AnoaTIK: Jurnal Teknologi Informasi dan Komputer e-ISSN: **2987-7652**
Vol. 1, No. 2, Desember 2023 *Hal. 54-58*

STRATEGI PROMOSI PERBANKAN DENGAN ANALISIS DAN IMPLEMENTASI ALGORITMA RANDOM FOREST

Ilham Indrastata

Universitas Mercu Buana
Email: ilhamindra79@gmail.com

Abstrak

Penelitian ini bertujuan untuk menganalisis dan mengimplementasikan algoritma *Random Forest* dalam konteks perbankan untuk menentukan strategi promosi yang efektif. Strategi promosi yang tepat sangat penting bagi lembaga keuangan seperti bank, karena dapat meningkatkan loyalitas nasabah, meningkatkan basis nasabah, dan mencapai pertumbuhan bisnis yang berkelanjutan. Pada tahap analisis, sejarah promosi dan respon pelanggan dari periode sebelumnya akan digunakan untuk membangun model. Algoritma *Random Forest* digunakan untuk mengidentifikasi variabel yang paling berpengaruh dalam menentukan efektivitas promosi. Penelitian ini diharapkan dapat memberikan wawasan baru bagi industri perbankan dalam menentukan strategi promosi yang lebih efektif dan menghasilkan pengelolaan sumber daya yang lebih efisien. Dengan memanfaatkan algoritma *Random Forest*, diharapkan bank dapat meningkatkan tingkat keberhasilan promosi, mengurangi biaya yang rumit, dan mencapai tujuan bisnis yang lebih baik. Namun, penelitian ini juga menyadari bahwa faktor lain seperti ekonomi makro, tren pasar, dan kebijakan perusahaan juga perlu diperhatikan dalam merumuskan strategi promosi yang komprehensif.

Kata kunci: *promosi, perbankan, random forest, algoritma*

Abstract

This research aims to analyze and implement the Random Forest algorithm in the context of banking to determine effective promotion strategies. Appropriate promotion strategies are crucial for financial institutions such as banks, as they can enhance customer loyalty, expand customer bases, and achieve sustainable business growth. In the analysis phase, historical promotion data and customer responses from previous periods will be used to build a model. The Random Forest algorithm will be employed to identify the most influential variables in determining promotion effectiveness. This study is expected to provide fresh insights to the banking industry in crafting more effective promotion strategies and achieving more efficient resource management. By leveraging the Random Forest algorithm, banks anticipate improving promotion success rates, reducing complex costs, and attaining better business objectives. However, this research also acknowledges that other factors such as macroeconomics, market trends, and corporate policies need to be considered in formulating a comprehensive promotion strategy.

Keywords: *promotion, strategy, random forest, algorithm*

1. PENDAHULUAN

Perbankan adalah sektor yang sangat kompetitif dan kompleks, di mana perusahaan-perusahaan perbankan harus terus menerapkan strategi promosi yang efektif untuk mempertahankan dan meningkatkan pangsa pasar mereka. Strategi promosi yang tepat dapat membantu perbankan untuk menarik nasabah baru, mempertahankan nasabah yang ada, dan membangun loyalitas pelanggan. Namun, dalam menghadapi persaingan yang semakin ketat, perbankan dihadapkan pada tantangan untuk mengidentifikasi strategi promosi yang paling sesuai untuk berbagai kelompok pelanggan dengan preferensi yang berbeda.

Dalam menghadapi tantangan ini, teknologi dan analisis data telah menjadi aspek yang semakin penting dalam pengembangan strategi promosi yang berhasil. Pemanfaatan algoritma machine learning, seperti *Random Forest* dan *K-Nearest Neighbors*, memberikan potensi untuk menganalisis data dengan lebih mendalam, mengidentifikasi pola-pola tersembunyi, dan menyusun strategi promosi yang lebih tepat dan terfokus.

Algoritma *Random Forest* adalah metode ensemble yang menggabungkan beberapa pohon keputusan untuk mencapai hasil yang lebih akurat dan konsisten dalam klasifikasi dan regresi. Sementara itu, algoritma *K-Nearest Neighbors* (K-NN) adalah metode pembelajaran tanpa pengawasan yang

berfokus pada pengelompokan data berdasarkan kesamaan atribut.

