

Pengembangan Perangkat Lunak Klasifikasi Biner Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Menggunakan Support Vector Machine

Oktavian Yudistira Putra

Informatika, Fakultas Ilmu Komputer dan Desain, Institut Teknologi dan Bisnis Kalbis
Jalan Pulomas Selatan Kav. 22, Jakarta 13210

Email: oktavianyudistiraputra@gmail.com

Abstract: This study aims to develop software for the binary classification of public sentiment towards the Covid-19 vaccination program which the Indonesian government began to organize on January 13, 2021. The data used comes from Twitter tweets with data crawling techniques. The tweets were classified binary through the training and testing stages which resulted in two classes of sentiment, namely positive and not positive. The feature extraction method used is Term Frequency-Inverse Document Frequency (TF-IDF), to weight the words in the form of a matrix based on the frequency of each word in the dataset. The method in the text classification process used is the Support Vector Machine (SVM). In the software development process, this research uses the incremental method. The results of the classification model training obtained model performance with an evaluation calculation of accuracy of 0.925, precision of 0.951, recall of 0.929, and f-measure of 0.94.

Keywords: binary classification, incremental, svm, tf-idf, twitter

Abstrak: Penelitian ini bertujuan untuk mengembangkan perangkat lunak klasifikasi biner sentimen masyarakat terhadap program vaksinasi Covid-19 yang mulai diselenggarakan pemerintah Indonesia pada tanggal 13 Januari 2021. Data yang digunakan berasal dari cuitan Twitter dengan teknik crawling data. Cuitan tersebut dilakukan klasifikasi biner melalui tahap pelatihan dan pengujian yang menghasilkan dua kelas sentimen, yaitu positif dan bukan positif. Metode ekstraksi fitur yang digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF), untuk melakukan pembobotan kata berbentuk matriks berdasarkan frekuensi setiap kata pada dataset. Metode dalam proses klasifikasi teks yang digunakan adalah Support Vector Machine (SVM). Dalam proses pengembangan perangkat lunak, penelitian ini menggunakan metode inkremental. Hasil dari pelatihan model klasifikasi memperoleh performansi model dengan perhitungan evaluasi accuracy sebesar 0,925, precision sebesar 0,951, recall sebesar 0,929, dan f-measure sebesar 0,94.

Kata kunci: inkremental, klasifikasi biner, svm, tf-idf, twitter

I. PENDAHULUAN

A. Latar Belakang

Di masa pandemi Covid-19 ini, negara Indonesia masih memasuki tingkat kritis dalam penanganan wabah ini, yaitu hingga mencapai 1,17 juta kasus (per-10 Februari 2021) [1]. Untuk mengatasi pandemi Covid-19, pemerintah RI telah mendistribusikan

vaksin yang bernama *Sinovac* ke berbagai daerah di seluruh Indonesia sebagai upaya pencegahan meningkatnya kasus pandemi Covid-19.

Seperti yang telah diketahui, bahwa sebelumnya pada tanggal 13 Januari 2021 bapak presiden RI Joko Widodo melaksanakan vaksinasi pertama kali yang disiarkan secara *live streaming* di akun *Youtube* sekretariat presiden maupun televisi [2].

Berbagai komentar terlampaikan dalam jejaring sosial di kalangan

masyarakat atas kebijakan yang dilakukan pemerintah. Ada komentar negatif maupun positif yang dipublikasi setelah beredar berita vaksinasi berdampak mengancam nyawa seseorang di negara lain. Diikuti dengan berbagai cuitan (*tweet*) pada media sosial *Twitter* yang sempat *trending* pada awal bulan Februari 2021 dengan tagar Vaksinasi Nasional.

Melihat keadaan tersebut pada penelitian ini berupaya memecahkan permasalahan dengan melakukan analisis sentimen untuk mengetahui kecenderungan masyarakat terhadap vaksinasi Covid-19. Pada kasus ini, topik kasus yang diangkat ialah opini masyarakat terhadap Vaksinasi Nasional yang akan dilaksanakan pemerintah di berbagai daerah. Analisis Sentimen merupakan studi yang terdiri dari *Natural Language Processing* (NLP), komputasi linguistik, dan analisis teks untuk mengidentifikasi sentimen teks [3].

Penelitian terdahulu yang dilakukan oleh Umi Rofiqoh, Rizal Setya Perdana, dan M. Ali Fauzi [4] melakukan analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler Indonesia. Metode yang digunakan yaitu *Support Vector Machine* (SVM) dan *Lexicon Based Features*.

Adapun tahap yang dilakukan antara lain, yaitu praproses teks (*preprocessing*), pembobotan kata, pembobotan *lexicon based features*, dan klasifikasi SVM. Tahap *preprocessing* yang dilakukan terdiri dari *tokenization*, *cleansing*, *case folding*, filterisasi, dan *stemming*. Pembobotan kata dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Kemudian dinormalisasikan dengan dengan metode *min-max*. Kumpulan data diambil dari media sosial *Twitter* dengan jumlah 300 data, dengan pembagian 70% sebagai data latih dan 30% sebagai data uji.

Hasil dari analisis sentimen terhadap tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler menggunakan metode SVM dan *Lexicon Based Features* adalah mendapatkan nilai *accuracy* sebesar 79%, *precision* sebesar 65%, *recall* sebesar 97%, dan *f-measure* sebesar 78%.

Berdasarkan penelitian terdahulu tersebut, pada penelitian ini mengusulkan analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM). SVM merupakan salah satu metode dalam *supervised learning* yang pada umumnya digunakan untuk klasifikasi (*Support Vector Classification*) dan regresi (*Support Vector Regression*). SVM sejenis model vektor berbasis *classifier* yang mengharuskan sebuah teks harus diubah menjadi vektor sebelum digunakan klasifikasi [5].

Metode tersebut dipilih karena metode SVM memiliki akurasi yang lebih baik, jika dibandingkan dengan metode *Naïve Bayes* dan *Decision Tree* berdasarkan komparasi metode dari studi literatur yang dilakukan oleh Umi, Rizal, dan Ali.

Rumusan masalah pada penelitian ini adalah bagaimana melakukan analisis sentimen masyarakat terhadap vaksinasi Covid-19 dengan mengembangkan model klasifikasi biner menggunakan metode *Support Vector Machine*?

Adapun tujuan pada penelitian ini, bermaksud untuk mengetahui kecenderungan sentimen masyarakat terhadap vaksinasi Covid-19 dengan mengembangkan model klasifikasi biner menggunakan *Support Vector Machine*.

Langkah awal yang perlu dilakukan ialah melakukan pengumpulan data dari media sosial *Twitter* menggunakan API *Twitter* dan *library Tweepy* dengan kata kunci “Vaksinasi” dan “Vaksin”. Data akan dikumpulkan dengan panjang 1000 data, yang bertujuan agar dalam perancangan model klasifikasi dapat

memberikan akurasi yang baik. Lalu data akan dikumpulkan dalam bentuk dokumen *Comma Separated Values* (CSV) yang akan dipersiapkan sebagai masukan (*input*) perangkat lunak.

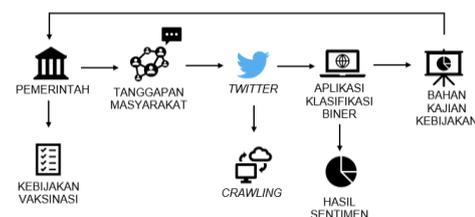
Kemudian kumpulan data (*dataset*) tersebut akan dilakukan *preprocessing* untuk membersihkan kata tidak baku dan penyeragaman frasa kata dengan melakukan normalisasi fitur (*cleansing*), *case folding*, *tokenization*, *stopwords removal*, dan *stemming*. Lalu data akan dipisah menjadi data uji sebesar 20% dan data latih sebesar 80%.

