

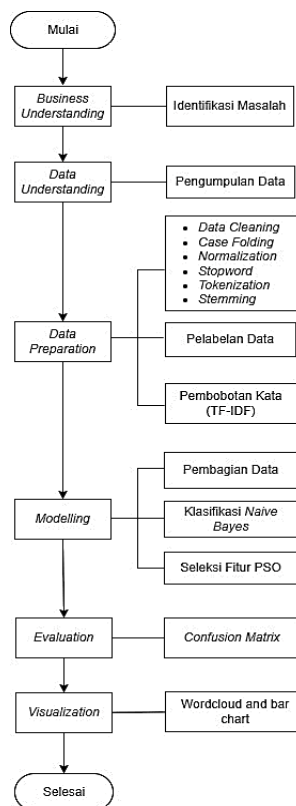
BAB III METODE PENELITIAN

3.1 Objek Penelitian

Objek penelitian adalah sebuah atribut atau target ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu mengenai sesuatu hal yang objektif, valid, dan reliabel terkait dengan suatu variabel tertentu (Sugiyono, 2018). Objek pada penelitian ini adalah *dataset* komentar pengguna aplikasi Jenius terhadap pelayanan yang telah diberikan di *website Google Play Store*. Variabel yang digunakan dalam penelitian ini adalah *review* dan *label*.

3.2 Desain Penelitian

Dalam penelitian ini, penulis membuat desain penelitian yang terdiri dari beberapa tahapan yang dilakukan untuk memenuhi tujuan penelitian. Desain penelitian dirancang menggunakan metode *Cross Industry Standard Process for Data Mining* (CRISP-DM) dan dikembangkan oleh penulis. Adapun tahapan-tahapan penelitian dapat dilihat pada Gambar 3.1 berikut :



Gambar 3.1 Desain Penelitian

Sumber : Daniel T.Larose (2015)

3.2.1 Business Understanding

Pada fase ini dilakukan identifikasi permasalahan berdasarkan topik yang telah ditentukan sebelumnya yaitu mengenai klasifikasi ulasan pengguna aplikasi Jenius menggunakan algoritma *Naive Bayes*. Proses ini bertujuan untuk menentukan tujuan dan batasan pada permasalahan yang menjadi ruang lingkup pada penelitian ini.

3.2.2 Data Understanding

Fase ini bertujuan untuk mengumpulkan, mengidentifikasi, memahami data, serta memverifikasi kebenaran dan kelengkapan data yang dimiliki. Pada penelitian ini, pengumpulan data didapatkan dari ulasan pengguna aplikasi Jenius di situs *Google Play Store* menggunakan metode *scrapping data*. Data mentah yang didapatkan dari hasil *scrapping data* disimpan kedalam *file* dengan format Excel untuk memudahkan proses analisis data.

3.2.3 Data Preparation

Pada fase ini dilakukan proses penyiapan data yang bertujuan untuk memperoleh data yang bersih dan siap digunakan dalam penelitian. Tahap ini sering digunakan untuk mengubah data yang awalnya tidak terstruktur menjadi data yang terstruktur. Berikut merupakan tahapan yang dilakukan dalam *pre-processing* data:

- a. *Data Cleaning*, yaitu proses membersihkan dan menghapus karakter yang tidak dibutuhkan, seperti angka, tanda baca, simbol, dan sebagainya.
- b. *Case Folding*, yaitu proses mengubah seluruh teks ulasan menjadi huruf kecil (*lowercase*) dengan untuk menyeragamkan seluruh karakter.
- c. *Normalization*, yaitu proses mengkonversi kata yang tidak baku, disingkat, atau salah eja menjadi kata baku yang sesuai dengan KBBI.
- d. *Stopwords Removal*, yaitu proses untuk menghapus kata yang tidak bermakna penting untuk klasifikasi menggunakan *stopwords* sastrawi.
- e. *Tokenization*, yaitu proses memotong kalimat pada teks ulasan menjadi bentuk satuan kata dengan spasi (*whitespace*) sebagai pemisahannya.

- f. *Stemming*, yaitu proses mengkonversi kata yang memiliki imbuhan menjadi kata dasar. Pada penelitian ini akan digunakan algoritma *stemming* Nazief dan Adriani untuk mengatasi risiko dari proses *stemming* yang memungkinkan hilangnya informasi data yang di *stem*.

