Perbandingan Algoritma *Support Vector Machine* Dan Algoritma *Random Forest* Dalam Prediksi Hipertensi

1st Syaloom Zefanya Yuni Br Manurung
Teknik Informatika
Telkom University Purwokerto
Purwokerto, Indonesia
syaloomzym@student.telkomuniversity.ac.id

2nd Dasril Aldo, S.Kom., M.Kom
Teknik Informatika
Telkom University Purwokerto
Purwokerto, Indonesia
dasrilaldo@telkomuniversity.ac.id

3rd Aminatus Sa'adah, S.Si., M.Si.
Teknik Informatika
Telkom University Purwokerto
Purwokerto, Indonesia
aminatuss@telkomuniversity.ac.id

***Abstrak* — Hipertensi merupakan salah satu penyakit tidak menular yang berpotensi menimbulkan komplikasi serius dan menunjukkan tren peningkatan prevalensi baik secara global maupun nasional. Deteksi dini terhadap kondisi ini sangat penting guna mencegah dampak kesehatan yang berbahaya. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma machine learning, yaitu Support Vector Machine (SVM) dan Random Forest (RF), dalam memprediksi hipertensi menggunakan data rekam medis dari Puskesmas Purwokerto Timur I. Karena data hipertensi biasanya memiliki distribusi kelas yang tidak seimbang, penelitian ini menerapkan teknik Oversampling untuk menyeimbangkan data. Tahapan penelitian mencakup preprocessing data, pembangunan model menggunakan algoritma SVM dan RF, serta evaluasi model dengan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa algoritma RF memberikan hasil terbaik dengan akurasi mencapai 98,92%, sementara SVM menghasilkan akurasi tertinggi sebesar 83,91%. Berdasarkan temuan tersebut, dapat disimpulkan bahwa algoritma RF lebih efektif dalam melakukan prediksi hipertensi pada data yang tidak seimbang, dan penerapan teknik Oversampling secara signifikan dapat meningkatkan performa model. Penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem prediksi hipertensi yang lebih akurat untuk mendukung upaya pencegahan dan pengelolaan kesehatan masyarakat.**

***Kata kunci*—hipertensi, Prediksi, Oversampling, Random Forest, Support Vector Machine**

I. PENDAHULUAN

Hipertensi dapat juga diartikan sebagai kondisi dengan tekanan darah secara konsisten melebihi batas normal. Tekanan darah sistolik dikatakan berada pada ambang batas atas jika mencapai 130 mmHg, sedangkan untuk tekanan darah diastolik batas atasnya adalah 80 mmHg [1]. Jika tekanan darah tinggi tidak dikendalikan dalam waktu yang lama, hal ini dapat menimbulkan berbagai komplikasi serius, termasuk terganggunya aliran darah, kerusakan pada pembuluh darah, serta meningkatnya risiko terkena penyakit degeneratif. Kondisi ini bisa menimbulkan dampak serius apabila tidak ditangani dengan benar [2]. Selain menjadi beban kesehatan fisik, hipertensi juga dapat menimbulkan beban psikologis yang memengaruhi kualitas hidup penderitanya [3]

Data World Health Organization (WHO) tahun 2015 menunjukkan sekitar 1,13 miliar orang di dunia mengalami hipertensi, dan jumlah ini terus meningkat. Jika kesadaran individu tidak membaik, WHO memprediksi pada 2025 penderita hipertensi bisa mencapai 1,5 miliar orang. [4]. Pada data Riset Kesehatan Dasar (Riskesdas) menyatakan bahwa pada tahun 2018, terdapat 658.201 orang berusia di atas 18 tahun di Indonesia yang menderita hipertensi. [5]. Fakta ini menunjukkan bahwa hipertensi telah menjadi permasalahan kesehatan global yang perlu mendapat perhatian khusus, terutama dalam upaya deteksi dan penanganan dini.

