JEMSI: Jurnal Ekonomi Manajemen Sistem Informasi
E-ISSN: 2686-5238
P-ISSN: 2686-4916
https://dinastirev.org/JEMSI dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/jemsi.v7i4
https://creativecommons.org/licenses/by/4.0/
Received: 9 Maret 2026 | **Revised:** 4 April 2026 | **Accepted:** 20 April 2026

Analisis Sentimen Konsumen Sebagai Dasar Evaluasi Produk Industri Hiburan: Perbandingan Metode Random Forest dan Support Vector Machine Pada Drama Korea Vagabond

Anggelina Karolina Teti[1], Edhy Sutanta[2], Febriani Astuti[3], Inggit Fatika[4], Retno Widiastuti[5]
[1]Program Studi Statistika, Fakultas Sains dan Teknologi Informasi, Universitas AKPRIND, Indonesia, karolinateti24@gmail.com
[2]Program Studi Informatika, Fakultas Sains dan Teknologi Informasi, Universitas AKPRIND, Indonesia, edhy_sst@akprind.ac.id
[3]Program Studi Manajemen Bisnis Syariah, Fakultas Ekonomi dan Bisnis Islam, UIN Raden Mas Said Surakarta, Indonesia, peppydua@gmail.com
[4]Program Studi Statistika, Fakultas Sains dan Teknologi Informasi, Universitas AKPRIND, Indonesia, inggitfatika@akprind.ac.id
[5]Program Studi Teknik Industri, Fakultas Teknik, Universitas Sarjanawiyata Tamansiswa, Indonesia, dias.rw@ustjogja.ac.id

Corresponding Author: inggitfatika@akprind.ac.id[4]

Abstract: *Audience reviews are an important source of information for the entertainment industry in understanding consumer satisfaction and supporting business decision-making. This study aims to compare the performance of the Random Forest and Support Vector Machine (SVM) methods in sentiment analysis of audience satisfaction toward the Korean drama Vagabond. The data were collected through a Likert-scale questionnaire (1–5) representing overall satisfaction and various aspects of the viewing experience. Model evaluation was conducted using accuracy, sensitivity, precision, and error rate. The results indicate that Random Forest outperforms SVM, achieving an accuracy of 85.88%, a precision of 73.65%, and an error rate of 14.12%, while SVM obtained an accuracy of 76.47%, a precision of 47.7%, and an error rate of 23.53%. Although SVM demonstrates higher sensitivity, Random Forest is considered superior overall due to its ability to provide more accurate and stable classification of audience satisfaction. These findings suggest that Random Forest is more effective as an analytical tool to support marketing strategies and product management in the consumer data–driven entertainment industry.*

Keywords: *Sentiment Analysis, Random Forest, Support Vector Machine, Viewer Satisfaction, Vagabond.*

Abstrak: Ulasan penonton merupakan sumber informasi penting bagi industri hiburan dalam memahami kepuasan konsumen dan mendukung pengambilan keputusan bisnis. Penelitian ini bertujuan membandingkan kinerja metode *Random Forest* dan *Support Vector Machine* (SVM) dalam analisis sentimen kepuasan penonton drama Korea *Vagabond*. Data diperoleh melalui

kuesioner skala Likert (1–5) yang merepresentasikan kepuasan keseluruhan serta berbagai aspek pengalaman menonton. Evaluasi model dilakukan menggunakan akurasi, *sensitivity*, *precision*, dan *error rate*. Hasil penelitian menunjukkan bahwa *Random Forest* menghasilkan kinerja yang lebih baik dengan akurasi 85,88%, precision 73,65%, dan *error rate* 14,12%, dibandingkan SVM yang memperoleh akurasi 76,47%, precision 47,7%, dan error rate 23,53%. Meskipun SVM memiliki *sensitivity* yang lebih tinggi, *Random Forest* dinilai lebih unggul secara keseluruhan karena mampu memberikan klasifikasi kepuasan penonton yang lebih akurat dan stabil. Temuan ini menunjukkan bahwa *Random Forest* lebih efektif digunakan sebagai alat analitik dalam mendukung strategi pemasaran dan pengelolaan produk di industri hiburan berbasis data konsumen.

Kata Kunci: Analisis Sentimen, *Random Forest*, *Support Vector Machine*, Kepuasan Penonton, Vagabond.

PENDAHULUAN

Dalam era digital yang semakin berkembang, ulasan pengguna terhadap produk hiburan seperti film dan drama menjadi sumber informasi penting yang mampu memengaruhi persepsi publik. Melalui platform streaming seperti Netflix maupun media sosial, penonton dapat dengan mudah menyampaikan penilaian dan tanggapan mereka. Ulasan-ulasan ini dapat bersifat positif, netral, atau negatif (Rahman, et al., 2024). Ulasan tersebut tidak hanya memengaruhi minat calon penonton, tetapi juga menjadi masukan berharga bagi industri hiburan dalam meningkatkan kualitas produksi serta merumuskan strategi pemasaran.

