ANALISIS PERBANDINGAN RANDOM FOREST DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI TINGKAT RISIKO KECANDUAN JUDI ONLINE BERBASIS K-MEANS CLUSTERING PADA POLA TRANSAKSI PENGGUNA

Annisa Olivia Rafidah, Kaisyah Azzahra, Trya Andini Banowati,
Winda Kurnia Sari, Ken Ditha Tania, Fathoni
Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang – Prabumulih KM. 32 Indralaya, Kabupaten Ogan Ilir, Sumatera Selatan
09031382328130@student.unsri.ac.id

ABSTRAK

Perkembangan teknologi digital telah meningkatkan akses publik ke berbagai platform hiburan daring, termasuk judi *online*, yang berpotensi menimbulkan risiko kecanduan dengan dampak negatif pada kondisi finansial, psikologis, dan sosial pengguna. Penelitian yang secara spesifik menganalisis tingkat risiko kecanduan judi *online* berdasarkan pola transaksi masih sangat terbatas, khususnya yang menggabungkan teknik *clustering* dan klasifikasi secara bersamaan. Penelitian ini bertujuan menganalisis dan membandingkan performa algoritma *Random Forest* dan *Support Vector Machine (SVM)* dalam mengklasifikasikan tingkat risiko kecanduan judi online berdasarkan pola transaksi pengguna. Data dari 2.286 pengguna terlebih dahulu dikelompokkan menggunakan *K-Means Clustering* (k=2) untuk menghasilkan label kelas, kemudian diklasifikasikan menggunakan kedua algoritma dengan penanganan ketidakseimbangan kelas melalui parameter *class_weight='balanced'*. Hasil evaluasi menunjukkan bahwa *Random Forest* mencapai akurasi 98,47%, mengungguli *Support Vector Machine (SVM)* dengan kernel RBF yang mencapai 98,03%. *Random Forest* juga unggul pada 7 dari 9 metrik evaluasi dan menghasilkan lebih sedikit kesalahan prediksi (7 dari 458 data uji) dibandingkan *Support Vector Machine (SVM)* (9 dari 458). *Random Forest* direkomendasikan sebagai model yang lebih efektif untuk klasifikasi risiko kecanduan judi *online* berbasis pola transaksi pengguna.

Kata kunci: *Random Forest, Support Vector Machine, K-Means Clustering, Klasifikasi, Kecanduan Judi Online, Pola Transaksi.*

1. PENDAHULUAN

Kemajuan teknologi digital yang pesat membantu perkembangan berbagai platform hiburan daring, termasuk layanan situs judi *online*. Salah satu dampak sosial dari fenomena ini adalah risiko kecanduan judi *online* bagi pengguna platform digital yang semangkin meningkat. Peningkatan jumlah pengguna aktivitas judi *online* ini dipicu oleh kemudahan akses melalui perangkat elektronik ke situs judi *online*. Hal ini menimbulkan kekhawatiran tentang kemungkinan dampak negatif terhadap perilaku pengguna, terutama kecanduan yang berpengaruh ke kondisi psikologis, ekonomi, dan sosial pengguna. Menurut penelitian sebelumnya, aktivitas perjudian *online* mampu berkontribusi terhadap seberapa parah gangguan perjudian dan bagaimana seseorang pulih dari tantangan tersebut [1]. Selain itu, fenomena perjudian *online* semakin banyak dibahas dalam berbagai penelitian yang melihat bagaimana tanggapan dan perspektif masyarakat atas aktivitas judi pada situs perjudian *online* [2]. Akibatnya, dibutuhkan metode analisis yang dapat mendukung dalam mendeteksi ancaman kecanduan judi online sejak dini.

Dengan meningkatnya prevalensi perjudian *online*, semakin banyak data tentang aktivitas pengguna di platform perjudian digital yang tersedia, yang berkontribusi pada pemahaman perilaku pengguna. Data ini mencakup, frekuensi bermain, ukuran taruhan, deposit, dan metode transaksi. Data transaksi ini menunjukkan keterlibatan individu dalam perjudian *online* dan kemungkinan adannya perilaku perjudian yang bermasalah. Beberapa penelitian telah menunjukkan bahwa pembelajaran mesin dapat digunakan untuk memprediksi perilaku yang bermasalah dalam perjudian dengan menggunakan data yang dikumpulkan dari situs perjudian *online* [3] [4]. Selain itu, pembelajaran mesin telah digunakan untuk menganalisis berbagai aspek perjudian digital, seperti menemukan situs web yang ilegal dengan menganalisis data yang dikumpulkan dari internet [5].

Berbagai bidang penelitian menggunakan algoritma klasifikasi untuk memproses dan menganalisis data dalam bidang analisis data. *Support Vector Machine (SVM)* dan *Random* Forest adalah beberapa algoritma umum yang dikenal memiliki tingkat klasifikasi data yang tinggi. Studi sebelumnya telah menunjukkan bahwa algoritma ini dapat memberikan hasil perbandingan yang efektif dalam berbagai skenario data, seperti analisis ulasan aplikasi digital dan analisis opini publik di media sosial [6][7] [8][9]. Selain itu, pembelajaran mesin juga digunakan dengan sangat baik dalam bidang lain, seperti prediksi penyakit dan analitik digital [10].

