

3. membuat ringkasan ilmiah 1 halaman perartikel (5 ringkasan)

Identitas artikel

Judul Perbandingan performa SVM dan naive bayes pada analisis sentimen aplikasi game online.

Penulis Galang paksi permana,danang aditya nugraha,heri santoso

Tahun 2024

Sumber publishing-widyagama.ac.id/ejournal-v2/indeks.php/jointecs/article/view/4854/3030

Latar belakang

Industri game online saat ini berkembang sangat pesat, terutama game berbasis mobile yang mudah diakses oleh berbagai kalangan masyarakat. Game online tidak hanya menjadi sarana hiburan, tetapi juga telah membentuk komunitas besar dengan jutaan pengguna aktif di seluruh dunia. Popularitas yang tinggi membuat persaingan antar game semakin ketat, sehingga kepuasan dan pengalaman pengguna menjadi faktor penting bagi keberhasilan sebuah game.

Salah satu cara untuk mengetahui tingkat kepuasan pengguna adalah dengan menganalisis **ulasan (review) dan rating** yang diberikan pada platform distribusi aplikasi seperti Google Play Store, App Store, maupun forum komunitas game. Ulasan ini berisi pendapat, kritik, dan saran yang dapat menjadi **sumber informasi berharga bagi pengembang** untuk meningkatkan kualitas game. Namun, jumlah ulasan yang sangat besar membuat analisis manual menjadi sulit dilakukan. Oleh karena itu, diperlukan suatu metode analisis otomatis yang mampu mengklasifikasikan sentimen pengguna secara cepat dan akurat.

Analisis sentimen merupakan teknik pengolahan bahasa alami (Natural Language Processing/NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini pengguna ke dalam kategori tertentu, seperti **positif, negatif, atau netral**. Dalam penelitian analisis sentimen, terdapat berbagai algoritma klasifikasi yang dapat digunakan. Dua algoritma yang paling banyak digunakan karena efektivitasnya dalam pengolahan teks adalah **Naïve Bayes (NB)** dan **Support Vector Machine (SVM)**.

Naïve Bayes adalah algoritma berbasis probabilistik yang sederhana, cepat, dan efisien dalam mengolah data dalam jumlah besar. Algoritma ini bekerja dengan prinsip teorema Bayes dan mengasumsikan bahwa setiap fitur bersifat independen. Kelebihan Naïve Bayes adalah kecepatan komputasi dan kemudahan implementasi, sehingga banyak digunakan sebagai baseline dalam analisis teks. Namun, kelemahannya terletak pada asumsi independensi yang sering tidak sesuai dengan kenyataan, terutama pada data ulasan game yang kaya akan korelasi kata.

Support Vector Machine (SVM) adalah algoritma yang bekerja dengan mencari hyperplane terbaik untuk memisahkan data ke dalam kelas yang berbeda. Algoritma ini terkenal kuat dalam menangani data berdimensi tinggi, seperti representasi teks menggunakan bag-of-words atau TF-IDF. SVM umumnya menghasilkan akurasi yang lebih tinggi dibandingkan Naïve Bayes, meskipun waktu komputasi training bisa lebih lama.

Mengacu pada karakteristik tersebut, analisis sentimen terhadap **ulasan game online** menjadi menarik untuk dilakukan dengan membandingkan performa kedua algoritma. Hal ini penting karena ulasan pengguna game sering mengandung bahasa informal, singkatan, emotikon, hingga istilah khas dunia gaming yang dapat memengaruhi hasil klasifikasi. Dengan adanya perbandingan ini, dapat diketahui algoritma mana yang lebih sesuai untuk memberikan hasil analisis sentimen yang **cepat, akurat, dan relevan**, sehingga bermanfaat bagi pengembang game maupun peneliti di bidang data mining dan NLP.

Tujuan

Penelitian mengenai perbandingan performa **Support Vector Machine (SVM)** dan **Naive Bayes** dalam analisis sentimen pada aplikasi game online memiliki beberapa tujuan penting, baik secara teoritis maupun praktis. Secara umum, tujuan penelitian ini adalah untuk mengetahui sejauh mana kedua algoritma klasifikasi teks tersebut mampu memberikan hasil yang optimal dalam mengelompokkan sentimen pengguna terhadap aplikasi game online, khususnya dari data ulasan yang tersedia pada platform distribusi aplikasi seperti Google Play Store maupun App Store.

Secara lebih rinci, tujuan penelitian ini adalah sebagai berikut:

1. **Mengevaluasi efektivitas algoritma Naive Bayes dalam analisis sentimen ulasan aplikasigameonline.**

Penelitian ini bertujuan untuk mengukur seberapa baik Naive Bayes dalam melakukan klasifikasi sentimen ke dalam kategori positif, negatif, atau netral. Evaluasi dilakukan melalui pengujian akurasi, presisi, recall, dan F1-score agar diperoleh gambaran menyeluruh tentang kelebihan dan keterbatasan metode probabilistik ini pada data teks yang beragam.

