

Peningkatan Keberagaman Data untuk Klasifikasi Penyakit Diabetes Berbasis Stacking Ensemble Learning

Nur Kholis Majid^{*)}, Catur Supriyanto, Aris Marjuni,

Program Studi Magister Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang

Info Artikel

Riwayat Artikel:

Received 2024-08-07

Revised 2024-12-17

Accepted 2024-12-18

Corresponding Author:

Nur Kholis Majid

Email:p31202302543@mhs.dinus.ac.id



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstract –Diabetes cases are becoming more common in the late years. Diabetes attacks not only parents, but also children. The development of diabetes is not far from the lifestyle and diet that we live on a daily basis. Therefore, early detection of diabetes is essential because the earlier the disease is detected, the easier it is to treat. In the process of detecting disease based on factors, the cause can be predicted with data mining. The aim of this research is to increase data diversity so that it can be processed to the maximum in data mining. In the process of data upgrading, we used the imbalance learning method SMOTE-ENN combined with the method Stacking Ensemble Learning. In the search for a powerful stacking model, seven classification algorithms were involved in the experiments carried out on this study, namely: Random Forest, Decision Tree, Gradient Boosting, Naïve Bayes, Extreme Gradient Boost, Logistic Regression, and k-Nearest Neighbor. Four algorithms were used to be classifiers level 0 (base model), namely kNN, Gradient Boosting, decision tree, and random forest, while Random Forest was used again to be classifier level 1. (meta model). With these combinations, the accuracy obtained is 97.3%. These are the highest results when compared to individual algorithms.

Keywords: Diabetes, Imbalance Data, Classification, Machine Learning, Stacking Ensemble Methods.

Abstrak – Kasus penyakit diabetes semakin banyak ditemukan pada akhir – akhir ini. Diabetes menyerang tidak hanya orang tua, melainkan juga menyerang anak – anak. Perkembangan penyakit diabetes tidak jauh dari pola hidup dan pola makan yang kita jalani sehari – hari. Oleh karena itu, deteksi sejak dini terhadap penyakit diabetes sangat dibutuhkan karena semakin awal penyakit terdeteksi maka semakin mudah perawatannya. Dalam proses pendeteksian penyakit berdasarkan faktor – faktor penyebabnya dapat diprediksi dengan data mining. Algoritma – algoritma dalam data mining dapat diolah sedemikian rupa sehingga bisa menghasilkan performa yang bagus. Tujuan dari penelitian ini adalah untuk meningkatkan keberagaman data sehingga dapat diproses maksimal dalam data mining. Dalam proses peningkatan data digunakan metode imbalance learning yaitu SMOTE-ENN yang digabung dengan metode Stacking Ensemble Learning. Dalam mencari model stacking yang kuat, dalam eksperimen ini tujuh algoritma klasifikasi digunakan, yaitu: Random Forest, Decision Tree, Gradient Boosting, Naïve Bayes, Extreme Gradient Boosting, Logistic Regression, k-Nearest Neighbor. Empat algoritma digunakan untuk menjadi classifier level 0 (base model) yaitu kNN, Gradient Boosting, Decision Tree, dan Random Forest, sementara Random Forest digunakan lagi untuk menjadi classifier level 1 (meta model). Menggunakan gabungan algoritma tersebut mendapatkan akurasi sebesar 97.3%. Dibandingkan dengan setiap algoritma yang diolah secara terpisah, akurasi gabungan algoritma tersebut merupakan yang paling tinggi. Kombinasi tersebut juga merupakan hasil tertinggi dari eksperimen kombinasi yang lainnya.

Kata Kunci: Diabetes, Imbalance Data, Klasifikasi, Machine Learning, Metode Stacking Ensemble.

I. PENDAHULUAN

Diseluruh dunia, 537 juta orang dewasa (20-79 tahun) hidup dengan diabetes pada tahun 2021. Diabetes bertanggung jawab atas 6,7 juta kematian dan 541 juta kematian orang dewasa mengalami gangguan toleransi glukosa, yang merupakan salah satu faktor risiko utama diabetes melitus [1]. Selain itu, penderita diabetes menghadapi risiko 60% lebih tinggi terkena penyakit kematian dini dibandingkan mereka yang tidak menderita diabetes dan peningkatan kerentanan terhadap penyakit komplikasi seperti kebutaan, gagal ginjal, serangan jantung, stroke, dan amputasi anggota badan [2]. Diabetes adalah penyakit kronis yang terjadi ketika tubuh tidak bisa menggunakan gula darah (glukosa) dengan baik. Glukosa merupakan sumber energi utama bagi sel tubuh manusia. Ada beberapa faktor yang menyebabkan diabetes, diantaranya yang paling sering dijumpai adalah obesitas, umur, dan tensi tinggi [3].

Maka dari itu, identifikasi tepat waktu dan pendeteksian sejak dini terhadap faktor-faktor yang menyebabkan diabetes sangat penting untuk tindakan pencegahan dan penanganan pasien lebih tepat sasaran. Machine learning muncul sebagai salah satu solusi dalam pendeteksian diabetes berdasarkan dari faktor-faktor penyebabnya. Machine learning dapat memproses kumpulan data yang rumit dan membedakan pola yang kompleks [4]. Metode machine learning pada data kesehatan, khususnya dalam tugas klasifikasi penyakit sudah sering digunakan [5]. Masalahnya adalah kualitas kumpulan data mempengaruhi performa metode machine learning dalam klasifikasi penyakit [6].

Secara umum, Sebagian besar data kesehatan, khususnya data penyakit memiliki masalah pada keberagaman data, seperti diabetes [7], jantung [8], dan kanker payudara [9]. Jika masalah kurang beragamnya data pada kumpulan data kesehatan tidak diatasi, hal ini dapat mempengaruhi kinerja metode klasifikasi sehingga membuat hasil prediksi menjadi bias. Dengan data yang beragam, metode klasifikasi dapat bekerja secara maksimal dalam konteks performa.

