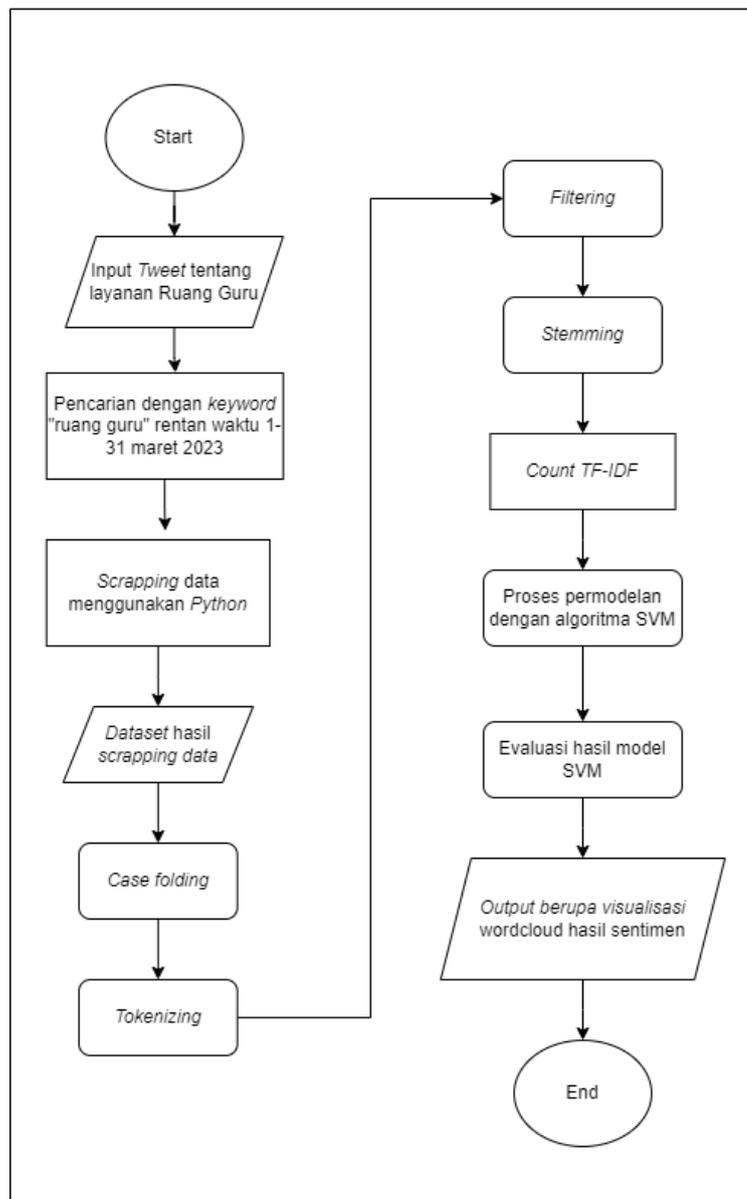


BAB III METODE PENELITIAN

3.1. Desain Penelitian

Desain penelitian dalam penelitian ini melibatkan serangkaian tahapan yang terstruktur dan sistematis desain penelitian ini dapat dilihat dari gambar 3.1 berikut.



Gambar 3. 1 Desain penelitian (Giovani et al., 2020)

Pada Gambar 3.1 dijelaskan mengenai alur dari proses penelitian dari awal hingga akhir. Proses penelitian dimulai dengan pengumpulan data menggunakan teknik *web scraping* dari *Twitter* dengan menggunakan kata kunci "ruang guru" dan memanfaatkan fitur *advanced search* untuk mengambil *tweet* yang diposting dalam periode waktu tertentu, yaitu 1 – 31 Maret 2023. Data yang diperoleh kemudian akan melalui *preprocessing*, yang mencakup pembersihan teks dari karakter khusus dan normalisasi, serta tokenisasi teks *tweet* untuk mempersiapkannya menjadi *dataset* yang siap digunakan.

Selanjutnya, *dataset* yang telah dibentuk akan dianotasi manual oleh peneliti untuk mengkategorikan sentimen dari setiap *review* pengguna sebagai positif atau negatif. Proses pembentukan *dataset* ini penting untuk melatih dan menguji model SVM. Algoritma Support Vector Machine akan diterapkan dalam tahap pelatihan model, di mana *dataset* yang telah dianotasi akan digunakan untuk melatih model SVM dengan tujuan untuk mengklasifikasikan sentimen dari *review* pengguna. Setelah model dilatih, tahap klasifikasi sentimen akan dilakukan, di mana model akan digunakan untuk menganalisis sentimen dari *tweet-tweet* yang belum pernah dilihat sebelumnya.