Oleh sebab itu dibutuhkan pemanfaatan teknologi kecerdasan buatan atau machine learning yang membantu untuk memprediksi penyakit hipertensi secara tepat dan akurat untuk kepentingan dalam memastikan kesehatan masyarakat [6]. Penelitian ini membandingkan dua algoritma pembelajaran mesin yang sering digunakan, adalah Support Vector Machine dan Random Forest, pada konteks prediksi hipertensi. Support Vector Machine dikenal karena keandalannya dalam pengelolaan data yang kompleks dan memiliki dimensi tinggi dengan menentukan hyperplane terbaik untuk memisahkan antara dua kelas [7]. Sebaliknya, Random Forest adalah metode ensemble learning dengan menggabungkan prediksi dari berbagai pohon keputusan secara acak, sehingga mampu mengatasi masalah overfitting dan meningkatkan akurasi prediksi [8].

Salah satu tantangan besar dalam membangun model prediksi di bidang medis adalah adanya ketimpangan distribusi data antar kelas (class imbalance), di mana salah satu kelas memiliki jumlah sampel yang jauh lebih besar dibandingkan kelas lainnya. Kondisi ini dapat membuat model lebih memihak pada kelas yang dominan, sehingga mengurangi kemampuan model dalam mengklasifikasikan kelas dengan data yang lebih sedikit. Untuk mengatasi

masalah ini, dapat digunakan metode oversampling, yaitu dengan menambahkan jumlah data pada kelas minoritas agar proporsi data antar kelas menjadi seimbang. [9] .

Berbagai penelitian sebelumnya menunjukkan efektivitas teknik resampling dalam meningkatkan performa klasifikasi. Penelitian oleh Syukron et al. [10] membuktikan bahwa penerapan SMOTE pada model Random Forest dalam prediksi gagal jantung meningkatkan akurasi hingga 88,1% dengan AUC sebesar 94,7%. Dalam studi Ghozali et al. [11], perbandingan SVM dan Random Forest untuk klasifikasi diabetes menunjukkan bahwa RF menghasilkan akurasi lebih tinggi yaitu 98%, dibanding SVM yang mencapai 92%. Sementara itu, penelitian Faruqziddan et al. [12] pada klasifikasi risiko kambuh kanker tiroid menunjukkan bahwa penggunaan Oversampling meningkatkan akurasi SVM dari 88% menjadi 91%.

II. KAJIAN TEORI

A. *Hipertensi*

Hipertensi yaitu salah satu penyakit yang termasuk penyakit tidak menular paling umum terjadi pada masalah kesehatan dan penyebab kematian tertinggi secara global karena frekuensinya yang relatif tinggi dan meningkat terus-menerus [13]. Hipertensi sering tidak disadari karena tidak adanya tanda – tanda ataupun gejala yang berpotensi mengakibatkan masalah penyakit yang serius. Pada umumnya, penderita hipertensi akan menyadari dirinya terkena hipertensi apabila gejala yang dialami sudah tidak bisa diatasi oleh diri sendiri. Biasanya gejala yang kerap muncul meliputi pusing, lemas, sakit kepala, mual, muntah, hingga bisa terjadi penurunan kesadaran sehingga penyakit ini dapat menjadi komplikasi atau masalah serius yang berakibat fatal [14].

Faktor-faktor yang mungkin menyebabkan diabetes diantaranya sebagai berikut berikut [15]:

1. Usia: Menurut penelitian Sudarso mendeskripsikan bahwa pertambahan usia merupakan salah satu penyebab timbulnya hipertensi akibat adanya penurunan elastisitas pada dinding arteri serta meningkatnya kekakuan pembuluh darah sistemik sebagai dampak dari proses penuaan. Berdasarkan data penelitian menunjukkan bahwa jumlah kasus hipertensi paling tinggi berada di usia 56-60 tahun (29,7%). Hal tersebut menjelaskan bahwa semakin usia seseorang bertambah maka peluang terjadinya hipertensi semakin besar juga.
2. Jenis Kelamin: Prevalensi hipertensi terjadi lebih tinggi pada perempuan, terutama pada kelompok usia dewasa tua dan lansia. Salah satu faktor yang mendasarinya adalah penurunan kadar hormon estrogen pada perempuan menjelang menopause, yang umumnya dimulai sekitar usia 45-55 tahun.
3. Faktor Genetik: Faktor genetik atau riwayat keluarga menjadi resiko terjadinya hipertensi, seperti salah satu orang tua yang terkena hipertensi dapat turun ke anaknya. Hal ini menyatakan bahwa tekanan darah orang tua dapat turun ke anaknya.
4. Aktifitas fisik: Aktivitas fisik ataupun jasmani sangat berpengaruh terhadap keseimbangan tekanan darah. Aktivitas fisik yang ketidakteraturan dapat memicu terjadinya hipertensi.
5. Obesitas
   Kenaikan berat badan dapat menyebabkan peningkatan tekanan darah. Semakin tinggi berat badan seseorang, maka semakin besar volume darah yang dibutuhkan untuk mengalirkan oksigen ke seluruh bagian tubuh. Darah yang menyebar melewati pembuluh darah mengakibatkan tekanan darah yang meningkat.