Salah satu drama Korea yang banyak menarik perhatian adalah Vagabond. Drama ini mengisahkan perjuangan Cha Dal-gun, seorang stuntman yang berusaha mengungkap kebenaran di balik kecelakaan pesawat yang menewaskan keponakannya. Selain menghadirkan aksi dan ketegangan, Vagabond juga menyinggung isu-isu sosial dan politik yang relevan, sehingga memunculkan beragam tanggapan dari penonton. Ulasan penonton terhadap drama ini menyimpan informasi penting terkait alur cerita, pengembangan karakter, serta pesan moral yang disampaikan. Untuk mengetahui kecenderungan sentimen dari ulasan tersebut, diperlukan analisis data berupa analisis sentimen.

Analisis sentimen merupakan pendekatan yang memanfaatkan pemrosesan bahasa alami, analisis teks, dan linguistik komputasional untuk mengidentifikasi serta mengekstraksi informasi subjektif dari data berbentuk teks. Salah satu objek yang umum digunakan dalam analisis sentimen adalah ulasan film (Putra, et al., 2023). Analisis sentimen pada penelitian ini bertujuan mengklasifikasikan opini pengguna ke dalam kategori positif, netral, atau negatif (Rahman, et al., 2024). Proses ini semakin akurat dengan bantuan algoritma pembelajaran mesin (machine learning). Dua metode yang sering digunakan dalam melakukan analisis sentimen adalah *Support Vector Machine* (SVM) dan *Random Forest* (RF). Hasil analisis sentimen yang optimal dapat diperoleh melalui penerapan algoritma yang digunakan sebagai pembanding dalam proses analisis (Ardana & Kambau, 2023). Algoritma Support Vector Machine (SVM) banyak digunakan dalam klasifikasi data karena memiliki tingkat akurasi yang tinggi serta proses komputasi yang efisien (Al-shufi & Erfina, 2021). SVM dikenal efektif dalam menangani data berdimensi tinggi dengan margin pemisah yang optimal (Fitriyana, et al., 2023), sedangkan RF merupakan metode ensemble berbasis pohon keputusan yang mampu memberikan klasifikasi stabil serta mengurangi risiko *overfitting* (Effendi & Ernawati, 2025).

Hasil penelitian sebelumnya menunjukkan bahwa performa kedua metode dapat berbeda bergantung pada karakteristik data. Sidauruk, et al., (2023) menemukan bahwa SVM memiliki akurasi lebih tinggi daripada RF dalam analisis sentimen ulasan aplikasi *KAI Access.* Temuan serupa juga dilaporkan oleh (Syafia, et al., 2023). Namun, penelitian lain oleh Adrian, et al., (2021) menunjukkan bahwa RF justru memiliki performa lebih baik dibandingkan SVM.

Perbedaan temuan ini mengindikasikan perlunya penelitian lebih lanjut untuk menguji kedua metode dalam konteks yang berbeda.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan kinerja metode RF dan SVM dalam analisis sentimen ulasan penonton drama Korea Vagabond. Permasalahan utama yang dikaji adalah metode mana yang memiliki kinerja lebih baik dalam mengklasifikasikan sentimen penonton. Hasil penelitian ini diharapkan dapat memberikan kontribusi praktis bagi pelaku industri hiburan dan pemasaran digital dalam memahami preferensi konsumen, serta kontribusi akademik dalam pengembangan penerapan pembelajaran mesin untuk analisis sentimen di bidang ekonomi dan bisnis.

METODE

Desain Penelitian

Penelitian ini bertujuan untuk membandingkan kinerja dua metode *machine learning*, yaitu *Random Forest* dan *Support Vector Machine*, dalam menganalisis sentimen ulasan drama Korea "Vagabond". Analisis sentimen dilakukan untuk menentukan apakah ulasan yang diberikan oleh pengguna bersifat positif atau negatif (Munandar, et al., 2023). Data yang digunakan dalam penelitian ini diperoleh melalui kuesioner yang disebarkan kepada penonton drama "Vagabond". Kuesioner ini dirancang untuk mengumpulkan opini dan perasaan penonton terhadap drama tersebut.

Lokasi Penelitian

Lokasi penelitian ini dilakukan dengan menyebarkan kuesioner melalui media sosial dan langsung kepada teman-teman yang merupakan penggemar drama Korea.

Sumber Data

Data yang digunakan dalam penelitian ini adalah data primer, yaitu data yang diperoleh melalui kuesioner yang disebarkan kepada responden yang merupakan penggemar drama Korea "Vagabond" yang ditayangkan pada tahun 2019.

Variabel

Variabel penelitian ini terdiri dari satu variabel dependen, yaitu Kepuasan Keseluruhan (KK) yang mengukur tingkat kepuasan responden terhadap drama Vagabond secara umum, serta delapan variabel independen. Variabel independen tersebut meliputi: Frekuensi Menonton (FM) yang menunjukkan intensitas menonton drama, Penilaian Atas Cerita (PAC) terkait daya tarik alur, Penilaian Kekuatan Karakter (PKK) mengenai kekuatan peran karakter, Penilaian Perkembangan Karakter (PPK) tentang perkembangan karakter sepanjang cerita, Kepuasan terhadap Akting Pemeran Utama (KAPU), Kepuasan terhadap Akting Pemeran Pembantu (KAPP), Keberhasilan Penyampaian Pesan (KPP) terkait efektivitas penyampaian pesan moral atau emosional, serta Elemen yang Disukai (EYD) yang menilai jumlah elemen favorit dalam drama (alur, akting, visual, dll). Semua variabel diukur menggunakan skala Likert, dengan rentang nilai sesuai indikator masing-masing variabel.