Dalam pengolahan data, teknik *clustering* juga sering digunakan untuk mengelompokkan data berdasarkan fitur tertentu. *K-means* adalah salah satu metode *clustering* yang paling umum, yang memungkinkan data dibagi menjadi beberapa kelompok berdasarkan jarak. Metode ini telah

digunakan dalam berbagai penelitian kombinasi data, termasuk penelitian tentang data sosial, kesehatan, dan layanan sosial [11][12]. Proses pengelompokan data ini berkontribusi pada pemahaman yang lebih sistematis tentang karakteristik data dan dengan demikian mendukung pendekatan analitis baru menggunakan metode komparatif.

Beberapa penelitian telah menggunakan algoritma pembelajaran mesin untuk menganalisis perilaku pengguna, tetapi sebagian besar berkonsentrasi pada analisis sentimen atau evaluasi statistik data teks dari ulasan aplikasi dan media sosial. Penelitian yang secara khusus menganalisis tingkat risiko kecanduan judi *online* berdasarkan pola transaksi pengguna masih relatif terbatas. Selain itu, tidak banyak penelitian yang mengekslorasi cara mengkategorikan dan mengklasifikasikan perilaku perjudian *online* berdasarkan pola transaksi pengguna.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pendekatan analisis dengan memanfaatkan metode *clustering K-Means* untuk mengelompokkan pola transaksi pengguna berdasarkan karakteristik tertentu. Hasil dari Penggabungan data tersebut berfungsi sebagai dasar untuk klasifikasi menggunakan *Random Forest* dan *Support Vector Machine (SVM)*. Tujuan studi ini adalah untuk menganalisis dan membandingkan bagaimana *Random Forest* dan *SVM* bekerja dalam mengklasifikasikan risiko kecanduan judi *online* berdasarkan pola transaksi. Kinerja model akan ditentukan oleh metrik *accuracy, precision, recall,* dan *F1-score* guna menentukan tingkat performa dari masing-masing metode pada proses klasifikasi. Dengan demikian, penelitian ini bertujuan untuk memberikan gambaran umum tentang seberapa efektif algoritma klasifikasi dalam mengidentifikasi masalah kecanduan judi *online* dan membantu dalam pemahaman yang lebih baik tentang karakteristik pengguna yang berpartisipasi dalam perjudian *online.*

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu telah menguji metode pembelajaran mesin untuk menganalisis pola perilaku perjudian *online.* Dalam salah satu penelitian, algoritma kecerdasan buatan digunkan untuk menebak perilaku perjudian yang memiliki masalah dengan menggunakan data kegiatan pemain di situs web judi *online*. Ini memperlihatkan bahwa inormasi tentang perilaku pengguna bisa dimanfaatkan untuk mengenali risiko kecanduan judi pada tahap awal [3].

Penelitian lain, menganalisis perubahan perilaku pengguna dari waktu ke waktu pada situs judi *online* dengan memanfaatkan data akun pengguna standar. Metode pembelajaran mesin dipakai untuk menemukan perilaku pemain yang menyebabkan perjudian bermasalah, sehingga dapat mendukung diagnosis yang lebih tepat [4]. Selain itu, sebuah penelitian juga memanfaatkan pembelajaran mesin untuk menemukan situs judi *online* ilegal melalui berbagai sumber data. Hasilnya menampilkan bahwa model pembelajaran mesin dapat secara optimal menidentifikasi beberapa fitur situs judi ilegal [5].

Selain itu, beberapa penelitian berfokus untuk mengevaluasi kinerja berbagai algoritma klasifikasi dalam analisis data. Salah satu penelitian meneliti berbagai algoritma pembelajaran mesin dimanfaatkan untuk prediksi dalam basis data kesehatan, seperti regresi logistik, *Random Forest, Support Vector Machine,* dan *k-Nearest Neighbor.* Hasilnya memperlihatkan bawah metode *Support Vector Machine (SVM)* dan *Random Forest* memiliki kinerja yang kompetitif pada proses klasifikasi [10]. Menurut penelitian lain, *Support Vector Machine, Random Forest,* dan *XGBoos*t memiliki kinerja yang sama dalam klasifikasi data, dan hasil dari penelitian mereka menampilkan bahwa akurasi setiap algoritma bergantung pada properti dataset yang digunakan [8].

