akurasi yang diperoleh berdasarkan metode Lexicon-based adalah 39% dan metode SVM sebesar 83% (Najib et al., 2019).

Terakhir dari Digna Tata Lukmana dkk yang melakukan analisis sentiment menggunakan metode *Support Vector Machine (SVM)* pada media sosial twitter dengan topik sentiment calon presiden 2019. Penelitian ini menuliskan bahwa hasil dari klasifikasi didapatkan akurasi sebesar 86%.

Adapun detail dari rujukan penelitian ditunjukan pada Tabel 2.1

Tabel 2.1 Review Penelitian Relevan

No	Judul	Objek	Penelitian, Publikasi, Tahun	Metode	Tools	Hasil
1	Analisa Perbandingan Tingkat Performansi Metode *Support Vector Machine* Dan Naïve Bayes Classifier Untuk Klasifikasi Jalur Minat SMA (Arifin & Sasongko, 2018)	Data Siswa SMA Tahun 2013-2014	Arifin, Theopilus Bayu Sasongko, Seminar Nasional Teknologi Informasi dan Multimedia, 2018	*Support Vector Machine* dan Naïve Bayes	-	Hasil akurasi SVM lebih tinggi dengan akurasi sebesar 96,88% dan Naïve Bayes dengan akurasi sebesar 89,63%
2	Perbandingan Metode Naive Bayes, *Support Vector Machine* Dan Id3 Dalam Penetapan Status Penanganan Kecelakaan Kerja (Anisa Eka Puridewi, 2018)	Data Kecelakaan Kerja	Anisa Eka Puridewi, Jaka Nugraha, Seminar Nasional Matematika dan Pendidikan Matematika, 2018	Naive Bayes, *Support Vector Machine* Dan Id3	-	Algoritma Naïve Bayes dan *Support Vector Machine* memiliki akurasi yang sama sebesar 0,6758.

3	Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features (Rofiqoh et al., 2017)	Data Layanan Telekomunikasi seluler pada Twitter	Umi Rofiqoh, Rizal Setya Perdana, M. Ali Fauzi, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2017	SVM dan Lexicon Based Features	-	Hasil terbaik pada skema tanpa lexicon based features dengan akurasi sebesar 84%, precision 76%, recall 86% dan F-Measure 80%.
4	Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi *Support Vector Machine* (Tineges et al., 2020)	Data tweet user indihome	Rian Tineges, Agung Triayudi, Ira Diana Sholihati, Jurnal Media Informatika Budidarma, 2020	*Support Vector Machine*	Python	Setelah dilakukan pengujian dengan metode SVM hasilnya adalah accuracy87%, precision 86%, recall 95%, error rate 13%, dan f1-score 90%.

5	Implementasi *Support Vector Machine* Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet (Sari & Haranto, 2019)	Tweet pengguna Telkom dan Biznet	Fadholi Fat Haranto, Bety Wulan Sari, Jurnal PILAR Nusa Mandiri, 2019	*Support Vector Machine*	-	Algoritma *Support Vector Machine* mendapatkan hasil persentase nilai positif sebesar 41,2% dan negatif sebesar 58,8% untuk Telkom dan persentase nilai positif sebesar 35,2% dan negatif sebesar 64,8% untuk Biznet
6	Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode *Support Vector Machine* (Santoso et al., 2017)	Hasil kuesioner untuk evaluasi dosen	Valonia Inge Santoso, Gloria Virginia, Yuan Lukito, Jurnal Transformatika, 2017	*Support Vector Machine*	-	Akurasi tertinggi SVM pada sistem ini yaitu 67,83%. Akurasi tertinggi dicapai pada sistem yang tidak menerapkan perubahan bobot pada dokumen uji dan menggunakan range > 0 untuk

						kelas sentimen positif, < 0 untuk kelas sentimen negatif, dan 0 untuk kelas sentimen netral, serta pelatihan menggunakan set datalatih ke 6.
7	Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, Knn, Dan Svm (Widaningsih, 2019)	Data mahasiswa Teknik Informatika pada tahun 2008 hingga tahun 2013.	Sri Widaningsih, Jurnal Tekno Instentif, 2019	C4.5, Naïve Bayes, Knn, Dan Svm	-	Naïve Bayes memiliki akurasi tertinggi sebesar 76,79% diikuti SVM, C4.5 dan KNN.
8	Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen	Tweet user maskapai penerbangan	Arsya Monica Pravina , Imam Cholissodin , Putra Pandu Adikara, Jurnal	*Support Vector Machine*		Hasil accuracy sebesar 40%, precision 40%, 100% recall, dan f-measure sebesar 57,14%.

	Twitter Menggunakan Algoritme *Support Vector Machine (SVM)* (Arsya Monica Pravina, Imam Cholissodin, 2019)		Pengembangan Teknologi Informasi dan Ilmu Komputer, 2019			
9	Sentiment Analysis on Twitter Data using KNN and SVM (Rezwanul et al., 2017)	Data Tweet	Mohammad Rezwanul Huq, Ahmad Ali, Anika Rahman, (IJACSA) International Journal of Advanced Computer Science and Applications,2017	KNN and SVM	-	Membangun model yang sederhana, beberapa fitur seperti fitur n-gram, fitur pola, fitur tanda baca, berbasis kata kunci. Hasil perbandingan algoritma pengklasifikasi (SCA) berkinerja lebih baik daripada SVM.
10	Sentiment Analysis of Tweets using SVM (Ahmad & Ali, 2017)mun	Twitter dataset for self-driving cars and Twitter dataset for Apple products	Munir Ahmad, Shabib Aftab, Iftikhar Ali, International Journal of	*Support Vector Machine*	Weka	Hasil diukur dalam hal presisi,

			Computer Applications, 2017			recall dan f-measure. Menurut hasil, untuk dataset pertama file presisi rata-rata, recall dan f-measure 55,8%, 59,9% dan 57,2% masing-masing. Untuk set data kedua, Presisi rata-rata, Recall dan F-Measure adalah 70,2%, 71,2% dan 69,9% masing-masing.
11	Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan *Support Vector Machine (SVM)*	Jaringan testbed	Mercury Fluorida Fibrianda, Adhitya Bhawiyuga, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2018	Naïve Bayes dan *Support Vector Machine*	Python	Persentase akurasi tertinggi diperoleh SVM Polynomial, sedangkan Naive Bayes menghasilkan persentase akurasi terendah.

	(Fibrianda & Bhawiyuga, 2018)					
12	ANALISIS PERBANDINGAN DETECTION TRAFFIC ANOMALY DENGAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (SVM) (Riadi et al., 2019)	Data traffic anomaly	Imam Riadi, Rusydi Umar, Fadhilah Dhinur Aini, Ilkom Jurnal Ilmiah, 2019	Naïve Bayes dan *Support Vector Machine (SVM)*	-	Hasil Naïve Bayes melalui sampel data grafik Distributions dan Radviz memiliki nilai probabilitas 0.1 dan nilai probabilitas paling tinggi yaitu 0.8. Untuk *Support Vector Machine (SVM)* menghasilkan grafik yang memiliki lebih besar nilai akurasinya.
13	Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme *Support Vector*	Komentar sosial media	Arsya Monica Pravina ,Imam Cholissodin ,Putra Pandu Adikara, Jurnal Pengembangan	*Support Vector Machine (SVM)*	-	Dengan digunakan parameter C bernilai 10 dan learning rate bernilai 0,03 serta

	Machine (SVM) (Arsya Monica Pravina, Imam Cholissodin, 2019)		Teknologi Informasi dan Ilmu Komputer, 2019			digunakan Lexicon Based Features dengan iterasi sebanyak 50 kali memberikan hasil accuracy sebesar 40%, precision 40%, 100% recall, dan f-measure sebesar 57,14%.
14	Analisis Sentimen Review Barang Berbahasa Indonesia Dengan Metode *Support Vector Machine* Dan Query Expansion (Haryanto et al., 2018)	Data review pelanggan	Dimas Joko Haryanto , Lailil Muflikhah , Mochammad Ali Fauzi, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2018	*Support Vector Machine (SVM)*	-	Berdasarkan hasil pengujian, diperoleh akurasi metode *Support Vector Machine* dan Query Expansion sebesar 96,25% dan akurasi menggunakan metode *Support Vector Machine* tanpa Query Expansion sebesar 94,75%.

15	Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi *Support Vector Machine* (Luqyana et al., 2018)	Data cyberbullying.	Wanda Athira Luqyana , Imam Cholissodin , Rizal Setya Perdana, Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 2018	*Support Vector Machine (SVM)*	-	Berdasarkan pengujian yang dilakukan didapatkan parameter terbaik pada metode SVM yaitu dengan nilai degree kernel polynomial sebesar 2, nilai learning rate sebesar 0,0001, dan jumlah iterasi maksimum yang digunakan adalah 200 kali. Dari pengujian tersebut didapatkan hasil akurasi tertinggi sebesar 90% pada komposisi data latih 50% dan komposisi data uji 50%.

16	Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter (Nasution & Hayaty, 2019)	Data Media Sosial Twitter	Muhammad Rangga Aziz Nasution, Mardhiya Hayaty, Jurnal Informatika, 2019	*K-Nearest Neighbour, Support Vector Machine (SVM)*	-	Perhitungan akurasi menunjukkan bahwa metode *Support Vector Machine* lebih unggul dengan nilai 89,70% tanpa K-Fold Cross Validation dan 88,76% dengan K-Fold Cross Validation. Sedangkan pada perhitungan waktu proses metode K-Nearest Neighbor lebih unggul dengan waktu proses 0.0160s tanpa K-Fold Cross Validation dan 0.1505s dengan K-Fold Cross Validation.

17	Analisis Sentimen Perusahaan Listrik Negara Cabang Ambon Menggunakan Metode *Support Vector Machine* dan Naive Bayes Classifier (Tuhuteru & Iriani, 2018)	Data Media Sosial Twitter	Hennie Tuhuteru , Ade Iriani, Jurnal Pengembangan IT, 2018	Naïve Bayes dan *Support Vector Machine (SVM)*	-	Penelitian ini menunjukan rata-rata tingkat akurasi metode klasifikasi SVM yang lebih baik dari pada metode NBC, yaitu sebesar 76.42%.
18	Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode *Support Vector Machine (SVM)*(Alvianda & Adikara, 2019)	Data Media Sosial Twitter	Ferdi Alvianda , Indriati , Putra Pandu Adikara , Jurnal Pengembangan IT, 2019	*Support Vector Machine (SVM)*	-	Tingkat akurasi tertinggi yang dihasilkan dari penelitian ini adalah 70% dengan nilai parameter λ sebesar 0,1, nilai konstanta γ sebesar 0,1, iterasi maksimum 5 dengan data latih sebanyak 80 dokumen (60 dokumen negatif dan 20

						dokumen positif) untuk data latih dan 20 dokumen (15 dokumen negatif dan 5 dokumen positif) untuk data uji.
19	Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter (Najib et al., 2019)	Data Media Sosial Twitter	Ahmad Choirun Najib , Akhmad Irsyad , Ghiffari Assamar Qandi , Nur Aini Rakhmawati, Fountain of Informatics Journal, 2019.	*Support Vector Machine (SVM)*	-	Berdasarkan hasil penelitian yang dilakukan akurasi yang diperoleh berdasarkan metode Lexicon-based adalah 39% dan metode SVM sebesar 83%. Dari penelitian ini diketahui bahwa SVM mempunyai performa yang lebih baik dibandingkan dengan Lexicon-based. Hasil

						Lexicon-based menunjukkan bahwa sentimen pada mayoritas atribut berupa netral.
20	ANALISIS SENTIMEN TERHADAP CALON PRESIDEN 2019 DENGAN SUPPORT VECTOR MACHINE DI TWITTER (Lukmana et al., 2019)	Data Media Sosial Twitter	Digna Tata Lukmana, Sri Subanti, Yuliana Susanti, Seminar Nasional Penelitian Pendidikan Matematika, 2019	*Support Vector Machine (SVM)*	R Studio	Penambahan fungsi Kernel pada Support Vector Machine berguna untuk mengatasi data yang tidak terpisah secara linier. Hasil dari klasifikasi didapatkan akurasi sebesar 86%.

2.1. LANDASAN TEORI

2.2.1. Data Mining

Data Mining merupakan proses untuk menemukan hubungan atau gambaran dari ratusan atau ribuan *field* dari sebuah relasional basis data yang besar. *Data Mining* juga sering disebut sebagai serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini belum diketahui. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data. *Data Mining* terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases*. (Hasan, 2017)

Fungsi-fungsi dalam *data mining* terdapat 6 fungsi yaitu : (Masripah, 2015)

1. **Fungsi deskripsi (d*escription*)**

   Fungsi deskripsi merupakan langkah menggambarkan sekumpulan data secara ringkas. Banyak cara yang dapat digunakan untuk memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya. Contoh dari penggambaran fungsi deksripsi yaitu Deskripsi Grafis, Deskripsi Lokasi, dan Deskripsi Keragaman.

2. **Fungsi estimasi (*estimation*)**

   Fungsi estimasi diidentifikasi untuk membuat perkiraan suatu hal yang sudah memiliki data. Fungsi estimasi terdiri dari dua cara yaitu Estimasi Titik dan Estimasi Selang Kepercayaan.

3. **Fungsi prediksi (*prediction*)**

Fungsi prediksi adalah memperkirakan hasil dari hal yang belum diketahui, dan digunakan untuk mendapatkan hal baru yang akan muncul selanjutnya. Contoh prediksi seperti Regresi Linier.