B. Support Vector Machine

Support Vector Machine merupakan algoritma pembelajaran mesin yang digunakan dalam tugas klasifikasi dan regresi, serta dikenal efektif dalam mengolah data berdimensi tinggi. Prinsip kerja SVM adalah dengan menentukan hyperplane terbaik yang dapat memisahkan dua kelas data secara optimal dengan margin maksimal. [16] [17]. Titik-titik data yang paling dekat dengan hyperplane dinamakan support vectors, yang berperan penting dalam menentukan posisi dan orientasi hyperplane. Pada kasus data non-linear, SVM memanfaatkan fungsi kernel seperti Radial Basis Function (RBF) untuk mentransformasikan data ke ruang berdimensi lebih tinggi sehingga memungkinkan pemisahan secara linear. [17].

C. Random Forest

Random Forest adalah algoritma ensemble berbasis pohon keputusan yang menyatukan output dari sejumlah pohon untuk meningkatkan ketepatan prediksi dan meminimalkan overfitting. Algoritma ini menggunakan pendekatan bagging (bootstrap aggregating), di mana masing-masing pohon dilatih menggunakan subset data yang dipilih secara acak [18] [19]. Hasil akhir prediksi diperoleh melalui voting mayoritas dari seluruh pohon. Random Forest cocok digunakan untuk klasifikasi data kompleks, termasuk dalam dunia medis [19].

D. *Confussion Matrix*

Confusion Matrix adalah proses yang diterapkan untuk menganalisis ketepatan model klasifikasi dalam mengidentifikasi data dengan berbagai kelas. Dengan mengukur tingkat keakurasian dari data maka dapat diketahui performa dari suatu model klasifikasi yang telah dibuat diketahui [20]

Nilai Aktual	Prediksi *Positive*	Prediksi *Negative*
Positive	*True Positive (TP)*	*False Negative (FN)*
Negative	*False Positive (FP)*	*True Negative (TN)*

Keterangan :
True Positive (TP) = memiliki nilai prediksi yang benar positif
False Positive (FP) = memiliki nilai salah positif
False Negative (FN) = memiliki nilai salah negative
True Negative (TN) = memiliki nilai prediksi benar negative

Berikut adalah rumus Confusion Matrix guna menghitung Accuracy, Precision, Recall, dan F1-Score [21]

a. $Accuracy = \frac{TP+TN}{TOTAL}$ (1)

b. $Precision = \frac{TP}{TP+FP}$ (2)

c. $Recall = \frac{TP}{TP+FN}$ (3)

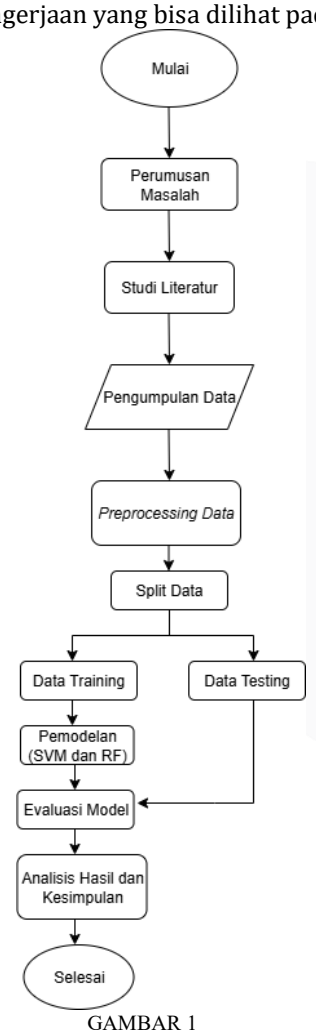
d. $F1 - score = \frac{(2 \times recall \times Precision)}{Recall+Precision}$ (4)

