

# Analisis Kinerja Algoritma Winnowing dan Machine Learning dalam Sistem Penilaian Esai Otomatis Berbasis Teks

Muhammad Fahmi Luthfi<sup>1</sup>, Kusrini<sup>2</sup>

Magister Teknik Informatika, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

Email: \*[m.f.luthfi21@students.amikom.ac.id](mailto:m.f.luthfi21@students.amikom.ac.id), [kusrini@amikom.ac.id](mailto:kusrini@amikom.ac.id)

(Naskah masuk: 5 Agustus 2025, diterima untuk diterbitkan: 20 Januari 2026)

**Abstrak:** Penelitian ini bertujuan untuk mengevaluasi penerapan teknik ekstraksi fitur Winnowing yang dipadukan dengan model pembelajaran mesin K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM) dalam sistem penilaian esai otomatis berbasis teks. Topik ini penting karena penilaian otomatis dapat mempercepat proses evaluasi esai dan mengurangi potensi bias manusia. Dalam penelitian ini, esai diklasifikasikan menggunakan teknik K-gram dan fingerprinting Winnowing untuk meningkatkan akurasi dan representasi fitur dari teks. Hasil evaluasi menunjukkan bahwa model Winnowing + SVM menghasilkan performa yang paling unggul dengan akurasi sebesar 0.876 dan F1-score sebesar 0.868430. Meskipun membutuhkan waktu komputasi yang lebih tinggi (5.181894 detik), SVM secara konsisten mengklasifikasikan esai dengan presisi dan stabilitas yang lebih baik. Sebaliknya, model Winnowing + KNN menunjukkan performa yang lebih rendah dengan akurasi sebesar 0.707 dan F1-score sebesar 0.686699, namun memiliki keunggulan dalam waktu komputasi yang jauh lebih cepat (0.077723 detik). Penggunaan Winnowing terbukti efektif dalam meningkatkan kinerja kedua model, dengan dampak yang lebih signifikan terhadap SVM. Kesimpulannya, SVM lebih direkomendasikan untuk sistem penilaian esai otomatis yang mengutamakan akurasi, sementara KNN lebih cocok untuk aplikasi yang memprioritaskan efisiensi waktu.

**Kata Kunci:** Winnowing; K-Nearest Neighbors; Support Vector Machine; Ekstraksi Fitur; Penilaian Esai Otomatis.

## Performance Analysis of Winnowing Algorithm and Machine Learning in Teks-based Automated Essay Grading System

**Abstract:** This research aims to evaluate the application of Winnowing feature extraction technique combined with K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) machine learning models in a text-based automated essay grading system. This topic is important because automated scoring can speed up the essay evaluation process and reduce the potential for human bias. In this study, essays were classified using K-grams and Winnowing fingerprinting techniques to improve accuracy and feature representation of the text. The evaluation results showed that the Winnowing + SVM model produced the most superior performance with an accuracy of 0.876 and F1-score of 0.868430. Despite requiring higher computation time (5.181894 seconds), SVM consistently classified the essays with better precision and stability. In contrast, the Winnowing + KNN model showed lower performance with an accuracy of 0.707 and F1-score of 0.686699, but had the advantage of much faster computation time (0.077723 seconds). The use of Winnowing proved effective in improving the performance of both models, with a more significant impact on SVM. In conclusion, SVM is more recommended for automated essay scoring systems that prioritize accuracy, while KNN is more suitable for applications that prioritize time efficiency.

**Keywords:** Winnowing; K-Nearest Neighbors; Support Vector Machine; Feature Extraction; Automated Essay Grading

### 1. PENDAHULUAN

Penilaian esai memiliki peran penting dalam pendidikan karena dapat mengukur kemampuan berpikir kritis, analisis, dan keterampilan komunikasi tertulis peserta didik. Namun, penilaian esai secara manual menghadapi beberapa tantangan, seperti subjektivitas waktu yang lama [1], dan ketidakkonsistenan. Untuk mengatasi tantangan-tantangan tersebut, sistem Automated Essay Scoring (AES) atau penilaian esai otomatis muncul sebagai solusi yang lebih objektif dan efisien [2].

AES memungkinkan penilaian esai dilakukan secara otomatis menggunakan algoritma dan model pembelajaran mesin, yang dapat mengurangi ketergantungan pada penilai manusia dan meningkatkan skalabilitas dalam proses penilaian [3]. Salah satu tantangan utama dalam AES adalah mengekstraksi fitur teks yang dapat mewakili kualitas tulisan esai secara akurat [4].