4. Fungsi klasifikasi (*classification*)

Fungsi klasifikasi merupakan proses menggolongkan suatu data. Contoh algoritma klasifikasi : Algoritma *Mean Vector*, Algoritma *K-nearest Neighbor*, Algoritma ID3, Algoritma C4.5, dan Algoritma C5.0

5. Fungsi pengelompokan (*cluster*)

Fungsi pengelompokan merupakan sebuah proses pengelompokan data. Data yang dikelompokan disebut objek atau catatan yang memiliki kemiripan atribut kemudian dikelompokan pada kelompok yang berbeda. Contoh algoritma yang digunakan seperti : Algoritma *Hirarchical Clustering*, Algoritma *Partitional Clustering*, Algoritma *Single Linkage*, Algoritma *Complete Linkage*, Algoritma *Average Linkage*, Algoritma *K- Means*.

6. Fungsi asosiasi (*association*)

Fungsi asosiasi difungsikan untuk menemukan aturan asosiasi (*association rule*) yang berguna dalam mengidentifikasi item-item yang menjadi objek. Algoritma yang digunakan seperti algoritma *Generalized Association Rules*, *Quantitative Association Rule*, *Asynchronous Parallel Mining*.

2.2.2. Text Mining

Text Mining dapat didefinisikan sebagai penemuan informasi baru yang sebelumnya tidak diketahui, secara otomatis mengekstrak informasi dari sumber – sumber teks tak terstruktur. Atau dalam lebih singkatnya *text mining* disebut sebagai suatu proses menganalisa teks untuk mendapatkan informasi yang berguna untuk tujuan tertentu. Perbedaan mendasar dari *text mining* dan *data mining* terletak pada sumber data yang digunakan. Pada *data mining* data yang diekstrak berasal dari pola-pola tertentu dan terstruktur, sedangkan *text*

mining sumber data yang digunakan berasal dari teks yang relatif tidak testruktur karena menggunakan bahasa manusia atau biasa disebut sebagai *natural language*. Tentu ada banyak jenis dari bahasa manusia yang digunakan, sehingga menjadikan teks mining menjadi salah satu ranah tersendiri dari *data mining*. Dalam *text mining* terdapat beberapa langkah untuk memproses data teks tersebut. (Purbo, 2019)

1. **Case Folding dan Tokenizing**

   Case Folding didefinisikan sebagai sebuah proses penyeragaman kata dengan cara mengubah seluruh kata menjadi huruf kecil (*lowercase*). Hanya huruf a sampai z yang dapat diterima karakter selain huruf dihilangkan. Disisi lain, terdapat juga kata-kata tertentu yang harus sesuai dengan kaidah yang tidak bisa dilakukan penyeragaman kata seperti kata lembaga atau institusi yang selalui diawali huruf kapital dan juga nama gelar. Proses penyeragaman ini juga bergantung dari sumber data yang digunakan untuk diproses. Sedangkan, tokenizing merupakan suatu tahapan pemotongan string kata berdasarkan penyusunan kata.

2. **Filtering**

   Filtering yakni pengambilan kata-kata penting dari hasil *Tokenizing* atau sering di isitilahkan dengan proses eliminasi kata-kata sesuai dengan kaidahnya. Algoritma *stop-word removal* merupakan salah satu yang digunakan untuk melakukan tahapan *filtering.*

3. ***Stemming***

   Stemming yaitu proses untuk memecah varian kata menjadi kata dasar sesuai dengan kata yang sedang diproses. Sebagai contoh kata yang diproses adalah Bahasa Indonesia, untuk memecah varian kata menjadi kata dasar harus sesuai dengan aturan Bahasa Indonesia atau sesuai dengan KBBI. Salah satu algoritma yang digunakan adalah Nazief & Adriani.

4. **Analyzing**

Analyzing merupakan tahapan menganlisa data teks yang sedang diproses untuk menentukan kemiripan (similaritas) antar dokumen teks. Jenis metode yang dapat digunakan adalah *cosine similarity*.

2.2.3. Analisis Sentimen

Analisis sentimen biasa disebut *opinion mining* ialah salah satu bagian dari *text mining*. *Text mining* melakukan studi tentang pendapat yang muncul dari orang-orang, sentimen, evaluasi, tingkah laku dan emosi terhadap suatu entitas seperti produk, layanan, organisasi, individu, permasalahan, topik, acar. Analisis sentimen biasa diterapkan untuk menganalisis komentar-komentar di sosial media seperti *Facebook*, *Twitter* untuk selanjutnya dilakukan proses penerjemahan menjadi sesuatu yang lebih bermakna, salah satunya dalam bentuk rating. Rating menjadi sangat penting dalam dunia bisnis disebabkan rating merupakan salah satu indikator kesuksesan. Di sisi lain, rating masih mejadi komoditas monopoli beberapa perusahaan seperti Nielsen, sehingga objektivitasnya menjadi kurang. Berdasarkan hal tersebut, ini menjadi sebuah celah ini yang dapat dimanfaatkan penulis untuk mencoba mengaplikasikan analisis sentimen pada *Facebook* untuk membuat sistem rating berdasar komentar. (Monarizqa et al., 2014)

2.2.4. Klasifikasi

Klasifikasi adalah sebuah proses menemukan model atau fungsi yang membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Dalam klasifikasi terdapat dua proses, yakni proses *training* dan proses *testing*. Proses *training* menggunakan *training set* yang telah diketahui label-labelnya yang berfungsi untuk membangun model. *Testing* digunakan untuk menguji keakuratan model yang telah dibangun saat proses *training*. (Dicky Nofriansyah, S.Kom., 2017)

2.2.5. *Support Vector Machine (SVM)*

SVM merupakan metode pembelajaran terawasi yang menghasilkan pemetaan dari fungsi input-output dari sekumpulan data training. Fungsi pemetaan ini bisa berupa fungsi klasifikasi atau fungsi regresi. SVM menggunakan ruang hipotesis berupa fungsi-fungsi linier di dalam feature space. SVM dilatih menggunakan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan learning bias yang berasal dari teori pembelajaran statistik (Christianini, 2000). Teori mengenai SVM sudah berkembang sejak tahun 1960an, namun pada tahun 1992 Vapnik, Boser, dan Guyon memperkenalkan kembali teori ini dan semenjak saat itu SVM berkembang dengan pesat. Gambar dibawah ini akan digunakan oleh penulis untuk mempermudah dalam memahami metode SVM. (Soman, K. P., Loganathan, R., & Ajay, 2009)

Model persamaan SVM ditunjukan pada persamaan 1.

f:w.x+b=0 (1)

Dimana :

w = parameter hyperplane yang dicari (garis yang tegak lurus antara garis hyperplane dan titik support vector)

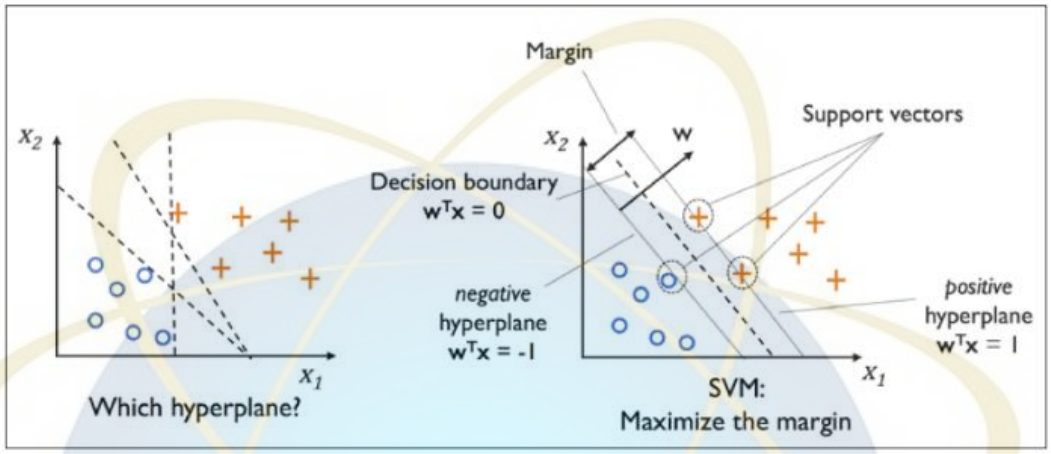
x = data input SVM (x1 = index kata, x2= bobot kata)

b = parameter hyperplane yang dicari (nilai bias)

f = fungsi Hyperplane

Support Vector Machine (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi (seperti *Support Vector Classification*) dan regresi (*Support Vector Regression*). Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara matematis dibandingkan dengan teknik-teknik klasifikasi lainnya. SVM juga dapat mengatasi masalah klasifikasi dan regresi dengan linear maupun non linear. SVM digunakan untuk mencari hyperplane terbaik dengan memaksimalkan jarak antar kelas. Hyperplane adalah sebuah fungsi yang dapat digunakan untuk pemisah antar kelas. Dalam 2-D fungsi

yang digunakan untuk klasifikasi antar kelas disebut sebagai line whereas, fungsi yang digunakan untuk klasifikasi antas kelas dalam 3-D disebut plane similarly, sedangan fungsi yang digunakan untuk klasifikasi di dalam ruang kelas dimensi yang lebih tinggi di sebut hyperplane.



Gambar 2.1. Metode SVM

Hyperplane yang ditemukan SVM diilustrasikan seperti Gambar 1 posisinya berada ditengah-tengah antara dua kelas, artinya jarak antara hyperplane dengan objek-objek data berbeda dengan kelas yang berdekatan (terluar) yang diberi tanda bulat kosong dan positif. Dalam SVM objek data terluar yang paling dekat dengan hyperplane disebut support vector. Objek yang disebut support vector paling sulit diklasifikasikan dikarenakan posisi yang hampir tumpang tindih (overlap) dengan kelas lain. Mengingat sifatnya yang kritis, hanya support vector inilah yang diperhitungkan untuk menemukan hyperplane yang paling optimal oleh SVM.

Y	X1	X2
1	1	1
-1	1	-1
-1	-1	1
-1	-1	-1

Dari contoh diatas, didapatkan plot contoh data yang dijelaskan pada Gambar 2.2.

Gambar 2.2. Metode SVM

Pada Gambar 2 menjelaskan bahwa terdapat 2 kelas yang terdiri dari -1 ditunjukkan dengan warna biru dan 1 ditunjukkan dengan warna orange. Pada masing-masing titik tersebut digunakan untuk mencari pemisah antara data positif dan data negatif. Penyelesaian sebagai berikut :

Diketahui :

Dengan syarat :

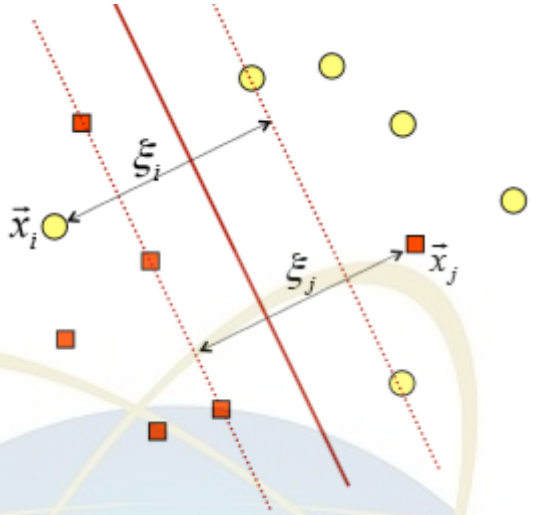
yi(xi . w +b) -1 ≥ 0, i = 1,2,3,….,n yi(x1.w1 + x2.w2 + b) ≥ 1

sehingga ditemukan persamaan sebagai berikut:

($w1 + w2 +$) ≥ 1 untuk y1 = 1, x1=1, x2=1 ($-w1 + w2 -$) ≥ 1

untuk y2 = -1, x1=1, x2=-1 ($w1 - w2 - b$) ≥ 1 untuk y3 = -1, x1=-1, x2=1 ($w1 + w2 - b$) ≥ 1 untuk y1 = -1, x1=-1, x2=-1

Berdasarkan persamaan diatas, maka didapatkan nilai dari setiap variabel. Dari persamaan 1 dan 2 didapatkan:



Gambar 2.3. SVM

Parameter C dipilih untuk mengontrol trade off antara margin dan error klasifikasi ξ atau nilai kesalahan pada klasifikasi. Parameter C ditentukan dengan mencoba beberapa nilai dan dievaluasi efeknya terhadap akurasi yang dicapai oleh SVM misalnya dengan cara Cross Validation. Nilai C yang besar berarti akan memberikan penalti yang lebih besar terhadap error klasifikasi tersebut. Pada umumnya permasalahan data tidak dapat dipisahkan secara Linear dalam ruang input, soft margin SVM tidak dapat menemukan pemisah dalam hyperplane sehingga tidak dapat memiliki akurasi yang besar dan tidak menggeneralisasi dengan baik. Oleh karena itu, dibutuhkan kernel untuk mentransformasikan data ke ruang dimensi yang lebih tinggi yang disebut ruang kernel yang berguna untuk memisahkan data secara Linear. Secara umum, fungsi kernel yang sering digunakan adalah kernel Linear, Polynomial dan Radial Basis Function (RBF).