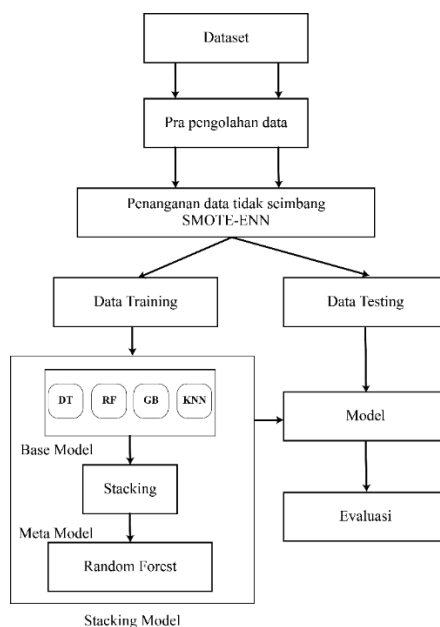
Data mining merupakan suatu proses mengekstrak pengetahuan dari tumpukan data guna menemukan pola atau hubungan padata data [10]. Dengan proses penambangan yang dilakukan terhadap data-data yang lampau (data historis), banyak hal yang bisa dilakukan mulai dari kegiatan klasifikasi, klusterisasi, forecasting, dan juga estimasi. Salah satu penerapan data mining dalam dunia nyata adalah pada bidang kesehata, yaitu untuk membantu praktisi kesehatan membuat keputusan dalam mendiagnosa sebuah penyakit. Proses diagnosa penyakit dapat dilakukan dengan salah satu metode didalam data mining, yaitu klasifikasi [11].

Klasifikasi adalah proses menemukan model untuk mengelompokkan data kedalam kelas yang telah ditentukan berdasarkan kriteria tertentu [12]. Beberapa contoh algoritma klasifikasi yang populer sering digunakan adalah Neural Network [13], kNN [14], Decision Tree [15], Random Forest [16], dan lain-lain. Klasifikasi penyakit penting dilakukan untuk membantu perawatan pasien berdasarkan tingkat resiko atau tingkatannya. Dengan demikian pasien dengan resiko rendah dan tinggi mendapat perawatan atau pengobatan yang berbeda, dengan tujuan untuk mengurangi tingkat kematian terhadap penyakit tersebut.

Pada penelitian ini, penggabungan metode untuk mengatasi data yang kurang beragam seperti oversampling dan undersampling digunakan untuk meningkatkan keberagaman data pada dataset public untuk melihat hasil prediksi yang lebih kuat. Ketika terdapat ketidakberagaman kelas dalam data pelatihan model machine learning, biasanya akan mengklasifikasikan kelas yang lebih besar secara berlebihan karena probabilitas sebelumnya yang meningkat [17]. Akibatnya, algoritma machine learning cenderung salah mengklasifikasikan kelas minoritas. Dampak lebih lanjut, model machine learning akan menghasilkan tingkat akurasi prediksi yang rendah pada kelas minoritas dan tinggi pada kelas mayoritas [18]. Keberagaman data dapat membantu memperbesar perbedaan kelas mayoritas dan minoritas, sehingga menghasilkan prediksi yang lebih kuat dan performa generalisasi yang lebih baik [19]. Untuk meningkatkan akurasi klasifikasi, peneliti menggunakan metode stacking ensemble learning.

II. METODE

Dalam menyelesaikan penelitian ini, beberapa tahapan dilakukan, mulai dari Koleksi data, Pra pengolahan data, Penanganan data tidak seimbang, Pemodelan metode, dan Evaluasi (Gambar 1).

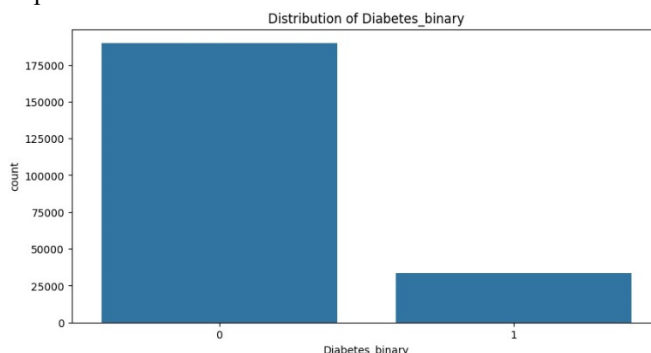


Gambar 1. Metode Penelitian

A. Koleksi Data

Dataset faktor penyakit diabetes yang digunakan untuk eksperimen dalam penelitian ini menggunakan dataset umum yang terbuka untuk publik berasal dari Kaggle. Dataset berasal dari survey yang dikoleksi oleh sistem yang masih berlanjut dengan nama BRFS atau Behavioral Risk Factor Surveillance System. BRFS sendiri adalah sistem

survey telepon terkait kesehatan utama yang mengumpulkan data negara bagian tentang penduduk AS mengenai perilaku berisiko terkait kesehatan, kondisi kesehatan kronis, dan penggunaan layanan pencegahan. BRFSS menyelesaikan lebih dari 400.000 wawancara orang dewasa setiap tahunnya, menjadikannya sistem survei kesehatan berkelanjutan terbesar di dunia [20]. Dataset berjumlah total 223243 baris dengan label pada fitur ‘Diabetes_binary’. Pembagian label dapat dilihat pada Gambar 2.



Gambar 2. Pembagian label

Dataset dibagi menjadi dua bagian, 90% untuk data training, dan 10% untuk data testing. Dataset berisi 22 fitur yang dijelaskan pada tabel 1. Variabel atau fitur kelas Diabetes Binary adalah variabel biner yang menunjukkan apakah pasien menderita diabetes. Lebih tepatnya, “0” menunjukkan tidak diabetes, dan “1” menunjukkan pradiabetes atau diabetes. Selain itu, penelitian ini menggunakan seluruh rangkaian fitur untuk pelatihan model yang diusulkan.

TABEL 1
 PENJELASAN DATASET YANG DIGUNAKAN

Fitur	Deskripsi	Rentang
Diabetes_binary	Tidak diabetes atau Diabetes	0,1
HighBP	Tekanan darah tinggi	0,1
HighChol	Kolesterol tinggi	0,1
CholCheck	Cek kolesterol dalam 5tahun terakhir	0,1
BMI	Index massa darah	12 – 99
Smoker	Perokok	0,1
Stroke	Memiliki penyakit stroke	0,1
HeartDiseaseorAttack	Kelainan atau penyakit jantung	0,1
PhysicalActivity	Aktifitas fisik dalam 30hari terakhir	0,1
Fruits	Konsumsi buah	0,1
Veggies	Konsumsi sayur	0,1
HvyAlcoholConsump	Konsumsi berat alcohol	0,1
AnyHealthCare	Perlindungan Kesehatan	0,1
NoDocbcCost	Biaya cek Kesehatan	0,1
GenHlth	Tingkat Kesehatan	[1,....,5]
MentHlth	Kesehatan mental	[0,....,30]
PhysHlth	Kesehatan fisik	[0,....,30]
DiffWalk	Kesulitan dalam berjalan	0,1
Sex	Jenis kelamin	0,1
Age	Umur	[1,....,13]
Education	Pendidikan	[1,....,6]
Income	Pendapatan	[1,....,11]

Pemilihan Dataset BRFSS berdasarkan penelitian sebelumnya yang sudah menggunakan untuk penelitian lainya seperti Xie, dkk [7]. Meneliti tentang pembangunan prediksi resiko diabetes. Kemudian Bhola, dkk [21] meneliti tentang teknik machine learning untuk prognosis penyakit kronis.

