Implementasi *Ensemble Stacking* Dengan Pendekatan Sampling dan *Logistic Regression* Untuk Meningkatkan Akurasi Klasifikasi Penyakit Jantung

Novandi Catur Aditya, Aprilia Baiin, Ardiyansyah, Yeni Mustika

Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika, Jl. Abdul Rahman Saleh No.18A, Bangka Belitung Laut, Kec. Pontianak Tenggara, Kota Pontianak, Kalimantan Barat, Indonesia, 78124.
Email : 19220364@bsi.ac.id

ABSTRAK

Model kategorisasi yang akurat diperlukan untuk membantu identifikasi dini penyakit jantung, yang merupakan salah satu penyebab kematian utama di dunia. Permasalahan utama dalam klasifikasi adalah ketidakseimbangan data yang dapat menurunkan kinerja model. Penelitian ini bertujuan meningkatkan akurasi klasifikasi penyakit jantung menggunakan metode *ensemble stacking* dengan pendekatan teknik sampling. Dataset yang digunakan adalah *Heart Failure Prediction Dataset* dari Kaggle sebanyak 918 data. Metode yang digunakan terdiri dari tiga *base learner*, yaitu *K-Nearest Neighbor, Naive Bayes, dan Random Forest*, serta *Logistic Regression* sebagai *meta-learner*. Penanganan data tidak seimbang dilakukan dengan empat skenario, yaitu tanpa sampling, *oversampling, undersampling, dan SMOTE*. Evaluasi dilakukan menggunakan *10-fold cross validation*. Hasil penelitian menunjukkan bahwa metode stacking memberikan performa terbaik dibandingkan model tunggal. Akurasi tertinggi diperoleh pada skenario *oversampling* sebesar 95,57% dengan nilai *precision, recall, dan F1-score* sebesar 0,956. Metode ini efektif meningkatkan kinerja klasifikasi penyakit jantung

Kata Kunci : Ensemble Stacking, Teknik Sampling, Klasifikasi Penyakit Jantung, Logistic Regression, Data Tidak Seimbang

ABSTRACT

An accurate categorisation model is necessary to assist early identification of heart disease, which is one of the top causes of death globally. A major challenge in classification is data imbalance, which can reduce model performance. This study aims to improve heart disease classification accuracy using an ensemble stacking method combined with sampling techniques. The dataset used is the Heart Failure Prediction Dataset from Kaggle, consisting of 918 instances. The proposed method employs three base learners, namely K-Nearest Neighbor, Naive Bayes, and Random Forest, with Logistic Regression as the meta-learner. To address class imbalance, four scenarios are applied: without sampling, oversampling, undersampling, and SMOTE. Model evaluation is conducted using 10-fold cross validation. The results show that the stacking method outperforms individual models across all scenarios. The best performance is achieved using oversampling, with an accuracy of 95.57% and precision, recall, and F1-score of 0.956. This approach effectively improves heart disease classification performance.

Keywords : *Ensemble Stacking, Sampling Techniques, Heart Disease Classification, Logistic Regression, Imbalanced Data*

1. PENDAHULUAN

Salah satu penyebab kematian utama di seluruh dunia adalah penyakit jantung. Statistik Organisasi Kesehatan Dunia menunjukkan bahwa penyakit kardiovaskular bertanggung jawab atas lebih dari 19,8 juta kematian, atau sekitar 32% dari seluruh kematian di dunia. Hal ini menjadikannya penyebab utama kematian di seluruh dunia. Bidang kedokteran telah banyak menggunakan teknik pembelajaran mesin untuk membantu prosedur klasifikasi karena kemajuan teknologi. Dengan bantuan pembelajaran mesin, sistem dapat menemukan tren dalam data dan membuat prediksi untuk membantu pengambilan keputusan.

Berbagai algoritma klasifikasi telah digunakan dalam penelitian sebelumnya, di antaranya *K-Nearest Neighbor (KNN), Random Forest, dan Naive Bayes*(Setiawan, Yasin, and Desianti 2025). KNN bekerja berdasarkan kedekatan jarak antar data untuk menentukan kelas suatu objek, namun sensitif terhadap skala data dan pemilihan parameter *k*. *Random Forest* menggunakan pendekatan ensemble berbasis pohon keputusan yang mampu meningkatkan akurasi dan mengurangi *overfitting*, meskipun membutuhkan komputasi yang lebih tinggi dan kurang interpretatif. Sementara itu, *Naive Bayes* merupakan metode probabilistik yang cepat dan efisien, tetapi memiliki kelemahan pada asumsi independensi antar fitur yang tidak selalu sesuai dengan kondisi data nyata. Meskipun cukup efektif, masing-masing algoritma memiliki keterbatasan dalam menangani kompleksitas data, terutama pada data yang bersifat non-linear dan tidak seimbang, sehingga diperlukan pendekatan lain untuk meningkatkan kinerja klasifikasi(Airi, Suprapti, and Bahtiar 2023). Salah satu pendekatan yang dapat digunakan untuk meningkatkan performa model adalah metode ensemble learning, khususnya stacking. Metode stacking memungkinkan penggabungan beberapa algoritma sebagai base learner untuk menghasilkan prediksi yang lebih akurat melalui model tingkat kedua (*meta-learner*)(Faska et al. 2023).

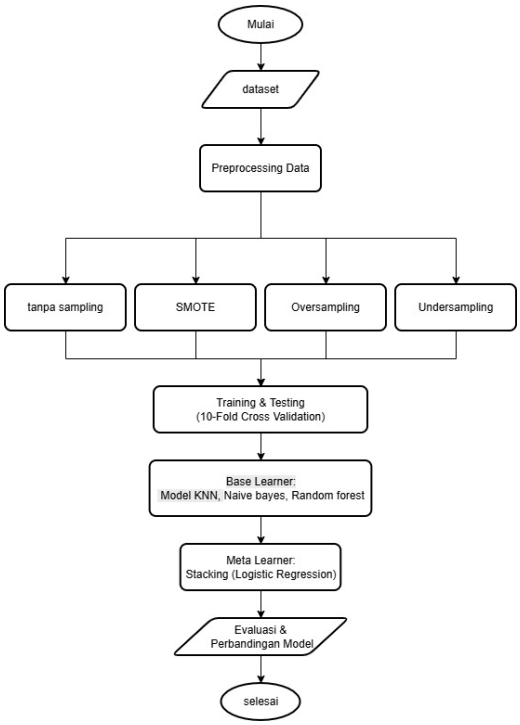
Dalam penelitian ini, digunakan *Logistic Regression* sebagai meta-learner pada metode stacking untuk mengkombinasikan hasil prediksi dari model dasar (*base learner*) yaitu *K-Nearest Neighbor (KNN), Random Forest, dan Naive Bayes. Logistic Regression* berperan sebagai model tingkat kedua yang menerima output prediksi dari masing-masing base learner sebagai input, kemudian mempelajari pola kombinasi terbaik dari prediksi tersebut untuk menghasilkan keputusan akhir yang lebih akurat.