Dalam penelitian lain, algoritma support vector machines dan forest random dibandingkan untuk menemukan konten berbahaya di situs web. Hasilnya menunjukkan bahwa kedua metode tersebut dapat menunjukan kemampuan klasifikasi yang efektif pada proses identifikasi data [13]. Selanjutnya, metode *Support Vector Machine (SVM)* dan *Random Forest* diaplikasikan dalam penelitian guuna mengklasifikasikan penyakit hipertensi yang berdasarkan pada data Kesehatan. Hasil dari penelitian ini menyatakan bahwa kedua metode tersebut menghasilkan capaian yang setara saat memproses data Kesehatan [14].

Selain metode klasifikasi, Teknik clustering juga sering dimanfaatkan pada proses analisis data. Pada suatu penelitian, algoritma K-Means yanng fleksibel untuk analisis dengan dataset besar. Hasil dari penelitian ini memperlihatkan bahwa metode K-Means cocok untuk mengkategorikan data berdasarkan variable yang serupa [12]. Adapun penelitian lain yang menerapkan algoritma K-Means untuk mengumpulkan data penerima bantuan keuangan. Hasil dari penelitian ini memperlihatkan bahwa metode K-Means mampu secara optimal dalam memetakan data sehingga bisa membantu proses evaluasi dan pengambilan Keputusan [11].

2.2. Kecanduan Judi *Online*

Kemudahan Akses situs judi *online* melalui perangkat digital tidak luput dari pengaruh teknologi di era digital yang berkembang pesat. Kondisi ini meningkatkan kemungkinan munculnya masalah judi atau kecanduan judi, yang dapat mempengaruhi kondisi sosial, ekonomi, dan psikologis seseorang.

Penelitian lain menemukan bahwa aktivitas perjudian *online* memiliki beberapa ciri-ciri yang berbeda dari perjudian konvensional dan dapat meningkatkan risiko gangguan perjudian. Pengguna yang berpartisipasi dalam forum perjudian *online* di internet mempunyai kecenderungan lebih besar untuk mengalami gangguan perjudian serta memperlihatkan

pola perilaku perjudian yang dapat berdampak pada proses pemulihan dalam waktu yang panjang [1].

2.3. Algoritma *K-Means Clustering*

Clustering adalah metode yang digunakan untuk mengkategorikan data berdasarkan seberapa mirip karakteristiknya. Salah satu metode clustering yang paling populer adalah *K-Means*, yang memecah data ke beberapa *cluster* berdasarkan rentang antara *centroid cluster* yang terbentuk.

Algoritma *K-Means* memiliki beberapa kelebihan pada bagian kesederhanaan dan efisiensi proses pengkategorian data. Algoritma ini melakukan proses berulang guna memperbarui i agar tetap konsisten. Akan tetapi, metode *K-Means* juga mempunyai kekurangan seperti tingkat kepekaan terhadap pemilihan *centroid* awal serta kemungkinan terkunci pada *local optimum* [12].

2.4. *Machine Learning*

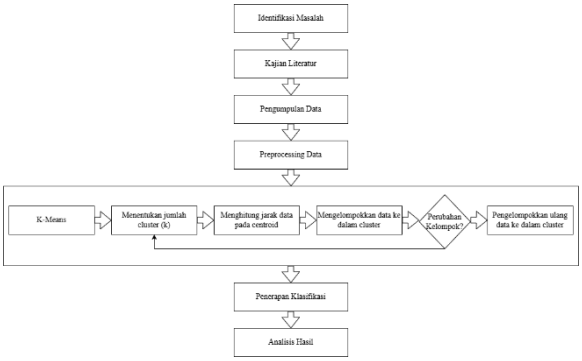
Machine Learning termasuk ke bagian dari teknologi kecerdasan buatan yang mendukung sistem komputer dalam menganalisis karakteristik dengan memanfaatkan data dan menciptakan keputusan atau perkiraan tanpa dirancang secara jelas. Teknik ini banyak diaplikasikan dalam bermacam sektor seperti analisis data, pengamatan pola perilaku pengguna, serta sistem prakiraan berlandaskan data digital.

Random Forest adalah algoritma klasifikasi yang berlandaskan *ensemble learning* yang menciptakan keputusan klasifikasi yang lebih tepat dengan mengembangkan pohon keputusan dan mengintegrasikan hasil prediksi dari setiap pohon. Metode ini dapat mengatasi data dengan jumlah atribut yang besar dan juga dapat menurunkan risiko *overfitting*. *Random Forest* sudah dimanfaatkan dalam penelitian terkait analisis perilaku judi *online* guna memperkirakan tingkat risiko masalah perjudian berdasarkan data aktivitas pengguna, dan memperlihatkan kinerja klasifikasi yang efektif dalam menentukan pengguna yang berisiko ketergantungan pada judi *online* [3].

Support Vector Machine (SVM) adalah salah satu metode *machine learning* yang paling umum diimplementasikan dalam proses klasifikasi data. metode ini mengidentifikasi *hyperplane* yang paling tepat, yang dapat mengelompokkan data menjadi dua atau lebih kategori dengan jarak pemisah tertinggi. *SVM* dapat mengatasi dataset yang sangat besar dan menghasilkan model klasifikasi dengan kemampuan adaptasi model yang berkualitas, sehingga metode ini banyak diterapkan dalam bermacam penelitian analisis data berlandaskan *machine learning* [15].