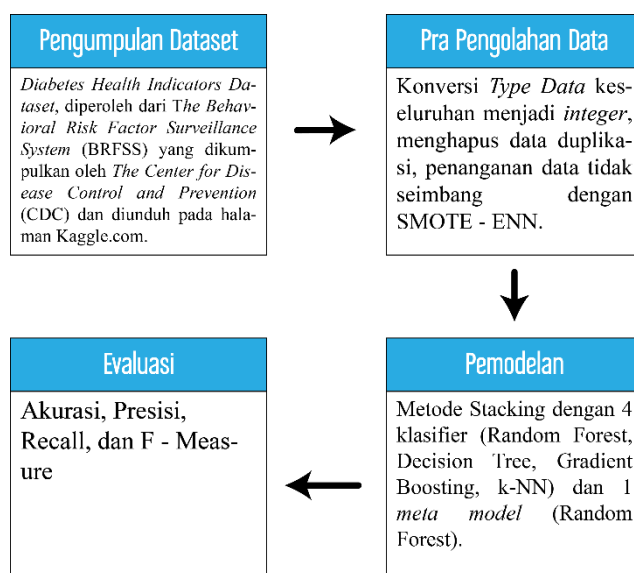
B. Tahapan Penelitian

Diagram alur dan rancangan dimulai dari langkah-langkah berurutan sebagai berikut: identifikasi masalah penelitian, studi literatur, pemodelan, analisis hasil, kesimpulan, seperti yang ditampilkan pada Gambar 3.

Pemodelan sampling untuk mengatasi data yang kurang beragam diusulkan menggunakan gabungan teknik oversampling dan undersampling yaitu SMOTE dan ENN. Metode SMOTE-ENN dipilih karena terbukti handal untuk menangani data yang kurang seimbang. Terbukti pada penelitian Hairani, dkk [22]. Metode SMOTE-ENN dapat meningkatkan performa metode klasifikasi SVM dan Random Forest untuk prediksi diabetes. Selanjutnya penelitian Latief, dkk [23]. SMOTE-ENN digunakan untuk menyeimbangkan data pasien kanker paru-paru.

Kemudian untuk pemodelan klasifikasi yang nantinya akan menghasilkan prediksi diusulkan menggunakan teknik stacking ensembling dimana pada level 0 atau base levelnya berisi Random Forest, Decision Tree, Gradient

Boosting, dan k-Nearest Neighbor, selanjutnya untuk level 1 atau meta levelnya menggunakan metode Random Forest. Metode Random Forest dipilih karena dinilai bagus dalam tugas klasifikasi. Salah satunya adalah penelitian yang dilakukan Song, dkk [24]. Random Forest pada penelitian tersebut dapat dengan andal mengklasifikasikan penyakit Alzheimer menggunakan fitur MRI otak, mengidentifikasi hipokampus, amigdala, dan ventrikel lateral inferior sebagai penanda biologis utama. Random Forest juga dapat meningkatkan kinerja dan mengurangi overfitting, ini berguna untuk menangani kumpulan data berdimensi tinggi dan dapat menangani feature kategorikal dan numerical. Pemilihan metode berdasarkan beberapa penelitian sebelumnya sudah membuktikan bahwa metode tersebut sudah menunjukkan performanya yang bagus terhadap sebuah dataset yang sejenis [8], [25], [22]. Pemilihan metode juga berdasarkan eksperimen terhadap beberapa metode yang dilakukan oleh peneliti dengan dataset yang digunakan, sehingga pemilihan metode dipilih berdasarkan performa yang dihasilkan. Metode gradient boosting dipilih karena metode tersebut bagus dalam tugas prediksi dan untuk meningkatkan performa dan mengurangi overfitting. Seperti pada penelitian Lu, dkk [26]. Penggunaan metode gradient boosting untuk menganalisis dan memprediksi kecelakaan. Secara spesifik, Decision Tree dapat menangani feature kategorikal dan numerical. Menggabungkan kelebihan pada model dasar yang diadopsi di tingkat ke-dua (meta model) memastikan bahwa fitur-fitur kategorikal telah ditangani, kemungkinan overfitting diminimalkan, dan kelas-kelas yang tidak seimbang diatasi [27].



Gambar 3. Tahapan penelitian

C. Pra Pengolahan Data

Pada tahap pra pengolahan data ini, terdapat sebuah langkah penting agar dataset siap digunakan pada tahap pemodelan metode yang digunakan.

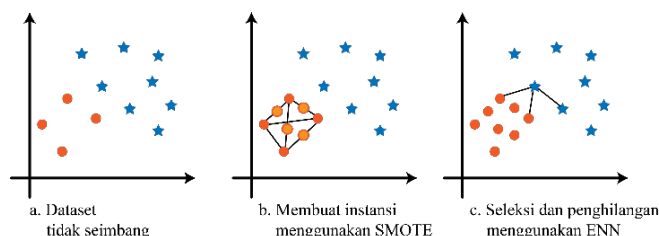
1) Konversi Tipe Data: Didalam tahap pra pengolahan data ini, semua atribut dalam data diubah tipe datanya menjadi integer. Ada 13 atribut yang sebelumnya bertipe float diubah menjadi tipe integer. Perubahan tersebut bertujuan untuk membantu integrasi dengan sistem atau atribut lain yang mungkin memerlukan data dalam format numerik. Selain itu, ada beberapa keuntungan lain dari konversi tersebut diantaranya adalah membantu pengolahan data lebih efektif. Keuntungan yang lain adalah sebagai optimasi dalam penyimpanan dimana integer lebih efisien dibandingkan string, terutama untuk dataset yang besar [28].

2) Menghapus data duplikasi: Data yang terdapat duplikasi nilai dihilangkan, tujuannya untuk mendapat nilai yang lebih representative dan mengurangi beban waktu pada proses pengolahan data [29]. Data semula berjumlah 236378 baris, kemudian terindikasi data duplikat sejumlah 13135 baris kemudian seluruh data duplikat dihapus dan pada akhirnya data menjadi 223243 baris.

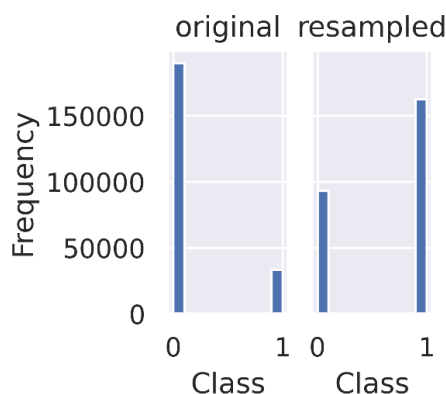
3) Cek nilai yang kosong: Dataset yang digunakan sudah termasuk bersih, dimana tidak ada data yang memiliki nilai kosong.