2.2.6. Twitter

Twitter adalah salah satu layanan microblogging yang cukup terkenal dan memungkinkan para penggunanya untuk menulis atau membuat status yang sering dinamakan kicauan atau tweet. Media sosial Twitter digunakan untuk mengutarakan berbagai pendapat atau opini akan sebuah produk, layanan atau hal lainnya. Twitter diciptakan oleh Jack Dorsey di tahun 2006 dan pertama meluncur di dunia maya saat Juli 2006

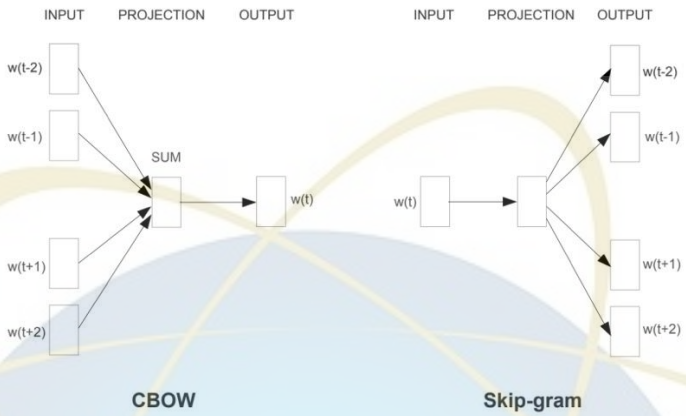
dengan alamat http://www.Twitter.com yang masih digunakan hingga saat ini. Pengguna dapat menulis pesan berdasarkan topik dengan menggunakan tanda #(hashtag). Sedangkan untuk menyebutkan atau membalas pesan dari pengguna lain bisa menggunakan tanda @.

2.2.7. Processing Text Menggunakan Word embedding

Word embeddings adalah proses konversi kata yang berupa karakter alphanumeric kedalam bentuk vector. Setiap kata adalah vector yang merepresentasikan sebuah titik pada space dengan dimensi tertentu. Dengan word embedding, kata-kata yang memiliki properti tertentu, misalnya berada pada konteks yang sama, atau memiliki semantic meaning yang sama berada tidak jauh satu sama lain pada space tersebut (Girsang, 2020).

Word2vec adalah model shallow neural network yang merubah representasi kata yang merupakan kombinasi dari karakter alphanumeric menjadi vector. Representasi vector tersebut memiliki properti relationship terhadap kata-kata yang berkaitan melalui proses training. Arsitektur neural network yang digunakan pada word2vec adalah neural network dengan sebuah hidden layer yang disebut dengan projection layer dan di-training menggunakan stochastic gradient descent dengan algoritma backpropagation. Projection layer adalah mapping dari kata yang ada pada konteks n-gram ke dalam bentuk continuous vectors. Kata-kata yang muncul bersamaan atau berulang kali pada konteks N-gram memiliki tendensi untuk teraktifasi oleh weight yang sama, sehingga terjadi korelasi antar kata-kata. Weights menghubungkan input layer dengan hidden layer, dan hidden layer dengan output layer. Weights diantara input layer dan hidden layer direpresentasikan dengan matriks W berukuran V x N, dimana V adalah dimensi dari input layer dan N adalah dimensi dari hidden layer. Sedangkan diantara hidden layer dan output layer, matriks W direpresentasikan dengan matriks berukuran N x V.

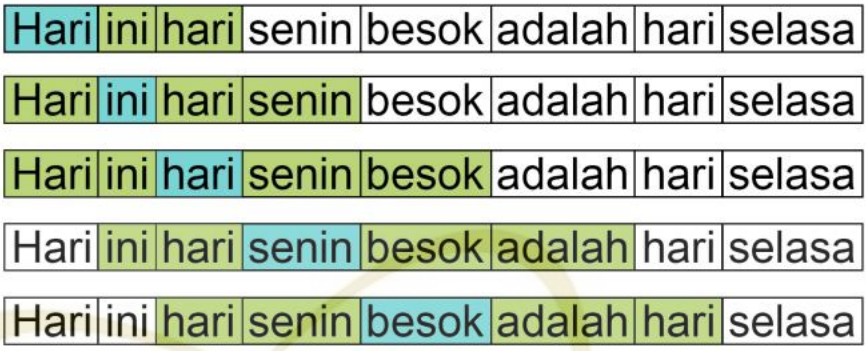
Terdapat dua model arsitektur yang dapat digunakan pada word2vec, yaitu CBOW dan Skip-Gram. Kedua model tersebut dapat dilihat pada gambar II.1 berikut:



(Girsang, 2018)

Gambar 2.4. Arsitekrur Word2vec

Pada model CBOW, word2vec menggunakan kata-kata yang ada di sebelah kiri dan kanan kata target dan dibatasi dengan window untuk memprediksi kata target tersebut. Sedangkan skip-gram menggunakan sebuah kata untuk memprediksi kata-kata yang ada di sebelah kiri dan kanan kata tersebut yang dibatasi oleh window. Masing-masing kata yang digunakan sebagai input di-encode ke dalam one-hot vector. Perbedaan dari kedua model tersebut adalah model prediksi kata. Pada CBOW, terdapat intermediate layer yang akan melakukan kalkulasi average pada vector kata-kata input karena CBOW menerima sejumlah n kata sebagai input. Window digunakan sebagai kernel untuk memperoleh input dan target words. Window digeser dari awal sampai akhir susunan kata. Ilustrasi dari window dapat dilihat pada gambar II.2 sebagai berikut:



(Girsang, 2018)

Gambar 2.5. Ilustrasi

Pada gambar II.2, ukuran dari window adalah 2. Kolom berwarna biru adalah center word, sedangkan kolom berwarna hijau adalah context words. Pada CBOW, center word digunakan sebagai target, dan context words digunakan sebagai input pada setiap iterasi. Sedangkan pada skip gram, center word digunakan sebagai target, dan context words digunakan sebagai target pada setiap iterasi.

2.2.8 Python

Python adalah bahasa pemrograman interpretatif yang dapat digunakan di berbagai platform dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode dan merupakan salah satu bahasa populer yang berkaitan dengan Data Science, Machine Learning, dan Internet of Things (IoT). Keunggulan Python yang bersifat interpretatif juga banyak digunakan untuk prototyping, scripting dalam pengelolaan infrastruktur, hingga pembuatan website berskala besar. Bahasa Python menjadi keharusan untuk Anda yang ingin mempelajari dasar-dasar scripting dan pengolahan data atau machine learning. Bahasa Python digunakan secara luas, masuk dalam 3 besar bahasa pemrograman yang digunakan dalam beberapa tahun belakangan. Pustaka (Library) yang luas, memungkinkan Anda mengembangkan ke bidang-bidang lainnya. Beberapa library atau framework terpopuler data science dan machine learning menggunakan

Python antara lain: Scikit-Learn, TensorFlow, PyTorch. Bahasa Python memiliki kurva pembelajaran (learning-curve) yang sangat landai, cocok untuk dipelajari sebagai bahasa pemrograman pertama - dengan kemudahan pembacaan dan kemudahan mempelajari sintaksisnya.

Sebenarnya, Python bukanlah bahasa pemrograman baru. Menurut Geeksforgeeks, Python sendiri sudah ada cukup lama semenjak tahun 1991. Bahasa pemrograman yang dikembangkan oleh Guido van Rossum ini terus mengalami pembaruan hingga saat ini. Nama Python sendiri diambil dari prorgam televisi favoritnya yang bernama "Monty Python Flying Circus". Python adalah bahasa pemrograman yang populer digunakan di seluruh dunia untuk mengembangkan situs web, algoritma dan menyederhanakan proses otomatisasi. Melalui bahasa pemrograman Python, setiap program akan menjadi lebih ringkas jika dibandingkan bahasa pemrograman lain. Tak hanya itu, Python bertujuan untuk menghasilkan kode yang lebih jelas dan lebih logis untuk berbagai keperluan. Proses pengkodean Python sangat sederhana sehingga memberikan keleluasaan bagi developer untuk mengembangkan fitur baru dari suatu situs atau aplikasi. Python banyak diaplikasikan pada berbagai sistem operasi seperti Linux, Microsoft Windows, Mac OS, Android, Symbian OS, Amiga, Palm dan lain-lain. Dengan kemudahan yang diberikan, tak heran Python lebih mudah dipelajari oleh pemula. Dalam perkembangannya, Python tak hanya digunakan dalam dunia teknologi, namun juga dalam hal lain khususnya analisis. Saking mudah dan uniknya Python, tak heran jika banyak raksasa teknologi juga menggunakannya.

Seperti dijelaskan sebelumnya, pentingnya Python tak hanya terbatas untuk urusan teknologi saja. Terdapat banyak hal yang dapat dipermudah dengan menerapkan bahasa pemrograman tersebut di dalamnya. Berikut adalah beberapa kegunaan Python sehingga penting untuk dipelajari.

1. Pengembangan website Dalam membangun bisnis, adanya website tentu menjadi unsur penting di dalamnya. Selain sebagai pemberi informasi

kepada calon konsumen, website juga menunjukkan kredibilitas perusahaan. Tak jarang, berbagai website untuk bisnis pun diberikan berbagai macam fitur agar pengunjung dapat lebih memahami bisnis yang dijalankan. Untuk mengembangkan suatu website agar lebih intuitif dan menarik, menggunakan Python akan mempermudah prosesnya.

2. Pengembangan IoT Hal lain yang menunjukan pentingnya Python adalah pengembangan internet of things (IoT). Internet of things adalah sebuah sistem di mana berbagai benda atau peralatan dapat berkomunikasi satu sama lain dengan piranti internet. Untuk mengembangkan hal tersebut, Python digunakan karena berbagai kemudahan dan fleksibilitasnya.

3. Penambangan data Hal lain yang termasuk dalam kegunaan Python adalah pengaturan dan pembersihan data. Python dianggap sebagai salah satu bahasa pemrograman terbaik untuk mengerjakannya. Selain itu, pembelajaran mesin dengan Python menyederhanakan analisis data dengan menggunakan algoritma.

4. Pengembangan machine learning Masih berkaitan dengan penambangan data, dalam menjalankan machine learning akan dibutuhkan beragam data untuk diinput. Beragam data yang masuk kemudian diolah untuk menjadi suatu tindakan yang dilakukan oleh mesin tersebut. Dalam prosesnya, untuk mempermudah proses ini bahasa pemrograman Python-lah yang digunakan.

5. Pengembangan game Ternyata, Python juga berguna untuk mengembangkan game yang kamu mainkan. Dalam Python, terdapat program yang bernama GUI. Antarmuka pengguna grafis (graphical user interface/GUI) memungkinkan orang untuk berinteraksi dengan komputer menggunakan elemen visual seperti ikon atau gambar alih-alih perintah berbasis teks. Inilah yang membuat game yang kamu mainkan menjadi lebih atraktif dan menantang.

6. Python untuk fintech Pentingnya Python tergambar juga dalam pengembangan fintech. Dengan menggunakan Python, aplikasi dan

berbagai fitur yang ada di fintech akan lebih aman. Selain itu, karena fintech memerlukan pengoperasian yang cepat, Python-lah yang digunakan seiring dengan kemampuannya untuk mendukung hal itu.

2.2.9 Cross Validation

Validation adalah proses untuk mengevaluasi keakuratan prediksi dari model. Validasi digunakan untuk memperoleh prediksi menggunakan model yang ada dan kemudian membandingkan hasil tersebut dengan hasil yang sudah diketahui, ini mewakili langkah paling penting dalam proses membangun sebuah model. *Cross Validation* adalah teknik validasi dengan membagi data secara acak ke dalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Dalam *Cross Validation*, jumlah tetap lkhususstan atau partisi dari data ditentukan sendiri. Cara standar untuk memprediksi error rate dari teknik pembelajaran dari sebuah sampel data tetap adalah dengan menggunakan tenfold cross validation. Dengan *cross validation*, data akan dibagi secara acak menjadi n bagian, dimana class diwakili (kurang lebih) proporsi yang sama seperti pada dataset yang penuh. Setiap bagian mendapatkan gilirannya dan skema pembelajaran dilatih pada sisa sembilan persepuluh; kemudian error rate dihitung pada *holdout set*.

Validation merupakan teknik *validasi* dengan membagi data secara acak kedalam bagian dan masing-masing bagian akan dilakukan proses klasifikasi (Ningtyas et al., 2019). *Validasi* dilakukan dalam melakukan pengukuran hasil prediksi, sedangkan *Evaluasi* digunakan dalam melakukan pengamatan dan menganalisa hasil kerja pada *RapidMiner* (Rahayu et al., 2019). Akurasi dapat diartikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual, presisi menunjukan tingkat ketepatan atau ketelitian dalam klasifikasi, sedangkan *recall* berfungsi untuk mengukur perbandingan positif aktual yang benar (Arifin & Sasongko, 2018).

Evaluasi dalam mengukur akurasi, presisi, dan *recall* biasanya digunakan *confusion matrix* (Ningtyas et al., 2019). *Confusion matrix*

merupakan alat ukur berbentuk *matrix* yang digunakan untuk menghasilkan jumlah ketepatan klasifikasi terhadap *class* dengan algoritma yang digunakan (Arifin & Sasongko, 2018), (Sulaehani, 2016).

Tabel 2.2 *Confusion matrix*

***CLASS*IFICATION**		**Predicted *Class***	
		***CLASS* = YES**	***CLASS* = NO**
OBERVED *CLASS*	***CLASS* = YES**	True *Positif* (TP)	False *Negative* (FN)
	***CLASS* = NO**	False *Positif* (FP)	True *Negative* (TN)

Sumber : (Sulaehani, 2016)

Keterangan :

True *Positif* (TP) = Perbandingan sampel bernilai *true* (benar) yang diprediksi secara benar

True Negative (TN) = Perbandingan sampel bernilai *false* (salah) yang diprediksi secara benar

False Positif (FP) = Perbandingan sampel bernilai *false* (salah) yang salah diprediksi sebagai sampel bernilai benar.

