

**KLASIFIKASI PENYAKIT BERBASIS BAGGING
MENGUNAKAN *NAÏVE BAYES*, *DECISION TREE*, DAN
*SUPPORT VECTOR MACHINE***

Diajukan Sebagai Syarat Untuk Menyelesaikan Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Jonathan Immanuel
NIM: 09021282126098

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024**

**KLASIFIKASI PENYAKIT BERBASIS *BAGGING*
MENGUNAKAN *NAÏVE BAYES*, *DECISION TREE*, DAN
*SUPPORT VECTOR MACHINE***

*Diajukan Sebagai Syarat Untuk Menyelesaikan Pendidikan Program Strata-1
Pada Jurusan Teknik Informatika*



Oleh:

Jonathan Immanuel
NIM: 09021282126098

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024**

LEMBAR PENGESAHAN SKRIPSI

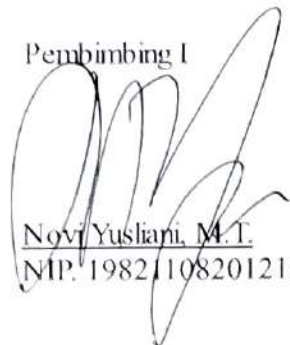
KLASIFIKASI PENYAKIT BERBASIS *BAGGING* DENGAN
NAÏVE BAYES, *DECISION TREE*, DAN *SUPPORT VECTOR*
MACHINE

Oleh:

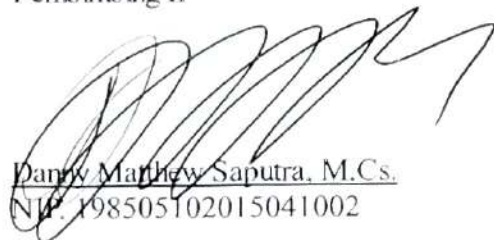
Jonathan Immanuel
NIM: 09021282126098

Palembang, 31 Desember 2024

Pembimbing I


Novi Yusliani, M.T.
NIP. 198211082012122001

Pembimbing II


Danny Matthew Saputra, M.Cs.
NIP. 198505102015041002

Mengetahui,
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERSETUJUAN

Telah dipublikasi di Prosiding Konferensi Bereputasi Internasional *International Conference on Electrical Engineering and Computer Science (ICECOS)* pada:

Hari : Jum'at
Tanggal : 20 Desember 2024

Nama : Jonathan Immanuel
NIM : 09021282126098
Judul : *Bagging-Based Disease Prediction Using Naïve Bayes, Decision Tree, and Support Vector Machine on Kaggle Dataset*

1. Penguji I

Dian Palupi Rini, Ph.D
NIP. 197802232006042002


.....

2. Penguji II

Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003


.....

3. Pembimbing I

Novi Yusliani, M.T.
NIP. 198211082012122001


.....

4. Pembimbing II

Danny Matthew Saputra, M.Cs.
NIP. 198505102015041002


.....

Mengetahui
Ketua Jurusan Teknik Informatika,



Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

PERNYATAAN INTEGRITAS

Yang bertanda tangan dibawah ini:

Nama : Jonathan Immanuel

NIM : 09021282126098

Judul : Klasifikasi Penyakit Berbasis Bagging Dengan Naïve Bayes,
Decision Tree, Dan Support Vector Machine

Hasil Pengecekan Software iThenticate/Turnitin : 8%

Menyatakan bahwa Skripsi saya merupakan hasil karya sendiri didampingi tim pembimbing dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam Skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai aturan yang berlaku.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa ada paksaan dari siapapun.

