

INTERPRETASI MODEL *LOGISTIC REGRESSION* MENGGUNAKAN
METODE SHAP DALAM KLASIFIKASI KELAYAKAN
PENERIMA ZAKAT DI DESA SEI RAMPAH

SKRIPSI

DISUSUN OLEH

ANGGI SORAYA

2209020073



UMSU

Unggul | Cerdas | Terpercaya

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2026

INTERPRETASI MODEL *LOGISTIC REGRESSION* MENGGUNAKAN
METODE SHAP DALAM KLASIFIKASI KELAYAKAN
PENERIMA ZAKAT DI DESA SEI RAMPAH

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom)
dalam Program Studi Teknologi Informasi, pada Fakultas Ilmu Komputer dan Teknologi
Informasi, Universitas Muhammadiyah Sumatera Utara

ANGGI SORAYA

2209020073

PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2026

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Judul Skripsi : INTERPRETASI MODEL *LOGISTIC REGRESSION*
MENGUNAKAN METODE SHAP DALAM
KLASIFIKASI KELAYAKAN
PENERIMA ZAKAT DI DESA SEI RAMPAH

Nama Mahasiswa : ANGGI SORAYA
NPM : 2209020073
Program Studi : TEKNOLOGI INFORMASI

Menyetujui
Komisi Pembimbing



(Zuli Agustina Gultom, M.Si.)
NIDN. 0130089003

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom, M.Kom.)
NIDN. 0117019301



Dekan



(Dr. Al-Khoarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

INTERPRETASI MODEL *LOGISTIC REGRESSION* MENGGUNAKAN METODE SHAP DALAM KLASIFIKASI KELAYAKAN PENERIMA ZAKAT DI DESA SEI RANPAH

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 15 April 2026

Yang membuat pernyataan



Anggi Soraya

NPM. 2209020073

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Anggi Soraya
NPM : 2209020073
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

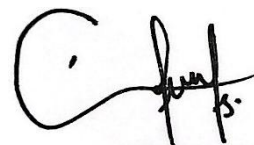
INTERPRETASI MODEL *LOGISTIC REGRESSION* MENGGUNAKAN
METODE SHAP DALAM KLASIFIKASI KELAYAKAN
PENERIMA ZAKAT DI DESA SEI RAMPAH

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 15 April 2026

Yang membuat pernyataan



Anggi Soraya

NPM. 2209020073

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Anggi Soraya
Tempat dan Tanggal Lahir : Tg.Beringin, 11 Januari 2004
Alamat Rumah : Perumahan Residence Bedagai No.3
Telepon/Faks/HP : 085262845835
E-mail : anggisoraya60@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SDN 106217 LR.AMALIUN TAMAT: 2016
SMP : SMP N1 SEI RAMPAH TAMAT: 2019
SMA : MAN 1 MEDAN TAMAT: 2022

KATA PENGANTAR



Assalamualaikum Warahmatullahi Wabarakatuh

Alhamdulillah, penulis lantunkan kehadiran Allah SWT, yang telah memberikan rahmat dan karunia-Nya kepada penulis, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Interpretasi Model Logistic Regression Menggunakan Metode SHAP Dalam Klasifikasi Kelayakan Penerima Zakat di Desa Sei Rampah”. Shalawat dan salam kepada Nabi Muhammad SAW, yang telah membawa kita dari zaman kegelapan menuju zaman yang terang benderang seperti sekarang ini. Penyusunan skripsi ini bertujuan untuk memenuhi salah satu syarat guna mencapai gelar sarjana komputer di Universitas Muhammadiyah Sumatera Utara. Dalam proses penyusunan skripsi ini, penulis menyadari bahwa penyelesaian skripsi ini tidak terlepas dari bantuan, bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Akrim, M.Pd., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firaahmi Rizky, M.Kom. Selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatra Utara.
4. Bapak Mhd. Basri, S.Si, M.Kom. Wakil Dekan III Program Studi Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi.

5. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom. Ketua Program Studi Teknologi Informasi.
6. Bapak Okvi Nugroho S.Kom., M.Kom. Sekretaris Program Studi Teknologi Informasi.
7. Ibu Zuli Agustina Gultom, M.Si Dosen Pembimbing yang senantiasa telah membimbing dan mengarahkan penulis sehingga dapat menyelesaikan skripsi dengan sebaik-baiknya.
8. Seluruh dosen dan staf Fakultas Ilmu Komputer dan Teknologi Informasi yang telah dengan tulus memberikan ilmu pengetahuan, bimbingan, bantuan, serta pelayanan yang baik kepada penulis selama menempuh masa perkuliahan, sehingga penulis dapat belajar, berkembang, dan menyelesaikan pendidikan ini dengan baik.
9. Ucapan terima kasih ini penulis persembahkan kepada kedua orang tua tercinta, Bapak Zainal Abidin dan Ibu Lela Hermayanti Lubis yang selalu menjadi sumber kekuatan, tempat bersandar dan berkeluh kesah. Terima kasih juga atas doa yang tak pernah putus, pengorbanan yang tak usai dalam setiap langkah penulis. Terima kasih juga selalu ada disisi penulis, hal ini menjadi penguat bagi penulis untuk segera menyelesaikan penulisan skripsi untuk mendapatkan gelar sarjana.
10. Ucapan terima kasih ini penulis persembahkan kepada saudara kandung tercinta, Safa Aulia Maika dan Dimas Hermawan, yang selalu memberikan dukungan, doa, dan semangat dalam proses penulisan skripsi ini, serta kepada kedua almarhumah kakak kandung penulis, Ainun Mardiah dan

Abdila Utami, yang senantiasa menjadi inspirasi, kenangan, dan motivasi bagi penulis untuk terus berjuang dan menyelesaikan karya ini.

11. Ucapan terima kasih ini penulis persembahkan kepada teman seperjuangan saya circle GALAXY, yang telah menjadi bagian cerita indah semasa perkuliahan. Terima kasih sudah menjadi tempat berkeluh kesah, tempat bercerita, tempat mengaduh segala permasalahan dan terima kasih juga selalu ada disisi penulis baik suka maupun duka. Terima kasih karena selalu menemani, mendukung, menghibur, dan berjalan bersama dalam setiap proses yang telah dilalui. Semoga persahabatan ini senantiasa terjaga, dan langkah kita semua selalu dimudahkan untuk meraih mimpi-mimpi terbaik di masa depan.
12. Terima kasih ini saya ucapkan kepada seluruh tim kkn mekar sawit atas dukungan, dan motivasi kepada penulis untuk segera menyelesaikan skripsi, semoga pertemanan kita senantiasa terjaga dan kita semua dimudahkan untuk meraih mimpi-mimpi terbaik dimasa yang akan datang.
13. Ucapan terima kasih yang tulus penulis sampaikan kepada teman-teman kelas C1 Para sahabat penulis dibangku perkuliahan yang selalu membersamai dalam empat tahun ini yang memberikan banyak kenangan indah selama perkuliahan. Terima kasih atas kebersamaan, dukungan, dan bantuan yang diberikan, terutama saat penulis mengalami kebuntuan dalam menyelesaikan skripsi ini.
14. Penulis juga mengucapkan terima kasih kepada teman-teman seangkatan penulis sejak awal perkuliahan Syahdah Ruri Ariska, Tiara Amanda, Dea Salsabila Gorat dan Aulia Ramadani, serta pihak-pihak lain yang tidak dapat penulis

sebutkan satu per satu, namun telah hadir dan memberikan dukungan, semangat, perhatian, serta bantuan kepada penulis, terutama pada masa-masa sulit dan titik terendah dalam proses penyusunan skripsi ini. Kehadiran dan kebaikan kalian menjadi kekuatan tersendiri bagi penulis untuk tetap bertahan, bangkit, dan menyelesaikan skripsi ini hingga akhir.

15. Terakhir, ucapan terima kasih kepada diri sendiri atas perjuangan, kesabaran, dan ketekunan dalam menyelesaikan skripsi ini. Tidak sedikit rintangan, keraguan, kelelahan, bahkan air mata yang menyertai setiap proses penulisan skripsi ini. Namun penulis tetap berusaha dan bertahan hingga sampai ketahap ini. Semoga setiap perjalanan ini mejadi pengingat bagi penulis bahwa setiap perjuangan tidak akan pernah sia-sia dan akan selalu memberikan makna serta pembelajaran yang berharga dimasa depan.

**INTERPRETASI MODEL *LOGISTIC REGRESSION* MENGGUNAKAN
METODE SHAP DALAM KLASIFIKASI KELAYAKAN
PENERIMA ZAKAT DI DESA SEI RAMPAH**

ABSTRAK

Penyaluran zakat yang tepat sasaran memerlukan proses penentuan penerima yang objektif, konsisten, dan dapat dipertanggungjawabkan. Penelitian ini bertujuan untuk membangun model Logistic Regression dalam mengklasifikasikan kelayakan penerima zakat di Desa Sei Rampah serta menginterpretasikan hasil prediksi menggunakan metode SHAP (Shapley Additive Explanations). Data yang digunakan merupakan data sekunder calon penerima zakat yang diperoleh dari pihak pengelola atau panitia penyaluran zakat di Desa Sei Rampah sebanyak 1.000 data, dengan variabel yang meliputi pendapatan, jumlah tanggungan, pekerjaan, kondisi tempat tinggal, dan kategori asnaf.

Tahapan penelitian meliputi preprocessing data, standarisasi struktur data, pembersihan data, encoding variabel kategorik, normalisasi fitur numerik, pembagian data latih dan data uji dengan proporsi 80:20, pembangunan model Logistic Regression, serta evaluasi model menggunakan accuracy, precision, recall, dan F1-score. Hasil evaluasi menunjukkan bahwa model menghasilkan accuracy sebesar 99%, precision sebesar 100%, recall sebesar 98,28%, dan F1-score sebesar 99,13%.

Selain menghasilkan performa klasifikasi yang sangat baik, penelitian ini juga menunjukkan bahwa metode SHAP mampu memberikan interpretasi yang transparan terhadap keputusan model, baik secara global maupun lokal. Pada interpretasi global, fitur yang paling berpengaruh terhadap prediksi kelayakan

adalah pendapatan, jumlah tanggungan, kondisi tempat tinggal, dan pekerjaan. Pada interpretasi lokal, SHAP mampu menjelaskan kontribusi masing-masing fitur terhadap prediksi individu tertentu. Dengan demikian, kombinasi Logistic Regression dan SHAP dapat digunakan sebagai pendekatan yang efektif dalam membantu proses klasifikasi kelayakan penerima zakat secara akurat, transparan, dan dapat dipertanggungjawabkan.

Kata kunci: Logistic Regression, SHAP, Klasifikasi, Kelayakan Penerima Zakat, Machine Learning

Interpretation of the Logistic Regression Model Using the SHAP Method in the Classification of Zakat Recipient Eligibility in Sei Rampah Village

ABSTRACT

Accurate zakat distribution requires a recipient determination process that is objective, consistent, and accountable. This study aims to build a Logistic Regression model for classifying zakat recipient eligibility in Sei Rampah Village and to interpret the prediction results using the SHAP (Shapley Additive Explanations) method. The data used in this study is secondary data of zakat recipient candidates obtained from the zakat management committee in Sei Rampah Village, consisting of 1,000 records, with variables including income, number of dependents, occupation, housing condition, and asnaf category.

The research stages include data preprocessing, data structure standardization, data cleaning, categorical variable encoding, numerical feature normalization, splitting the dataset into training and testing sets with an 80:20 proportion, building the Logistic Regression model, and evaluating the model using accuracy, precision, recall, and F1-score. The evaluation results show that the model achieved 99% accuracy, 100% precision, 98.28% recall, and 99.13% F1-score.

In addition to achieving excellent classification performance, this study also shows that the SHAP method is able to provide transparent interpretations of the model's decisions, both globally and locally. In the global interpretation, the most influential features in predicting eligibility are income, number of dependents, housing condition, and occupation. In the local interpretation, SHAP is able to explain the contribution of each feature to the prediction for specific individuals. Therefore, the combination of Logistic Regression and SHAP can be used as an effective approach

to support the classification of zakat recipient eligibility in an accurate, transparent, and accountable manner.

Keywords: Logistic Regression, SHAP, Classification, Zakat Recipient Eligibility, Machine Learning

DAFTAR ISI

LEMBAR PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI	iv
RIWAYAT HIDUP	v
KATA PENGANTAR	vi
ABSTRAK	x
ABSTRACT	xii
DAFTAR ISI	xiv
DAFTAR TABEL	xvii
DAFTAR GAMBAR	xviii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
BAB II	7
LANDASAN TEORI	7
2.1 Konsep Zakat dan Penentuan Kelayakan Penerima Zakat	7
2.2 Data Mining dan Klasifikasi	7
2.2.1 Tahapan Data Mining	8
2.3 Logistic Regression	9
2.4 SHAP (Shapley Additive Explanations)	11
2.5 Penelitian Terdahulu	12
2.6 ANALISIS GAP	15
BAB III	16
METODOLOGI PENELITIAN	16
3.1 Jenis Penelitian	16
3.2 Analisis Alat Penelitian	16
3.2.1 Perangkat Keras(Hardware)	17
3.2.2 Perangkat Lunak (Software)	17
3.3 Objek Penelitian dan Sumber Data	18
3.4 Proses Pembuatan Model	19
3.4.1 Variabel Penelitian	19

3.4.2	Preprocessing Data	19
3.4.3	Pembagian Data Latih & Uji.....	22
3.4.4	Evaluasi Kinerja Model.....	22
3.5	Diagram Alur	25
3.5.1	Diagram.....	25
3.5.2	Flowchart	27
3.5.3	Usecase Admin dan Pengguna	28
3.6	Tampilan UI/UX	29
3.6.1	Halaman Login.....	29
3.6.2	Halaman Dashboard.....	30
3.6.3	Halaman Data Calon Penerima Zakat	30
3.6.4	Halaman Proses Model (<i>Logistic Regression</i>)	31
3.6.5	Halaman Interpretasi Metode SHAP.....	32
3.6.7	Halaman Login Pengguna	33
3.7	Jadwal Penelitian	34
BAB IV	35
HASIL DAN PEMBAHASAN	35
4.1	Gambaran Umum Data Penelitian	35
4.2	Hasil Preprocessing Data	36
4.2.1	Tampilan Data Awal Sebelum Preprocessing.....	36
4.2.2	Standarisasi Struktur Data dan Penamaan Kolom	36
4.2.3	Pembersihan Data dan Pemeriksaan Missing Value	37
4.2.4	Validasi Nilai Tidak Wajar dan Konsistensi Data	39
4.2.5	Seleksi Fitur	39
4.2.6	Encoding Variabel Kategorik.....	41
4.2.7	Normalisasi Fitur Numerik	42
4.2.9	Pembagian Data Latih dan Data Uji.....	43
4.3	Hasil Pemodelan Logistic Regression.....	44
4.3.1	Pelatihan Model	44
4.3.2	Hasil Prediksi pada Data Uji.....	44
4.4.1	Confusion Matrix	51
4.5	Interpretasi Hasil Menggunakan SHAP	54
4.5.1	Interpretasi Global (Global Explanation).....	54
4.5.2	Interpretasi Lokal (Local Explanation)	56
4.6.2	Kontribusi SHAP dalam Meningkatkan Transparansi	59
4.6.3	Implikasi Praktis bagi Pengelolaan Zakat di Desa Sei Rampah.....	60
4.6.4	Keterbatasan Penelitian.....	61
BAB V	51

KESIMPULAN DAN SARAN.....	51
5.1 Kesimpulan	51
5.2 Saran	52
DAFTAR PUSTAKA	54
Lampiran 1. Sk Penetapan Dosen Pembimbing.....	58
Lampiran 2. Hasil Cek Turnitin	60

DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian Terdahulu	12
Tabel 3.1 Spesifikasi Laptop.....	17
Tabel 3.2 Jadwal Penelitian.....	34
Tabel 4.1 Tampilan Data Awal Sebelum Preprocessing.....	36
Tabel 4.2 Hasil Pemeriksaan Missing Value	37
Tabel 4.3 Hasil Pemeriksaan Data Duplikat	38
Tabel 4.4 Hasil Validasi Nilai Tidak Wajar.....	39
Tabel 4.5 Hasil Seleksi Fitur.....	40
Tabel 4.6 Hasil Standarisasi Fitur Numerik.....	42
Tabel 4.7 Hasil Standarisasi Fitur Numerik.....	43
Tabel 4.8 Hasil Metrik Evaluasi Model Logistic Regression	54

DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian.....	25
Gambar 3. 2 flowchart Logistic Regression dan SHAP.....	27
Gambar 3. 3 Usecase Admin dan Pengguna	28
Gambar 3. 4 Halaman Login.....	29
Gambar 3. 5 Halaman Dashboard	30
Gambar 3. 6 Halaman Data Calon Penerima Zakat	30
Gambar 3. 7 Proses Model (Logistic Regression)	31
Gambar 3. 8 Halaman Interpretasi Metode SHAP.....	32
Gambar 3. 9 Halaman Hasil	32
Gambar 3. 10 Halaman Login Pengguna Calon Penerima Zakat	33
Gambar 4. 1 Tampilan Data Awal Sebelum Preprocessing.....	36
Gambar 4. 2 Hasil Encoding Variabel Kategorik	41
Gambar 4. 3 Tampilan Data Akhir Setelah Preprocessing	43
Gambar 4. 4 Confusion Matrix Model Logistic Regression.....	51
Gambar 4. 5 Tampilan Grafik SHAP Lokal	56

BAB I

PENDAHULUAN

1.1 Latar Belakang

Zakat merupakan kewajiban bagi umat Islam yang memiliki fungsi sosial-ekonomi penting, terutama untuk membantu mengurangi kesenjangan dan mendukung kesejahteraan masyarakat melalui penyaluran kepada pihak yang berhak. Dalam konteks desa, persoalan utama penyaluran zakat sering kali bukan semata pada ketersediaan dana, melainkan pada ketepatan penentuan penerima. Ketika proses seleksi masih mengandalkan penilaian manual, keputusan berpotensi dipengaruhi subjektivitas panitia serta tidak didasarkan pada perhitungan yang konsisten, sehingga dapat menimbulkan ketidaktepatan sasaran dan memunculkan pertanyaan dari masyarakat (Aristanto et al., 2024). Ketepatan dalam menentukan penerima zakat menjadi hal yang sangat penting karena zakat tidak hanya berfungsi sebagai bentuk ibadah, tetapi juga sebagai instrumen pemerataan kesejahteraan sosial di masyarakat. Apabila penyaluran tidak dilakukan secara tepat sasaran, maka tujuan zakat untuk membantu masyarakat yang benar-benar membutuhkan dapat menjadi kurang optimal. Oleh sebab itu, diperlukan suatu pendekatan yang mampu mendukung proses penentuan penerima zakat secara lebih objektif, terukur, dan konsisten.

