



**ALGORITMA MISSFOREST UNTUK PERBAIKAN *MISSING VALUE*
PADA DATASET HEPATITIS MENGGUNAKAN KLASIFIKASI
*NAÏVE BAYES***

Skripsi

**Untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD FAIESAL ANDHINI

NIM 1711016310014

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

APRIL 2023

**ALGORITMA MISSFOREST UNTUK PERBAIKAN *MISSING VALUE*
PADA DATASET HEPATITIS MENGGUNAKAN KLASIFIKASI
*NAÏVE BAYES***

Skripsi

**Untuk memenuhi persyaratan
dalam menyelesaikan Strata-1 Ilmu Komputer**

Oleh

MUHAMMAD FAIESAL ANDHINI

NIM 1711016310014

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

APRIL 2023

SKRIPSI

ALGORITMA MISSFOREST UNTUK PERBAIKAN *MISSING VALUE* PADA DATASET HEPATITIS MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES*

Oleh :

MUHAMMAD FAIESAL ANDHINI
1711016310014

Telah dipertahankan di depan Dosen Penguji pada hari kamis tanggal 06 April 2023

Susunan Dosen Penguji :

Pembimbing I



Muliadi, S.Kom., M.Cs.

NIP. 197804222010121002

Dosen Penguji I



Radityo Adi Nugroho, S.T., M.Kom.

NIP. 198212042008011006

Pembimbing II



Dodon Turianto Nugrahadi, S.Kom, M.Eng

NIP. 1980011122009121002

Dosen Penguji II



Muhammad Itqan Mazdadi, S.Kom., M.Kom

NIP. 199006122019031013



Banjarbaru,

2023

Ketua Program Studi Ilmu Komputer

Irwanto Budiman, S.T., M.Kom

NIP. 197703252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 06 April 2023



Muhammad Faiesal Andhini
NIM. 1711016310014

ABSTRAK

ALGORITMA MISSFOREST UNTUK PERBAIKAN *MISSING VALUE* PADA DATASET HEPATITIS MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES* (Oleh: Muhammad Faiesal Andhini; Pembimbing: Muliadi, S.Kom., M.Cs. dan Dodon Turianto Nugrahadi, S.Kom., M.Eng.; 2023; 65 halaman)

Hepatitis adalah penyakit peradangan hati yang disebabkan oleh virus hepatitis. Penyakit ini menjadi ancaman kesehatan utama di seluruh dunia, termasuk di Indonesia. Kumpulan data (*dataset*) tersebut perlu diklasifikasikan sebelum dapat digunakan untuk tujuan medis. Namun, dataset yang tidak lengkap (*missing value*) dapat menghambat proses klasifikasi. Untuk mengatasi masalah ini, pendekatan yang dapat dilakukan adalah dengan menghapus data yang memiliki *missing value* atau melakukan perbaikan imputasi dengan metode tertentu. Dalam penelitian ini, terdapat 155 data dan dilakukan dua skenario untuk mengetahui akurasi klasifikasi terbaik. Pada skenario pertama, klasifikasi *naïve bayes* dilakukan dengan menghapus 75 data yang memiliki *missing value*. Sedangkan pada skenario kedua, klasifikasi *naïve bayes* dilakukan setelah melakukan perbaikan imputasi dengan algoritma Missforest pada 75 data dari 155 data. Pengujian dilakukan dengan komposisi data 50:50, 60:40, dan 70:30. Hasil penelitian menunjukkan bahwa pada skenario pertama, akurasi klasifikasi berturut-turut sebesar 87,50%, 87,50%, dan 83,33%. Sementara pada skenario kedua, akurasi klasifikasi berturut-turut sebesar 83,33%, 90,32%, dan 93,62%. Dari hasil ini dapat disimpulkan bahwa penggunaan algoritma Missforest mampu meningkatkan performa pada akurasi klasifikasi *naïve bayes*.