Palembang, 31 Desember 2024



Jonathan Immanuel

NIM. 09021282126098

ABSTRACT

In classification, finding the optimal model to handle a specific problem is crucial. Various algorithms, such as *Naïve Bayes*, Decision Tree, and Support Vector Machines (SVM), each have their own strengths and weaknesses. One commonly used technique to enhance model performance is Bagging, the ensemble technique. Bagging combines weak models into a stronger model by reducing bias and variance. This research explores the application of the Bagging aggregation technique using *Naïve Bayes*, Decision Tree, and SVM for disease classification in the Multiple Disease Prediction dataset obtained from Kaggle. The purpose is to develop a Bagging-based classification system that can improve model performance and demonstrate an alternative method for enhancing model performance. After preprocessing the dataset, it was found that the model's performance was subpar, so the training and testing sets were combined, shuffled, and split again to provide optimal conditions. The results show that Bagging improves model performance, especially the Bagged Decision Tree model, achieving the highest F1-Score of 0.971. However, other models did not show the same results. Models Bagged with *Naïve Bayes* and SVM showed improvement compared to the normal models, but the performance of the superior algorithm, Decision Tree, decreased, making the overall performance suboptimal. The largest performance drop was observed in the Bagged *Naïve Bayes* and SVM model, with a score of 0.625. The findings indicate that the Bagging-based classification system has been successfully implemented, but clear results regarding the improvement of model performance using the Bagging aggregation technique have not yet been obtained.

Keywords: Classification, Bagging, *Naïve Bayes*, Decision Tree, Support Vector Machine

ABSTRAK

Dalam klasifikasi, pencarian model yang optimal untuk menangani suatu masalah tertentu sangatlah penting. Berbagai algoritma klasifikasi, seperti *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machines* (SVM), masing-masing memiliki kelebihan dan kelemahan tersendiri. Salah satu teknik yang umum digunakan untuk meningkatkan kinerja model adalah teknik ensemble *Bagging*. *Bagging* dapat menggabungkan model-model yang lemah menjadi satu model yang lebih baik dengan menurunkan bias dan varians. Penelitian ini akan menelusuri penerapan teknik agregasi *Bagging* menggunakan *Naïve Bayes*, *Decision Tree*, dan SVM untuk klasifikasi penyakit dalam dataset *Multiple Disease Prediction* yang didapatkan dari Kaggle. Tujuannya adalah untuk mengembangkan sistem klasifikasi berbasis *Bagging* yang dapat meningkatkan kinerja model serta menunjukkan metode alternatif dalam peningkatan performa model. Setelah praproses dataset, ditemukan bahwa kinerja model kurang efektif sehingga set pelatihan dan pengujian digabungkan, diacak, dan dibagi kembali untuk memberikan kondisi yang optimal. Hasilnya menunjukkan bahwa *Bagging* meningkatkan performa model, terutama model *Bagged Decision Tree*, yang mencapai *F1-Score* tertinggi sebesar 0,971. Namun, model-model lainnya tidak menunjukkan hasil yang sama. Model yang di-*Bagging* dengan *Naïve Bayes* dan SVM menunjukkan peningkatan dibandingkan dengan model yang normal, tetapi kinerja algoritma yang unggul, yaitu *Decision Tree*, menurun sehingga performa model tidak maksimal. Penurunan performa terbesar tampak pada model *Bagged Naïve Bayes* dan SVM dengan skor 0,625. Hasil yang didapatkan menunjukkan sistem klasifikasi berbasis *Bagging* telah berhasil dilakukan, namun belum mendapatkan hasil yang jelas mengenai peningkatan performa model menggunakan teknik agregasi *Bagging*.

Kata Kunci: Klasifikasi, *Bagging*, *Naïve Bayes*, *Decision Tree*, *Support Vector Machine*