Beberapa penelitan terkait yang telah dilakukan oleh para peneliti sebelumnya mencoba untuk memprediksi Kasus Terkonfirmasi, Meninggal, dan Sembuh COVID-19 di India Menggunakan Model *Random Forest* yang dilakukan oleh Gupta V (2021), kemudian Zailani A. menerapkan algoritma random forest ini untuk menetukan kelayakan pemberian kredit pada koperasi (2020). selain itu Lediyana H. menggunakan algoritma berbeda yaitu KNN untuk penentuan resiko kredit kepemilikan kendaraan bermotor.

Oleh karena itu, penelitian ini dilakukan dengan tujuan untuk mengimplementasikan kinerja algoritma Random Forest dalam menentukan strategi promosi yang paling efektif dalam konteks perbankan. Dengan menggunakan data historis promosi dan respons pelanggan, penelitian ini akan menganalisis variabel-variabel yang paling berpengaruh dalam efektivitas promosi, serta mengidentifikasi kelompok-kelompok pelanggan dengan perilaku serupa untuk segmentasi pasar yang lebih baik.

2. METODE PENELITIAN

Dalam mengimplementasikan metode kuantitatif pada penelitian ini, langkah pertama adalah mengumpulkan dataset yang relevan dari sumber data yang terpercaya, yaitu Bank Marketing *dataset* yang tersedia secara publik melalui tautan https://archive.ics.uci.edu/ml/datasets/bank+marketing. *Dataset* ini mencakup berbagai atribut yang terkait dengan kampanye pemasaran langsung yang dilakukan oleh sebuah lembaga perbankan di Portugal. Informasi ini termasuk data demografis pelanggan, kontak telepon, serta hasil dari kampanye pemasaran tersebut, di mana pelanggan diminta untuk berlangganan produk deposito bank ('yes') atau tidak ('no').

Setelah data dikumpulkan, langkah selanjutnya adalah melakukan *preprocessing*. Ini mencakup langkah-langkah seperti membersihkan data dari nilai yang hilang atau tidak valid, menormalkan variabel, dan mengubah format data jika diperlukan. Tujuan dari tahap *preprocessing* ini adalah untuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang baik dan siap digunakan dalam proses selanjutnya.

Setelah tahap *preprocessing* selesai, penelitian ini dilanjutkan dengan mengaplikasikan algoritma *Random Forest* untuk perhitungan. Algoritma ini digunakan untuk membangun model prediksi yang dapat mengidentifikasi faktor-faktor yang paling berpengaruh dalam menentukan apakah pelanggan akan berlangganan produk deposito bank atau tidak. Kemudian, ini akan membantu lembaga perbankan dalam mengambil keputusan yang lebih baik dalam merancang strategi pemasaran mereka.

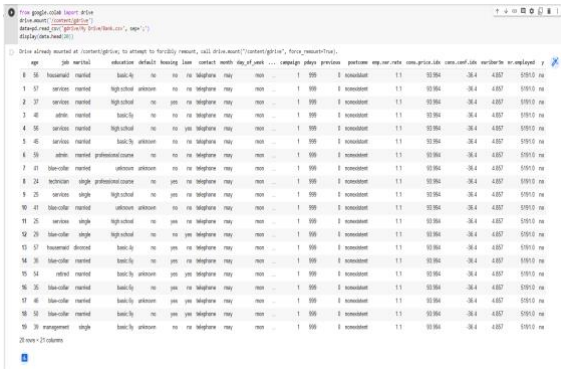
Setelah hasil perhitungan dengan menggunakan algoritma *Random Forest* didapatkan, penelitian akan beralih ke tahap analisis hasil. Ini melibatkan evaluasi model, interpretasi variabel yang signifikan, dan pengambilan wawasan yang dapat digunakan untuk merumuskan strategi promosi yang lebih efektif dalam konteks perbankan.

Terakhir, penelitian ini dilanjutkan dengan tahap implementasi algoritma. Hasil dan wawasan dari analisis akan diterapkan dalam praktik perbankan, memungkinkan lembaga tersebut untuk mengoptimalkan kampanye pemasaran mereka dan meningkatkan efisiensi sumber daya dalam upaya untuk meningkatkan jumlah pelanggan yang berlangganan produk deposito bank.

