



**IMPLEMENTASI *RANDOM FOREST* DENGAN *HYPER-PARAMETER*
TUNING BAYESIAN SEARCH UNTUK MEMPREDIKSI GEJALA DEMAM
TIFOID**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

**SHEGA SEPTIA AGESTATAMA
NIM 1711016310027**

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

JULI 2023

SKRIPSI
IMPLEMENTASI RANDOM FOREST DENGAN HYPER-PARAMETER
TUNING BAYESIAN SEARCH UNTUK MEMPREDIKSI GEJALA DEMAM
TIFOID

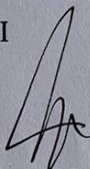
Oleh :

SHEGA SEPTIA AGESTATAMA
1711016310027

Telah dipertahankan di depan Dosen Penguji pada hari Rabu tanggal 05 Juli 2023

Susunan Dosen Penguji :

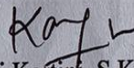
Pembimbing I



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

NIP. 199006122019031013

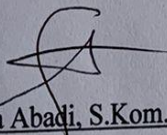
Dosen Penguji I



Dwi Kartini, S.Kom., M.Kom

NIP. 198704212012122003

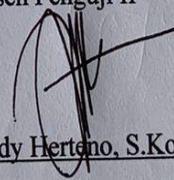
Pembimbing II



Friska Abadi, S.Kom., M.Kom

NIP. 19880913201612110001

Dosen Penguji II



Rudy Herteno, S.Kom., M.Kom

NIP. 198809252022031003

Banjarbaru, 05 Juli 2023

Ketua Program Studi Ilmu Komputer



Rawanudin, 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, 05 Juli 2023

Yang Menyatakan,



Shega Septia Agestatama

NIM.1711016310027

ABSTRAK

PREDIKSI GEJALA DEMAM TIFOID MENGGUNAKAN *RANDOM FOREST* DENGAN *HYPER-PARAMETER TUNING BAYESIAN SEARCH* (Oleh : Shega Septia Agestatama; Pembimbing: Muhammad Itqan Mazdadi, S.Kom., M.Kom. dan Friska Abadi, S.Kom, M.Kom.; 2023; 56 halaman)

Diagnosis demam tifoid yang tepat waktu dan akurat memiliki peran krusial dalam pengobatan yang efektif dan pencegahan penyebaran penyakit. Oleh karena itu, perlu dilakukan upaya untuk meminimalkan kesalahan dalam deteksi dan menghindari keterlambatan dalam diagnosis demam tifoid. Dalam jurnal ini, membahas berbagai strategi dan pendekatan yang dapat membantu meningkatkan akurasi diagnosis demam tifoid serta mengurangi risiko penyebaran infeksi, dilakukan penerapan dan pemanfaatan teknik data mining. Penerapan dan pemanfaatan teknik data mining mampu menemukan pola. Dampaknya telah mendorong pengembangan sistem kecerdasan dan penunjang keputusan di bidang kesehatan. Keuntungannya termasuk kemampuan untuk mendiagnosis penyakit dengan akurasi lebih tinggi, mampu memprediksi tingkat keparahan berbagai penyakit dan melakukan pengawasan kesehatan dari jarak jauh.. *Data mining* adalah rangkaian proses untuk menggali nilai tambah dari data dengan cara yang tidak diketahui sebelumnya secara manual. Penelitian ini menggunakan klasifikasi dengan metode *Random Forest* yang dikombinasikan dengan metode *hyperparameter tuning*. Tujuan dari penelitian ini adalah untuk mengevaluasi performa model yang sudah diolah. Model tersebut dilatih dengan dataset dan metode *hyperparameter tuning* yang digunakan adalah *Bayesian Search*, dengan validasi menggunakan *10-fold Cross Validation* dalam 5 iterasi. Beberapa *hyper-parameter* yang dikonfigurasi pada *Random Forest* antara lain *n_estimators*, *min_samples_split*, *max_depth*, *min_samples_leaf*, *max_features*, dan *bootstrap*. Hasil pengujian *Random Forest* tanpa *tuning hyper-parameter* memiliki akurasi sebesar 93,10%, sedangkan dengan *tuning Hyper-parameter* menggunakan *Bayesian Search*, akurasi meningkat menjadi 96,55%.