Analisis Sentimen

Analisis sentimen merupakan pendekatan berbasis teknologi yang digunakan untuk mengidentifikasi respons, sudut pandang, serta polaritas opini individu baik positif, negatif, maupun netral dalam data tekstual yang bersumber dari media sosial (Kumar & Jaiswal, 2020). Pada dasarnya, analisis sentimen termasuk ke dalam proses klasifikasi teks (text classification) (Anwar, et al., 2021). Penerapannya dapat dilakukan pada tiga level analisis, yaitu tingkat dokumen, tingkat kalimat, serta tingkat aspek atau entitas (Sharma, et al., 2014).

Dalam penelitian ini, analisis sentimen difokuskan pada pengukuran kepuasan penonton terhadap drama Korea Vagabond melalui pendekatan klasifikasi berbasis machine learning.

Data diperoleh dari kuesioner yang menggunakan skala Likert 1–5 untuk merepresentasikan tingkat kepuasan secara keseluruhan maupun pada berbagai aspek pengalaman menonton, seperti alur cerita, kualitas akting, sinematografi, dan pengembangan karakter. Skor Likert tersebut kemudian dikategorikan ke dalam kelas sentimen (misalnya positif, netral, dan negatif) untuk membentuk variabel target dalam proses klasifikasi.

Penggunaan skala Likert dipilih karena mampu mengukur sikap dan persepsi responden secara terstruktur serta memberikan representasi kuantitatif terhadap opini yang bersifat subjektif. Selain itu, skala Likert memudahkan proses transformasi data ke dalam format numerik yang kompatibel dengan algoritma klasifikasi seperti Random Forest dan Support Vector Machine (SVM). Dengan pendekatan ini, analisis sentimen tidak hanya mengidentifikasi polaritas opini, tetapi juga memungkinkan evaluasi komparatif kinerja model dalam mengklasifikasikan tingkat kepuasan penonton secara lebih objektif dan terukur.

a) Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan metode supervised learning yang membangun fungsi pemetaan antara data masukan dan keluaran berdasarkan sekumpulan data latih yang telah diberi label (Wang, 2005). Pada metode ini, model dilatih menggunakan data dengan kategori yang telah diketahui untuk selanjutnya digunakan dalam melakukan prediksi atau klasifikasi terhadap data baru. Prinsip utama SVM terletak pada upaya meminimalkan kesalahan generalisasi dengan cara memaksimalkan jarak (margin) antara hyperplane pemisah dan titik data terdekat dari masing-masing kelas (Hardle & Simar, 2015). Margin yang lebih lebar diharapkan mampu menghasilkan model klasifikasi yang lebih optimal dan andal dalam menangani data yang belum pernah digunakan pada tahap pelatihan (Saigal, 2021).

b) Random Forest (RF)

Random Forest Classifier merupakan pengembangan dari algoritma decision tree, di mana setiap pohon keputusan dibangun menggunakan sampel data yang berbeda, serta pemilihan atribut dilakukan secara acak pada setiap proses pembentukan pohon (Herjanto & Carudin, 2024). Algoritma ini memiliki keunggulan dalam meningkatkan akurasi prediksi, khususnya ketika menghadapi data yang tidak lengkap, serta menunjukkan ketahanan terhadap keberadaan outlier dan efisiensi dalam penggunaan sumber daya penyimpanan (Efendi, et al., 2024). Selain itu, mekanisme seleksi fitur yang efektif memungkinkan identifikasi fitur-fitur paling relevan, sehingga secara keseluruhan mampu meningkatkan kinerja model klasifikasi (Manullang, et al., 2023).

c) Tahapan Analisis

Metode yang digunakan dalam penelitian ini adalah *Random Forest* dan *Support Vector Machine* dengan tahapan-tahapan sebagai berikut:

1. Melakukan Proses dimulai dari identifikasi dan persiapan data yang akan dianalisis.
2. Data dikumpulkan melalui penyebaran kuisioner kepada responden sesuai dengan topik penelitian. Data yang diperoleh masih dalam bentuk mentah dan belum siap untuk dianalisis langsung.
3. Melakukan uji validitas dan reliabilitas terhadap instrumen kuisioner untuk memastikan kelayakan item pertanyaan, diikuti dengan pembersihan dan penyesuaian data, yang meliputi: pemeriksaan kelengkapan data, pengubahan skala Likert menjadi bentuk numerik, serta pengkategorian dan standarisasi variabel sesuai kebutuhan analisis.
4. Data yang telah dibersihkan kemudian diacak menggunakan fungsi *set.seed*() dan dibagi menjadi data *training* dan data *testing*. Proses ini dilakukan dengan menggunakan tiga variasi seed yang sama untuk semua metode klasifikasi, yaitu: *set.seed* = 415, *set.seed* = 463, dan *set.seed* = 179.
5. Melakukan analisis klasifikasi dengan metode *Random Forest* (Biau & Scornet, 2016) (Story, 1986).