3. METODE PENELITIAN

Tahapan penelitian yang dilakukan dalam studi ini terdiri dari beberapa proses yaitu, Kajian literatur, pengumpulan dataset, preprocessing data, pelabelan menggunakan K-means clustering, dan penerapan klasifikasi dan dilanjutkan dengan evaluasi perbandingan, serta yang terakhir yaitu analisis hasil model klasifikasi.



Gambar 1. Tahapan Penelitian

3.1. Identifikasi Masalah

Tahap pertama dalam penelitian ini adalah melakukan identifikasi masalah yang berkaitan dengan meningkatnya aktivitas judi *online* yang dapat berpengaruh pada perilaku finansial pengguna. Pada tahap ini dilakukan observasi pada pola transaksi pengguna yang berpeluang memperlihatkan tingkat risiko kecanduan judi *online*. Permasalahan utama yang diangkat dalam penelitian ini adalah bagaimana mengelompokkan tingkat risiko kecanduan judi *online* berdasarkan pola transaksi serta membandingkan performa algoritma klasifikasi yang digunakan.

3.2. Kajian Literatur

Tahap awal dalam penelitian ini adalah mempelajari berbagai referensi seperti jurnal, artikel ilmiah, buku, dan publikasi yang berkaitan dengan topik penelitian yang membahas tentang kecanduan, judi *online*, mesin pembelajaran, *clustering*, serta metode klasifikasi *Random Forest* dan *Support Vector Machine (SVM)*. Kajian literatur ini dilakukan untuk mendapatkan pemahaman tentang gambaran dasar dan metode yang sesuai dengan topik penelitian yang dapat digunakan sebagai landasan atau acuan dalam penelitian [13].

3.3. Dataset

Langkah selanjutnya yaitu pengumpulan dataset. Dataset yang digunakan dalam penelitian ini didapatkan dari platform *Mendeley* Data https://data.mendeley.com/datasets/ yang secara terbuka memfasilitasi untuk penelitian. Dataset ini meliputi informasi dari aktivitas pengguna perjudian online. Dataset ini terdiri dari beragam variabel yang menjelaskan perilaku pengguna seperti jumlah deposit, total deposit, nilai deposit terbesar, lamanya waktu aktivitas pengguna, rata rata aktivitas pengguna, dan juga total kerugian bersih yang didapat pengguna.

3.4. Pra-pemrosesan Data

Prapemrosesan Data adalah langkah penting sebelum melakukan proses data mining ataupun machine learning dikarenakan kualitas data sangat mempengaruhi hasil dari analisis yang terbentuk oleh

model. Prapemrosesan data mencakup Penanganan data yang hilang (*missing value*), pembersihan data (*data cleaning*), transformasi data, dan normalisasi data untuk menyelaraskan skala antar atribut [16]. Pada penelitian ini, proses dilakukan dengan tahapan sebagai berikut:

a. Penanganan *missing value*: Mengatasi data yang kosong agar tidak mengganggu proses analisis dan pelatihan model mesin pembelajaran.
b. *Skewness Check*: Digunakan untuk mengetahui apakah distribusi data mempunyai kemiringan (*skewed*).
c. Transformasi logaritmik: diterapkan untuk mengurangi pengaruh nilai ekstrem dan menormalkan distribusi data.
d. Normalisasi data: Menggunakan *MinMaxScaler* untuk menyamakan skala fitur dalam rentang 0 hingga 1. Penting untuk *K-means* dan *Support Vector Machine (SVM).*
e. *Label Encoding*: Mengubah data kategorikal menjadi numerikal agar bisa diproses oleh algoritma.
f. Pembagian Data: Dataset dibagi menjadi data latih dan data uji dengan perbandingan 80:20 menggunakan *stratified sampling*.
g. *Data Scaling*: dilakukan menggunakan data latih untuk mencegah data leakage pada model.

Dengan melakukan tahap *preprocessing*, data yang sebelumnya masih mentah bisa diolah menjadi data yang konsisten dan siap digunakan pada tahap selanjutnya yaitu proses clustering menggunakan algoritma *K-Means* serta proses klasifikasi menggunakan algoritma *Random Forest* dan *Support Vector Machine.*

3.5. *Clustering K-Means*

K-Means Clustering merupakan salah satu metode *unsupervised learning* dalam data mining yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan antar data. Metode ini bekerja dengan cara membagi data ke dalam beberapa kelompok (*cluster*) berdasarkan jarak terdekat terhadap pusat *cluster* (*centroid*). Algoritma ini umumnya digunakan pada data angka karena proses pengelompokannya berdasarkan pada perhitungan jarak antar data [17]. Langkah-langkah proses *K-Means Clustering* adalah sebagai berikut:

a. Menentukan jumlah *cluster* (k) yang akan dibentuk.
b. Menentukan titik pusat *cluster* awal (*centroid*) sebanyak k dengan acak dari data yang ada.
c. Menghitung jarak setiap data terhadap masing-masing *centroid* menggunakan metode *Euclidean Distance*.
d. Mengelompokkan setiap data ke dalam *cluster* yang mempunyai jarak paling dekat dengan *centroid*.
e. Menghitung ulang nilai *centroid* baru berdasarkan rata-rata data yang berada pada *cluster* itu.
f. Melakukan ulang proses perhitungan jarak dan pembentukan *cluster* sampai posisi *centroid* tidak berubah lagi atau sudah meraih kondisi memusat atau menyatu.