Kata kunci : Klasifikasi, Missforest, *Naïve Bayes*

ABSTRACT

MISSFOREST ALGORITHM FOR IMPROVING MISSING VALUE ON HEPATITIS DATASET USING NAÏVE BAYES CLASSIFICATION (By: Muhammad Faiesal Andhini; Mentor: Muliadi, S.Kom., M.Cs. and Dodon Turianto Nugrahadi, S.Kom., M.Eng.; 2023; 65 Pages)

Hepatitis is an inflammatory liver disease caused by the hepatitis virus. This disease is a major health threat worldwide, including in Indonesia. The dataset has to be classified before it can be used for medical purposes. However, incomplete datasets (missing values) can impede the classification process. An approach that can be taken is by removing instances with missing values or performing imputation repairs using certain methods. In this study, there are 155 instances of hepatitis patients and two scenarios are carried out to determine the best classification accuracy. In the first scenario, naïve bayes classification is carried out by removing 75 instances with missing values. Meanwhile, in the second scenario, naïve bayes classification is carried out after performing imputation repairs with the Missforest algorithm on 75 instances of hepatitis patients out of 155 instances. Testing is conducted with data compositions of 50:50, 60:40, and 70:30. The results show that in the first scenario, the classification accuracy is 87.50%, 87.50%, and 83.33% respectively. Meanwhile, in the second scenario, the classification accuracy is 83.33%, 90.32%, and 93.62% respectively. From these results, it can be concluded that the use of the Missforest algorithm can improve performance on the accuracy naïve bayes classification.

Keywords : Classification, Missforest, Naïve Bayes

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur kepada ALLAH SWT atas rahmat dan karunia-Nya sehingga penulis bisa menyelesaikan skripsi yang berjudul “*Algoritma Missforest Untuk Perbaikan Missing Value Pada Dataset Hepatitis Menggunakan Klasifikasi Naïve Bayes*” untuk memenuhi syarat untuk menyelesaikan Pendidikan Program Strata-1 di jurusan Ilmu Komputer pada Fakultas Matematika dan Ilmu Pengetahuan Alam di Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak terkait yang sudah membantu dalam penyelesaian skripsi ini, diantaranya:

1. Allah SWT, karena atas limpahan karunia-Nya penulis dapat menyelesaikan skripsi ini.
2. Keluarga tercinta, terkhusus kedua orang tua yang selalu memberikan doa, bantuan, semangat, dukungan, dan selalu memotivasi untuk terus berjuang dalam proses penyelesaian skripsi ini.
3. Bapak Muliadi, S.Kom., M.Cs. selaku dosen pembimbing utama dan Bapak Dodon Turianto Nugrahadi, S.Kom., M.Eng. selaku dosen pembimbing pendamping yang telah memberikan ilmu, waktu, dan bimbingannya kepada penulis dalam pengerjaan skripsi.
4. Bapak Radityo Adi Nugroho, S.T., M.Kom. dan Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom. selaku dosen penguji yang telah memberikan masukan untuk penyajian skripsi ini.
5. Bapak Irwan Budiman, S.T., M.Kom. selaku ketua Program Studi S-1 Ilmu Komputer beserta seluruh dosen dan karyawan/staff pegawai atas segala bantuan yang diberikan selama penulis mengikuti studi.
6. Seluruh Panitia Skripsi dan Dosen serta Staf Program Studi Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat, atas bantuan dan ilmu yang diberikan selama ini yang sangat bermanfaat.

7. Seluruh mahasiswa Ilmu Komputer 2017 yang menjadi teman perjuangan dalam perkuliahan di Prodi Ilmu Komputer yang selalu menjadi keluarga dan selalu bersedia memberikan informasi–informasi mengenai perkuliahan.
8. Noor Hidayah yang telah memberikan support, motivasi dan semangat sehingga mampu dalam menyelesaikan penelitian ini.
9. Lutfi Islami, Muhammad Syaoki Faradisa, Noor Hidayah, Sartika Dewi, Digun Fitriadi, Ahmad Nur Rixon, Rahmat Hidayat, dan Muhammad Imam Sufiazi yang selalu membantu dalam menyelesaikan penelitian ini.
10. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, maka kritik dan saran dari semua pihak sangat diharapkan demi penyempurnaan selanjutnya. Akhirnya, hanya kepada Tuhan Yang Maha Esa kita kembalikan semua urusan dan diharapkan skripsi ini dapat bermanfaat bagi semua pihak dan dapat menambah pengetahuan kita semua.