a. Menentukan parameter *mtry* (m) dan banyak pohon atau *ntree* (k).
b. Membangun model pada data *training* data *training* baru yang terbentuk dari proses *bootstrap* melalui tiga kali pengacakan.
c. Pemilihan model terbaik berdasarkan akurasi tertinggi.
d. Membangun pohon keputusan berdasarkan model terbaik sebanyak k kali sehingga diperoleh k buah pohon acak.
e. Melakukan prediksi klasifikasi berdasarkan data *testing*. Setiap pohon klasifikasi akan menghasilkan satu keputusan sehingga didapatkan k buah keputusan. Penentuan klasifikasi didasarkan pada keputusan terbanyak (*majority vote*)
f. Mengevaluasi ketepatan klasifikasi
g. Mengevaluasi kinerja klasifikasi

6. Melakukan analisis klasifikasi dengan metode *Support Vector Machine* (Suthaharan, 2016) (Kecman, 2005).
   a. Pra-pemrosesan Data
   b. Melakukan standarisasi data.
   c. Membangun model pada data *training* data *training* baru yang terbentuk dari tiga kali pengacakan.
   d. Melakukan analisis pada data *training* dengan pendekatan *Non-Linearly Separable Data*.
   e. Membangun model menggunakan kernel *Radial Basis Function* (RBF)
   f. Mengevaluasi Kinerja Model
7. Perbandingan kinerja hasil klasifikasi pada Metode *Random Forest* dan *Support Vector Machine*.

HASIL DAN PEMBAHASAN

Gambaran dan Karakteristik Data

Berdasarkan hasil analisis deskriptif, variabel Kepuasan Keseluruhan (KK) didominasi oleh kategori sangat puas sebanyak 284 sentimen, diikuti puas sebanyak 88 sentimen, sementara kategori lainnya relatif sedikit dan tidak ada yang memilih tidak puas sama sekali. Variabel Frekuensi Menonton (FM) menunjukkan mayoritas responden berada pada kategori sangat sering sebanyak 208 sentimen dan sering sebanyak 140 sentimen, sedangkan kategori tidak pernah tidak dipilih. Pada variabel Penilaian Alur Cerita (PAC), sebagian besar responden menilai alur cerita sangat menarik sebanyak 340 sentimen dan menarik sebanyak 78 sentimen, dengan kategori kurang menarik dan tidak menarik sama sekali tidak dipilih. Variabel Penilaian Kekuatan Karakter (PKK) didominasi oleh kategori sangat kuat sebanyak 258 sentimen dan kuat sebanyak 140 sentimen, sementara kategori lemah dan sangat lemah tidak ada. Selanjutnya, variabel Penilaian Perkembangan Karakter (PPK) terbagi seimbang antara kategori sangat berkembang dan berkembang masing-masing sebanyak 196 sentimen, dengan 30 sentimen pada kategori cukup berkembang. Pada variabel Kepuasan terhadap Akting Pemeran Utama (KAPU), mayoritas responden memilih kategori sangat puas sebanyak 354 sentimen dan puas sebanyak 60 sentimen, dengan hanya sedikit yang menilai cukup puas atau kurang puas. Variabel Kepuasan terhadap Akting Pemeran Pembantu (KAPP) didominasi kategori sangat puas sebanyak 228 sentimen dan puas sebanyak 154 sentimen, serta 40 sentimen pada kategori cukup puas. Untuk variabel Keberhasilan Penyampaian Pesan (KPP), kategori yang paling banyak dipilih adalah sangat berhasil sebanyak 230 sentimen dan berhasil sebanyak 154 sentimen, dengan sedikit responden menilai cukup berhasil. Terakhir, pada variabel Elemen yang Disukai (EYD), distribusi jawaban relatif merata, dengan kategori terbanyak adalah 2 elemen sebanyak 146 sentimen, disusul 3 elemen sebanyak 94 sentimen, 1 elemen sebanyak 114 sentimen, serta 4 elemen sebanyak 68 sentimen, sementara kategori 0 elemen tidak ada yang memilih.

Pembagian Data Training dan Testing

1. Pembagian Data *Training* dan Data *Testing* Proporsi 80:20 *Set.seed* (415)
   Dengan menggunakan *set.seed* 415, diperoleh total 337 data *training* dan 85 data *testing*. Pada kategori data, distribusi 80:20 menghasilkan 225 data *training* dan 59 data *testing* untuk kategori sangat puas, 68 data *training* dan 20 data *testing* untuk kategori puas, 22 data *training* dan 2 data *testing* untuk kategori cukup puas, serta 22 data *training* dan 4 data *testing* untuk kategori kurang puas. Tidak terdapat data pada kategori tidak puas sama sekali.
2. Pembagian Data *Training* dan Data *Testing* Proporsi 80:20 *Set.seed* (463)
   Dengan *set.seed* 463, jumlah data *training* tetap 337 dan data *testing* 85. Distribusi kategori menunjukkan 227 data *training* dan 57 data *testing* untuk kategori sangat puas, 68 data *training* dan 20 data *testing* untuk kategori puas, 20 data *training* dan 4 data *testing* untuk kategori cukup puas, serta 22 data *training* dan 4 data *testing* untuk kategori kurang puas. Tidak ada data pada kategori tidak puas sama sekali.
3. Pembagian Data *Training* dan Data *Testing* Proporsi 80:20 *Set.seed* (179)
   Dengan *set.seed* 179, jumlah data yang diperoleh tetap konsisten yaitu 337 data *training* dan 85 data *testing*. Kategori sangat puas terdiri dari 230 data *training* dan 54 data *testing*; kategori puas terdiri dari 73 data *training* dan 15 data *testing*; kategori cukup puas terdiri dari 16 data *training* dan 8 data *testing*; dan kategori kurang puas terdiri dari 18 data *training* dan 8 data *testing*. Tidak terdapat data pada kategori tidak puas sama sekali.

