

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

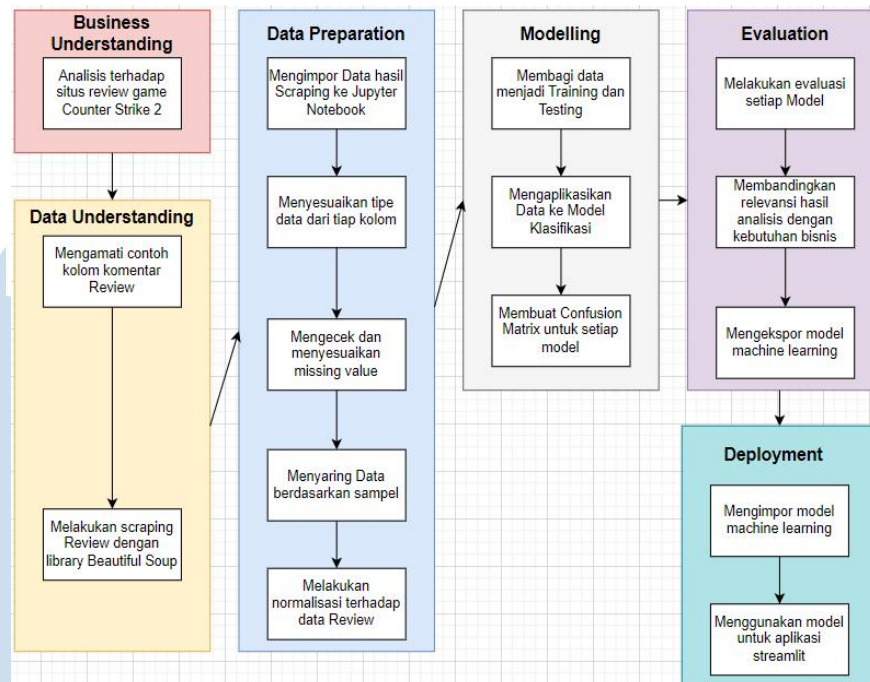
Penelitian ini akan menggunakan data sekunder dengan mengambil berbagai ulasan yang sudah diunggah pemain terhadap permainan *Counter Strike 2* dari *platform Steam* yang akan dinilai untuk diketahui teks manakah yang menjadi indikator terbesar dalam mempengaruhi seorang pemain untuk memberikan ulasan positif dan juga negatif.

Peranan teks yang menentukan suka atau tidaknya pemain akan diukur dengan menggunakan *scraping* data terhadap halaman portal Steam atas ulasan dari berbagai pemain akan dijadikan objek penelitian ini. Setelah melakukan ulasan pada *platform steam*, maka data akan diambil dan dianalisis sentimennya dengan bantuan algoritma. Setelah mendapatkan hasil analisis, maka didapati kata-kata kunci yang mempengaruhi ketidakpuasan dari pemain dalam memberikan ulasan.

3.2 Metode Penelitian

3.2.1 Alur Penelitian

Secara garis besar, penelitian ini dibagi menjadi tiga fase dimana fase pertama adalah analisis kebutuhan penelitian, dan setelah itu akan dilakukan implementasi terhadap hasil ulasan yang terdaftar di Steam dengan melakukan *scraping* data dan setelah mendapatkan data, maka analisis kuantitatif penelitian akan dilakukan untuk menyimpulkan hasil dari penelitian.



Gambar 3. 1 Alur penelitian

Gambar 3.1 Mengilustrasikan alur penelitian yang dilakukan. Berikut merupakan penjelasan dari setiap tahapan yang tercantum pada diagram:

1. *Business Understanding*

Tahapan ini merupakan tahapan pertama dalam penelitian ini. Analisis Kebutuhan diperlukan supaya peneliti dapat mendapatkan gambaran umum mengenai keperluan dan kegiatan yang harus dipenuhi untuk mengembangkan sebuah sistem. Dalam penelitian ini, maka hal yang pertama kali harus dilakukan adalah untuk mempersiapkan akun Steam yang sudah dimiliki untuk dapat mengakses halaman yang berisikan berbagai macam ulasan yang nantinya akan diteliti. Setelah memasuki halaman ulasan, maka halaman *review* akan diamati sejenak untuk memastikan bahwa *review* yang nantinya akan diambil berasal dari *game* yang benar.