Untuk mengevaluasi kinerja model yang dihasilkan, metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* akan digunakan. Seluruh proses metodologi ini diharapkan dapat memberikan wawasan mendalam mengenai sentimen pengguna terhadap layanan Ruang Guru berdasarkan analisis data yang dilakukan dengan pendekatan algoritma *Support Vector Machine*.

3.2. Pengumpulan Data

3.2.1. Sumber Data

Dalam rangkaian penelitian ini, informasi yang relevan dihimpun dari komentar-komentar dalam bahasa Indonesia yang berasal dari platform media sosial populer, yaitu *Twitter*, yang mencakup aspek-aspek tertentu mengenai layanan Ruang Guru. Proses pengumpulan data dilakukan dengan memanfaatkan pendekatan teknis dalam bentuk teknik pengambilan data dari sumber online yang dikenal sebagai *web scraping*, yang diimplementasikan menggunakan bahasa

Bagus Haryanto, 2023

Penerapan Algoritma *Support Vector Machine* untuk Analisis Sentimen *Review* Pengguna pada *Twitter* Ruang Guru

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

pemrograman *Python* menggunakan *tools Google Colab*, sebuah platform berbasis *cloud computing* yang dikembangkan oleh *Google* (Guntara, 2023).

3.2.2. Teknik Pengumpulan Data

Proses pengumpulan data dilakukan dengan mengambil *dataset* yang bersumber dari platform media sosial *Twitter*, dengan menerapkan *keyword* "ruang guru" sebagai parameter utama. Penerapan *keyword* tersebut dilakukan melalui penggunaan fitur *advanced search* yang disediakan oleh *Twitter*, yang memungkinkan penyaringan secara tepat terhadap informasi yang relevan. Dalam upaya untuk memastikan bahwa hanya data yang memiliki relevansi waktu yang terkait dengan konteks penelitian yang diambil, batasan kronologis diterapkan dengan mengfokuskan pada *tweet* yang terpublikasikan dalam jangka waktu spesifik, yaitu antara tanggal 1-31 Maret 2023.

Pada tahap pencarian ini, dilakukan langkah-langkah untuk mengidentifikasi dan mengumpulkan data yang sesuai dengan kriteria waktu yang telah ditetapkan. Proses ini diilustrasikan secara rinci dalam Gambar 3.2 yang disajikan dalam penelitian ini. Gambar tersebut memberikan gambaran mengenai proses pencarian dan pengumpulan data, termasuk langkah-langkah teknis yang dilakukan dalam setiap fase. Dengan demikian, tahapan pengumpulan data yang digunakan dalam penelitian ini mencerminkan komitmen terhadap pendekatan metodologis yang sistematis dan terstruktur dalam memperoleh sumber informasi yang relevan.



Gambar 3. 2 Teknik pengumpulan data (Giovani et al., 2020)

3.3. Penerapan Metode CRISP-DM

Metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) adalah sebuah kerangka kerja yang sering digunakan dalam proyek *data mining*, termasuk analisis sentimen pada data sosial media seperti *Twitter*. Dalam proyek

Bagus Haryanto, 2023

Penerapan Algoritma *Support Vector Machine* untuk Analisis Sentimen Review Pengguna pada *Twitter Ruang Guru*

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

analisis sentimen untuk *review* pengguna pada layanan Ruang Guru di *Twitter*, tahapan-tahapan dalam CRISP-DM digunakan untuk mengelola proyek secara sistematis. Tahapan-tahapan ini meliputi *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*. Berikut penerapan algoritma *Support Vector Machine* (SVM) dilakukan dalam tahapan pemodelan untuk mengidentifikasi sentimen *review* pengguna pada *Twitter* Ruang Guru:

3.3.1. *Business Understanding*

Dalam tahapan awal yang disebut sebagai pemahaman bisnis, dilaksanakan suatu proses dalam rangka memperoleh pemahaman yang mendalam mengenai penelitian yang sedang dilakukan, dengan tujuan mencapai hasil yang konsisten dan sesuai dengan ekspektasi yang telah ditetapkan. Proses pemahaman ini memerlukan langkah-langkah yang strategis, yang meliputi upaya untuk menemukan dan menggali sumber-sumber informasi yang beragam dan relevan yang berkaitan dengan objek penelitian yang menjadi fokus.