Selanjutnya, akan dilakukan ekstraksi fitur dan teknik pembobotan kata menggunakan metode TF-IDF. Setelah itu, dilakukan proses pelatihan dan pengujian model klasifikasi dengan metode SVM. Selain itu, penelitian ini juga merancang tampilan antarmuka pengguna berbasis web menggunakan kerangka kerja *Flask*.

Hasil akhir dari perangkat lunak ini adalah model klasifikasi sentimen, performansi dari pelatihan model klasifikasi, persentase diagram *pie*, hasil klasifikasi sentimen, dan tampilan antarmuka pengguna berbasis web.

Dengan penelitian ini diharapkan agar pemerintah dapat membuat kebijakan yang relevan terhadap masyarakat dengan mempertimbangkan kecenderungan sentimen dari masyarakat.

II. METODE PENELITIAN

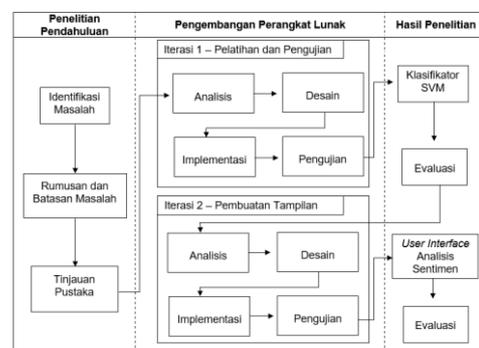


Gambar 1 Kerangka Pemikiran

Kerangka pemikiran pada penelitian ini yang terdapat pada Gambar 1,

menggambarkan bagaimana peran perangkat lunak klasifikasi biner dalam memproses suatu kumpulan data (*dataset*) tanggapan masyarakat terhadap kebijakan vaksinasi Covid-19 oleh pemerintah yang diambil dari cuitan *Twitter*. Setelah itu, kumpulan data cuitan dimasukkan ke dalam perangkat lunak klasifikasi biner untuk menghasilkan polaritas positif dan bukan positif.

Hasil dari sentimen tersebut dapat dijadikan bahan kajian untuk pemerintahan, agar kebijakan vaksinasi dapat dievaluasi berdasarkan sentimen langsung dari masyarakat. Selain itu, hasil analisis sentimen ini juga dapat berfungsi sebagai media komunikasi masyarakat secara *non-verbal*, sehingga tanpa ada komunikasi dari dewan perwakilan rakyat, pemerintah dapat membaca kondisi masyarakat dari dampak kebijakan pemerintah melalui rasio analisis sentimen.



Gambar 2 Proses Penelitian

Proses penelitian yang terdapat pada Gambar 2 merupakan alur dalam mengembangkan perangkat lunak klasifikasi biner pada penelitian ini.

A. Penelitian Pendahuluan

Diawali dengan melakukan pembuatan rumusan dan identifikasi permasalahan yang akan dikembangkan solusinya.

Pada tahun awal tahun 2020 adalah masa-masa tersulitnya seluruh bangsa

yang ada di dunia, termasuk negara Indonesia yaitu dengan datangnya wabah pandemi Covid-19. Kemudian berbagai negara berlomba-lomba untuk menciptakan vaksin untuk melakukan pencegahan dari wabah tersebut.

Di akhir tahun 2020, pemerintahan negara Indonesia telah memasok vaksin yang telah dibeli dari negara China yang bernama *Sinovac* [6]. Kemudian pada awal tahun, tepatnya pada tanggal 13 Januari 2021, Presiden Republik Indonesia Ir. Joko Widodo melakukan vaksinasi di istana kepresidenan yang disiarkan langsung melalui *Youtube* sekretariat presiden [2].

Banyak opini masyarakat terhadap vaksinasi yang akan diprogramkan dan didistribusikan ke berbagai wilayah Indonesia. Hingga media sosial *Twitter* mengangkat topik terhangatnya yang bertagar Vaksinasi Nasional. Banyak masyarakat yang menyumbang opininya dalam konteks positif maupun negatif.

Berdasarkan topik permasalahan tersebut, sentimen masyarakat terhadap vaksinasi Covid-19 pada media sosial *Twitter* menjadi identifikasi masalah pada penelitian ini, dengan rumusan masalah yaitu bagaimana melakukan analisis sentimen dengan mengembangkan model klasifikasi biner sentimen masyarakat terhadap vaksinasi Covid-19 menggunakan *Support Vector Machine*?

Penelitian diawali dengan pencarian tinjauan pustaka mengenai analisis sentimen, yang berasal dari cakupan berbagai sumber penelitian yaitu melalui jurnal, *e-book*, artikel, dan prosiding.

Setelah tinjauan pustaka dilakukan, selanjutnya melakukan pengembangan perangkat lunak. Dalam hal ini, metode pengembangan perangkat lunak yang digunakan adalah inkremental. Inkremental digunakan karena memiliki model manajemen perancangan yang sederhana dan efektif dalam perancangan suatu pengembangan perangkat lunak.

Dalam penelitian ini, inkremental memiliki dua iterasi dalam pengembangannya, yaitu pada iterasi satu melakukan pelatihan dan pengujian terhadap model klasifikasi biner menggunakan algoritma SVM dan pada iterasi dua melakukan pembuatan tampilan antarmuka pengguna menggunakan kerangka kerja *Flask*.

B. Pengembangan Perangkat Lunak

1. Inkremental

Metode Inkremental adalah salah satu model *Software Development Life Cycle* (SDLC) yang digunakan untuk melakukan rekayasa pengembangan perangkat lunak (aplikasi) yang melibatkan tahapan analisis, desain, implementasi, dan pengujian. Dalam pengembangannya, metode ini hanya fokus terhadap inti produk yang dikembangkan. Secara umum metode inkremental melakukan penggabungan antara elemen aliran proses linier dan paralel di setiap iterasi (bersiklus). Pada setiap iterasi menghasilkan suatu produk sistem atau fitur dan fungsionalitas yang di evaluasi, kemudian dilanjutkan dengan iterasi yang baru. Proses inkremental dilakukan secara bertahap dan meningkat berbentuk aliran linier yang terintegrasi, hingga produk pengembangan perangkat lunak yang diharapkan bisa tercapai.

Dalam perancangannya, metode ini menghasilkan fitur secara bertahap atau bersiklus. Mulai dari fitur yang dasar, hingga fitur yang memerlukan banyak integrasi dari keseluruhan program. Metode ini hanya fokus terhadap inti produk yang dikembangkan [7].

Adapun penjelasan tahapan dari metode inkremental yang digunakan pada penelitian ini sebagai berikut :

1. Analisis

Tahapan ini melakukan analisis pada kebutuhan dari pengembangan

perangkat lunak. Hasil analisis memberikan informasi pada perangkat lunak yang dikembangkan berupa kekurangan, kelebihan, fungsi, dan perencanaan.

2. Desain

Tahapan ini melakukan perancangan arsitektur perangkat lunak yang menggambarkan komponen-komponen yang diwujudkan berupa *prototype*.

3. Implementasi

Tahapan ini melakukan tahap merancang kode program atau pemrograman, pembangunan sistem, dan realisasi desain dalam pengembangan perangkat lunak

4. Pengujian

Tahapan ini melakukan pengujian dari hasil implementasi pengembangan perangkat lunak, meliputi pengujian program, pengujian sistem, dan pengujian desain yang berpotensi menimbulkan *error* pada perangkat lunak.

Metode inkremental digunakan dalam penelitian ini, karena metode ini sederhana dan efektif dalam mengembangkan perangkat lunak, tanpa memerlukan kerja tim dan menghasilkan fitur program yang konsisten untuk diimplementasikan.

Pada penelitian ini terdapat dua iterasi dalam melakukan pengembangan perangkat lunak, yaitu iterasi satu untuk melakukan pelatihan dan pengujian model menggunakan metode SVM dan iterasi dua untuk melakukan pembuatan tampilan menggunakan kerangka kerja *Flask*.