E. Random Over Sampling

Random Oversampling adalah salah satu metode yang diterapkan dalam mengatasi dataset dengan persebaran kelas yang tidak seimbang. Melalui metode ini, data dari kelas minoritas diperbanyak dengan cara menggandakan secara acak. Tujuan dari pendekatan ini yaitu untuk menyamakan jumlah data antara kelas mayoritas dan minoritas, sehingga model machine learning dapat mempelajari kedua kelas secara proporsional selama proses pelatihan [9]

III. METODE

Penelitian ini melakukan klasifikasi terhadap dataset Hipertensi yang berasal dari dataset campuran yang akan dilakukan pengujian terhadap seberapa besar ketepatan model dalam melakukan klasifikasi terdapat data. Pengujian model dilakukan berdasarkan penggunaan algoritma *Support Vector Machine* dan *Random Forest*. dengan alur pengerjaan yang bisa dilihat pada gambar .



GAMBAR 1
(ALUR PENELITIAN)

1. Pengumpulan data

Proses awal dari penelitian ini dilakukan dengan mengumpulkan dataset pasien hipertensi dari Puskesmas Purwokerto Timur 1. Dataset tersebut memuat 7 fitur dan terbagi dalam 11 kelas diagnosis, dengan total sebanyak 1.209 data [22].

2. *Preprocessing*

Proses *preprocessing* data ini dilakukan cleaning data untuk mengatasi data yang hilang atau kosong, normalisasi data, dan class balancing [34] [48].

3. Split Data

Langkah selanjutnya yaitu membagi data menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan untuk membangun serta melatih model, sementara data uji digunakan untuk mengevaluasi performa model dalam melakukan prediksi. Proporsi pembagian data ditetapkan sebesar 80% untuk pelatihan dan 20% untuk pengujian. [23].

4. Pemodelan

Penelitian ini melakukan pemodelan menggunakan dua algoritma machine learning, yaitu Support Vector Machine (SVM) dan Random Forest (RF), dengan tujuan utama membandingkan performa keduanya dalam tugas klasifikasi [44]. Algoritma SVM menerapkan kernel Radial Basis Function (RBF) untuk mengatasi data yang tidak dapat dipisahkan secara linear, sementara Random Forest memanfaatkan 100 pohon keputusan guna meningkatkan akurasi serta mengurangi risiko overfitting. Kedua model dilatih menggunakan data latih sebanyak 80% dari total data yang telah diseimbangkan sebelumnya menggunakan teknik Random Oversampling, dan diuji dengan sisa 20% data.

5. Evaluasi Model

Tahapan berikutnya adalah evaluasi, di mana peneliti menilai hasil klasifikasi dengan tujuan mengukur tingkat akurasi dari algoritma Support Vector Machine dan Random Forest. Penilaian diterapkan dengan menggunakan confusion matrix sebagai alat ukur performa model dalam memprediksi hipertensi [23].

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan data

Tahap awal penelitian dilakukan yaitu pengumpulan dataset. Pada gambar 2 merupakan hasil dari pengumpulan dataset yang akan digunakan. (Hasil pengambilan data tersebut mendapatkan total 1209 kolom dengan 7 fitur dan terbagi dalam 11 kelas diagnosis.

NO	USIA	KATEGORI	JENIS KELAMIN	TB	BB	SISTOLE	DIASTOLE	DIAGNOSA
1	57	Dewasa	Laki-laki	150	58	158	93	Cerebral infarction, unspecified
2	60	Lansia	Perempuan	150	43	120	67	Other cerebral infarction
3	62	Lansia	Laki-laki	165	59	119	81	Essential (primary) hypertension
4	64	Lansia	Laki-laki	147	45	122	70	Hypertensive heart disease
5	40	Dewasa	Laki-laki	162	66	152	103	Cerebral infarction, unspecified
6	54	Dewasa	Perempuan	146	79	160	80	Essential (primary) hypertension
7	81	Lansia	Perempuan	165	58	117	85	Essential (primary) hypertension
8	69	Lansia	Laki-laki	170	50	179	106	Essential (primary) hypertension
9	60	Lansia	Perempuan	151	51	161	94	Cerebral infarction, unspecified
10	78	Lansia	Laki-laki	159	46	190	118	Essential (primary) hypertension