False Negative (FN) = Perbandingan sampel bernilai *true* (benar) yang salah diprediksi sebagai sampel bernilai salah.

menghitung nilai *accuracy*, *precision* , dan *recall* menggunakan rumus perhitungan sebagai berikut (Sulaehani, 2016) :

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

$$Precision = \frac{TP}{TP+FP}$$

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

Receiver Operating Characteristic (ROC) biasa digunakan untuk pembelajaran mesin dan penelitian *data mining*. Dalam masalah klasifikasi, ROC merupakan 40entim untuk memvisualisasikan, mengatur dan memilih pengklasifikasian berdasarkan pada kinerja mereka (Gorunecu, 2011). *Receiver Operating Characteristic* (ROC) adalah grafik yang dapat digunakan dalam menilai suatu model. Dalam klasifikasi *data mining* , nilai AUC dapat dibagi menjadi beberapa kelompok diantaranya (Sulaehani, 2016) :

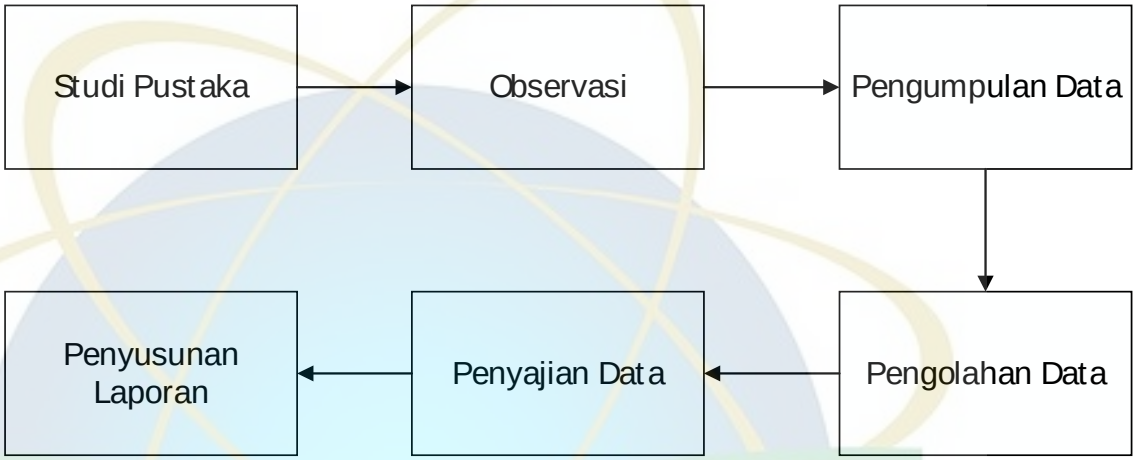
1. 0.90 – 1.00 = Excellent *class*ification
2. 0.80 – 0.90 = Good *class*ification
3. 0.70 – 0.80 = Fair *class*ification
4. 0.60 – 0.70 = Poor *class*ification
5. 0.50 – 0.60 = Failure

BAB III

METODE PENELITIAN

3.1.Alur Penelitian

Adapun alur dalam penelitian ini ditunjukan pada Gambar 3.1.



Gambar 3.1. Alur Penelitian

Penjelasan masing-masing tahapan :

1. **Studi Pustaka**

   Proses studi pustaka dilakukan untuk mencari referensi terkait penelitian yang pernah dilakukan sebelumnya, dan juga rujukan-rujukan dalam bentuk buku dan informasi-informasi mengenai kasus terkait. Hal ini dijadikan landasan untuk menemukan permasalahan yang akan diselesaikan dalam penelitian ini.

2. **Observasi**

   Observasi dilakukan dengan mengamati tren dari objek penelitian di sosial media. Pengamatan dilakukan dengan memperhatikan intensitas interaksi yang terjadi dan akan dilakukan penelitian lebih jauh dengan menganalisis sentiment yang muncul.

3. **Pengumpulan Data**

   Pengumpulan data yang dilakukan adalah dengan melakukan pengambilan data di sosial media, khususnya *twitter* dan kaggle disini yang diambil untuk menjadi studi kasus.

4. **Pengolahan Data**

   Pengolahan data dilakukan setelah proses pengambilan data di sosial media berhasil dilakukan. Pada proses pengolahan data sendiri dilakukan proses pembersihan data, agar data yang diolah benar-benar data bersih bukan data mentah (data kotor).

5. **Penyajian Data**

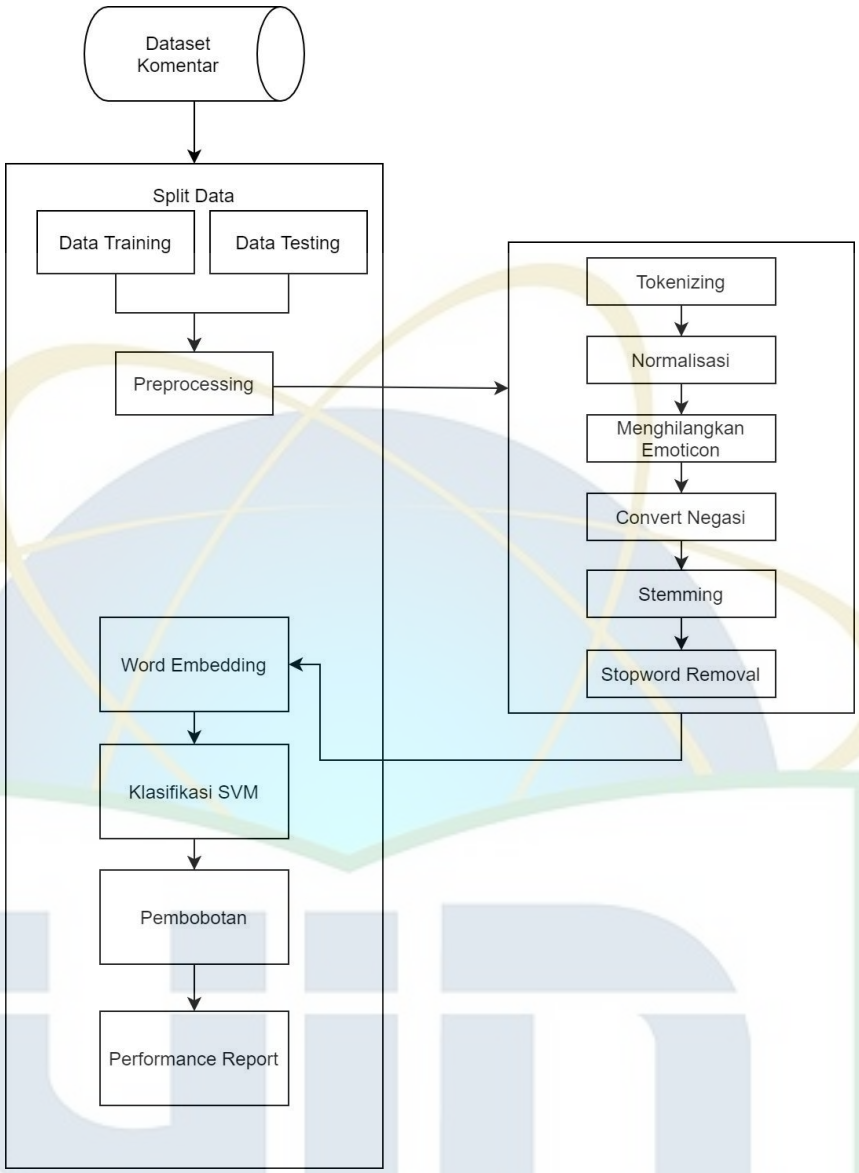
   Penyajian data dilakukan untuk menampilkan hasil pengolahan data. Data yang sudah diolah dengan data *cleaning, preprocessing* bisa dilakukan penyajian dalam bentuk informasi yang baru dan berguna. Penyajian data membutuhkan bantuan algoritma untuk melakukan klasifikasi sentiment yang muncul. Data yang sudah siap akan dilakukan proses pelabelan data, kemudian dilakukan klasifikasi berapa persen dari masing-masing sentiment yang dihasilkan, sehingga menghasilkan data dan informasi yang baru.

6. **Penyusunan Laporan**

   Setelah semua tahapan penelitian dilakukan, yang terakhir adalah menyusunnya ke dalam laporan penelitian. Laporan ini berisikan tahapan dari awal pengumpulan, pengolahan, penyajian data hingga penarikan kesimpulan dari penelitian yang dilakukan.

3.2.Metode Usulan

Pada penelitian ini peneliti mengusulkan suatu metode dalam analisis sentimen guna mengetahui sentiment tanggapan pemberian vaksin, dengan menggunakan metode klasifikasi *Support Vector Machine* serta menggunakan pemrosesan data awal dengan beberapa tahapan yang perlu dilalui. Adapun model yang ditawarkan ditunjukan pada



Gambar 3.2. Model Usulan

Adapun penjelasan dari model yang diusulkan adalah sebagai berikut :

1. **Data Awal Komentar**

   Data awal didapatkan dengan mencari data di sosial media. Saat ini sosial media bisa menjadi sumber data mentah yang dapat dilakukan banyak intepretasi hasil bermacam-macam. Data awal yang digunakan adalah data komentar di sosial media twitter. Data komentar atau biasa

disebut cuitan dari pengguna twitter tentu bermacam-macam jenis atau topik pembicaraannya. Untuk mempersempit area data yang akan diambil datanya maka perlu ditentukan topik apa yang akan diambil. Untuk kasus pada tema ini adalah mengenai sentimen penggunaan vaksin, sentimen penggunaan atau penyuntikan vaksin menjadi tema atau topik yang akan digali. Sehingga dapat menggunakan kata kunci sesuai dengan yang dibutuhkan seperti vaksinasi, vaksincovid dan sebagainya. Dataset bersumber dari Kaggle Repository.

2. **Split Data**

   Proses pertama setelah datanya ditemukan, maka dilakukan split data. Hal ini bertujuan untuk membagi dataset kedalam data training dan data testing. Pembagian dapat dilakukan dengan membagi 80% data training dan 20% data testing atau 70% data training dan 30% data testing lalu akan ditentukan mana pembagian data yang paling tepat berdasarkan akurasi tertinggi.

3. **Preprocessing**

   Preprocessing merupakan tahapan untuk menghilangkan elemen yang tidak diperlukan pada dataset seperti imbuhan, symbol, dan kata negasi. Preprocessing yang digunakan adalah tokenisasi, normalisasi, menghilangkan emoticon, convert negasi, *stemming*, dan stopword removal.

4. **Klasifikasi**

   klasifikasi disini dilakukan dengan menganalisis kata per kata atau dalam kalimat utuh untuk mendapatkan hasil dari kelas sentimen yang ada, jika kata cenderung menyudutkan atau berkonotasi buruk seperti kata payah, buruk dan sebagainya adalah salah satu kata yang dapat membuat sebuah kalimat menjadi memiliki sentiment negatif, atau sebaliknya jika dia memiliki pujian seperti bagus, mantap dan

sebagainya akan mudah dilakukan di kelompokan sebagai data kelas sentimen positif.

Setelah dilakukan proses pelabelan data, maka data yang sudah dilabeli Akan dilakukan klasifikasi menggunakan metode *Support Vector Machine (SVM)* metode ini Akan menemukan akurasi seberapa besar akurasi yang didapatkan dari pengelompokan yang dilakukan oleh metode. Dalam skema ujinya Akan dilakukan pengacakan kedalam bagian sebagai data testing dan juga data training. Setelah dikelompokan Akan mudah dalam mengenali sejauh mana klasifikasi yang benar dilakukan oleh algoritma.

5. **Performance Report**

   Interpretasi hasil dengan membaca dari serangkaian pengujian dan proses klasifikasi yang dilakukan dan menghasilkan akurasi, hal itu akan membuat penulis dapat menyimpulkan sejauh mana algoritma dapat bekerja dalam melakukan klasifikasi. Setelah melalui serangkaian uji coba, atau tahapan pengetesan, Akan didapatkan hasil untuk dilakukan pembacaan data, dan penarikan kesimpulan.

3.3.Pengujian

Data yang sudah diolah dan dilakukan analisis sentimen, pada tahap selanjutnya Akan di lakukan pengujian. Pengujian merupakan tahapan dimana sistem Akan dijalankan. Tahap pengujian diperlukan sebagai ukuran bahwa sistem dapat dijalankan sesuai dengan tujuan. Pengujian sistem analisis sentimen Twitter ini dilakukan dengan model pengujian akurasi dimana Akan dicari nilai dari pengujian klasifikasi mengenai nilai F-Measure, Precision, dan Recall. Selain itu pengujian menggunakan cross validation yang merupakan suatu metode tambahan dari teknik data mining yang bertujuan untuk memperoleh hasil akurasi yang maksimal. Metode ini sering juga disebut dengan k-fold cross validation dimana percobaan sebanyak k kali untuk satu model dengan parameter yang sama

BAB IV

IMPLEMENTASI EKSPERIMEN

4.1.Pengambilan Data

Data yang diambil adalah data dari sosial media *Twitter*. Sosial media *Twitter* merupakan salah satu sosial media berbasis teks yang memiliki banyak pengguna aktif. Pengambilan data dilakukan dengan mengambil data dari sumber public yaitu Kaggle repository dengan url berikut: *https://www.kaggle.com/rpnugroho/indonesian-vaccination-tweets*. Dataset yang digunakan terdiri dari 30.000 record dengan 16 atribut. Atribut dalam data tersebut seperti pada Tabel 4.1.