Untuk mengurangi subjektivitas tersebut, berbagai penelitian telah mengembangkan Sistem Pendukung Keputusan (SPK) berbasis kriteria. SPK membuat proses seleksi lebih terstruktur karena setiap calon penerima dinilai berdasarkan indikator yang ditetapkan, lalu dihitung menjadi nilai akhir atau peringkat prioritas. Salah satu pendekatan yang sering digunakan adalah metode

pembobotan seperti *SMART* (*simple multi-attribute rating technique*) untuk menghasilkan skor kelayakan yang menjadi dasar penentuan penerima zakat (Okta Felani, 2024). Namun, SPK berbobot umumnya masih sangat bergantung pada penentuan bobot awal dan asumsi penilai. Ketika karakteristik warga berubah, jumlah variabel bertambah, atau pola kemiskinan di masyarakat semakin beragam, pendekatan berbobot dapat menjadi kurang adaptif dan berisiko mempertahankan bias penilaian yang sama dalam bentuk yang “terhitung”. Seiring meningkatnya ketersediaan data sosial-ekonomi di tingkat desa, seleksi penerima zakat fitrah dituntut semakin konsisten, transparan, dan dapat dipertanggungjawabkan. Karena itu, pendekatan machine learning mulai dimanfaatkan untuk mengklasifikasikan kelayakan berdasarkan pola pada data dan kinerjanya dapat diuji menggunakan metrik evaluasi. Namun, agar keputusan dapat dipertanggungjawabkan kepada warga, hasil klasifikasi perlu disertai penjelasan sehingga SHAP digunakan untuk menunjukkan variabel yang paling berpengaruh secara umum dan alasan prediksi pada tiap individu.

Salah satu penelitian menggunakan algoritma *Naïve Bayes* untuk mengelompokkan calon penerima ke dalam kelas “layak” dan “tidak layak” serta mengevaluasi kinerja model menggunakan *confusion matrix* (Salsabila & Feli Ramury, 2024). Penelitian lainnya menunjukkan bahwa kinerja algoritma klasifikasi bisa berbeda di setiap dataset, sehingga pemilihan model sebaiknya didasarkan pada hasil pengujian dan perbandingan performa. Hal ini juga terlihat pada penelitian Ayatulloh dkk. (2025) yang membandingkan lima algoritma klasifikasi pada sentimen ulasan pengguna Netflix dan memperoleh variasi akurasi, dengan *Logistic Regression* dan *Random Forest* mencapai akurasi tertinggi sebesar

76% (Ayatulloh et al., 2025). Temuan tersebut menunjukkan bahwa *Logistic Regression* dapat bersifat kompetitif dibanding algoritma lain pada studi tertentu. Dalam konteks data sosial-ekonomi, Melani dkk. (2024) membandingkan *Naive Bayes* dan *Logistic Regression* pada klasifikasi penerima bantuan iuran jaminan kesehatan (BPJS-PBI) di NTB, dengan akurasi *Naive Bayes* sebesar 63,90% dan *Logistic Regression* 62,26%. Hasil ini menegaskan bahwa pemilihan model untuk klasifikasi kelayakan pada data sosial-ekonomi perlu diuji pada data setempat (Melani et al., 2024).

Oleh karena itu, penelitian ini memilih *Logistic Regression* sebagai model utama untuk mengklasifikasikan kelayakan penerima zakat, kemudian menginterpretasikan hasilnya menggunakan SHAP agar keputusan “layak/tidak layak” tidak hanya berupa label, tetapi juga dapat dijelaskan faktor-faktor yang mendorongnya. Keunggulan SHAP terletak pada kemampuannya memberikan penjelasan baik secara global (variabel yang paling berpengaruh pada model) maupun secara lokal (alasan prediksi pada individu tertentu). Selain itu, sifatnya yang aditif dan konsisten memungkinkan kontribusi tiap variabel dijumlahkan hingga membentuk prediksi akhir, sehingga interpretasi menjadi lebih transparan dan mudah dipertanggungjawabkan kepada pihak terkait (Kristanaya et al., 2025).

1.2 Rumusan Masalah

Penelitian ini merumuskan permasalahan yang akan dibahas sebagai dasar pelaksanaan penelitian. Berikut ini rumusan masalah dalam penelitian ini:

1. Bagaimana membangun model *Logistic Regression* untuk mengklasifikasikan kelayakan penerima zakat di Desa Sei Rampah?

2. Bagaimana kinerja model *Logistic Regression* dalam mengklasifikasikan kelayakan penerima zakat berdasarkan metrik evaluasi (misalnya akurasi, *precision*, *recall*, *F1-score*)
3. Bagaimana metode SHAP menginterpretasikan hasil prediksi model, baik untuk mengetahui variabel paling berpengaruh (interpretasi global) maupun alasan prediksi pada individu tertentu (interpretasi lokal)?

1.3 Batasan Masalah

Penelitian ini memiliki batasan permasalahan yang dibahas agar dapat menyelesaikan permasalahan utama. Berikut ini batasan masalah penelitian ini:

1. Objek penelitian dibatasi pada klasifikasi kelayakan penerima zakat di Desa Sei Rampah menggunakan data calon penerima zakat yang tersedia pada periode penelitian, dengan keluaran dua kelas yaitu “layak” dan “tidak layak” sesuai label/acuan pada data.
2. Model yang digunakan dibatasi pada *Logistic Regression* sebagai model utama untuk melakukan klasifikasi kelayakan penerima zakat.
3. Interpretasi hasil model dibatasi menggunakan metode SHAP (global dan lokal), serta evaluasi kinerja model dibatasi pada metrik evaluasi klasifikasi (misalnya akurasi, *precision*, *recall*, *F1-score*) berdasarkan skema pembagian data latih dan data uji.

1.4 Tujuan Penelitian

Penelitian ini memiliki tujuan yang ingin dicapai sebagai arah dan fokus pelaksanaan penelitian. Berikut ini tujuan penelitian dalam penelitian ini:

1. Membangun model *Logistic Regression* untuk mengklasifikasikan kelayakan penerima zakat di Desa Sei Rampah berdasarkan data calon penerima zakat yang tersedia.
2. Mengevaluasi kinerja model *Logistic Regression* menggunakan metrik evaluasi klasifikasi (misalnya akurasi, precision, recall, F1-score).
3. Menginterpretasikan hasil prediksi model menggunakan SHAP untuk mengetahui variabel yang paling berpengaruh (interpretasi global) serta alasan prediksi pada individu tertentu (interpretasi lokal) sebagai dasar penjelasan yang lebih transparan dan dapat dipertanggungjawabkan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat baik secara praktis maupun akademis sesuai dengan hasil yang diperoleh. Berikut ini manfaat penelitian ini:

1. Memberikan gambaran dan panduan dalam membangun model *Logistic Regression* untuk mengklasifikasikan kelayakan penerima zakat di Desa Sei Rampah.
2. Memberikan informasi mengenai kinerja model *Logistic Regression* dalam mengklasifikasikan kelayakan penerima zakat berdasarkan metrik evaluasi (misalnya akurasi, precision, recall, F1-score), sehingga hasilnya dapat dijadikan pertimbangan dalam evaluasi kinerja model.
3. Memberikan penjelasan yang lebih transparan melalui metode SHAP untuk mengetahui variabel yang paling berpengaruh (interpretasi global) serta alasan prediksi pada individu tertentu (interpretasi lokal), sehingga keputusan “layak/tidak layak” lebih mudah dipahami dan dipertanggungjawabkan,

termasuk membantu panitia menjelaskan dasar keputusan kepada warga/pihak terkait.

BAB II

LANDASAN TEORI

2.1 Konsep Zakat dan Penentuan Kelayakan Penerima Zakat

Zakat adalah kewajiban bagi umat Islam yang memiliki peran signifikan dalam mengurangi kemiskinan meningkatkan kesejahteraan Masyarakat (Utami et al., 2024). Dalam praktik pendistribusiannya, khususnya di tingkat desa, tantangan tidak hanya terletak pada keterbatasan dana, tetapi juga pada ketepatan penentuan kelayakan penerima agar bantuan tepat sasaran. Rizki & Samsudin. (2024) menegaskan bahwa pendistribusian zakat sering menghadapi tantangan dalam menentukan penerima yang tepat dan alokasi yang optimal, sehingga diperlukan pendekatan yang lebih sistematis dalam pengambilan keputusan. Penentuan kelayakan penerima zakat umumnya dilakukan berdasarkan indikator sosial-ekonomi yang menggambarkan kondisi kebutuhan warga, seperti pendapatan, pekerjaan, jumlah tanggungan, kondisi tempat tinggal, dan kondisi kesehatan. Penggunaan indikator ini membuat proses seleksi lebih jelas dan terukur karena keputusan didasarkan pada kriteria yang dapat diperiksa dan dibandingkan, sehingga mengurangi ketergantungan pada penilaian subjektif (Rizki & Samsudin, 2024).

2.2 Data Mining dan Klasifikasi

Data mining adalah proses untuk mengekstraksi informasi atau pengetahuan baru dari kumpulan data berukuran besar dengan memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan machine learning, sehingga menghasilkan pola yang bermanfaat dan dapat digunakan dalam pengambilan keputusan. Dalam prosesnya, data mining umumnya berjalan melalui rangkaian langkah mulai dari

pemilihan data, preprocessing untuk meningkatkan kualitas data, transformasi data, proses pemodelan (data mining), hingga interpretasi dan evaluasi hasil agar pengetahuan yang diperoleh dapat dipahami dan digunakan (Faran & Aldisa, 2023).

Dalam penelitian ini, tugas data mining yang digunakan adalah klasifikasi, yaitu membangun model untuk memetakan data calon penerima zakat ke dalam kelas target berdasarkan atribut sosial-ekonomi. Klasifikasi termasuk kategori supervised learning karena membutuhkan variabel target/label (misalnya “layak” dan “tidak layak”) untuk membentuk model yang dapat memprediksi kelas pada data baru (Faran & Aldisa, 2023). Secara umum, klasifikasi juga merupakan teknik penting dalam data mining untuk menghasilkan informasi yang berguna dari data dalam jumlah besar melalui pembentukan model prediktif (Yoga Religia, Agung Nugroho, 2021).

2.2.1 Tahapan Data Mining

- a. Data Pemilihan (*selection*) menentukan data yang relevan dari sumber yang tersedia (data calon penerima zakat) dan menetapkan fitur sosial-ekonomi yang dipakai serta target kelas “layak” dan “tidak layak”.
- b. Data *Cleaning/Preprocessing* membersihkan data agar kualitasnya baik, misalnya menangani data hilang (*missing*), data duplikat, nilai tidak wajar, serta merapikan penulisan kategori agar konsisten (Faran & Aldisa, 2023).
- c. Data *Transformation* mengubah data ke bentuk yang siap diproses model, misalnya *encoding* variabel kategorik menjadi numerik dan melakukan standarisasi/normalisasi fitur numerik bila diperlukan sebelum pemodelan klasifikasi (Faran & Aldisa, 2023).

- d. *Data Mining/Modeling* melatih model klasifikasi (pada penelitian kamu menggunakan *Logistic Regression*) untuk mempelajari pola hubungan antara variabel sosial-ekonomi dan label kelayakan.
- e. *Evaluasi Model (Evaluation)* mengukur performa model menggunakan metrik evaluasi klasifikasi seperti akurasi, *precision*, *recall*, dan *F1-score* pada data uji agar terlihat kemampuan generalisasi model (Nabilawati et al., 2025).
- f. *Penyajian Pengetahuan (Knowledge Presentasion)* menyajikan hasil model agar mudah dipahami pihak terkait, misalnya melalui interpretasi fitur dan alasan prediksi menggunakan SHAP sehingga keputusan model lebih transparan dan dapat dijelaskan.

2.3 Logistic Regression

Logistic Regression adalah metode klasifikasi (terutama klasifikasi biner) yang digunakan ketika variabel target memiliki dua kategori, misalnya 1 = “layak” dan 0 = “tidak layak”. Dalam regresi logistik biner, tujuan model bukan menghasilkan nilai kontinu seperti pada regresi linear, melainkan memperkirakan peluang (probabilitas) suatu data termasuk ke kelas 1 berdasarkan variabel prediktor. Oleh karena itu, regresi logistik digunakan ketika keluaran hanya memiliki dua kategori dan ingin diperoleh output berupa probabilitas yang terletak pada rentang 0–1. Dalam penerapannya, regresi logistik biner dapat digunakan untuk menilai pengaruh berbagai faktor terhadap status kelayakan (“layak/tidak layak”), yang dapat disertai evaluasi kecocokan model (misalnya uji Hosmer–Lemeshow), pengujian signifikansi koefisien parameter (misalnya uji Wald), serta

interpretasi besaran pengaruh variabel melalui *odds ratio* atau $\text{Exp}(B)$ sebagai perbandingan peluang kejadian terhadap non-kejadian (Gultom et al., 2024).

a. Bentuk model logistic regression sebagai berikut:

$$\pi(x_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi})} \quad (1)$$

Keterangan:

$\pi(x_i)$ = probabilitas kejadian/kelas target pada data ke- i

\exp = fungsi eksponensial (basis e)

β_0 = konstanta (intercept)

$\beta_1, \beta_2, \dots, \beta_p$ = koefisien regresi untuk variabel bebas ke-1 sampai ke- p

$x_{1i}, x_{2i}, \dots, x_{pi}$ = nilai variabel bebas ke-1 sampai ke- p pada data ke- i

b. Dengan fungsi logitnya yaitu:

$$g(x_i) = \ln \left[\frac{\pi(x_i)}{1 - \pi(x_i)} \right] = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} \quad (2)$$

Keterangan:

$g(x_i)$ = fungsi logit (log-odds) untuk data ke- i

\ln = logaritma natural

$\pi(x_i)$ = probabilitas kejadian/kelas target pada data ke- i

$1 - \pi(x_i)$ = probabilitas bukan kelas target pada data ke- i

$\frac{\pi(x_i)}{1 - \pi(x_i)}$ = odds (perbandingan peluang kelas target terhadap non-target)

β_0 = konstanta (intercept)

$\beta_1, \beta_2, \dots, \beta_p$ = koefisien regresi untuk masing-masing variabel bebas

$x_{1i}, x_{2i}, \dots, x_{pi}$ = nilai variabel bebas ke-1 sampai ke- P pada data ke- i

Pada penelitian ini, bentuk probabilitas dan *logit* digunakan untuk memodelkan peluang status kelayakan (layak/tidak layak) berdasarkan variabel-variabel sosial-ekonomi sebagai prediktor. (Innassuraiya et al., 2022).