GAMBAR 2
(DATASET HIPERTENSI)

B. Preprocessing

1. Cleaning Data

Pada tahap ini dilakukan hasil pemeriksaan missing value pada setiap kolom dataset.

```
Jumlah missing value per kolom:
USIA             0
KATEGORI         0
JENIS KELAMIN    0
TB               0
BB               0
SISTOLE          0
DIASTOLE         0
DIAGNOSA         0
dtype: int64

Process finished with exit code 0
```

GAMBAR 3
(CLEANING DATA)

Terlihat bahwa tidak ada data yang hilang (nilai 0) di semua kolom seperti Usia, Kategori, Jenis Kelamin, TB, BB, Sistole, Diastole, dan Diagnosa, sehingga data siap digunakan untuk tahap selanjutnya.

2. Normalisasi Data

Pada tahap ini dilakukan hasil transformasi data menggunakan StandardScaler, yaitu teknik normalisasi yang mengubah setiap fitur menjadi skala standar dengan rata-rata 0 dan standar deviasi 1.

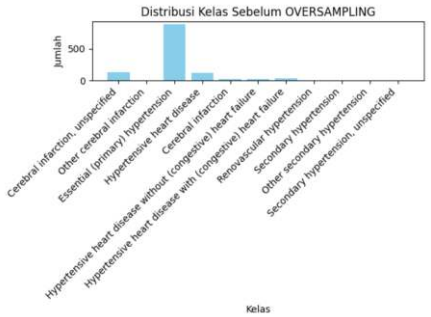
```
Contoh hasil StandardScaler (X_train_scaled):
       USIA  KATEGORI  JENIS KELAMIN        TB        BB   SISTOLE  DIASTOLE
0 -1.020166 -1.083135       0.768414 -0.173102  0.181870 -0.081897  1.290723
1  1.286089  0.908104      -1.301381  0.480338  0.181870  0.358737  0.098053
2 -0.031771  0.908104      -1.301381 -1.153262 -0.044473 -0.522531 -0.796449
3  0.380060  0.908104       0.768414 -1.153262 -1.176185 -0.682761 -0.200114
4 -1.843828 -1.083135       0.768414  0.480338  1.615372  0.759314  0.843472

Process finished with exit code 0
```

GAMBAR 4
(NORMALISASI DATA)

Nilai-nilai pada kolom seperti USIA, TB, dan lainnya telah dinormalisasi agar memiliki distribusi yang seragam, memudahkan proses pelatihan model machine learning.

3. Class Balancing

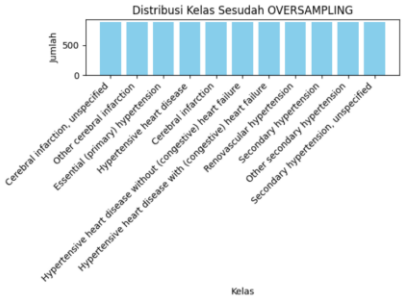
Pada tahap ini dilakukan distribusi jumlah data pada masing-masing kelas diagnosis sebelum dilakukan proses oversampling.



GAMBAR 5
(CLASS BALANCING SEBELUM OVERSAMPLING)

Terlihat bahwa kelas Essential (primary) hypertension mendominasi data, sementara kelas lainnya memiliki jumlah sampel yang jauh lebih sedikit. Hal ini menunjukkan adanya ketidakseimbangan kelas yang perlu diatasi agar model pembelajaran mesin tidak bias terhadap kelas mayoritas. Pada penelitian ini menggunakan metode oversampling untuk mengatasi ketidakseimbangan kelas.

Pada tahap ini dilakukan distribusi kelas pada dataset setelah dilakukan proses oversampling.



GAMBAR 6
(CLASS BALANCING SESUDAH OVERSAMPLING)

Terlihat bahwa seluruh kelas diagnosis hipertensi telah memiliki jumlah data yang seimbang, yaitu masing-masing sebanyak 881 sampel. Hal ini menandakan bahwa teknik oversampling berhasil diterapkan untuk mengatasi ketidakseimbangan data, sehingga model yang dibangun nantinya tidak akan condong atau bias terhadap kelas mayoritas saja. Keseimbangan ini penting agar algoritma dapat belajar secara adil terhadap semua kelas yang ada.