4) Penanganan data tidak seimbang: Terlihat pada gambar 2. Menunjukkan bahwa label pada dataset yang digunakan tidak seimbang. Maka dari itu perlu penyeimbangan data atau meragamkan data menggunakan SMOTE-ENN. SMOTE-ENN adalah metode sampling yang komprehensif diajukan oleh Batista, dkk [30]. Mengkombinasikan

SMOTE dengan ENN. SMOTE adalah metode pengambilan oversampling, dan ide metode utamanya adalah untuk membentuk contoh kelas minoritas baru dengan melakukan interpolasi antara beberapa contoh kelas minoritas yang terletak bersama. Meskipun secara efektif dapat meningkatkan akurasi klasifikasi model, hal ini juga dapat menghasilkan sampel noise dan sampel batas. Untuk membuat cluster kelas yang terdefinisi dengan lebih baik, ENN digunakan sebagai metode pembersihan data yang dapat menghapus apapun contoh yang label kelasnya setidaknya dua dari tiga tetangga terdekatnya berbeda dari kelasnya. SMOTE-ENN mengurangi kemungkinan terjadinya overfitting [30]. Puri dan Gupta mengatakan bahwa teknik ini memiliki kinerja yang lebih baik ketika kumpulan data yang digunakan memiliki ketidakseimbangan yang tinggi [31]. Karena beberapa alasan tersebut maka digunakanlah metode SMOTE-ENN pada penelitian ini. Ilustrasi dari SMOTE-ENN digambarkan pada Gambar 4.



Gambar 4. Proses SMOTE-ENN



Gambar 5. Hasil sampling dataset

D. Modelling

Terdapat beberapa pendekatan untuk mengatasi data yang kurang beragam. 1) Data Balancing, dimana pada tahap pre-processing data diproses menggunakan sampling untuk membuat data menjadi lebih beragam. 2) Modifikasi algoritma. 3) Pendekatan metode ensemble. 4) Feature Selection & Evaluation Measures [32]. Penggunaan cross-validation digunakan untuk generalisasi model learning [15]. Untuk mendapatkan model yang dijadikan data statistik menggunakan K-Fold Cross Validation. Nilai k yang digunakan pada metode K-Fold Cross Validation adalah k=5. Selanjutnya, pada penelitian ini kami menggunakan dua pendekatan untuk mengatasi data yang kurang beragam, yaitu data balancing dan pendekatan metode ensemble. Untuk data balancing kami menggunakan metode SMOTE-ENN dan pendekatan metode ensemble kami menggunakan metode stacking ensemble learning.

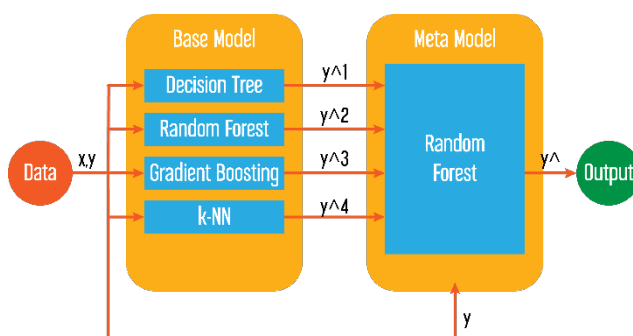
Machine learning adalah bagian dari kecerdasan buatan yang digunakan untuk memperkenalkan program komputer. Program-program tersebut diperkenalkan oleh sekumpulan data ekstensif menggunakan beberapa algoritma dan statistik [33]. Machine learning membantu untuk melacak data. Beberapa teknik machine learning menggunakan set data yang sangat luas atau besar, membuatnya menjadi semakin spesifik dalam menetapkan pola berdasarkan data lampau, dan dapat memprediksi tentang masa yang akan mendatang.

Ensemble learning adalah metode meta machine learning yang mencari kinerja prediktif yang lebih baik dengan menggabungkan prediksi dari beberapa model [34]. Ada tiga kelas utama pada ensemble learning yaitu bagging, boosting, dan stacking [35]. Persamaan dari 3 kelas tersebut adalah semuanya menggunakan beberapa base model atau model dasar berupa beberapa model machine learning. Singkatnya, ensemble learning adalah penggabungan model classifier atau model pakar untuk mengatasi masalah seperti regresi dan klasifikasi. Tujuan penggabungan model classifier ini adalah untuk mendapatkan hasil yang lebih baik.

Stacking ensemble learning adalah salah satu metode ensemble learning yang menggunakan beberapa base model dengan melakukan pembelajaran secara paralel dan independen pada tiap base model dan kemudian dikombinasikan dengan menggunakan satu algoritma meta learning untuk memberikan output dari hasil kombinasi base model. Stacking adalah menggabungkan beberapa pembelajaran tingkat dasar (base model) ke pembelajaran tingkat kedua (meta model) [36].

Decision Tree merupakan algoritma populer, baik untuk keperluan klasifikasi maupun regresi. Secara spesifik, Decision Tree dapat menangani fitur kategorikal dan numerikal, Random Forest menggabungkan beberapa pohon keputusan untuk meningkatkan kinerja dan mengurangi overfitting, ini berguna untuk menangani kumpulan data berdimensi tinggi dan dapat menangani fitur kategorikal dan numerikal. Selanjutnya penggunaan Gradient Boost untuk meningkatkan performa dan mengurangi overfitting. Menggabungkan kelebihan pada model dasar yang diadopsi di tingkat ke-dua (meta model) memastikan bahwa fitur-fitur kategorikal telah ditangani, kemungkinan overfitting diminimalkan, dan kelas-kelas yang tidak seimbang diatasi [27].

Langkah selanjutnya adalah menentukan algoritma sebagai meta model. Algoritma meta model akan melakukan proses pelatihan sekaligus training dari data yang digunakan untuk mendapatkan hasil prediksi dari base model. Algoritma Random Forest dipilih sebagai meta model pada penelitian ini. Algoritma Random Forest dirasa bisa diandalkan sebagai meta model. Dengan terbukti dari penelitian sebelumnya yang juga menggunakan Random Forest sebagai meta model dan diaplikasikan pada bidang berbeda-beda [37], [38], [39].