C. Iterasi Satu

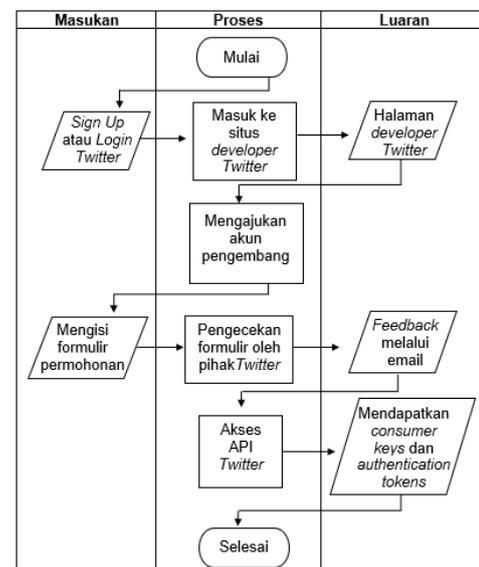
Iterasi satu pada penelitian ini mengembangkan klasifikasi biner

dengan melakukan pelatihan dan pengujian model untuk menghasilkan model yang dapat melakukan analisis sentimen.

1. Text Mining

Text mining adalah penambangan teks dengan cara menganalisis teks yang tidak terstruktur yang diperoleh dari kumpulan dokumen untuk mengekstraksi informasi dan secara otomatis mengidentifikasi pola serta hubungan yang menarik dalam data tekstual atau biasa disebut dengan teknik *crawling*. *Text mining* berfokus pada teks dari kumpulan dokumen yang tidak terstruktur untuk menyiapkan informasi yang terstruktur sebelum suatu kumpulan data tersebut dipersiapkan khusus untuk proses analisis [8].

Dalam hal ini data yang digunakan untuk penelitian ialah data yang diambil dari media sosial *Twitter*, menggunakan *API Twitter* dan *library Tweepy*.



Gambar 3 Proses memperoleh API *Twitter*

Proses *crawling* data dapat dilakukan dengan diperlukannya *API Twitter*. Adapun cara mendapatkan *API Twitter* dapat diketahui berdasarkan *flowchart* yang terdapat pada Gambar 3.

Untuk melakukan *crawling* data diperlukan pendekatan sesuai prosedur yang ditentukan oleh pihak *Twitter*, yaitu memerlukan API *Twitter* berupa *consumer keys* dan *authentication tokens* untuk mengakses data cuitan yang akan dikumpulkan. Untuk memperoleh API *Twitter*, diperlukan akun *Twitter* terlebih dahulu. Lalu membuka situs pengembang *Twitter* sebagai jalur akses dalam melakukan permintaan API *Twitter*. Setelah itu melakukan pengajuan permohonan menjadi akun pengembang dengan mengisi formulir yang disediakan oleh pihak pengembang *Twitter*. Setelah selesai, pihak *Twitter* melakukan pengecekan terhadap pengajuan yang dikirimkan, lalu diberikan respon berupa *email* dengan konteks apakah diterima atau tidak pengajuan tersebut. Jika pengajuan diterima, maka API *Twitter* sudah bisa diakses melalui situs pengembang *Twitter*.

Setelah semua sudah dilakukan, maka akses API *Twitter* berupa *consumer keys* dan *authentication tokens* dapat ditampilkan dan digunakan untuk pengembangan aplikasi pada penelitian ini. Pada penelitian ini melakukan *crawling* sebanyak 1000 data cuitan.

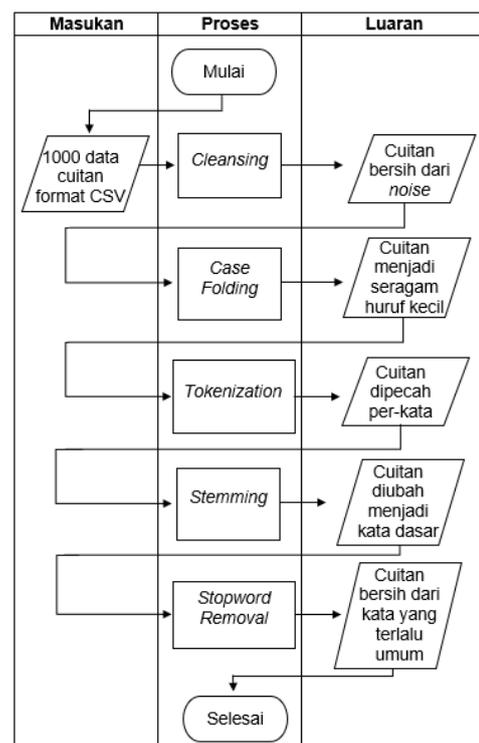
2. Text Preprocessing

Text Preprocessing adalah proses dalam *text mining* yang mengubah kumpulan data yang masih mentah serta tidak terstruktur menjadi data yang siap untuk di analisis. Proses ini dilakukan supaya atribut yang tidak memiliki pengaruh khusus dapat dihilangkan dan dijadikan data yang seragam. Hasil proses ini dapat membantu mempersiapkan *dataset* agar bersih dari *noise* atau atribut yang tidak diperlukan saat pengambilan informasi [9].

Tahap *text preprocessing* pada penelitian ini memiliki beberapa komponen proses yang dilakukan, yaitu

dengan normalisasi cuitan teks (*cleansing*), mengubah cuitan teks menjadi huruf kecil (*case folding*), melakukan pemecahan kalimat menjadi kata untuk membersihkan macam-macam tanda baca (*tokenization*), melakukan perubahan kata menjadi bentuk baku (*stemming*), melakukan penghapusan kata yang terlalu umum (*stopword removal*).

Setelah itu, *dataset* dilakukan ekstraksi ke dalam format dokumen *Comma Separated Values (CSV)*, agar dapat menjadi masukan (*input*) dari pelatihan model klasifikasi biner. Adapun *flowchart* dari tahap *preprocessing* terdapat pada Gambar 4.



Gambar 4 Tahap *preprocessing*

3. Klasifikasi Teks

Klasifikasi Teks adalah teknik otomatis pengkategorian teks ke dalam suatu kelas yang telah ditentukan berdasarkan isi dari teks tersebut. Misalkan pada artikel, jika suatu artikel membahas tentang pengetahuan atau studi kasus tentang kebijakan

pemerintah, pernyataan kepala daerah, pemilu, dan sebagainya, maka artikel tersebut dikategorikan dalam kelas politik. Klasifikasi teks berguna agar mempermudah pengkategorian berdasarkan teks dari data yang telah terkumpul, sehingga teks yang belum diketahui kategorinya dapat lebih efektif untuk diklasifikasi [10].

Pada umumnya suatu kumpulan data dapat diklasifikasikan ke dalam satu kelas (*single-label*) ataupun menjadi beberapa kelas (*multi-label*). Pada *single label*, suatu dokumen hanya dikategorikan ke dalam satu kelas. Kalau *multi label*, suatu dokumen dapat dikategorikan ke dalam beberapa kelas secara terpisah. Dalam pengembangannya, klasifikasi teks memiliki berbagai macam algoritma yang umum digunakan dalam penelitian, antara lain seperti *Naïve Bayes Classifier*, *Support Vector Machine (SVM)*, *Neural Network*, *Decision Tree*, *Random Forest*, *K-Nearest Neighbor (KNN)*, dan lain sebagainya [10].

Pada penelitian ini, klasifikasi teks sangat berhubungan erat dalam pengembangan perangkat lunak klasifikasi biner, yaitu dengan mengklasifikasikan suatu teks kedalam kelas sentimen yang telah ditentukan menggunakan metode klasifikasi SVM.

3.1 Term Frequency-Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) adalah salah satu metode algoritma pembobotan pada fitur kata berdasarkan kemunculan kata. Algoritma TF-IDF memiliki dua bagian utama yaitu TF dan IDF. *Term Frequency (TF)* merupakan proses perhitungan jumlah kemunculan atau frekuensi kata (*term*) pada suatu dokumen, sehingga memberikan bobot lebih terhadap kata yang lebih sering muncul dalam dokumen [4].