Kata kunci: Demam Tifoid, *Random Forest*, *Hyper-parameter Tuning*, *Bayesian Search*

ABSTRACT

PREDICTION OF TYPHOID FEVER SYMPTOMS USING RANDOM FOREST WITH BAYESIAN SEARCH HYPERPARAMETER TUNING (By : Shega Septia Agestatama Supervisor: Muhammad Itqan Mazdadi, S.Kom., M.Kom. dan Friska Abadi, S.Kom, M.Kom.; 2023; 56 pages)

Timely and accurate diagnosis of typhoid fever plays a crucial role in effective treatment and prevention of disease transmission. Therefore, efforts are needed to minimize errors in detection and avoid delays in diagnosing typhoid fever. This journal discusses various strategies and approaches that can help improve the accuracy of typhoid fever diagnosis and reduce the risk of infection spread through the application and utilization of data mining techniques. The implementation and utilization of data mining techniques have been able to discover patterns, leading to the development of intelligent systems and decision support in the field of healthcare. Its benefits include the ability to diagnose diseases with higher accuracy, predict the severity of various illnesses, and enable remote health monitoring. Data mining is a series of processes to extract previously unknown value from data in an automated manner. This research employs classification using the Random Forest method combined with hyperparameter tuning. The goal of this study is to evaluate the performance of the processed model. The model is trained with a dataset, and the hyperparameter tuning method used is Bayesian Search, with validation using 10-fold Cross Validation in 5 iterations. Several hyperparameters configured for Random Forest include `n_estimators`, `min_samples_split`, `max_depth`, `min_samples_leaf`, `max_features`, and `bootstrap`. The testing results of Random Forest without hyperparameter tuning show an accuracy of 93.10%, whereas with Hyper-parameter tuning using Bayesian Search, the accuracy increases to 96.55%.

Keywords: Typhoid Fever, Random Forest, Hyper-parameter Tuning, Bayesian Search.

PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **“IMPLEMENTASI *RANDOM FOREST* DENGAN *HYPER-PARAMETER TUNING BAYESIAN SEARCH* UNTUK MEMPREDIKSI GEJALA DEMAM TIFOID”** untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini, terutama untuk Bapak, Ibu, Nisa dan Nazwa.
2. Bapak Muhammad Itqan Mazdadi S.Kom., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Friska Abadi, S.Kom., M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Ucapan terima kasih kepada Ryan, Barok, Arif, Bowo, Aldo, Irvan, Ubai, Ilham, Zamzam, Yudhit, Fauzi,, Haikal, Said, Eddy, Azmi, Luthfi, Didit, Rizqon,

Muflih, Noval, Apri, lastri, yang memberikan dukungan dalam proses mengerjakan skripsi.

8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 05 Juli 2023



Shega Septia Agestatama

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA	
2.1 Kajian Terdahulu	5
2.2 Landasan Teori	10
2.2.1 Demam Tifoid	10
2.2.2 <i>Data Mining</i>	11
2.2.3 <i>Decission tree</i>	12
2.2.4 <i>Random Forest</i>	13

2.2.5	<i>Split Validation</i>	14
2.2.6	<i>Hyper-parameter</i>	15
2.2.7	<i>Hyper-parameter tuning</i>	17
2.2.8	<i>Bayesian Search</i>	17
2.2.9	<i>Cross Validation</i>	19
2.2.10	<i>Confusion Matrix</i>	20
 BAB III METODE PENELITIAN		
3.1	Bahan Penelitian	22
3.2	Alat Penelitian	22
3.3	Variabel Penelitian.....	22
3.4	Prosedur Penelitian	22
 BAB IV HASIL DAN PEMBAHASAN		
4.1	Hasil.....	25
4.1.1	Pengumpulan Dataset.....	25
4.1.2	<i>Preprocessing Data</i>	27
4.1.3	Pembagian <i>Data Training</i> dan <i>Data Testing</i>	30
4.1.4	<i>Klasifikasi</i>	30
4.1.5	<i>Klasifikasi Random Forest</i> tanpa <i>Hyper-parameter tuning</i>	30
4.1.6	<i>Klasifikasi Random Forest</i> dengan <i>Hyper-parameter tuning</i>	46
4.2	Pembahasan	51
 BAB V PENUTUP		
5.1	Kesimpulan	55
5.2	Saran	55
DAFTAR PUSTAKA		56
LAMPIRAN		58