Gambar 6. Stacking model yang digunakan

E. Evaluasi

Eksperimen yang dilakukan untuk studi ini berakhir dengan evaluasi, yang melibatkan mencari tahu tingkat akurasi, presisi, recall, dan f - measure. Untuk melihat dengan jelas seberapa sering model memprediksi atau mengklasifikasikan, akurasi diperlukan (formula 1).

$$\text{Akurasi} = \frac{TP+TN}{(TP+FP+TN+FN)} \quad (1)$$

Presisi merupakan rasio prediksi benar mengalami diabetes dibandingkan dengan keseluruhan hasil yang diprediksi positif. Nilai presisi (formula 2) diperoleh dari pembagian True Positive dengan penjumlahan True Positive dan False Positive. Presisi sangat dibutuhkan pada penelitian dibidang kesehatan [40].

$$\text{Presisi} = \frac{TP}{(TP+FP)} \quad (2)$$

Recall merupakan rasio prediksi benar mengalami diabetes dengan tepat terhadap keseluruhan sampel deteksi diabetes (formula 3). Nilai recall yang tinggi menentukan dengan benar kasus yang didiagnosis.

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (3)$$

Pengukuran performa yang terakhir adalah F - Measure. F - Measure diperoleh dari perhitungan recall dan precision (formula 4).

$$\text{F - Measure} = \frac{2 * \text{Recall} * \text{Presisi}}{\text{Recall} + \text{Presisi}} \quad (4)$$

Ke-empat metode evaluasi tersebut akan digunakan sebagai tolak ukur performa dari masing-masing model klasifikasi secara terpisah, juga mengukur beberapa performa model klasifikasi stacking untuk mencari yang terbaik.

III. HASIL DAN PEMBAHASAN

Dalam penelitian ini digunakan tujuh metode klasifikasi data mining. Ketujuh metode tersebut adalah: k-Nearest Neighbor, Random Forest, Decision Tree, Logistic Regression, Gradient Boosting, Naïve Bayes, Extreme Gradient Boosting. Setiap algoritma dijalankan dengan data sampel dan tanpa sampling SMOTE-ENN pada kumpulan data yang disiapkan. Hasil akurasi, presisi, dan recall tanpa sampling disajikan pada Tabel 2. Sedangkan hasil performa dengan sampling disajikan pada Tabel 3.

TABEL 2
 HASIL PERFORMA ALGORITMA SECARA TERPISAH DAN TANPA SAMPLING

Algoritma	Akurasi	Presisi	Recall	F-Measure
k-Nearest Neighbor	0.836	0.400	0.212	0.277
Random Forest	0.845	0.445	0.163	0.239
Decision Tree	0.774	0.273	0.306	0.288
Gradient Boosting	0.855	0.552	0.158	0.245
Logistic Regression	0.853	0.516	0.151	0.233
Extreme Gradient Boosting	0.853	0.533	0.169	0.256
Naïve Bayes	0.757	0.324	0.568	0.413

Percobaan pertama dilakukan dengan mengolah data tanpa sampling atau tanpa melalui balancing data menggunakan SMOTE-ENN. Terlihat pada table 2, meskipun masing-masing metode klasifikasi sudah menunjukkan akurasi yang tinggi. Akan tetapi nilai evaluasi presisi, recall, dan f-measure masih sangat rendah. Ini menunjukkan bahwa model masih belum akurat. Presisi sangat penting dalam klasifikasi untuk menganalisa hasil dari luaran klasifikasi tersebut [41].

TABEL 3
 HASIL PERFORMA ALGORITMA SECARA TERPISAH DENGAN SAMPLING SMOTE-ENN

Algoritma	Akurasi	Presisi	Recall	F-Measure
k-Nearest Neighbor + SMOTE-ENN	0.953	0.936	0.994	0.994
Random Forest SMOTE-ENN	0.947	0.939	0.980	0.959
Decision Tree SMOTE-ENN	0.921	0.922	0.957	0.939
Gradient Boosting SMOTE-ENN	0.875	0.883	0.926	0.904
Logistic Regression SMOTE-ENN	0.862	0.875	0.913	0.894
Extreme Gradient Boosting SMOTE-ENN	0.893	0.903	0.932	0.917
Naïve Bayes SMOTE-ENN	0.832	0.859	0.880	0.869

Setelah melakukan percobaan, terlihat perbedaan yang signifikan pada tabel 2 dan tabel 3 terjadi pada hasil performa yang menggunakan data sampling SMOTE-ENN dan yang tidak menggunakan data sampling. Ke-tujuh algoritma yang dijalankan dengan data sampling menggunakan SMOTE-ENN mendapatkan nilai akurasi, presisi, recall, dan f - measure yang termasuk tinggi, yaitu diatas 90%. Nilai tertinggi dimiliki oleh k-Nearest Neighbor memakai SMOTE-ENN, dengan akurasi 95,3%, presisi 93,6%, recall 99,4%, dan f – measure 96,4%. Penggunaan data sangat memiliki pengaruh besar dalam nilai performa yang dihasilkan oleh suatu algoritma. Karena dalam hal ini objek yang digunakan adalah data yang berhubungan dengan kesehatan khususnya data diabetes, yang dimana hasil false negative mempunyai resiko yang sangat tinggi, sehingga angka akurasi sebesar 95,3% masih perlu ditingkatkan. Dengan adanya angka tersebut, berarti masih ada kemungkinan sebesar 4,7% bahwa alat pendeteksi akan salah mengklasifikasikan suatu kelompok sebagai mengidap penyakit atau tidak. Bagaimanapun diperlukan upaya lebih agar angka akurasi 95,3% yang dimiliki oleh k-Nearest Neighbor bisa ditingkatkan lagi. Begitupun dengan nilai presisi, recall, dan f – measure secara keseluruhan. Pengoptimasian algoritma klasifikasi untuk meningkatkan performa adalah salah satunya dengan menggunakan teknik ensemble, terdapat beberapa tipe ensemble learning yaitu bagging, boosting, voting, stacking, dan blending [42]. Pada penelitian ini peneliti menggunakan metode stacking.

Menggunakan metode stacking, nilai performa skor akurasi, presisi, dan f - measure untuk klasifikasi penyakit diabetes berhasil meningkat, meskipun nilai yang meningkat tidak terlalu tinggi karena skor yang diperoleh dari base model sudah sangat tinggi. Perbandingan dari hasil performa pada metode stacking dengan 7 model klasifikasi dasar dapat dilihat pada tabel 4.

Berdasarkan uji coba yang telah dilakukan, terdapat beberapa peningkatan seperti pada akurasi yang berhasil meningkat sebesar 0.020, sedangkan untuk presisi berhasil meningkat sebesar 0.038, adapun untuk recall sedikit menurun dengan nilai 0.011, untuk nilai f – measure meningkat sebesar 0.015. Peningkatan nilai-nilai ini penting mengingat semakin mendekat ke angka sempurna yaitu 100%.