Jika suatu *term* memiliki kesamaan di lain kalimat atau data, maka bobot TF akan bertambah. Dan jika tidak ada kesamaan dilain kalimat, maka bobot tidak akan ditambahkan. Adapun persamaan dari pembobotan TF sebagai berikut [4]:

$$W_{tf_{t,d}} = \begin{cases} 1 + \log_{10}(tf_{t,d}), & tf_{t,d} > 0 \\ 0, & tf_{t,d} = 0 \end{cases} \quad (1)$$

Keterangan :

$W_{tf_{t,d}}$: bobot *term frequency* dari suatu dokumen

$tf_{t,d}$: *term* yang sering muncul

Inverse Document Frequency (IDF) merupakan proses perhitungan bobot kata yang jarang muncul atau tidak mewakili suatu dokumen. Semakin sedikit jumlah dokumen yang mengandung term, maka nilai IDF akan semakin besar. IDF merupakan kebalikan dari nilai TF, yang bertujuan melakukan kompensasi terhadap *term* yang tidak sering muncul dalam dokumen [4].

Jadi, *term* yang tidak memiliki kemunculan di kalimat lain akan diberikan bobot yang lebih rendah dibandingkan dengan *term* yang memiliki kesamaan di kalimat lain. Adapun persamaan dari IDF sebagai berikut [4]:

$$idf_t = \log_{10}(N/df_t) \quad (2)$$

Keterangan :

idf_t : bobot *term* yang jarang muncul

N : jumlah dokumen teks

df_t : jumlah dokumen yang mengandung *term*

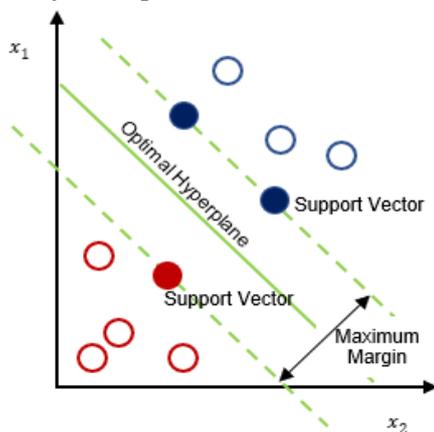
Penerapan TF-IDF akan merancang dokumen agar dapat menetapkan bobot setiap kalimat terhadap suatu dokumen atau *dataset*. Lalu setelah setiap kalimat atau data diberi bobot, kalimat akan diurutkan berdasarkan bobot kalimat paling besar. adapun persamaan dari penerapan TF-IDF sebagai berikut [11]:

$$TF_IDF = W_{tf_{t,d}} * \log_2 \frac{N}{idf_t} \quad (3)$$

Pada penelitian ini menggunakan metode algoritma TF-IDF untuk mengubah *dataset* menjadi lebih optimal dengan pembobotan kata dan melakukan ekstraksi fitur sebagai masukan data dari model klasifikasi biner sentimen.

3.2. Support Vector Machine

Support Vector Machine (SVM) adalah metode *machine learning* untuk model *supervised learning* yang memiliki ruang hipotesis berupa fungsi linier dalam ruang fitur (*feature space*) dengan memaksimalkan *hyperplane* terbaik yang dapat melakukan klasifikasi data dan analisis regresi. *Hyperplane* adalah sebuah fungsi yang dapat memisahkan data antar kelas pada dimensi yang tinggi, di atasnya fungsi *line whereas* (2-D) dan *plane similarly* (3-D). SVM lebih dikenal sebagai *maximum margin classifier*, disebabkan SVM melakukan klasifikasi melalui pendekatan alternatif untuk memperoleh *margin* yang maksimal yaitu dengan cara mencari *support vector*. SVM pertama kali diperkenalkan pada tahun 1992 oleh Vladimir Vapnik, Bernhard Boser, dan Isabelle Guyon [9]. Adapun ilustrasi metode klasifikasi SVM ditunjukkan pada Gambar 2.



Gambar 5 Ilustrasi Metode *Support Vector Machine*
(Diadaptasi dari sumber [4])

Pada dasarnya SVM berprinsip memprediksi pada dasar kasus klasifikasi secara linier (*linier classifier*), namun pengembangan SVM juga menghasilkan masukan konsep kernel pada *feature space* atau ruang fitur berdimensi tinggi, sehingga dapat memproses pada *problem non-linier* [12]. Adapun persamaan dari metode klasifikasi SVM sebagai berikut :

$$f(x) = w \cdot x + b \quad (4)$$

atau

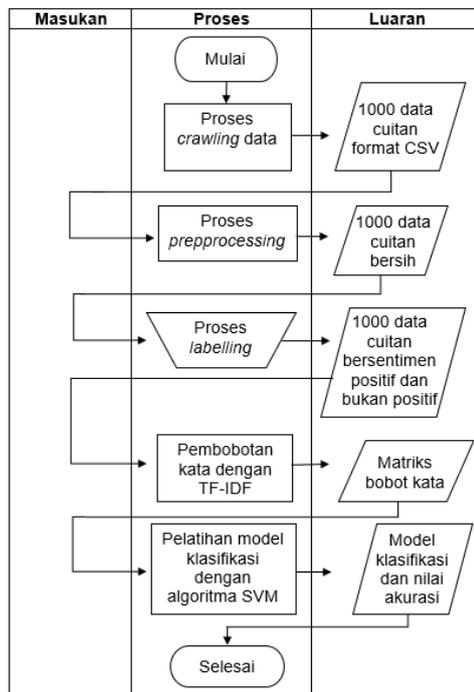
$$f(x) = \sum_{i=1}^m a_i y_i K(x, x_i) + b \quad (5)$$

Keterangan :

- w : parameter *hyperplane* yang dicari (garis yang tegak lurus antara garis *hyperplane* dan titik *support vector*)
- x : titik data masukan *Support Vector Machine* (SVM)
- a_i : nilai bobot setiap titik data
- $K(x, x_i)$: fungsi kernel
- b : parameter *hyperplane* yang dicari (nilai bias)

Pada penelitian ini menggunakan metode klasifikasi SVM untuk tahap pelatihan model klasifikasi. Penelitian ini mengusulkan metode SVM, karena pada penelitian yang dilakukan oleh Umi Rofiqoh, Rizal Setya Perdana, dan M. Ali Fauzi [4] terdapat analisis komparasi metode klasifikasi yang telah dilakukan pada penelitian sebelum mereka. Hasil analisis tersebut menunjukkan bahwa SVM menjadi metode klasifikasi yang lebih dominan dari sisi performansi model dibandingkan dengan metode klasifikasi lainnya.

4. Pelatihan dan Pengujian Model



Gambar 6 Pelatihan model klasifikasi biner

Pada iterasi satu, tahap pelatihan model klasifikasi yang terdapat pada Gambar 6 diperlukan untuk merancang model klasifikasi yang dapat menjadi pembelajaran mesin untuk melakukan klasifikasi teks pada data teks yang ingin diketahui sentimennya.

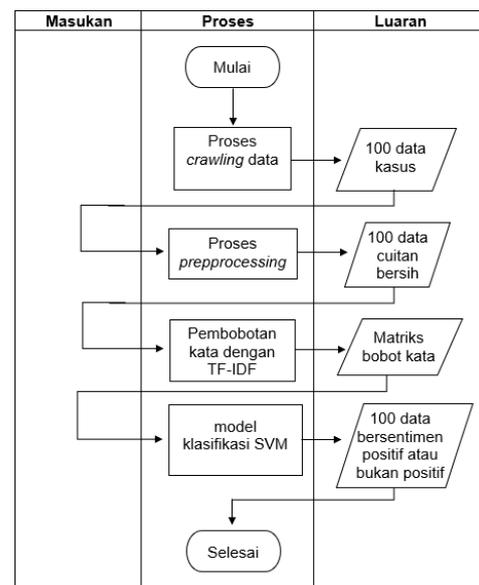
Perancangan model klasifikasi sentimen pada data teks cuitan *Twitter* diawali dengan melakukan pengumpulan *dataset* yaitu dengan teknik *crawling* data. *dataset* pada penelitian ini yang dikumpulkan untuk dilakukan pelatihan model sebanyak 1000 data dengan format CSV.