Dengan menggunakan perhitungan jarak tersebut, setiap data akan diletakkan pada *cluster* yang mempunyai jarak paling dekat dengan *centroid* sehingga terbentuk kelompok data yang mempunyai karakteristik sejenis.

3.6. Penerapan Klasifikasi

Setelah proses *clustering* selesai, tahapan selanjutnya yang dilakukan pada penelitian ini yaitu, proses klasifikasi yang dilakukan menggunakan algoritma *Random Forest* dan *Support Vector Machine (SVM)* yang diterapkan menggunakkan bahasa pemrograman *Python*. Kedua algoritma ini dipilih karena sering digunakan dalam penelitian *machine learning* dan dapat menangani pola data yang cukup kompleks. *Random Forest* beroperasi dengan membentuk sejumlah pohon keputusan (*decision tree*) kemudian setiap pohon akan menghasilkan perkiraan masing-masing untuk menentukan kelas akhirnya, sedangkan *Support Vector Machine (SVM*) bekerja dengan mencari garis pemisah terbaik *(hyperplane)* yang dapat memisahkan data antar kelas dengan *margin* yang paling besar [14].

Dalam penerapannya, model *Random Forest* dikonfigurasi menggunakan beberapa parameter untuk mengatur jumlah pohon dan proses pembentukan cabang pada setiap pohon keputusan. Sementara itu, model *Support Vector Machine (SVM)* memanfaatkan fungsi kernel untuk membantu memetakan data ke ruang dimensi yang lebih tinggi sehingga data yang tidak dapat dipisahkan secara linear dapat diklasifikasikan dengan lebih baik. Kedua model kemudian dilatih menggunakan fungsi *fit*() pada data latih, lalu digunakan untuk melakukan prediksi pada data uji guna memperoleh hasil klasifikasi dari data yang telah diproses sebelumnya [14].

3.7. Analisis Hasil

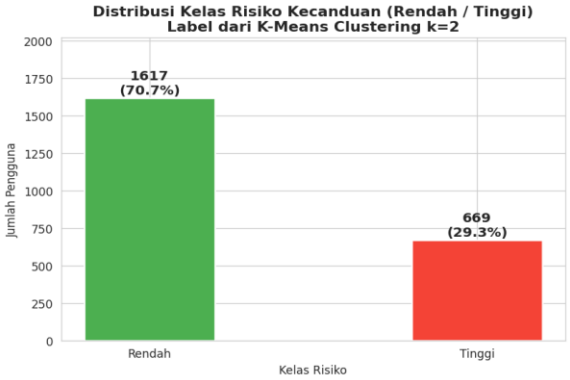
Tahap terakhir dalam penelitian ini adalah melakukan analisis terhadap hasil klasifikasi yang diperoleh dari kedua algoritma. Evaluasi model dilakukan dengan menggunakan *confusion matrix*, yaitu tabel yang digunakan untuk melihat perbandingan antara hasil perkiraan model dengan data asli. Dari *confusion matrix* tersebut selanjutnya dihitung beberapa metrik evaluasi seperti *accuracy, precision, recall*, dan *F1-score* untuk menghitung kinerja model dengan lebih terperinci.

Accuracy digunakan untuk mengetahui persentase keseluruhan perkiraan yang benar dari total data. *Precision* menampilkan seberapa tepat model dalam memperkirakan suatu kelas tertentu, sedangkan *recall* mendeskripsikan kemampuan model dalam

menemukan seluruh data yang memang termasuk dalam kelas tersebut. Sementara itu, *F1-score* merupakan nilai gabungan antara *precision* dan *recall* yang digunakan untuk memberikan gambaran keselarasan performa model, terutama ketika data yang digunakan mempunyai jumlah kelas yang tidak seimbang.

4. HASIL DAN PEMBAHASAN

4.1. Distribusi Kelas



Gambar 2. Distribusi Kelas Risiko Kecanduan

Menurut Gambar 1, hasil *K-Means Clustering* dengan k=2 menghasilkan dua kategori risiko kecanduan judi *online*. Dari 2.286 pengguna, 1.617 pengguna (70,7%) dikategorikan sebagai Risiko Rendah, sedangkan 669 pengguna (29,3%) dikategorikan sebagai Risiko Tinggi. Distribusi ini selaras dengan karakteristik demografi pengguna kasino *online* secara keseluruhan, di mana sebagian besar pengguna bersifat kasual dengan keterlibatan minimal, sedangkan sebagian kecil menunjukkan tanda-tanda perilaku kecanduan. Ketidakseimbangan kelas diatasi dengan menggunakan parameter *class_weight='balanced'* untuk kedua model.