Table 4. 1
Deskripsi Atribut *Dataset*

No	Atribut	Deskripsi
1	Id	Id tweet
2	Date	Tanggal posting
3	*Text*	Isi narasi tanggapan tentang vaksin sinovac
4	*Hastag*	*Hastag* pada postingan
5	user_name	Orang yang melakukan posting
6	User location	Lokasi user saat memposting tweet
7	User description	Deskripsi user
8	User created	Tanggal
9	User followers	Jumlah followers user
10	User friends	Jumlah following user
11	User favorites	Jumlah tweet yang difavoritkan oleh user
12	Retweet	Jumlah postingan di retweet
13	Favorite	Jumlah postingan di favoritkan

14	Source	Sumber tweet yang diposting (Android, Web, Iphone dll)
15	Is retweet	Apakah diretweet? TRUE/False
16	Reply To Status	Jumlah reply status

4.2.Pemrosesan Awal

Pada tahapan ini, membutuhkan *eksplorasi* atau pendalaman terhadap *dataset*. *Eksplorasi* dilakukan dengan tujuan untuk menunjukan pada semua atribut dan *class* dalam *dataset* tersebut *valid,* sehingga bisa digunakan untuk objek penelitian yang baik. Maka dari itu, tujuan untuk mengetahui hasil analisis sentimen terbaik pada *tweet*.

1. *Data Tranformation*

Dataset yang digunakan memiliki 16 atribut original dari sumbernya, akan tetapi tidak semua atribut atau *features* tersebut akan digunakan karena terdapat *features* yang tidak dapat membantu dalam proses *sentiment analysis* ini sehingga perlu dilakukan *feature selection*. Dari 16 atribut, fitur yang akan digunakan hanya berjumlah 3 atribut yaitu seperti pada Tabel 4.2.

Table 4 2.
Atribut *Dataset* Yang Digunakan

No	Atribut	Deskripsi
1	Id	Id tweet
2	*Text*	Isi narasi tanggapan tentang vaksin sinovac

Selain menghilangkan feature yang tidak penting, diperlukan juga *preprocessing text* untuk merubah *dataset* yang masih original dan masih dalam keadaan kotor dan belum siap dilakukan klasifikasi. Adapun bisa dilakukan klasifikasi, memungkinkan akurasinya rendah.

2. *Sampling*

Untuk pengujian model yang digunakan, data akan dibagi menjadi dua bagian antara lain *data training* dan *data testing* dengan menggunakan *Split Validation*. Dengan besaran pembagian data yaitu 70% untuk *data training* dan 30% untuk *data testing*. *Data training* ini untuk pengembangan pada model dan *data testing* ini untuk pengujian model.

4.3. Perangkat Lunak yang Digunakan

Pada pembuatan simulasi untuk implementasi *SVM* pada kasus sentiment analisis ini memerlukan beberapa perangkat lunak sebagai berikut:

1. Sistem Operasi *Windows* 10 Pro

Proses pembuatan sistem ini dilakukan melalui *windows 10.*

2. Jupyter Notebook

Jupyter Notebook digunakan untuk melakukan penulisan program, dan pengaturan – pengaturan *source code* di Bahasa pemrograman *python* melalui sistem operasi *windows* 10 Pro.

***3. Browser* (Google Chrome)**

Browser digunakan untuk pengetesan dan pengecheckkan implementasi model.

4. Python

Proses pembuatan aplikasi *IT Helpdesk* dengan pengkodingan yang berbahasa pemrograman *Python* 3.7.

a. Library

Pada implementasi model *Support Vector Machine* dan *Word embedding* untuk kasus analisis *sentiment* tanggapan tentang vaksin sinovac ini, dibutuhkan *library* yang mendukung. Pada penelitian ini, akan dibuat sebuah simulasi dalam menguji model *Support Vector Machine* untuk menghasilkan akurasi yang baik dengan menggunakan Bahasa pemrograman *python* dan *library* seperti pada Tabel 4.3.

Tabel **4.3**.
Teknologi Yang Digunakan

No	**Nama *Library***	**Kegunaan**
1	Numpy	Membaca *dataset*, memproses *dataset*
2	NLTK	Melakukan *preprocessing* seperti tokenizer dan yang lainnya
3	Keras	*Library* untuk word *SVM*
4	Tensorflow	*Library* untuk implementasi *SVM*
5	Sastrawi	
6	Seaborn	*Library* untuk memunculkan grafik
7	Gensim	*Library* untuk *Fasttext* (*Word embedding*) dan melabeli positif maupun negatif

4.4. Cara Kerja Model

Proses dari model yang diusulkan sesuai dengan Gambar 3.2. berikut penjelasannya yang diimplementasikan dengan teknologi pada Tabel 4.3.

A. Dataset

Dataset yang digunakan pada penelitian ini disimpan dalam format *Comma Separated Values* (CSV) dan berjumlah 30.000 *dataset*, sebagai contoh berikut ditampilkan Sebagian *dataset* pada Tabel 4.4. dan lebih lengkapnya terdapat di Lampiran A1.

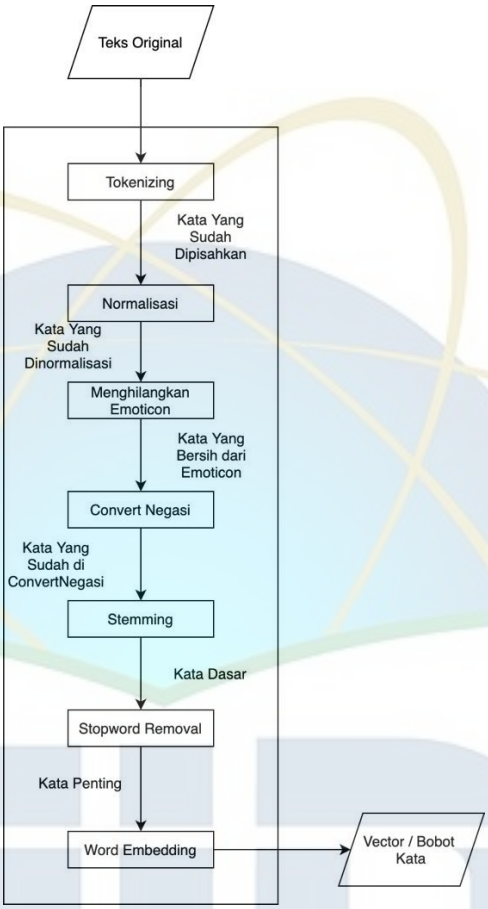
Tabel 4.4.
Dataset

Id	*Text*
1348286903527768065	#vaksin untuk #indonesia
1348302168248340481	Gak Kenal maka Gak Kebal. Vaksin jadi bukti manusia berjuang menjaga eksistensinya di dunia ini.
1348307055040348160	Sistem satu data mendukung Vaksinasi di Indonesia #Vaksin #Covid19 #CoronaVirus #Kesehatan
1348322660925669377	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…
1348425614584266753	Apa BedanyaVaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…
1348427123594207233	Dua Lagi Obat yang Dapat NyelamatinNyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…
1348434452775657475	Vaksin bikinanTiongkok kembali dipertanyakan, orang tua gakboleh suntik!
1348434452775657337	Walaupun di vaksin, tetangga guemasih terkena covid juga tuh #vaksingagal

B. Tahap *Preprocessing*

Tahap ini dilakukan untuk membersihkan data. *Dataset* yang kita ambil dari repository dengan bentuk original tentunya belum tentu siap untuk digunakan

untuk membedakan mana tweet positif, *negatif* dan netral. Data kotor tersebut seperti terdapat teks kosong, teks duplikat, dan kata yang mempunyai multi penafsiran. Berikut merupakan tahap *preprocessing text* pada penelitian ini:



Gambar 4.1. Proses *Preprocessing*

1. *Tokenizing*

Tokenizing merupakan sebuah upaya dalam melakukan persiapan *dataset* untuk diklasifikasi. Proses ini adalah memisahkan setiap kata pada sebuah kalimat dengan sebuah pemisah yang dapat berupa spasi, koma, titik koma maupun titik. Pada penelitian ini, delimiter yang digunakan adalah spasi. Sehingga sesuai dengan Tabel 4.4 maka hasil setelah teks dilakukan *tokenizing* seperti pada Tabel 4.5.

Tabel 4.5.
Hasil *Tokenizing*

Id	*Text*	Hasil *Tokenizing*
1348286903527768065	#vaksin untuk #indonesia ☺	#vaksin

		Untuk Indonesia ☺
1348302168248340481	Gak Kenal maka Gak Kebal. Vaksin jadi bukti manusia berjuang menjaga eksistensinya di dunia ini. HEHEHE :D	Gak Kenal Maka Gak Kebal. Vaksin Jadi Bukti Manusia Berjuang Menjaga Eksistensinya Di Dunia Ini. HEHEHE :D
1348307055040348160	Sistem satu data mendukung Vaksinasi di Indonesia =D #Vaksin #Covid19 #CoronaVirus #Kesehatan	Sistem Satu Data mendukung Vaksinasi di Indonesia =D #Vaksin #Covid19 #CoronaVirus #Kesehatan
1348322660925669377	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…
1348425614584266753	Apa BedanyaVaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…	Apa Bedanya Vaksin China Sinovac dan Sinopharm serta Merek Lain?

		#Vaksin #Perusahaanfarmasi #Virus #Farmasi…
1348427123594207233	Dua Lagi Obat yang Dapat NyelamatinNyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…	Dua Lagi Obat yang Dapat Nyelamatin Nyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…
1348434452775657475	Vaksin bikinanTiongkok kembali dipertanyakan, orang tua gakboleh suntik!	Vaksin bikinan Tiongkok kembali dipertanyakan, orang tua gak boleh suntik!
1348434452775657337	Walaupun di vaksin, tetangga guemasih terkena covid juga tuh #vaksingagal	Walaupun di vaksin, tetangga gue masih terkena covid juga tuh #vaksingagal

2. Normalisasi

Normalisasi Teks adalah proses pengolahan teks yang bertujuan untuk mengubah struktur atau bentuk teks yang asalnya sulit dimengerti komputer hingga akhirnya mudah dimengerti dan diolah lebih lanjut. Salah satu upaya normalisasi adalah mengembalikan kata gaul jadi kata yang baku, mengubah semua teks kapital jadi *lowercase*, dan menjabarkan angka. Setelah dilakukan *tokenizasi* pada Tabel 4.5 maka hasil setelah teks dilakukan normalisasi seperti pada Tabel 4.6.

Tabel 4.6.
Hasil Normalisasi

Id	*Text*	Hasil Normalisasi
1348286903527768065	#vaksin untuk Indonesia ☺	#vaksin untuk Indonesia ☺
1348302168248340481	Gak Kenal Maka Gak Kebal. Vaksin Jadi Bukti Manusia Berjuang Menjaga Eksistensinya Di Dunia Ini. HEHEHE :D	Tidak Kenal Maka Tidak Kebal. Vaksin Jadi Bukti Manusia Berjuang Menjaga Eksistensinya Di Dunia Ini. HEHEHE :D
1348307055040348160	Sistem Satu Data mendukung Vaksinasi di Indonesia =D #Vaksin #Covid19 #CoronaVirus #Kesehatan	Sistem Satu Data mendukung Vaksinasi di Indonesia =D #Vaksin #Covid19 #CoronaVirus #Kesehatan
1348322660925669377	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…
1348425614584266753	Apa Bedanya Vaksin China	Apa Bedanya Vaksin China

	Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…	Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…
1348427123594207233	Dua Lagi Obat yang Dapat Nyelamatin Nyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…	Dua Lagi Obat yang Dapat Selamatkan Nyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…
1348434452775657475	Vaksin bikinan Tiongkok kembali dipertanyakan, orang tua gak boleh suntik!	Vaksin buatan Tiongkok kembali dipertanyakan, orang tua tidak boleh suntik!
1348434452775657337	Walaupun di vaksin, tetangga gue masih terkena covid juga tuh #vaksingagal	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal

3. Menghilangkan Emoticon

Terkadang emoticon menjadi sebuah value yang cukup menggangu dalam proses klasifikasi, kebanyakan orang menambahkan emoticon dalam *tweet*nya untuk menunjukan emosi yang dirasakan, Akan tetapi dalam kasus *sentiment analysis* emoticon tidak dapat mempengaruhi *negatif*, netral ataupun positif suatu tanggapan sehingga emoticon perlu kita hilangkan. Proses menghilangkan emoticon pada Bahasa pemrograman *python* tentunya kita menggunakan kamus yang menjadi pembanding dan berisi *pattern* emoji yang mungkin ada pada sebuah *tweet.* Setelah dilakukan *normalisasi* pada Tabel 4.6 maka hasil setelah teks dilakukan penghilangan emoticon seperti pada Tabel 4.7.