TABEL 4
 HASIL PERFORMA ALGORITMA SECARA INDIVIDU DENGAN ALGORITMA STACKING

Algoritma	Akurasi	Presisi	Recall	F-Measure
Stacking	0.855	0.554	0.148	0.234
Stacking + SMOTE-ENN	0.973	0.974	0.983	0.979
k-Nearest Neighbor + SMOTE-ENN	0.953	0.936	0.994	0.994
Random Forest SMOTE-ENN	0.947	0.939	0.980	0.959
Decision Tree SMOTE-ENN	0.921	0.922	0.957	0.939
Gradient Boosting SMOTE-ENN	0.875	0.883	0.926	0.904
Logistic Regression SMOTE-ENN	0.862	0.875	0.913	0.894

Pada tabel 4 tertulis bahwa ada peningkatan metode stacking yang menggunakan data sampling hasil pengolahan metode SMOTE-ENN. Penggunaan model Stacking dengan SMOTE-ENN pada data tidak beragam sangat diandalkan, karena menghasilkan performa yang tinggi.

Peneliti juga melakukan uji coba yang lain dengan mengganti base model dan meta model dengan beragam algoritma terlihat pada tabel 5.

TABEL 5
 PERBANDINGAN STACKING DENGAN BASE DAN META MODEL YANG BERBEDA

Base Model	Meta Model	Akurasi	Presisi	Recall	F-Measure
RF, DT, GB, NB	KNN	0.916	0.920	0.951	0.935
RF, DT, XGB, kNN	DT	0.956	0.967	0.964	0.965
RF, DT, GB, KNN	LR	0.971	0.967	0.988	0.977
RF, DT, GB, KNN	RF	0.973	0.974	0.983	0.979

Dari hasil diatas, diperoleh nilai akurasi, presisi, recall, dan f-measure yang ternyata tidak lebih baik dari Random Forest, Decision Tree, Gradient Boosting, dan kNN sebagai base model dan Random Forest sebagai meta model. Bagaimanapun, berdasarkan hasil yang tersaji pada tabel IV, dapat disimpulkan bahwa hasil performa yang didapat dengan metode stacking dapat bekerja lebih baik dari algoritma-algoritma klasifikasi dasar yang di eksekusi secara individu dengan menghasilkan performa yang lebih baik. Meskipun pada penelitian sebelumnya yang menggunakan Logistic Regression sebagai meta model menghasilkan hasil yang cukup baik [43], [44], [45], [46], [47], [48], [49], akan tetapi pada eksperimen kami yang mendapatkan tingkat performa yang lebih baik adalah Random Forest sebagai meta model.

TABEL 6
 PERBANDINGAN DENGAN PENELITIAN TERKAIT

Peneliti	Tahun	Dataset	Metode	Akurasi
Xie, dkk	2019		NN	82.4%
Bhola, dkk	2021	BRFSS	Random Forest	86.8%
Ullah, dkk	2022		k-NN	93.38%
Proposed			Stacking+SMOTE-ENN	97,3%

Pada penelitian yang dilakukan oleh ullah, dkk [25], nilai akurasi tertinggi hanya 93.38% pada kasus deteksi resiko tinggi penyakit diabetes menggunakan dataset BRFSS tahun 2015 menggunakan algoritma kNN dan data diseimbangkan menggunakan metode SMOTE-ENN. Pada penelitian lain yang dilakukan [7], akurasi tertinggi yang didapat adalah 82.4% menggunakan dataset diabetes BRFSS tahun 2014. Bhola, dkk [21] melakukan penelitian terhadap beberapa penyakit kronis dan salah satunya adalah diabetes menggunakan dataset BRFSS tahun 2017. Dimana algoritma Random Forest terbukti handal dengan menghasilkan skor 86.8%. Chowdhury, dkk [50] menginvestigasi terkait penanganan data yang tidak seimbang dengan pendekatan beberapa metode seperti ENN, SMOTE-N, SMOTE-ENN, SMOTE-Tomek menggunakan dataset BRFSS 2021.

Meskipun hasil yang didapat sudah termasuk tinggi, penelitian ini masih terdapat beberapa batasan diantaranya adalah proses pengolahan data yang lama. Terhitung dengan eksperimen yang dikerjakan dengan google colab pro dengan runtime T4GPU menghabiskan waktu sebanyak 549menit. Beberapa cara untuk mengatasi masalah tersebut adalah dengan menggunakan GPU premium untuk komputasi tambahan.

IV. SIMPULAN

Perkembangan penyakit diabetes yang massive dengan tingkat resiko yang berbeda-beda akan sangat mengkhawatirkan jika tidak ditangani secara tepat sesuai tingkat resiko dan secara cepat. Alasan tersebut membuat

para peneliti untuk ikut ambil andil dalam pengolahan data terkait perkembangan penyakit, salah satunya dengan mencari metode yang optimal untuk pendeteksian terhadap indikasi penyakit diabetes atau tidak. Langkah ini sangat penting untuk menyisir seseorang dengan indikasi terkena penyakit diabetes berdasarkan faktor – faktor dalam kehidupan supaya seseorang tersebut dapat memperbaiki pola hidup atau menjalani perawatan sesuai kondisinya. Disisi lain, bagi praktisi kesehatan dapat terbantu dengan pengelompokkan resiko diabetes ini dengan menerapkan perawatan dan pengobatan sesuai resiko yang diderita pasien. Dataset publik digunakan dalam penelitian ini, dengan terlebih dahulu melakukan pra pemrosesan data agar data tersebut bersih untuk diproses selanjutnya. Setelah data bersih, kemudian data disergamkan tingkatnya pada label yang digunakan menggunakan algoritma SMOTE-ENN. Selanjutnya data siap untuk masuk pada tahap pengklasifikasian, 7 algoritma klasifikasi (Random Forest, Decision Tree, Gradient Boosting, Naïve Bayes, Extreme Gradient Boosting, Logistic Regression, k-Nearest Neighbor) diproses secara mandiri untuk dilihat performanya, selanjutnya dibandingkan dengan metode stacking. Hasilnya, metode stacking menggunakan SMOTE-ENN sebagai penyeragam datanya dengan meta model Random Forest sukses mendapatkan nilai yang lebih unggul dari 7 algoritma yang dieksekusi secara individu. Terakhir, stacking menggunakan SMOTE-ENN sebagai penyeragam datanya dengan Random Forest sebagai meta model juga dibandingkan dengan metode stacking lain, tetapi menggunakan base dan meta model yang berbeda, yaitu kNN, Decision Tree, Logistic Regression, dan Random Forest. Hasilnya metode stacking menggunakan SMOTE-ENN sebagai penyeragam datanya dengan meta classifier Random Forest masih menjadi yang terbaik, dengan tingkat akurasi 97.3%, presisi 97.4%, recall 98.3%, dan f – measure 97.9%. Meskipun nilai performa yang sudah mencapai rata – rata 97%, bagaimanapun hasil tersebut masih perlu untuk ditingkatkan, mengingat dalam dunia kesehatan perbedaan 0,001% akan sangat berarti. Maka dari itu, nilai performa secara keseluruhan akan sangat mempengaruhi keakuratan pendeteksian penyakit diabetes. Penyempurnaan tingkat performa sampai 100% dapat dijadikan sebagai tujuan pada penelitian selanjutnya. Beberapa cara yang dapat dilakukan untuk meningkatkan performa tersebut bisa dengan pre-processing dengan feature selection, kemudian mengatur hyperparameter pada classifier sehingga classifier dapat bekerja secara optimal.