2.4 SHAP (Shapley Additive Explanations)

SHAP merupakan salah satu metode interpretasi model (*explainability*) yang bertujuan menjelaskan kontribusi setiap variabel terhadap prediksi (Kristanaya et al., 2025). SHAP merupakan salah satu metode interpretasi model (*explainability*) yang bertujuan menjelaskan kontribusi setiap variabel terhadap prediksi. Dengan demikian, model tidak hanya menghasilkan label, tetapi juga memberikan alasan fitur mana yang paling memengaruhi prediksi, baik secara umum maupun pada setiap individu. SHAP merupakan pendekatan berbasis teori permainan yang menghitung kontribusi setiap variabel terhadap output model untuk masing-masing data. Melalui analisis SHAP nilai SHAP yang positif menunjukkan bahwa suatu variabel meningkatkan kecenderungan prediksi menuju kelas tertentu, sedangkan nilai SHAP yang negatif menunjukkan bahwa variabel tersebut menurunkan kecenderungan prediksi ke kelas tersebut. Dengan adanya penjelasan ini, SHAP memberikan nilai tambah dalam interpretasi hasil model karena keputusan “layak/tidak layak” dapat dijelaskan faktor pendorongnya secara lebih transparan dan mudah dipertanggungjawabkan (Hadi et al., 2021).

a. Rumus aditif *SHAP*

$$[f(x) = E[f(X)] + \sum_{j=1}^p \phi_j] \quad (3)$$

Keterangan:

$f(x)$ = output/nilai prediksi model untuk data x

$E[f(X)]$ = nilai harapan (rata-rata teoretis) dari keluaran fungsi F ketika masukan X merupakan peubah acak sesuai distribusinya.

$\sum_{j=1}^p \phi_j$ = total kontribusi seluruh fitur terhadap prediksi, sehingga jika dijumlahkan dengan ϕ_0 menghasilkan prediksi akhir $f(x)$

b. Rumus *Shapley* (nilai SHAP)

$$[\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)]] \quad (4)$$

Keterangan:

ϕ_j = nilai SHAP (nilai *Shapley*) untuk fitur/variabel ke- j

$\sum_{S \subseteq F \setminus \{j\}}$ = penjumlahan untuk semua subset S dari himpunan fitur F yang tidak memuat fitur j

$\frac{|S|!(|F| - |S| - 1)!}{|F|!}$ = bobot/koeffisien *Shapley*

$f_S(x_S)$ = output model dengan fitur subset S

$f_{S \cup \{j\}}(x_{S \cup \{j\}})$ = output model saat subset S ditambah fitur j (Khairunnisa, 2023).

2.5 Penelitian Terdahulu

Tabel 2.1 Ringkasan Penelitian Terdahulu

NO	Judul dan peneliti	Pembahasan	Metode	Kelebihan atau Kekurangan
1.	Sistem Pendukung Keputusan pendistribusian zakat menggunakan metode simple multi attribute rating technique (smart) (Okta Felani, 2024).	Penelitian ini membangun sistem pendukung keputusan untuk membantu proses penentuan kelayakan penerima zakat berdasarkan beberapa kriteria sosial-ekonomi. Sistem menghasilkan	<i>SMART (simple multi-attribute rating technique)</i>	Kelebihan: Perhitungan sederhana, mudah diterapkan, dan menghasilkan skor/ranking yang jelas. Kekurangan: Sangat bergantung pada bobot kriteria yang ditetapkan penilai, sehingga berpotensi subjektif. Belum menggunakan model klasifikasi yang belajar dari

NO	Judul dan peneliti	Pembahasan	Metode	Kelebihan atau Kekurangan
		nilai akhir tiap calon penerima sehingga dapat dibuat peringkat sebagai dasar keputusan.		data (misalnya <i>Logistic Regression</i>) dan belum menyediakan interpretasi kontribusi fitur per individu seperti SHAP.
2.	Sistem Pendukung Keputusan Pendistribusian Zakat Menggunakan Metode <i>SMART (simple multi-attribute rating technique)</i> dan <i>SAW (simple additive weighting)</i> (INFORMASI) (Rizki & Samsudin, 2024).	Penelitian ini menggabungkan pembobotan kriteria dan perhitungan skor total untuk merangking calon penerima zakat. Tujuannya agar seleksi lebih sistematis dan terdokumentasi dibanding penilaian manual.	<i>SMART (simple multi-attribute rating technique)</i>	Kelebihan: Proses seleksi lebih terukur dan konsisten karena keputusan didasarkan pada skor akhir yang dihitung dari kriteria yang ditetapkan. Kekurangan: Masih sensitif pada penentuan bobot awal dan kurang adaptif ketika kondisi/karakteristik data warga berubah. Belum menekankan interpretasi alasan keputusan pada level individu (<i>local explanation</i>).
3.	Sistem Pendukung Keputusan Penentuan Penerima Zakat Menggunakan Metode <i>AHP (analytic hierarchy process)</i> (Hadr omi et al., 2025).	Penelitian ini membuat sistem pendukung keputusan untuk menentukan peringkat calon penerima zakat dengan menentukan bobot kriteria melalui perbandingan berpasangan. Kriteria yang digunakan menggambarkan kondisi sosial-	<i>AHP (analytic hierarchy process)</i> .	Kelebihan: Pembobotan lebih sistematis karena menggunakan perbandingan berpasangan dan uji konsistensi, sehingga mengurangi subjektivitas dibanding penentuan bobot secara langsung. Kekurangan: Tetap bergantung pada penilaian ahli/penentu bobot. Belum menggunakan pendekatan machine learning seperti

NO	Judul dan peneliti	Pembahasan	Metode	Kelebihan atau Kekurangan
		ekonomi (misalnya penghasilan, pekerjaan, tanggungan, kondisi tempat tinggal, kondisi kesehatan).		<i>Logistic Regression</i> dan belum ada penjelasan faktor pendorong keputusan secara global-lokal seperti SHAP.
4.	Implementasi Data Mining dengan Algoritma <i>Naive Bayes</i> untuk Klasifikasi Kelayakan Penerima Bantuan Zakat Fitrah (Salsabila & Feli Ramury, 2024).	Penelitian ini menerapkan data mining untuk mengklasifikasikan kelayakan penerima zakat fitrah menjadi dua kelas (layak dan tidak layak). Kinerja model dievaluasi menggunakan confusion matrix dan metrik klasifikasi (misalnya <i>akurasi</i> , <i>precision</i> , <i>recall</i>).	<i>Naive Bayes</i> .	Kelebihan: Implementasi sederhana dan komputasinya cepat, cocok sebagai model klasifikasi dasar. Kekurangan: Asumsi independensi antar fitur sering kurang sesuai untuk data sosial-ekonomi yang saling berkaitan. Fokus utama pada hasil klasifikasi dan metrik evaluasi, belum menekankan interpretasi keputusan secara rinci (global dan per individu) seperti SHAP.
5.	Implementasi Metode AHP (<i>analytic hierarchy process</i>) dan SMART (<i>simple multi-attribute rating technique</i>) untuk Penentuan Keputusan Calon Penerima Zakat Produktif (Rafiq et al., 2024).	Penelitian ini mengombinasikan AHP (<i>analytic hierarchy process</i>) untuk menentukan bobot kriteria dan SMART (<i>simple multi-attribute rating technique</i>) untuk menghitung nilai akhir alternatif sehingga menghasilkan peringkat calon penerima zakat produktif.	AHP (<i>analytic hierarchy process</i>) dan SMART (<i>simple multi-attribute rating technique</i>)	Kelebihan: Bobot kriteria ditentukan lebih terstruktur AHP (<i>analytic hierarchy process</i>) dan hasil keputusan berupa nilai akhir/ranking yang jelas SMART (<i>simple multi-attribute rating technique</i>) Kekurangan: Keputusan tetap sangat dipengaruhi bobot kriteria. Belum menggunakan model klasifikasi yang belajar dari data (misalnya <i>Logistic Regression</i>) dan belum ada interpretasi kontribusi variabel pada tiap individu seperti SHAP.

2.6 ANALISIS GAP

Penelitian terdahulu menunjukkan bahwa upaya mengurangi subjektivitas seleksi penerima zakat banyak dilakukan melalui Sistem Pendukung Keputusan (SPK) berbasis pembobotan kriteria. (Okta Felani, 2024) menyebut panitia masih menggunakan “persepsi atau anggapan” dalam memilih calon penerima, sehingga mereka membangun SPK pendistribusian zakat menggunakan metode *SMART* (*simple multi-attribute rating technique*). Hasil pengujian sistem menghasilkan nilai akhir untuk beberapa calon penerima, misalnya Wardoyo (0,6125), Mugiyem (0,8875), dan Kawit (0,5306) sebagai dasar keputusan kelayakan.

Penelitian lainnya oleh Salsabila dkk. (2024) menerapkan data mining dengan algoritma *Naïve Bayes* untuk mengklasifikasikan kelayakan penerima zakat fitrah (kelas “layak” dan “tidak layak”). Evaluasi menggunakan *confusion matrix* menunjukkan akurasi 76,47%, *precision* 92,85%, dan *recall* 81,25%. Akurasi menggambarkan proporsi prediksi yang benar dibanding seluruh data uji, sedangkan *precision* dan *recall* menunjukkan ketepatan prediksi pada kelas target serta kemampuan model menangkap kasus yang benar di kelas tersebut (Salsabila & Feli Ramury, 2024). Namun, penelitian ini masih menggunakan *Naïve Bayes* dan konteks data dari pengurus masjid, sehingga belum menguji *Logistic Regression* sebagai model utama dan belum menekankan interpretasi keputusan (alasan “layak/tidak layak” pada tiap individu) dengan pendekatan seperti SHAP yang menjadi fokus pada penelitian ini di Desa Sei Rampah.

BAB III

METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif terapan yang menggunakan pendekatan komputasional untuk mengklasifikasikan kelayakan penerima zakat di Desa Sei Rampah berdasarkan data sosial-ekonomi calon penerima. Pendekatan kuantitatif dipilih karena data yang digunakan diolah melalui tahapan data mining, kemudian direpresentasikan dalam bentuk numerik (misalnya pendapatan, jumlah tanggungan, kondisi tempat tinggal, pekerjaan, dan variabel sosial-ekonomi lain) agar dapat diproses oleh algoritma *Logistic Regression* untuk menghasilkan klasifikasi “layak” dan “tidak layak” secara otomatis. Seluruh proses analisis dilakukan secara sistematis dan terukur, mulai dari pengumpulan data, preprocessing, pembagian data latih dan data uji, pelatihan model, hingga evaluasi menggunakan metrik kinerja klasifikasi seperti akurasi, precision, recall, F1-score. Selain menghasilkan prediksi, penelitian ini juga menekankan transparansi keputusan dengan menginterpretasikan hasil model menggunakan SHAP untuk menjelaskan kontribusi variabel secara global maupun alasan prediksi pada tingkat individu. Dengan demikian, penelitian ini tidak hanya berfokus pada ketepatan klasifikasi, tetapi juga pada kemampuan model untuk memberikan penjelasan yang dapat dipertanggungjawabkan dalam konteks penyaluran zakat.

3.2 Analisis Alat Penelitian

Pada proses analisis ini dijelaskan alat yang digunakan dalam penelitian, yang terbagi menjadi perangkat keras (*hardware*) dan perangkat lunak (*software*).

3.2.1 Perangkat Keras(Hardware)

Perangkat keras yang digunakan dalam penelitian ini yaitu:

1. Laptop

Menggunakan laptop HP dengan spesifikasi sebagai berikut:

Tabel 3.1 Spesifikasi Laptop

Komponen	Spesifikasi
Mode Sistem	HP PROBOOK 440
Prosesor	Intel COREi5 Gen 11
Storage	256 GB SSD
Sistem Operasi	Windows 11
Dimensi	32.19 x 21.39 x 1.99 cm
Memori	8 GB RAM

3.2.2 Perangkat Lunak (Software)

Dalam penelitian ini digunakan beberapa perangkat lunak, antara lain:

1. Python

Python digunakan sebagai bahasa pemrograman utama dalam penelitian ini. Python dimanfaatkan untuk melakukan preprocessing data, seperti pembersihan data, penanganan missing value, encoding variabel kategorik, dan pembagian data latih serta data uji. Selain itu, Python juga digunakan untuk membangun model Logistic Regression, menghitung metrik evaluasi seperti akurasi, precision, recall, dan F1-score, serta menampilkan interpretasi hasil prediksi menggunakan metode SHAP.

2. Visual Studio Code

Visual Studio Code digunakan sebagai editor kode untuk menulis, mengedit, dan menjalankan program yang dibuat menggunakan bahasa Python. Dalam penelitian ini, Visual Studio Code berfungsi untuk membantu proses preprocessing data, pembuatan model Logistic Regression, evaluasi kinerja model, serta implementasi interpretasi hasil menggunakan metode SHAP. Selain itu, perangkat lunak ini juga digunakan dalam pengembangan antarmuka sistem agar proses klasifikasi kelayakan penerima zakat dapat dijalankan dengan lebih terstruktur dan efisien.

3.3 Objek Penelitian dan Sumber Data

Objek pada penelitian ini adalah data calon penerima zakat di Desa Sei Rampah yang digunakan untuk memprediksi kelas kelayakan, yaitu “layak” (1) atau “tidak layak” (0). Sumber data berasal dari dataset sekunder yang diperoleh dari pihak pengelola/panitia penyaluran zakat di Desa Sei Rampah, yang berisikan variabel sosial-ekonomi sebagai fitur analisis, seperti pendapatan/penghasilan, jenis pekerjaan, jumlah tanggungan keluarga, kondisi tempat tinggal, serta indikator lain yang tersedia pada data. Label kelayakan pada penelitian ini mengacu pada keputusan/ketetapan yang sudah ada pada data (acuan penilaian pengelola zakat). Apabila pada data belum tersedia label kelayakan, maka label dibentuk secara operasional berdasarkan kriteria kelayakan yang ditetapkan pada periode penelitian (misalnya berdasarkan batas penghasilan dan indikator kondisi ekonomi tertentu) agar terbentuk target klasifikasi biner yang konsisten.

3.4 Proses Pembuatan Model

3.4.1 Variabel Penelitian

1. Variabel Independen (Variabel Bebas)
 - a. Indikator sosial-ekonomi calon penerima zakat (misalnya pendapatan atau penghasilan, jenis pekerjaan jumlah tanggungan keluarga, kondisi tempat tinggal, serta indikator lain yang tersedia pada data).
 - b. Karakteristik dan kelengkapan data calon penerima zakat (jumlah data yang digunakan, variasi kondisi sosial-ekonomi warga, kelengkapan atribut serta kualitas data seperti adanya nilai kosong atau data tidak konsisten).
 - c. Pada penelitian ini, model Logistic Regression digunakan dengan parameter bawaan dari sistem, dengan pembagian data latih dan data uji sebesar 80:20 serta threshold klasifikasi 0,5.

2. Variabel Dependen (Variabel Terikat)

Variabel ini menunjukkan status kelayakan calon penerima zakat berdasarkan indikator sosial-ekonomi yang digunakan dalam penelitian. Kelayakan diklasifikasikan ke dalam dua kelas, yaitu: Layak (1) dan Tidak Layak (0). Nilai variabel terikat ini menjadi target yang diprediksi oleh model *Logistic Regression* dan kemudian dijelaskan alasannya menggunakan SHAP.

3.4.2 Preprocessing Data

Preprocessing dilakukan untuk memastikan data calon penerima zakat bersih, konsisten, dan siap dipakai pada pemodelan *Logistic Regression* untuk klasifikasi kelayakan (“layak/tidak layak”). Langkah preprocessing yang disesuaikan dengan kondisi dataset penelitian:

1. Standarisasi struktur data dan penanam kolom pada dataset calon penerima zakat, dilakukan penyeragaman format agar data mudah diproses, yaitu:
 - a. Penerima nama kolom (trim/strip spasi, menyamakan huruf besar-kecil, menghapus symbol yang tidak perlu).
 - b. Memastikan setiap kolom memiliki nama yang jelas (menghapus kolom “Unnamed”/kolong kosong jika ada).
 - c. Menyeragamkan penulisan nilai kategori (misalnya pada pekerjaan/status rmas/kondisi rumah) agar tidak terjadi duplikasi kategori karena perbedaan penulisan.
2. Pembersihan data dan pemeriksaan missing value/duplikasi
Dataset diperiksa untuk memastikan tidak ada nilai kosong (missing) dan data ganda (duplikat). Jika ditemukan missing value, penanganannya dapat berupa imputasi (mengisi nilai) atau menghapus baris sesuai kebutuhan dan alasan metodologis. Jika ditemukan duplikasi (data orang tercatat lebih dari sekali), data duplikat dihapus agar tidak menyebabkan bias.
3. Validasi nilai tidak wajar dan konsistensi data sosial-ekonomi dilakukan
Pengecekan nilai yang tidak logis, misalnya:
 - a. Pendapatan < 0 atau pendapatan sangat ekstrem (outlier).
 - b. Jumlah tanggungan < 0 atau tidak masuk akal.
 - c. Usia (jika ada) di luar batas wajar.
 - d. Data kategori yang tidak sesuai (misalnya status rumah di luar pilihan yang ditetapkan) jika ditemukan nilai tidak wajar, dilakukan koreksi berdasarkan sumber data atau penanganan outlier (misalnya winsorizing atau penghapusan data ekstrem) agar model tidak bias.

