



**PENERAPAN *SHAPLEY ADDITIVE EXPLANATIONS* (SHAP) PADA
MODEL *RANDOM FOREST* UNTUK KLASIFIKASI PEMILIHAN
FASILITAS KESEHATAN**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
ANGELINA IVANNA GENARDI
NIM. 2211017120002**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
NOVEMBER 2025**



**PENERAPAN *SHAPLEY ADDITIVE EXPLANATIONS* (SHAP) PADA
MODEL *RANDOM FOREST* UNTUK KLASIFIKASI PEMILIHAN
FASILITAS KESEHATAN**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
ANGELINA IVANNA GENARDI
NIM. 2211017120002**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
NOVEMBER 2025**

SKRIPSI

PENERAPAN *SHAPLEY ADDITIVE EXPLANATIONS* (SHAP) PADA MODEL *RANDOM FOREST* UNTUK KLASIFIKASI PEMILIHAN FASILITAS KESEHATAN

Oleh
Angelina Ivanna Genardi
NIM. 2211017120002

Telah dipertahankan pada hari Senin, tanggal 10-11-2025 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



Selvi Annisa, S.Si., M.Si.
NIP. 199212262022032016

Penguji I



Dewi Anggraini, S.Si., M.App.Sci., Ph.D.
NIP. 198303282005012001

Pembimbing II



Rifqi Adhya Rahman, S.Mat., M.Si.
NIP. 199708072024061003

Penguji II



Oni Soesanto, S.Si., M.Si.
NIP. 197301262005011003

Banjarbaru, 10 November 2025

Mengetahui,

Koordinator Program Studi Statistika
EMIPA ULM



Dewi Anggraini, S.Si., M.App.Sci., Ph.D.
NIP. 198303282005012001

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, 10 November 2025



Angelina Ivanna Genardi

NIM. 2211017120002

ABSTRAK

Penerapan *Shapley Additive Explanations* (SHAP) pada Model *Random Forest* untuk Klasifikasi Pemilihan Fasilitas Kesehatan (Oleh: Angelina Ivanna Genardi; Pembimbing: Selvi Annisa dan Rifqi Aulya Rahman, 2025; 52 halaman)

Pemerataan layanan kesehatan bergantung pada perilaku masyarakat dalam memilih fasilitas kesehatan. Akses yang merata belum tercapai karena masih terdapat kesenjangan pemanfaatan antar kelompok masyarakat. Pemilihan fasilitas kesehatan dipengaruhi oleh berbagai faktor sosial ekonomi. Sebagian besar penelitian sebelumnya menggunakan model klasifikasi klasik. Penelitian ini bertujuan mengembangkan model klasifikasi pemilihan fasilitas kesehatan di Kelurahan Basirih menggunakan *Random Forest* dengan interpretasi *Shapley Additive Explanations* (SHAP). Analisis deskriptif dilakukan untuk menggambarkan karakteristik sosial ekonomi keluarga berdasarkan pilihan fasilitas kesehatan. Ketidakseimbangan kelas pada data diatasi menggunakan *Synthetic Minority Oversampling Technique - Nominal* (SMOTE-N). SHAP digunakan untuk menjelaskan kontribusi setiap prediktor terhadap hasil klasifikasi. Hasil analisis deskriptif menunjukkan keluarga yang memilih Puskesmas umumnya kepala keluarga berpendidikan dasar, jarak rumah <1 km, tidak memiliki asuransi atau hanya asuransi swasta, terdapat riwayat hipertensi, dan pengeluaran <1 juta per bulan. Penerapan SMOTE-N meningkatkan performa model *Random Forest* dengan akurasi seimbang 78.53% dan F1-Score 62.86%. Nilai SHAP mengidentifikasi jarak tempat tinggal, kepemilikan jaminan kesehatan, dan besaran pengeluaran sebagai prediktor yang paling berkontribusi terhadap hasil klasifikasi. Penelitian ini merekomendasikan agar pemerintah dapat meningkatkan akses geografis ke Puskesmas, memperluas jangkauan layanan bagi pengguna jaminan kesehatan, dan memberikan subsidi bagi keluarga kurang mampu.