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Kumpulan data ini menjelaskan hasil promosi pemasaran langsung dari lembaga perbankan Portugis. Kampanye pemasaran didasarkan pada panggilan telepon. Promosi yang dilakukan sebagian besar didasarkan pada panggilan telepon langsung, menawarkan klien bank untuk menempatkan deposito berjangka. Jika setelah semua upaya penandaan klien telah setuju untuk menempatkan deposit - variabel target ditandai 'ya', jika tidak 'tidak'.



Gambar 1. *Dataset* Penelitian

Data klien bank yang digunakan dalam penelitian ini mencakup berbagai atribut yang memberikan wawasan tentang profil pelanggan dan interaksi mereka dengan layanan perbankan. Pertama, atribut umur (*age*) mencatat usia pelanggan secara numerik. Selanjutnya, atribut jenis pekerjaan (*job*) menggambarkan beragam jenis pekerjaan yang dapat dimiliki oleh klien, dengan kategori seperti '*admin*', '*blue-collar*', '*services*', dan lainnya. Status perkawinan (*marital*) menjadi atribut yang mencatat apakah klien tersebut '*divorced*', '*married*', '*single*', atau '*unknown*'.

Selanjutnya, atribut pendidikan (*education*) mencatat tingkat pendidikan klien, mulai dari '*basic 4y*' hingga '*university degree*'. Atribut '*default*' mencatat apakah klien memiliki kredit *default* dengan kategori 'tidak', 'ya', atau 'tidak diketahui'. Kemudian, atribut '*housing*' mencatat apakah klien memiliki pinjaman perumahan dengan pilihan 'tidak', 'ya', atau 'tidak diketahui'. Atribut '*loan*' mencatat apakah klien

memiliki pinjaman pribadi dengan pilihan 'tidak', 'ya', atau 'tidak diketahui'.

Berkaitan dengan kontak terakhir promosi saat ini, atribut '*contact*' mencatat jenis komunikasi kontak, dengan pilihan '*celluler*' atau '*telephone*'. Atribut '*month*' mencatat bulan kontak terakhir dalam setahun, seperti 'jan' atau 'feb'. Selanjutnya, atribut '*day_of_week*' mencatat hari kontak terakhir dalam seminggu, seperti 'mon', 'tue', dan lainnya. Atribut '*duration*' mencatat durasi kontak terakhir dalam detik; namun, penting untuk dicatat bahwa atribut ini sangat mempengaruhi target keluaran dan sebaiknya hanya digunakan untuk tujuan *benchmarking*.

Atribut lainnya termasuk '*campaign*' yang mencatat jumlah kontak yang dilakukan selama promosi ini, '*pdays*' yang mencatat jumlah hari yang berlalu sejak klien terakhir dihubungi dari promosi sebelumnya (999 berarti klien tidak pernah dihubungi sebelumnya), 'previous' yang mencatat jumlah kontak yang dilakukan sebelum promosi ini, dan '*poutcome*' yang mencatat hasil promosi pemasaran sebelumnya dengan kategori 'gagal', 'tidak ada', atau 'berhasil'.

Terakhir, atribut konteks sosial dan ekonomi mencakup 'emp.var.rate', 'cons.price.idx', 'cons.conf.idx', 'euribor3m', dan 'nr.employed', yang semuanya adalah indikator-indikator yang relevan dalam konteks ekonomi dan kepercayaan konsumen.

Variabel keluaran (target) dalam penelitian ini adalah atribut 'y', yang mencatat apakah klien telah berlangganan deposito berjangka dengan pilihan biner 'ya' atau 'tidak'. Seluruh dataset ini akan digunakan dalam analisis dan pemodelan untuk mengidentifikasi faktor-faktor yang memengaruhi keputusan pelanggan dalam berlangganan deposito berjangka.

Proses *preprocessing data* merupakan langkah krusial dalam pengolahan data, di mana data diolah dan diseleksi untuk mengatasi berbagai masalah seperti *noise*, *missing value*, dan data yang tidak konsisten. Tujuan utama dari *preprocessing* adalah untuk mempersiapkan data agar dapat digunakan dalam perhitungan serta memastikan keandalan dan kualitas data yang digunakan dalam analisis maupun pemodelan.