2. *Data Understanding*

Setelah melakukan berbagai hal yang dibutuhkan pada tahapan *Business Understanding*, maka proses *Data Understanding* dapat dilaksanakan, dimana pertama-tama komentar pemain akan diamati untuk dilihat jenis data apa saja yang kira-kira akan ditarik. Dengan mengoptimalkan *library BeautifulSoup* dari *Python* yang bekerja dengan cara mengidentifikasi dan menyalin data berdasarkan elemen *Javascript* dari halaman *web*, maka diambil seluruh komentar, jam terbang, keputusan rekomendasi, tanggal, hingga nama pemain sebagai respon API. Seluruh aspek komentar tersebut nantinya akan diubah menjadi file dataset *Excel* yang nantinya akan digunakan untuk menuju tahapan selanjutnya, yaitu *Data Preparation*.

3. *Data Preparation*

Pada fase *Data Preparation*, maka seluruh data yang didapatkan dalam dataset akan dijabarkan dan dipreparasikan dengan menggunakan tools *Jupyter Notebook* dan bahasa pemrograman *Python*. Tahapan pertama yang dilakukan pada saat preparasi data adalah untuk memastikan bahwa data yang digunakan adalah data yang bersih, oleh karena itu berbagai komentar yang sama sekali tidak memiliki huruf alphabet harus disingkirkan. Hal ini juga dilakukan untuk memastikan algoritma analisis teks dapat berjalan dengan baik. Setelah data dibersihkan, maka akan dilakukan penyaringan lagi dimana data komentar yang akan digunakan didalam penelitian ini harus hanya merupakan komentar yang telah dibuat setelah rilisnya permainan *Counter Strike 2*, yaitu 27 September 2023 sampai 1 April 2024. Persiapan data terakhir akan dilakukan dengan cara menyaring lagi data tersebut dimana sampel yang akan diambil harus merupakan pemain yang memiliki jam terbang lebih besar dari 100 jam supaya dapat dipastikan bahwa pemain tersebut sudah memiliki pengalaman bermain yang cukup di versi sebelumnya juga. Data kolom komentar yang sudah diproses nantinya akan dinormalisasi lagi dengan menggunakan *Natural Language Processing (NLP)* dengan proses *tokenized*, *stopwords*, *lemmatized*, dan *normalize*. Setelah seluruh langkah-langkah tersebut dilakukan, maka data dapat dikatakan siap untuk dianalisis.

4. *Modelling*

Tahapan ini adalah proses dimana data yang sudah dipreparasikan dan dibersihkan akan dijadikan sebagai objek untuk keempat model *machine learning* yang ada. Setelah diuji maka akan dibentuk diagram berupa *Confusion Matrix* untuk memvisualisasikan hasil prediksi model.

5. *Evaluation*

Dalam tahapan ini, setiap model yang sudah dibuat akan diukur dan diuji tingkat keberhasilannya. Tingkat keberhasilan dapat dikatakan bagus apabila model memiliki nilai akurasi yang tinggi. Selain itu, hasil dari analisis data juga akan dibandingkan lagi dengan kebutuhan bisnis, dimana apabila sudah sesuai, maka model siap untuk menjalani fase *Deployment*.

6. *Deployment*:

Model yang sudah diekspor di tahapan sebelumnya akan diimplementasikan kedalam platform berbasis *web* yang dapat memprediksi sentimen dan jenis rekomendasi berdasarkan input dari penggunaanya

3.2.2 Metode Data Mining

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi umum yang digunakan dalam *data mining* untuk mengatasi proyek-proyek analisis data yang kompleks. Tabel 3.1 menjelaskan mengenai perbandingan antara metode *data mining CRISP-DM* dengan metode sejenis lainnya:

Tabel 3. 1 Perbandingan metode Data Mining

Nama	Kelebihan	Kekurangan
<i>Crisp-DM</i>	1. <i>CRISP-DM</i> menyediakan kerangka kerja yang terstruktur dan sistematis untuk mengelola proyek <i>data mining</i> dari awal hingga akhir. Ini membantu proyek untuk tetap terorganisir dan fokus pada langkah-	1. <i>CRISP-DM</i> mungkin kurang responsif terhadap perubahan lingkungan atau persyaratan proyek yang mungkin muncul selama proses. Ini bisa menjadi masalah jika proyek data mining berlangsung