Klasifikasi Menggunakan Metode Random Forest

Dalam membangun model *Random Forest* yang optimal, penentuan parameter utama yaitu jumlah pohon keputusan (*ntree*) dan jumlah variabel prediktor acak pada setiap pemisahan cabang (*mtry*) menjadi langkah penting. Berdasarkan perhitungan, nilai *mtry* yang digunakan adalah 1, 2, dan 5, sedangkan *ntree* ditetapkan secara fleksibel dengan variasi 50, 100, 500, dan 1000, sehingga menghasilkan 12 percobaan model untuk setiap nilai *set.seed*. Data dibagi dengan proporsi 80% untuk latih dan 20% untuk uji, mengikuti praktik umum dalam machine learning agar pelatihan model tetap efektif sekaligus memungkinkan evaluasi yang objektif. Selain itu, pembagian data dilakukan secara acak menggunakan tiga nilai pengacakan (*set.seed* = 415, 463, dan 179) untuk memastikan keandalan hasil analisis.

1. Pemilihan Model Terbaik Metode *Random Forest*

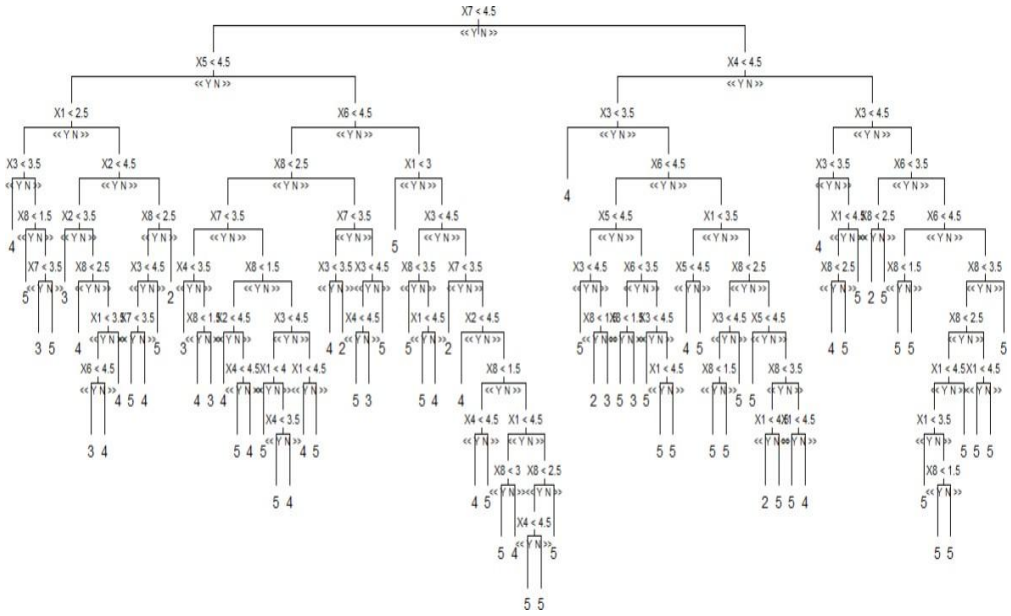
Tabel 1. Nilai Akurasi Model *Random Forest* untuk Proporsi 80:20

Pengacakan	Rata-rata Akurasi
Set.seed (415)	0,7693
Set.seed (463)	0,7826
Set.seed (179)	0,7878

Sumber: Data Riset

Berdasarkan Tabel 1, diperoleh nilai rerata tertinggi pada proporsi *data training* 80% menggunakan *set.seed* (179) yaitu sebesar 0,7878 atau 78,78% sehingga model terbaik yang diperoleh terdapat pada *mtry* = 5 dan *ntree* = 100 dengan nilai akurasi sebesar 0,8487 atau 84,87%. Berdasarkan hal tersebut maka penerapan metode *Random Forest* untuk mengklasifikasi Analisis Sentimen Ulasan Drama Korea *Vagabond* menggunakan model dengan *mtry* = 5 dan *ntree* = 100 pada proporsi 80%:20% dengan *set.seed* (179).

2. Pembentukan Pohon Klasifikasi Metode *Random Forest*



Sumber: Hasil Riset

Gambar 1. Pohon Klasifikasi Pertama Model m=5, dan k=100

Berdasarkan Gambar 1, maka interpretasi pohon klasifikasi pertama (*k = 1*) pada model *mtry* = 5 dan *k* = 100 sebagai berikut: Kepuasan keseluruhan terhadap ulasan Drama Korea *Vagabond* dikategorikan sebagai puas (Label 4) apabila memenuhi salah satu dari kriteria, Jika nilai X7 (Keberhasilan penyampaian pesan) < 4,5, X5 (Kepuasan terhadap akting pemeran utama) < 4,5, X1 (Frekuensi menonton) < 2,5, X3 (Penilaian kekuatan karakter dalam peran di drama) < 3,5.