UCAPAN TERIMAKASIH

Ucapan rasa terima kasih yang besar kepada Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi karena telah mendanai penelitian ini melalui skema Penelitian Tesis Magister tahun 2024

DAFTAR PUSTAKA

- [1] “Diabetes estimates (20-79 y): People with diabetes, in 1,000s.” Accessed: Mar. 22, 2024. [Online]. Available: <https://diabetesatlas.org/data/>
- [2] “Diabetes basics,” International Diabetes Federation. Accessed: Jun. 15, 2024. [Online]. Available: <https://idf.org/about-diabetes/what-is-diabetes/>
- [3] “Complications,” International Diabetes Federation. Accessed: Jun. 15, 2024. [Online]. Available: <https://idf.org/about-diabetes/diabetes-complications/>
- [4] X. Hu, L. Chu, J. Pei, W. Liu, and J. Bian, “Model Complexity of Deep Learning: A Survey,” Aug. 02, 2021, *arXiv*: arXiv:2103.05127. doi: 10.48550/arXiv.2103.05127.
- [5] A. Holzinger, “Machine Learning for Health Informatics,” vol. 9605, 2016, pp. 1–24. doi: 10.1007/978-3-319-50478-0_1.
- [6] R. Longadge and S. Dongre, “Class Imbalance Problem in Data Mining Review,” May 07, 2013, *arXiv*: arXiv:1305.1707. doi: 10.48550/arXiv.1305.1707.
- [7] Z. Xie, O. Nikolayeva, J. Luo, and D. Li, “Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques,” *Prev Chronic Dis*, vol. 16, p. E130, Sep. 2019, doi: 10.5888/pcd16.190109.
- [8] A. Daza Vergaray *et al.*, “Stacking Ensemble Based Hyperparameters for the Prediction of Heart Disease,” Sep. 06, 2023, *Rochester, NY*: 4553047. doi: 10.2139/ssrn.4553047.
- [9] A. Carriero, L. Groenhoff, E. Vologina, P. Basile, and M. Albera, “Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024,” *Diagnostics*, vol. 14, no. 8, Art. no. 8, Jan. 2024, doi: 10.3390/diagnostics14080848.
- [10] J. Han and M. Kamber, *Data mining: concepts and techniques*, 3rd ed. Burlington, MA: Elsevier, 2012.
- [11] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2022.
- [12] K. Lai, N. Twine, A. O’Brien, Y. Guo, and D. Bauer, “Artificial Intelligence and Machine Learning in Bioinformatics,” in *Encyclopedia of Bioinformatics and Computational Biology*, S. Ranganathan, M. Gribskov, K. Nakai, and C. Schönbach, Eds., Oxford: Academic Press, 2019, pp. 272–286. doi: 10.1016/B978-0-12-809633-8.20325-7.
- [13] J. Amin, M. A. Anjum, M. Sharif, S. Kadry, and J. Kim, “Explainable Neural Network for Classification of Cotton Leaf Diseases,” *Agriculture*, vol. 12, no. 12, Art. no. 12, Dec. 2022, doi: 10.3390/agriculture12122029.
- [14] M. Beckmann, N. F. F. Ebecken, and B. S. L. P. de Lima, “A KNN Undersampling Approach for Data Balancing,” *Journal of Intelligent Learning Systems and Applications*, vol. 7, no. 4, Art. no. 4, Sep. 2015, doi: 10.4236/jilsa.2015.74010.
- [15] L. Rokach and O. Maimon, “Decision Trees,” in *The Data Mining and Knowledge Discovery Handbook*, vol. 6, 2005, pp. 165–192. doi: 10.1007/0-387-25465-X_9.
- [16] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [17] T. Pan, J. Zhao, W. Wu, and J. Yang, “Learning imbalanced datasets based on SMOTE and Gaussian distribution,” *Information Sciences*, vol. 512, pp. 1214–1233, Feb. 2020, doi: 10.1016/j.ins.2019.10.048.
- [18] D. Elreedy and A. F. Atiya, “A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance,” *Information Sciences*, vol. 505, pp. 32–64, Dec. 2019, doi: 10.1016/j.ins.2019.07.070.
- [19] M. O. Ojo and A. Zahid, “Improving Deep Learning Classifiers Performance via Preprocessing and Class Imbalance Approaches in a Plant Disease Detection Pipeline,” *Agronomy*, vol. 13, no. 3, Art. no. 3, Mar. 2023, doi: 10.3390/agronomy13030887.