Selain itu, basis data yang berkaitan dengan penyakit kardiovaskular sering menunjukkan ketidakseimbangan statistik, di mana satu kategori pasien terwakili secara berlebihan(Rahim, Pratiwi, and Fikri 2023). Kapasitas model untuk mengidentifikasi kelas minoritas dapat berkurang sebagai akibat dari bias terhadap kelas mayoritas ini. Beberapa metode pengambilan sampel, termasuk *oversampling, undersampling, dan Synthetic Minority Oversampling Technique (SMOTE)*, dapat digunakan untuk mengatasi masalah ini. Terdapat manfaat dan kekurangan pada setiap metode untuk menangani data yang terdistribusi tidak merata (Wongvorachan, He, and Bulut 2023).

Penelitian ini bertujuan untuk meningkatkan kinerja klasifikasi penyakit jantung dengan mengatasi permasalahan ketidakseimbangan data melalui penerapan metode ensemble stacking. Penelitian ini membandingkan beberapa teknik sampling, yaitu tanpa sampling, *oversampling, undersampling,* dan *SMOTE*, untuk menganalisis pengaruhnya terhadap peningkatan akurasi model. Hasil penelitian diharapkan dapat menunjukkan kombinasi metode yang paling efektif dalam menghasilkan model klasifikasi yang lebih akurat dan andal.

2. METODE

Pemrosesan data numerik dan pengukuran kinerja model menggunakan berbagai metrik penilaian merupakan fokus utama penelitian kuantitatif ini yang menggunakan pendekatan eksperimental. Strategi eksperimental digunakan untuk menguji dan membandingkan beberapa model kategorisasi dalam berbagai pengaturan pengobatan, sedangkan pendekatan kuantitatif dipilih karena hasil yang objektif dan terukur(Ara et al. 2026). Penelitian ini bertujuan untuk menganalisis pengaruh penerapan teknik sampling dalam mengatasi ketidakseimbangan data terhadap kinerja metode ensemble stacking dalam meningkatkan akurasi klasifikasi penyakit jantung. Untuk mencapai tujuan tersebut, dilakukan serangkaian percobaan dengan mengombinasikan berbagai metode sampling dan algoritma klasifikasi guna memperoleh model dengan performa yang optimal dan hasil yang dapat dipertanggungjawabkan (Sijabat, Aji, and Prabowo 2026).



Gambar 1. Diagram Alir Penelitian

Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari Kaggle, yaitu dataset *Heart Failure Prediction* https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction. Analisis penyakit jantung menggunakan kumpulan data ini, yang mencakup fitur-fitur yang relevan. Sebanyak 918 titik data yang mewakili berbagai karakteristik medis pasien digunakan dalam penelitian ini. Faktor-faktor yang berkaitan dengan kesehatan kardiovaskular dicatat, seperti usia, jenis kelamin, sifat ketidaknyamanan dada, tekanan darah, kolesterol, glukosa, dan temuan elektrokardiografi.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
2	40	M	ATA	140	289	0	Normal	172	N	0	Up	NO
3	49	F	NAP	160	180	0	Normal	156	N	1	Flat	YES
4	37	M	ATA	130	283	0	ST	98	N	0	Up	NO
5	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	YES
6	54	M	NAP	150	195	0	Normal	122	N	0	Up	NO
7	39	M	NAP	120	339	0	Normal	170	N	0	Up	NO
8	45	F	ATA	130	237	0	Normal	170	N	0	Up	NO
9	54	M	ATA	110	208	0	Normal	142	N	0	Up	NO
10	37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	YES
11	48	F	ATA	120	284	0	Normal	120	N	0	Up	NO
12	37	F	NAP	130	211	0	Normal	142	N	0	Up	NO
13	58	M	ATA	136	164	0	ST	99	Y	2	Flat	YES
14	39	M	ATA	120	204	0	Normal	145	N	0	Up	NO
15	49	M	ASY	140	234	0	Normal	140	Y	1	Flat	YES
16	42	F	NAP	115	211	0	ST	137	N	0	Up	NO
17	54	F	ATA	120	273	0	Normal	150	N	1.5	Flat	NO
18	38	M	ASY	110	196	0	Normal	166	N	0	Flat	YES
19	43	F	ATA	120	201	0	Normal	165	N	0	Up	NO
20	60	M	ASY	100	248	0	Normal	125	N	1	Flat	YES
21	36	M	ATA	120	267	0	Normal	160	N	3	Flat	YES
22	43	F	TA	100	223	0	Normal	142	N	0	Up	NO
23	44	M	ATA	120	184	0	Normal	142	N	1	Flat	NO
24	49	F	ATA	124	201	0	Normal	164	N	0	Up	NO

Gambar 2. Atribut Dataset Yang Digunakan Dalam Penelitian

Preprocessing Data

Tujuan dari langkah pra-perlakuan data dalam penelitian ini adalah untuk menjamin data berkualitas tinggi untuk fase pemodelan. Untuk menemukan data yang mungkin tidak relevan, algoritma ini memeriksa nilai yang hilang, data duplikat, dan nilai unik untuk setiap karakteristik. Hasil pemeriksaan menunjukkan bahwa dataset tidak mengandung nilai yang hilang maupun data yang terduplikasi, serta tidak ditemukan nilai unik yang tidak relevan, sehingga tidak diperlukan proses pembersihan data lanjutan.