KATA PENGANTAR

Segala puji syukur kepada Tuhan atas segala karunia, rahmat, dan kasih sayang-Nya sehingga penulis dapat menyelesaikan skripsi dengan baik dan lancar. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan program Strata-1 (S1) pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Selama penyusunan dan penulisan tugas akhir ini, penulis menyadari bahwa tidak lepas dari bimbingan, bantuan, dan dukungan dari berbagai pihak sehingga akhirnya segala hambatan dapat diatasi dengan baik. Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Allah yang telah memberikan rahmat dan anugerah, berupa ilmu yang berlimpah, kesempatan, dan kesehatan jasmani dan rohani sehingga penulis dapat melaksanakan dan menyelesaikan tugas akhir.
2. Kedua Orang tua dan kedua adik-adik yang senantiasa mendukung, mendoakan, dan memberikan semangat untuk melakukan yang terbaik.
3. Bapak Prof. Dr. Erwin, S. Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Hadipurnawan Satria, Ph.D. selaku Ketua Jurusan Teknik Informatika.
5. Bapak Osvari A., S.Kom., M.T. selaku Sekretaris Jurusan Teknik Informatika yang telah membimbing penulis dalam menyelesaikan skripsi ini.
6. Ibu Novi Yusliani, M.T., selaku Dosen Pembimbing I dan Bapak Danny Matthew Saputra, M.Cs., selaku Dosen Pembimbing II yang telah bersedia meluangkan waktu dan tenaga serta memberikan bimbingan, saran, arahan, serta ilmu yang bermanfaat selama proses penulisan tugas akhir.
7. Ibu Dian Palupi Rini, Ph.D., selaku Dosen Penguji I dan Ibu Alvi Syahrini Utami, M.Kom., selaku Dosen Penguji II yang telah bersedia meluangkan waktu dan tenaga untuk melakukan ujian komprehensif serta memberikan bimbingan dan saran selama masa revisi tugas akhir.
8. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmunya kepada penulis selama proses perkuliahan.

9. Seluruh staf jurusan Teknik Informatika yang telah membantu memastikan kelancaran proses administrasi dan akademik selama masa perkuliahan.
10. Seluruh rekan-rekan kelas dan semua pihak yang terlibat dalam proses pembuatan skripsi ini yang telah menghabiskan waktu, menghibur, memberikan bantuan, memotivasi, dan berjuang bersama penulis semasa perkuliahan.

Penulis sadar bahwa masih ada banyak kekurangan dalam penyusunan tugas akhir ini yang disebabkan keterbatasan pengetahuan dan pengalaman, maka dari itu penulis sangat mengharapkan kritik dan saran yang konstruktif dapat menjadi masukan untuk penelitian kedepannya. Semoga tugas akhir ini dapat bermanfaat.

Palembang, Desember 2024
Penulis

Jonathan Immanuel

DAFTAR ISI

Halaman

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	ii
HALAMAN PERSETUJUAN.....	iii
PERNYATAAN INTEGRITAS	iv
ABSTRACT.....	v
ABSTRAK.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	I-1
1.1. Pendahuluan.....	I-1
1.2. Latar Belakang	I-1
1.3. Rumusan Masalah.....	I-3
1.4. Tujuan	I-4
1.5. Manfaat	I-4
1.6. Batasan Masalah	I-4
1.7. Sistematika Penulisan	I-5
1.8. Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-1
2.1. Pendahuluan.....	II-1
2.2. Landasan Teori.....	II-1
2.2.1. <i>Bagging (Bootstrap Aggregating)</i>	II-1
2.2.2. <i>Naïve Bayes</i>	II-2
2.2.3. <i>Decision Tree</i>	II-4
2.2.4. <i>Support Vector Machine (SVM)</i>	II-5
2.2.5. <i>Confusion Matrix</i>	II-7
2.3. Penelitian Lain yang Relevan	II-9
2.4. Kesimpulan	II-10