3. Evaluasi Kinerja Klasifikasi Model *Random Forest*

Tabel 2. Tabel Klasifikasi pada *Random Forest*

Data Testing	**Prediksi**			
	Kurang Puas	Cukup Puas	Puas	Sangat Puas
Kurang Puas	5	0	0	0
Cukup Puas	0	5	0	2
Puas	0	2	11	0
Sangat Puas	3	1	4	52

Sumber: Data Riset

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ data} = \frac{5+5+11+52}{85} = \frac{73}{85} = 0{,}8588 \text{ atau } 85{,}88\%$$

$$Average\ Recall = \frac{1 + 0{,}714 + 0{,}846 + 0{,}867}{4} = 0{,}8568$$

$$Average\ Precision = \frac{0{,}625 + 0{,}625 + 0{,}733 + 0{,}963}{4} = 0{,}7365$$

$$Erorr\ Rate = 1 - Akurasi = 1 - 0{,}8588 = 0{,}1412\ atau\ 14{,}12\%$$

Berdasarkan hasil perhitungan dari Tabel 2 di atas, model klasifikasi sentimen menggunakan metode Random Forest menghasilkan tingkat akurasi sebesar 85,88% dengan *error rate* sebesar 14,12%. Capaian ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan sentimen ulasan penonton secara tepat. Nilai *sensitivity* (*recall*) rata-rata sebesar 85,68% mengindikasikan bahwa Random Forest efektif dalam mengidentifikasi sentimen yang muncul pada ulasan, sehingga berpotensi meminimalkan kesalahan dalam menangkap persepsi penonton. Sementara itu, nilai *precision* rata-rata sebesar 73,65% menunjukkan bahwa sebagian besar prediksi sentimen yang dihasilkan model

tergolong akurat. Dari perspektif ekonomi dan bisnis, performa ini menegaskan bahwa Random Forest cukup andal untuk digunakan sebagai alat analitik dalam memahami respons konsumen terhadap produk industri kreatif, seperti drama televisi, yang selanjutnya dapat dimanfaatkan sebagai dasar pengambilan keputusan strategis terkait pemasaran, pengembangan konten, dan evaluasi kualitas produk hiburan.

Klasifikasi Menggunakan Metode Support Vector Machine

1. Pembagian Data *Training* dan *Testing*

   Pada analisis *Support Vector Machine* ini, digunakan proporsi data yang sama seperti pada metode *Random Forest*, yaitu 80% data *training* dan 20% untuk data *testing*. Pembagian data dilakukan dengan menggunakan nilai *set.seed* = 179, yaitu *seed* terbaik yang diperoleh dari hasil pengujian pada metode *Random Forest*, agar perbandingan kinerja kedua model dilakukan secara adil pada dataset yang identik.
2. Penentuan Analisis Linearitas

   Berikut merupakan hasil akurasi yang diperoleh dari penggunaan metode *Support Vector Machine* dengan kernel linear dalam melakukan klasifikasi sentimen terhadap ulasan drama korea *vagabond*. Nilai akurasi tersebut ditampilkan pada Tabel 6.

Tabel 3. Hasil Akurasi SVM Linear

Akurasi SVM dengan Kernel Linear	0,6941

Sumber: Data Riset

Berdasarkan karakteristik data, penelitian ini menggunakan pendekatan *Non-Linearly Separable Data* karena data tidak dapat dipisahkan secara linier di ruang fitur aslinya. Hal ini dibuktikan dari uji awal dengan kernel linear pada *Support Vector Machine* yang hanya menghasilkan akurasi 69,41%, sehingga pemisahan kelas secara linier kurang optimal. Oleh karena itu, dipilih kernel *Radial Basis Function (RBF)* yang mampu memetakan data ke ruang berdimensi lebih tinggi sehingga menghasilkan batas keputusan lebih baik pada data dengan pemisahan non-linear, sekaligus efektif menangani data kompleks dan banyak digunakan dalam praktik.

3. Klasifikasi Data *Non-Linearly Separable* Menggunakan Kernel RBF

Tabel 4. Tabel Parameter Model SVM

Komponen	Nilai
SVM-*Type*	*C-classification*
SVM-*Kernel*	*Radial Basis Function (RBF)*
Parameter Cost (*C*)	1
Parameter Gamma (γ)	0,1
Support Vector	228

Sumber: Data Riset

Berdasarkan Tabel 4. model SVM yang digunakan bertipe *C-classification* dengan kernel *Radial Basis Function* (RBF), dipilih karena data tidak dapat dipisahkan secara linier dan kernel ini mampu memetakan data ke ruang berdimensi lebih tinggi untuk menghasilkan pemisahan kelas yang lebih optimal. Parameter *cost* (C) = 1 menunjukkan keseimbangan antara margin dan akurasi, sedangkan gamma = 0,1 memberi pengaruh cukup luas dari tiap data point terhadap pembentukan *hyperplane*, sesuai dengan data yang tidak terlalu tajam atau lokal. Model ini menghasilkan 228 support vector, menandakan struktur data relatif kompleks dan adanya kemungkinan tumpang tindih antar kelas sehingga banyak titik data diperlukan dalam menentukan batas keputusan.