3.3.2. *Data Understanding*

Pada tahap ini, dilakukan proses "*scrapping data*" dari platform media sosial *Twitter* untuk mengumpulkan data yang berisi sentimen atau pendapat pengguna tentang layanan Ruang Guru. Proses pengumpulan data dilakukan dengan mengambil teks dari tweet-tweet yang berkaitan dengan Ruang Guru menggunakan bahasa pemrograman *Python* dan perangkat lunak *Google Colab*.

Data yang berhasil dikumpulkan melalui proses *scrapping* kemudian akan diberikan label sentimen (positif dan negatif) secara manual menggunakan alat bantu perangkat lunak *Excel*. Selain itu, tahap ini juga mencakup eksplorasi data untuk memahami karakteristik dan kualitas data yang dihasilkan. Hal ini bertujuan untuk mengidentifikasi potensi batasan dan aspek-aspek yang relevan yang perlu diperhatikan dalam analisis sentimen selanjutnya.

3.3.3. *Data Preparation*

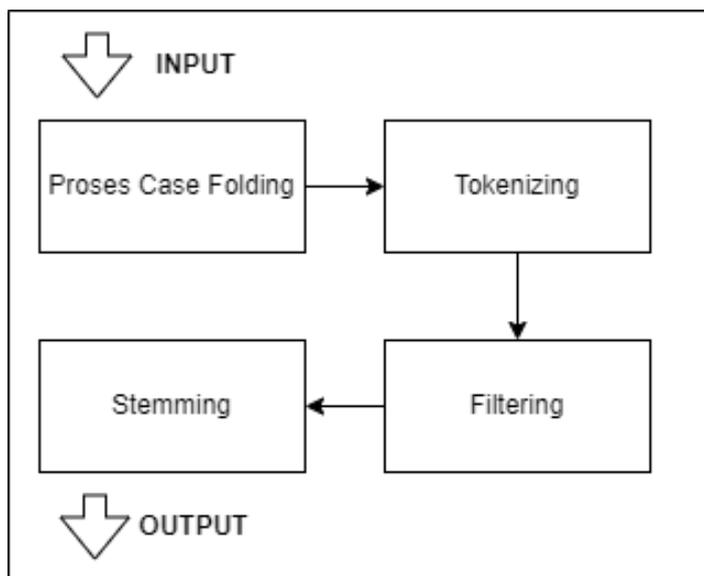
Tahap *data preparation* dilakukan sebagai langkah persiapan sebelum *dataset* dimasukkan ke dalam model. Proses ini melibatkan beberapa tahapan, antara lain:

Bagus Haryanto, 2023

Penerapan Algoritma *Support Vector Machine* untuk Analisis Sentimen *Review* Pengguna pada *Twitter* Ruang Guru

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

- 1) *Case folding*, bertujuan untuk mengebuah semua huruf dalam teks menjadi huruf kecil dan membersihkan tanda baca serta spasi ganda.
- 2) *Tokenizing*, bertujuan untuk memecah kalimat menjadi kata-kata individual yang membentuk kalimat tersebut.
- 3) *Filtering*, bertujuan untuk menghilangkan kata-kata yang penghubung yang tidak memiliki makna (stopwords).
- 4) *Stemming*, bertujuan untuk mengubah setiap kata menjadi kata dasar.



Gambar 3. 3 Alur kerja *data preparation* (Giovani et al., 2020)

3.3.4. Modelling

Dalam tahap ini, penerapan algoritma *Support Vector Machine* (SVM) digunakan untuk melakukan analisis sentimen pada data *Twitter* yang telah melewati tahap persiapan sebelumnya. SVM adalah salah satu algoritma klasifikasi yang banyak digunakan dalam analisis sentimen karena kemampuannya yang unggul dalam mengklasifikasikan data ke dalam kategori sentimen yang berbeda. Pada fase ini dilakukan perhitungan TF-IDF, dan dilakukan juga proses *splitting* dataset menjadi dua bagian, yaitu data uji (*testing*) dan data latih (*training*).

Pada langkah selanjutnya dalam penelitian ini, berfokus pada penerapan algoritma *Support Vector Machine* (SVM) sebagai landasan utama dalam menjalankan proses analisis sentimen. Salah satu kelebihan khusus dari SVM

terletak pada kapasitasnya yang unggul dalam melakukan pengelompokan data ke dalam kategori sentimen yang beragam, sebuah kualitas yang sangat relevan dalam konteks analisis sentimen yang dilakukan dalam penelitian ini.