4. Seleksi fitur (menghapus kolom yang tidak relevan)

Kolom yang tidak berhubungan dengan penentuan kelayakan dihapus, misalnya nomor urut, ID internal, catatan bebas yang tidak terstruktur, atau kolom administrasi yang tidak digunakan sebagai prediktor. Setelah itu, fitur (X) ditetapkan berdasarkan indikator sosial-ekonomi yang tersedia, misalnya pendapatan, pekerjaan, jumlah tanggungan, kondisi tempat tinggal, dan variabel lain yang ada pada dataset. Label target (Y) ditetapkan sebagai kelas “layak” dan “tidak layak”.

5. Pembentukan/validasi label kelayakan (jika diperlukan)

Jika dataset sudah memiliki label “layak/tidak layak”, maka dilakukan pengecekan konsistensi label terhadap aturan/kriteria yang berlaku. Jika label belum tersedia, maka label dibentuk secara operasional berdasarkan acuan/kriteria penyaluran zakat yang digunakan di Desa Sei Rampah, sehingga target klasifikasi menjadi jelas dan terukur.

6. Encoding variable kategorik

Karena Logistic Regression membutuhkan input numerik, variabel kategorik seperti pekerjaan, status rumah, jenis tempat tinggal, atau kategori kondisi rumah diubah menjadi numerik. Umumnya digunakan One-Hot Encoding agar kategori tidak diperlakukan memiliki urutan tertentu (misalnya “buruh” tidak dianggap lebih besar dari “petani”).

7. Normalisasi/standarisasi fitur numerik

Jika skala fitur numerik berbeda jauh (misalnya pendapatan jauh lebih besar dibanding jumlah tanggungan), dilakukan standarisasi/normalisasi agar pelatihan *Logistic Regression* lebih stabil dan koefisien model lebih mudah

dibandingkan. Proses ini biasanya dilakukan dalam pipeline agar konsisten antara data latih dan data uji.

8. Pembagian data latih dan data uji

Setelah preprocessing selesai, dataset dibagi menjadi data latih dan data uji (misalnya 80:20 atau 90:10). Data latih digunakan untuk membangun model *Logistic Regression*, sedangkan data uji digunakan untuk mengukur performa model secara objektif sebelum dilakukan interpretasi menggunakan SHAP.

3.4.3 Pembagian Data Latih & Uji

Dataset pada penelitian ini dibagi menjadi:

1. Data latih (*training*) untuk melatih model *Logistic Regression* dalam mengklasifikasikan kelayakan penerima zakat.
2. Data uji (*testing*) untuk menguji kemampuan generalisasi model pada data yang belum pernah dilihat saat pelatihan. Mengikuti praktik umum pada penelitian klasifikasi, pembagian data dapat menggunakan 80% untuk training dan 20% untuk testing (atau alternatif 90%:10% jika jumlah data terbatas). Pembagian sebaiknya dilakukan dengan stratified splitting agar proporsi kelas “layak” dan “tidak layak” pada data latih dan data uji tetap seimbang, sehingga evaluasi model lebih adil dan tidak bias pada salah satu kelas.

3.4.4 Evaluasi Kinerja Model

Tahap evaluasi sangat penting untuk mengetahui sejauh mana model dapat memprediksi dengan benar. Evaluasi dilakukan menggunakan confusion matrix, yang membandingkan hasil prediksi model dengan data yang sebenarnya. Metrik

utama yang sering digunakan untuk evaluasi adalah akurasi, presisi, recall dan F1-Score.

1. Akurasi (*Accuracy*)

Akurasi mengukur seberapa banyak prediksi yang benar dibandingkan dengan total keseluruhan data. Ini dihitung dengan rumus :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Diketahui :

- a. TP = True Positive (prediksi positif yang benar)
- b. TN = True Negative (prediksi negative yang benar)
- c. FP = False Positive (prediksi positif yang salah)
- d. FN = False Negative (prediksi negative yang salah)

2. Presisi (*Precision*)

Presisi menunjukkan seberapa akurat prediksi positif dari model. Artinya, dari semua data yang diprediksi sebagai positif, berapa persen yang benar-benar positif.

Rumusnya:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3. Recall

Recall mengukur seberapa baik model dalam menemukan semua data yang benar-benar positif. Rumusnya :

$$Recall = (TP)/(TP + FN) \quad (7)$$

4. F1-Score

Fokus utama dalam penelitian ini adalah pada metrik F1-score, karena F1-score memberikan penilaian yang seimbang antara presisi dan recall. Presisi mengukur

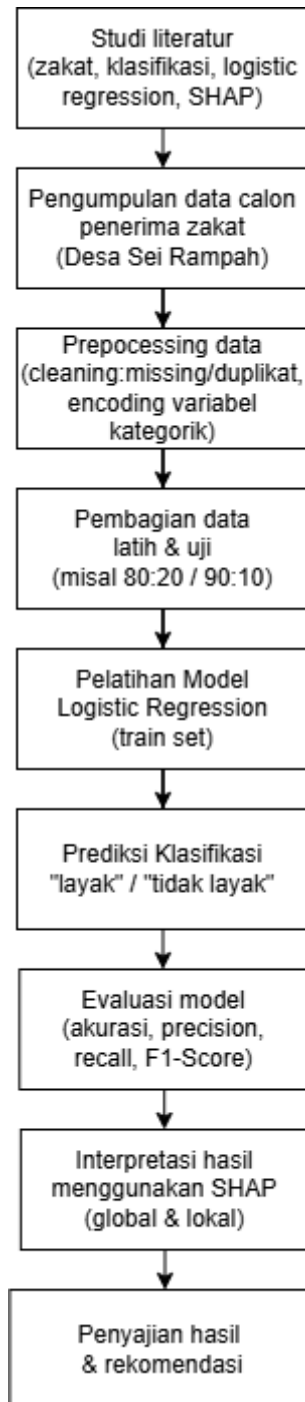
berapa banyak prediksi positif yang benar, sedangkan recall mengukur berapa banyak data positif yang berhasil teridentifikasi. Rumus F1-score adalah :

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

F1-score sangat berguna dalam kondisi data tidak seimbang, di mana akurasi bisa menyesatkan (Abdi et al., 2023).

3.5 Diagram Alur

3.5.1 Diagram Alur Penelitian

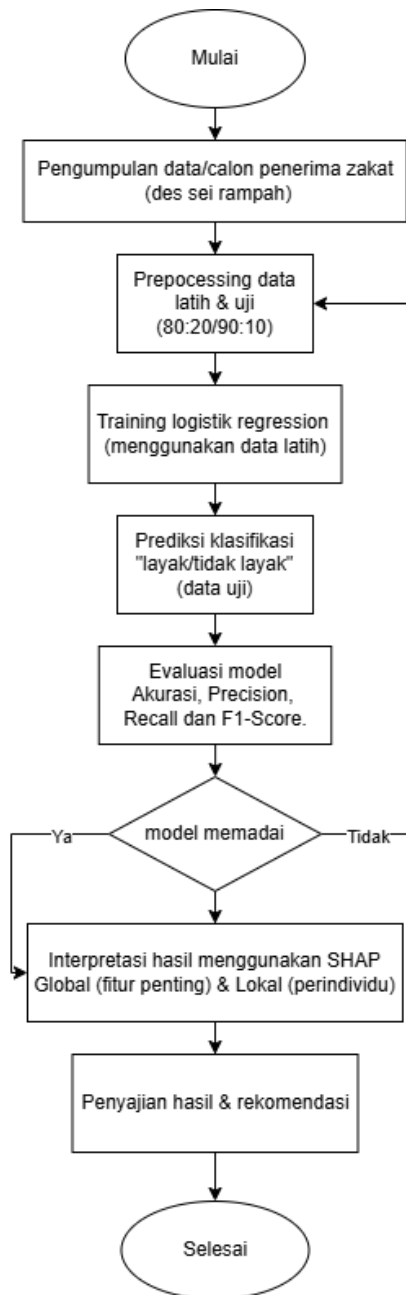


Gambar 3. 1 Diagram Alur Penelitian

Diagram alur penelitian ini menunjukkan tahapan penelitian yang dimulai dari studi literatur mengenai zakat, klasifikasi, Logistic Regression, dan SHAP.

Selanjutnya dilakukan pengumpulan data calon penerima zakat di Desa Sei Rampah, kemudian data masuk ke tahap preprocessing untuk proses cleaning, penanganan missing value dan duplikat, serta encoding pada variabel kategorik. Setelah itu, data dibagi menjadi data latih dan data uji, lalu dilakukan pelatihan model Logistic Regression menggunakan data latih. Model yang telah dilatih kemudian digunakan untuk melakukan prediksi klasifikasi layak dan tidak layak, yang selanjutnya dievaluasi menggunakan akurasi, precision, recall, dan F1-score. Tahap berikutnya adalah interpretasi hasil menggunakan SHAP secara global dan lokal, kemudian hasil akhir disajikan dalam bentuk hasil dan rekomendasi.

3.5.2 Flowchart

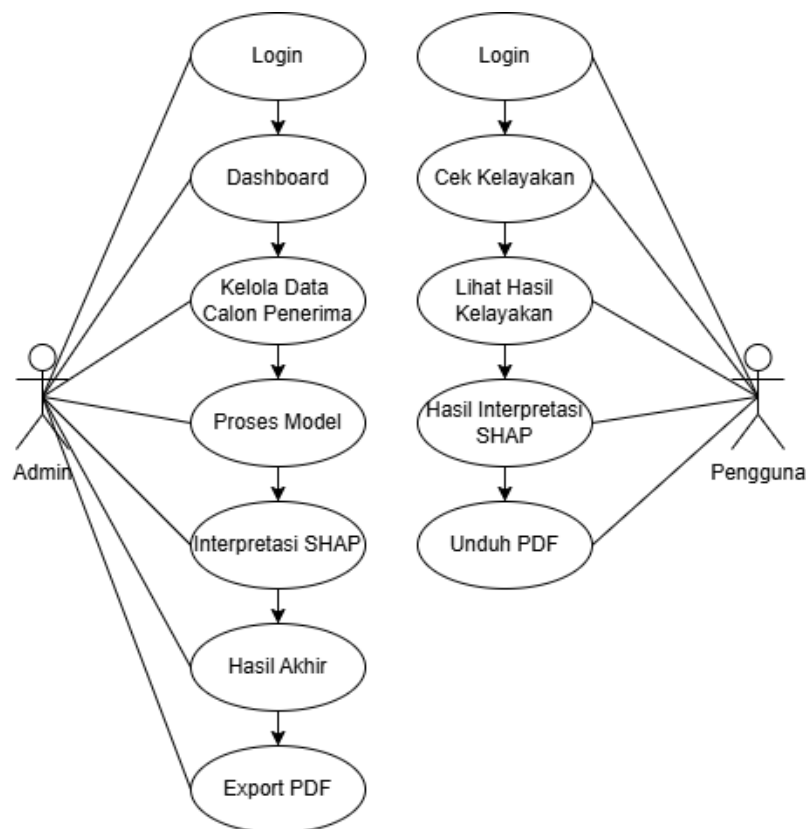


Gambar 3. 2 flowchart Logistic Regression dan SHAP

Flowchart ini menggambarkan alur proses sistem dimulai dari pengumpulan data calon penerima zakat, kemudian data melalui tahap preprocessing dan dibagi menjadi data latih dan data uji dengan komposisi 80:20. Selanjutnya dilakukan training Logistic Regression menggunakan data latih, lalu model digunakan untuk

melakukan prediksi klasifikasi layak atau tidak layak pada data uji. Hasil prediksi kemudian dievaluasi menggunakan metrik akurasi, precision, recall, dan F1-score. Jika model dinilai belum memadai, proses kembali ke tahap pengolahan data, sedangkan jika model sudah memadai maka dilanjutkan ke tahap interpretasi hasil menggunakan SHAP, baik secara global untuk melihat pengaruh fitur secara umum maupun secara lokal untuk melihat penjelasan pada individu tertentu. Setelah itu, sistem menampilkan hasil dan rekomendasi sebelum proses berakhir pada tahap selesai.

3.5.3 Usecase Admin dan Pengguna



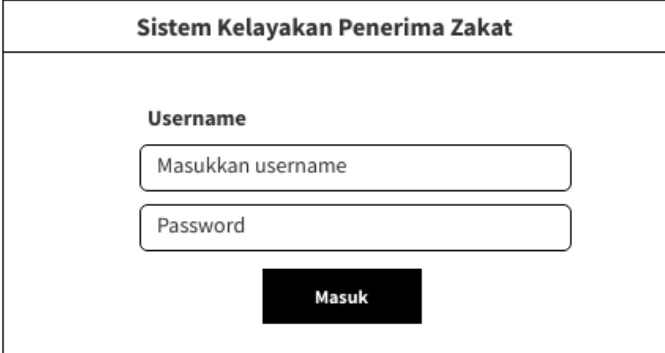
Gambar 3.3 Usecase Admin dan Pengguna

Diagram usecase ini menggambarkan interaksi antara dua aktor, yaitu Admin dan Pengguna, dalam sistem kelayakan penerima zakat. Admin memiliki akses untuk login, melihat dashboard, mengelola data calon penerima, menjalankan

proses model, melihat interpretasi SHAP, menampilkan hasil akhir, dan melakukan export PDF. Sementara itu, Pengguna dapat login, melakukan cek kelayakan, melihat hasil kelayakan, melihat hasil interpretasi SHAP, serta mengunduh PDF hasil pemeriksaan. Diagram ini menunjukkan bahwa admin berperan dalam pengelolaan dan pengolahan data, sedangkan pengguna berfokus pada pengecekan dan melihat hasil kelayakan.

3.6 Tampilan UI/UX

3.6.1 Halaman Login



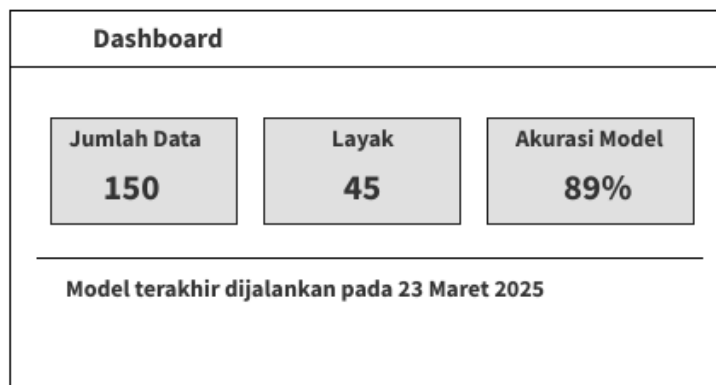
The image shows a login form for a system titled "Sistem Kelayakan Penerima Zakat". The form contains the following elements:

- A header bar with the text "Sistem Kelayakan Penerima Zakat".
- A label "Username" above a text input field containing the placeholder "Masukkan username".
- A label "Password" above a text input field containing the placeholder "Password".
- A black button with the white text "Masuk" centered below the input fields.

Gambar 3. 4 Halaman Login

Halaman login berfungsi untuk mengautentikasi admin sebelum masuk ke sistem. Form login hanya terdiri atas dua input utama, yaitu username dan password, serta satu tombol masuk.

3.6.2 Halaman Dashboard



Gambar 3. 5 Halaman Dashboard

Halaman Dashboard merupakan halaman utama yang muncul setelah admin login. Pada halaman ini ditampilkan Dashboard, tiga ringkasan informasi yaitu Jumlah Data, Layak, dan Akurasi Model yang berfungsi untuk memberikan informasi singkat mengenai total data, jumlah data yang dinyatakan layak, serta tingkat akurasi model. Serta keterangan di bagian bawah yang menunjukkan bahwa model terakhir dijalankan pada 23 Maret 2025.

3.6.3 Halaman Data Calon Penerima Zakat

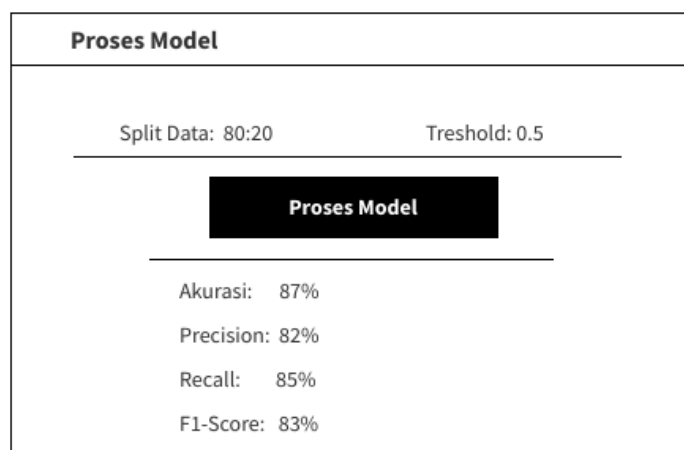
Data Calon Penerima				
		Upload CSV	Tambah Data	
ID	Nama	Pendapatan	Tanggungan	Pekerjaan
001	Ahmad	1,200,000	3	Petani
002	Siti	800,000	2	Rumah Tangga
		Edit	Hapus	

Gambar 3. 6 Halaman Data Calon Penerima Zakat

Halaman Data Calon Penerima digunakan untuk menampilkan data calon penerima zakat yang telah dimasukkan ke dalam sistem. Pada halaman ini tersedia

tombol Upload CSV untuk menambahkan data secara massal dan Tambah Data untuk input data secara manual. Data yang tersimpan ditampilkan dalam bentuk tabel dengan kolom ID, Nama, Pendapatan, Tanggungan, dan Pekerjaan, serta disediakan tombol Edit dan Hapus untuk mengelola data.