4. Evaluasi Kinerja Klasifikasi Model *Support Vector Machine*

Tabel 5. Tabel Klasifikasi pada *Support Vector Machine*

Data Testing	Prediksi			
	Kurang Puas	Cukup Puas	Puas	Sangat Puas
Kurang Puas	2	0	0	0
Cukup Puas	0	1	0	0
Puas	0	1	8	0
Sangat Puas	6	6	7	54

Sumber: data Riset

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ data} = \frac{2+1+8+54}{85} = \frac{65}{85} = 0.7647 \text{ atau } 76.47\%$$

$$Average\ Recall = \frac{1 + 1 + 0{,}888 + 0{,}7397}{4} = 0{,}9069$$

$$Average\ Precision = \frac{0{,}25 + 0{,}125 + 0{,}533 + 1}{4} = 0{,}477$$

$$Erorr\ Rate = 1 - Akurasi = 1 - 0{,}7647 = 0{,}2353\ atau\ 23{,}53\%$$

Berdasarkan hasil perhitungan, model klasifikasi menggunakan metode SVM menghasilkan tingkat akurasi sebesar 76,47% dengan error rate 23,53%. Nilai *sensitivity* (*recall*) yang tinggi sebesar 90,69% menunjukkan bahwa model sangat efektif dalam mengidentifikasi sentimen dominan pada ulasan konsumen, sehingga berpotensi mendukung perusahaan dalam menangkap persepsi pasar secara luas. Namun, nilai *precision* yang relatif rendah sebesar 47,7% mengindikasikan bahwa ketepatan model dalam memprediksi sentimen tertentu masih terbatas, yang dapat berdampak pada kurang optimalnya segmentasi konsumen jika digunakan sebagai dasar pengambilan keputusan bisnis secara spesifik.

a) Perbandingan Kinerja Klasifikasi Metode *Random Forest* dan *Support Vector Machine*

Tabel 6. Evaluasi performa klasifikasi

	Random Forest	*Support Vector Machine*
Akurasi	85,88%	76,47%
Sensitivity	85,68%	90,96%
Precision	73,65%	47,7%
Erorr Rate	14,12%	23,53%

Sumber: Data Riset

Berdasarkan Tabel 6, metode Random Forest menunjukkan kinerja klasifikasi yang unggul pada data pengujian dengan tingkat akurasi sebesar 85,88%. Capaian ini mengindikasikan bahwa model mampu mengidentifikasi tingkat kepuasan penonton secara tepat, sehingga berpotensi memberikan gambaran yang andal bagi pengelola industri hiburan dalam memahami respons pasar terhadap suatu produk. Nilai *sensitivity* sebesar 85,68% menunjukkan kemampuan model dalam menangkap mayoritas persepsi konsumen yang sebenarnya, yang penting dalam mengidentifikasi sentimen dominan sebagai dasar perumusan strategi pemasaran dan pengembangan konten. Selain itu, nilai precision sebesar 73,65% mencerminkan ketepatan model dalam menghasilkan prediksi kepuasan, sehingga risiko kesalahan interpretasi preferensi konsumen relatif rendah. Rendahnya *error rate* sebesar 14,12% semakin menegaskan bahwa Random Forest mampu memberikan hasil klasifikasi yang stabil dan dapat diandalkan untuk mendukung pengambilan keputusan bisnis berbasis data.

Sementara itu, metode SVM menghasilkan tingkat akurasi sebesar 76,47%, yang meskipun masih tergolong baik, namun lebih rendah dibandingkan Random Forest. Keunggulan SVM terletak pada nilai *sensitivity* yang tinggi, yaitu 90,96%, yang menunjukkan kemampuan model dalam menangkap hampir seluruh respon kepuasan penonton. Hal ini bermanfaat dalam konteks bisnis ketika tujuan utama adalah meminimalkan potensi kehilangan

informasi terkait respons konsumen. Namun, nilai *precision* yang relatif rendah sebesar 47,7% mengindikasikan tingginya kemungkinan kesalahan dalam mengklasifikasikan kepuasan penonton, sehingga berisiko menghasilkan interpretasi pasar yang kurang akurat. *Error rate* yang lebih tinggi, yaitu 23,53%, juga menunjukkan bahwa model ini kurang optimal untuk digunakan sebagai dasar utama dalam pengambilan keputusan strategis.

Secara keseluruhan, hasil perbandingan menunjukkan bahwa Random Forest merupakan metode yang lebih efektif dalam mengklasifikasikan kepuasan penonton pada penelitian ini. Keunggulan *Random Forest* dalam hal akurasi, ketepatan prediksi, dan stabilitas model menjadikannya lebih sesuai untuk mendukung analisis perilaku konsumen, perencanaan strategi pemasaran, serta pengelolaan produk di industri hiburan yang berorientasi pada data dan kepuasan pelanggan.

KESIMPULAN

Berdasarkan hasil analisis, dapat disimpulkan bahwa metode Random Forest menunjukkan kinerja yang lebih unggul dibandingkan Support Vector Machine dalam mengklasifikasikan sentimen ulasan penonton Drama Korea *Vagabond*. Random Forest menghasilkan tingkat akurasi yang lebih tinggi, precision yang lebih baik, serta error rate yang lebih rendah, sehingga mampu memberikan hasil klasifikasi yang lebih stabil dan andal. Meskipun Support Vector Machine memiliki keunggulan pada nilai sensitivity yang tinggi, rendahnya precision dan tingginya error rate menunjukkan bahwa metode ini masih kurang optimal dalam menghasilkan prediksi kepuasan penonton yang akurat. Oleh karena itu, Random Forest dinilai lebih sesuai untuk digunakan sebagai alat analitik dalam memahami respons konsumen dan mendukung pengambilan keputusan strategis di industri hiburan berbasis data.