C. Split Data

Di tahap ini, data yang sudah melalui proses oversampling dibagi menjadi dua bagian, yakni 80% sebagai data pelatihan dan 20% sebagai data pengujian. Data pelatihan berfungsi untuk melatih model, sedangkan data pengujian digunakan untuk menilai kinerja model terhadap data yang belum pernah dikenali sebelumnya. Dari hasil oversampling diperoleh total 9.691 data, dengan rincian 7.752 data dipakai untuk pelatihan dan 1.939 data dipakai untuk pengujian.

D. Pemodelan

Proses pemodelan dilakukan dengan menggunakan dua algoritma, yaitu Support Vector Machine (SVM) dan Random Forest (RF). Model SVM dibangun menggunakan kernel Radial Basis Function (RBF) dengan parameter default dari pustaka scikit-learn, sedangkan model Random Forest menggunakan 100 pohon keputusan dengan nilai random_state ditetapkan agar hasil dapat direproduksi. Sebelum pelatihan, data latih diseimbangkan menggunakan metode Random Oversampling untuk memastikan bahwa distribusi kelas seimbang. Kedua model dilatih menggunakan 80% data, sedangkan 20% sisanya digunakan sebagai data uji.

E. Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja model dalam memprediksi kelas hipertensi. Metrik yang digunakan meliputi akurasi, presisi, recall, dan F1-score, yang dihitung dari confusion matrix. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki performa yang lebih baik dibandingkan SVM. Model Random Forest menghasilkan akurasi sebesar 98,92% sedangkan Model SVM menghasilkan akurasi sebesar 83,91%. Performa tinggi yang ditunjukkan oleh Random Forest menunjukkan bahwa metode ini lebih mampu menangani data medis yang kompleks dan tidak seimbang. Hal ini konsisten dengan hasil beberapa penelitian terdahulu yang juga menyimpulkan bahwa Random Forest memiliki keunggulan dalam klasifikasi data medis setelah dilakukan resampling.

Berikut merupakan tabel perbandingan SVM dan Random Forest:

	Diagnosa	*SVM*	*RF*
Accuracy		83,91%	98,92%
Precision	Cerebral infarction	0.70	1.00
	Cerebral infarction, unspecified	0.66	0.92
	Essential (primary) hypertension	0.68	1.00
	Hypertensive heart disease	0.69	1.00
	Hypertensive heart disease with (congestive) heart failure	0.81	0.99
	Hypertensive heart disease without (congestive) heart failure	0.78	1.00
	Other cerebral infarction	0.99	1.00
	Other secondary hypertension	1.00	1.00
	Renovascular hypertension	0.98	0.99
	Secondary hypertension	0.89	1.00
	Secondary hypertension, unspecified	1.00	1.00
Recall	Cerebral infarction	0.97	1.00
	Cerebral infarction, unspecified	0.43	1.00
	Essential (primary) hypertension	0.55	0.89
	Hypertensive heart disease	0.55	1.00
	Hypertensive heart disease with (congestive) heart failure	0.85	1.00
	Hypertensive heart disease without (congestive) heart failure	0.91	1.00
	Other cerebral infarction	1.00	1.00
	Other secondary hypertension	1.00	1.00
	Renovascular hypertension	1.00	1.00
	Secondary hypertension	1.00	1.00
	Secondary hypertension, unspecified	1.00	1.00
F1-Score	Cerebral infarction	0.81	1.00
	Cerebral infarction, unspecified	0.52	0.96
	Essential (primary) hypertension	0.61	0.94
	Hypertensive heart disease	0.61	1.00
	Hypertensive heart disease with (congestive) heart failure	0.83	1.00
	Hypertensive heart disease without (congestive) heart failure	0.84	1.00
	Other cerebral infarction	1.00	1.00
	Other secondary hypertension	1.00	1.00
	Renovascular hypertension	0.99	1.00
	Secondary hypertension	0.94	1.00
	Secondary hypertension, unspecified	1.00	1.00