Metode *Winnowing*, yang awalnya dikembangkan untuk deteksi plagiarisme, memiliki potensi untuk mengubah esai menjadi fingerprint yang dapat digunakan untuk klasifikasi teks oleh algoritma pembelajaran mesin [4]. Kombinasi *Winnowing* dengan teknik Machine Learning (ML), seperti Support Vector Machine (SVM), K-Nearest Neighbors (KNN), dan Naive Bayes, dapat menjadi alternatif yang menarik untuk mengatasi masalah ekstraksi fitur dalam AES [5].

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental untuk menganalisis kinerja algoritma *Winnowing* dan machine learning (SVM, KNN) dalam sistem penilaian esai otomatis berbasis teks. Tahapan penelitian dimulai dari pengumpulan data esai berlabel skor dari dataset standar dari Kaggle dilanjutkan dengan preprocessing teks (pembersihan, tokenisasi, stemming), ekstraksi fitur menggunakan *Winnowing* dengan teknik k-gram dan fingerprinting, klasifikasi dengan tiga algoritma machine learning, serta evaluasi performa berbasis akurasi, F1-score, dan waktu komputasi untuk menentukan pendekatan terbaik.



Gambar 1. Alur Penelitian

### 2.1. Pengumpulan Data

Sumber Data Sekunder Dataset *Automated Student Assessment Prize (ASAP) AES* dari Kaggle yang terdiri dari 12.976 esai berbahasa Inggris dengan skor berdasarkan rubrik yang telah ditentukan [6]. <https://www.kaggle.com/competitions/learning-agency-lab-automated-essay-scoring-2/code>

### 2.2. Preprocessing

preprocessing dalam penelitian ini mengadaptasi berbagai teknik yang telah terbukti efektif dalam penelitian sebelumnya. Proses tokenisasi dan penghapusan stopword dilakukan menggunakan library NLTK [7] dengan daftar stopword bahasa Indonesia yang dikembangkan oleh [8] Untuk stemming, diterapkan dua pendekatan berbeda: algoritma Porter Stemmer (Porter, 1980) untuk teks bahasa Inggris dan library Sastrawi [9] untuk teks bahasa Indonesia guna mereduksi kata ke bentuk dasarnya. Selain itu, penelitian ini memanfaatkan SymSpell [10] untuk melakukan koreksi ejaan otomatis guna menormalisasi variasi penulisan kata yang mengandung kesalahan ketik. Setiap tahap preprocessing dipilih berdasarkan pertimbangan efektivitas dan kesesuaian dengan karakteristik data penelitian.

### 2.3. Ekstraksi Fitur

Proses ekstraksi fitur dilakukan dengan menggunakan algoritma *Winnowing*. Algoritma ini memanfaatkan teknik k-gram untuk membagi teks esai menjadi potongan-potongan kecil, yang kemudian diubah menjadi nilai hash menggunakan metode polynomial rolling hash. Nilai hash ini kemudian dipilih berdasarkan window untuk menghasilkan fingerprint yang mewakili fitur esai dalam bentuk vektor numerik [11], [12]

#### 2.4. Klasifikasi

Model klasifikasi dilatih menggunakan tiga algoritma Machine Learning yang populer. Pertama, Support Vector Machine (SVM) yang menggunakan kernel linear atau Radial Basis Function (RBF) dengan optimasi parameter C dan gamma melalui teknik GridSearchCV untuk meningkatkan performa model [13]. Terakhir, K-Nearest Neighbors (KNN) diimplementasikan dengan pengoptimalan parameter K menggunakan validasi silang untuk menentukan nilai optimal berdasarkan akurasi prediksi [14]. Ketiga algoritma ini dipilih karena keunggulannya dalam menangani berbagai karakteristik data.

#### 2.5. Evaluasi

Evaluasi kinerja model akan dilakukan dengan metrik akurasi, precision, recall, dan F1-Score untuk mengukur efektivitas prediksi, ketepatan klasifikasi data positif, dan kemampuan mengidentifikasi semua data positif. Waktu komputasi dihitung untuk menilai efisiensi training dan testing. Untuk memastikan generalisasi dan mengurangi bias, 5-fold cross-validation digunakan dengan membagi data menjadi lima bagian. Paired t-test diterapkan untuk membandingkan kinerja antar-model, memastikan perbedaan kinerja signifikan secara statistik dan bukan kebetulan.