BAB III	METODE PENELITIAN	III-1
3.1.	Pendahuluan	III-1
3.2.	Pengumpulan Data	III-1
3.2.1.	Jenis dan Sumber Data	III-1
3.2.2.	Metode Pengumpulan Data	III-1
3.3.	Tahapan Penelitian	III-2
3.3.1.	Pengumpulan Data.....	III-3
3.3.2.	Perancangan Kerangka Kerja Penelitian	III-3
3.3.3.	Penentuan Kriteria Pengujian	III-5
3.3.4.	Penentuan Format Data Pengujian	III-5
3.3.5.	Penentuan Alat Bantu Penelitian	III-5
3.3.6.	Pengujian Penelitian	III-7
3.3.7.	Analisis Pengujian dan Penyimpulan	III-7
3.4.	Metode Pengembangan Perangkat Lunak.....	III-7
3.5.	Manajemen Proyek Penelitian	III-8
3.6.	Kesimpulan	III-9
BAB IV	PENGEMBANGAN PERANGKAT LUNAK	IV-1
4.1.	Pendahuluan	IV-1
4.2.	<i>Iterative Enhancement</i>	IV-1
4.3.	Eksplorasi dan Implementasi Sederhana	IV-1
4.3.1.	Eksplorasi	IV-2
4.3.2.	Penyelesaian Masalah-Masalah.....	IV-8
4.4.	Eksperimen dan Pengembangan Model-Model	IV-11
4.5.	Pengujian dan Analisis.....	IV-14
4.6.	Kesimpulan	IV-15
BAB V	HASIL DAN ANALISIS PENELITIAN.....	V-1
5.1.	Pendahuluan	V-1
5.2.	Hasil Penelitian	V-1
5.2.1.	Skenario Pengujian	V-1
5.2.2.	Data Hasil Pengujian Set Pelatihan dengan Duplikat dan Tanpa Duplikat	V-2
5.2.3.	Data Hasil Pengujian Set Pelatihan Pembagian Baru....	V-10
5.3.	Analisis Hasil Penelitian	V-17
5.4.	Kesimpulan	V-18

BAB VI KESIMPULAN DAN SARAN	VI-1
6.1. Kesimpulan	VI-1
6.2. Saran	VI-2
DAFTAR PUSTAKA	xv
DAFTAR LAMPIRAN	xvii

DAFTAR TABEL

Halaman

Tabel III-1. Susunan Model.....	III-4
Tabel III-2. <i>Confusion Matrix</i>	III-5
Tabel III-3. Alat Bantu yang Digunakan dalam Penelitian	III-6
Tabel III-4. <i>F1-Score</i> Model.....	III-7
Tabel III-5. Rencana Jadwal Penelitian	III-8
Tabel IV-1. Hasil Algoritma IV-1.....	IV-2
Tabel IV-2. Nilai Atribut Satu Titik Data Pelatihan yang Diinterpolasi.....	IV-4
Tabel IV-3. Jumlah Nilai Kelas <i>Disease</i> Dataset Pelatihan	IV-6
Tabel IV-4. Jumlah Nilai Kelas <i>Disease</i> Dataset Pengujian.....	IV-7
Tabel IV-5. Jumlah Nilai Kelas <i>Disease</i> Dataset Pelatihan Baru	IV-10
Tabel IV-6. Jumlah Nilai Kelas <i>Disease</i> Dataset Pengujian Baru	IV-11
Tabel IV-7. Konfigurasi Model-Model.....	IV-12
Tabel V-1. <i>Confusion Matrix Decision Tree</i> pada Set Pengujian Awal.....	V-3
Tabel V-2. <i>Confusion Matrix SVM</i> pada Set Pengujian Awal	V-3
Tabel V-3. <i>Confusion Matrix Naïve Bayes</i> pada Set Pengujian Awal	V-4
Tabel V-4. <i>Confusion Matrix Bagged Decision Tree</i> pada Set Pengujian Awal	V-4
Tabel V-5. <i>Confusion Matrix Bagged SVM</i> pada Set Pengujian Awal	V-5
Tabel V-6. <i>Confusion Matrix Bagged Naïve Bayes</i> pada Set Pengujian Awal ..	V-5
Tabel V-7. <i>Confusion Matrix Bagged Decision Tree</i> dan SVM pada Set Pengujian Awal	V-6
Tabel V-8. <i>Confusion Matrix Bagged Decision Tree</i> dan <i>Naïve Bayes</i> pada Set Pengujian Awal	V-7
Tabel V-9. <i>Confusion Matrix Bagged SVM</i> dan <i>Naïve Bayes</i> pada Set Pengujian Awal	V-7
Tabel V-10. <i>Confusion Matrix Bagged Decision Tree</i> , SVM, dan <i>Naïve Bayes</i> pada Set Pengujian Awal	V-8
Tabel V-11. Hasil Pengujian Model-Model Set Pelatihan Tanpa Duplikat	V-8
Tabel V-12. <i>Confusion Matrix Decision Tree</i> pada Set Pengujian Baru	V-10
Tabel V-13. <i>Confusion Matrix SVM</i> pada Set Pengujian Baru	V-11
Tabel V-14. <i>Confusion Matrix Naïve Bayes</i> pada Set Pengujian Baru	V-11
Tabel V-15. <i>Confusion Matrix Bagged Decision Tree</i> pada Set Pengujian Baru.....	V-12
Tabel V-16. <i>Confusion Matrix Bagged SVM</i> pada Set Pengujian Baru	V-12
Tabel V-17. <i>Confusion Matrix Bagged Naïve Bayes</i> pada Set Pengujian Baru ..	V-13