4.2. Perbandingan *Random Forest dan Support Vector Machine (SVM)*

```
Classification Report:
              precision    recall  f1-score   support

      Rendah       0.99      0.99      0.99       324
      Tinggi       0.97      0.98      0.97       134

    accuracy                           0.98       458
   macro avg       0.98      0.98      0.98       458
weighted avg       0.98      0.98      0.98       458

Cross-Validation (5-fold): 98.73% ± 0.54%
```

Gambar 3. Laporan Klasifikasi *Random Forest.*

Tahap pengelompokkan menggunakan algoritma *Random Forest* mendapatkan nilai dari *precision*, *recall*, f1-score, dan *support*. Hasil Akurasi *Random Forest* mencapai akurasi total 98,47% pada data testing.

```
Classification Report:
              precision    recall  f1-score   support

      Rendah       0.99      0.98      0.99       324
      Tinggi       0.95      0.99      0.97       134

    accuracy                           0.98       458
   macro avg       0.97      0.98      0.98       458
weighted avg       0.98      0.98      0.98       458

Cross-Validation (5-fold): 98.16% ± 0.98%
```

Gambar 4. Laporan Klasifikasi *Support Vector Machine (SVM).*

Tahap pengelompokkan menggunakan algoritma *Support Vector Machine* mendapatkan nilai dari *precision*, *recall*, *f1-score*, dan *support. Support Vector Machine* yang menggunakan kernel RBF mencapai akurasi total 98,03%.

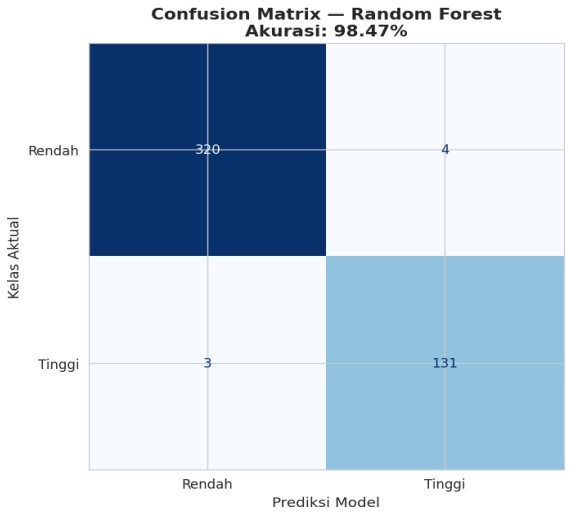
```
📌 Kelas: Rendah
  Metrik                RF        SVM       Unggul
  ------------------------------------------------
  Precision          0.9907    0.9937        SVM  (selisih: 0.0030)
  Recall             0.9877    0.9784         RF  (selisih: 0.0093)
  F1-Score           0.9892    0.9860         RF  (selisih: 0.0032)

📌 Kelas: Tinggi
  Metrik                RF        SVM       Unggul
  ------------------------------------------------
  Precision          0.9704    0.9496         RF  (selisih: 0.0207)
  Recall             0.9776    0.9851        SVM  (selisih: 0.0075)
  F1-Score           0.9740    0.9670         RF  (selisih: 0.0069)
```

Gambar 5. Laporan Perbandingan Klafikasi Random Forest dan *Support Vector Machine (SVM).*

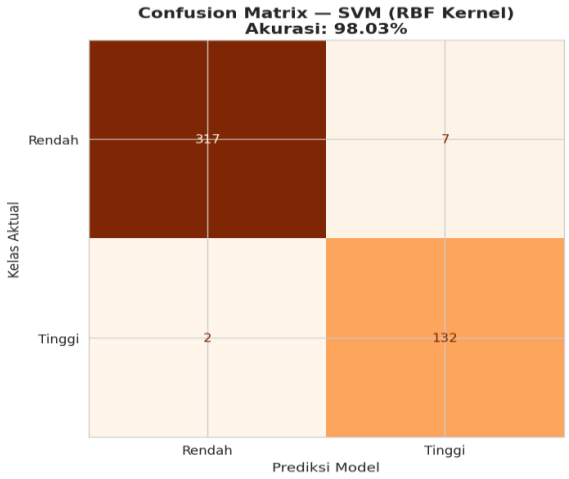
Pada perbandingan per kelas, *Random Forest* mengungguli *Support Vector Machine* dalam 7 dari 9 metrik. Pada kelas Rendah, *Random Forest* mengungguli *Support Vector Machine* dalam *Recall* (0,9877 vs 0,9784, selisih 0,0093) dan *F1-Score* (0,9892 vs 0,9860, selisih 0,0032), sedangkan *Support Vector Machine* sedikit lebih unggul dari *Random Forest* dalam *Precision* (0,9937 vs 0,9907, selisih 0,0030). Pada kelas Tinggi, *Random Forest* mengungguli *Support Vector Machine* dalam *Precision* (0,9704 dibandingkan dengan 0,9496, selisih 0,0207) dan *F1-Score* (0,9740 dibandingkan dengan 0,9670, selisih 0,0069), sedangkan *Support Vector Machine* lebih unggul dari *Random Forest* dalam *Recall* (0,9851 dibandingkan dengan 0,9776, selisih 0,0075).