Selain itu, dilakukan penyesuaian pada atribut kelas (target) agar sesuai dengan kebutuhan proses klasifikasi. Nilai kelas yang semula berbentuk numerik, yaitu 0 dan 1, diubah menjadi bentuk kategorikal, yaitu *no* dan *yes*. Proses ini bertujuan untuk memastikan data dapat digunakan secara optimal dalam proses pemodelan klasifikasi.

Penanganan Ketidakseimbangan Data

Ketika terdapat kelebihan data dari satu kategori dibandingkan kategori lainnya, mengatakan bahwa data tersebut tidak seimbang. Karena masalah ini, model klasifikasi dapat menjadi condong ke arah kelas dominan dan menjadi kurang efektif dalam mengidentifikasi kelas minoritas. Akibatnya, agar model dapat menghasilkan kinerja yang lebih seimbang, diperlukan metode khusus untuk menangani ketidakseimbangan data(Salmi et al. 2024).

1. Tanpa Sampling

Pada skenario ini, dataset digunakan dalam kondisi asli tanpa dilakukan perubahan atau penyesuaian terhadap distribusi kelas. Dengan demikian, proporsi antara kelas mayoritas dan kelas minoritas tetap seperti pada data awal. Skenario ini berfungsi sebagai pembanding dasar (*baseline*) untuk mengevaluasi kinerja model sebelum diterapkan teknik penanganan ketidakseimbangan data. Hasil dari skenario ini digunakan sebagai acuan untuk menilai sejauh mana metode sampling yang diterapkan mampu meningkatkan performa klasifikasi.

2. *SMOTE*

SMOTE (*Synthetic Minority Over-sampling Technique*) Ketika terjadi ketidakseimbangan data antara kelas mayoritas dan minoritas, SMOTE adalah teknik pemrosesan data pembelajaran mesin yang dapat memperbaikinya. Alih-alih hanya menyalin data seperti yang dilakukan pendekatan oversampling tradisional, strategi ini

menciptakan data sintetis baru untuk kelas minoritas berdasarkan seberapa dekat data saat ini(Syukron et al. 2023). Dengan menggunakan metode ini, SMOTE mampu meningkatkan kemampuan model untuk mempelajari pola representatif dengan meningkatkan variasi data dalam kelas minoritas. Dipercaya bahwa strategi ini, ketika diterapkan, akan meningkatkan kemampuan klasifikasi, terutama dalam mengenali data dari kelas minoritas, dan mengurangi bias model terhadap kelas mayoritas(Song and Yang 2026).

3. *Oversampling*

Salah satu teknik modifikasi data yang membantu dalam pekerjaan klasifikasi ketika terdapat ketidakseimbangan antar kelas adalah oversampling(Sapriadi and Nasution 2024). Teknik ini bekerja dengan cara menghasilkan atau menambah sampel baru (data sintetis) pada kelompok minoritas agar jumlahnya setara dengan kelompok mayoritas(Taskiran et al. 2025). Dalam penelitian ini, teknik oversampling digunakan untuk membuat distribusi data lebih seimbang dengan meningkatkan representasi kelas minoritas. Dengan kondisi data yang lebih proporsional, model diharapkan mampu mempelajari pola klasifikasi dengan lebih baik dan mengurangi kecenderungan bias terhadap kelas mayoritas.

4. *Undersampling*

Undersampling atau mengurangi jumlah data dalam kelas mayoritas untuk membuat distribusi antar kelas lebih seimbang adalah salah satu pendekatan untuk menangani ketidakseimbangan data(Qadrini 2023). Pendekatan ini bertujuan untuk mengurangi kekuatan kelas dominan, yang dapat membuat model menjadi bias dan kurang mahir dalam melihat tren di dalam kelas minoritas. Dengan mengurangi sebagian data mayoritas, model diharapkan dapat lebih fokus dalam mempelajari karakteristik dari kedua kelas secara lebih proporsional(Sir and Soepranoto 2022). Meskipun demikian, teknik undersampling memiliki kelemahan yaitu berpotensi menghilangkan informasi penting dari data mayoritas, sehingga dapat mempengaruhi performa model jika tidak dilakukan dengan tepat(Zhou et al. 2024). Oleh karena itu, penerapan undersampling dalam penelitian ini digunakan sebagai salah satu skenario untuk dibandingkan dengan teknik lainnya dalam meningkatkan kinerja klasifikasi.

Pembangunan Model Klasifikasi

Pada tahap ini dilakukan pembangunan model klasifikasi untuk mendeteksi penyakit jantung menggunakan pendekatan *ensemble learning* dengan metode *stacking*. Kemampuan untuk mengintegrasikan banyak metode kategorisasi ke dalam satu model menyebabkan model ini terpilih sebagai pendekatan yang lebih disukai. Dalam penelitian ini, proses pembangunan model dilakukan dalam dua tingkat, yaitu *base learner* dan *meta-learner*.

1. Base-Learner

Pada tahap *base learner* digunakan tiga algoritma klasifikasi, yaitu *K-Nearest Neighbor (KNN), Naive Bayes,* dan *Random Forest*. Ketiga algoritma tersebut dipilih karena memiliki pendekatan klasifikasi yang berbeda sehingga dapat saling melengkapi dalam mempelajari pola data.

a. . K-Nearest Neighbor (KNN)

Algoritma *KNN* melakukan klasifikasi berdasarkan kedekatan jarak antar data. Data baru akan diklasifikasikan berdasarkan mayoritas tetangga terdekat(Sejati et al. 2022), didefinisikan sebagai berikut (1):

$$dist(x, y) = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2} \quad \textbf{(1)}$$