Dataset yang telah terkumpul, selanjutnya dilakukan proses *preprocessing* agar *dataset* dapat menjadi masukan yang bersih dari *noise* atau kata/karakter yang tidak bisa dilakukan pembelajaran mesin dan menjadi data yang terstruktur.

Kemudian data dilakukan *labelling*, guna menjadi masukan yang dapat dilatih pada model klasifikasi. Pada proses ini, *dataset* akan terbagi menjadi dua kelas, yaitu kelas positif dan bukan

positif. *Labelling* dilakukan dengan cara manual, berdasarkan kontekstual dan objektif terhadap vaksinasi dan vaksin. Kelas positif diinisialisasikan dengan angka 1 dan kelas bukan positif diinisialisasikan dengan angka 0.

Setelah seluruh tahap persiapan data telah dilakukan semua, selanjutnya *dataset* yang sudah siap masuk ke dalam tahap pelatihan model klasifikasi biner. Pelatihan model diawali dengan melakukan pembobotan kata dan ekstraksi fitur menjadi vektor dengan menggunakan metode TF-IDF. Lalu data yang telah di ekstraksi menjadi vektor dilakukan tahap klasifikasi biner dengan menggunakan metode SVM. Sehingga, hasil dari pemodelan klasifikasi biner dapat dijadikan model yang bisa menentukan sentimen pada tahap pengujian.



Gambar 7 Tahap klasifikasi sentimen

Pada tahap pengujian dengan dilakukannya tahap klasifikasi sentimen yang terdapat pada Gambar 7, menjadi tahap akhir pada iterasi satu. Dengan mengikuti logika pada tahap pelatihan model yang telah dikembangkan sebelumnya, tanpa melibatkan tahap *labelling* data dan tahap pelatihan klasifikasi SVM.

Pada tahap ini, data kasus yang dikumpulkan sebanyak 100 data cuitan. Lalu dilakukan proses *preprocessing* untuk menjadikan data yang terstruktur dan TF-IDF untuk menjadikan data sebagai masukan data vektor yang dapat dilakukan pembelajaran mesin. Kemudian, data tersebut akan ditentukan sentimennya, melalui model klasifikasi biner yang telah dikembangkan dengan menggunakan metode SVM.

Pada tahap klasifikasi menghasilkan polaritas sentimen berdasarkan pembelajaran mesin yang telah dilatih, yaitu sentimen positif dan bukan positif.

C. Iterasi Dua

1. Flask

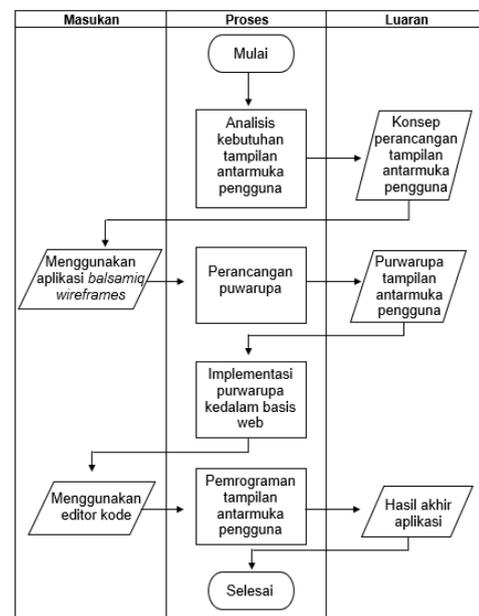
Flask adalah kerangka kerja web berbasis mikro yang menggunakan dependensi atau *library jinja* dan *werkzeug*. *Flask* dapat digunakan dengan penulisan bahasa pemrograman *Python*. *Flask* disebut web berbasis mikro karena kerangka kerja ini memiliki sistem yang sederhana dan mudah untuk diimplementasikan, namun memiliki fungsionalitas yang relevan dan fleksibel dalam pengembangan web. Kerangka kerja ini secara *default* tidak memberikan akses abstraksi *database*, validasi formulir, dsb. Tetapi kerangka kerja *Flask* membuka akses untuk mendukung ekstensi dari kerangka kerja lain ataupun bahasa pemrograman lain dalam melakukan integrasi pengembangan web [13].

Pada penelitian ini melakukan perancangan tampilan antarmuka pengguna menggunakan kerangka kerja *Flask* yang ditampilkan dalam bentuk web. *Flask* diusulkan sebagai tampilan antarmuka pengguna pada penelitian ini, karena bahasa pemrograman yang selaras dengan tahap pengembangan pelatihan model klasifikasi dan juga memiliki konsep yang sederhana dalam perancangan sistem.

2. Pembuatan Tampilan

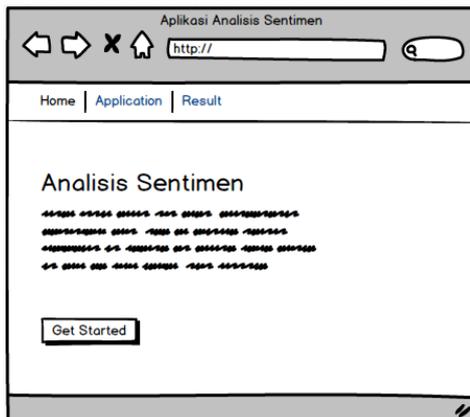
Pada iterasi dua dilakukan pembuatan tampilan menggunakan kerangka kerja *Flask*, agar fitur klasifikasi sentimen dapat sederhana digunakan oleh pengguna (*user*).

Berdasarkan analisis yang dilakukan, tampilan antarmuka pengguna yang digunakan adalah perangkat lunak berbasis web. Adapun tahapan perancangan tampilan antarmuka pengguna terdapat pada Gambar 8.



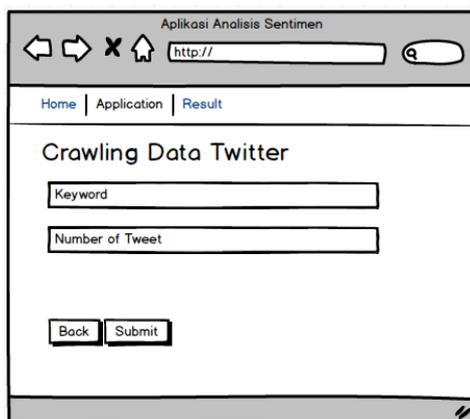
Gambar 8 Perancangan tampilan antarmuka pengguna

Kemudian hasil dari tahapan perancangan tampilan antarmuka pengguna divisualisasikan kedalam bentuk purwarupa. Yang bertujuan untuk membuat desain tampilan yang akan dirancang dan direalisasikan kedalam perangkat lunak klasifikasi biner. Adapun desain dari purwarupa untuk perangkat lunak terdapat pada Gambar 9 sampai Gambar 11.



Gambar 9 Purwarupa: halaman utama

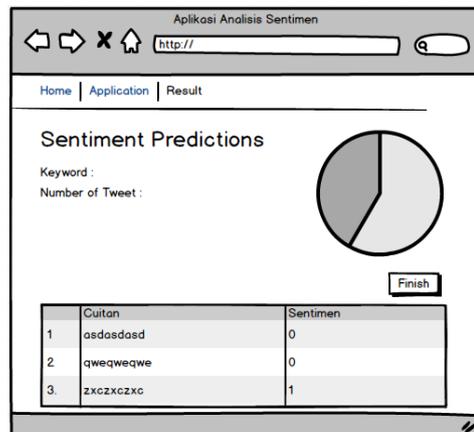
Pada tampilan purwarupa halaman utama yang terdapat pada Gambar 9 memiliki tiga menu navigasi yang ditampilkan pada bagian *header* (atas). Pada halaman utama akan memberikan antarmuka yang cukup sederhana, dengan sedikit penjelasan dari analisis sentimen. Lalu terdapat ada tombol yang berlabel *get started* yang artinya jika pengguna ingin menjalankan aplikasi, maka supaya dilakukan klik saja tombol tersebut.