Tabel 4.7.
Hasil Menghilangkan Emoticon

Id	*Text*	Hasil Normalisasi
1348286903527768065	#vaksin untuk Indonesia ☺	#vaksin untuk Indonesia
1348302168248340481	Tidak Kenal Maka Tidak Kebal. Vaksin Jadi Bukti Manusia Berjuang Menjaga Eksistensinya Di Dunia Ini. HEHEHE :D	Tidak Kenal Maka Tidak Kebal. Vaksin Jadi Bukti Manusia Berjuang Menjaga Eksistensinya Di Dunia Ini. HEHEHE
1348307055040348160	Sistem Satu Data mendukung Vaksinasi di Indonesia =D #Vaksin #Covid19 #CoronaVirus	Sistem Satu Data mendukung Vaksinasi di Indonesia #Vaksin #Covid19 #CoronaVirus #Kesehatan

	#Kesehatan	
1348322660925669377	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…
1348425614584266753	Apa Bedanya Vaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…	Apa Bedanya Vaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…
1348427123594207233	Dua Lagi Obat yang Dapat Selamatkan Nyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…	Dua Lagi Obat yang Dapat Selamatkan Nyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…
1348434452775657475	Vaksin buatan Tiongkok kembali dipertanyakan, orang tua tidak boleh suntik!	Vaksin buatan Tiongkok kembali dipertanyakan, orang tua tidak boleh suntik!

1348434452775657337	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal

4. *Convert negasi*

Dalam sebuah teks atau postingan seseorang terdapat kata-kata yang mengandung makna *negasi* seperti kata "Tidak" dan "Jangan". Kata ini merupakan value yang penting dalam menentukan *sentiment analysis* sebuah tanggapan sehingga dalam proses *preprocessing* kata tersebut jangan sampai hilang. Pada kasus penelitian ini, terdapat proses yaitu *stop word removal* dimana dengan kamus *stop word* list, kata *negasi* akan dihilangkan karena masuk ke dalam kamus *stop word*, maka perlu adanya *convert negasi* agar kata "tidak" dan "jangan" tidak hilang. Langkah dalam melakukan *convert negasi* adalah menggabungkan kata *negasi* dengan kata depannya, misalkan terdapat kalimat "Tidak Sembuh" maka akan berubah menjadi "TidakSembuh". Setelah dilakukan menghilangkan emoji pada Tabel 4.7 maka hasil setelah teks dilakukan *convert negasi* seperti pada Tabel 4.8.

Tabel 4.8.
Hasil *Convert negasi*

Id	*Text*	Hasil Normalisasi
1348286903527768065	#vaksin untuk Indonesia	#vaksin untuk Indonesia
1348302168248340481	Tidak Kenal Maka Tidak Kebal. Vaksin Jadi Bukti Manusia	TidakKenal Maka TidakKebal. Vaksin Jadi Bukti Manusia Berjuang Menjaga

	Berjuang Menjaga Eksistensinya Di Dunia Ini. HEHEHE	Eksistensinya Di Dunia Ini. HEHEHE
1348307055040348160	Sistem Satu Data mendukung Vaksinasi di Indonesia #Vaksin #Covid19 #CoronaVirus #Kesehatan	Sistem Satu Data mendukung Vaksinasi di Indonesia #Vaksin #Covid19 #CoronaVirus #Kesehatan
1348322660925669377	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…
1348425614584266753	Apa Bedanya Vaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…	Apa Bedanya Vaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…
1348427123594207233	Dua Lagi Obat yang Dapat Selamatkan Nyawa Pasien	Dua Lagi Obat yang Dapat Selamatkan Nyawa Pasien

	COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…	COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…
1348434452775657475	Vaksin buatan Tiongkok kembali dipertanyakan, orang tua tidak boleh suntik!	Vaksin buatan Tiongkok kembali dipertanyakan, orang tua tidakboleh suntik!
1348434452775657337	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal

5. *Stemming*

Stemming merupakan salah satu upaya yang penting dalam persiapan teks. *Stemming* adalah Langkah menghilangkan imbuhan atau mengembalikan sebuah kata menjadi kata dasarnya. Dalam penelitian ini, untuk melakukan *stemming* diperlukan sebuah *library* dengan nama Sastrawi dimana *library* ini mengandung semua kata dasar dan imbuhan. Ketika terdeteksi sebuah kata yang berimbuhan, maka *python* akan menghilangkan imbuhan tersebut. Sesuai yang diketahui bahwa imbuhan pada Bahasa Indonesia dapat didetailkan sebagai berikut:

Kata masukan memiliki dua awalan (prefiks) dan tiga akhiran (sufiks). Berikut detail dari asumsi dari metode Arifin Setiono:

[AW1] + [AW2] + KD + [AK3] + [AK2] + [AK1]

Dimana AW merupakan awalan, KD merupakan kata dasar dan AK merupakan akhiran. Tahapan yang dilakukan pada metode tersebut adalah sebagai berikut [41]:

1. Melakukan pemeriksaan pada kamus kata dasar, jika terdapat kata yang sesuai pada kamus kata dasar maka proses cukup samapai disini, jika tidak Akan dilanjutkan dengan mempersiapkan variabel awalan [0], awalan [1], akhiran [0], akhiran [1], akhiran [2] untuk menampung imbuhan yang telah terpisah dari kata dasar.
2. Pemotangan dilakukan secara berurutan sebagai berikut:
    a. AW I, hasil disimpan pada awalan [0].
    b. AW II, hasil disimpan pada awalan [1].
    c. AK I, hasil disimpan pada akhiran [0].
    d. AK II, hasil disimpan pada akhiran [1].
    e. AK III, hasil disimpan pada akhiran [2].

    Dalam setiap proses pemotongan di atas selalu dilakukan pemeriksaan dengan kamus kata dasar. Hal tersebut dilakukan untuk mengetahui apakah hasil pemotongan tersebut sudah ada dalam bentuk kata dasar. Apabila sudah dalam bentuk kata dasar, maka proses dinyatakan selesai dan tidak perlu melanjutkan proses pemotongan selanjutnya.
3. Namun jika sampai tahap dua belum menemukan kata dasar, maka akan dilakukan proses kombinasi. Kata dasar yang telah dihasilkan akan dikombinasikan dengan imbuhan-imbuhan dalam 12 konfigurasi berikut:
    a. KD
    b. KD + AK III
    c. KD + AK III + AK II
    d. KD + AK III + AK II + AK I
    e. AW I + AW II + KD
    f. AW I + AW II + KD + AK III
    g. AW I + AW II + KD + AK III + AK II
    h. AW I + AW II + KD + AK III + AK II + AK I

i. AW II + KD
j. AW II + KD + AK III
k. AW II + KD + AK III + AK II
l. AW II + KD + AK III + AK II + AK I

Pemeriksaan dalam 12 kombinasi di atas selalu dilakukan pemeriksaan dengan kamus kata dasar. Hal tersebut dilakukan untuk mengetahui apakah hasil pemotongan tersebut sudah ada dalam bentuk kata dasar. Apabila sudah dalam bentuk kata dasar, maka proses dinyatakan selesai, namun jika sampai pada tahap akhir tidak ditemukan kata dasar yang tepat maka akan dikembalikan pada kata semula. Setelah dilakukan *convert negasi* pada Tabel 4.8. maka hasil setelah teks dilakukan *stemming* seperti pada Tabel 4.9.

Tabel 4.9.
Hasil *Stemming*

Id	Teks	Hasil *Stemming*
1348286903527768065	#vaksin untuk Indonesia	#vaksin untuk Indonesia
1348302168248340481	TidakKenal Maka TidakKebal. Vaksin Jadi Bukti Manusia Berjuang Menjaga Eksistensinya Di Dunia Ini. HEHEHE	TidakKenal Maka TidakKebal. Vaksin Jadi Bukti Manusia Berjuang jaga Eksistensi Di Dunia Ini. HEHEHE
1348307055040348160	Sistem Satu Data mendukung Vaksinasi di Indonesia #Vaksin #Covid19 #CoronaVirus	Sistem Satu Data mendukung Vaksinasi di Indonesia #Vaksin #Covid19 #CoronaVirus

	#sehat	#sehat
1348322660925669377	Siap menerima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	Siap terima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…
1348425614584266753	Apa Bedanya Vaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…	Apa Beda Vaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…
1348427123594207233	Dua Lagi Obat yang Dapat Selamatkan Nyawa Pasien COVID-19 Ditemukan #Infus #Virus #Wabah #Radang #Vaksin…	Dua Lagi Obat yang Dapat Selamat Nyawa Pasien COVID-19 temu #Infus #Virus #Wabah #Radang #Vaksin…
1348434452775657475	Vaksin buatan Tiongkok kembali dipertanyakan, orang tua tidakboleh suntik!	Vaksin buat Tiongkok kembali tanya, orang tua tidakboleh suntik!

1348434452775657337	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal

6. *Stop word removal*

Stop word Removal adalah proses *filtering,* pemilihan kata-kata penting dari hasil token yaitu kata-kata apa saja yang di gunakan untuk mewakili dokumen. Pada penelitian ini, acuan *stop word removal* adalah sebuah kamus yang diambil dari *library* Sastrawi yaitu *stop word* list berbahasa Indonesia. Tahapan ini akan dilakukan pencarian kata dan membuang kata yang ada dalam *stop word removal*. Dalam NLP (*Natural Language Processing*) *stop word* merupakan kata yang diabaikan dalam pemrosesan, kata-kata ini biasanya disimpan ke dalam stop lists. Karakteristik utama dalam pemilihan *stop word* biasanya adalah kata yang mempunyai frekuensi kemunculan yang tinggi misalnya kata penghubung seperti "dan", "atau", "tapi", "akan" dan lainnya. Tidak ada aturan pasti dalam menentukan *stop word* yang akan digunakan, penentuan *stop word* bisa disesuaikan dengan kasus yang sedang diselesaikan. Setelah dilakukan *stemming* pada Tabel 4.9 maka hasil setelah teks dilakukan *stop word removal* seperti pada Tabel 4.10.

Tabel 4.10.
Hasil *Stop word Removal*

Id	**Hasil *Stemming***	**Hasil *Stop word Removal***
1348286903527768065	#vaksin untuk Indonesia	#vaksin Indonesia
1348302168248340481	TidakKenal Maka TidakKebal.	TidakKenal TidakKebal. Vaksin

	Vaksin Jadi Bukti Manusia Berjuang jaga Eksistensi Di Dunia Ini. HEHEHE	Bukti Manusia Berjuang jaga Eksistensi Dunia
1348307055040348160	Sistem Satu Data mendukung Vaksinasi di Indonesia #Vaksin #Covid19 #CoronaVirus #sehat	Sistem Satu Data mendukung Vaksinasi Indonesia #Vaksin #Covid19 #CoronaVirus #sehat
1348322660925669377	Siap terima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	Siap terima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…
1348425614584266753	Apa Beda Vaksin China Sinovac dan Sinopharm serta Merek Lain? #Vaksin #Perusahaanfarmasi #Virus #Farmasi…	Apa Beda Vaksin China Sinovac Sinopharm Merek Lain #Vaksin #Perusahaanfarmasi #Virus #Farmasi
1348427123594207233	Dua Lagi Obat yang	Dua Obat Selamat Nyawa

	Dapat Selamat Nyawa Pasien COVID-19 temu #Infus #Virus #Wabah #Radang #Vaksin…	Pasien COVID-19 temu #Infus #Virus #Wabah #Radang #Vaksin…
1348434452775657475	Vaksin buat Tiongkok kembali tanya, orang tua tidakboleh suntik!	Vaksin buat Tiongkok kembali tanya, orang tua tidakboleh suntik!
1348434452775657337	Walaupun di vaksin, tetangga saya masih terkena covid juga tuh #vaksingagal	Walaupun vaksin, tetangga masih terkena covid #vaksingagal

7. *Word embedding*

Word embedding adalah istilah yang digunakan untuk teknik mengubah sebuah kata menjadi sebuah vektor atau array yang terdiri dari kumpulan angka. Ketika membuat model *machine learning* yang menerima *input* sebuah teks, tentu *machine learning* tidak bisa langsung menerima mentah-mentah teks yang kita miliki, kata tersebut harus diubah dulu menjadi angka dengan acuan sebuah kamus kata. Biasanya jika tidak menggunakan *word embedding*, setiap kata akan diubah menjadi angka dalam bentuk Integer sesuai dengan posisi angka tersebut dalam kamus, misalkan kata "Sembuh" diubah menjadi angka "4" dan kata "Meninggal" diubah menjadi angka "7".Angka-angka tersebut kita ubah lagi menjadi sebuah vektor (array 1 dimensi) yang memiliki panjang sepanjang banyak kata

yang kita miliki di kamus. Array tersebut hanya akan bernilai 1 atau 0 (disebut *one hot encoding*). Nilai 1 diposisikan pada indeks yang merupakan nomor kata tersebut sedangkan elemen lainnya bernilai 0.

Contohnya untuk kata "sembuh", dengan banyak kosakata yang kita miliki adalah 100 kata, maka dari kata tersebut kita akan memperoleh sebuah vektor dengan panjang 100 yang berisi 0 semua kecuali pada posisi ke 3 yang bernilai 1.

Jika kamus yang dimiliki mempunyai ukuran mencapai 10,000 kata, maka untuk setiap katanya akan di-*convert* menjadi vektor ukuran 10,000 yang hampir semua elemennya bernilai 0 semua. Metode ini selain kurang efisien dalam memori juga tidak memberikan banyak informasi. Dengan metode *word embedding*, kata dapat diubah menjadi sebuah vektor yang berisi angka-angka dengan ukuran yang cukup kecil untuk mengandung informasi yang lebih banyak. Informasi yang diperoleh akan cukup banyak sampai-sampai vektor kita akan dapat mendeteksi makna, seperti kata "marah" dan "mengamuk" itu lebih memiliki kedekatan nilai ketimbang kata "marah" dengan "bahagia".