Banjarbaru, 06 April 2023



Muhammad Faiesal Andhini

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan	3
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu.....	5
2.2 Keaslian Penelitian.....	7
2.3 Hepatitis	9
2.4 Data Mining	10
2.5 <i>Machine Learning</i>	11
2.6 <i>Supervised Learning</i>	12
2.7 <i>Unsupervised Learning</i>	12
2.8 <i>Semi-supervised Learning</i>	12
2.9 <i>Reinforcement Learning</i>	12
2.10 <i>Missing Value</i>	13
2.11 <i>Preprocessing Data</i>	14
2.12 <i>Missforest</i>	15

2.13	<i>Data Training dan Data Testing</i>	16
2.14	Klasifikasi <i>Naïve Bayes</i>	17
2.15	<i>Confusion Matrix</i>	20
BAB III METODE PENELITIAN		22
3.1	Alat Penelitian	22
3.2	Bahan Penelitian.....	22
3.3	Variabel Penelitian	22
3.4	Prosedur Penelitian.....	22
BAB IV HASIL DAN PEMBAHASAN		26
4.1	Hasil	26
4.1.1	Pengumpulan Data	26
4.1.2	<i>Preprocessing Data</i>	28
4.1.3	Pembagian Komposisi <i>Data Training dan Data Testing</i> ...	37
4.1.4	Klasifikasi	39
4.1.5	Evaluasi.....	46
4.2	Pembahasan.....	57
BAB V PENUTUP.....		61
5.1	Kesimpulan	61
5.2	Saran.....	61
DAFTAR PUSTAKA		62
DAFTAR LAMPIRAN		66
LAMPIRAN		

DAFTAR GAMBAR

Gambar 1. Alur Penelitian	23
Gambar 2. Pola <i>Missing Value</i> Pada Dataset Hepatitis	27
Gambar 3. Perbandingan Distribusi Kelas <i>Live</i> Dan Kelas <i>Die</i>	28
Gambar 4. Pohon Prediksi Fitur Malaise	34
Gambar 5. <i>Confusion Matrix Naïve Bayes</i> 50:50	47
Gambar 6. <i>Confusion Matrix Naïve Bayes</i> 60:40	48
Gambar 7. <i>Confusion Matrix Naïve Bayes</i> 70:30	50
Gambar 8. Grafik Perbandingan Nilai Akurasi <i>Naïve Bayes</i>	51
Gambar 9. <i>Confusion Matrix Naïve Bayes</i> + Missforest 50:50	52
Gambar 10. <i>Confusion Matrix Naïve Bayes</i> + Missforest 60:40	54
Gambar 11. <i>Confusion Matrix Naïve Bayes</i> + Missforest 70:30	55
Gambar 12. Grafik Perbandingan Nilai Akurasi <i>Naïve Bayes</i> + Missforest	57
Gambar 13. Perbandingan Akurasi Model <i>Naïve Bayes</i>	60