Kata Kunci: *Shapley Additive Explanations*, *Random Forest*, Pemilihan Fasilitas Kesehatan, SMOTE-N

ABSTRACT

Application of Shapley Additive Explanations (SHAP) on the *Random Forest Model for Classification of Health Facility Selection* (By: Angelina Ivanna Genardi; Supervisors: Selvi Annisa and Rifqi Aulya Rahman, 2025; 52 pages)

Health service equity depends on community behavior in choosing healthcare facilities. Equal access has not been achieved due to utilization disparities among population groups. The choice of healthcare facilities is influenced by various socioeconomic factors. Most previous studies used classical classification models. This study aims to develop a classification model for healthcare facility selection in Basirih Subdistrict using the Random Forest algorithm with Shapley Additive Explanations (SHAP). Descriptive analysis was conducted to describe family socioeconomic characteristics based on healthcare facility choices. Class imbalance was handled using the Synthetic Minority Oversampling Technique – Nominal (SMOTE-N). SHAP was applied to explain each predictor's contribution to classification results. Descriptive analysis shows that families choosing community health centers (Puskesmas) are generally headed by individuals with basic education, live less than 1 km away, lack or only have private insurance, have a history of hypertension, and monthly expenditures below one million rupiah. Applying SMOTE-N improved Random Forest performance with a balanced accuracy of 78.53% and an F1-score of 62.86%. SHAP identified residential distance, health insurance ownership, and expenditure level as the most influential predictors. This study recommends improving access to Puskesmas, expanding coverage for insured users, and providing subsidies for low-income families.

Keywords: Shapley Additive Explanations, Random Forest, Health Facility Selection, SMOTE-N

PRAKATA

Puji syukur kepada Tuhan Yang Maha Esa karena atas berkat, kasih, dan penyertaan-Nya penulis dapat menyelesaikan Tugas Akhir yang berjudul “Penerapan *Shapley Additive Explanations* (SHAP) pada Model *Random Forest* untuk Klasifikasi Pemilihan Fasilitas Kesehatan”. Tugas Akhir ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan program sarjana Strata-1 (S-1) pada Program Studi Statistika FMIPA ULM. Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada pihak-pihak yang selalu mendukung, membantu, serta memberikan perhatian selama proses penulisan Tugas Akhir ini, diantaranya:

1. Kedua orang tua penulis yang memberikan dukungan, perhatian, doa yang tanpa henti dan telah menjadi sumber semangat dan kekuatan bagi penulis sepanjang perjalanan pendidikan hingga berada di tahap ini;
2. Anastasia Devinna Genardi, adik tercinta yang selalu setia memberikan dukungan, perhatian, doa, dan menjadi pendengar yang baik di setiap suka maupun duka;
3. Ibu Selvi Annisa, S.Si., M.Si. dan Bapak Rifqi Aulya Rahman, S.Mat., M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, motivasi, dan dukungan sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik;
4. Ibu Dewi Anggraini, S.Si., M.App.Sci., Ph.D. dan Bapak Oni Soesanto, S.Si., M.Si. selaku dosen penguji yang telah meluangkan waktu dan memberikan masukan untuk perbaikan penulisan Tugas Akhir ini;
5. Bapak/Ibu Dosen serta Tenaga Administrasi di Program Studi Statistika FMIPA ULM yang telah memberikan ilmu, bimbingan, dan dukungan sepanjang masa perkuliahan;
6. Teman-teman Funtastic yang selalu menjadi teman seperjuangan, saling menguatkan, berbagi cerita dalam suka maupun duka sepanjang perjalanan perkuliahan ini;
7. Teman-teman *Intern* BI Booyah atas kebersamaan dan semangat saling menguatkan selama berjuang menyusun Tugas Akhir;
8. Semua pihak yang tidak bisa penulis sebutkan satu per satu, tetapi kontribusinya memberikan makna tersendiri bagi penulis.

Penulis menyadari sepenuhnya bahwa penulisan Tugas Akhir ini masih banyak kekurangan. Oleh karena itu, dengan kerendahan hati penulis menerima segala bentuk kritik dan saran yang dapat menjadi bahan perbaikan tulisan di masa mendatang. Besar harapan penulis Tugas Akhir ini dapat memberikan manfaat, menambah wawasan, dan menjadi referensi yang berguna bagi pembaca di masa mendatang.