Tahapan awal dalam *preprocessing data* melibatkan proses *Exploratory Data Analysis* (EDA), di mana data dianalisis secara menyeluruh untuk memahami karakteristiknya. Selama analisis ini, berbagai fitur atau atribut dalam data diperiksa secara rinci. Hasil dari EDA memungkinkan peneliti untuk merumuskan hipotesis mengenai faktor-faktor individu (fitur) yang mungkin memengaruhi hasil yang ingin diprediksi, yaitu apakah seorang klien akan berlangganan deposito berjangka atau tidak.

Berdasarkan analisis distribusi data yang dilakukan, beberapa temuan penting dapat diidentifikasi. Pertama, terdapat temuan bahwa staf administrasi dan spesialis teknis memiliki tingkat pembukaan deposito yang lebih tinggi. Secara relatif, proporsi pensiunan dan pelajar juga dapat dianggap cukup tinggi. Temuan kedua adalah meskipun konsumen yang sudah menikah secara absolut lebih sering menyetujui layanan deposito, para lajang cenderung merespons lebih baik secara relatif.

Selain itu, hasil analisis menunjukkan bahwa saluran komunikasi terbaik untuk mempromosikan deposito adalah melalui panggilan seluler. Temuan keempat adalah adanya perbedaan yang signifikan antara konsumen yang sudah menggunakan jasa bank dan yang memiliki pinjaman dalam hal tanggapan terhadap layanan deposito. Terakhir, terlihat bahwa kepemilikan rumah (*housing*) tidak memiliki pengaruh yang signifikan terhadap kinerja perusahaan pemasaran dalam hal promosi deposito berjangka.

Berdasarkan hasil analisis matriks korelasi, beberapa temuan penting dapat diidentifikasi. Pertama, durasi panggilan (call duration) merupakan fitur yang paling berkorelasi dengan variabel target. Korelasi yang signifikan ini menunjukkan bahwa durasi panggilan yang terjadi selama interaksi dengan klien memiliki dampak yang kuat pada keputusan klien untuk berlangganan deposito berjangka. Untuk mengurangi pengaruh durasi panggilan pada model, dapat dipertimbangkan untuk mengubah atau menormalkan nilai durasi panggilan. Sebagai contoh, pengelompokan durasi panggilan menjadi kategori seperti 'pendek', 'sedang', atau 'panjang' atau penggunaan metode penskalaan lainnya dapat menjadi langkah yang relevan.

Kedua, terdapat beberapa fitur yang menunjukkan korelasi yang signifikan di antara mereka sendiri, yaitu tingkat pekerjaan (job), indeks kepercayaan konsumen (cons.conf.idx), dan indeks harga konsumen (cons.price.idx). Korelasi tinggi ini mengindikasikan bahwa ketiga fitur ini saling berkaitan erat dan memberikan wawasan tentang berbagai aspek sosial-ekonomi klien. Selain itu, korelasi yang tinggi dengan variabel target menunjukkan bahwa ketiga fitur ini bersama-sama memberikan informasi yang berharga bagi model dalam memprediksi keputusan klien. Variabilitas yang diperkenalkan oleh fitur-fitur ini dapat meningkatkan kemampuan model untuk melakukan generalisasi dengan lebih baik.

Namun, perlu dicatat bahwa dataset ini didominasi oleh variabel kategoris, dan jumlah variabel numerik yang memiliki korelasi signifikan dengan target tidak lebih dari empat. Oleh karena itu, diperlukan transformasi dan perubahan pada variabel kategoris untuk meningkatkan kemampuan model dalam menggeneralisasi data. Menghapus variabel kategoris tidak merupakan opsi yang layak dalam konteks ini.

Secara khusus, fitur durasi panggilan dan beberapa variabel kategori yang dapat diubah menjadi biner memerlukan perhatian khusus. Oleh karena itu, disarankan untuk menggunakan teknik seperti *binning* dan transformasi yang sesuai,

mungkin dengan memetakan nilai menjadi 0 dan 1, untuk menghadapinya.