UCAPAN TERIMA KASIH

Publikasi ini didanai oleh Direktorat Jenderal Riset dan Pengembangan, Kementerian Pendidikan Tinggi, Sains, dan Teknologi melalui skema Penelitian Fundamental Tahun Anggaran 2025 dengan nomor kontrak SP DIPA-139.04.1.693320/2025 revisi 04 tertanggal 30 April 2025.

REFERENSI

Adrian, M., Putra, M., Rafialdy, M., & Rakhmawati, N. (2021). Perbandingan Metode Klasifikasi Random Forest dan SVM pada Analisis Sentimen PSBB. *Jurnal Informatika UPGRIS, 7*(1), 36-40.

Al-shufi, M., & Erfina, A. (2021). Sentimen Analisis Mengenai Aplikasi Streaming Film menggunakan Algoritma Support Vector Machine di Play Store. *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)* (hal. 156-162). Sukabumi: Universitas Nusa Putra.

Anwar, M. T., Permana, A. K., Ambarwati, L., & Agustin, D. (2021). Analyzing Public Opinion Based on Emotion Labeling Using Transformers. *2021 2nd International Conference on Innovative and Creative Information Technology (ICITech)*, (hal. 74-78). Salatiga.

Ardana, Y., & Kambau, R. (2023). Perbandingan Analisis Sentimen Algoritma Support Vector Machine dan Naive Bayes Terhadap Tanggapan Publik Tentang Pembelajaran Online di Masa Pandemi COVID-19. *Jaournal of Artificial Intelligence & Data Science, 3*(1), 1-7.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST 25*, 197–227.

Efendi, M., Sarwido, & Zyen, A. (2024). Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk. *RESOLUSI : Rekayasa Teknik Informatika dan Informasi, 5*(1), 12-20.

Effendi, P., & Ernawati, T. (2025). Analisis Sentimen Ulasan Aplikasi Game Hay Day Menggunakan Algoritma Random Forest. *JITET(Jurnal Informatika dan Teknik Elektro Terapan), 13*(3S1), 1-8.

Fitriyana, V., Hakim, L., Novitasari, D., & Asyhar, h. (2023). Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine. *Jurnal Buana Informatika, 14*(1), 40-49.

Hardle, W. K. (2015). *Applied Multivariate Statistical Analysis.* Berlin: Springer.

Herjanto, M., & Carudin. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap pada Play Store Menggunakan Algoritma Random Forest Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*, 1204-1210.

Kecman, V. (2005). Support Vector Machines: Theory and Applications. *Studies in Fuzziness and Soft Computing, 177*, 1-47.

Kumar, A., & Jaiswal, A. (2020). Systematic Literature Review of Sentiment Analysis On Twitter Using Soft Computing Technique. *Concurr. Comput. Pract. Exp., 32*(1).

Manullang, O., Prianto, C., & Harani, N. H. (2023). Analisis Sentimen untuk Menampilkan Hasil Calon Pemilu Presiden Menggunakan Lexicon Based dan Random Forest. *JIF (Jurnal Ilmiah Informatika), 11*(2), 159-169.

Munandar, A., Yaasin, M., & Firdaus, R. (2023). Analisis Sentimen Netizen Indonesia Mengenai Boikot Produk. *Tauhidinomics: Journal of Islamic Banking and Economics, 3*(1), 23-40.

Putra, F., Fadilah, F., & Enri, U. (2023). Analisis Sentimen Ulasan Film Oppenheimer Pada Situs Imdb Menggunakan Metode Naive Bayes. *Majalah Ilmiah UNIKOM, 21*(2), 87-94.

Rahman, I., Hasanah, A., & Heryana, N. (2024). Analisis Sentimen Ulasan Pengguna Aplikasi SAMSAT Digital Nasional (SIGNAL) dengan Menggunakan Metode Naive Bayes Classifier. *Jurnal Informatika Dan Teknik Elektro Terapan, 12*(2), 963-969.

Sharma, R., Nigam, S., & Jain, R. (2014). Polarity Detection At Sentence Level. *Int. J. Comput. Appl., 86*(11).

Sidauruk, N., Riza, N., & Fatonah, R. (2023). Penggunaan Metode SVM dan Random Forest untuk Analisis Sentimen Ulasan Pengguna Terhadap KAI Access di Google Playstore. *Jurnal Mahasiswa Teknik Informatika (JATI), 7*(3), 1901-1906.

Story, M. &. (1986). Remote Sensing Brief Accuracy Assesment: A User's Perspective. *Photogrammetric Engineering and Remote Sensing, 52*(3), 397-399.

Suthaharan, S. (2016). Support Vector Machine. In: Machine Learning Models and Algorithms for Big Data Classification. *Integrated Series in Information Systems, 36*, 207-235.

Syafia, A., Hidayattullah, M., & Suteddy, W. (2023). Studi Komparasi Algoritma SVM dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS. *Jurnal Informatika: Jurnal pengembangan IT (JPIT), 8*(3), 207-212.

Wang, L. (2005). *Support Vector Machines: Theory and Applications.* Berlin: Springer.