#### 2.6. Pengacuan Pustaka

Dalam Penelitian pertama menganalisis sentimen ulasan aplikasi SIREKAP 2024 [16] menggunakan algoritma Naive Bayes (NB) dan SVM dengan pendekatan kuantitatif. Data dikumpulkan melalui web scraping, kemudian diproses dengan tahap preprocessing yang meliputi pembersihan teks, tokenisasi, dan stopword removal, serta ekstraksi fitur menggunakan TF-IDF. Hasil penelitian menunjukkan bahwa SVM mencapai akurasi 95%, lebih tinggi dibandingkan NB yang memperoleh akurasi 94%, dengan SVM juga memiliki recall dan F1-score yang lebih baik, menjadikannya lebih efektif dalam mengidentifikasi sentimen yang lebih kompleks. Kelebihan dari penelitian ini terletak pada kecepatan implementasi Naive Bayes dan kemampuan SVM dalam menangani data berdimensi tinggi, namun terdapat kelemahan terkait ketidakseimbangan kelas yang mempengaruhi kinerja pada kelas minoritas. Selain itu, penelitian ini tidak mengeksplorasi algoritma Winnowing, yang berpotensi meningkatkan analisis kualitas teks dalam Automated Essay Scoring (AES). Teknik preprocessing seperti TF-IDF yang digunakan dalam penelitian ini dapat diadaptasi untuk persiapan data esai dalam AES. Keunggulan SVM dalam klasifikasi teks yang kompleks mendukung penerapannya dalam AES untuk menilai kualitas esai dengan lebih efektif. Gap yang ditemukan adalah peluang untuk mengintegrasikan Winnowing dalam analisis fingerprinting esai, yang dapat meningkatkan akurasi dalam penilaian esai otomatis.

Penelitian kedua, Penelitian berjudul "Penilaian Esai Pendek Otomatis Berdasarkan Similaritas Semantik dengan SBERT" oleh Nurul Chamidah [17], Universitas Pembangunan Nasional Veteran Jakarta, bertujuan untuk mengevaluasi kinerja model penilaian esai otomatis dengan membandingkan esai jawaban mahasiswa dengan kunci jawaban evaluator berdasarkan kesamaan semantik. Penelitian ini menggunakan dataset bahasa Indonesia dari ujian mata kuliah Pengantar Basis Data dan memanfaatkan model Siamese-BERT (SBERT) untuk menghasilkan representasi numerik dari teks, yang kemudian dianalisis menggunakan Cosine Similarity untuk mengukur kesamaan semantik. Evaluasi model dilakukan dengan menggunakan metrik Mean Absolute Error (MAE) dan Pearson Correlation, dengan hasil MAE rata-rata 0.26 dan korelasi rata-rata 0.78. Model menunjukkan kinerja yang baik, meskipun ada tantangan dalam mengurangi error, yang tercermin dalam nilai MAE yang masih cukup besar. Penelitian ini menyimpulkan bahwa penggunaan SBERT efektif untuk penilaian esai otomatis dan menyarankan pelatihan ulang model dengan dataset yang lebih besar untuk meningkatkan akurasi.

Penelitian ketiga, Penelitian berjudul "Membandingkan Kinerja Metode Ekstraksi Fitur dan Model Pembelajaran Mesin dalam Penilaian Esai" oleh Lihua Yao [18] bertujuan untuk mengevaluasi kinerja berbagai metode ekstraksi fitur dan model pembelajaran mesin dalam sistem penilaian esai otomatis (AES). Penelitian ini menggunakan data dari Kaggle dan dataset ASAP, serta menerapkan NLP dan BERT untuk pemrosesan korpus dan ekstraksi fitur. Metodologi penelitian melibatkan empat langkah utama: pra-pemrosesan data, ekstraksi fitur, pelatihan model, dan evaluasi kinerja model menggunakan metrik seperti presisi, recall, akurasi, F-score, dan Quadratic Weighted Kappa (QWK). Hasil penelitian menunjukkan bahwa model jaringan saraf tiruan (ANN) yang dilatih pada semua data prompt memberikan kinerja terbaik dengan QWK keseluruhan 0.9724. Penambahan fitur seperti skor keterbacaan dari Spacy Textstat secara signifikan meningkatkan hasil prediksi, sementara model NLP menunjukkan kinerja lebih baik dibandingkan BERT. Namun, penelitian ini menghadapi keterbatasan terkait waktu komputasi dan kualitas data, serta tantangan dalam mengoptimalkan parameter fine tuning. Penulis menyarankan eksplorasi lebih lanjut terhadap jumlah fitur, teknik penambahan data, dan penerapan model transformer seperti GPT-2, GPT-3, T5, dan Roberta untuk meningkatkan akurasi penilaian esai otomatis. Penelitian ini memberikan wawasan penting mengenai penggabungan data pelatihan yang lebih komprehensif dan pemilihan model yang tepat untuk sistem penilaian esai otomatis.