Untuk variabel kategori dengan lebih dari tiga jenis kategori (seperti perkawinan dan pendidikan), pendekatan yang disarankan adalah menggunakan *encoding target*. Pendekatan ini akan memungkinkan penghubungan yang lebih akurat antara nilai-nilai kategori dengan variabel target dan memungkinkan penggunaan informasi kategoris dalam bentuk numerik dalam model.

Hasil analisis matriks korelasi memberikan pandangan yang berharga dalam persiapan data untuk pemodelan. Hal ini akan memungkinkan pengoptimalan variabel dan fitur dalam upaya untuk mengembangkan model yang lebih baik dalam memprediksi keputusan klien terkait deposito berjangka.

Selanjutnya dilakukan pengujian terhadap algoritma *Random Forest* dengan menghitung ROC_AUC (*Area Under the Receiver Operating Characteristic Curve*) dan *accuracy rate*:

```
# models that we iterate over
look_for = [ gs_rf, gs_knn]
# dict for later use
model_dict = {0:'RandomForest', 1:'Knn'}
```

```
[48] ''' Function to iterate over models and obtain results'''
     # set empty dicts and list
     result_acc = {}
     result_auc = {}
     models = []

     for index, model in enumerate(look_for):
             start = time.time()
             print()
             print('+++++++ Start New Model ++++++++++++++++++++++')
             print('Estimator is {}'.format(model_dict[index]))
             model.fit(X_train, y_train)
             print('-----------------------------------------------')
             print('best params {}'.format(model.best_params_))
             print('best score is {}'.format(model.best_score_))
             auc = roc_auc_score(y_test, model.predict_proba(X_test)[:,1])
             print('-----------------------------------------------')
             print('ROC_AUC is {} and accuracy rate is {}'.format(auc, model.score(X_test, y_test)))
             end = time.time()
             print('It lasted for {} sec'.format(round(end - start, 3)))
             print('++++++++ End Model ++++++++++++++++++++++++++')
             print()
             print()
             models.append(model.best_estimator_)
             result_acc[index] = model.best_score_
             result_auc[index] = auc
```

Gambar 2. Pengujian algoritma *Random Forest*

Kemudian, dilakukan proses evaluasi hasil yang dilakukan dengan mempertimbangkan nilai ROC_AUC dan *accuracy rate* dari algoritma *Random Forest* yang telah digunakan. Hasil yang diperoleh adalah sebagai berikut:

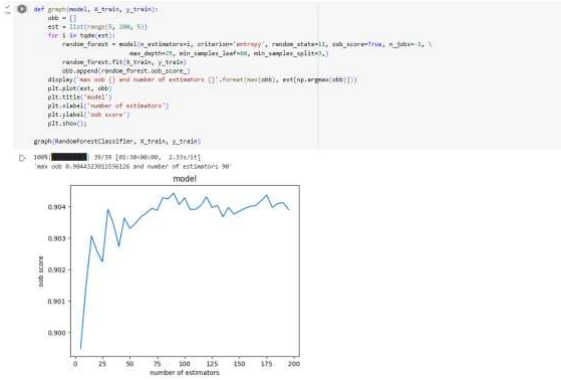
```
+++++++ Start New Model ++++++++++++++++++++
Estimator is RandomForest
-----------------------------------------------
best params {'rf__criterion': 'entropy', 'rf__max_depth': 25, 'rf__min_samples_leaf': 80, 'rf__min_samples_split': 3, 'rf__n_estimators': 70}
best score is 0.9015622707953857
-----------------------------------------------
ROC_AUC is 0.926898425815701 and accuracy rate is 0.9038251366120219
It lasted for 98.722 sec
++++++++ End Model ++++++++++++++++++++++++++


+++++++ Start New Model ++++++++++++++++++++
Estimator is Knn
-----------------------------------------------
best params {'knn__n_neighbors': 10}
best score is 0.9051364857044[illegible]
-----------------------------------------------
ROC_AUC is 0.9021804711321758 and accuracy rate is 0.9023679417122204
It lasted for 26.577 sec
++++++++ End Model ++++++++++++++++++++++++++
```

Gambar 3. Hasil Pengujian Algoritma *Random Forest*

Hasil penelitian menunjukkan bahwa algoritma *Random Forest* menghasilkan metrik ROC_AUC sebesar 0.926898425815701 dan tingkat *accuracy rate* sebesar 0.9038251366120219.