2. **Mengevaluasi efektivitas algoritma Support Vector Machine (SVM) dalam analisis sentimenulasanaplikasigameonline.**

Penelitian ini juga bertujuan untuk menilai kinerja SVM dalam mengolah data ulasan dengan dimensi tinggi (high dimensional text data) yang dihasilkan oleh representasi kata, misalnya TF-IDF. Melalui proses pengujian, penelitian ingin mengetahui apakah SVM dapat memberikan hasil klasifikasi yang lebih stabil, akurat, dan konsisten dibandingkan Naive Bayes.

3. **Membandingkan performa SVM dan Naive Bayes pada domain analisis sentimen gameonline.**

Tujuan utama penelitian adalah membandingkan kedua algoritma berdasarkan

berbagai metrik evaluasi, sehingga dapat diketahui algoritma mana yang lebih unggul untuk kasus analisis sentimen ulasan pengguna game online. Perbandingan ini penting karena setiap algoritma memiliki kelebihan dan kekurangan masing-masing; Naive Bayes unggul dalam kecepatan, sedangkan SVM unggul dalam akurasi.

4. **Memberikan rekomendasi algoritma terbaik untuk analisis sentimen pada aplikasi gameonline.**

Hasil penelitian diharapkan dapat memberikan panduan praktis bagi pengembang aplikasi, peneliti, maupun pihak industri game dalam memilih metode klasifikasi yang paling sesuai dengan kebutuhan. Jika akurasi menjadi prioritas utama, maka SVM mungkin lebih disarankan. Namun, jika keterbatasan sumber daya komputasi menjadi kendala, maka Naive Bayes dapat menjadi alternatif yang lebih efisien.

5. **Memberikan kontribusi ilmiah dalam bidang text mining dan machine learning.**

Tujuan lain yang tidak kalah penting adalah memberikan kontribusi akademik berupa kajian perbandingan dua algoritma populer dalam machine learning pada domain khusus, yaitu ulasan aplikasi game online. Hal ini diharapkan dapat memperkaya literatur penelitian di bidang analisis sentimen, serta menjadi dasar untuk pengembangan penelitian selanjutnya, misalnya dengan metode ensemble atau deep learning.

Pendahuluan

Perkembangan teknologi informasi telah membawa dampak besar terhadap berbagai bidang kehidupan, salah satunya adalah industri hiburan digital seperti **game online**. Game online tidak hanya berfungsi sebagai sarana hiburan, tetapi juga telah membentuk komunitas global dengan jutaan pemain aktif yang saling berinteraksi setiap harinya. Pertumbuhan pesat industri ini mendorong pengembang untuk terus meningkatkan kualitas produk agar mampu bersaing dengan game lain di pasar yang semakin kompetitif.

Salah satu sumber informasi penting yang dapat digunakan pengembang untuk mengevaluasi kualitas game adalah **ulasan (review) dan rating pengguna** yang tersedia pada platform distribusi aplikasi, seperti Google Play Store atau App Store. Melalui ulasan, pengguna menyampaikan pendapat, kritik, maupun saran mengenai pengalaman mereka saat bermain game. Ulasan tersebut umumnya bersifat subjektif dan dapat berupa sentimen positif, negatif, maupun netral. Analisis terhadap sentimen ini menjadi sangat penting karena dapat membantu pengembang memahami tingkat kepuasan pengguna, mengidentifikasi kelemahan aplikasi, serta menentukan strategi pengembangan yang tepat.

Namun, jumlah ulasan yang sangat besar membuat analisis manual menjadi tidak efektif. Oleh karena itu, dibutuhkan metode analisis otomatis yang dapat mengolah data teks dalam jumlah besar dengan cepat dan akurat. Salah satu pendekatan yang banyak digunakan dalam bidang ini adalah **analisis sentimen**. Analisis sentimen merupakan cabang dari *Natural Language*

Processing (NLP) yang bertujuan untuk mengklasifikasikan opini atau pernyataan pengguna ke dalam kategori tertentu, seperti positif, negatif, atau netral.

Dalam penelitian analisis sentimen, terdapat berbagai algoritma klasifikasi yang dapat digunakan. Dua algoritma yang paling populer adalah **Naïve Bayes (NB)** dan **Support Vector Machine (SVM)**.

- **Naïve Bayes** merupakan algoritma berbasis probabilistik yang sederhana, cepat, dan efisien dalam menangani data teks berukuran besar. Algoritma ini menggunakan prinsip teorema Bayes dengan asumsi independensi antar fitur. Meski sederhana, Naïve Bayes sering digunakan sebagai baseline karena kecepatan komputasinya. Namun, kelemahannya terletak pada asumsi independensi yang kurang sesuai dengan kondisi nyata, terutama dalam ulasan game yang kaya akan korelasi antar kata.
- **Support Vector Machine (SVM)** adalah algoritma klasifikasi yang bekerja dengan mencari hyperplane terbaik untuk memisahkan data ke dalam kelas tertentu. SVM dikenal memiliki performa yang baik dalam mengolah data berdimensi tinggi seperti teks, terutama dengan representasi *bag-of-words* atau *TF-IDF*. Walaupun waktu pelatihan (training) lebih lama dibandingkan Naïve Bayes, SVM umumnya memberikan hasil akurasi yang lebih tinggi.