Nilai *n* menunjukkan jumlah variabel yang digunakan dalam perhitungan. Variabel *xi* dan *yi* merupakan elemen dari vektor *x* dan *y* yang berada dalam suatu ruang vektor. Vektor tersebut dapat dinyatakan sebagai *x = (x1,x2,x3,…,xn) dan y = (y1,y2,y3,…,yn).*

b. Naïve Bayes

Teknik klasifikasi berbasis probabilitas, Naive Bayes, menemukan kemungkinan data termasuk ke dalam kelas tertentu menggunakan Teorema Bayes(Harun and Fahmi 2024). Untuk menggunakan teknik Naive Bayes, seseorang harus terlebih dahulu menentukan kemungkinan setiap fitur data yang telah ditentukan sebelumnya. Kemudian, menggunakan label yang telah ditentukan sebagai dasar untuk prediksi data, nilai kemungkinan dihitung dengan mengalikan semua probabilitas ini(Sejati et al. 2022). Pada persamaan (2) rumus untuk menghitung nilai probalitas:

$$p(c|x) = \frac{p(x|c)}{p(x)} p(c) \quad \textbf{(2)}$$

Pada persamaan (2), *P(c/x)* menunjukkan probabilitas posterior suatu kelas atau target setelah mempertimbangkan informasi dari atribut prediktor. *P(c)* merupakan probabilitas awal (*prior probability*) dari kelas sebelum adanya pengamatan terhadap data. Selanjutnya, *P(x/c)* menggambarkan kemungkinan munculnya atribut prediktor berdasarkan kelas tertentu. Adapun *P(x)* adalah probabilitas awal dari atribut atau data prediktor yang digunakan dalam proses klasifikasi.

c. Random forest

Random Forest beroperasi dengan membentuk sejumlah besar pohon keputusan (*decision tree*). Pada setiap simpul, kualitas pembagian data dievaluasi menggunakan ukuran *Gini Impurity*(Hendrawinata, Jasmir, and Gunardi 2026), sebagaimana dinyatakan dalam Persamaan (3):

$$Gini = 1 - \sum_{i=1}^{c} (pi)^2 \quad \textbf{(3)}$$

Pi digunakan untuk menunjukkan tingkat dominasi suatu kelas pada simpul yang sedang dihitung.

2. Meta-Learner

Logistic Regression merupakan salah satu algoritma dalam *supervised learning*. Pada metode *supervised learning*, model dilatih menggunakan dataset yang telah memiliki label, di mana setiap data latih terdiri dari fitur sebagai input dan label sebagai target output. Regresi logistik digunakan sebagai metode statistik untuk menyelesaikan permasalahan klasifikasi biner, yaitu ketika tujuan penelitian adalah memprediksi probabilitas dari dua kemungkinan kelas, seperti “ya” atau “tidak”, serta “positif” atau “negatif”(Kumar et al. 2024).

Model ini berfungsi untuk menggabungkan hasil prediksi dari seluruh *base learner* sehingga menghasilkan keputusan akhir yang lebih optimal. Proses stacking dilakukan dengan menjadikan output prediksi dari *KNN, Naive Bayes*, dan *Random Forest* sebagai input bagi *Logistic Regression*.

Validasi Model

Penerapan prosedur validasi silang 10-fold membantu mengurangi kemungkinan overfitting dan menjamin hasil evaluasi yang independen. Dataset dibagi menjadi sepuluh bagian yang sama; satu bagian berfungsi sebagai data uji, sedangkan bagian lainnya berfungsi sebagai data pelatihan. Proses ini dilakukan sebanyak 10 iterasi, dan hasil evaluasi dirata-ratakan untuk mendapatkan performa model yang stabil(Putri and Suparwito 2023).

Evaluasi Kinerja Model

Kinerja model dievaluasi menggunakan beberapa metrik yang umum digunakan dalam klasifikasi, yaitu *accuracy, precision, recall,* dan *F1-score*. Metrik-metrik ini dihitung berdasarkan *confusion matrix* yang terdiri dari nilai *true positive, true negative, false positive,* dan *false negative*(Rainio, Teuho, and Klén 2024).

1. *Accuracy* : Digunakan untuk mengevaluasi akurasi klasifikasi keseluruhan model, yang diukur berdasarkan persentase titik data yang diprediksi dengan benar(Sujon et al. 2025).
2. *Precision* : Persentase data positif yang benar-benar positif merupakan ukuran akurasi model dalam memprediksi kelas positif(Sujon et al. 2025).
3. *Reccal* : Memeriksa seberapa baik model dapat mengidentifikasi semua data positif, atau lebih spesifiknya, berapa banyak titik data positif yang berhasil ditebak dengan benar oleh model(Sujon et al. 2025).
4. *F1*-score : Mengambil rata-rata harmonik dari recall dan precision adalah cara yang baik untuk mengukur seberapa seimbang model tersebut, yang sangat berguna saat berurusan dengan data yang tidak seimbang(Sujon et al. 2025).

3. HASIL DAN PEMBAHASAN

Pada bagian ini, hasil pengujian model klasifikasi penyakit jantung disajikan berdasarkan beberapa skenario pengujian, yaitu tanpa sampling, *oversampling, undersampling*, dan *Synthetic Minority Over-sampling Technique* (*SMOTE*). Pengujian dilakukan menggunakan metode *ensemble stacking* dengan *Logistic Regression* sebagai *meta-learner* serta *K-Nearest Neighbo*r (*KNN*), *Naive Bayes*, dan *Random Forest* sebagai *base learner*. Evaluasi model dilakukan menggunakan metode *10-fold cross validation* dengan metrik *accuracy, precision, recall, dan F1-score*.

Hasil Pengujian Tanpa Sampling

Pada skenario ini, dataset digunakan dalam kondisi asli tanpa dilakukan penyesuaian distribusi kelas. Hasil pengujian model tanpa sampling dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pengujian Model Tanpa Sampling

Model	Akurasi	Precesion	Recall	F1-score
KNN	82,89%	0,829	0,829	0,829
Naïve Bayes	86,05%	0,860	0,861	0,860
Random Forest	86,05%	0,861	0,861	0,860
Stacking (Logistic)	87,79%	0,878	0,878	0,878

Berdasarkan tabel 1, hasil pengujian, diperoleh nilai akurasi *KNN* sebesar 82,89%, *Naive Bayes* sebesar 86,05%, *Random Forest* sebesar 86,05%, dan *Stacking Logistic Regression* sebesar 87,79%. Hasil tersebut menunjukkan bahwa metode *stacking* memiliki performa terbaik dibandingkan metode lainnya pada kondisi dataset asli tanpa penyeimbangan data.