DAFTAR TABEL

Tabel 1. Keaslian Penelitian	7
Tabel 2. Penelitian Yang Diusulkan	9
Tabel 3. Data Fitur Pada Dataset Hepatitis	10
Tabel 4. Contoh <i>Missing Value</i> Pada Dataset Hepatitis	13
Tabel 5. <i>Confusion Matrix</i>	21
Tabel 6. Daftar Fitur Pada Dataset Hepatitis.....	26
Tabel 7. Contoh Dataset Hepatitis.....	27
Tabel 8. Jumlah Distribusi Kelas.....	27
Tabel 9. Transformasi Data Kategorikal Menjadi Data Numerikal	28
Tabel 10. Hasil <i>Ordinal Encoding</i> Pada Dataset.....	29
Tabel 11. Hasil <i>Cleaning Data</i> Pada <i>Missing Value</i>	30
Tabel 12. Contoh Perhitungan Nilai Rata-Rata.	31
Tabel 13. Data Hasil Imputasi Dengan Nilai Rata-Rata Dan Nilai Modus.....	31
Tabel 14. Urutan Data Atribut Berdasarkan Jumlah <i>Missing Value</i>	32
Tabel 15. Contoh Data Yang Akan Diterapkan Model <i>Random Forest</i>	33
Tabel 16. Hasil Imputasi Data Train 50:50 Dengan Missforest.....	34
Tabel 17. Hasil Imputasi Data Test 50:50 Dengan Missforest.....	35
Tabel 18. Hasil Imputasi Data Train 60:40 Dengan Missforest.....	35
Tabel 19. Hasil Imputasi Data Test 60:40 Dengan Missforest.....	35
Tabel 20. Hasil Imputasi Data Train 70:30 Dengan Missforest.....	36
Tabel 21. Hasil Imputasi Data Test 70:30 Dengan Missforest.....	36
Tabel 22. Jumlah Distribusi Kelas Pada Data Training Dan Data Testing	37
Tabel 23. Jumlah Distribusi Kelas Pada Data Training Dan Data Testing	38
Tabel 24. Contoh 5 Data Skenario Pertama	39
Tabel 25. Contoh 5 Data Skenario Kedua	40
Tabel 26. Contoh Tabel Probabilitas <i>Naïve Bayes</i> $x1$	40
Tabel 27. Contoh Tabel Probabilitas <i>Naïve Bayes</i> + Missforest $x1$	41
Tabel 28. <i>Confusion Matrix Naïve Bayes</i> 50:50.....	47
Tabel 29. Hasil <i>Misspredict Naïve Bayes</i> 50:50.....	47

Tabel 30. Perhitungan Akurasi <i>Naïve Bayes</i> 50:50.....	48
Tabel 31. <i>Confusion Matrix Naïve Bayes</i> 60:40.....	49
Tabel 32. Hasil <i>Misspredict Naïve Bayes</i> 60:40.....	49
Tabel 33. Perhitungan Akurasi <i>Naïve Bayes</i> 60:40.....	49
Tabel 34. <i>Confusion Matrix Naïve Bayes</i> 70:30.....	50
Tabel 35. Hasil <i>Misspredict Naïve Bayes</i> 70:30.....	50
Tabel 36. Perhitungan Akurasi <i>Naïve Bayes</i> 70:30.....	51
Tabel 37. <i>Confusion Matrix Naïve Bayes + Missforest</i> 50:50	52
Tabel 38. Hasil <i>Misspredict Naïve Bayes + Missforest</i> 50:50	52
Tabel 39. Perhitungan Akurasi <i>Naïve Bayes + Missforest</i> 50:50.....	53
Tabel 40. <i>Confusion Matrix Naïve Bayes + Missforest</i> 60:40	54
Tabel 41. Hasil <i>Misspredict Naïve Bayes + Missforest</i> 60:40	54
Tabel 42. Perhitungan Akurasi <i>Naïve Bayes + Missforest</i> 60:40.....	55
Tabel 43. <i>Confusion Matrix Naïve Bayes + Missforest</i> 70:30	56
Tabel 44. Hasil <i>Misspredict Naïve Bayes + Missforest</i> 70:30	56
Tabel 45. Perhitungan Akurasi <i>Naïve Bayes + Missforest</i> 70:30.....	56
Tabel 46. Hasil Akurasi Model Skenario Klasifikasi <i>Naïve Bayes</i>	59
Tabel 47. Hasil Akurasi Model Skenario Klasifikasi <i>Naïve Bayes + Missforest</i> ...	59

DAFTAR LAMPIRAN

- Lampiran 1.** Source Code *Installing*
- Lampiran 2.** Source Code *Importing*
- Lampiran 3.** Source Code *Naïve Bayes*
- Lampiran 4.** Source Code *Confusion Matrix Naïve Bayes*
- Lampiran 5.** Source Code *Naïve Bayes + Missforest*
- Lampiran 6.** Source Code *Confusion Matrix Naïve Bayes + Missforest*
- Lampiran 7.** Source Code *Pohon Missforest*
- Lampiran 8.** Source Code *MissPredict*
- Lampiran 9.** Dataset *Hepatitis*
- Lampiran 10.** Dataset Hasil *Ordinal Encoding*
- Lampiran 11.** Dataset Hasil *Skenario Pertama*
- Lampiran 12.** Dataset Hasil *Skenario Kedua*
- Lampiran 13.** Riwayat *Hidup Penulis*