### 3. HASIL DAN PEMBAHASAN

#### 4.1 Deskripsi Data dan Preprocessing

##### 4.1.1 Sumber Data

Dataset Automated Student Assessment Prize (ASAP) AES dari Kaggle terdiri dari 12.976 esai berbahasa Inggris yang diberi skor berdasarkan rubrik yang telah ditentukan. Dataset ini dapat diakses melalui kompetisi Kaggle Automated Essay Scoring. Dalam penelitian ini, digunakan 5.000 data dengan rincian 4.000 data untuk training dan 1.000 data untuk testing. Berikut adalah contoh tabel yang menggambarkan format data dalam dataset:

**Tabel 1.** Dataset

essay_id	full_text	score
555	Many people have come to believe that the "fac..."	4
3491	Think of how many car crashes a day steal love...	4
527	You may have seen Venus, or sometimes called t...	4
3925	I do not feel driverless cars are needed, because...	3
2989	Do aliens live on Mars? Some people believe ...	3

Penjelasan di atas menggambarkan sumber dataset, jumlah data, pembagian antara data pelatihan dan pengujian, serta contoh format tabel yang digunakan dalam penelitian ini.

##### 4.1.2 Tahapan Preprocessing

Menjelaskan langkah-langkah preprocessing yang dilakukan pada data, seperti pembersihan teks, tokenisasi, stemming, dan penghapusan stopwords.

**Tabel 2** Hasil sebelum dan sesudah preprosesing

Raw Text	Preprocessed Text
<i>What is the electoral college? The electoral college process plays a key role in the U.S. Presidential election.</i>	<i>electoral college electoral college process plays key role us presidential election</i>
<i>Driverless cars are not a good idea for the future. They may pose a safety risk and cause job losses.</i>	<i>driverless cars good idea future create safety risks cause job losses</i>
<i>Dear State Senator, I'm NOT in favor of the current healthcare plan.</i>	<i>dear state senator im favor keeping electoral health care plan</i>

<i>I'm going to tell you some positive and some negative ways that smart cars will affect the world.</i>	<i>im going tellin positive neative ways smart cars affect world</i>
<i>Is the face on Mars an alien artifact or just a natural landform?</i>	<i>face mars alien artifact natural landform far unknown</i>
<i>Driverless cars could be a good thing, I totally agree with the idea.</i>	<i>driverless cars could good thing totally agree idea</i>
<i>There are many advantages of limiting car usage. It decreases pollution and can improve overall health.</i>	<i>many advantages limiting car usage decreasing pollution improve health</i>
<i>The face on Mars is a landform in the region of Mars called Cydonia.</i>	<i>face mars landform region mars called cydonia</i>
<i>The challenge of exploring Venus. Exploring Venus would help us understand its atmosphere and geology.</i>	<i>challange exploring venus exploring venus would help understand atmosphere geology</i>
<i>Though it may seem amazing to think about having driverless cars on the road, the technology is still not quite ready.</i>	<i>though may seem amazing think driveless cars technology not ready</i>

Untuk menjelaskan proses ekstraksi fitur menggunakan Winnowing, kita dapat memecahnya dalam dua bagian utama sesuai dengan struktur yang Anda sebutkan: Teknik K-gram dan Fingerprinting.

#### 4.2 Ekstraksi Fitur Menggunakan Winnowing

**Winnowing** adalah salah satu teknik dalam pemrosesan teks yang digunakan untuk mengekstrak fitur penting dari teks. Dalam konteks ini, kita akan menjelaskan bagaimana Winnowing digunakan untuk mempercepat pencocokan kata, dengan menerapkan dua teknik utama: K-gram dan Fingerprinting.

##### 4.2.1 Teknik K-gram

**K-gram** adalah teknik dalam ekstraksi fitur yang memecah teks menjadi substrings atau segmen-segmen berukuran tetap, yang disebut "gram". Misalnya, untuk kata "electoral college" yang terdapat dalam raw text, kita bisa menghasilkan 2-gram atau 3-gram sebagai contoh. Dalam penelitian ini, Ramli et al. menjelaskan bahwa penggunaan K-gram dalam algoritma winnowing untuk deteksi plagiarisme dapat meningkatkan efektivitas sistem dalam mengenali dokumen yang memiliki kesamaan [15].

##### Penerapan K-gram dalam Winnowing:

Untuk setiap kata atau frasa dalam teks, kita memecahnya menjadi subsekuen berukuran tetap (misalnya  $K=3$ ), yang dikenal sebagai n-gram.