Metode *stacking* mampu meningkatkan performa klasifikasi karena menggabungkan keunggulan beberapa *base learner* sehingga menghasilkan prediksi yang lebih stabil dan akurat. Sementara itu, algoritma *KNN* memperoleh akurasi paling rendah karena sensitif terhadap distribusi data yang tidak seimbang.

Selain nilai akurasi, hasil *weighted average precision, recall, dan f-measure* juga menunjukkan bahwa metode stacking memperoleh nilai tertinggi yaitu sebesar 0,878. Hal ini menandakan bahwa model *stacking* mampu memberikan keseimbangan yang baik dalam proses klasifikasi data penyakit jantung.

Hasil Pengujian Menggunakan *SMOTE*

Untuk mengatasi ketidakseimbangan data, pendekatan SMOTE digunakan untuk menghasilkan data sintetis pada kelas minoritas dalam skenario ini. Hasil pengujian dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pengujian Model Menggunakan *SMOTE*

Model	Akurasi	Precesion	Recall	F1-score
KNN	83,85%	0,839	0,839	0,839
Naïve Bayes	86,31%	0,863	0,863	0,863
Random Forest	88,09%	0,881	0,881	0,881
Stacking (Logistic)	88,38%	0,884	0,884	0,884

Berdasarkan Tabel 2, Hasil pengujian menunjukkan bahwa penerapan *SMOTE* mampu meningkatkan performa beberapa algoritma. Nilai akurasi *KNN* meningkat menjadi 83,85%, *Naive Bayes* menjadi 86,31%, *Random Forest* menjadi 88,09%, dan *Stacking Logistic Regression* menjadi 88,38%.

Berdasarkan hasil tersebut, metode stacking kembali memperoleh nilai akurasi tertinggi dibandingkan algoritma lainnya. Peningkatan akurasi setelah penerapan *SMOTE* menunjukkan bahwa teknik ini efektif dalam mengurangi pengaruh ketidakseimbangan data pada proses klasifikasi.

Random Forest juga mengalami peningkatan performa yang cukup signifikan karena algoritma ini mampu memanfaatkan distribusi data baru hasil *SMOTE* dengan lebih baik. Selain itu, nilai *weighted average precision, recall, dan f-measure* metode *stacking* mencapai 0,884 yang menunjukkan kualitas klasifikasi yang semakin baik dibandingkan pengujian tanpa sampling.

Hasil Pengujian Menggunakan *Oversampling*

Pada skenario ini, dilakukan penambahan jumlah data pada kelas minoritas untuk menyeimbangkan distribusi data. Hasil pengujian dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian Model Menggunakan Oversampling

Model	Akurasi	Precesion	Recall	F1-score
KNN	93,22%	0,933	0,932	0,932
Naïve Bayes	85,75%	0,858	0,858	0,858
Random Forest	94,89%	0,949	0,949	0,949
Stacking (Logistic)	95,57%	0,956	0,956	0,956

Berdasarkan Tabel 3, hasil pengujian menunjukkan peningkatan akurasi yang sangat signifikan pada beberapa algoritma. *KNN* memperoleh akurasi sebesar 93,22%, *Naive Bayes* sebesar 85,75%, *Random Forest* sebesar 94,89%, dan *Stacking Logistic Regression* sebesar 95,57%.

Metode *stacking* dengan *Logistic Regression* berhasil memperoleh performa terbaik dengan akurasi sebesar 95,57%. Hasil ini menunjukkan bahwa kombinasi *ensemble stacking* dan teknik *oversampling* sangat efektif dalam meningkatkan kualitas klasifikasi penyakit jantung.

Random Forest juga menunjukkan performa yang sangat tinggi karena algoritma berbasis *ensemble* ini mampu memanfaatkan data tambahan hasil *oversampling* secara optimal. Sementara itu, *Naive Bayes* mengalami sedikit penurunan dibanding pengujian *SMOTE,* yang menunjukkan bahwa algoritma probabilistik kurang optimal dalam menangani penambahan data duplikasi.

Nilai *weighted average precision, recall, dan f-measure stacking* sebesar 0,956 memperlihatkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dan stabil pada seluruh kelas data.

Hasil Pengujian Menggunakan *Undersampling*

Pada skenario ini, dilakukan pengurangan jumlah data pada kelas mayoritas untuk menghasilkan distribusi data yang lebih seimbang. Hasil pengujian dapat dilihat pada Tabel 4.

Table 4. Hasil Pengujian Model Menggunakan Undersampling

Model	Akurasi	Precesion	Recall	F1-score
KNN	81,95%	0,820	0,820	0,820
Naïve Bayes	85,48%	0,855	0,855	0,855
Random Forest	85,97%	0,861	0,860	0,860
Stacking (Logistic)	87,80%	0,878	0,878	0,878

Berdasarkan Tabel 4, hasil pengujian diperoleh nilai akurasi *KNN* sebesar 81,95%, *Naive Bayes* sebesar 85,48%, *Random Forest* sebesar 85,97%, dan *Stacking Logistic Regression* sebesar 87,80%.

Hasil tersebut menunjukkan bahwa metode *stacking* tetap memperoleh performa terbaik dibandingkan metode lainnya. Akan tetapi, teknik *undersampling* tidak memberikan peningkatan akurasi yang signifikan dibandingkan metode *sampling* lainnya.

Penurunan performa pada beberapa algoritma disebabkan oleh berkurangnya jumlah data mayoritas sehingga informasi penting dalam dataset ikut hilang. Kondisi ini mempengaruhi kemampuan model dalam mempelajari pola data secara menyeluruh.

Nilai *weighted average precision, recall, dan f-measure* metode stacking sebesar 0,878 menunjukkan bahwa performa model masih cukup baik meskipun tidak seoptimal teknik *oversampling* maupun *SMOTE.*

Hasil Pengujian Perbandingan Seluruh Metode

Untuk melihat perbandingan kinerja model secara keseluruhan, hasil pengujian pada setiap skenario ditampilkan dalam bentuk tabel dan diagram. Tabel digunakan untuk menyajikan nilai evaluasi secara rinci, sedangkan diagram digunakan untuk memperjelas perbandingan performa antar metode.

