

## Metode ( ringkas/data/algoritma/prosedur

### Metode penelitian

Penelitian ini menggunakan data hasil dari scraping ulasan pengguna aplikasi game online Clash of Clans pada Google Play Store dengan bantuan Google Colab menggunakan bahasa pemrograman python. Scraping adalah teknik untuk mendapatkan informasi secara otomatis tanpa harus menyalinnya secara manual[13]. Ulasan yang digunakan terbatas hanya ulasan yang berasal dari negara Indonesia dan berjumlah 1000 ulasan. Berikut adalah diagram alur dari metode penelitian seperti pada dataset yang sudah bersih kemudian akan dibagi menjadi dua bagian yaitu data latih dan data uji. Metode yang digunakan dalam penelitian ini adalah Support Vector Machine dan Naive Bayes sebagai perbandingan untuk mendapatkan hasil akurasi terbaik dalam pengklasifikasian data teks sentimen ulasan.

### Data

Pengumpulan Data Dataset yang peneliti gunakan pada penelitian ini merupakan dataset ulasan dari pengguna aplikasi Clash of Clans dari hasil scraping pada Google Play Store. Data ini memiliki 10 atribut dan 1000 data ulasan. Langkah pertama yang harus dilakukan adalah menginstall google play scraper, kemudian install library pandas dan numpy. Arahkan google play scraper ke "com.supercell.clashofclans" untuk melakukan scraping data ulasan pada aplikasi Clash of Clans. Data yang diambil berjumlah 1000 data ulasan newest atau terbaru dengan bahasa dan negara Indonesia pada 5 Januari 2023 sampai 8 Januari 2023.

### Algoritma

Algoritma Naïve Bayes merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik. Metode ini pertama kali dikenalkan oleh ilmuwan Inggris Thomas Bayes, yaitu digunakan untuk memprediksi peluang yang terjadi di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes[20]. Model naïve bayes yang digunakan pada penelitian ini adalah gaussian naïve bayes yang merupakan sebuah teknik klasifikasi yang digunakan dalam machine learning dengan menggunakan metode probability dan distribusi gaussian. Model ini mengasumsikan bahwa setiap feature pada data memiliki pengaruh yang independen dalam memprediksi target. Kombinasi prediksi dari seluruh parameter adalah prediksi akhir dengan probability dari target variable yang diklasifikasikan ke dalam dua kelas. Klasifikasi akhirnya adalah hasil probability yang lebih tinggi dari grup target  $P(X|Y)=P(Y|X).P(X)P(Y)$  (4) Pada rumus 4 digunakan untuk mendapatkan nilai token dari setiap kata menggunakan model naïve bayes. Dimana X,Y adalah kejadian,  $P(X|Y)$  merupakan probability untuk X ketika Y benar.  $P(Y|X)$  adalah probability untuk Y ketika X benar.  $P(X)$ ,  $P(Y)$  adalah probability independent untuk X dan Y.

## Prosedur penelitian

**Pengumpulan Data** Dataset yang peneliti gunakan pada penelitian ini merupakan dataset ulasan dari pengguna aplikasi Clash of Clans dari hasil scraping pada Google Play Store. Data ini memiliki 10 atribut dan 1000 data ulasan. Langkah pertama yang harus dilakukan adalah menginstall google play scraper, kemudian install library pandas dan numpy. Arahkan google playscraper ke “com.supercell.clashofclans” untuk melakukan scrapping data ulasan pada aplikasi Clash of Clans.

**Preprocessing Data** yang didapat dari hasil scraping ulasan pengguna aplikasi game online Clash of Clans inilah yang akan dilakukan tahapan preprocessing. Tahap Preprocessing yang terdapat dalam penelitian ini adalah case folding, filtering (stopword removal), stemming dan word normalization. Semua fungsi yang telah dibuat pada tahapan preprocessing kemudian dijalankan pada satu fungsi yaitu Text Processing Pipeline.

**Ekstraksi dan Seleksi Fitur** Setelah proses preprocessing selesai, dilakukan ekstraksi fitur dan seleksi fitur. Ekstraksi fitur menjadi bagian yang sangat penting dalam pengolahan dokumen pada mesin pencari karena sangat menentukan keberhasilan proses text mining. Jika nilai fitur yang dihasilkan tidak tepat, maka informasi yang digali dalam text mining tidak bisa memenuhi kriteria yang diinginkan. Akibatnya, informasi yang ditampilkan oleh mesin pencari tidak akan memenuhi keinginan pengguna[16]. Metode ekstraksi fitur yang digunakan untuk mendapatkan fitur terpilih pada dataset adalah TF-IDF dan N-Gram. Sedangkan Seleksi Fitur adalah proses dimana seleksi dilakukan secara otomatis atau manual memilih fitur-fitur yang berkontribusi paling besar pada variabel prediksi atau output. Memilih fitur yang tidak relevan dalam dataset dapat mengurangi keakuratan model dan membuat model belajar berdasarkan fitur yang tidak relevan. Manfaat melakukan seleksi fitur sebelum memodelkan data yaitu mengurangi overfitting, meningkatkan akurasi dan mengurangi waktu pelatihan[17]. Chi-Square dan SelectKBest adalah metode seleksi fitur yang akan diterapkan pada penelitian ini.

## Pembagian Dataset

Dataset dibagi menjadi **data latih (training set)** dan **data uji (testing set)**, misalnya dengan perbandingan 80:20 atau menggunakan **k-fold cross validation**.

## Implementasi Algoritma

**Naïve Bayes:** menghitung probabilitas setiap kelas berdasarkan distribusi kata.

**SVM:** mencari hyperplane terbaik yang memisahkan kelas positif, negatif, dan netral.

## **Pengujian dan Evaluasi**

Melakukan klasifikasi pada data uji dengan kedua algoritma.

Mengukur performa menggunakan metrik evaluasi seperti:

**Accuracy** (tingkat ketepatan)

**Precision** (ketepatan prediksi positif)

**Recall** (cakupan prediksi positif)

**F1-score** (rata-rata harmonis precision dan recall)

Membandingkan hasil antara SVM dan Naïve Bayes.

## **Analisis Hasil**

Menjelaskan kelebihan dan kelemahan tiap algoritma berdasarkan hasil pengujian.

Memberikan interpretasi apakah SVM lebih unggul karena mampu menangani data dengan dimensi tinggi, atau Naïve Bayes lebih cocok karena sederhana dan cepat.

## **Kesimpulan dan Rekomendasi**

Menyimpulkan algoritma yang paling efektif untuk analisis sentimen ulasan game online.

Memberikan saran untuk penelitian selanjutnya, misalnya mencoba algoritma lain (Random Forest, LSTM) atau menambah jumlah data.