Misalnya, untuk kalimat "electoral college":

**K=2 (2-gram):** ['electoral college']

**K=3 (3-gram):** ['electoral college process']

Dengan cara ini, kita dapat membuat representasi fitur berbasis urutan kata-kata yang relevan di dalam teks yang memungkinkan kita untuk menangkap pola yang ada dalam teks. Menurut penelitian yang dilakukan oleh Narasimhulu et al., winnowing memungkinkan seleksi fitur yang efisien dalam klasifikasi biner, yang berguna dalam berbagai aplikasi deteksi plagiarisme dan pemrosesan teks [16]. Pendekatan ini memungkinkan sistem untuk fokus pada elemen-elemen yang paling berarti dalam teks, sehingga meningkatkan akurasi pencocokan kata dan efisiensi proses pencarian.

##### 4.2.2 Fingerprinting

Fingerprinting adalah teknik yang digunakan untuk menghasilkan representasi ringkas dari teks dengan tujuan mempercepat pencocokan teks atau deteksi duplikasi. Dalam konteks Winnowing, fingerprinting memungkinkan untuk mempercepat pencocokan kata dengan mengubah setiap segmen teks (misalnya, 2-gram atau 3-gram) menjadi nilai hash atau tanda tangan unik. Setelah teks dipecah menjadi K-gram, setiap segmen teks tersebut di-hash menggunakan fungsi hash tertentu untuk menghasilkan "fingerprint". Fingerprint ini kemudian digunakan untuk

mempbandingkan segmen-segmen teks tanpa harus membandingkan teks secara langsung, yang menghemat waktu dan sumber daya. Sebagai contoh, untuk frasa "electoral college" yang telah diproses dengan K-gram (misalnya, 3-gram), fingerprinting akan menghasilkan nilai hash untuk segmen-segmen tersebut, dan hanya hash yang dicocokkan, bukan seluruh kalimat atau kata, sehingga mempercepat proses pencocokan teks.

**Tabel 3** Fingerprinting

<i>Raw Text</i>	<i>K-gram</i>	<i>Fingerprint (Hash)</i>
"What is the electoral college?"	"electoral college"	Hash(electoral college)
"Driverless cars are not a good idea for the future"	"driverless cars", "good idea", "the future"	Hash(driverless cars), Hash(good idea), Hash(the future)

### 4.3 Penerapan Algoritma Machine Learning

#### 4.3.1 Algoritma K-Nearest Neighbors (KNN)

Algoritma **K-Nearest Neighbors (KNN)** digunakan untuk melakukan klasifikasi esai berdasarkan kedekatannya dengan data yang sudah dilabeli. Setelah melakukan ekstraksi fitur menggunakan **Winnowing**, yang melibatkan teknik k-gram dan fingerprinting, fitur yang diperoleh digunakan sebagai input untuk model KNN. **KNN** bekerja dengan cara menghitung jarak antar data yang tidak terlabel (misalnya esai yang belum diklasifikasikan) dengan data pelatihan yang sudah terlabel. Dalam hal ini, esai yang terdekat dengan titik data yang sudah dilabeli akan diklasifikasikan ke dalam kategori yang sama. Pada kode ini, nilai **k** ditetapkan sebesar 5, yang berarti model ini akan mencari lima tetangga terdekat untuk menentukan kategori dari setiap esai. Dengan memanfaatkan **Winnowing**, fitur yang diekstraksi menjadi lebih relevan dan efisien, memungkinkan KNN untuk bekerja dengan baik meskipun pada dataset yang besar dan beragam.

#### 4.3.2 Support Vector Machine (SVM)

Model Support Vector Machine (SVM) dengan kernel linear diterapkan untuk klasifikasi esai setelah ekstraksi fitur menggunakan Winnowing. SVM bekerja dengan cara mencari sebuah hyperplane terbaik yang membagi data ke dalam dua kelas yang berbeda (misalnya, skor rendah dan tinggi). Dalam hal ini, SVM digunakan untuk memisahkan esai ke dalam kategori Low, Medium, atau High berdasarkan fitur yang telah diekstraksi. Kernel linear dipilih karena kemampuannya yang efektif dalam menangani data dengan dimensi yang lebih tinggi dan distribusi yang lebih kompleks. Setelah data esai diproses dan fitur diekstraksi, SVM dengan kernel linear menghitung margin terbaik yang memisahkan kelas, menghasilkan model klasifikasi yang mampu mengklasifikasikan esai dengan akurasi tinggi. Teknik ini sangat berguna ketika data memiliki banyak fitur dan interaksi yang kompleks, seperti teks dalam esai [17].