4.3. *Confusion Matrix*

Confusion matrix Random Forest menunjukkan bahwa dari 324 pengguna Risiko Rendah dalam data test, 320 diklasifikasikan secara akurat sebagai Rendah, sedangkan 4 salah diidentifikasi sebagai Tinggi (*false positive*). Dari 134 pengguna Risiko Tinggi, 131 diidentifikasi secara akurat sebagai Tinggi, sedangkan 3 salah diklasifikasikan sebagai Rendah (*false negative*). Kesalahan *Random Forest* keseluruhan berjumlah 7 dari 458 sampel.



Gambar 6. *Confussion Matrix Random Forest*



Gambar 7. *Confusion Matrix Support Vector Machine (SVM).*

Confusion matrix Support Vector Machine menunjukkan bahwa dari 324 pengguna Risiko Rendah, 317 diklasifikasikan secara akurat sebagai Risiko Rendah, sedangkan 7 salah diidentifikasi sebagai Risiko Tinggi (. Dari 134 pengguna Risiko Tinggi, 132 diprediksi secara akurat sebagai Risiko Tinggi, sedangkan hanya 2 yang salah diklasifikasikan sebagai Risiko Rendah (*false negative*). Tingkat kesalahan keseluruhan untuk *Support Vector Machine* adalah 9 dari 458 sampel.

Jenis Kesalahan	Random Forest	SVM
True Negative (Rendah → Rendah)	320	317
False Positive (Rendah → Tinggi)	4	7
False Negative (Tinggi → Rendah)	3	2
True Positive (Tinggi → Tinggi)	131	132
Total Prediksi BENAR	451	449
Total Prediksi SALAH	7	9
Akurasi	98.47%	98.03%

Gambar 8. Perbandingan *Confussion Matrix Random Forest* dan *Support Vector Machine (SVM).*

Menurut tabel perbandingan *confusion matrix*, *Random Forest* menghasilkan total 7 kesalahan prediksi dari 458 data uji, lebih sedikit dibandingkan 9 kesalahan *Support Vector Machine*. *Random Forest* menunjukkan keunggulan dalam menekan *false positive* (4 dibandingkan 7), menunjukkan bahwa *Random Forest* lebih presisi dalam menghindari tindakan intervensi yang tidak perlu pada pengguna berisiko rendah. Satu-satunya keunggulan *Support Vector Machine* adalah lebih sedikitnya *false negative* (2 dibandingkan 3), namun selisihnya hanya satu kasus sehingga tidak bermakna secara praktis. Dalam konteks deteksi kecanduan judi *online*, *false negative* merupakan kesalahan yang lebih kritis karena pengguna yang benar-benar kecanduan dapat terlewat dari program intervensi. Meskipun demikian, secara keseluruhan *Random Forest* tetap unggul sebagai model yang lebih baik dengan total kesalahan lebih sedikit dan akurasi lebih tinggi.

5. KESIMPULAN DAN SARAN

Penelitian ini secara efektif mengkategorikan risiko kecanduan judi *online* melalui integrasi teknik K-*Means Clustering* dan algoritma *supervised learning*. *K-Means* secara efektif mengkategorikan 2.286 pengguna ke dalam dua kategori risiko: Risiko Rendah (70,7%) dan Risiko Tinggi (29,3%). Dalam membandingkan kedua algorima untuk klasifikasi, *Random Forest* mengungguli *Support Vector Machine*, mencapai akurasi 98,47%, didukung oleh perolehan *precision*, *recall*, dan *F1-Score* yang lebih baik pada sebagian besar kategori kelas.. Temuan penelitian ini menunjukkan bahwasanya *Random Forest* sedikit lebih efisien dalam mengidentifikasi risiko kecanduan judi *online* dibandingkan dengan *Support Vector Machine*. Studi tambahan disarankan untuk menggabungkan berbagai atribut perilaku pengguna yang lebih luas dan memakai dataset yang cakupannya lebih besar untuk meningkatkan generalisasi model ke skenario dunia nyata.

DAFTAR PUSTAKA

[1] G. Challet-Bouju *et al.*, “Impact of Gambling on the Internet on Middle-Term and Long-Term Recovery from Gambling Disorder: A 2-Year Longitudinal Study,” *J. Gambl. Stud.*, vol. 41, no. 2, pp. 567–592, Jun. 2025, doi: 10.1007/s10899-024-10328-0.