	<p>langkah yang diperlukan dalam setiap tahap.</p> <p>2. CRISP-DM menekankan pemahaman yang mendalam terhadap kebutuhan bisnis dalam proses <i>data mining</i>. Ini membantu memastikan bahwa hasil dari proyek data mining secara langsung relevan dan bermanfaat bagi organisasi[22]</p>	<p>dalam lingkungan yang berubah dengan cepat.[38].</p>
<i>KDD</i>	<p>1. Kemampuan dalam menemukan pola, tren, dan pengetahuan baru yang mungkin tidak terlihat sebelumnya dalam data. Ini dapat membantu organisasi dalam mengidentifikasi peluang baru, meningkatkan efisiensi, atau memahami perilaku pelanggan[48].</p>	<p>1. Proses KDD sering kali kompleks dan memerlukan keterlibatan dari berbagai disiplin ilmu, termasuk statistik, matematika, dan pemodelan. Hal ini bisa memakan waktu dan biaya, terutama jika peneliti tidak memiliki sumber daya yang cukup[48].</p>
<i>SEMMA</i>	<p>1. <i>SEMMA</i> menekankan pemahaman yang mendalam terhadap kebutuhan bisnis dalam proses data mining. Ini memastikan bahwa hasil dari proyek data mining relevan dan bermanfaat bagi organisasi, dan membantu dalam mengidentifikasi peluang atau tantangan yang dapat dihadapi[49].</p>	<p>1. Metode <i>Data Mining Semma</i> hanya membahas hingga ke sisi <i>Assesment</i>, sehingga kurangnya [49].</p>

Berdasarkan perbandingan yang telah dilakukan pada Tabel 3.1, maka dipilihlah teknik *data mining CRISP-DM* untuk diaplikasikan kedalam penelitian ini. Hal ini dikarenakan salah satu dari hasil penelitian ini adalah dibentuknya sebuah aplikasi yang diimplementasikan berdasarkan model algoritma klasifikasi terbaik, dan juga ketentuan dan tujuan dilakukannya *data mining* sudah jelas disebutkan dari awal, yaitu untuk mengetahui tanggapan pemain terhadap *Counter Strike 2* berdasarkan komentar ulasan dari *Steam*.

Teknik lainnya seperti *KDD* juga tidak cocok untuk diimplementasikan di penelitian ini dikarenakan memakan waktu dan sumber daya yang banyak, dimana hal itu terbatas didalam penelitian ini. *SEMMA* dalam hal lain, juga kurang cocok untuk diterapkan didalam penelitian ini, dikarenakan dalam *SEMMA* tidak ada tahapan *Deployment* seperti yang dimiliki oleh *CRISP-DM*.

3.3 Teknik Pengumpulan Data

Dalam prosesnya, riset ini akan melibatkan pihak pemain yang harus memiliki pengalaman bermain selama lebih dari 100 jam dan juga ulasan yang akan diambil sebagai bahan analisis juga harus merupakan komentar yang dibuat setelah 27 September 2023 hingga 1 April 2024. Oleh karena itu, penelitian ini akan menggunakan teknik sampling Probability Sampling dengan metode *Criteria Sampling*. Alasan mengapa metode ini dipilih dapat terlihat pada objek penelitian yang memiliki berbagai syarat untuk bisa diikutsertakan pendapatnya sebagai objek didalam penelitian ini[40]. Biarpun demikian, data yang dikumpulkan dan digunakan dalam penelitian ini adalah data sekunder karena data yang diambil merupakan kumpulan ulasan yang sebelumnya sudah pernah diunggah oleh pemain.

3.3.1 Populasi dan Sampel

Populasi dalam penelitian ini mencakup seluruh pemain game Counter Strike yang memberikan *review* pada game tersebut. Populasi ini mencakup semua pemain yang memenuhi kriteria tersebut, tidak peduli seberapa banyak atau sedikit pemain tersebut, yaitu sebanyak 13289 data. Dengan memanfaatkan metode penghitungan sampel *Slovin* untuk margin kesalahan 0.15, maka jumlah sample yang dihasilkan yaitu sebanyak 3330 data. Penyaringan Sampel akan dilakukan lagi dengan menetapkan kriteria seleksi yang spesifik, seperti pemain yang memiliki pengalaman bermain minimal 100 jam dan memberikan *review* dalam 27 September 2023 sampai 1 April 2024. Total Sampel akhir yang didapat pada penelitian ini adalah sebanyak 1916 data.

3.3.2 Periode Pengambilan Data

Seperti dengan halnya yang sudah dijelaskan dalam bagian sampel, maka periode pengumpulan data akan dilakukan tepatnya di tanggal penelitian dilakukan, yaitu pada April 2024 dan pengumpulan data akan dilakukan oleh mesin secara cepat dalam bentuk *review scraping*.

3.4 Teknik Analisis Data

Penelitian ini akan dilaksanakan dengan mengoptimalkan teknik analisis data kuantitatif. Hal ini dikarenakan analisis sentimen dengan algoritma klasifikasi mengutamakan pemrosesan data secara matematis dan statistik. Melalui penerapan algoritma *Machine Learning*, teks dapat diklasifikasikan ke dalam kategori sentimen yang berbeda, seperti positif, negatif, atau netral, yang tercermin dalam bentuk nilai numerik atau proporsi. Metode ini memungkinkan pengolahan volume data besar dengan cepat dan efisien, serta mengandalkan kekuatan komputasi untuk melakukan analisis otomatis.