### 4.4 Hasil Evaluasi Performa Model

#### 4.4.1 Evaluasi Model Winnowing + KNN

Model K-Nearest Neighbors (KNN) yang menggunakan metode ekstraksi fitur Winnowing menunjukkan performa yang relatif moderat dibandingkan model lainnya. Dalam hal akurasi, model ini mencapai nilai sebesar 0.707, yang berarti sekitar 70,7% prediksi kategori skor berhasil diklasifikasikan dengan benar. Nilai F1-Score yang diperoleh sebesar 0.686699 menunjukkan bahwa keseimbangan antara precision dan recall masih tergolong cukup, namun belum optimal. Nilai precision sebesar 0.686783 menandakan bahwa dari seluruh prediksi yang dikategorikan ke dalam

suatu kelas, sekitar 68,7% adalah benar, sedangkan nilai recall sebesar 0.707 mengindikasikan bahwa sekitar 70,7% dari seluruh data aktual berhasil dikenali oleh model.

#### 4.4.2 Evaluasi Model WInnowing + SVM

Model WInnowing + Support Vector Machine (SVM) menunjukkan performa yang unggul dalam evaluasi klasifikasi kategori skor. Dengan akurasi sebesar 0.876, model ini mampu mengklasifikasikan sekitar 87,6% data uji secara tepat. F1-Score yang mencapai 0.868430 mencerminkan keseimbangan yang sangat baik antara nilai precision dan recall. Nilai precision yang tinggi, yaitu 0.891291, menunjukkan bahwa sebagian besar prediksi yang dilakukan oleh model adalah benar, sementara nilai recall sebesar 0.876 menunjukkan bahwa sebagian besar data aktual berhasil dikenali dan diklasifikasikan sesuai kelasnya.

#### 4.4.3 Perbandingan Kinerja: WInnowing + KNN vs WInnowing + SVM

Berdasarkan hasil yang ditunjukkan dalam Tabel 4, terlihat adanya perbedaan kinerja yang cukup signifikan antara model WInnowing + KNN dan WInnowing + SVM. Model WInnowing + SVM menunjukkan performa yang lebih unggul dalam hampir seluruh metrik evaluasi. Akurasi model SVM mencapai 0.876, jauh di atas model KNN yang hanya memperoleh akurasi sebesar 0.707. Begitu pula pada metrik F1-Score, precision, dan recall, SVM konsisten menunjukkan keunggulan yang mencerminkan kemampuannya dalam menangkap pola klasifikasi yang kompleks secara lebih efektif. Hasil ini menunjukkan bahwa SVM lebih andal untuk digunakan dalam konteks klasifikasi kategori skor esai berbasis fitur winnowing.

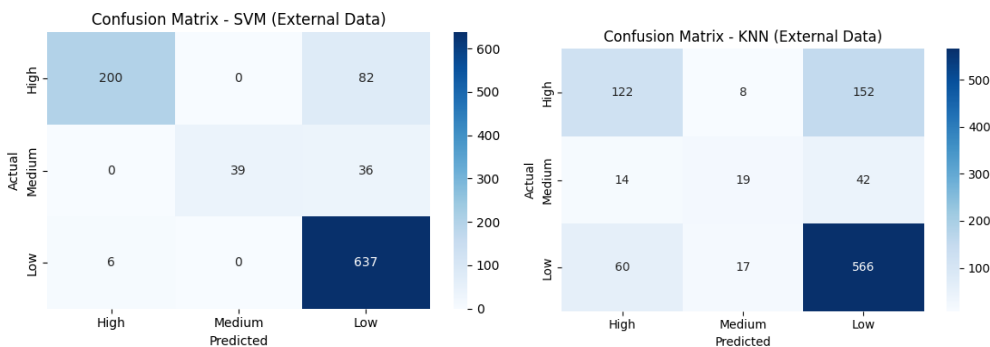
Dari segi efisiensi komputasi, model WInnowing + KNN memiliki waktu komputasi yang jauh lebih singkat, yakni hanya 0.077723 detik, dibandingkan dengan WInnowing + SVM yang memerlukan waktu sekitar 5.181894 detik. Namun, trade-off antara kecepatan dan ketepatan menjadi pertimbangan penting, terutama dalam sistem penilaian otomatis yang mengutamakan akurasi. Dengan demikian, meskipun KNN lebih ringan secara waktu proses, model SVM lebih direkomendasikan karena memberikan hasil klasifikasi yang lebih presisi dan stabil sebagaimana juga diungkapkan oleh Gomiasti et al. yang menunjukkan kelebihan SVM dalam mengatasi data dengan banyak fitur [18], terutama dalam pengolahan data teks dengan representasi fitur yang kompleks seperti WInnowing.