Berdasarkan karakteristik kedua algoritma tersebut, penelitian ini menjadi penting untuk dilakukan dengan tujuan **membandingkan performa SVM dan Naïve Bayes pada analisis sentimen ulasan game online**. Perbandingan ini diharapkan dapat memberikan gambaran yang jelas mengenai keunggulan dan kelemahan masing-masing algoritma, baik dari segi akurasi, kecepatan komputasi, maupun ketahanan terhadap data yang kompleks.

Dengan adanya hasil perbandingan ini, penelitian dapat memberikan kontribusi dalam memilih algoritma yang paling sesuai untuk analisis sentimen ulasan game online. Selain itu, hasil penelitian juga bermanfaat bagi pengembang game sebagai bahan evaluasi untuk meningkatkan kualitas produk berdasarkan persepsi pengguna.

Metode penelitian

Penelitian ini menggunakan data hasil dari scraping ulasan pengguna aplikasi game online Clash of Clans pada Google Play Store dengan bantuan nGoogle Colab menggunakan bahasa pemrograman python. Scraping adalah teknik untuk mendapatkan informasi secara otomatis tanpa harus menyalinnya secara manual. Ulasan yang digunakan terbatasnya 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.

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. Data yang diambil berjumlah 1000 data ulasan newest atau terbaru dengan bahasa dan negara Indonesia pada 5 Januari 2023 sampai 8 Januari 2023.

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. Dilakukan optimasi pada ekstraksi fitur dan seleksi fitur dengan cara melakukan 6 kali pengujian kombinasi fitur dengan tujuan untuk mendapatkan model klasifikasi yang paling optimal dan menghasilkan hasil akurasi yang lebih tinggi dan akurat.

Klasifikasi Support Vector Machine Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari jumlah kelas yang tersedia. Metode SVM adalah metode terbaik yang digunakan dalam melakukan prediksi berdasarkan kerangka kerja statistik yang dikembangkan oleh seorang peneliti yang bernama Chervonenkis dan Vapnik. SVM di dalam machine learning merupakan salah satu algoritme pembelajaran yang terawasi sehingga dapat digunakan dalam melakukan analisis untuk kebutuhan klasifikasi dan regresi. Konsep SVM bermula dari masalah klasifikasi dua kelas sehingga membutuhkan training set positif dan negatif. SVM berusaha menemukan hyperplane (pemisah) terbaik untuk memisahkan ke dalam dua kelas dan memaksimalkan margin antara dua kelas tersebut. Klasifikasi SVM dinotasikan sebagai

rumus 1. $f(x) = W^T X + b$ (1) Dalam klasifikasi SVM, fungsi klasifikasi dilambangkan dengan $f(x)$. Fungsi $f(x)$ memprediksi kelas target berdasarkan fitur input x . W adalah vektor bobot, X adalah vektor fitur masukan, dan b adalah bias. Sehingga diperoleh rumus 2 dan rumus 3. $[(W^T \cdot x_i) + b] \geq 1$ untuk $y_i = +1$ (2) $[(W^T \cdot x_i) + b] \leq -1$ untuk $y_i = -1$ (3) Klasifikasi dilakukan dengan dataset latih yang dinotasikan dengan $f(x)$ = himpunan data training, ke $i = 1, 2, \dots, n$ dan, y_i = label kelas dari x_i . Pengelompokkan menggunakan hyperplane linier seperti terlihat pada rumus. Parameter w menotasikan vector terhadap hyperplane, b adalah offset.

2.5. Klasifikasi Naïve Bayes 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. 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.

Evaluasi Tahapan evaluasi dilakukan di akhir proses penelitian. Pada dataset ulasan pengguna aplikasi game online Clash of Clans di Google Play Store yang telah melalui tahap ekstraksi fitur dan seleksi fitur. Uji akurasi dilakukan untuk menentukan model klasifikasi yang paling optimal, dengan menggunakan 6 kombinasi fitur dan tiga presentase rasio pengujian, yaitu 70:30, 80:20, dan 90:10 untuk data training dan data testing, dengan total hasil pengujian akurasi sebanyak 36 hasil akurasi. Tahapan evaluasi pada penelitian ini dilakukan dengan metode Confusion Matrix, Classification Report dan mesin uji. Confusion Matrix dan Classification Report merupakan metode yang digunakan dalam melakukan perhitungan akurasi dan evaluasi model klasifikasi. Sementara mesin uji berfungsi untuk melakukan pengujian model yang berjalan pada algoritma dengan menghasilkan label positif dan negatif, yaitu 0 dan 1.