3.6.4 Halaman Proses Model (*Logistic Regression*)



Gambar 3. 7 Proses Model (Logistic Regression)

Halaman Proses Model digunakan untuk menjalankan pengujian model pada data yang telah tersedia. Pada halaman ini ditampilkan pembagian data 80:20 dan nilai threshold 0,5 sebagai parameter proses, kemudian admin dapat menekan tombol Proses Model untuk menjalankan model. Setelah proses selesai, sistem menampilkan hasil evaluasi berupa akurasi 87%, precision 82%, recall 85%, dan F1-score 83% sebagai informasi kinerja model.

3.6.5 Halaman Interpretasi Metode SHAP

Interpretasi SHAP

Cari ID>Nama (misal:001 atau Ahmad)

SHAP Global ▼

Klik untuk melihat grafik kontribusi fitur seluruh data

SHAP Lokal ▼

Data Hasil

ID : 001 Nama : Ahmad Statu : Layak Probabilitas : 76%	Penjelasan: <ul style="list-style-type: none"> • Pendapatan rendah meningkatkan peluang layak • Tanggungan tinggi meningkatkan peluang layak
---	--

Gambar 3. 8 Halaman Interpretasi Metode SHAP

Halaman Interpretasi SHAP digunakan untuk menjelaskan hasil prediksi model secara lebih rinci. Pada halaman ini tersedia kolom pencarian untuk mencari data berdasarkan ID atau nama, kemudian bagian SHAP Global dapat diklik untuk menampilkan grafik kontribusi fitur secara keseluruhan, sedangkan bagian SHAP Lokal dapat diklik untuk menampilkan grafik penjelasan prediksi pada data individu yang dipilih. Di bagian bawah, Data Hasil menampilkan identitas penerima, status kelayakan, nilai probabilitas, serta faktor-faktor yang memengaruhi hasil prediksi.

3.6.6 Halaman Hasil

Hasil Akhir

Cari Id>Nama (contoh: 001 atau Ahmad)

ID	Nama	Probabilitas	Status
001	Ahmad	0.76	Layak
002	Siti	0.30	Tidak Layak

Export PDF

Gambar 3. 9 Halaman Hasil

Halaman Hasil Akhir digunakan untuk menampilkan hasil akhir klasifikasi dalam bentuk tabel yang berisi ID, Nama, Probabilitas, dan Status kelayakan penerima zakat. Pada halaman ini juga tersedia kolom pencarian untuk mencari data berdasarkan ID atau nama, serta tombol Export PDF untuk mengekspor hasil yang telah diperoleh.

3.6.7 Halaman Login Pengguna

The image shows two side-by-side screenshots of a web application interface. The left screenshot, titled "Cek Kelayakan Penerima Zakat", contains four input fields: "Nama Lengkap" (Masukkan Nama Lengkap), "Desa" (Masukkan nama desa), "Penghasilan" (Masukkan penghasilan bulanan), and "Jumlah Tanggungan" (Masukkan Jumlah Tanggungan). Below these fields is a black button labeled "Cek Kelayakan". The right screenshot, titled "Hasil Kelayakan", displays the results for a user named Ahmad. It shows "Status : ✓ Layak" and "Probabilitas : 76%". Below this, a "Keterangan:" section explains that the data meets the eligibility criteria based on system prediction. At the bottom, there are two buttons: "Lihat Interpretasi SHAP" and "Unduh PDF".

Gambar 3. 10 Halaman Login Pengguna Calon Penerima Zakat

Halaman ini digunakan oleh pengguna untuk melakukan pengecekan kelayakan penerima zakat dengan mengisi data nama lengkap, desa, penghasilan, dan jumlah tanggungan. Setelah menekan tombol Cek Kelayakan, sistem akan menampilkan hasil kelayakan berupa Nama, status, dan nilai probabilitas, disertai keterangan singkat hasil prediksi. Selain itu, pengguna juga dapat menekan tombol Lihat Interpretasi SHAP untuk melihat penjelasan hasil prediksi dan tombol Unduh PDF untuk menyimpan hasil pemeriksaan.

3.7 Jadwal Penelitian

Tabel 3.2 Jadwal Penelitian

No	Kegiatan	Waktu				
		September	Oktober	November	Desember	Januari
1	Studi Literatur					
2	Pengajuan Judul					
3	Riset Awal					
4	Pembuatan Proposal					

BAB IV

HASIL DAN PEMBAHASAN

4.1 Gambaran Umum Data Penelitian

Penelitian ini menggunakan data calon penerima zakat di Desa Sei Rampah yang diperoleh dari pihak pengelola atau panitia penyaluran zakat setempat. Data yang digunakan merupakan data sekunder yang memuat informasi sosial-ekonomi calon penerima zakat dan digunakan sebagai dasar untuk melakukan klasifikasi status kelayakan ke dalam dua kelas, yaitu layak dan tidak layak.

Dataset awal yang diperoleh berjumlah 1.000 baris data dengan 9 atribut. Variabel yang digunakan dalam penelitian ini mencakup indikator sosial-ekonomi yang relevan terhadap penentuan kelayakan penerima zakat, seperti pendapatan atau penghasilan, jumlah tanggungan keluarga, jenis pekerjaan, kondisi tempat tinggal, serta variabel lain yang tersedia pada dataset. Adapun variabel target pada penelitian ini adalah status kelayakan calon penerima zakat, dengan kode 1 untuk kategori layak dan 0 untuk kategori tidak layak.

Setelah dilakukan tahapan preprocessing, diperoleh dataset akhir yang siap digunakan dalam pemodelan Logistic Regression. Dataset akhir berjumlah 1.000 data dengan 7 fitur prediktor dan satu label target. Data inilah yang kemudian digunakan untuk proses pelatihan, pengujian, evaluasi model, dan interpretasi hasil menggunakan metode SHAP.

4.2 Hasil Preprocessing Data

4.2.1 Tampilan Data Awal Sebelum Preprocessing

No	Nama Calon Penerima	Alamat	Desa	Pendapatan/Penghasilan	Jenis Pekerjaan	Jumlah Tanggungan	Kondisi Tempat Tinggal	Kategori Asnaf
1	SUYITNO	DUSUN III CEMPEDAK LOE	CEMPEDAK LOBANG	Rp3.232.952	Petani	4	Sewa/Kontrak	Miskin
2	MASLIK	DSN I BELIDAN	CEMPEDAK LOBANG	Rp2.699.081	Buruh Pabrik	6	Sewa/Kontrak	Miskin
3	JUNAI	DUSUN IV	CEMPEDAK LOBANG	Rp3.148.520	Petani	4	Sewa/Kontrak	Miskin
4	JUNAI	DUSUN III CEMPEDAK LOE	CEMPEDAK LOBANG	Rp4.459.495	Wiraswasta	4	Milik Pribadi/Sendiri	Tidak Termasuk Asnaf
5	JAKA SYAHPUTRA	DUSUN V	CEMPEDAK LOBANG	Rp1.905.243	Buruh Harian	4	Sewa/Kontrak	Fakir
6	NURHAYSAH	DSN II CEMP LOBANG	CEMPEDAK LOBANG	Rp4.068.066	Wiraswasta	1	Milik Pribadi/Sendiri	Tidak Termasuk Asnaf
7	JUMEN	DUSUN II CEMPEDAK LOB	CEMPEDAK LOBANG	Rp2.982.217	Wiraswasta	5	Sewa/Kontrak	Miskin

Gambar 4. 1 Tampilan Data Awal Sebelum Preprocessing

Pada Gambar 4.1 diatas, menunjukkan data awal sebelum dilakukan preprocessing. Pada tahap ini, data masih disajikan sesuai dengan format awal dari dataset, seperti nomor urut, nama calon penerima, alamat, desa, pendapatan, pekerjaan, jumlah tanggungan, kondisi tempat tinggal, dan kategori asnaf. Sebelum digunakan pada model Logistic Regression, data tersebut perlu melalui proses preprocessing agar struktur data menjadi lebih rapi, konsisten, dan sesuai dengan kebutuhan pemodelan.

4.2.2 Standarisasi Struktur Data dan Penamaan Kolom

Pada tahap awal, dilakukan pemeriksaan terhadap struktur dataset untuk memastikan konsistensi format dan penamaan kolom. Proses ini mencakup penghapusan spasi berlebih pada nama kolom, penyeragaman penggunaan huruf kapital, serta penghapusan simbol-simbol yang tidak diperlukan. Kolom-kolom yang tidak memiliki nama (unnamed) atau kosong dihapus dari dataset untuk menghindari kesalahan dalam pemrosesan selanjutnya.

Tabel 4. 1 Tampilan Data Awal Sebelum Preprocessing

No	Nama Kolom Sebelum Standarisasi	Nama Kolom Setelah Standarisasi
1	Nama Calon Penerima	nama
2	Desa	desa

3	Pendapatan/Penghasilan	pendapatan
4	Jumlah Tanggungan	jumlah_tanggungan
5	enis Pekerjaan	pekerjaan
6	Kondisi Tempat Tinggal	kondisi_tempat_tinggal
7	Kategori Asnaf	asnaf

Pada Tabel 4.1 diatas, penamaan kolom pada dataset telah diseragamkan agar lebih mudah diproses menggunakan Python. Penyeragaman dilakukan dengan mengubah huruf menjadi huruf kecil, mengganti spasi dengan tanda garis bawah, serta menyederhanakan nama kolom tanpa mengubah makna dari atribut tersebut. Hasil dari tahapan ini adalah dataset dengan struktur kolom yang rapi dan konsisten, siap untuk diproses pada tahap berikutnya.

4.2.3 Pembersihan Data dan Pemeriksaan Missing Value

Dataset diperiksa secara menyeluruh untuk mengetahui apakah terdapat nilai kosong (*missing value*) maupun data ganda atau duplikat. Pemeriksaan *missing value* dilakukan pada setiap kolom yang terdapat dalam dataset, seperti nama, desa, pendapatan, jumlah tanggungan, pekerjaan, kondisi tempat tinggal, dan asnaf. Selain itu, pemeriksaan data duplikat juga dilakukan untuk memastikan bahwa tidak ada calon penerima zakat yang tercatat lebih dari satu kali dalam dataset. Berdasarkan hasil pemeriksaan, tidak ditemukan adanya *missing value* pada seluruh atribut yang digunakan.

Tabel 4. 2 Hasil Pemeriksaan Missing Value

No	Nama Kolom	Jumlah Missing Value
1	Nama	0

2	Desa	0
3	Pendapatan	0
4	Jenis Pekerjaan	0
5	Jumlah Tanggungan	0
6	Kondisi Tempat Tinggal	0
7	Asnaf	0

Pada Tabel 4.2 diatas, seluruh kolom yang digunakan dalam penelitian tidak memiliki nilai kosong. Hal ini ditunjukkan dengan jumlah *missing value* pada setiap kolom bernilai 0. Dengan demikian, dataset dapat dilanjutkan ke tahap preprocessing berikutnya tanpa perlu dilakukan pengisian nilai kosong atau penghapusan baris data.

Tabel 4. 3 Hasil Pemeriksaan Data Duplikat

Keterangan	Jumlah Data
Jumlah data awal	1000
Data duplikat	0
Jumlah data setelah penghapusan duplikat	1000

Hal ini menunjukkan bahwa data calon penerima zakat telah terisi dengan lengkap. Selain itu, tidak ditemukan pula data duplikat, sehingga setiap calon penerima zakat hanya terwakili oleh satu baris data. Dengan demikian, dataset dinyatakan layak untuk dilanjutkan ke tahap preprocessing berikutnya tanpa perlu dilakukan imputasi nilai kosong maupun penghapusan data duplikat.

4.2.4 Validasi Nilai Tidak Wajar dan Konsistensi Data

Tabel 4. 4 Hasil Validasi Nilai Tidak Wajar

No	Variabel	Pemeriksaan	Hasil
1	Pendapatan	Nilai negatif	Tidak Ditemukan
2	Pendapatan	Nilai ekstrem/outlier	Tidak ditemukan outlier
3	Jumlah Tanggungan	Nilai negatif	Tidak Ditemukan
4	Jumlah Tanggungan	Nilai tidak wajar	Tidak Ditemukan

Pada tahap ini, dilakukan pemeriksaan terhadap nilai-nilai yang dianggap tidak wajar pada dataset. Pemeriksaan dilakukan untuk memastikan bahwa atribut numerik seperti pendapatan dan jumlah tanggungan tidak memiliki nilai negatif atau nilai yang tidak sesuai dengan kondisi sebenarnya. Pada variabel pendapatan, pemeriksaan dilakukan untuk memastikan tidak terdapat nilai negatif maupun nilai yang terlalu ekstrem. Sementara itu, pada variabel jumlah tanggungan, pemeriksaan dilakukan untuk memastikan bahwa jumlah tanggungan tidak bernilai negatif dan masih berada dalam rentang yang wajar. Berdasarkan hasil validasi, tidak ditemukan nilai negatif maupun nilai tidak wajar pada atribut yang diperiksa. Dengan demikian, data dinyatakan konsisten dan dapat digunakan pada tahap preprocessing berikutnya.

4.2.5 Seleksi Fitur

Seleksi fitur dilakukan untuk menentukan atribut-atribut yang digunakan dalam proses pemodelan Logistic Regression. Pada tahap ini, atribut yang berhubungan dengan penentuan kelayakan penerima zakat dipertahankan, sedangkan atribut yang tidak berpengaruh langsung terhadap proses klasifikasi

tidak digunakan sebagai prediktor. Kolom seperti nomor urut dan alamat tidak digunakan karena hanya berfungsi sebagai informasi administratif dan tidak memberikan kontribusi langsung terhadap keputusan model. Adapun atribut nama tetap disimpan sebagai identitas calon penerima zakat, tetapi tidak digunakan sebagai fitur dalam proses pemodelan. Fitur yang digunakan dalam penelitian ini meliputi desa, pendapatan, pekerjaan, jumlah tanggungan, kondisi tempat tinggal, dan asnaf.

Tabel 4. 5 Hasil Seleksi Fitur

No	Nama Atribut	Keterangan	Status
1	Nomor Urut	Penomoran data calon penerima zakat	Tidak Digunakan
2	Nama	Identitas calon penerima zakat	Digunakan Sebagai Identitas
3	Alamat	Informasi alamat lengkap calon penerima zakat	Tidak Digunakan
4	Desa	Informasi desa calon penerima zakat	Digunakan
5	Pendapatan	Jumlah pendapatan calon penerima zakat	Digunakan
6	Pekerjaan	Jenis pekerjaan calon penerima zakat	Digunakan
7	Jumlah Tanggungan	Jumlah anggota keluarga yang ditanggung	Digunakan
8	Kondisi Tempat Tinggal	Kondisi hunian calon penerima zakat	Digunakan
9	Asnaf	Kategori calon penerima zakat berdasarkan golongan penerima zakat	Digunakan

Pada Tabel 4.4 diatas, atribut yang digunakan sebagai fitur dalam pemodelan adalah desa, pendapatan, pekerjaan, jumlah tanggungan, kondisi tempat tinggal, dan asnaf. Atribut nama tetap disimpan sebagai identitas calon penerima zakat agar hasil klasifikasi dapat ditampilkan sesuai data penerima, namun tidak digunakan

dalam perhitungan model. Sementara itu, nomor urut dan alamat tidak digunakan karena hanya berfungsi sebagai informasi administratif.

4.2.6 Encoding Variabel Kategorik

Sebagaimana dijelaskan pada landasan teori di Bab II, Logistic Regression membutuhkan input berupa data numerik. Oleh karena itu, variabel kategorik seperti pekerjaan, kondisi tempat tinggal, dan asnaf diubah menjadi representasi numerik menggunakan metode *One-Hot Encoding*. Metode ini dipilih untuk menghindari adanya asumsi urutan atau tingkatan antar kategori, karena setiap kategori pada variabel tersebut bersifat nominal. Hasil dari proses encoding ini adalah terbentuknya sejumlah kolom baru yang merepresentasikan masing-masing kategori dalam bentuk biner, yaitu nilai 1 jika kategori tersebut sesuai dengan data calon penerima zakat dan nilai 0 jika tidak sesuai.