Setelah dilakukan *pre-preprocessing* data, dilanjutkan dengan tahap pelabelan terhadap data ulasan pengguna aplikasi Jenius. Proses pelabelan data dilakukan secara manual. Setelah dilakukan pelabelan data, langkah selanjutnya adalah tahap pembobotan kata. Pada tahap ini, setiap kata atau *term* yang ada pada dokumen akan diberi bobot dan dilakukan perhitungan dengan menggunakan metode pembobotan *Term Frequency - Inverse Document Frequency* (TF-IDF). Tahapan ini dilakukan dengan tujuan untuk mentransformasikan data ulasan menjadi data berupa angka. Perhitungan dapat dilakukan dengan menggunakan rumus persamaan yang ada pada bab 2 persamaan (2.7), (2.8), dan (2.9).

3.2.4 Modelling

Pada fase ini dilakukan proses pembentukan model klasifikasi data dengan menggunakan algoritma *Naive Bayes* dan seleksi fitur *Particle Swarm Optimization* (PSO) Namun, sebelum dilakukan klasifikasi akan dilakukan pembagian data latih dan data uji terlebih dahulu dengan presentasi 80% data latih dan 20% data uji. Data latih adalah data yang digunakan selama tahap pelatihan untuk melakukan pengenalan pola data dan pembelajaran data. Sementara itu, data uji adalah data yang digunakan selama fase pengujian untuk melakukan proses pengujian berdasarkan data pengenalan dan pembelajaran yang diperoleh selama proses pelatihan (Yunhasnawa & Mardhika, 2019). Berdasarkan penelitian (Yunhasnawa & Mardhika, 2019), sebuah sistem akan lebih baik dalam mengidentifikasi pola data yang baru jika jumlah pembagian data latih lebih besar daripada jumlah pembagian data uji.

Untuk mengoptimasi nilai akurasi dari algoritma klasifikasi *Naive Bayes* dilakukan penambahan seleksi fitur algoritma *Particle Swarm Optimization*. Proses ini untuk mengelompokkan kata atau fungsi yang dapat digunakan sebagai perwakilan yang signifikan dari dokumen yang ada. PSO akan dimulai dengan koleksi yang berisi sejumlah *term* solusi, kemudian pembangkitan *term* dilakukan secara acak dan dengan kecepatan acak, diikuti dengan pencarian solusi perbaikan

term terbaik untuk kategori yang dipilih. Setelah menentukan local best dan global best dengan cara membandingkan nilai *local best* pertama dengan nilai *fitness* pada iterasi *local best* berikutnya, maka *global best* ditentukan dengan cara membandingkan nilai *local best* dengan nilai *fitness* dari *local best*. Selanjutnya, apabila nilai *fitness* terbaik sudah tercapai atau iterasi sudah maksimal, maka algoritma akan berakhir.

3.2.5 Evaluation

Pada fase ini dilakukan evaluasi dengan menggunakan *confusion matrix* untuk mengukur performa model algoritma *Naive Bayes* dan seleksi fitur PSO yang telah dibentuk. Performa klasifikasi akan diukur dengan menghitung nilai *accuracy*, *precision*, *recall*, *f-1 score* dan akan dilakukan perbandingan hasil evaluasi klasifikasi algoritma *Naive Bayes* tanpa PSO dengan klasifikasi algoritma *Naive Bayes* menggunakan seleksi fitur PSO.

3.2.6 Visualization

Setelah dilakukan evaluasi terhadap model, fase selanjutnya yaitu tahap visualisasi data dalam bentuk *word cloud* dan grafik *bar chart* untuk memudahkan pembaca dalam memahami hasil penelitian. Visualisasi menggunakan *word cloud* bertujuan untuk mengetahui frekuensi kemunculan kata yang paling sering muncul dari setiap kelas positif dan kelas negatif.

3.3 Metode Penelitian

3.3.1 Jenis Penelitian dan Metode yang digunakan

Jenis penelitian ini adalah penelitian kuantitatif dengan pendekatan deskriptif. Penelitian ini menggunakan pendekatan kuantitatif dikarenakan dalam penelitian melibatkan penggunaan angka-angka dalam seluruh tahapan penelitian, mulai dari pengumpulan data, penafsiran data, hingga visualisasi data yang dihasilkan (Sugiyono, 2018). Kemudian dideskripsikan dengan deduksi yang berangkat dari teori-teori yang bersifat umum, diikuti dengan observasi untuk menilai validitas keabsahan teori tersebut. Hasil penelitian selanjutnya akan dinyatakan secara deskriptif karena akan digunakan untuk mendeskripsikan data yang diperoleh dan menjawab rumusan masalah.