Tabel V-18. <i>Confusion Matrix Bagged Decision Tree</i> dan SVM pada Set Pengujian Baru.....	V-13
Tabel V-19. <i>Confusion Matrix Bagged Decision Tree</i> dan <i>Naïve Bayes</i> pada Set Pengujian Baru.....	V-14
Tabel V-20. <i>Confusion Matrix Bagged SVM</i> dan <i>Naïve Bayes</i> pada Set Pengujian Baru.....	V-14
Tabel V-21. <i>Confusion Matrix Bagged Decision Tree</i> , SVM, dan <i>Naïve Bayes</i> pada Set Pengujian Baru	V-15
Tabel V-22. Hasil Tes Model-Model Set Pelatihan Pembagian Baru.....	V-15

DAFTAR GAMBAR

Halaman

Gambar II-1. Diagram Cara Kerja <i>Bagging</i>	II-1
Gambar II-2. <i>Confusion Matrix</i> dengan Jumlah Marginal	II-7
Gambar III-1. Diagram Tahapan Penelitian	III-2
Gambar III-2. Diagram Arsitektur Sistem	III-3
Gambar IV-1. <i>Source Code</i> Pemuatan Pustaka dan Informasi Dataset	IV-2
Gambar IV-2. <i>Source Code</i> Penjumlahan Nilai Kosong pada Setiap Atribut ..	IV-5
Gambar IV-3. <i>Source Code</i> Penjumlahan Data Duplikat	IV-6
Gambar IV-4. <i>Source Code</i> Penjumlahan Nilai Kelas <i>Disease</i>	IV-6
Gambar IV-5. <i>Source Code</i> Pembersihan Data Duplikat	IV-8
Gambar IV-6. <i>Source Code</i> Pencabutan Baris Data Kelas Ekstra	IV-8
Gambar IV-7. <i>Source Code</i> Penggabungan Dataset	IV-9
Gambar IV-8. <i>Source Code</i> Pembagian Dataset Gabungan	IV-9
Gambar IV-9. <i>Source Code</i> Pembagian Fitur dan Label	IV-10
Gambar IV-10. <i>Source Code</i> <i>Encoding</i> Kelas	IV-11
Gambar IV-11. <i>Source Code</i> Pengembangan Model-Model	IV-12
Gambar IV-12. <i>Source Code</i> Contoh Kode Pelatihan Model	IV-13
Gambar IV-13. <i>Source Code</i> Contoh Kode Pengujian Model	IV-14
Gambar IV-14. <i>Source Code</i> Contoh Keluaran Kode Gambar IV-13	IV-14
Gambar V-1. Graf Tabel Hasil Pengujian Model-Model Set Pelatihan Tanpa Duplikat	V-9
Gambar V-2. Graf Tabel Hasil Pengujian Model-Model Set Pelatihan Pembagian Baru	V-16

BAB I

PENDAHULUAN

1.1. Pendahuluan

Pada Bab ini akan dijelaskan mengenai latar belakang, rumusan masalah, tujuan dan manfaat penelitian, dan batasan masalah. Secara keseluruhan, skripsi ini menjelaskan bagaimana mengembangkan sistem klasifikasi *Bagging* dengan menggunakan algoritma-algoritma *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machines* (SVM) untuk mengklasifikasi dataset *Multiple Disease Prediction* yang tersedia di Kaggle. Metode ini dapat digunakan sebagai alternatif untuk peningkatan kinerja model klasifikasi terhadap beragam dataset.