Gambar 10 Purwarupa: halaman aplikasi

Selanjutnya tampilan antarmuka pengguna memiliki halaman aplikasi yang terdapat pada Gambar 10. Pada halaman aplikasi meminta dua masukan kedalam kolom formulir untuk melakukan proses *crawling* data *Twitter*. Adapun kolom formulir tersebut terdiri dari kolom kata kunci dan kolom jumlah data cuitan yang akan dikumpulkan. Lalu juga ada tombol *submit* yang akan

melakukan proses dari permintaan tersebut.



Gambar 11 Purwarupa: halaman hasil

Tampilan akhir dari antarmuka pengguna adalah halaman hasil yang terdapat pada Gambar 11. Pada halaman ini, hasil dari pemrosesan klasifikasi biner akan dikirim dan ditampilkan dalam bentuk data maupun visualisasi. Adapun data yang dihasilkan berupa data cuitan yang berhasil dikumpulkan dan data yang divisualisasikan dalam bentuk diagram *pie*. Selain itu, ada tombol yang berlabel *finish* yang digunakan untuk kembali ke halaman utama dari aplikasi ini.

III. HASIL DAN PEMBAHASAN

A. Iterasi Satu

1. Hasil

Iterasi satu memberikan hasil pelatihan model klasifikasi biner yang dapat melakukan analisis sentimen berdasarkan pelatihan model yang telah dilakukan menggunakan metode SVM.

Berdasarkan hasil rancangan pelatihan model klasifikasi biner, tahap awal yang dilakukan adalah tahap *crawling* data. Pada tahap ini memberikan hasil pengumpulan data sebanyak 4766 data cuitan. Sebab,

parameter yang diambil sebanyak 100 data cuitan per halaman terbaru tanpa batas. Adapun hasil dari tahap *crawling* data terdapat pada Gambar 12.

cuitan	
1	b'KAPOLRES Dampingi Gubri Tinjau Vaksinasi Mas...
2	b'Vaksinasi kapan kok nggak nyambung tahun per...
3	b'Ada 17 ribu perusahaan yang mengikuti progra...
4	b'Vaksinisasi Tahap II Bagi Pejabat Karantina ...
5	b'Koordinator Vaksinasi RSNU Banyuwangi dr Riz...
...	...
4762	b'Bisa2nya kelewatan jadwal vaksin'
4763	b'hadeh hari ini vaksin kedua euy males pol'
4764	b'Besok vaksin kedua, semoga ga kerasa aneh2'
4765	b'Takut mau vaksin, ntar tambah mager kan baha...
4766	b'Alhamdulillah hari ke-2 setelah vaksin udh g...

4766 rows × 1 columns

Gambar 12 Hasil *crawling* data

Kemudian, dilanjutkan ke tahap *preprocessing*. Tahap *preprocessing* yang dirancang dapat memberikan hasil *dataset* yang bersih dari kata atau karakter yang tidak diperlukan dalam melakukan analisis sentimen. Sehingga, *dataset* dapat menjadi masukan data yang terstruktur. Adapun rancangan tersebut terdiri dari proses *cleansing*, *case folding*, *tokenization*, *stemming*, dan *stopword removal*. Lalu pada tahap ini juga melakukan filter data, untuk mencapai data yang ditargetkan pada penelitian ini. Sehingga, setelah dilakukan filter, *dataset* memiliki data sebanyak 1000 data. Adapun persamaannya dari tahap *preprocessing* terdapat pada Gambar 13.

cuitan	
1	ruf amin senior muslim indonesia contoh vaksin...
2	alm vaksin minggu april stlh vaksin astrazanec...
3	vaksin covid jokowi selar tindak negara kaya s...
4	ribu usaha catat program vaksin gotong royong ...
5	usaha juta kerja terima vaksin covid
...	...
996	syarikat mohon bantu wang raja rana tekan kes ...
997	uts didik seni dosen vaksin uts senin serempak
998	biar test drive vaksin covid udh kena covid
999	pemkab sleman gelar giat vaksinasi covid muka ...
1000	efek vaksin rasain yaituuu ngantuk pollll tpi...

1000 rows × 1 columns

Gambar 13 Hasil tahap *preprocessing*

Selanjutnya, untuk tahap *labelling* memberikan hasil *dataset* yang memiliki polaritas berupa sentimen positif dan bukan positif. Tahap ini dilakukan dengan cara manual, dengan penentuan cuitan data yang berobjektif terhadap kalimat Vaksinasi dan Vaksin. Adapun *labelling* dilakukan dengan cara, inisialisasi angka 1 dianggap sebagai sentimen positif dan inisialisasi angka 0 dianggap sebagai sentimen bukan positif. Hasil dari tahap *labelling* terdapat pada Gambar 14.

sentimen		cuitan	
1	0	aah kamu vaksin pake nasi wqwqwqwq vaksinasi b...	
2	0	baru aja nih lihat berita inews dua orang meni...	
3	0	zul takut mcm apa pandang ingat kite boleh tol...	
4	0	bpom eropa menemukan hubungan pembekuan darah ...	
5	0	habis vaksin masih pusing mual begini ditambah...	
...	
996	1	meski mengandung babi wapres minta vaksinasi m...	
997	1	menunggu vaksin kedua	
998	1	metode jemput bola akan digunakan pemerintah k...	
999	1	meski belum pasti dalam rangka menjalankan per...	
1000	1	mettanews arab saudi perbolehkan umrah dan haj...	

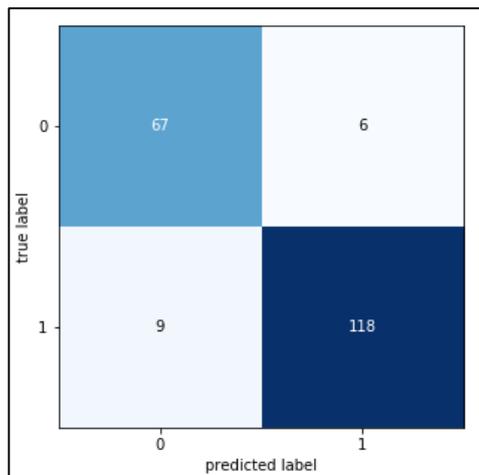
1000 rows × 2 columns

Gambar 14 Hasil *labelling*

Kemudian, selanjutnya data dilakukan pembelajaran mesin dengan metode pembobotan kata TF-IDF dan metode klasifikasi SVM. Tahap ini diawali dengan melakukan pemecahan data (*splitting*) dengan persebaran 80% sebagai data latih dan 20% sebagai data uji.

Berikutnya, data dilakukan ekstraksi fitur kata menjadi vektor dan pembobotan kata menggunakan metode TF-IDF. Setelah itu dilakukan pelatihan model klasifikasi biner menggunakan metode klasifikasi SVM.

Hasil laporan performansi dari pelatihan model di tampilkan ke dalam bentuk tabel *confusion matrix*, yang terdapat pada Gambar 15.



Gambar 15 Hasil *confusion matrix*

Confusion matrix pada penelitian ini menghasilkan 67 sebagai nilai prediksi tepat pada kelas positif, 118 sebagai nilai prediksi tepat pada kelas negatif (bukan positif), 6 sebagai nilai prediksi keliru pada kelas positif, dan 9 sebagai nilai prediksi keliru pada kelas negatif. Tabel ini menunjukkan, bahwa lebih banyak prediksi yang tepat pada kelas sentimen negatif, dari pada ketepatan prediksi terhadap kelas sentimen positif.