Dari tabel perbandingan ini, dapat disimpulkan bahwa model Random Forest secara konsisten mengungguli model SVM di hampir semua kategori diagnosa hipertensi. Random Forest menunjukkan kemampuan yang luar biasa dalam mencapai presisi, recall, dan F1-score yang sangat tinggi, bahkan sempurna, di sebagian besar jenis hipertensi. Hal ini menunjukkan bahwa Random Forest mempunyai kemampuan generalisasi yang lebih unggul dalam mengklasifikasikan beragam kelas hipertensi yang terdapat dalam dataset. Meskipun SVM menunjukkan kinerja yang sangat baik pada beberapa diagnosa spesifik (misalnya, "Other secondary hypertension" dan "Secondary hypertension, unspecified"), kinerja keseluruhannya cenderung lebih bervariasi dan tidak seoptimal Random Forest, terutama pada diagnosa yang lebih kompleks atau bervariasi. Oleh karena itu, berdasarkan metrik-metrik per diagnosa ini, Random Forest merupakan pilihan model yang lebih baik dalam prediksi hipertensi.

V. KESIMPULAN

Berdasarkan hasil Penelitian bahwa Algoritma Support Vector Machine (SVM) mampu memberikan hasil klasifikasi yang cukup baik dengan akurasi 83,91%, meskipun kurang stabil pada kelas dengan jumlah data yang sedikit. Sebaliknya, algoritma Random Forest (RF) menunjukkan performa yang lebih optimal dan konsisten dengan akurasi 98,92%, serta nilai presisi, recall, dan F1-score yang tinggi.

Secara keseluruhan, Random Forest lebih unggul dibandingkan SVM dalam memprediksi hipertensi pada data yang tidak seimbang.

REFERENSI

[1] R. Putra Pridiptama, W. Wasono, F. Deny, and T. Amijaya, "Perbandingan Algoritma Support Vector Machine dan Na¨ıveNa¨ıve Bayes pada Klasifikasi Penyakit Tekanan Darah Tinggi (Studi Kasus: Klinik Polresta Samarinda)," 2024. [Online]. Available: http://jurnal.fmipa.unmul.ac.id/index.php/Basis

[2] I. L. Maria, A. S. Yusnitasari, C. Lifoia, A. Tiara Aurelia Annisa, and S. Mulyani, "Determinants of Hypertension Incidence in the Work Areas of the Bone and Barru District Health Centers in 2022," *Media Kesehatan Masyarakat Indonesia*, vol. 18, no. 3, pp. 83–89, Sep. 2022.

[3] M. A. Aprianur, "Peran Artificial intelligence Dalam Deteksi Penyakit Hipertensi: Sistematik Review," 2022. [Online]. Available: https://www.researchgate.net/publication/377086205

[4] K. Abdul Khalim, U. Hayati, and A. Bahtiar, "Perbandingan Prediksi Penyakit Hipertensi Menggunakan Metode Random Forest dan Naive Bayes," 2023.

[5] A. Wulandari, S. Atika Sari, and Ludiana, "Penerapan Relaksasi Benson Terhadap Tekanan Darah pada Pasien Hipertensi di RSUD Jendral Ahmad Yani Kota Metro Tahun 2022," *Jurnal Cendikia Muda*, vol. 3, no. 2, 2023.

[6] P. Purwono, P. Dewi, S. K. Wibisono, and B. P. Dewa, "Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network," *Insect (Informatics and Security): Jurnal Teknik Informatika*, vol. 7, no. 2, pp. 82–90, Mar. 2022.

[7] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA Jurnal*, vol. 10, no. 02, 2020.

[8] L. Sari, A. Romadloni, and R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," *Infotekmesin*, vol. 14, no. 1, pp. 155–162, Jan. 2023.

[9] H. Hasbi and T. B. Sasongko, "Optimasi Performa Random Forest dengan Random Oversampling dan SMOTE pada Dataset Diabetes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, p. 1756, Jul. 2024, doi: 10.30865/mib.v8i3.7855.