1.2. Latar Belakang

Dalam ranah klasifikasi, pencarian model yang optimal untuk menangani masalah tertentu sangatlah penting. Berbagai algoritma klasifikasi, seperti *Naïve Bayes*, *Decision Tree*, dan *Support Vector Machines* (SVM), masing-masing datang dengan kelebihan dan kelemahan tersendiri. Adapun juga teknik *Bagging*, teknik *ensemble* yang umum digunakan dalam pembelajaran mesin, dimana metode ini dapat menggabungkan model-model yang lemah menjadi satuan model yang lebih baik. Namun, sebuah ide yang relevan muncul: Algoritma-algoritma klasifikasi, ketika diagregasi, dapat mengalami peningkatan kinerja melalui teknik *ensemble* yang dikenal dengan *Bagging*.

Dalam penelitian sebelumnya, Bauer dan Kohavi (1999) menunjukkan bahwa teknik seperti *Bagging* dan *Boosting* dapat meningkatkan akurasi model dengan mengurangi varians dan bias. Diterapkan *Bagging* pada MC4—sebuah *decision tree inducer* yang dasar dan didapatkan bahwa metode ini secara konsisten

meningkatkan kinerja model terhadap semua dataset, tanpa ada peningkatan eror, dan mencapai pengurangan rata-rata relatif eror sebesar 4%.

Kemudian, Wu, Ma, dan Toe (2023) mengusulkan model jaringan syaraf konvolusional (CNN) yang memanfaatkan metode *ensemble Bagging* dan *stacking*, menghasilkan model *Bagging Stacked Ensemble CNN* (BSECNN), untuk klasifikasi kanker payudara. Hasil penelitian mereka menunjukkan bahwa, dalam klasifikasi biner, pendekatan ini mencapai akurasi sebesar 98,94%, sementara dalam klasifikasi lima kelas, akurasinya mencapai 98,34%. Modelnya juga mengungguli VGG16 dan ResNet-50 dalam hal akurasi, dengan perbedaan sebesar 8,22% dan 6,33% masing-masing.

Teknik *Bagging* sangat efektif dalam meningkatkan kinerja model dengan mengurangi bias dan varians. Proses ini dilakukan dengan membagi data menjadi beberapa sampel, kemudian setiap sampel tersebut digunakan untuk melatih model yang direplikasi sebanyak jumlah sampel. Hasil dari setiap model kemudian diagregasi dengan metode mayoritas atau nilai rata-rata hasil keluaran. (Jafarzadeh et al., 2021).

Kelebihan utama dari *Bagging* adalah kemampuannya untuk meningkatkan stabilitas dan akurasi model. Dengan menggabungkan hasil dari beberapa model, *Bagging* dapat mengurangi risiko *overfitting* dan meningkatkan generalisasi model terhadap data baru. Selain itu, *Bagging* juga dapat menangani data yang bervariasi dan mengurangi pengaruh dari *outlier*, sehingga menghasilkan model yang lebih fleksibel dan andal (Jafarzadeh et al., 2021).

Bagging diterapkan pada algoritma *Naïve Bayes*, *Decision Tree*, dan SVM karena masing-masing algoritma ini memiliki karakteristik yang dapat

dioptimalkan dengan teknik *Bagging*. *Naïve Bayes*, yang sederhana dan cepat, dapat memperoleh manfaat dari pengurangan varians melalui agregasi beberapa model. *Decision Tree*, yang rentan terhadap *overfitting*, dapat distabilkan dengan *Bagging*, menghasilkan model yang lebih generalis. SVM, yang sensitif terhadap skala data dan pemilihan kernel, dapat memperoleh peningkatan kinerja dengan mengurangi bias dan varians. Dengan demikian, *Bagging* membantu meningkatkan kinerja keseluruhan dari ketiga algoritma ini.