2. Pengujian

Setelah melakukan pelatihan model klasifikasi menggunakan metode klasifikasi SVM, selanjutnya dilakukan pengujian menggunakan data kasus yang didapatkan melalui teknik *crawling* data. Pada pengujian ini, dilakukan pengumpulan data sebesar 100 data *cutitan*.

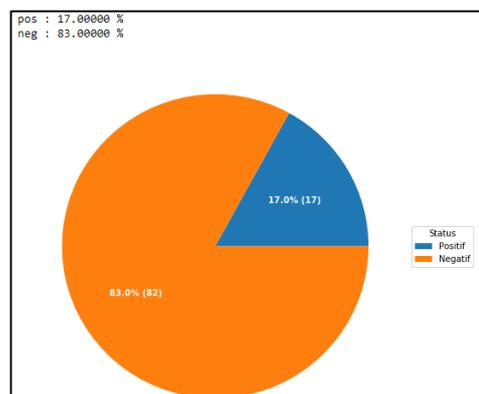
Setelah data dikumpulkan, data dilakukan sama seperti data latih pada

pelatihan model klasifikasi, yaitu dengan melewati tahap *preprocessing* dan pembobotan kata TF-IDF, tanpa melalui tahap *labelling* data dan metode klasifikasi SVM. Setelah tahap TF-IDF, data dilakukan klasifikasi biner menggunakan model klasifikasi SVM yang telah dilatih menggunakan 1000 data. Hasil dari analisis sentimen terdapat pada Gambar 16.

	cutitan	cutitan_prep	sentimen
0	b'Ariangga Hartarto: Harga Vaksin Mandiri Seg...	airiangga hartarto harga vaksin mandiri	0
1	b'Update / Sebaran Vaksin 3 Mei 2021 Pukul 17...	update sebar vaksin mei wita	0
2	b'Update / Sebaran Vaksin 3 Mei 2021 Pukul 17...	update sebar vaksin mei wita	0
3	b'Panggilan vaksin cepat je jawa: wPanggilan ...	panggil vaksin cepat jawa wPanggilan tuhan rel...	0
4	b'Sebagai tanda sokongan kepada Program Imunis...	tanda sokong program imunisasi covid bangsa pi...	0
...
95	b'I was thinking, kan ramai dah daftar nak vak...	was thinking ramai dah daftar nak vaksin notf...	0
96	b'Alur Vaksinasi Covid-19 Dipangkas Jadi 2 Mej...	alur vaksinasi covid pangkas meja	1
97	b'Sejak Vaksinasi, Ada 143 Kasus KIP di Jawa ...	vaksinasi kipi jawa barat	1
98	b'dokfess! Jadi koas kok begini banget yatuD!	dokfess koas banget jaga malam jam jam pagi ja...	0
99	b'Ahli Parlimen DAP pesa penjelasan pecahan da...	ahli parlimen dap pesa pecah dana kwan vaksin ...	0

Gambar 16 Hasil klasifikasi biner

Hasil dari analisis sentimen tersebut dipresentasikan kedalam bentuk diagram *pie*. Hasil analisis sentimen pada tahap pengujian menggunakan 100 data, memberikan hasil persentase 83% untuk sentimen bukan positif dan 17% untuk sentimen positif. Adapun hasil persentasi terdapat pada Gambar 17.



Gambar 17 Hasil diagram *pie*

3. Pembahasan

Berdasarkan pelatihan dan pengujian yang telah dilakukan pada iterasi satu, hasil pelatihan model klasifikasi biner pada data latih sebanyak 1000 data dari *cutitan Twitter*, memperoleh performansi model dengan skor akurasi 0,925.

Data latih yang terkumpul memiliki polaritas perbandingan 611 data untuk cuitan bersentimen positif dan 389 data untuk cuitan bersentimen bukan positif.

Adapun pelatihan model klasifikasi dilakukan menggunakan pembelajaran mesin dari *library scikit-learn* yakni algoritma SVM. Algoritma SVM merupakan metode klasifikasi atau klasifikator dalam penelitian ini, yang dapat memberikan jarak maksimal antar kelas pada pelatihan model klasifikasi menggunakan *hyperplane*.

Performansi dari pelatihan model klasifikasi dilanjutkan dengan dilakukannya evaluasi terhadap model. Evaluasi model bertujuan untuk mengetahui kinerja mesin pembelajaran dari suatu model klasifikasi. Dalam penelitian ini evaluasi terhadap model klasifikasi yang digunakan adalah *confusion matrix*. *Confusion matrix* memberikan hasil perhitungan evaluasi berupa *accuracy*, *precision*, *recall*, dan *f-measure*. Adapun hasil evaluasi pada pelatihan model klasifikasi terdapat pada Tabel 1.

Tabel 1 Hasil laporan evaluasi pelatihan model

<i>Evaluasi</i>	<i>Skor</i>
<i>Accuracy</i>	0,925
<i>Precision</i>	0,951
<i>Recall</i>	0,929
<i>F-Measure</i>	0,94

Kemudian, pada tahap pengujian dilakukan menggunakan data kasus dengan melakukan pengumpulan cuitan sebanyak 100 data. Menghasilkan perbandingan 17% sentimen positif dan 83% sentimen bukan positif. Setiap data yang diambil satu kali waktu tidak bisa diambil untuk suatu kesimpulan, membutuhkan data beberapa periode waktu yang berbeda untuk memastikan sentimen masyarakat dalam menanggapi vaksinasi Covid-19 di negara Indonesia. Dalam penelitian ini model yang telah dilatih dapat memberikan implikasi berupa prediksi polaritas, baik sentimen positif maupun bukan positif.

B. Iterasi Dua

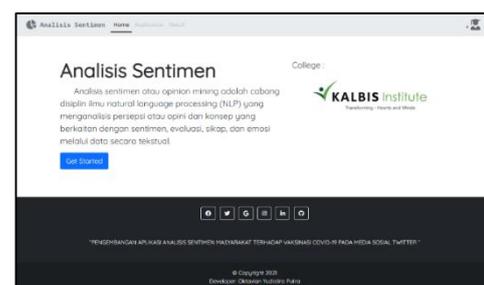
1. Hasil

Iterasi dua memberikan hasil berupa tampilan antarmuka pengguna, yang memiliki 3 halaman web dari 5 file pembangunan kode program.

Tampilan pada penelitian ini, dibangun menggunakan kerangka kerja *Flask*.yang ditulis menggunakan bahasa pemrograman *Python* dan *HTML*. Pada penulisan kode program, penelitian ini menggunakan fitur *template inheritance* yang dimiliki oleh *Flask*. Fitur ini memberikan kemudahan dalam pengembangan tampilan antarmuka pengguna.

Tampilan antarmuka pengguna yang dibuat terdiri dari 3 halaman web, yaitu halaman utama, halaman aplikasi, dan halaman hasil.

Hasil pada halaman utama, terdiri dari definisi analisis sentimen, nama perguruan tinggi tempat penelitian, dan tombol “*Get Started*” untuk memulai menjalankan aplikasi. Hasil tampilan antarmuka pengguna halaman utama terdapat pada Gambar 4.

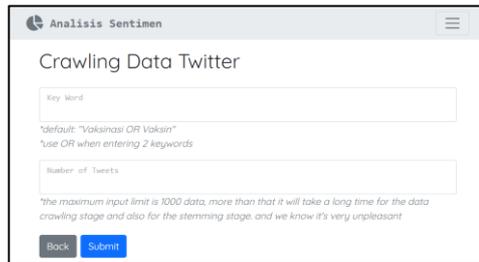


Gambar 18 Hasil halaman utama

Hasil pada halaman aplikasi terdapat komponen kolom formulir dan tombol. Pada kolom formulir terdiri dari kata kunci dan jumlah cuitan yang membutuhkan masukan untuk diproses. Tombol halaman aplikasi terdiri dari dua buah, tombol berlabel “*Back*” berfungsi untuk mengembalikan ke halaman utama dan tombol berlabel “*Submit*” berfungsi untuk melakukan pengolahan

masukan yang diisi pada kolom formulir.