[10] A. Syukron, E. Saputro, and P. Widodo, "Penerapan Metode Smote Untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Gagal Jantung," 2023. [Online]. Available: https://doi.org/10/25047/jtit.v10i1.312

[11] A. Ghozali, H. Pratiwi, S. Sulistijowati Handajani Program Studi Statistika, F. Matematika dan Ilmu Pengetahuan Alam, and U. Sebelas Maret, "IMPLEMENTASI DATA MINING MENGGUNAKAN METODE RANDOM FOREST DAN SUPPORT VECTOR MACHINE DALAM KLASIFIKASI PENYAKIT DIABETES 1)," Bulan Juli Hal. [Online]. Available: https://jurnal.unikal.ac.id/index.php/Delta/index

[12] P. Budi Utomo, M. Faruqziddan, E. Herdika Septa Aulia, and S. Dini Azzahra, "Perbandingan Skenario Balancing Oversampling dan Undersampling dalam Klasifikasi Resiko Kambuh Kanker Tiroid menggunakan Algoritma SVM Linear," *JAMI: Jurnal Ahli Muda Indonesia*, vol. 5, no. 2, pp. 172–182, Dec. 2024, doi: 10.46510/jami.v5i2.320.

[13] A. A. Anggraini, V. S. Putri, and Z. Nuranti, "Pengaruh Pendidikan Kesehatan dan Pemberian Daun Seledri pada Pasien dengan Hipertensi di Wilayah RT 10 Kelurahan Murni," *Jurnal Abdimas Kesehatan (JAK)*, vol. 2, no. 1, p. 30, Jan. 2020, doi: 10.36565/jak.v2i1.89.

[14] S. C. Nurzanah, S. Alam, and T. I. Hermanto, "Analisis Association Rule untuk Identifikasi Pola Gejala Penyakit Hipertensi Menggunakan Algoritma Apriori (Studi kasus : Klinik rafina Medical Center)," *Jurnal Informatika dan Komputer*, vol. 5, no. 2, 2022.

[15] Wiwin Vidiyastana Afifah, Irfansyah Baharuddin Pakki, and Tanti Asrianti, "Analisis Faktor Risiko Kejadian Hipertensi Pada Lansia Di Wilayah Kerja Puskesmas Rapak Mahang Kecamatan Tenggarong Kabupaten Kutai Kartanegara," *Wal'afiat Hospital Journal*, vol. 3, no. 1, 2022.

[16] H. Apriyani, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," 2020. [Online]. Available: https://journal-computing.org/index.php/journal-ita/index

[17] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 640, Apr. 2021.

[18] S. Budiman, A. Sunyoto, and A. Nasiri, "Analisa Performa Penggunaan Feature Selection untuk Mendeteksi Intrusion Detection Systems dengan Algoritma Random Forest Classifier," 2021. [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[19] N. Hadi and J. Benedict, "Implementasi Machine Learning untuk Prediksi Harga Rumah Menggunakan Algoritma Random Forest," 2024. [Online]. Available: https://www.kaggle.com/harlfoxem/housesalesprediction

[20] Y. Afrillia, L. Rosnita, and D. Siska, "Analisis Sentimen Ciutan Twitter Terkait Penerapan Permendikbudristek Nomor 30 Tahun 2021 Menggunakan TextBlob dan Support Vector Machine," *G-Tech: Jurnal Teknologi Terapan*, vol. 6, no. 2, pp. 387–394, Oct. 2022.

[21] D. Ismafillah, T. Rohana, and Y. Cahyana, "Implementasi Model Support Vector Machine dan

Logistic Regression Untuk Memprediksi Penyakit Stroke," *Jurnal Riset Komputer)*, vol. 10, no. 1, pp. 2407–389, 2023, [Online]. Available: http://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

[22] F. Refindha, A. Harianto, Z. Alawi, and I. A. Sa'ida, "PENGARUH KOMPOSISI SPLIT DATA PADA AKURASI KLASIFIKASI PENDERITA DIABETES MENGGUNAKAN ALGORITMA MACHINE LEARNING," *Jurnal Sistem Informasi dan Informatika (Simika)*, vol. 8, no. 1, 2025.

[23] M. Dava Maulana *et al.*, "ALGORITMA XGBOOST UNTUK KLASIFIKASI KUALITAS AIR MINUM," 2023. [Online]. Available: https://www.kaggle.com/datasets/adityak