[2] M. Febrian As Shidiq, D. Alita, and L. Ratu Kec Kedaton Kota Bandar Lampung, “ANALISIS SENTIMEN MASYARAKAT TERHADAP KASUS JUDI ONLINE MENGGUNAKAN DATA DARI MEDIA SOSIAL X PENDEKATAN NAIVE BAYES DAN SVM,” *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, 2025.

[3] M. Auer and M. D. Griffiths, “Using artificial intelligence algorithms to predict self-reported problem gambling with account-based player data in an online casino setting,” *J. Gambl. Stud.*,

vol. 39, no. 3, pp. 1273–1294, Sep. 2023, doi: 10.1007/s10899-022-10139-1.

[4] S. Andersson, P. Carlbringorcid, K. Lyonorcid, M. Bermell, and P. Lindner, "Insights into the temporal dynamics of identifying problem gambling on an online casino: A machine learning study on routinely collected individual account data," *J. Behav. Addict.*, vol. 14, no. 1, pp. 490–500, Mar. 2025, doi: 10.1556/2006.2025.00013.

[5] M. Min and D. A. Lee, "Illegal online gambling site detection using multiple resource-oriented machine learning," *J. Gambl. Stud.*, vol. 40, no. 4, pp. 2237–2255, 2024.

[6] N. Febriyanti and A. F. Rozi, "Komparasi Algoritma Naïve Bayes, Support Vector Machine, dan Random Forest Untuk Analisis Sentimen Ulasan Pengguna Aplikasi CGV Cinemas Indonesia," *Technology and Science (BITS)*, vol. 7, no. 1, 2025, doi: 10.47065/bits.v7i1.7459.

[7] I. Nur Amalina, Norhikmah, and W. Miftahul Ashari, "Sistemasi: Jurnal Sistem Informasi Analysis of the Performance Comparison between Random Forest and SVM RBF in Detecting Cyberbullying on Imbalanced Data with the SMOTE Approach," vol. 14, no. 6, 2025, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[8] N. Epriyanti, A. Meiriza, and D. Yunika Hardiyanti, "Perbandingan Kinerja SVM, Random Forest dan XGBoost pada Aplikasi Access by KAI Menggunakkan ADASYN," *Jurnal Riset Komputer)*, vol. 12, no. 5, pp. 2407–389, 2025, doi: 10.30865/jurikom.v12i5.9134.

[9] F. Naifah Firzatullah and Nuroji, "Analisis Sentimen Pengguna Aplikasi Byond BSI Pada Google Play Store Menggunakan Algoritma SVM Dan Random Forest," vol. 9, p. 2025, doi: 10.47002/metik.v9i2.1089.

[10] N. Nasution, M. A. Hasan, and F. Bakri Nasution, "Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset," *IT Journal Research and Development*, vol. 9, no. 2, pp. 140–150, Apr. 2025, doi: 10.25299/itjrd.2025.17941.

[11] N. Nurahman and J. Susanto, "Klasterisasi Data Penerima Bantuan Langsung Tunai Menggunakan Algoritma K-Means," *JURIKOM (Jurnal Ris. Komputer)*, vol. 10, no. 2, p. 461, 2023.

[12] H. Hu, J. Liu, X. Zhang, and M. Fang, "An effective and adaptable K-means algorithm for big data cluster analysis," *Pattern Recognit.*, vol. 139, p. 109404, 2023.

[13] H. Syahputra and A. Wibowo, "Comparison of Support Vector Machine (SVM) and Random Forest Algorithm for Detection of Negative Content on Websites," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 9, no. 1, pp. 165–173, Mar. 2023, doi: 10.26555/jiteki.v9i1.25861.

[14] S. Alia Azhaar, T. Al Mudzakir, H. Yulia Novita, and S. Faisal, "Implementasi Algoritma Support Vector Machine (SVM) dan Random Forest Untuk Klasifikasi Penyakit Hipertensi Berdasarkan Data Kesehatan," *Jurnal Riset Komputer)*, vol. 12, no. 4, pp. 2407–389, 2025, doi: 10.30865/jurikom.v12i4.8744.

[15] V. Stavropoulos *et al.*, "Deep learning(s) in gaming disorder through the user-avatar bond: A longitudinal study using machine learning," *J. Behav. Addict.*, vol. 12, no. 4, pp. 878–894, Dec. 2023, doi: 10.1556/2006.2023.00062.

[16] S. K. Dirjen *et al.*, "Terakreditasi SINTA Peringkat 2 Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *masa berlaku mulai*, vol. 1, no. 3, pp. 117–122, 2017.

[17] D. Aldo, "Data Mining Sales of Skin Care Products Using the K-Means Method," *Sinkron*, vol. 8, no. 1, pp. 295–304, Jan. 2023, doi: 10.33395/sinkron.v8i1.12007