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Struktur <i>Decision Tree</i>	12
Gambar 2. Visualisasi Metode <i>Random Forest</i>	13
Gambar 3. Ilustrasi <i>Split Validation</i>	15
Gambar 4. Visual <i>Bayesian Search</i>	18
Gambar 5. Visual kedua <i>Bayesian Search</i>	19
Gambar 6. <i>Cross Validation 10-fold</i>	20
Gambar 7. Alur Penelitian.....	23
Gambar 8. Perbandingan persentase kelas tifoid dan tidak tifoid.....	27
Gambar 9. Alur <i>Preprocessing</i>	27
Gambar 10. Ilustrasi Pembagian Dataset	30
Gambar 11. Ilustrasi Pohon pertama <i>Random Forest</i>	32
Gambar 12. Ilustrasi Pohon kedua <i>Random Forest</i>	33
Gambar 13. Ilustrasi Pohon ketiga <i>Random Forest</i>	34
Gambar 14. <i>Root node</i> pohon keputusan <i>Random Forest</i>	36
Gambar 15. <i>Node 1.2</i> Pohon Keputusan <i>Random Forest</i>	38
Gambar 16. Pohon Keputusan pertama <i>Random Forest</i>	39
Gambar 17. Pohon Keputusan kedua <i>Random Forest</i>	41
Gambar 18. Pohon Keputusan Ketiga <i>Random Forest</i>	42
Gambar 19. <i>Confusion Matrix Random Forest</i> tanpa <i>Hyper-parameter tuning</i>	45
Gambar 20. Pembagian data <i>Cross Validation</i>	48
Gambar 21. <i>Confusion Matrix Random Forest</i> dengan <i>Hyper-parameter tuning</i>	50
Gambar 22. Perbandingan Kinerja Model	54

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian.....	7
Tabel 2. Perancangan Penelitian	9
Tabel 3. Kelebihan dan kekurangan <i>decision tree</i>	13
Tabel 4. Parameter <i>Default Random Forest</i>	16
Tabel 5. <i>Confusion Matrix</i>	26
Tabel 6. Fitur dan Kelas Dataset	26
Tabel 7. Nama Fitur	27
Tabel 8. Dataset Rekam Medis Demam Tifoid.....	25
Tabel 9. Jumlah Distribusi Kelas	26
Tabel 10. Dataset <i>transform label encoder</i>	28
Tabel 11. Dataset setelah dilakukan <i>label encoder</i>	29
Tabel 12. <i>Hyper-parameter Random Forest</i>	31
Tabel 13. Data untuk model <i>Random Forest</i>	35
Tabel 14. Nilai <i>Entropy</i> dan <i>gain</i> pada pohon pertama	36
Tabel 15. Data dari atribut Demam(hari) yang bernilai 1	37
Tabel 16. Nilai <i>Entropy</i> dan <i>gain</i> pada <i>node 1.2</i>	37
Tabel 17. Data dari atribut muntah bernilai “1”.....	38
Tabel 18. Nilai <i>Entropy</i> dan <i>gain</i> pada <i>node 2.2</i>	39
Tabel 19. Data kedua untuk model <i>Random Forest</i>	40
Tabel 20. Data ketiga untuk model <i>Random Forest</i>	41
Tabel 21. Hasil <i>Majority Voting</i>	43
Tabel 22. <i>Confusion Matrix</i>	44
Tabel 23. <i>Confusion Matrix</i> klasifikasi <i>Random Forest</i> tanpa <i>hyper-parameter tuning</i>	45
Tabel 24. Perhitungan akurasi klasifikasi <i>Random Forest</i> tanpa <i>Hyper-parameter tuning</i>	45
Tabel 25. <i>Hyper-parameter</i> yang dikonfigurasi.....	46
Tabel 26. Kandidat <i>Hyper-parameter Bayesian Search</i>	49

Tabel 27. <i>10-fold Cross validation</i>	49
Tabel 28. <i>Confusion Matrix</i> Klasifikasi <i>Random Forest</i> dengan <i>Hyper-parameter tuning</i>	50
Tabel 29. Perhitungan Akurasi klasifikasi <i>Random Forest</i> dengan <i>Hyper-parameter tuning</i>	50
Tabel 30. <i>Hyper-parameter</i> optimal	53
Tabel 31. Perbandingan <i>Hyper-parameter tuning</i>	53

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 *Source Code* Pembagian Dataset

Lampiran 2 *Source Code* Random Forest

Lampiran 3 *Source Code* Bayesian Search

Lampiran 4 *Source Code* label encoder