Untuk memastikan bahwa *hyper-parameter* kritis yang digunakan dalam algortima *Random Forest* telah dipilih dengan benar selama proses Pencarian *Grid*, dilakukan visualisasi dengan menggunakan grafik skor *Out-of-Bag* (OOB).

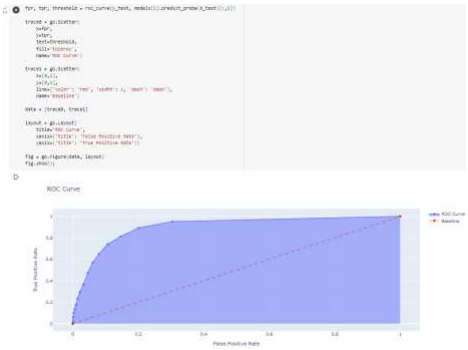


Gambar 4. Grafik skor *out-of-bag* (OOB)

Grafik ini menunjukkan bahwa terdapat kesamaan antara dua parameter, yaitu jumlah *estimator* sebanyak 80 dengan skor ROC AUC terbaik, dan jumlah estimator sebanyak 90 dengan skor OOB maksimum.

Grafik ini memberikan gambaran visual tentang kinerja algoritma Random Forest dengan variasi jumlah *estimator*. Fakta bahwa titik puncak skor ROC AUC yang sesuai dengan titik puncak skor OOB menunjukkan konsistensi dalam pemilihan hyper-parameter kritis, yaitu jumlah *estimator*. Hasil ini memberikan keyakinan bahwa jumlah estimator yang tepat telah dipilih, yang dapat menghasilkan kinerja algortima yang baik dalam hal kemampuan membedakan kelas positif dan negatif (ROC AUC) serta kemampuan untuk menggeneralisasi data yang belum pernah dilihat sebelumnya (OOB).

Hal tersebut merupakan indikasi positif bahwa algoritma *Random Forest* yang dihasilkan melalui Pencarian *Grid* memiliki kemampuan yang baik dalam mengatasi *overfitting* dan mampu melakukan generalisasi yang baik pada data yang tidak digunakan dalam pembuatan setiap pohon keputusan. Keselarasan antara skor ROC AUC terbaik dan skor OOB maksimum mengindikasikan bahwa pengoptimalan *hyper-parameter* telah dilakukan dengan baik, dan algoritm yang dihasilkan memiliki kinerja yang kuat.



Gambar 5. Kurva distribusi tingkat *False Positive Rate*

Berdasarkan analisis grafik yang telah disajikan, dapat dilihat bahwa kurva distribusi menunjukkan distribusi yang seimbang dengan kecenderungan

terhadap tingkat *False Positive Rate*. Nilai ROC AUC sebesar 0.9269 yang diperoleh dari model terbaik adalah nilai yang cukup tinggi, dan hal ini memberikan dasar yang kuat untuk mengambil asumsi lanjutan terkait dengan data ini.

3.2. Pembahasan

Berdasarkan hasil analisis yang telah dilakukan, terbukti bahwa algoritma *Random Forest* memiliki tingkat akurasi dan ROC AUC yang lebih unggul yang mana dapat mengintegrasikan algoritma *Random Forest* sebagai komponen integral dari strategi promosi dalam kampanye pemasaran bank.

Keputusan ini ditempuh mengingat *Random Forest* telah terbukti memberikan kinerja yang lebih superior dalam memprediksi dan mengklasifikasikan target pemasaran, yang pada gilirannya dapat meningkatkan efektivitas strategi promosi yang dijalankan. Dengan menerapkan algoritma *Random Forest*, bank akan memiliki kemampuan yang lebih kuat untuk mengidentifikasi calon pelanggan yang memiliki potensi tinggi untuk merespons kampanye pemasaran, berdasarkan analisis yang menyeluruh terhadap data yang relevan.