- [20] "CDC - BRFSS - BRFSS Frequently Asked Questions (FAQs)." Accessed: Jul. 19, 2024. [Online]. Available: https://www.cdc.gov/brfss/about/brfss_faq.htm
- [21] G. Bhola, A. Garg, and M. Kumari, "Comparative Study of Machine Learning Techniques for Chronic Disease Prognosis," in *Computer Networks and Inventive Communication Technologies*, S. Smys, R. Palanisamy, A. Rocha, and G. N. Beligiannis, Eds., Singapore: Springer Nature, 2021, pp. 131–144. doi: 10.1007/978-981-15-9647-6_10.
- [22] H. Hairani and D. Priyanto, "A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data," *International Journal of Advanced Computer Science and Applications*, vol. 14, pp. 585–590, Aug. 2023, doi: 10.14569/IJACSA.2023.0140864.
- [23] M. A. Latief, L. R. Nabila, W. Miftakhurrahman, S. Ma'rufatullah, and H. Tantyoko, "Handling Imbalance Data using Hybrid Sampling SMOTE-ENN in Lung Cancer Classification," *IJECSA*, vol. 3, no. 1, pp. 11–18, Feb. 2024, doi: 10.30812/ijecsa.v3i1.3758.
- [24] M. Song, H. Jung, S. Lee, D. Kim, and M. Ahn, "Diagnostic Classification and Biomarker Identification of Alzheimer's Disease with Random Forest Algorithm," *Brain Sciences*, vol. 11, no. 4, p. 453, Apr. 2021, doi: 10.3390/brainsci11040453.
- [25] Z. Ullah *et al.*, "Detecting High-Risk Factors and Early Diagnosis of Diabetes Using Machine Learning Methods," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–10, Sep. 2022, doi: 10.1155/2022/2557795.
- [26] P. Lu *et al.*, "A Gradient Boosting Crash Prediction Approach for Highway-Rail Grade Crossing Crash Analysis," *Journal of Advanced Transportation*, vol. 2020, pp. 1–10, Jun. 2020, doi: 10.1155/2020/6751728.
- [27] A. Ghasemieh, A. Lloyd, P. Bahrami, P. Vajar, and R. Kashef, "A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients," *Decision Analytics Journal*, vol. 7, p. 100242, Jun. 2023, doi: 10.1016/j.dajour.2023.100242.
- [28] "Cara Convert String ke Integer di Python: Panduan Lengkap 2024 | Revou." Accessed: Jul. 02, 2024. [Online]. Available: <https://revou.co/panduan-teknis/convert-string-to-int-python>
- [29] Y. Kwon, M. Lemieux, J. McTavish, and N. Wathen, "Identifying and removing duplicate records from systematic review searches," *J Med Libr Assoc*, vol. 103, no. 4, pp. 184–188, Oct. 2015, doi: 10.3163/1536-5050.103.4.004.
- [30] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.
- [31] A. Puri and M. Kumar Gupta, "Improved Hybrid Bag-Boost Ensemble With K-Means-SMOTE-ENN Technique for Handling Noisy Class Imbalanced Data," *The Computer Journal*, vol. 65, no. 1, pp. 124–138, Jan. 2020, doi: 10.1093/comjnl/bxab039.
- [32] H. Patel, D. Singh Rajput, G. Thippa Reddy, C. Iwendi, A. Kashif Bashir, and O. Jo, "A review on classification of imbalanced data for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 16, no. 4, p. 1550147720916404, Apr. 2020, doi: 10.1177/1550147720916404.
- [33] S. K. Singh, A. K. Tiwari, and H. K. Paliwal, "A state-of-the-art review on the utilization of machine learning in nanofluids, solar energy generation, and the prognosis of solar power," *Engineering Analysis with Boundary Elements*, vol. 155, pp. 62–86, Oct. 2023, doi: 10.1016/j.enganabound.2023.06.003.
- [34] A. A. Khan, O. Chaudhari, and R. Chandra, "A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation," *Expert Systems with Applications*, vol. 244, p. 122778, Jun. 2024, doi: 10.1016/j.eswa.2023.122778.
- [35] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241–258, Apr. 2020, doi: 10.1007/s11704-019-8208-z.
- [36] D. Wolpert, "Stacked Generalization," *Neural Networks*, vol. 5, pp. 241–259, Dec. 1992, doi: 10.1016/S0893-6080(05)80023-1.
- [37] N. V. S. D. Sripada, S. V, and M. Aghaei, "Detection of visual faults in photovoltaic modules using a stacking ensemble approach," *Heliyon*, vol. 10, no. 6, p. e27894, Mar. 2024, doi: 10.1016/j.heliyon.2024.e27894.
- [38] H. Qian, N. Bao, D. Meng, B. Zhou, H. Lei, and H. Li, "Mapping and classification of Liao River Delta coastal wetland based on time series and multi-source GaoFen images using stacking ensemble model," *Ecological Informatics*, vol. 80, p. 102488, May 2024, doi: 10.1016/j.ecoinf.2024.102488.
- [39] M. Ali, M. N. Haider, S. A. Lashari, W. Sharif, A. Khan, and D. A. Ramli, "Stacking Classifier with Random Forest functioning as a Meta Classifier for Diabetes Diseases Classification," *Procedia Computer Science*, vol. 207, pp. 3459–3468, Jan. 2022, doi: 10.1016/j.procs.2022.09.404.
- [40] S. A. Hicks *et al.*, "On evaluation metrics for medical applications of artificial intelligence," *Sci Rep*, vol. 12, p. 5979, Apr. 2022, doi: 10.1038/s41598-022-09954-8.
- [41] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, "Deep Convolution Neural Network sharing for the multi-label images classification," *Machine Learning with Applications*, vol. 10, p. 100422, Dec. 2022, doi: 10.1016/j.mlwa.2022.100422.
- [42] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, vol. 14, 2012. doi: 10.1201/b12207.
- [43] F. A. Rafrastara, C. Supriyanto, C. Paramita, and Y. P. Astuti, "Deteksi Malware menggunakan Metode Stacking berbasis Ensemble," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, no. 1, Art. no. 1, Jan. 2023, doi: 10.30591/jpit.v8i1.4606.
- [44] "A Novel Dynamic Android Malware Detection System With Ensemble Learning | IEEE Journals & Magazine | IEEE Xplore." Accessed: Aug. 01, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/8374030>
- [45] A. Ahad, *A Stacking-based Ensemble Learning Method for Outlier Detection*. 2021. doi: 10.21203/rs.3.rs-677845/v1.
- [46] Y. 1 Jia, H. 1 Zhang, H. 2 Liu, G. 1 Zhong, and G. 1 C. of C. A. Li, "Flight Delay Classification Prediction Based on Stacking Algorithm," 2021, doi: 10.1155/2021/4292778.
- [47] "Papers with Code - LTG-ST at NADI Shared Task 1: Arabic Dialect Identification using a Stacking Classifier." Accessed: Aug. 01, 2024. [Online]. Available: <https://paperswithcode.com/paper/ltg-st-at-nadi-shared-task-1-arabic-dialect>
- [48] H. Byeon, "Exploring Factors Associated with the Social Discrimination Experience of Children from Multicultural Families in South Korea by using Stacking with Non-linear Algorithm," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 12, no. 5, Art. no. 5, 58/31 2021, doi: 10.14569/IJACSA.2021.0120516.
- [49] Y. Wang, Q. Pan, X. Liu, and Y. Ding, "ET-MSF: a model stacking framework to identify electron transport proteins," *Front Biosci (Landmark Ed)*, vol. 27, no. 1, p. 12, Jan. 2022, doi: 10.31083/j.fb12701012.
- [50] M. M. Chowdhury, R. S. Ayon, and M. S. Hossain, "An investigation of machine learning algorithms and data augmentation techniques for diabetes diagnosis using class imbalanced BRFSS dataset," *Healthcare Analytics*, vol. 5, p. 100297, Jun. 2024, doi: 10.1016/j.health.2023.100297.