**Tabel 4** Perbandingan Model

Model	Akurasi	F1-Score	Precision	Recall	Waktu Komputasi (s)
WInnowing + SVM	0.876	0.868430	0.891291	0.876	5.181894
WInnowing + KNN	0.707	0.686699	0.686783	0.707	0.077723

### 4.5 Analisis Hasil

#### 4.5.1 Perbandingan Antara Model dan Metode



### Gambar 2. confusion matrix SVM dan KNN

Gambar 2 confusion matrix dari dua model klasifikasi, yaitu Winnowing + SVM dan Winnowing + KNN, yang masing-masing diterapkan pada data uji eksternal. Berdasarkan visualisasi ini, terlihat bahwa model Winnowing + SVM menunjukkan distribusi klasifikasi yang lebih konsisten dan akurat dibandingkan model Winnowing + KNN. Pada confusion matrix SVM, prediksi untuk kategori "Low" sangat dominan dan akurat dengan 637 prediksi yang benar. Selain itu, kategori "High" juga memiliki tingkat keberhasilan tinggi, yaitu 200 prediksi benar, dengan sedikit kesalahan klasifikasi ke kategori "Low" sebanyak 82 kasus. Sementara itu, untuk kelas "Medium", SVM berhasil memprediksi 39 data secara tepat.

Sebaliknya, confusion matrix dari model KNN menunjukkan lebih banyak kesalahan klasifikasi. Meskipun prediksi untuk kelas "Low" masih cukup tinggi (566 benar), terdapat penyimpangan yang cukup besar, seperti kelas "High" yang sebagian besar justru diklasifikasikan ke dalam kelas "Low" (132 salah klasifikasi), serta kelas "Medium" yang tidak hanya memiliki prediksi benar lebih rendah (19), tetapi juga salah dialokasikan ke kelas lain secara signifikan. Hal ini mengindikasikan bahwa KNN cenderung memiliki keterbatasan dalam membedakan fitur antar kelas ketika menggunakan representasi teks berbasis Winnowing.

Secara teori, SVM unggul karena merupakan model yang bekerja dengan baik dalam ruang berdimensi tinggi dan mampu membentuk hyperplane optimal yang memisahkan antar kelas secara maksimal, sehingga lebih cocok untuk data hasil ekstraksi teks seperti Winnowing. Seperti yang dijelaskan oleh Shin dan Gierl, SVM melakukan klasifikasi dengan mencari garis atau hyperplane di dalam suatu ruang multidimensi, yang dirancang untuk memisahkan objek menjadi kelas yang jelas dan berbeda [19]. Di sisi lain, KNN bergantung pada kedekatan jarak antar vektor dalam ruang fitur, sehingga lebih rentan terhadap noise dan tidak seefisien dalam menangani distribusi data yang kompleks. Oleh karena itu, kombinasi Winnowing + SVM terbukti lebih robust dan akurat untuk tugas klasifikasi skor esai dibandingkan Winnowing + KNN.

## 4. KESIMPULAN

Berdasarkan hasil evaluasi, dapat disimpulkan bahwa kombinasi metode ekstraksi fitur Winnowing dengan algoritma Support Vector Machine (SVM) memberikan performa yang lebih unggul dibandingkan dengan kombinasi Winnowing + K-Nearest Neighbors (KNN). Model Winnowing + SVM menunjukkan akurasi tertinggi sebesar 88 %, dengan nilai F1-Score sebesar 0.868 dan precision sebesar 0.891, yang mencerminkan kemampuannya dalam melakukan klasifikasi dengan presisi dan konsistensi yang tinggi. Meskipun waktu komputasi SVM lebih besar dibandingkan KNN, hasil klasifikasi yang lebih akurat menjadikannya pilihan yang lebih tepat untuk sistem penilaian esai otomatis.

Secara keseluruhan, SVM lebih cocok digunakan pada aplikasi yang mengutamakan akurasi meskipun sedikit mengorbankan kecepatan, sementara KNN lebih efisien dalam aplikasi yang memiliki keterbatasan waktu. Penelitian ini juga membuka kemungkinan pengembangan selanjutnya dengan mengeksplorasi teknik ekstraksi fitur lainnya atau melakukan penyempurnaan parameter pada kedua model menggunakan grid search untuk mencari konfigurasi terbaik, yang dapat meningkatkan kinerja dan waktu komputasi model secara keseluruhan.