Penerapan algoritma *Random Forest* dalam strategi promosi bank akan membantu dalam mengalokasikan sumber daya dan upaya pemasaran secara lebih efisien. Hal ini akan memungkinkan fokus pada segmen pelanggan yang memiliki potensi tinggi untuk berpartisipasi dan merespons tawaran pemasaran, yang pada akhirnya dapat membawa dampak positif pada hasil akhir kampanye pemasaran. Pendekatan ini akan mengoptimalkan penggunaan sumber daya dan meningkatkan tingkat kesuksesan inisiatif pemasaran secara keseluruhan.

4. KESIMPULAN

Dalam kesimpulan penelitian ini, setelah melalui analisis mendalam terhadap algoritma *Random Forest*, dapat disimpulkan bahwa algoritma ini menunjukkan tingkat akurasi dan ROC AUC yang superior. Oleh karena itu, rekomendasi kuat diberikan untuk mengadopsi algoritma *Random Forest* sebagai pilihan utama untuk analisis pemasaran di masa yang akan datang.

Hal ini mengindikasikan bahwa algoritma *Random Forest* memiliki potensi besar untuk memberikan hasil yang lebih akurat dan handal dalam mengidentifikasi pola serta tren dalam data pemasaran. Penerapannya dapat memberikan dukungan yang sangat berharga bagi perusahaan dalam membuat keputusan yang lebih terinformasi dan efektif dalam merancang strategi pemasaran yang tepat sasaran dan efisien. Oleh karena itu, untuk meningkatkan efektivitas analisis pemasaran, disarankan dengan tegas untuk memilih dan mengimplementasikan algoritma *Random Forest* dalam konteks penelitian dan praktik pemasaran di masa depan.

5. DAFTAR PUSTAKA

[1] V. K. Gupta, A. Gupta, D. Kumar, and A. Sardana, “Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model,” Big Data Mining and Analytics, vol. 4, no. 2, pp. 116–123, Jun. 2021, doi: 10.26599/BDMA.2020.9020016.

[2] A. U. Zailani And N. L. Hanun, “Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera,” Infotech: Journal Of Technology Information, Vol. 6, No. 1, Pp. 7–14, Jun. 2020, Doi: 10.37365/Jti.V6i1.61.

[3] H. Leidiyana, “Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor,” 2013.

[4] R. Supriyadi, W. Gata, N. Maulidah, A. Fauzi, I. Komputer, and S. Nusa Mandiri Jalan Margonda Raya No, “Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah,” vol. 13, no. 2, pp. 67–75, 2020, [Online]. Available: http://journal.stekom.ac.id/index.php/E-Bisnis-page67

[5] U. Khultsum and A. Subekti, “Penerapan Algoritma Random Forest dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat,” Jurnal Media Informatika Budidarma, vol. 5, no. 1, p. 186, Jan. 2021, doi: 10.30865/mib.v5i1.2624.

[6] M. Rivki and A. M. Bachtiar, “Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia,” Jurnal Sistem Informasi, vol. 13, no. 1, p. 31, May 2017, doi: 10.21609/jsi.v13i1.500.

[7] M. Khalilia, S. Chakraborty, and M. Popescu, “Predicting disease risks from highly imbalanced data using random forest,” BMC Med Inform Decis Mak, vol. 11, no. 1, 2011, doi: 10.1186/1472-6947-11-51.

[8] O. Kherif, Y. Benmahamed, M. Teguar, A. Boubakeur, and S. S. M. Ghoneim, “Accuracy Improvement of Power Transformer Faults Diagnostic using KNN Classifier with Decision Tree Principle,” IEEE Access, 2021, doi: 10.1109/ACCESS.2021.3086135.

[9] M. Pal and S. Parija, “Prediction of Heart Diseases using Random Forest,” in Journal of Physics: Conference Series, Mar. 2021, vol. 1817, no. 1. doi: 10.1088/1742-6596/1817/1/012009.

[10] L. Makurumidze, W. S. Manjoro, and W. Makondo, “Implementing Random Forest to Predict Churn,” International Journal of Computer Science and Mobile Computing, vol. 11, no. 2, pp. 75–84, Feb. 2022, doi: 10.47760/ijcsmc.2022.v11i02.009.