Banjarbaru, 10 November 2025

Angelina Ivanna Genardi

DAFTAR ISI

	Hlm.
PERNYATAAN	iv
ABSTRAK.....	v
ABSTRACT	vi
PRAKATA.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Kajian Penelitian Terdahulu.....	5
2.2 Kajian Teori.....	8
2.2.1 Fasilitas Pelayanan Kesehatan.....	8
2.2.2 Model Pemanfaatan Pelayanan Kesehatan	9
2.2.3 Profil Kelurahan Basirih.....	10
2.2.4 Statistika Deskriptif.....	12
2.2.5 <i>Synthetic Minority Oversampling Technique</i> -Nominal	13
2.2.6 Klasifikasi	14
2.2.7 <i>Random Forest</i>	14
2.2.8 <i>Hyperparameter Tuning</i> dan Validasi Silang Berstrata	16
2.2.9 <i>Confusion Matrix</i>	17
2.2.10 <i>Shapley Additive Explanations</i>	18
BAB III METODE PENELITIAN.....	22
3.1 Sumber Data.....	22
3.2 Variabel Penelitian.....	22
3.3 Prosedur Penelitian.....	23
BAB IV HASIL DAN PEMBAHASAN	28
4.1 <i>Preprocessing</i> Data.....	28
4.2 Statistika Deskriptif.....	30
4.3 Pembagian Data <i>Training</i> dan <i>Testing</i>	36
4.4 Teknik <i>Resampling</i> SMOTE-N.....	37
4.4.1 Menghitung Jarak VDM.....	37
4.4.2 Membuat Data Sintesis	38

4.5	Pembentukan Model <i>Random Forest</i>	38
4.5.1	<i>Bootstrap Sampling</i>	39
4.5.2	Pembuatan Pohon Keputusan.....	39
4.5.3	Hasil <i>Hyperparameter Tuning</i>	40
4.6	Evaluasi Model <i>Random Forest Data Testing</i>	40
4.7	SHAP	43
4.7.1	Interpretasi Lokal	43
4.7.2	Interpretasi Global	46
BAB V PENUTUP		51
5.1	Kesimpulan	51
5.2	Saran.....	51
DAFTAR PUSTAKA		53
LAMPIRAN.....		58
RIWAYAT HIDUP		85

DAFTAR GAMBAR

	Hlm.
Gambar 1.1 Wilayah Kelurahan Basirih	1
Gambar 2.1 Distribusi Penduduk Berdasarkan Kelompok Usia	10
Gambar 2.2 Distribusi Penduduk Berdasarkan Tingkat Pendidikan	11
Gambar 2.3 Distribusi Penduduk Berdasarkan Jenis Pekerjaan	11
Gambar 2.4 Ilustrasi Model <i>Random Forest</i>	15
Gambar 2.5 Ilustrasi Pembagian Data 10- Lipatan Validasi Silang Berstrata.	17
Gambar 2.6 Visualisasi Nilai SHAP Menggunakan <i>Bar Plot</i>	20
Gambar 2.7 Visualisasi Nilai SHAP Menggunakan <i>Beeswarm Plot</i>	20
Gambar 2.8 Visualisasi Nilai SHAP Menggunakan <i>Waterfall Plot</i>	21
Gambar 3.1 Prosedur Analisis	27
Gambar 4.1 Diagram Lingkaran Proporsi Pilihan Fasilitas Kesehatan	30
Gambar 4.2 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Tingkat Pendidikan	31
Gambar 4.3 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Jumlah Anggota Keluarga	31
Gambar 4.4 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Jarak	32
Gambar 4.5 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Sarana Transportasi	32
Gambar 4.6 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Kepemilikan Jaminan Kesehatan	33
Gambar 4.7 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Riwayat Batuk/Pilek	34
Gambar 4.8 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Riwayat Diare	34
Gambar 4.9 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Riwayat Hipertensi	35
Gambar 4.10 Proporsi Pilihan Fasilitas Kesehatan Berdasarkan Tingkat Pengeluaran	36
Gambar 4.11 <i>Waterfall Plot</i> Observasi ke- 1: (a) Kelas Bukan Puskesmas (b) Kelas Puskesmas	44
Gambar 4.12 <i>Bar Plot</i> Nilai SHAP Setiap Prediktor	46
Gambar 4.13 <i>Beeswarm Plot</i> Nilai SHAP Setiap Prediktor: (a) Kelas Bukan Puskesmas (b) Kelas Puskesmas	47

DAFTAR TABEL

	Hlm.
Tabel 2.1 Kajian Penelitian Terdahulu.....	5
Tabel 2.2 <i>Confusion Matrix</i>	17
Tabel 3.1 Variabel Penelitian.....	22
Tabel 4.1 Identifikasi Nilai yang Tidak Valid	28
Tabel 4.2 Pembentukan Variabel Jumlah Anggota Keluarga	29
Tabel 4.3 Reklasifikasi Variabel.....	29
Tabel 4.4 Proporsi Pilihan Fasilitas Kesehatan	30
Tabel 4.5 Rasio Pembagian Data.....	36
Tabel 4.6 Ilustrasi Data untuk Menghitung Jarak VDM.....	37
Tabel 4.7 Ilustrasi Data untuk Membuat Data Sintesis	38
Tabel 4.8 Proporsi Kelas Data Sebelum dan Sesudah <i>Resampling</i> SMOTE-N..	38
Tabel 4.9 <i>Hyperparameter Tuning</i>	39
Tabel 4.10 Performa Model pada Data <i>Training</i> Setelah <i>Hyperparameter Tuning</i>	40
Tabel 4.11 <i>Confusion Matrix</i> Model (100, 3, 25, 10) Tanpa <i>Resampling</i> SMOTE-N.....	40
Tabel 4.12 <i>Confusion Matrix</i> Model (100, 3, 25, 10) dengan <i>Resampling</i> SMOTE-N.....	41
Tabel 4.13 Perbandingan Performa Model Sebelum dan Sesudah <i>Resampling</i> SMOTE-N.....	43
Tabel 4.14. Nilai SHAP Setiap Prediktor pada Observasi Pertama	44