Proses analisis dilakukan dengan pendekatan yang lebih terperinci. Langkah pertama melibatkan perhitungan TF-IDF (*Term Frequency-Inverse Document Frequency*). Langkah ini memungkinkan representasi vektor dari kata-kata dalam dokumen-dokumen *Twitter* yang telah dipersiapkan sebelumnya. Setelah representasi vektor terbentuk, dataset yang telah tersedia akan dilakukan *splitting* menjadi dua subset, yaitu data uji (*testing*) dan data latih (*training*), untuk keperluan pelatihan dan evaluasi algoritma.

3.3.5. *Evaluation*

Pada fase ini, dilakukan suatu proses evaluasi yang komprehensif terhadap hasil-hasil dari pemodelan yang telah dilakukan. Evaluasi ini dilakukan melalui penerapan berbagai *matrix*, termasuk pembentukan *confusion matrix*, penghitungan *accuracy*, *precision*, *recall*, serta nilai *F1-score*. Proses evaluasi ini dirancang untuk menganalisis sejauh mana model yang telah dimodelkan dengan menggunakan algoritma *Support Vector Machine* (SVM) mampu mengatasi permasalahan analisis sentimen.

Selanjutnya, hasil dari pemodelan yang mengadopsi algoritma SVM akan dibandingkan dengan kinerja algoritma lainnya, seperti *Naïve Bayes* dan *Decision Tree*. Pendekatan perbandingan ini bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai keefektifan relatif dari algoritma SVM dalam memprediksi sentimen berdasarkan ulasan pengguna. Proses komparatif ini akan memberikan gambaran yang lebih terperinci mengenai keunggulan dan keterbatasan dari masing-masing algoritma dalam konteks analisis sentimen.

Dengan demikian, tahap evaluasi ini memainkan peran penting dalam mengukur validitas dan kualitas dari pemodelan yang telah dilakukan, serta memberikan pemahaman yang lebih komprehensif mengenai kemampuan algoritma SVM dalam memproses data sentimen. Melalui analisis ini, diharapkan dapat ditemukan wawasan berharga yang dapat menjadi panduan

dalam pengembangan dan pengoptimalan algoritma analisis sentimen di masa mendatang.

3.3.6. *Deployment*

Hasil dari analisis sentimen *review* pengguna pada *Twitter* Ruang Guru dijelaskan dalam bentuk visualisasi *wordcloud* setimen *review* dan laporan. Laporan ini berisi hasil temuan dari tahap pemahaman bisnis, analisis data, dan evaluasi model. Laporan tersebut dapat dipresentasikan kepada pihak-pihak yang berkepentingan dan membantu Ruang Guru dalam memahami perspektif pengguna serta mengambil langkah-langkah yang tepat berdasarkan hasil analisis tersebut.

Konsekuensi dari analisis mendalam terhadap sentimen yang terkandung dalam *review* pengguna mengenai platform Ruang Guru di *Twitter* tercermin dalam bentuk visualisasi yang informatif, termasuk pula dalam bentuk *wordcloud* yang menggambarkan secara visual sebaran kata-kata yang mencirikan sentimen dalam ulasan tersebut. Selain itu, hasil analisis ini juga dijelaskan secara lebih rinci melalui penyusunan suatu laporan yang komprehensif.

Secara lebih detail, laporan tersebut mengulas hasil temuan yang dihasilkan dari masing-masing tahap penting dalam penelitian ini. Tahap pemahaman bisnis mempersembahkan konteks yang mendalam mengenai lingkup serta tujuan analisis, sementara tahap analisis data merinci pola-pola yang terungkap dalam data sentimen *review* pengguna. Di sisi lain, tahap evaluasi model melibatkan kajian mendalam terhadap performa dan efektivitas model analisis sentimen yang diterapkan, termasuk pula perbandingannya dengan model-model alternatif seperti algoritma *Naïve Bayes* dan *Decision Tree*.

Laporan ini tidak hanya memiliki nilai penting dalam pengembangan bisnis, tetapi juga sebagai alat komunikasi yang strategis. Laporan tersebut dapat dipresentasikan kepada berbagai pihak yang memiliki kepentingan, termasuk tim internal Ruang Guru, para pengambil keputusan, serta untuk referensi bagi peneliti selanjutnya. Dengan demikian, laporan ini bukan hanya sekadar menjadi dokumentasi hasil penelitian, tetapi juga menjadi instrumen yang efektif dalam

memberikan pandangan yang mendalam dan mendukung pengambilan keputusan yang berlandaskan pada analisis yang akurat dan terperinci.