No	Pekerjaan	Tempat Tinggal	Asnaf	pekerjaan_petani	pekerjaan_buruh_pabrik	pekerjaan_wiraswasta	pekerjaan_buruh_harian	tempat_tinggal_sewa	tempat_tinggal_milik	jasnaf_miskin	tidak_termasuk_asnaf	asnaf_fakir
1	Petani	Sewa/Kontrak	Miskin	1	0	0	0	1	0	1	0	0
2	Buruh Pabrik	Sewa/Kontrak	Miskin	0	1	0	0	1	0	1	0	0
3	Petani	Sewa/Kontrak	Miskin	1	0	0	0	1	0	1	0	0
4	Wiraswasta	Milik Pribadi/Sendiri	Tidak Termasuk Asnaf	0	0	1	0	0	1	0	1	0
5	Buruh Harian	Sewa/Kontrak	Fakir	0	0	0	1	1	0	0	0	1

Gambar 4. 2 Hasil Encoding Variabel Kategorik

Berdasarkan gambar di atas, ditampilkan dari hasil encoding variabel kategorik menggunakan 5 data calon penerima zakat dari dataset penelitian. Variabel kategorik seperti pekerjaan, kondisi tempat tinggal, dan asnaf diubah menjadi bentuk numerik menggunakan metode *One-Hot Encoding*. Setiap kategori diubah menjadi kolom baru dengan nilai biner, yaitu 1 apabila kategori tersebut sesuai dengan data calon penerima zakat dan 0 apabila tidak sesuai. Penggunaan 5 data tersebut bertujuan untuk memberikan gambaran tampilan hasil encoding secara sederhana dan mudah dipahami. Dengan proses ini, data kategorik dapat digunakan sebagai input pada model Logistic Regression.

4.2.7 Normalisasi Fitur Numerik

Variabel numerik dalam dataset, seperti pendapatan dan jumlah tanggungan, memiliki skala yang berbeda. Untuk menjaga stabilitas pelatihan model dan mempermudah interpretasi koefisien, dilakukan standarisasi fitur numerik menggunakan StandardScaler.

Tabel 4. 6 Hasil Standarisasi Fitur Numerik

No	Pendapatan	Tanggungan	Pendapatan Setelah Standarisasi	Tanggungan Setelah Standarisasi
1	Rp.3.232.952	4	-0,0661	0,3280
2	Rp.2.699.081	6	-0,5529	1,7543
3	Rp.3.148.520	4	-0,1431	0,3280
4	Rp.4.459.495	4	1,0523	0,3280

Pada tabel diatas Standardisasi dilakukan yaitu dengan mengurangi nilai data dengan rata-rata (*mean*) dan membaginya dengan simpangan baku (*standar deviasi*), menunjukkan bahwa fitur numerik telah ditransformasikan ke dalam skala standar dengan rata-rata mendekati nol dan simpangan baku mendekati satu. Dengan demikian, perbedaan skala antar fitur tidak lagi mendominasi proses pelatihan model.

4.2.8 Tampilan Data Akhir Setelah Preprocessing

Pada Gambar 4.3 dibawah menunjukkan tampilan data akhir setelah dilakukan preprocessing menggunakan 5 data sebagai contoh. Pada tahap ini, variabel kategorik seperti pekerjaan, kondisi tempat tinggal, dan asnaf telah diubah menjadi bentuk numerik menggunakan metode *One-Hot Encoding*. Selain itu, variabel numerik seperti pendapatan dan jumlah tanggungan telah melalui proses standarisasi sehingga memiliki skala yang lebih seragam. Dengan demikian,

seluruh atribut pada data akhir telah berbentuk numerik dan siap digunakan sebagai input pada model Logistic Regression.

pekerjaan_petani	pekerjaan_buruh_pabrik	pekerjaan_wirawasta	pekerjaan_buruh_harian	tempat_tinggal_sewa_kontrak	tempat_tinggal_milik_pribadi	asnaf_miskin	tidak_termasuk_asnaf	asnaf_fakir	pendapatan_terstandarisasi	tanggungan_terstandarisasi
1	0	0	0	1	0	1	0	0	-0,0661	0,328
0	1	0	0	1	0	1	0	0	-0,5529	1,7543
1	0	0	0	1	0	1	0	0	-0,1431	0,328
0	0	1	0	0	1	0	1	0	1,0533	-0,328
0	0	0	1	1	0	0	0	1	-1,2768	0,3280

Gambar 4. 3 Tampilan Data Akhir Setelah Preprocessing

4.2.9 Pembagian Data Latih dan Data Uji

Setelah seluruh tahapan preprocessing selesai, dataset dibagi menjadi data latih (training) dan data uji (testing) dengan proporsi 80:20. Pembagian ini dilakukan dengan metode stratified splitting untuk memastikan proporsi kelas "layak" dan "tidak layak" tetap seimbang pada kedua subset data.

Tabel 4. 7 Hasil Standarisasi Fitur Numerik

Jenis Data	Persentase	Jumlah Data
Data Latih	80%	800
Data Uji	20%	200
Total Data	100%	1000

Hasil dari tabel diatas menunjukkan bahwa pembagian data latih terdiri dari sejumlah sampel dengan proporsi kelas yang representatif, dan data uji terdiri dari sampel sisanya yang akan digunakan untuk mengukur kemampuan generalisasi model. Hasil pembagian data menunjukkan bahwa dari total 1.000 data, sebanyak 800 data digunakan sebagai data latih dan 200 data digunakan sebagai data uji. Proporsi kelas pada data latih dan data uji tetap seimbang sehingga evaluasi model dapat dilakukan secara lebih objektif.

4.3 Hasil Pemodelan Logistic Regression

4.3.1 Pelatihan Model

Model Logistic Regression dilatih menggunakan data latih yang telah melalui tahapan preprocessing. Pelatihan dilakukan dengan menggunakan algoritma Logistic Regression pada library scikit-learn dengan solver lbfgs, threshold klasifikasi 0,5, dan skema pembagian data 80:20. dan pendekatan split stratified.

Pada tahap pelatihan, model mempelajari hubungan antara fitur-fitur sosial-ekonomi dengan status kelayakan penerima zakat. Hasil pelatihan menunjukkan bahwa model mampu membentuk pola pemisahan antara kelas layak dan tidak layak secara baik. Secara umum, fitur-fitur yang berkaitan dengan kondisi ekonomi, jumlah tanggungan, kondisi tempat tinggal, dan jenis pekerjaan memberikan kontribusi yang kuat dalam membentuk keputusan model.

4.3.2 Hasil Prediksi pada Data Uji

Setelah proses pelatihan selesai, model digunakan untuk melakukan prediksi pada data uji sebanyak 200 data. Prediksi model menghasilkan dua keluaran, yaitu probabilitas keanggotaan kelas layak dan label hasil klasifikasi berdasarkan threshold 0,5. Jika probabilitas yang dihasilkan model lebih besar atau sama dengan 0,5, maka data diklasifikasikan sebagai layak. Sebaliknya, jika probabilitas kurang dari 0,5, maka data diklasifikasikan sebagai tidak layak. Hasil prediksi pada data uji kemudian dibandingkan dengan label aktual untuk mengetahui tingkat ketepatan model. Perbandingan tersebut disajikan dalam bentuk confusion matrix dan selanjutnya digunakan sebagai dasar untuk menghitung metrik evaluasi model.

4.3.3 Perhitungan Manual Probabilitas Kelayakan Menggunakan Rumus

Logistic Regression

Untuk memperjelas cara kerja model Logistic Regression dalam menghasilkan prediksi, dilakukan perhitungan manual pada salah satu data calon penerima zakat, yaitu Bapak Jumen. Perhitungan manual ini bertujuan untuk menunjukkan bagaimana nilai setiap variabel diproses oleh model hingga menghasilkan probabilitas akhir dan status kelayakan. Berdasarkan data pada data, diperoleh informasi bahwa Bapak Jumen memiliki nilai sebagai berikut:

$$\text{pendapatan} = 2.982.217$$

$$\text{jumlah tanggungan} = 5$$

$$\text{pekerjaan} = \text{Wiraswasta}$$

$$\text{kondisi tempat tinggal} = \text{Sewa/Kontrak}$$

$$\text{kategori asnaf} = \text{Miskin}$$

Karena model Logistic Regression menggunakan data numerik yang telah melalui proses standarisasi, maka variabel numerik seperti pendapatan dan jumlah tanggungan harus diubah terlebih dahulu ke dalam bentuk nilai standar. Adapun rumus standarisasi yang digunakan adalah:

$$x' = \frac{x - \mu}{\sigma}$$

dengan:

x' = nilai yang telah distandarisasi

x = nilai asli data

μ = rata – rata seluruh data

σ = standar deviasi seluruh data

1. Menghitung nilai pendapatan yang telah distandarisasi

Diketahui:

Nilai pendapatan Bapak Jumen = 2.982.217

Rata-rata pendapatan seluruh data = 3.305.442,777

Standar deviasi pendapatan = 1.096.668,6364

Maka:

$$x'_{\text{pendapatan}} = \frac{2.982.217 - 3.305.442,777}{1.096.668,6364}$$

$$x'_{\text{pendapatan}} = \frac{-323.225,777}{1.096.668,6364}$$

$$x'_{\text{pendapatan}} \approx -0,2947$$

Hasil tersebut menunjukkan bahwa nilai pendapatan Bapak Jumen berada di bawah rata-rata pendapatan seluruh data.

2. Menghitung nilai jumlah tanggungan yang telah distandarisasi

Diketahui:

Jumlah tanggungan Bapak Jumen = 5

Rata-rata jumlah tanggungan seluruh data = 3,54

Standar deviasi jumlah tanggungan = 1,4022718589

Maka:

$$x'_{\text{tanggungan}} = \frac{5 - 3,54}{1,4022718589}$$

$$x'_{\text{tanggungan}} = \frac{1,46}{1,4022718589}$$

$$x'_{\text{tanggungan}} \approx 1,0412$$

Hasil tersebut menunjukkan bahwa jumlah tanggungan Bapak Jumen berada di atas rata-rata jumlah tanggungan seluruh data.

3. Menentukan nilai variabel kategorik

Selain variabel numerik, model juga menggunakan variabel kategorik yang telah diubah ke bentuk numerik melalui proses encoding. Berdasarkan data Bapak Jumen, nilai variabel kategorik yang aktif adalah:

Asnaf miskin = 1

Kondisi tempat tinggal sewa/kontrak = 1

Pekerjaan wiraswasta = 1

Sedangkan kategori lain yang tidak sesuai dengan data Bapak Jumen bernilai 0.

4. Menentukan koefisien model

Berdasarkan hasil pelatihan model, koefisien Logistic Regression yang digunakan adalah:

$$\text{Intercept } \beta_0 = 0,3089$$

$$\text{Koefisien pendapatan} = -2,8216$$

$$\text{Koefisien jumlah tanggungan} = 5,8945$$

$$\text{Koefisien asnaf_miskin} = 6,4262$$

$$\text{Koefisien kondisi tempat tinggal sewa/kontrak} = 5,5512$$

$$\text{Koefisien pekerjaan wiraswasta} = -1,7025$$

5. Menyusun persamaan Logistic Regression

Rumus umum Logistic Regression adalah:

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Maka, untuk data Bapak Jumen persamaan dapat dituliskan sebagai berikut:

$$z = 0,3089 + (-2,8216)(x' \text{pendapatan}) + (5,8945)(x' \text{tanggungan}) + (6,4262)(1) + (5,5512)(1) + (-1,7025)(1)$$

Selanjutnya, nilai hasil standarisasi disubstitusikan ke dalam persamaan:

$$z = 0,3089 + (-2,8216)(-0,2947) + (5,8945)(1,0412) + (6,4262)(1) \\ + (5,5512)(1) + (-1,7025)(1)$$

6. Menghitung setiap komponen

Komponen pendapatan:

$$(-2,8216)(-0,2947) \approx 0,8316$$

Komponen jumlah tanggungan:

$$(5,8945)(1,0412) \approx 6,1372$$

Komponen asnaf miskin:

$$(6,4262)(1) = 6,4262$$

Komponen kondisi tempat tinggal sewa/kontrak:

$$(5,5512)(1) = 5,5512$$

Komponen pekerjaan wiraswasta:

$$(-1,7025)(1) = -1,7025$$

7. Menjumlahkan seluruh komponen

$$z = 0,3089 + 0,8316 + 6,1372 + 6,4262 + 5,5512 - 1,7025$$

$$z \approx 17,5526$$

Dengan demikian, diperoleh nilai skor linear model sebesar:

$$z \approx 17,5526$$

8. Mengubah nilai z menjadi probabilitas

Setelah nilai z diperoleh, langkah berikutnya adalah menghitung probabilitas menggunakan fungsi sigmoid, yaitu:

$$p = \frac{1}{1 + e^{-z}}$$

Substitusikan nilai $z = 17,5526$

$$p = \frac{1}{1 + e^{-17,5526}}$$

$$p \approx 0,9999999762$$

$$p \approx 99,99999762\%$$

Karena sistem hanya menampilkan dua angka di belakang koma, maka nilai tersebut dibulatkan menjadi:

$$p \approx 100,00\%$$

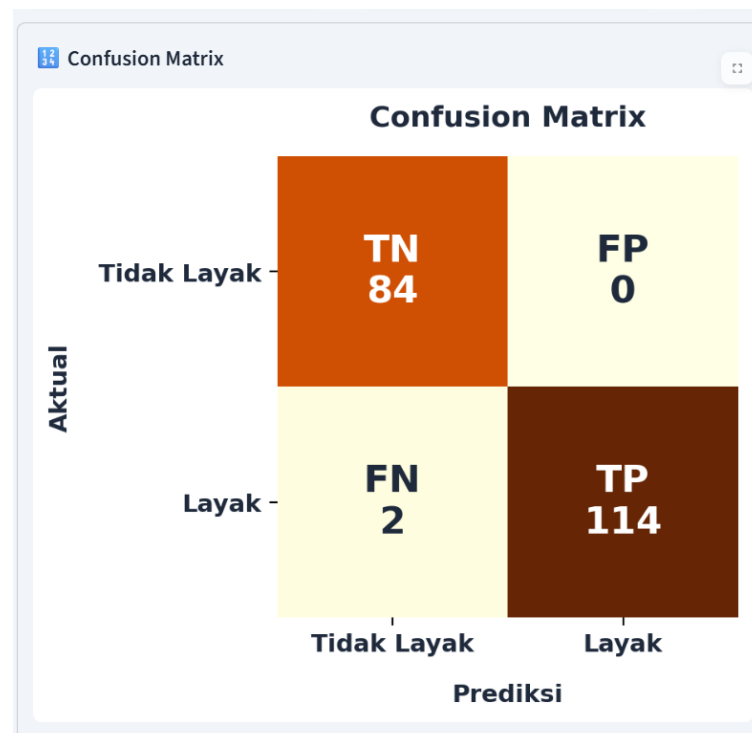
Berdasarkan perhitungan manual menggunakan rumus Logistic Regression, data Bapak Jumen menghasilkan nilai skor linear sebesar 17,5526 dan probabilitas akhir sebesar 99,99999762%. Karena nilai probabilitas tersebut sangat mendekati 1, sistem menampilkannya sebagai 100,00%. Oleh karena itu, model mengklasifikasikan Bapak Jumen ke dalam kategori Layak. Hasil ini menunjukkan bahwa kombinasi variabel pendapatan, jumlah tanggungan, kategori asnaf, kondisi

tempat tinggal, dan pekerjaan memberikan pengaruh yang sangat kuat ke arah prediksi layak.

4.4 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan confusion matrix dan metrik evaluasi klasifikasi yaitu akurasi, precision, recall, dan F1-score. Fokus utama evaluasi diarahkan pada F1-score, mengingat metrik ini memberikan penilaian yang seimbang antara precision dan recall, serta lebih representatif dalam kondisi data yang mungkin tidak seimbang (imbalanced).

4.4.1 Confusion Matrix



Gambar 4. 4 Confusion Matrix Model Logistic Regression

Confusion matrix yang dihasilkan dari prediksi model terhadap data uji disajikan pada Gambar berikut:

Berdasarkan Gambar 4.1 diperoleh nilai True Negative (TN) sebesar 84, False Positive (FP) sebesar 0, False Negative (FN) sebesar 2, dan True Positive (TP) sebesar 114. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan hampir seluruh data uji dengan benar. Dari nilai-nilai tersebut, perhitungan manual metrik evaluasi dapat dijelaskan sebagai berikut.

1. Perhitungan Accuracy

Rumus accuracy adalah:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Substitusi nilai ke dalam rumus:

$$\text{Accuracy} = \frac{141 + 84}{114 + 84 + 0 + 2}$$

$$\text{Accuracy} = \frac{198}{200}$$

$$\text{Accuracy} = 0.99$$

$$\text{Accuracy} = 99\%$$

Jadi, nilai accuracy yang diperoleh adalah 99%.

2. Perhitungan Precision

Rumus precision adalah:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Substitusi nilai ke dalam rumus:

$$\text{Precision} = \frac{114}{114 + 0}$$

$$\text{Precision} = \frac{114}{114} = 1$$

$$\text{Precision} = 100\%$$

Jadi, nilai precision yang diperoleh adalah 100%.