Word embedding yang akan kita gunakan pada kasus ini menggunakan *Fasttext*. *Fasttext* adalah *library* yang dikeluarkan oleh Facebook yang dapat digunakan untuk *word embedding*. Sebenarnya, *Fasttext* sendiri adalah pengembangan dari *library Word2vec* yang telah lebih lama terkenal sebagai *library* untuk *word embedding*. *Fasttext* memiliki keunggulan dibanding *Word2vec*. Salah satunya adalah kemampuan *Fasttext* untuk menangani kata yang tidak pernah kita jumpai sebelumnya (*Out Of Vocabulary word* atau dikenal OOV). Misalnya kata-kata yang tidak baku seperti "Pengoptimalisasian" tetap akan diperoleh vektornya. *Library Word2vec* ataupun teknik *one hot encoding* tradisional yang seperti dijelaskan sebelumnya akan menghasilkan eror ketika menerima kata yang tidak pernah ada di kamus.

Tahapan *Word embedding* menggunakan *fasttext* menentukan parameter betikut:

1. sg : parameter ini menentukan *learning* algorithm apa yang akan digunakan. Terdapat 2 pilihan *learning* algorithm, skip-gram atau CBOW.
2. Size: parameter ini menentukan dimensi dari *vector*.
3. Window: parameter ini menentukan jumlah kata sebelum dan sesudah kata tertentu yang digunakan sebagai pertimbangan konteks dalam satu kalimat.
4. Min_count: parameter ini menentukan jumlah minimum kemunculan suatu kata agar kata tersebut tidak diabaikan
5. Iter: parameter ini menentukan jumlah iterasi dilakukannya *training*.

6. Min_n: parameter ini menentukan panjang minimum character n-gram yang Akan digunakan untuk *training* representasi kata.
7. Max_n: parameter ini menentukan panjang maksimum *character n-gram* yang Akan digunakan untuk traning representasi kata.

Berikut merupakan contoh kata yang Akan diproses dengan *fasttext*.

Tabel 4.11.
Hasil *Word embedding*

Id	Teks	Bobot Kata dalam Kamus
1348286903527768065	#vaksin Indonesia	0.22 0.031
1348302168248340481	TidakKenal TidakKebal. Vaksin Bukti Manusia Berjuang jaga Eksistensi Dunia	0.452 0.32 0.22 0.049 0.11 0.12 0.45 0.120 0.033
1348307055040348160	Sistem Satu Data mendukung Vaksinasi Indonesia #Vaksin #Covid19 #CoronaVirus #sehat	0.352 0.52 0.022 0.149 0.22 0.031 0.22 0.120 0.121 0.342
1348322660925669377	Siap terima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	0.352 0.52 0.022 0.20 0.22 0.221 0.22 0.220 0.121
1348425614584266753	Apa Beda Vaksin China Sinovac Sinopharm Merek	0.352 0.52 0.22 0.450 0.22 0.321 0.32

	Lain #Vaksin #Perusahaanfarmasi #Virus #Farmasi	0.320 0.21 0.12 0.24 0.134
1348427123594207233	Dua Obat Selamat Nyawa Pasien COVID-19 temu #Infus #Virus #Wabah #Radang #Vaksin…	0.352 0.52 0.022 0.130 0.22 0.191 0.22 0.990 0.12 0.11 0.12 0.22
1348434452775657475	Vaksin buat Tiongkok kembali tanya, orang tua tidakboleh suntik!	0.22 0.12 0.322 0.230 0.22 0.91 0.22 0.90 0.2
1348434452775657337	Walaupun vaksin, tetangga masih terkena covid #vaksingagal	0.42 0.22 0.12 0.322 0.230 0.19 0.23

Agar mempermudah klasifikasi, maka teks harus dirubah ke bentuk *one-hot* matrix, tahap ini dilakukan karena setiap kata pada dasarnya mempunyai satu dimensi yang berbeda. Setiap array adalah panjang dari dictionary, dan setiap nilai dalam dictionary yg bukan merupakan nilai token maka diwakili dengan angka 0 sedangkan nilai token diwakili oleh angka 1. Hasil dari tahap *one-hot* matrix dengan *word embedding* seperti pada table 4.12

Tabel 4.12.
Hasil *Word embedding*

Id	**Teks**	***One hot encoding***
1348286903527768065	#vaksin	10

	Indonesia	01
1348302168248340481	TidakKenal TidakKebal. Vaksin Bukti Manusia Berjuang jaga Eksistensi Dunia	100000000 010000000 001000000 000100000 000010000 000001000 000000100 000000010 000000001
1348307055040348160	Sistem Satu Data mendukung Vaksinasi Indonesia #Vaksin #Covid19 #CoronaVirus #sehat	1000000000 0100000000 0010000000 0001000000 0000100000 0000010000 0000001000 0000000100 0000000010 0000000001
1348322660925669377	Siap terima vaksin covid 19 #vaksinhalal #vaksinsiap34propinsi #vaksicovid19 #jokowidodo…	100000000 010000000 001000000 000100000 000010000 000001000 000000100 000000010 000000001
1348425614584266753	Apa Beda Vaksin China Sinovac Sinopharm Merek Lain #Vaksin #Perusahaanfarmasi #Virus #Farmasi	100000000000 010000000000 001000000000 000100000000 000010000000 000001000000 000000100000 000000010000 000000001000 000000000100 000000000010 000000000001
1348427123594207233	Dua Obat Selamat Nyawa Pasien COVID-19 temu #Infus	100000000000 010000000000 001000000000 000100000000 000010000000 000001000000 000000100000 000000010000

	#Virus #Wabah #Radang #Vaksin…	000000001000 000000000100 000000000010 000000000001
1348434452775657475	Vaksin buat Tiongkok kembali tanya, orang tua tidakboleh suntik!	100000000 010000000 001000000 000100000 000010000 000001000 000000100 000000010 000000001
1348434452775657337	Walaupun vaksin, tetangga masih terkena covid #vaksingagal	1000000 0100000 0010000 0001000 0000100 0000010 0000001

C. Klasifikasi dengan SVM

Tahap utama dari penelitian ini adalah klasifikasi dengan menggunakan algoritma *Support Vector Machine*. Fitur yang sudah dipilih sebelumnya akan digunakan sebagai masukan perhitungan oleh *Support Vector Machine*, untuk mengklasifikasikan dokumen. Pada tahap ini digunakan dokumen *training* sebagai dokumen masukan. Teks dari setiap kalimat tentang vaksin sinovac covid sebelumnya telah ditransformasikan ke dalam representasi vektor kata menggunakan *word embedding* kemudian menjadi sebuah *input*an pada *layer* pertama dengan Maksimum panjang filter dari *input* adalah 1000, sehingga *input* Akan berupa matriks berukuran 1000 x 300.Adapun gambaran *input*an pada proses *SVM* sebagai berikut:

Tabel 4.13. *Input*an *SVM*

Kata	*One hot encoding*	*Input Layers*	Keterangan
#vaksin Indonesia	10 01		2 kernel dan 4 filter

Algoritma klasifikasi SVM menggunakan data latih untuk membentuk model *classifier*, model yang terbentuk akan digunakan

sebagai prediksi kelas data baru yang belum pernah ada sebelumnya. Data latih dan data uji yang digunakan adalah data yang telah memiliki label kelas, dengan perbandingan data latih dan data uji adalah 80% : 20%.

Tabel 4.14 Proporsi Kelas Sentimen Hasil Pelabelan Secara Manual Pada Data Latih dan Data Uji

Klasifikasi	Positif	Negatif	Jumlah
Data Latih	1025	326	1351
Data Uji	117	32	149
Jumlah	1142	358	1500

Tabel 4.15. Proporsi Kelas Sentimen Hasil Pelabelan dengan *Sentiment Scoring* Pada Data Latih dan Data Uji

Klasifikasi	Positif	Negatif	Jumlah
Data Latih	1058	293	1351
Data Uji	114	35	149
Jumlah	1172	328	1500

Pada penelitian ini digunakan metode *Support Vector Machine* (SVM) dengan fungsi *kernel* yang digunakan adalah *kernel linear* dan *kernel* RBF. Pengujian pengaruh parameter *Support Vector Machine* dilakukan untuk mengetahui nilai-nilai parameter SVM yang optimal untuk proses analisis sentimen. Pada *kernel linear* terdapat satu parameter yang diuji yaitu nilai *Cost* dengan nilai parameter *Cost* (C) : 0,01; 0,1; 1; 10; 100; 1000 untuk data latih. Hasil dari pengujian pengaruh nilai *Cost* pada model linier hasil pelabelan secara manual ditunjukkan pada Tabel 17, sedangkan pada model

linier hasil pelabelan dengan *sentiment scoring* ditunjukkan pada Tabel 18. Penelitian ini menggunakan 10-*cross validation* untuk menguji performa *machine* dalam membentuk klasifikasi.

Tabel 4.16. Nilai Akurasi Keseluruhan dan Akurasi Kappa pada Model Linier Hasil Pelabelan Secara Manual

Evaluasi Model	***Cost* (C)**					
	0,01	**0,1**	**1**	**10**	**100**	**1000**
Akurasi Keseluruhan	0,7919	0,7919	0,7852	0,7785	0,7785	0,7785
Akurasi *Kappa*	0,1105	0,1105	0,0966	0,0832	0,0832	0,0832

Tabel 4.17. Nilai Akurasi Keseluruhan dan Akurasi Kappa pada Model Linier Hasil Pelabelan dengan *Sentiment Scoring*

Evaluasi Model	***Cost* (C)**					
	0,01	**0,1**	**1**	**10**	**100**	**1000**
Akurasi Keseluruhan	0,7919	0,7718	0,7584	0,7651	0,7651	0,7651
Akurasi *Kappa*	0,21	0,1673	0,0948	0,0833	0,0571	0,0571

Nilai C yang paling optimal pada Tabel 7 adalah 0,01 dan 0,1 karena memperoleh nilai akurasi keseluruhan dan akurasi *kappa* paling besar yaitu 79,19%

dan 11,05%. Sedangkan pada Tabel 4.15 menunjukkan bahwa nilai C paling optimal adalah 0,01, karena memperoleh nilai akurasi keseluruhan dan akurasi *kappa* paling besar yaitu 79,19% dan 21%.

Berdasarkan persamaan (69) pada *kernel* RBF terdapat dua parameter yang diuji yaitu nilai *Cost* (C) dan *Gamma* (γ). *Gamma* (γ) yang digunakan adalah 0,00026 dan nilai parameter *Cost* (C): 0,01; 0,1; 1; 10; 100; 1000 untuk data latih. Hasil dari pengujian pengaruh nilai *Cost* pada performa *kernel* RBF dengan nilai *gamma* tetap ditunjukkan pada Tabel 21 dan Tabel 22. Penelitian ini menggunakan 10-*cross validation* untuk menguji performa *machine* dalam membentuk klasifikasi.

Tabel 4.18. Nilai Akurasi Keseluruhan dan Akurasi Kappa pada Model RBF Hasil Pelabelan Secara Manual

Evaluasi Model	***Cost* (C)**					
	0,01	**0,1**	**1**	**10**	**100**	**1000**
Akurasi Keseluruhan	0,7852	0,7919	0,7919	0,7919	0,7919	0,7919
Akurasi *Kappa*	0	0,0804	0,1105	0,1105	0,1105	0,1652

Tabel 4.19. Nilai Akurasi Keseluruhan dan Akurasi Kappa pada Model RBF Hasil Pelabelan dengan *Sentiment Scoring*

Evaluasi Model	***Cost* (C)**					
	0,01	**0,1**	**1**	**10**	**100**	**1000**
Akurasi Keseluruhan	0,7584	0,7852	0,7919	0,7919	0,7785	0,7584
Akurasi *Kappa*	-0,0132	0,1501	0,21	0,1881	0,1812	0,0948

Berdasarkan pada Tabel 9 dari hasil pengujian nilai *Cost* dengan nilai *gamma* tetap, nilai C yang paling optimal adalah 1000

karena memperoleh nilai akurasi keseluruhan dan akurasi *kappa* paling besar yaitu 79,19% dan 16,52%. Sedangkan pada Tabel 10 menunjukkan bahwa nilai C paling optimal adalah 1, karena memperoleh nilai akurasi keseluruhan dan akurasi *kappa* paling besar yaitu 79,19% dan 21%.

Setelah melakukan analisis dengan menggunakan *kernel linear* dan *kernel*

RBF diperoleh model terbaik dari masing-masing model sebagai berikut:

Tabel 4.20. Model Terbaik Kernel Linear dan Kernel Radial dari Hasil Pelabelan Data Secara Manual

Evaluasi Model	***Kernel Linear* (C=0.1)**	***Kernel* RBF (C=1000 dan γ=0.00026)**
Akurasi Keseluruhan	0,7919	0,7919
Akurasi *Kappa*	0,1105	0.1652

Tabel 4.21. Model Terbaik Kernel Linier dan Kernel Radial dari Hasil Pelabelan Data dengan *Sentiment Scoring*

Evaluasi Model	***Kernel Linear* (C=0.1)**	***Kernel* RBF (C=1 dan γ=0.00026)**
Akurasi Keseluruhan	0,7919	0,791 9
Akurasi *Kappa*	0,21	0.21

Berdasarkan Tabel 11 dan 12, model *kernel linear* dan *kernel* RBF dari hasil pelabelan data secara manual dan *sentiment scoring* memiliki akurasi keseluruhan tertinggi yang sama yaitu 79,19%. Pada pelabelan data secara manual nilai akurasi *kappa* pada *kernel*

RBF lebih besar daripada nilai akurasi *kappa* pada *kernel linear*, sedangkan pada pelabelan data dengan *sentiment scoring* mengasilkan akurasi *kappa* tertinggi yang sama yaitu 21%. Hal tersebut menunjukkan bahwa model dengan *kernel* RBF memiliki kecocokan hasil klasifikasi sentimen pada Gojek dengan benar dibandingkan menggunakan *kernel linear*.