Kolom formulir telah dilakukan teknik validasi yang dapat membantu aplikasi mendapatkan masukan formulir yang sesuai dari pengguna. Adapun hasil tampilan antarmuka pengguna halaman aplikasi terdapat pada Gambar 19.

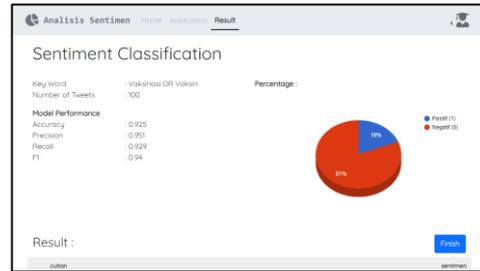


Gambar 19 Hasil halaman aplikasi

2. Pengujian

Setelah melakukan pembangunan tampilan antarmuka pengguna, selanjutnya perangkat lunak dilakukan pengujian pada halaman hasil. Pengujian dilakukan dengan melakukan masukan kata kunci “Vaksinasi OR Vaksin” dan mengambil data cuitan sebesar 100 data pada halaman aplikasi. Lalu, pada halaman hasil akan menampilkan hasil pengolahan masukan yang diberikan berupa performansi model yang telah dilakukan pelatihan berupa perhitungan evaluasi yang terdiri dari *accuracy*, *precision*, *recall*, *f-measure*. Selain itu, juga terdapat klasifikasi biner sentimen dari data yang telah dikumpulkan dan persentase diagram *pie*.

Pengujian pada tampilan antarmuka pengguna menghasilkan perangkat lunak yang sesuai berdasarkan perancangan yang dibangun, yaitu perangkat lunak berbasis web dapat melakukan analisis sentimen dengan model klasifikasi biner, seperti yang telah dikembangkan pada iterasi satu. Adapun hasil tampilan antarmuka pengguna halaman hasil terdapat pada Gambar 20.



Gambar 20 Hasil halaman hasil

3. Pembahasan

Berdasarkan pembangunan tampilan antarmuka pengguna yang telah dikembangkan, perangkat lunak klasifikasi biner dapat memberikan tampilan berbasis web yang dirancang menggunakan kerangka kerja *Flask*.

Penelitian ini memilih menggunakan kerangka kerja *Flask* karena memiliki persamaan dalam bahasa pemrograman. Sehingga dapat memudahkan pengembangan perangkat lunak dalam melakukan integrasi.

Hasil akhir dari pengembangan perangkat lunak klasifikasi biner yaitu memiliki tampilan antarmuka pengguna yang dapat digunakan untuk melakukan klasifikasi biner sentimen masyarakat terhadap vaksinasi Covid-19 pada cuitan *Twitter* yang telah dilakukan pengumpulan data.

IV. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, maka dapat diambil kesimpulan mengenai pengembangan perangkat lunak klasifikasi biner sentimen masyarakat terhadap vaksinasi Covid-19 menggunakan *support vector machine*, yaitu sebagai berikut :

1. Hasil dari pengembangan perangkat lunak dapat mengetahui kecenderungan sentimen masyarakat terhadap vaksinasi Covid-19 dengan model klasifikasi biner yang telah dirancang menggunakan metode *Support Vector Machine* (SVM).

- Adapun hasil pengujian pada penelitian ini mendapatkan persentase sentimen bukan positif 83% dan sentimen positif 17% dari 100 data cuitan.
2. Metode pengembangan perangkat lunak pada penelitian ini menggunakan metode inkremental. Metode inkremental pada penelitian ini memiliki dua iterasi, iterasi satu melakukan pelatihan dan pengujian model dan iterasi dua melakukan pembuatan tampilan.
 3. Iterasi satu memberikan hasil model klasifikasi biner yang dapat melakukan analisis sentimen dengan cara klasifikasi biner yang terbagi menjadi dua kelas, yaitu kelas positif dan kelas bukan positif.
 4. Iterasi dua memberikan hasil tampilan antarmuka pengguna berbasis web. Adapun pembuatan tampilan dirancang menggunakan kerangka kerja *Flask*.
 5. Hasil dari tahap pelatihan model klasifikasi biner memberikan perhitungan evaluasi dengan perolehan skor *accuracy* sebesar 0,925, *precision* sebesar 0,951, *recall* sebesar 0,929, dan *f-measure* sebesar 0,94.
 6. Rancangan pelatihan model klasifikasi biner terdiri dari *crawling* data, *preprocessing*, *labelling*, metode ekstraksi fitur TF-IDF, dan metode klasifikasi SVM.

DAFTAR RUJUKAN

- [1] Rahajeng Kusumo Hastuti, "Siap-Siap! Kasus Covid-19 RI Akan Meledak Seminggu Lagi," *10 Februari*, 2021. <https://www.cnbcindonesia.com/news/202101165503-4-222529/siap-siap-kasus-covid-19-ri-akan-meledak-seminggu-lagi> (diakses 01 April 2021).
- [2] F. S. Dewi, "Jokowi Disuntik Vaksin Covid-19 Sinovac Hari Ini, Ibu Negara Kapan?," *13 Januari*, 2021. <https://kabar24.bisnis.com/read/20210113/15/1342247/jokowi-disuntik-vaksin-covid-19-sinovac-hari-ini-ibu-negara-kapan> (diakses 01 April 2021).
- [3] M. A. Fauzi and S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking Image Processing View project Smart Wheelchair View project," *Researchgate.Net*, vol. 2, no. 8, hlm. 2766, 2018, [Online]. Available: <https://www.researchgate.net/publication/322959527>.
- [4] U. Rofiqoh, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexion Based Feature," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 1, no. 12, hlm. 1726, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/628>.
- [5] N. Saputra, T. B. Adji, and A. E. Permanasari, "Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming Menggunakan Metode Naive Bayes dan SVM," *J. Din. Inform.*, vol. 5, no. November, p. 12, 2015.
- [6] N. S. Sagita, "Vaksin Corona Sudah Sampai di Indonesia, Ini Asal Negara dan Tingkat Keampuhannya," *detikHealth*, 2020. <https://health.detik.com/berita-detikhealth/d-5285492/vaksin-corona-sudah-sampai-di-indonesia-ini-asal-negara-dan-tingkat-keampuhannya> (diakses 01 April 2021).
- [7] R. S. Pressman, *Software Engineering: A Practitioner's Approach*, Seventh. New York: the McGraw-Hill Companies, 2010.
- [8] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "Using text mining techniques for extracting information from research articles," *Stud. Comput. Intell.*, vol. 740, no. January, hlm. 374, 2018, doi: 10.1007/978-3-319-67056-0_18.
- [9] F. S. Jumeilah, "Penerapan Support Vector Machine (SVM) untuk Pengkategorian Penelitian," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 1, p. 19, 2017, doi: 10.29207/resti.v1i1.11.
- [10] S. Widaningsih and A. Suheri, "Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian Web Of Science Dengan Menggunakan Text Mining," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, hlm. 322, 2018.
- [11] B. P. Sharifi, D. I. Inouye, and J. K. Kalita, "Summarization of twitter microblogs," *Comput. J.*, vol. 57, no. 3, hlm. 7, 2014, doi: 10.1093/comjnl/bxt109.
- [12] P. A. Octaviani, Y. Wilandari, and D.

Ispriyanti, “Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang,” vol. 3, hlm. 811–820, 2014.

[13] A. Ronacher, “Flask,” *1 April*, 2010. <https://flask.palletsprojects.com/en/1.1.x/foreword/> (diakses 04 Mei 2021).