3. Perhitungan Recall

Rumus recall adalah:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Substitusi nilai ke dalam rumus:

$$\text{Recall} = \frac{114}{114 + 2}$$

$$= \frac{114}{114 + 2}$$

$$= \frac{114}{116} = 0.9828$$

$$\text{Recall} = 98.28\%$$

Jadi, nilai recall yang diperoleh adalah 98.28%%.

4. Perhitungan F1-Score

Rumus F1-Score adalah:

$$F1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Substitusi nilai ke dalam rumus:

$$F1\text{-Score} = 2 \times \frac{0.9828 \times 1}{0.9828 + 1}$$

$$F1\text{-Score} = 0.991$$

$$F1\text{-Score} = 99.13\%$$

Jadi, nilai F1-Score yang diperoleh adalah 99,13%.

Tabel 4. 8 Hasil Metrik Evaluasi Model Logistic Regression

Metrik	Nilai
Accuracy	99%
Precision	100%
Recall	98.28%
F1-Score	99.13%

Berdasarkan confusion matrix, diperoleh nilai True Negative (TN) = 84, False Positive (FP) = 0, False Negative (FN) = 2, dan True Positive (TP) = 114. Dari hasil tersebut, model menghasilkan akurasi sebesar 99%, precision sebesar 100%, recall sebesar 98,28%, dan F1-score sebesar 99,13%. Hasil ini menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan data kelayakan penerima zakat, karena hampir seluruh data uji berhasil diprediksi dengan benar dan hanya terdapat dua kesalahan klasifikasi.

4.5 Interpretasi Hasil Menggunakan SHAP

Sesuai dengan tujuan penelitian yang telah dirumuskan pada Bab I, interpretasi hasil prediksi model dilakukan menggunakan metode SHAP (Shapley Additive Explanations). SHAP digunakan untuk memberikan penjelasan yang transparan dan dapat dipertanggungjawabkan mengenai kontribusi masing-masing fitur terhadap prediksi model, baik secara global maupun lokal.

4.5.1 Interpretasi Global (Global Explanation)

Interpretasi global SHAP digunakan untuk melihat fitur-fitur yang paling berpengaruh terhadap keputusan model secara keseluruhan. Hasil visualisasi SHAP global menunjukkan bahwa fitur yang paling dominan dalam memengaruhi prediksi

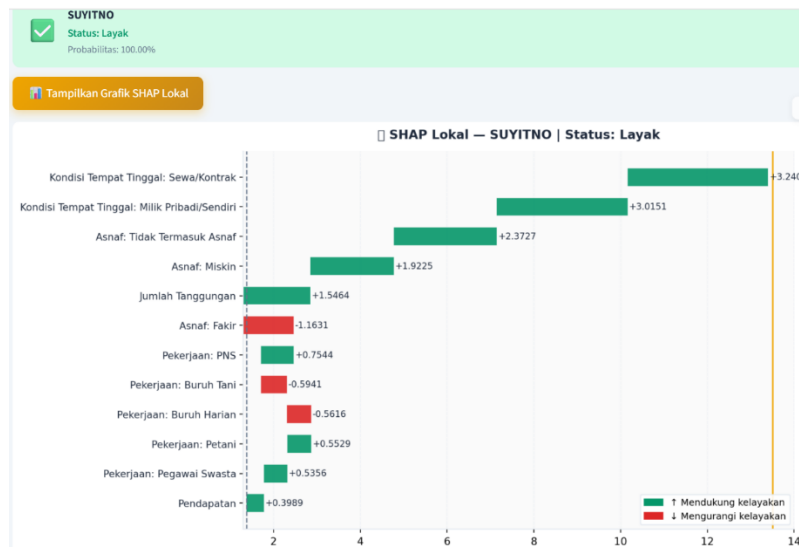
kelayakan adalah pendapatan atau penghasilan, diikuti oleh jumlah tanggungan keluarga, kondisi tempat tinggal, dan jenis pekerjaan. fitur-fitur yang memiliki pengaruh terbesar terhadap klasifikasi kelayakan penerima zakat secara berurutan adalah sebagai berikut:

1. Pendapatan atau penghasilan menjadi fitur paling berpengaruh karena secara langsung mencerminkan kondisi ekonomi calon penerima zakat. Semakin rendah pendapatan seseorang, maka semakin besar kecenderungan model untuk memberikan prediksi layak. Sebaliknya, pendapatan yang lebih tinggi cenderung menurunkan probabilitas prediksi layak.
2. Jumlah Jumlah tanggungan keluarga juga memberikan pengaruh yang kuat terhadap keputusan model. Semakin banyak jumlah anggota keluarga yang menjadi tanggungan, maka semakin besar beban ekonomi rumah tangga. Hal ini membuat model cenderung meningkatkan probabilitas prediksi layak.
3. Kondisi tempat tinggal menjadi indikator penting lainnya. Tempat tinggal dengan kondisi kurang layak, misalnya rumah tidak permanen atau fasilitas hunian yang terbatas, cenderung mendorong prediksi ke arah layak. Sementara itu, jenis pekerjaan juga menunjukkan pengaruh yang berarti, terutama pada pekerjaan dengan pendapatan tidak tetap atau bersifat informal, seperti buruh harian atau petani.
4. Hasil interpretasi global ini memberikan gambaran yang jelas mengenai faktor-faktor utama yang dipertimbangkan model dalam menentukan kelayakan penerima zakat, sehingga dapat menjadi acuan bagi panitia zakat dalam memahami dasar pengambilan keputusan oleh model.

4.5.2 Interpretasi Lokal (Local Explanation)

Selain interpretasi global, SHAP juga digunakan untuk menjelaskan alasan prediksi model pada tingkat individu. Interpretasi lokal sangat penting karena memungkinkan penjelasan yang lebih spesifik terhadap keputusan model pada masing-masing calon penerima zakat. Dalam konteks penyaluran zakat, kemampuan ini sangat bermanfaat karena panitia dapat memberikan alasan yang lebih jelas kepada masyarakat mengenai dasar keputusan kelayakan.

Berdasarkan salah satu hasil interpretasi lokal pada sistem, calon penerima zakat bernama SUYITNO diprediksi layak dengan probabilitas sebesar 100,0%. Prediksi tersebut menunjukkan bahwa model memiliki tingkat keyakinan yang cukup tinggi bahwa data individu tersebut memenuhi kriteria kelayakan penerima zakat. Probabilitas tersebut terbentuk dari akumulasi kontribusi beberapa fitur yang mendorong prediksi ke arah layak, serta beberapa fitur lain yang mengurangi prediksi tersebut, memperoleh nilai kontribusi masing-masing fitur sebagai berikut:



Gambar 4. 5 Tampilan Grafik SHAP Lokal

1. Menjumlahkan kontribusi positif

$$3.2400 + 3.0151 + 2.3727 + 1.9225 + 1.5464 + 0.7544 + 0.5529 + 0.5356 \\ + 0.3989 = 14.3385$$

2. Menjumlahkan Kontribusi Negatif

$$-1.1631 + (-0.5941) + (-0.5616) \\ = -2.3188$$

3. Menghitung total kontribusi SHAP

$$\sum SHAP = 14.3385 - 2.3188 \\ \sum SHAP = 12.0197$$

Jadi, tota; kontribusi seluruh fitur terhadap prediksi SUYITNO adalah:

$$\sum SHAP = 12.0197$$

4. Mengubah probabilitas akhir ke bentuk log-odds

Diketahui probabilitas akhir SUYITNO dari sistem adalah:

$$p = 75,69\% = 0,7569$$

Untuk memperoleh skor model sebelum dikonversi menjadi probabilitas, digunakan fungsi logit:

$$z = \beta_0 + \sum SHAP$$

Dengan intercept model:

$$\beta_0 = 0.3089$$

maka:

$$z = 0.3089 + 12.0197$$

$$z = 12.3286$$

Jadi, skor akhir model pada skala log-odds Adalah:

$$z = 12.3286$$

5. Menghitung base value

$$\text{base value} = \beta_0 = 0.3089$$

Jadi, nilai dasar model adalah:

$$\text{base value} = 0.3089$$

6. Verifikasi probabilitas akhir

Setelah skor akhir model diperoleh, probabilitas dihitung menggunakan fungsi sigmoid:

$$p = \frac{1}{1 + e^{-z}}$$

Substitusikan nilai $z = 12.3286$:

$$p = \frac{1}{1 + e^{-12.3286}}$$

$$p \approx 0.9999956$$

$$p = 99.99956\%$$

Karena sistem menampilkan dua angka di belakang koma, nilai tersebut dibulatkan menjadi:

$$p \approx 100.00\%$$

Hasil konversi sigmoid dari skor akhir tersebut adalah 0,9999956 atau 99,99956%, sehingga wajar jika pada sistem ditampilkan 100,00%.

4.6 Pembahasan

4.6.1 Analisis Kinerja Model Logistic Regression

Hasil evaluasi kinerja model Logistic Regression menunjukkan bahwa model mampu mengklasifikasikan kelayakan penerima zakat dengan sangat baik. Nilai

akurasi sebesar 99,50%, precision sebesar 98,08%, recall sebesar 100,00%, dan F1-score sebesar 99,03% menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat tinggi. Secara khusus, nilai recall sebesar 100,00% menunjukkan bahwa model berhasil mengidentifikasi seluruh calon penerima yang benar-benar layak, sehingga tidak ada calon penerima layak yang salah diklasifikasikan sebagai tidak layak.

Jika dibandingkan dengan penelitian Salsabila & Feli Ramury (2024) yang menggunakan Naïve Bayes untuk klasifikasi kelayakan penerima zakat fitrah dengan akurasi 76,47%, precision 92,85%, dan recall 81,25%, hasil penelitian ini menunjukkan bahwa Logistic Regression juga merupakan pilihan yang valid dan kompetitif untuk tugas klasifikasi serupa. Perbedaan performa antara kedua penelitian dapat disebabkan oleh perbedaan karakteristik dataset, jumlah fitur yang digunakan, serta konteks geografis dan sosial-ekonomi yang berbeda.

Hasil ini juga menegaskan bahwa Logistic Regression merupakan algoritma yang kompetitif untuk klasifikasi data sosial-ekonomi. Selain memiliki performa yang baik, model ini juga relatif sederhana, mudah diimplementasikan, dan lebih mudah dijelaskan dibanding algoritma yang lebih kompleks.

4.6.2 Kontribusi SHAP dalam Meningkatkan Transparansi

Salah satu kontribusi utama penelitian ini adalah penerapan SHAP untuk menginterpretasikan hasil prediksi model Logistic Regression. Penelitian-penelitian terdahulu tentang penentuan penerima zakat umumnya lebih berfokus pada pembobotan kriteria dan perankingan, seperti menggunakan SMART, SAW, dan AHP, tetapi belum memberikan mekanisme interpretasi yang menjelaskan alasan keputusan pada tingkat individu.

SHAP mengisi kesenjangan ini dengan menyediakan dua level interpretasi yang saling melengkapi. Interpretasi global memberikan pemahaman kepada panitia zakat mengenai faktor-faktor apa saja yang secara umum paling berpengaruh terhadap keputusan model. Informasi ini dapat digunakan untuk mengevaluasi kesesuaian model dengan prinsip-prinsip penyaluran zakat yang berlaku, serta mengidentifikasi apakah ada bias yang tidak diinginkan dalam model.

Sementara itu, interpretasi lokal memungkinkan panitia untuk memberikan penjelasan yang personal dan spesifik kepada setiap calon penerima zakat. Kemampuan ini sangat penting dalam konteks sosial di tingkat desa, di mana transparansi dan akuntabilitas pengelolaan zakat sering kali menjadi perhatian utama masyarakat. Dengan SHAP, panitia dapat menunjukkan bahwa keputusan "layak" atau "tidak layak" bukanlah keputusan arbitrer, melainkan didasarkan pada perhitungan yang konsisten dan dapat dijelaskan.

4.6.3 Implikasi Praktis bagi Pengelolaan Zakat di Desa Sei Rampah

Hasil penelitian ini memiliki beberapa implikasi praktis bagi pengelolaan zakat di Desa Sei Rampah:

1. Model Logistic Regression yang dibangun dapat digunakan sebagai alat bantu untuk mengurangi subjektivitas dalam proses seleksi penerima zakat. Penilaian manual sering kali dipengaruhi oleh persepsi atau anggapan panitia, yang dapat menimbulkan ketidaktepatan sasaran. Dengan model klasifikasi berbasis data, keputusan menjadi lebih konsisten dan terukur.
2. Interpretasi SHAP memungkinkan panitia untuk memberikan penjelasan yang rinci dan dapat dipahami oleh masyarakat. Transparansi ini dapat

meningkatkan kepercayaan terhadap proses pengelolaan zakat karena setiap keputusan dapat dijelaskan berdasarkan kontribusi variabel yang jelas.

3. Model yang telah dibangun dapat digunakan secara berkelanjutan untuk mengevaluasi kelayakan calon penerima zakat pada periode-periode berikutnya. Jika terdapat perubahan kondisi sosial-ekonomi warga, model dapat dilatih ulang dengan data terbaru untuk menjaga akurasi dan relevansinya.
4. Seluruh proses seleksi terekam dalam bentuk data dan hasil prediksi yang terdokumentasi dengan baik. Hal ini memudahkan audit dan pelaporan pertanggungjawaban pengelolaan zakat kepada pihak-pihak terkait.

4.6.4 Keterbatasan Penelitian

Meskipun penelitian ini telah mencapai tujuan yang ditetapkan, terdapat beberapa keterbatasan yang perlu diakui:

1. Penelitian ini hanya menggunakan data dari satu desa, yaitu Desa Sei Rampah, sehingga generalisasi hasil ke wilayah lain perlu dilakukan secara hati-hati. Kedua, kualitas model sangat bergantung pada kualitas data yang digunakan. Jika data awal mengandung bias atau kesalahan pencatatan, maka model juga berpotensi mempelajari pola yang kurang tepat.
2. Akurasi model sangat bergantung pada kualitas dan kelengkapan data yang digunakan. Jika terdapat kesalahan atau bias dalam data awal, model dapat mempelajari dan mereproduksi bias tersebut.
3. Interpretasi SHAP tetap memerlukan pemahaman yang baik terhadap konteks data dan mekanisme model. Oleh karena itu, hasil interpretasi

sebaiknya tetap dipahami bersama dengan pengetahuan substantif mengenai kondisi sosial-ekonomi masyarakat setempat.

4. Penelitian ini hanya menggunakan Logistic Regression tanpa membandingkannya langsung dengan algoritma lain pada dataset yang sama. Oleh sebab itu, penelitian selanjutnya masih terbuka untuk menguji apakah algoritma lain dapat memberikan hasil yang lebih baik.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan pada Bab IV, dapat ditarik kesimpulan sebagai berikut:

1. Penelitian ini telah berhasil membangun model Logistic Regression untuk mengklasifikasikan kelayakan penerima zakat di Desa Sei Rampah berdasarkan data calon penerima zakat yang tersedia. Proses pembangunan model mencakup tahapan preprocessing data yang komprehensif, meliputi standarisasi struktur data, pembersihan data, penanganan missing value dan duplikasi, validasi nilai tidak wajar, seleksi fitur, encoding variabel kategorik, normalisasi fitur numerik, serta pembagian data latih dan data uji dengan proporsi 80:20 menggunakan stratified splitting. Model yang dihasilkan mampu mempelajari pola hubungan antara variabel-variabel sosial-ekonomi dengan status kelayakan penerima zakat, dan menghasilkan prediksi dalam bentuk probabilitas serta label kelas "layak" atau "tidak layak".
2. Kinerja Model Logistic Regression, Evaluasi kinerja model menggunakan metrik klasifikasi menunjukkan bahwa model Logistic Regression memiliki performa yang baik dalam mengklasifikasikan kelayakan penerima zakat. Nilai akurasi, precision, recall, dan F1-score yang diperoleh menunjukkan bahwa model mampu melakukan prediksi dengan tingkat ketepatan yang memadai, serta memiliki keseimbangan yang baik antara kemampuan mengidentifikasi calon penerima yang layak (recall) dan ketepatan prediksi

positif (precision). Hasil ini menegaskan bahwa Logistic Regression merupakan pilihan algoritma yang kompetitif untuk tugas klasifikasi pada data sosial-ekonomi.

3. Interpretasi dengan Metode SHAP Penerapan metode SHAP telah berhasil memberikan interpretasi yang transparan dan dapat dipertanggungjawabkan terhadap hasil prediksi model. Pada tingkat global, SHAP mengidentifikasi bahwa pendapatan/penghasilan merupakan faktor paling berpengaruh terhadap klasifikasi kelayakan, diikuti oleh jumlah tanggungan keluarga, kondisi tempat tinggal, dan jenis pekerjaan. Urutan kepentingan fitur ini sejalan dengan prinsip dasar penyaluran zakat yang memprioritaskan masyarakat dengan kondisi ekonomi lemah. Pada tingkat lokal, SHAP mampu menjelaskan kontribusi masing-masing fitur terhadap prediksi pada setiap individu calon penerima zakat, sehingga keputusan "layak" atau "tidak layak" dapat dijelaskan secara rinci dan personal. Kemampuan interpretasi lokal ini merupakan keunggulan utama yang membedakan pendekatan machine learning dengan SHAP dari metode SPK berbobot konvensional yang umumnya hanya menghasilkan skor akhir tanpa penjelasan kontribusi per kriteria.