Tabel 5. Perbandingan Kinerja Model *Stacking* pada Setiap Teknik *Sampling*

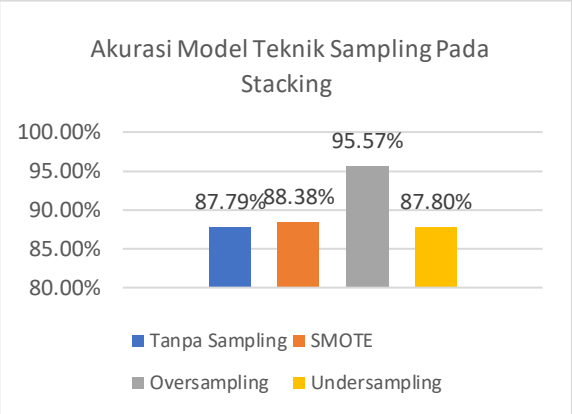
Model	Akurasi
Tanpa Sampling	87,79%
SMOTE	88,38%
Oversampling	95,57%
Undersampling	87,80%

Berdasarkan Tabel 5, dapat diketahui bahwa teknik *oversampling* memberikan hasil akurasi terbaik pada hampir seluruh algoritma, terutama pada metode *Stacking Logistic Regression* yang mencapai akurasi tertinggi sebesar 95,57%.

Metode *stacking* secara konsisten menjadi algoritma dengan performa terbaik pada seluruh skenario pengujian. Hal ini membuktikan bahwa pendekatan *ensemble stacking* mampu meningkatkan kemampuan klasifikasi melalui kombinasi beberapa *base learner*.

Selain itu, teknik *SMOTE* juga memberikan peningkatan performa yang cukup baik dibandingkan tanpa *sampling* dan *undersampling*. Sedangkan *undersampling* cenderung menghasilkan performa yang lebih rendah karena proses pengurangan data dapat menyebabkan hilangnya informasi penting pada dataset.

Berdasarkan hasil penelitian, dapat disimpulkan bahwa kombinasi teknik *oversampling* dengan metode *Ensemble Stacking* menggunakan *meta-learner Logistic Regression* merupakan pendekatan terbaik dalam klasifikasi penyakit jantung pada penelitian ini.



Gambar 3. Diagram Perbandingan Akurasi Model *stacking* Setiap Teknik *Sampling*

Berdasarkan Diagram tersebut akurasi metode *Ensemble Stacking Logistic Regression* menunjukkan perbandingan performa model pada setiap teknik *sampling* yang digunakan dalam penelitian, yaitu tanpa *sampling, SMOTE, oversampling, dan undersampling.*

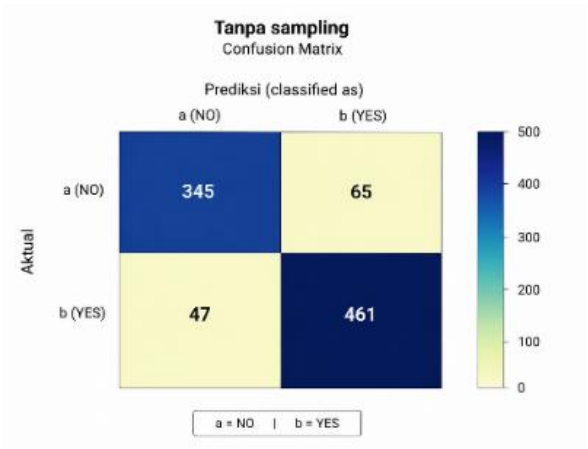
Metode *stacking* tanpa sampling memperoleh nilai akurasi sebesar 87,79%. Temuan ini menunjukkan bahwa strategi penumpukan dapat mencapai kinerja klasifikasi yang layak meskipun dataset tersebut memiliki ketidakseimbangan yang melekat.

Nilai akurasi meningkat menjadi 88,38% saat pengujian dengan pendekatan SMOTE. Alasan di balik peningkatan ini adalah SMOTE dapat menghasilkan data sintetis untuk kelas minoritas. Hal ini menyebabkan distribusi data yang lebih seimbang dan memungkinkan model untuk mempelajari pola data dengan lebih tepat.

Hasil akurasi tertinggi diperoleh pada pengujian menggunakan *oversampling* dengan nilai sebesar 95,57%. Diagram *chart* menunjukkan peningkatan yang sangat signifikan dibandingkan teknik lainnya. Hal ini membuktikan bahwa teknik *oversampling* sangat efektif dalam meningkatkan performa metode stacking karena mampu menambah jumlah data minoritas tanpa menghilangkan informasi penting pada dataset.

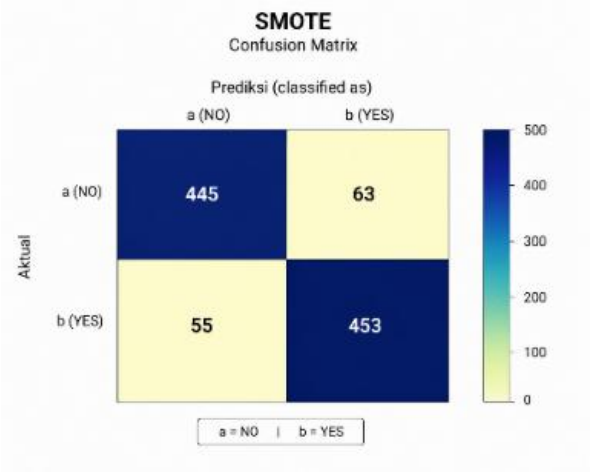
Sementara itu, pada pengujian menggunakan *undersampling*, metode *stacking* memperoleh nilai akurasi sebesar 87,80%. Nilai tersebut hampir sama dengan pengujian tanpa *sampling*. Hal ini menunjukkan bahwa pengurangan data pada kelas mayoritas dalam teknik *undersampling* menyebabkan sebagian informasi penting hilang sehingga peningkatan performa model tidak terlalu signifikan.