Untuk mendalami masalah klasifikasi ini, penelitian ini menggunakan dataset *Multiple Disease Prediction* yang dibuat oleh Aboelnaga pada tahun 2024. Dataset yang tersedia secara publik ini telah banyak digunakan dalam tugas klasifikasi. Setiap atribut direpresentasikan sebagai tipe data *floating-point* dan telah dinormalisasi. Oleh karena itu, dataset ini digunakan sebagai eksperimen yang cocok untuk mengevaluasi efektivitas *Naïve Bayes*, *Decision Tree*, dan SVM yang akan diterapkan teknik agregasi *Bagging*.

Penelitian ini diharapkan dapat memberikan wawasan baru tentang efektivitas teknik agregasi *Bagging* dengan algoritma *Naïve Bayes*, *Decision Tree*, dan SVM, terutama dalam ranah klasifikasi. Selain itu, penelitian ini juga bertujuan untuk berkontribusi pada pengembangan metode alternatif yang dapat meningkatkan kinerja model-model tersebut.

1.3. Rumusan Masalah

Berdasarkan latar belakang diatas, maka masalah yang ingin diteliti dalam penelitian ini adalah:

1. Bagaimana mengembangkan sistem klasifikasi berbasis *Bagging* menggunakan *Naïve Bayes*, *Decision Tree*, dan SVM?

2. Bagaimana kinerja sistem klasifikasi berbasis *Bagging* menggunakan *Naïve Bayes*, *Decision Tree*, dan SVM berdasarkan *F1-Score*.

1.4. Tujuan

Tujuan dari penelitian ini adalah untuk:

1. Menghasilkan sistem klasifikasi berbasis *Bagging* dengan *Naïve Bayes*, *Decision Tree*, dan SVM.
2. Mengetahui kinerja sistem klasifikasi berbasis *Bagging* menggunakan *Naïve Bayes*, *Decision Tree*, dan SVM berdasarkan *F1-Score*.

1.5. Manfaat

Adapun manfaat yang diharapkan dari penelitian ini, yaitu:

1. Menambah wawasan baru terkait teknik *Bagging*, agregasi model, dan penerapannya.
2. Menunjukkan metode alternatif untuk menyusun dan meningkatkan performa model terhadap masalah klasifikasi untuk penelitian kedepannya.

1.6. Batasan Masalah

Suatu batasan dalam penelitian perlu dikembangkan agar penyelesaian yang dicapai sesuai dengan permasalahan yang dihadapi dan tidak meluas ke permasalahan lain yang tidak terkait. Dari itu, penulis menyajikan bahasan terkait:

1. Dataset yang digunakan adalah *Multiple Disease Prediction*.
2. Bahasa pemrograman yang digunakan adalah Python dan pustaka yang digunakan adalah *pandas* dan *scikit-learn*.

1.7. Sistematika Penulisan

Sistematika penulisan yang digunakan pada penelitian ini mengikuti standar operasional penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya yakni:

BAB I. PENDAHULUAN

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan yang digunakan dalam penyusunan laporan akhir ini.

BAB II. KAJIAN LITERATUR

Bab ini menjelaskan dasar-dasar teori yang digunakan dalam penelitian, seperti definisi algoritma-algoritma *Naïve Bayes*, *decision tree*, dan SVM, pustaka-pustaka yang digunakan dalam pengembangan sistem, penilaian performa model dan visualisasi hasil prediksi, serta penelitian terkait sebelumnya yang berkaitan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Bab ini membahas mengenai tahapan-tahapan yang akan dilaksanakan selama penelitian seperti metode pengumpulan data hingga metode perancangan perangkat lunak. Setiap tahapan penelitian akan dijelaskan secara rinci sesuai kerangka kerja yang ditetapkan.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini membahas proses pengembangan perangkat lunak dimulai dari analisis kebutuhan perangkat lunak hingga pengujian model-model untuk evaluasi.