## DAFTAR PUSTAKA

- [1] D. Ramesh dan S. K. Sanampudi, "An Automated Essay Scoring Systems: A Systematic Literature Review," *Artificial Intelligence Review*, vol. 55, no. 3, hlm. 2495–2527, 2021, doi: 10.1007/s10462-021-10068-2.
- [2] D. Ifenthaler, "Automated Essay Scoring Systems," hlm. 1–15, 2022, doi: 10.1007/978-981-19-0351-9\_59-1.

- [3] C. T. Lim, C. H. Bong, W. S. Wong, dan N. K. Lee, "A Comprehensive Review of Automated Essay Scoring (AES) Research and Development," *Pertanika Journal of Science and Technology*, vol. 29, no. 3, 2021, doi: 10.47836/pjst.29.3.27.
- [4] X. Wu, S. Qiao, dan Q. Tan, "Destination Management for Ecotourism Activity Using Analytical Hierarchy Process," *Scientific Programming*, vol. 2022, hlm. 1–6, Jan 2022, doi: 10.1155/2022/4143060.
- [5] R. Pradeep dan M. Kowsalya, "An Investigation of Several Models for Machine Learning Based Automated Essay Grading System," 2022, doi: 10.21203/rs.3.rs-1278674/v1.
- [6] Learning Agency Lab, "Learning Agency Lab - Automated Essay Scoring 2.0." [Daring]. Tersedia pada: <https://www.kaggle.com/competitions/learning-agency-lab-automated-essay-scoring-2/code>
- [7] S. Bird, E. Klein, dan E. Loper, *Natural Language Processing with Python*. O'Reilly Media, 2009.
- [8] F. Z. Tala, "A study of stemming effects on information retrieval in bahasa Indonesia," Universitas Indonesia, 2009.
- [9] Sastrawi Team, "Sastrawi: Indonesian stemmer." [Daring]. Tersedia pada: <https://github.com/sastrawi/sastrawi>
- [10] W. Garbe, "SymSpell: 1 million times faster spelling correction algorithm." [Daring]. Tersedia pada: <https://github.com/wolfganggarbe/SymSpell>
- [11] D. P. Sari dan A. Rahman, "Implementasi Algoritma Winnowing untuk Deteksi Kemiripan," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 2, hlm. 123–130, 2021.
- [12] H. Prabowo dan A. Setiawan, "Analisis Kinerja Algoritma Winnowing pada Pendeteksian Plagiarisme," *Jurnal Ilmiah Komputer dan Informatika*, vol. 10, no. 1, hlm. 45–52, 2022.
- [13] S. Goyal, "Effective software defect prediction using support vector machines (SVMs)," *Int J Syst Assur Eng Manag*, vol. 13, no. 2, hlm. 681–696, Apr 2022, doi: 10.1007/s13198-021-01326-1.
- [14] S. Goyal, "Handling Class-Imbalance with KNN (Neighbourhood) Under-Sampling for Software Defect Prediction," *Artif Intell Rev*, vol. 55, no. 3, hlm. 2023–2064, Mar 2022, doi: 10.1007/s10462-021-10044-w.
- [15] M. S. Ramli, S. Cokrowibowo, dan M. F. Rustan, "Uji Plagiarism pada Tugas Mahasiswa Menggunakan Algoritma Winnowing," *JACOST*, vol. 2, no. 2, hlm. 108–112, Des 2021, doi: 10.52158/jacost.v2i2.177.
- [16] Y. Narasimhulu, P. Kolambkar, dan V. V. China, "Winnowing modified online feature selection algorithm for efficient binary classification," *Statistical Analysis*, vol. 17, no. 4, Agu 2024, doi: 10.1002/sam.11707.
- [17] J. O. Contreras, S. M. S. Hilles, dan Z. A. Bakar, "Essay Question Generator based on Bloom's Taxonomy for Assessing Automated Essay Scoring System," 2021, doi: 10.1109/ICSCEE50312.2021.9498166.
- [18] F. S. Gomiasti, W. Warto, E. Kartikadarma, J. Gondohanindijo, dan D. R. I. M. Setiadi, "Enhancing Lung Cancer Classification Effectiveness Through Hyperparameter-Tuned Support Vector Machine," *J. Comput. Theor. Appl.*, vol. 1, no. 4, hlm. 396–406, Mar 2024, doi: 10.62411/jcta.10106.
- [19] J. Shin dan M. J. Gierl, "More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms," *Language Testing*, vol. 38, no. 2, hlm. 247–272, Apr 2021, doi: 10.1177/0265532220937830.