Confusion Matrix Stacking Pada Teknik Sampling



Gambar 4. *Confusion Matrix Stacking* Tanpa *Sampling*

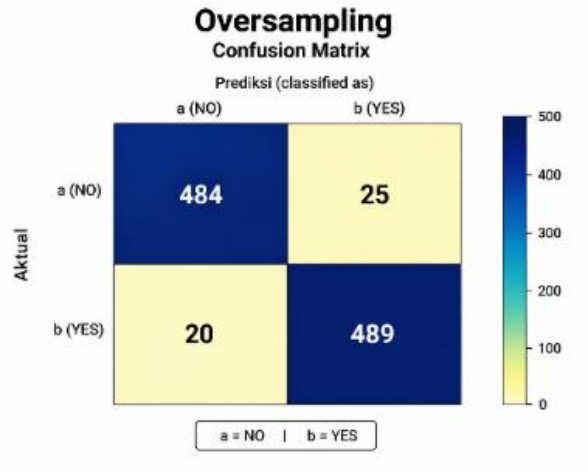
Berdasarkan gambar 4, skor TP sebesar 461 dan skor TN sebesar 345 dihasilkan oleh model ketika teknik tanpa pengambilan sampel digunakan. Hasil ini menunjukkan bahwa model berkinerja sangat baik dalam mengidentifikasi kelas YA. Meskipun demikian, masih ada kesalahan dalam klasifikasi, dengan 65 FP dan 47 FN. Temuan ini menunjukkan bahwa model masih belum mampu membedakan secara optimal antara kedua kelompok karena ketidakseimbangan data.



Gambar 5. *Confusion Matrix Stacking SMOTE*

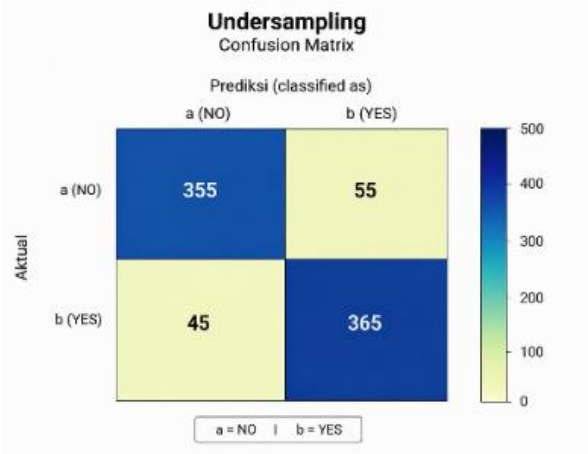
Berdasarkan gambar 5, skor TN meningkat menjadi 445 setelah menerapkan pendekatan SMOTE, yang meningkatkan kemampuan pengenalan kelas TIDAK pada model. Metode SMOTE mencapai tujuannya untuk distribusi data yang lebih merata dengan menghasilkan data sintetis untuk kelas minoritas. Model mempertahankan kinerja yang lebih konsisten meskipun terjadi peningkatan skor False Negative menjadi 55, berkat prediksi benar yang cukup tinggi dan terdistribusi secara merata untuk kedua kelas. Ini mem-

buktikan bahwa dibandingkan dengan teknik tanpa pengambilan sampel, *SMOTE* membantu model dalam mempelajari pola dalam data minoritas.



Gambar 6. *Confusion Matrix Stacking Oversampling*

Berdasrkan gambar 6, dari semua metode yang diuji, oversampling menghasilkan kinerja tertinggi. Dengan mempertimbangkan nilai False Negative yang rendah yaitu 20 dan nilai False Positive yang rendah yaitu 25, hal ini terbukti. Lebih lanjut, nilai True Negative model sebesar 484 dan True Positive sebesar 489 menunjukkan kemampuannya yang sangat baik dalam mengkategorikan kedua kelompok tersebut. Teknik *oversampling* meningkatkan jumlah data minoritas dengan menggandakan data yang ada sehingga informasi penting pada dataset tetap dipertahankan. Kondisi tersebut membuat model lebih optimal dalam mengenali pola data dan mengurangi tingkat kesalahan klasifikasi.



Gambar 7. *Confusion Matrix Stacking Undersampling*

Berdasarkan gambar 7, teknik *undersampling* menghasilkan performa yang relatif lebih rendah. Nilai *True Positive* hanya mencapai 365 dengan *True Negative* sebesar 355. Hal ini disebabkan undersampling mengurangi jumlah data pada kelas mayoritas sehingga sebagian informasi penting dalam dataset hilang. Akibatnya, kemampuan model dalam mengenali pola data menjadi menurun dan berdampak pada penurunan performa klasifikasi secara keseluruhan.

4. SIMPULAN

Berdasarkan hasil penelitian mengenai implementasi metode *Ensemble Stacking* dengan pendekatan teknik *sampling* dan *meta-learner Logistic Regression* pada klasifikasi penyakit jantung, diperoleh kesimpulan bahwa metode *Ensemble Stacking* mampu memberikan performa klasifikasi yang lebih baik dibandingkan algoritma *KNN, Naive Bayes,* dan *Random Forest* pada seluruh skenario pengujian.

Penerapan pendekatan pengambilan sampel meningkatkan kinerja klasifikasi, menurut hasil penelitian, khususnya pada dataset dengan ketidakseimbangan kelas. Pada pengujian tanpa *sampling*, metode *Ensemble Stacking* memperoleh akurasi sebesar 87,79%. Setelah diterapkan teknik *SMOTE*, akurasi meningkat menjadi 88,38%. Sementara itu, teknik *undersampling* menghasilkan akurasi sebesar 87,80%, yang menunjukkan bahwa pengurangan data mayoritas belum mampu memberikan peningkatan performa yang signifikan.

Pengujian terbaik diperoleh pada penerapan teknik *oversampling*, dimana metode *Ensemble Stacking* dengan *Logistic Regression* berhasil mencapai akurasi sebesar 95,57%. Hasil tersebut membuktikan bahwa teknik *oversampling* mampu meningkatkan kemampuan model dalam mengenali pola data pada kelas minoritas tanpa menghilangkan informasi penting pada dataset.

Berdasarkan keseluruhan hasil pengujian, dapat disimpulkan bahwa kombinasi teknik *oversampling* dan metode *Ensemble Stacking* dengan *meta-learner Logistic Regression* merupakan pendekatan terbaik dalam penelitian ini untuk klasifikasi penyakit jantung. Pendekatan tersebut terbukti mampu meningkatkan akurasi klasifikasi secara optimal dan menghasilkan performa yang lebih baik dibandingkan metode lainnya.