BAB V

HASIL DAN PEMBAHASAN

5.1.Interpretasi Hasil

Setelah model yang diusulkan dibuat dan dilakukan *training*, maka terdapat hasil dari performa model yang dibuat sebagai berikut:

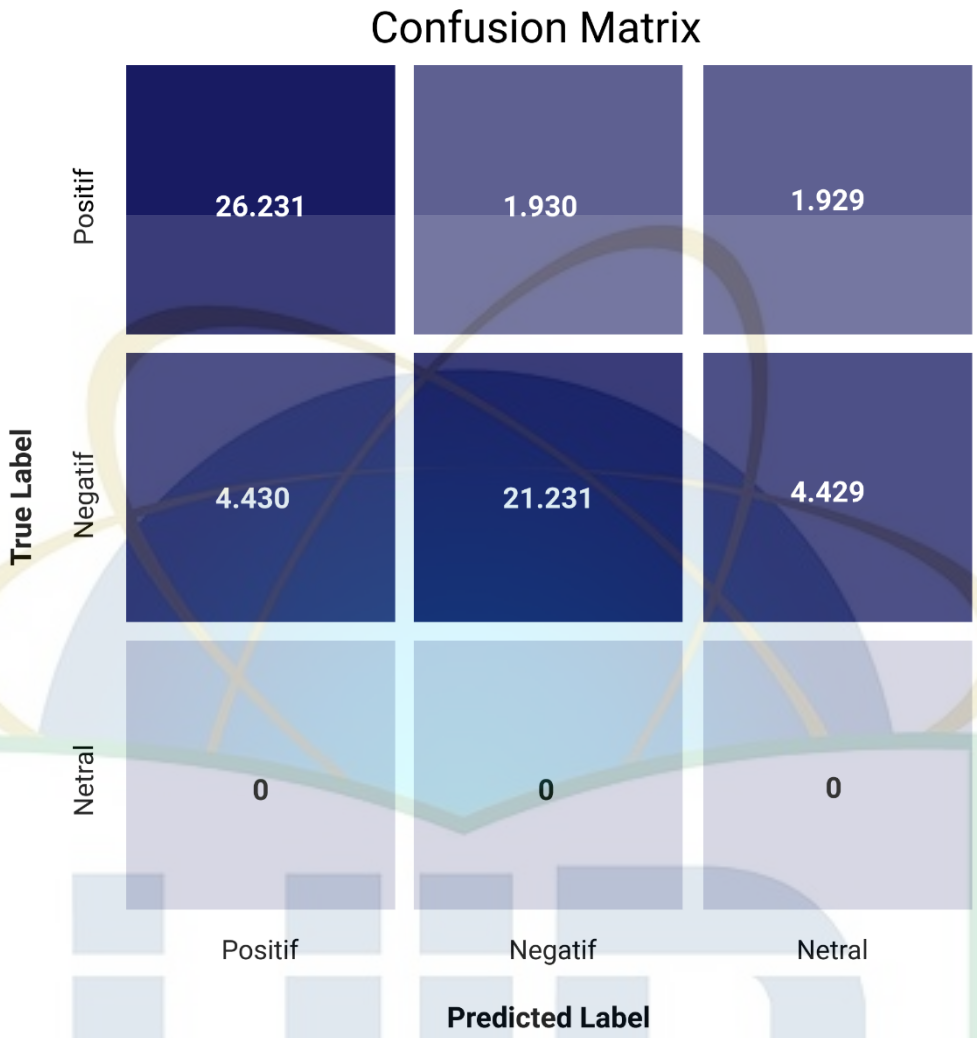
Tabel 5.1. Performa Model

No	Jumlah Epoch	Split	Akurasi
1	10	70:30	72%
2	10	80:20	74%
3	15	70:30	82%
4	15	80:20	83%
5	20	70:30	84%
6	20	80:20	85%
7	25	70:30	85%
8	25	80:20	85%

Berdasarkan Tabel 5.1. pada proses *training*, dilakukan perubahan jumlah epoch dan komposisi split *dataset* dan hasil akurasi terbaik didapat pada Jumlah Epoch 20 dan komposisi split 80:20. Adapun Ketika dinaikan epoch menjadi 25 akurasi tetap sama, maka tidak perlu lagi menambahkan jumlah *epoch*.

5.2. Evaluasi dan Pengujian Model

Pada evaluasi dan pengujian, akan dicari nilai precision dan recall. Nilai tersebut didapatkan dari jumlah *True Positif (TP), False Positif (FP), True Negatif (TN)* dan *False Negatif (FN)* yang dihasilkan oleh confussion matrix seperti pada Gambar berikut:



Gambar 5.1. Confussion Matrix

Berikut merupakan jumlah TP, FP, TN dan FN pada Tabel 14.

Tabel 5.2. Nilai TP TF TN dan FN

	Total *Class*	Jumlah
TP	10.000 Negatif 10.000 Positif 10.000 Netral	26.231
FP		3859
TN		21.231
FN		8859

Berdasarkan pada Tabel 4.12 didapatkan nilai berikut:

1. *Precision*

   Dengan rumus TP / TP + FP

   = 26231 / 26231 +3859

   = 26231 / 30090

   = **0.87175**

2. *Recall*

   Dengan rumus TP / TP + FP

   = 26231 / 26231 + 8859

   = 26231 / 35091

   = **0.7475**

5.3. Source Code

Proses pengkodingan untuk implementasi metode *SVM* dalam *sentiment analysis* ini sebagai berikut:.

```
import pandas as pd
import numpy as np
from datetime import datetime,timedelta
import pytz
import re
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import ast
import string
from wordcloud import WordCloud
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import itertools
import matplotlib.pyplot as plt
import seaborn as sns
```

Gambar 5. *2*. Code Import *Library*

```
def repair_exaggeration(x):
    word_tokens = word_tokenize(x)
    new_x =''
    for i in word_tokens:
        if (i == sinovac):
            new = re.sub(r'(\w)\1\1+',r'\1\1',i)
            new_x = new_x +new+' '
```

```
        elif(i ==sinovac):
            new = sinovac
            new_x = new_x +new+' '
        else:
            new = re.sub(r'(\w)\1\1\1+',r'\1',i)
            new_x = new_x +new+' '
    return new_x

def del_word(x,key_list):
    n = len(key_list)
    word_tokens = word_tokenize(x)
    new_x =''
    for word in word_tokens:
        if word not in key_list:
            new_x = new_x+word+' '
    return new_x

def clean_tweets(tweet):
    my_file = open("cleaning_source/combined_stop_words.txt",
"r")
    content = my_file.read()
    stop_words = content.split("\n")
    file_2  = open("cleaning_source/update_combined_slang_word
s.txt", "r")
    content2 = file_2.read()
    slang_words = ast.literal_eval(content2)
    my_file.close()
    file_2.close()

    tweet = tweet.lower()
    tweet = re.sub(r':', '', tweet)
    tweet = re.sub(r',Ä¶', '', tweet)
    tweet = re.sub(r'[^\x00-\x7F]+',' ', tweet)
    tweet = re.sub('[^a-zA-Z]', ' ', tweet)

    tweet=re.sub("&lt;/?.*?&gt;","&lt;&gt;",tweet)
```

```
    tweet=re.sub("(\\d|\\W)+"," ",tweet)

    #remove other symbol from tweet
    tweet = re.sub(r'â', '', tweet)
    tweet = re.sub(r'€', '', tweet)
    tweet = re.sub(r'¦', '', tweet)

    word_tokens = word_tokenize(tweet)
    for w in word_tokens:
        if w in slang_words.keys():
            word_tokens[word_tokens.index(w)] = slang_words[w]

    filtered_tweet = [w for w in word_tokens if not w in stop_
words]
    filtered_tweet = []

    for w in word_tokens:
        if w not in stop_words and w not in string.punctuation
:
            filtered_tweet.append(w.lower())
    return ' '.join(filtered_tweet)

def count_words(x):
    words = word_tokenize(x)
    n=len(words)
    return n
```

Gambar 5.3. Code Proses *Preprocessing Text*

BAB VI

PENUTUP

6.1. Kesimpulan

Setelah melakukan penelitian tentang Sentimen Analisis Vaksin, dapat ditarik kesimpulan bahwa :

a. Penelitian ini menerapkan algoritma *Support Vector Machine (SVM)* dalam melakukan klasifikasi sentiment di sosial media terhadap pemberikan vasksin Sinovac dengan baik dan dapat membantu mendeteksi sentiment dengan membagi setiap kata positif dan kata negarif sesuai dengan *library* dari NLTK dan gensim dan menghitung total jumlah kata positif dan negatif untuk menentukan sentiment setiap kalimat.
b. akurasi yang dihasilkan oleh metode *Support Vector Machine (SVM)* dalam melakukan klasifikasi sentiment sebesar 85% dan dikatakan baik.
c. SVM berhasil melakukan deteksi otomatis sentiment pada twitter tentang vaksin Sinovac dengan bantuan *word embeddi*ng dan pembobotan setiap kata yang mempunyai bobot berbeda.

6.2. Saran

Penelitian yang dilakukan masih jauh dari sempurna maka dari itu penulis memberikan saran. Adapun saran tersebut adalah :

a. Pengembangan dengan mencoba melakukan komparasi atau mencoba metode klasifikasi yang lain.
b. Bisa dilakukan pengembangan lebih lanjut dalam bentuk purwarupa atau prototype program

DAFTAR PUSTAKA

Ahmad, M., & Ali, I. (2017). *Sentiment Analysis of Tweets using SVM*. *177*(5), 25–29.

Anisa Eka Puridewi, J. N. (2018). Perbandingan metode naive bayes, *Support Vector Machine* dan id3 dalam penetapan status penanganan kecelakaan kerja. *Seminar Nasional Matematika Dan Pendidikan Matematika*, 130–137.

Arifin, O., & Sasongko, T. B. (2018). Analisa perbandingan tingkat performansi metode *Support Vector Machine* dan naïve bayes classifier. *Seminar Nasional Teknologi Informasi Dan Multimedia 2018*, *6*(1), 67–72.

Arsya Monica Pravina, Imam Cholissodin, P. P. A. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme *Support Vector Machine (SVM)*. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(3), 2789–2797. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793

Dicky Nofriansyah, S.Kom., M. K. D. I. G. W. N. M. S. 2017. (2017). *ALGORITMA DATA MINING DAN PENGUJIAN*. DEEPUBLISH.

Finance, D. (2021). 4 juta dosis vaksin corona siap disuntikkan ke orang ri mulai februari 2020. *Detik*.

Hasan, M. (2017). *Menggunakan Algoritma Naive Bayes Berbasis*. *9*, 317–324.

Hootsuite & We Are Social. (2020). *We Are Social & Hootsuite (2020)*. DIGITAL 2020 GLOBAL DIGITAL OVERVIEW.

Masripah, S. (2015). Evaluasi Penentuan Kelayakan Pemberian Kredit Koperasi Syariah Menggunakan Algoritma Klasifikasi C4.5. *Jurnal Pilar Nusa Mandiri*, *XI*(1), 1–10.

Monarizqa, N., Nugroho, L. E., & Hantono, B. S. (2014). Penerapan Analisis Sentimen Pada Twitter Berbahasa Indonesia Sebagai Pemberi Rating. *Jurnal Penelitian Teknik Elektro Dan Teknologi Informasi*, *1*, 151–155.

Purbo, O. W. (2019). *Text Mining*. Andi.

Rezwanul, M., Ali, A., & Rahman, A. (2017). Sentiment Analysis on Twitter Data using KNN and SVM. *International Journal of Advanced Computer Science and Applications*, *8*(6), 19–25. https://doi.org/10.14569/ijacsa.2017.080603

Rofiqoh, U., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode *Support Vector Machine* dan Lexion Based Feature. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *1*(12), 1725–1732. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628

Ruan, Y., Durresi, A., & Alfantoukh, L. (2018). Using Twitter trust network for stock market analysis. *Knowledge-Based Systems*, *0*, 1–12. https://doi.org/10.1016/j.knosys.2018.01.016

Santoso, V. I., Virginia, G., & Lukito, Y. (2017). Penerapan Sentiment Analysis Pada Hasil Evaluasi Dosen Dengan Metode *Support Vector Machine*. *Jurnal Transformatika*, *14*(2), 72. https://doi.org/10.26623/transformatika.v14i2.439

Sari, B. W., & Haranto, F. F. (2019). Implementasi *Support Vector Machine* Untuk Analisis Sentimen Pengguna Twitter Terhadap Pelayanan Telkom Dan Biznet. *Jurnal Pilar Nusa Mandiri*, *15*(2), 171–176. https://doi.org/10.33480/pilar.v15i2.699

Soman, K. P., Loganathan, R., & Ajay, V. (2009). *Machine learning with SVM and other kernel methods*. PHI Learning Pvt. Ltd.

Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi *Support Vector Machine (SVM)*. *Jurnal Media Informatika Budidarma*, *4*(3), 650. https://doi.org/10.30865/mib.v4i3.2181

Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, *13*(1), 16–25. https://doi.org/10.36787/jti.v13i1.78