Metode penelitian ini adalah *Cross Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM merupakan standar proses *data mining* yang umum digunakan sebagai strategi pemecahan masalah dalam bisnis dan penelitian. Terdapat 6 tahapan dalam metode CRISP-DM yaitu *business understanding, data understanding, data preparation, modelling, evaluation, dan deployment* (Daniel T.Larose, 2015).

3.3.2 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder adalah data yang akan digunakan dalam proses pengolahan data untuk mengetahui *feedback* atau persepsi pengguna terhadap layanan yang diberikan oleh Aplikasi Jenius pada situs *Google Play Store*. Data diambil menggunakan teknik *web scapping* dengan jumlah sampel 3047 *records* dan kemudian didapatkan rentang waktu dari tanggal 16 Januari 2022 hingga 13 April 2023.

3.3.3 Teknik Pengumpulan Data

3.3.3.1 Studi Literatur

Studi literatur dilakukan dengan mengumpulkan dan menggali informasi-informasi akurat yang berkaitan dengan topik penelitian serta mengkaji kemajuan metodologi yang berhubungan dengan penelitian yang akan dilakukan. Beberapa referensi dan literatur yang digunakan dalam penelitian ini berasal dari *e-book*, jurnal, skripsi, dan artikel terkait yang berkaitan dengan penerapan analisis sentimen, klasifikasi teks, *data preprocessing, scrapping data*, algoritma *Naive Bayes*, seleksi fitur *Particle Swarm Optimization*, dan CRISP-DM.

3.3.3.2 Web Scrapping

Dalam penelitian ini, data yang digunakan diperoleh menggunakan teknik *web scrapping*. *Web Scraping* merupakan metode pengambilan data dari sebuah *website* dan digunakan untuk analisis sentimen (Romadoni dkk., 2020). Proses *web scrapping* dilakukan dengan mengambil ulasan yang diberikan oleh pengguna Jenius melalui situs *Google Play Store* dengan bantuan salah satu *library python* yaitu *google play scapper*.

3.3.4 Teknik Analisis Data

Dalam penelitian ini, digunakan *tools Microsoft Excel* dan *Google Colaboratory* untuk proses analisis dan pengujian data. Terdapat beberapa teknik analisis data yang digunakan dalam penelitian ini, antara lain:

1. Analisis Deskriptif, teknik ini digunakan untuk memberikan gambaran mengenai ulasan aplikasi Jenius yang ada pada situs *Google Play Store*.
2. Analisis Sentimen. Teknik ini digunakan untuk mengklasifikasi data ulasan menjadi dua kelas yaitu kelas positif dan negatif.
3. *Text Mining*. Teknik ini digunakan untuk menemukan pola yang menarik pada sekumpulan data tekstual dengan jumlah yang besar.
4. Metode *Machine Learning*. Pada penelitian ini data akan diklasifikasikan ke dalam dua kelas dan membentuk sebuah model klasifikasi dengan menggunakan algoritma *Naive Bayes* berbasis *Particle Swarm Optimization* untuk optimasi kinerja algoritma klasifikasi.
5. *Wordcloud*, digunakan untuk melakukan visualisasi frekuensi kata yang paling sering muncul pada ulasan yang diberikan pengguna aplikasi Jenius dan dilakukan identifikasi terhadap faktor penyebab permasalahan yang didapatkan pada ulasan negatif dan faktor yang membuat aplikasi telah baik yang didapatkan dari ulasan positif sehingga dapat dilakukan pemecahan terhadap masalah yang dihadapi.

3.4 Alat dan Bahan Penelitian

Dalam melakukan penelitian ini, penulis menggunakan perangkat penelitian yang meliputi perangkat keras dan perangkat lunak untuk mendukung kegiatan penelitian, yaitu sebagai berikut:

3.4.1 Perangkat Keras (*Hardware*)

Peneliti menggunakan perangkat keras berupa sebuah laptop dengan spesifikasi sebagai berikut:

1. *Processor* : Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz (8 CPUs), ~1.8 GHz
2. *Memory* : 16 GB RAM
3. *Graphic* : NVIDIA GeForce MX130
4. *Storage* : *Harddisk* 1TB, SSD 128 GB

3.4.2 Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan penulis dalam penelitian ini adalah sebagai berikut:

1. Sistem Operasi : *Windows 10 Home Single Language* 64-bit (10.0, *Build* 19044)
2. Bahasa Pemrograman : *Python*
3. Aplikasi yang digunakan :
 - a. *Google Colaboratory*
 - b. *Microsoft Excel 365*