Dalam kasus ulasan *platform* Steam terhadap permainan *Counter Strike 2*, maka sentimen yang akan dinilai adalah sentimen positif dan negatif yang didapatkan dari ulasan pemain. Sentimen tersebut nantinya akan digunakan untuk membentuk model *Machine Learning* yang dapat mengukur secara numerik terhadap kata-kata yang terdapat didalam komentar.

3.5 Perbandingan Algoritma Klasifikasi

Teknik analisis data yang akan digunakan dalam penelitian ini adalah dengan mengoptimalkan berbagai algoritma berjenis klasifikasi, yakni *Naïve Bayes*, *Support Vector Machine*, *Random Forest*, dan *Logistic Regression*. Berikut adalah perbandingan kelebihan kekurangan dari keempat algoritma yang dipakai[23]:

Tabel 3. 2 Tabel Perbandingan *Naïve Bayes*, *SVM*, *Random Forest*, dan *Logistic Regression*

Nama	Kelebihan	Kekurangan
------	-----------	------------

Nama	Kelebihan	Kekurangan
<i>Naïve Bayes</i>	Kecepatan dan efisiensinya yang tinggi. Algoritma ini dapat beroperasi dengan cepat dan membutuhkan sedikit sumber daya komputasi, membuatnya cocok untuk memproses data teks besar dengan efisien[16].	<i>Naive Bayes</i> membuat asumsi bahwa fitur-fitur yang digunakan untuk analisis sentimen adalah independen satu sama lain, meskipun dalam konteks nyata, fitur-fitur ini mungkin saling terkait[17].
<i>Support Vector Machine</i>	<ul style="list-style-type: none"> - <i>SVM</i> efektif dalam menangani data dengan ruang fitur yang tinggi, terutama dalam analisis teks yang melibatkan vektor kata-kata atau representasi vektor - Dapat bekerja dengan baik dalam dimensi fitur yang besar dan kompleks, membuatnya cocok untuk analisis sentimen pada dataset teks yang kaya fitur[18] 	- Ketika menghadapi data teks yang tidak terstruktur, seperti kalimat yang panjang dan kompleks, bisa memerlukan langkah-langkah preprocessing tambahan dan mungkin tidak seefektif ketika bekerja dengan data terstruktur lainnya[18].
<i>Random Forest</i>	<i>Random Forest</i> memiliki kemampuan untuk menangani dataset dengan beragam fitur dan interaksi yang kompleks antara fitur-fitur tersebut, sehingga cocok untuk analisis sentimen di mana variasi dalam pemilihan kata atau frasa dapat dianggap sebagai fitur yang penting, dan interaksi antara fitur-fitur tersebut dapat kompleks[19].	<i>Random Forest</i> dapat cenderung <i>overfitting</i> , terutama pada dataset yang sangat terperinci atau mengandung noise. Model algoritma cenderung terlalu sesuai dengan data pelatihan dan gagal menggeneralisasi dengan baik pada data baru[19].
<i>Logistic Regression</i>	<i>Logistic Regression</i> Memiliki interpretabilitas yang tinggi, yang berarti mudah dipahami dan dijelaskan. Koefisien yang dihasilkan oleh model dapat memberikan informasi tentang sejauh mana setiap fitur berkontribusi terhadap prediksi sentimen positif atau negatif. Ini dapat membantu dalam menemukan korelasi langsung antara fitur-fitur tertentu dan sentimen yang diungkapkan	<i>Logistic Regression</i> bekerja dengan baik saat hubungan antara variabel independen dan dependen bersifat linier. Namun, jika hubungan tersebut bersifat nonlinear, <i>Logistic Regression</i> mungkin tidak dapat menangkap dengan baik. Ini dapat menjadi kelemahan ketika menghadapi data teks yang mungkin memiliki hubungan yang kompleks antara kata-kata atau

Nama	Kelebihan	Kekurangan
	dalam <i>review</i> [20].	frasa[20].

Berdasarkan tabel diatas, maka dapat dilihat bahwa seluruh algoritma klasifikasi yang didaftarkan dapat dikatakan cocok untuk digunakan dalam penelitian ini, mengingat bahwa penelitian ini akan membandingkan berbagai variabel yang bergantung antara satu sama lain. Meskipun demikian, tidaklah juga menutup kemungkinan apabila ada beberapa algoritma diatas yang diuntungkan dalam konteks penelitian ini dikarenakan tujuan dari penelitian ini sangat bersesuaian dengan kekuatan dari algoritma tersebut.