DAFTAR LAMPIRAN

	Hlm.
Lampiran 1. Surat Izin Penelitian dan Pengambilan Data.....	58
Lampiran 2. Data Penelitian	59
Lampiran 3. Data Duplikat	60
Lampiran 4. Data yang Tidak Valid	62
Lampiran 5. Data Penelitian Setelah <i>Preprocessing</i>	64
Lampiran 6. Kombinasi Model.....	65
Lampiran 7. Nilai Probabilitas Hasil Klasifikasi Model pada Data <i>Training</i>	67
Lampiran 8. Nilai SHAP Setiap Prediktor untuk Kelas Bukan Puskesmas.....	72
Lampiran 9. Nilai SHAP Setiap Prediktor untuk Kelas Puskesmas.....	74
Lampiran 10. <i>Syntax Python</i> untuk <i>Import Library</i> , <i>Data</i> , dan <i>Preprocessing</i> .	76
Lampiran 11. <i>Syntax Python</i> untuk Statistika Deskriptif.....	78
Lampiran 12. <i>Syntax Python</i> untuk <i>Tuning Hyperparameter</i> Model tanpa SMOTE-N Rasio 70:30	81
Lampiran 13. <i>Syntax Python</i> untuk <i>Tuning Hyperparameter</i> Model dengan SMOTE-N Rasio 70:30	82
Lampiran 14. <i>Syntax Python</i> untuk Evaluasi <i>Data Testing</i> Rasio 70:30	83
Lampiran 15. <i>Syntax Python</i> untuk Nilai SHAP	84

DAFTAR LAMBANG DAN SINGKATAN

$\delta(x_{aj}, x_{bj})$	jarak antar dua nilai kategorik pada prediktor ke- j
ϕ_j	nilai kontribusi <i>Shapley</i> dari prediktor X_j
$\Delta(a, b)$	jarak <i>Value Difference Metric</i> (VDM) antara observasi ke- a dan observasi ke- b
c	jumlah kelas pada variabel target
$\hat{f}(X)$	hasil prediksi akhir untuk observasi X
$f(x_{aj})$	frekuensi kemunculan nilai x_{aj}
$f(x_{aj}, y_m)$	frekuensi nilai x_{aj} yang termasuk ke dalam kelas target ke- m
$f(x_{bj})$	frekuensi kemunculan nilai x_{bj}
$f(x_{bj}, y_m)$	frekuensi nilai x_{bj} yang termasuk ke dalam kelas target ke- m
n	jumlah observasi dalam <i>node</i> N
n_t	jumlah observasi dalam subset N_t setelah di <i>split</i>
n_m	jumlah observasi dalam kelas target ke- m
p	jumlah variabel prediktor
r	parameter jarak
s	jumlah <i>subset</i> hasil pemisahan
w_a	bobot dari observasi ke- a
w_b	bobot dari observasi ke- b
x_{aj}	nilai variabel prediktor ke- j pada observasi ke- a
x_{bj}	nilai variabel prediktor ke- j pada observasi ke- b
$E_X(\hat{f}(X))$	ekspektasi prediksi model terhadap distribusi data
N	himpunan data dalam suatu <i>node</i>
N_t	<i>subset</i> ke- t setelah pemisahan
S	himpunan bagian dari seluruh variabel prediktor kecuali X_j
$val(S)$	prediksi untuk nilai prediktor pada himpunan S yang dimarjinalkan
$val(S \cup \{X_j\})$	ekspektasi prediksi ketika prediktor X_j ditambahkan ke S
CV	<i>Cross Validation</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
SHAP	<i>Shapley Additive Explanations</i>
SMD	Survei Mawas Diri
SMOTE	<i>Synthetic Minority Oversampling Technique</i>
SMOTE-N	<i>Synthetic Minority Oversampling Technique-Nominal</i>
SVM	<i>Support Vector Machine</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>