5. DAFTAR PUSTAKA

Airi, Fitri Adha Hariyati, Tati Suprapti, and Agus Bahtiar. 2023. "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke." 18:73–79. doi:https://doi.org/10.30587/e-link.v18i1.5271.

Ara, Jinat, Hanif Bhuiyan, Isfara Islam Roza, and Abrar Nahin Mohammad Shadman. 2026. "Importance of Balanced Datasets with Feature Selection and Ensemble Methods on Heart Disease Classification Using Distinctive Machine Learning Techniques : A Comparative Analysis." 1–29.

Faska, Zahra, Lahbib Khrissi, Khalid Haddouch, and Nabil El Akkad. 2023. "A Robust and Consistent Stack Generalized Ensemble - Learning Framework for Image Segmentation." *Journal of Engineering and Applied Science* 1–20. doi:10.1186/s44147-023-00226-4.

Harun, Rodyah Mulyani, and Faisal Fahmi. 2024. "Perbandingan Algoritma Naïve Bayes , KNN , Dan Decision Tree Pada Analisis Sentimen Terhadap Ulasan Aplikasi KitaLulus." 6(3). doi:10.47065/bits.v6i3.6367.

Hendrawinata, Suryadillah, Jasmir Jasmir, and Gunardi Gunardi. 2026. "Perbandingan Algoritma Machine Learning Untuk Prediksi Customer Churn Perbankan Berbasis Credit Score Comparison of Machine Learning Algorithms for Banking Customer Churn Prediction Based on Credit Score." 15:1574–83. doi:https://doi.org/10.32520/stmsi.v15i5.6148.

Kumar, Prof Neelam A., Laxmi Jangale, Vaishnavi Sathe, Aishwarya Shelke, and Tanvi Redij. 2024. "Study of Supervised Logistic Regression Algorithm." 13(2231):227–30.

Putri, A. K., and H. Suparwito. 2023. "Uji Algoritma Stacking Ensemble Classifier Pada Kemampuan Adaptasi Mahasiswa Baru Dalam Pembelajaran Online." 3(1):1–12. doi:https://doi.org/10.24002/konstelasi.v3i1.7009.

Qadrini, Laila. 2023. "Undersampling Dan K-Fold Random Forest Untuk Klasifikasi Kelas Tidak Seimbang." 4(4):1967–74. doi:10.47065/bits.v4i4.3141.

Rahim, Abd Mizwar A., Inggrid Yanuar Risca Pratiwi, and Muhammad Ainul Fikri. 2023. "Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Clasifier." *Indonesian Journal of Computer Science* 12(1):2995–3011. doi:https://doi.org/10.33022/ijcs.v12i5.3413.

Rainio, Oona, Jarmo Teuho, and Riku Klén. 2024. "Evaluation Metrics and Statistical Tests for Machine Learning." *Scientific Reports* 1–14. doi:10.1038/s41598-024-56706-x.

Salmi, Mabrouka, Dalia Atif, Diego Oliva, Ajith Abraham, and Sebastian Ventura. 2024. *Handling Imbalanced Medical Datasets : Review of a Decade of Research*. Vol. 57. Springer Netherlands.

Sapriadi, and Mardiah Nasution. 2024. "Oversampling Menggunakan P Endekatan Latin Hypercube Sampling Dan K-Nearest Neighbors Untuk Meningkatkan Kinerja." *Jurnal Komputer Terapan* (November):98–110.

doi:https://doi.org/10.35143/jkt.v10i2.6389.

Sejati, Puteri, Munawar, Marzuki Pilliang, and Habibullah Akbar. 2022. "Studi Komparasi Naive Bayes , K-Nearest Neighbor , Dan Random Forest Untuk Prediksi Calon Mahasiswa Yang Diterima Atau Mundur." *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)* 9(7):1341–48. doi:10.25126/jtiik.202296737.

Setiawan, Ismail, Ilham Fatah Yasin, and Yuninda Tri Desianti. 2025. "Komparasi Kinerja Algoritma Random Forest , Decision Tree , Naïve Bayes , Dan KNN Dalam Prediksi Tingkat Depresi Mahasiswa Menggunakan Student Depression Dataset." 6(1):47–58. doi:https://doi.org/10.35960/ikomti.v6i1.1756.

Sijabat, Glen Fierre, Wahyu Aji, and Eko Prabowo. 2026. "Studi Komparatif Model Machine Learning Untuk Klasifikasi Penyakit Jantung Dengan SMOTE Pada Data Imbalanced." 13(1):398–409. doi:10.30865/jurikom.v13i1.9485.

Sir, Yosua Alberth, and Agus H. H. Soepranoto. 2022. "Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas." 10(1):31–38. doi:10.35508/jicon.v10i1.6554.

Song, Shihao, and Sibo Yang. 2026. *An Improved SMOTE Method Based on Triangular Scoring Mechanism with Random Perturbation for Handling Class-Imbalanced Classification Problems*. Vol. 123.

Sujon, Khaled Mahmud, Rohayanti Hassan, Kwonhue Choi, and Abdus Samad. 2025. "Accuracy, Precision, Recall, F1-Score, or MCC? Empirical Evidence from Advanced Statistics, ML, and XAI for Evaluating Business Predictive Models." doi:https://doi.org/10.1186/s40537-025-01313-4.

Syukron, Akhmad, Eko Saputro, Sardiarinto, and Pudji Widodo. 2023. "Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung." 10(1):47–50. doi:https://doi.org/10/25047/jtit.v10i1.312.

Taskiran, Salimkan Fatma, Bahaeddin Turkoglu, Ersin Kaya, and Tunc Asuroglu. 2025. "OPEN A Comprehensive Evaluation of Oversampling Techniques for Enhancing Text Classification Performance." 1–20. doi:https://doi.org/10.1038/s41598-025-05791-7.

Wongvorachan, Tarid, Surina He, and Okan Bulut. 2023. "A Comparison of Undersampling , Oversampling , and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining." doi:https://doi.org/10.3390/info14010054.

Zhou, Wensheng, Chen Liu, Peng Yuan, and Lei Jiang. 2024. "Applied Sciences An Undersampling Method Approaching the Ideal Classification Boundary for Imbalance Problems." doi:https:// doi.org/10.3390/app14135421.