BAB V. HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan hasil penelitian berdasarkan langkah dan metode yang direncanakan sebelumnya. Analisis diberikan sebagai basis dari kesimpulan yang diambil dari penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penelitian teknik agregasi *Bagging* ini.

1.8. Kesimpulan

Dengan uraian yang telah dijelaskan pada subbab-subbab sebelumnya, penelitian ini akan membahas mengenai penerapan teknik agregasi *Bagging* dengan algoritma *Naïve Bayes*, *Decision Tree*, dan *SVM* untuk mengklasifikasi dataset *Multiple Disease Prediction*.

DAFTAR PUSTAKA

- Alelyani, S. (2021). "Stable bagging feature selection on medical data." *Journal of Big Data*. 8(11).
- Aniruddha, A. (2024). ML Handbook - F1 Score. (ML Handbook, GitHub). Tersedia: https://bsc-iitm.github.io/ML_Handbook/pages/f1.html. Retrieved 4 Mei 2024.
- Bauer, E. & Kohavi, R. (1999). "An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants." *Machine Learning*. 36, 105–139.
- Brownlee, J. (2020). A Gentle Introduction to Scikit-Learn: A Python Machine Learning Library. (Python Machine Learning, Online). Tersedia: <https://machinelearningmastery.com/a-gentle-introduction-to-scikit-learn-a-python-machine-learning-library/>. Retrieved 4 Mei 2024.
- Cortes, C. & Vapnik, V. (1995). "Support-vector networks." *Machine Learning*. 20(3), 273-297.
- DataCamp. (2023, 20 November). What is Bagging in Machine Learning? A Guide With Examples. (DataCamp, Online) Tersedia: <https://www.datacamp.com/tutorial/what-bagging-in-machine-learning-a-guide-with-examples>. Retrieved 28 Oktober 2024.
- Downey, A. (2012). *Think Python*. Sebastopol, CA: O'Reilly Media, Inc.
- Düntsch, I. & Gediga, G. (2019). "Confusion matrices and rough set data analysis." *Journal of Physics: Conference Series*. 1229(1), 12-55.
- Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., & Homayouni, S. (2021). "Bagging and Boosting Ensemble Classifiers for Classification of Multispectral, Hyperspectral, and PolSAR Data: A Comparative Evaluation." *Remote Sensing*. 13(21), 4405.
- Nelli, F. (2023). pandas: Reading and Writing Data. Python Data Analytics. Berkeley, CA: Apress.
- Ranshing, S. (2023, 20 Desember). Iterative Enhancement Model in Software Development. (GeeksForGeeks, Online). Tersedia:

- <https://www.geeksforgeeks.org/iterative-enhancement-model-in-software-development/>. Retrieved 2 Desember 2024.
- Rish, I. (2001, 02 November). "An empirical study of the naive Bayes classifier." *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*. 3.
- Rokach, L. & Maimon, O. (2014). *Data Mining with Decision Trees: Theory and Applications*. Israel: World Scientific.
- scikit-learn: machine learning in Python. (scikit-learn 1.4.2 documentation, Online). Tersedia: <https://scikit-learn.org/stable/index.html>. Retrieved 4 May 2024.
- Ting, K. M. (2011). "Confusion Matrix." *Encyclopedia of Machine Learning*. C. Sammut and G. I. Webb, Eds. Boston, MA: Springer.
- Wu, P., Ma, R., & Toe, T. T. (2023). "Stacking-Enhanced Bagging Ensemble Learning for Breast Cancer Classification with CNN." *2023 3rd International Conference on Electronic Engineering (ICEEM)*. 1-6.
- Zararsiz, G., Elmali, F., & Ozturk, A.. (2012). "Bagging Support Vector Machines for Leukemia Classification." *International Journal of Computer Science Issues*, 9(6), 355-358.