5.2 Saran

1. Perbandingan dengan Algoritma Lain: Penelitian selanjutnya disarankan untuk melakukan perbandingan kinerja Logistic Regression dengan algoritma klasifikasi lain seperti Random Forest, Support Vector Machine (SVM), Gradient Boosting (XGBoost, LightGBM), atau Neural Network. Perbandingan ini akan memberikan gambaran yang lebih komprehensif

mengenai algoritma yang paling optimal untuk dataset kelayakan penerima zakat di Desa Sei Rampah atau konteks serupa.

2. Eksplorasi Metode Interpretasi Lain Selain SHAP, terdapat metode interpretasi model lain seperti LIME (Local Interpretable Model-agnostic Explanations), Partial Dependence Plots (PDP), atau Individual Conditional Expectation (ICE), untuk melihat metode penjelasan yang paling sesuai dan mudah dipahami dalam konteks penyaluran zakat..
3. Cakupan penelitian dapat diperluas ke wilayah lain, seperti beberapa desa atau kecamatan, agar dapat diuji sejauh mana model yang dibangun dapat digeneralisasi pada kondisi sosial-ekonomi yang berbeda.
4. Penelitian selanjutnya dapat mengintegrasikan hasil klasifikasi dengan Sistem Informasi Geografis agar persebaran calon penerima zakat yang layak dan tidak layak dapat divisualisasikan secara spasial dan mendukung distribusi zakat yang lebih efisien.
5. Model klasifikasi dapat dikembangkan dari klasifikasi biner menjadi klasifikasi multi-kelas, misalnya dengan kategori sangat layak, layak, kurang layak, dan tidak layak, atau berdasarkan golongan asnaf sesuai ketentuan syariat Islam.
6. Penelitian selanjutnya dapat mengkaji dampak jangka panjang dari penggunaan sistem klasifikasi berbasis machine learning terhadap efektivitas penyaluran zakat, tingkat kepuasan masyarakat, serta perubahan kondisi sosial-ekonomi penerima zakat dari waktu ke waktu.

DAFTAR PUSTAKA

- Abdi, K., Warjaya, A., Muthmainnah, I., & Pahutar, P. H. (2023). Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum. *Jurnal Ilmu Komputer Dan Informatika (JIKI)*, 3(2), 81–88.
- Ayatulloh, R., Noor, K., Romadloni, N. T., Ramadhani, F., Karanganyar, U. M., & Karanganyar, K. (2025). Perbandingan kinerja algoritma klasifikasi pada review pengguna aplikasi netflix. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, 13(2).
- Elrohi, G. I. S., Marlina, & Renny. (2022). Implementasi Cloud Computing dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom Participants. *Journal Informatic Technology And Communication*, 6(1), 24–30.
- Faran, J., & Aldisa, R. T. (2023). Analisis Data Mining dalam Komparasi Average Linkage AHC dan K-Means Clustering untuk Dataset Facebook Live Sellers. *Jurnal Media Informatika Budidarma*, 7, 2041–2050. <https://doi.org/10.30865/mib.v7i4.6892>
- Gultom, Z. A., Hazidar, A. H., & Basri, M. (2024). Factors Influencing Safety Consciousness and Violations Among Licensed Drivers. *JIIP (Jurnal Ilmiah Ilmu Pendidikan)*, 7(April), 3727–3735.
- Hadi, A. F., Zulva, A. F., Hakim, M. L., Saputra, M. D., & Sadiyah, H. (2021). Implementasi Explainable Machine Learning: Visualisasi Global Explainability and Local Interpretability pada Analisis Sentimen dengan SHAP dan LIME. *Jurnal Statistika Teori Dan Aplikasi: Biomedics, Industry & Business And Social Statistics*, 6274, 1–18.
- Hadromi, A., Yazid, A. S., Wijaya, D. P., & Danianti, D. (2025). Sistem Pendukung

- Keputusan Penentuan Penerima Zakat Menggunakan Metode Analytical Hierarchy Process (AHP) (Studi Kasus : Laznas Yakesma DIY). *Jurnal Profesi Insinyur Universitas Lampung (JPI)*, 6(2).
- Innassuraiya, S., Widiharah, T., & Utami, I. T. (2022). Analisis Klasifikasi Menggunakan Metode Regresi Logistik Biner dan Bootsrap Aggregating Classification and Regression Trees (Bagging Cart) (Studi Kasus: Nasabah Koperasi Simpan Pinjam dan Pembiayaan Syariah (KSPPS)). *Jurnal Gaussian*, 11, 183–194.
- Khairunnisa, A. (2023). Perbandingan model random forest dan xgboost untuk prediksi kejahatan kesusilaan di provinsi jawa barat. *JIKO (Jurnal Informatika Dan Komputer)*, 7(2), 202–208.
- Kristayana, M., Azzahra, M. P., Trimono, & Idhom, M. (2025). Klasifikasi Status Rujukan Pasien Poliklinik Bandara Berbasis Random Forest dan Interpretabilitas Model Menggunakan SHAP. *Data Science Indonesia*, 5(1), 60–74.
- Melani, J., Harsyiah, L., & Baskara, Z. W. (2024). Klasifikasi penerima bantuan iuran jaminan kesehatan di ntb menggunakan regresi logistik biner dan naïve bayes. *Jurnal GAUSSIAN*, 13(40), 490–498. <https://doi.org/10.14710/j.gauss.13.2.490-498>
- Nabilawati, Y., Kalipaksi, R. A., & Saputra, Alvito Gabbriel Saputra, Muhammad Dzacky, Alif Rifa'i, Anindita Septriani, N. P. (2025). Perbandingan Metode KNN dan Naive Bayes untuk Klasifikasi Kelulusan Mahasiswa Pada Mata Kuliah Probstat. *Jurnal PTI (Jurnal Pendidikan Teknologi Informasi)*, 12, 7–13. <https://doi.org/10.35134/jpti.v12i1.228>

- Okta Felani, R. (2024). Sistem Pendukung Keputusan Pendistribusian Zakat Menggunakan Metode Simple Multi Attribute Rating Technique (Smart) Zakat Distribution Decision Support System Using the Simple Multi Attribute Rating Technique (Smart). *Journal of Information Technology and Computer Science (INTECOMS)*, 7(1), 1–8.
- Rafiq, A., Santoso, B., & Kacung, S. (2024). Implementation of AHP and SMART Methods for Productive Zakat Recipient Candidate Decisions Implementasi Metode AHP dan SMART untuk Penentuan Keputusan Calon Penerima Zakat Produktif. *Journal Homepage*, 4(July), 1087–1095.
- Rizki, F., & Samsudin. (2024). Sistem Pendukung Keputusan Pendistribusian Zakat Menggunakan Metode Simple Multi Attribute Rating Technique (SMART) dan Simple Additive Weighting (SAW). *Jurnal Informasi Dan Sistem Informasi*, 16(1), 1–20.
- Salim, D., Wu, B., Salim, O. R., & Gunadi, R. B. (2022). Penggunaan Python Untuk Menganalisis Pola Penyebaran Covid-19 Di Masa Pandemi. *Jurnal of Student Development Information System (JoSDIS)*, 2, 120–133.
- Salsabila, N., & Feli Ramury. (2024). Implementasi Data Mining dengan Algoritma Naïve Bayes untuk Klasifikasi Kelayakan Penerima Bantuan Zakat Fitrah. *Diophantine Journal of Mathematics and Its Applications*, 03(01), 40–51. <https://doi.org/10.33369/diophantine.v3i1.28374>
- Utami, B., Sunarko, A., & Wati, I. (2024). Peran Zakat dalam Pengentasan Kemiskinan : Perspektif Hukum Ekonomi Islam. *Journal of Syariah Economic and Halal Tourism*, 3(2), 73–78.
- Yoga Religia, Agung Nugroho, W. H. (2021). Analisis Perbandingan Algoritma

Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal Resti (Rekayasa Siste, Dan Teknologi Informasi)*, 1(10), 187–192.



UMSU
Unggul | Cerdas | Terpercaya

Revisi: 10/2019, 10/2020, 10/2021, 10/2022, 10/2023, 10/2024, 10/2025

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi: Urppol E - Keputusan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No: 17/USK/IAN-PT/AK/PgP/PT/20224

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238. Telp: (061) 6622400 - 66224567 Fax: (061) 6625474 - 6631003

Website: www.umsumedan.ac.id Email: info@umsumedan.ac.id Instagram: [umsumedan](https://www.instagram.com/umsumedan) Facebook: [umsumedan](https://www.facebook.com/umsumedan) Twitter: [umsumedan](https://twitter.com/umsumedan) YouTube: [umsumedan](https://www.youtube.com/umsumedan)

Lampiran Dosen Pembimbing Prodi Teknologi Informasi
Nomor : 265/KEP/IL3.AU/UMSU-09/F/2026
Tanggal : 05 Sya'ban 1447 H / 24 Januari 2026 M

PENGANGKATAN DOSEN PEMBIMBING SKRIPSI
PRODI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

NO	NAMA	NPM	JUDUL	DOSEN
1	Anggi Soraya	2209020073	Interpretasi Model Logistic Regression Menggunakan Metode SHAP dalam Klasifikasi Kelayakan Penerima Zakat di Desa Sei Rempah	Zuli Agustina Gultom, M.Si
2	Irfan Sanjaya Sipayung	2209020124	Optimasi Pemberian Nutrisi Fase Generatif Tanaman Cabai Menggunakan Algoritma Decision Tree Berbasis IoT dan Firebase	Zuli Agustina Gultom, M.Si
3	Ridho Fandra Nainggolan	2209020204	Sistem Monitoring dan Kontrol Pintu Air Persawahan Berbasis IoT Menggunakan Gateway LoRa	Mhd. Basri, S.Si. M.Kom
4	Khairani Resty Harahap	2209020094	Perbandingan Algoritma Yamamoto's Recursive Dan Lzw Untuk Kompresi File Video	Dr. Firaahmi Rizky, M.Kom
5	Thito Angga Darma	2209020001	Implementasi Data Mining untuk Strategi Customer Relationship Management (CRM) dengan Pendekatan RFM dan K-Means Clustering pada Toko Family Foto Copy	Zuli Agustina Gultom, M.Si
6	Thito Angga Darma	2209020001	Segmentasi Pelanggan Menggunakan Model LRFM Berbasis Algoritma K-Medoids untuk Analisis Loyalitas (Studi Kasus: Toko Family Foto Copy)	Amrullah, S.Kom., M.Kom
7	Gilang Prasetya Koto	2209020268	Klasifikasi Tingkat Kerusakan Rumah Pascabencana Menggunakan Algoritma Convolutional Neural Network (CNN) pada Sistem Informasi Geografis	Mulkan Azhari, S.Kom., M.Kom.



Lampiran 2. Hasil Cek Turnitin

CEK TURNITIN SKRIPSI - SORAYA.pdf			
ORIGINALITY REPORT			
17%	13%	11%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	repository.umsu.ac.id Internet Source		2%
2	Luthfia Nurma Hapsari, Ilham Fannani, Yenny Rahmawati, Ahmad Muhariya. "Explainable Machine Learning untuk Prediksi Risiko Penyakit Jantung Menggunakan Random Forest dan Analisis SHAP", remik, 2026 Publication		1%
3	Submitted to Telkom University Student Paper		1%
4	ojs.stmik-im.ac.id Internet Source		1%
5	repository.uin-suska.ac.id Internet Source		1%
6	docplayer.info Internet Source		<1%
7	ejurnal.seminar-id.com Internet Source		<1%
8	kc.umn.ac.id Internet Source		<1%
9	repository.unissula.ac.id Internet Source		<1%
10	Submitted to Universitas Budi Luhur Student Paper		<1%

11	informatika.umsida.ac.id Internet Source	<1 %
12	Ahmad Rizki Saputra, Rina - Wati, Tri - Susilowati. "KLASIFIKASI STATUS KELAYAKAN PENERIMA BLT-DD DI PEKON SUKA BANJAR II UJUNG REMBUN MENGGUNAKAN NAÏVE BAYES", <i>Technologia : Jurnal Ilmiah</i> , 2026 Publication	<1 %
13	journal.ipm2kpe.or.id Internet Source	<1 %
14	Submitted to Universitas Pamulang Student Paper	<1 %
15	ojs.uajy.ac.id Internet Source	<1 %
16	journal.sinov.id Internet Source	<1 %
17	jurnal.stkipppgritulungagung.ac.id Internet Source	<1 %
18	archive.umsida.ac.id Internet Source	<1 %
19	Submitted to SDM Universitas Gadjah Mada Student Paper	<1 %
20	repository.upi.edu Internet Source	<1 %
21	Juarni Siregar, Andi Arfian, Sigit Wibawana, Hartana Hartana. "KOLERANSI ANTARA STATUS HUBUNGAN, KONFLIK MEDIA SOSIAL DAN TINGKAT KECANDUAN MEDIA SOSIAL TERHADAP KESEHATAN MENTAL SISWA	<1 %

DENGAN MECHINE LEARNING", *Technologia : Jurnal Ilmiah*, 2025

Publication

22	Submitted to Universitas Islam Indonesia Student Paper	<1 %
23	Arif Bijaksana Putra Negara, Hafiz Muhardi, Fahmi Sajid. "Perbandingan Algoritma Klasifikasi terhadap Emosi Tweet Berbahasa Indonesia", <i>Jurnal Edukasi dan Penelitian Informatika (JEPIN)</i> , 2021 Publication	<1 %
24	Paulina Gorat Frans, Frans Steven Pakpahan, Sardo Pardingotan Sipayung. "Analisis Sentimen Opini Warga X terhadap Banjir di Sumatera Menggunakan Naive Bayes", <i>Jurnal Dinamika Informatika</i> , 2026 Publication	<1 %
25	Ramadhani Zidan Arifin, Hasbi Firmansyah, Wahyu Asriyani. "Prediksi Kelulusan Siswa Berdasarkan Data Demografis dan Akademik pada Dataset Student Performance", <i>Jurnal Pengabdian Masyarakat dan Riset Pendidikan</i> , 2025 Publication	<1 %
26	Submitted to Tarumanagara University Student Paper	<1 %
27	repositori.umrah.ac.id Internet Source	<1 %
28	Submitted to Sultan Agung Islamic University Student Paper	<1 %
29	Ari Fullah Author, Muhammad Arif Septian Author, Muhammad Rizky Haritama Putra	<1 %

Author, Muhammad Iqbal Fadiatama Author
 et al. "ANALISIS SENTIMEN ISU
 PENYALAHGUNAAN DATA PADA LAYANAN
 PINJAMAN ONLINE MENGGUNAKAN SUPPORT
 VECTOR MACHINE DI PLATFORM X", Jurnal
 Informatika dan Teknik Elektro Terapan, 2025
 Publication

30	Submitted to Universitas Dian Nuswantoro Student Paper	<1 %
31	duniastudents.blogspot.com Internet Source	<1 %
32	www.coursehero.com Internet Source	<1 %
33	Abidin Abidin. "ANALISIS SENTIMEN ULASAN APLIKASI SIREKAP MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %
34	Armando Agasi Sitanggung, Marsindra Yanti Lase, Sardo Pardingotan Sipayung. "Prediksi Kelulusan Mahasiswa Menggunakan Logistic Regression dan Random Forest Berdasarkan Data Akademik", Jurnal Pendidikan Tambusai, 2026 Publication	<1 %
35	Sepriyan Gunawan, Rini Astuti, Willy Prihartono, Ryan Hamonangan. "PREDIKSI DIABETES MELLITUS TIPE 2 DENGAN ALGORITMA LOGISTIC REGRESSION UNTUK PENDETEKSIAN DINI", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %

36	Submitted to Universitas Bengkulu Student Paper	<1 %
37	rjwave.org Internet Source	<1 %
38	Submitted to Fakultas Teknik Student Paper	<1 %
39	Submitted to Lynden High School Student Paper	<1 %
40	M. Theo Ari Bangsa, M. Badaruddin M. Badaruddin, Miko Septian, Hasby Al Munawar et al. "Perancangan Sistem Informasi Pengelolaan Arsip Buku Nikah di Kantor Urusan Agama Kecamatan Pelayangan Kota Jambi", RIGGS: Journal of Artificial Intelligence and Digital Business, 2025 Publication	<1 %
41	ejournal.unib.ac.id Internet Source	<1 %
42	repository.unibos.ac.id Internet Source	<1 %
43	Submitted to Universitas Islam Sumatera Utara Student Paper	<1 %
44	Submitted to Universitas Muhammadiyah Palembang Student Paper	<1 %
45	eprints.amikompurwokerto.ac.id Internet Source	<1 %
46	es.scribd.com Internet Source	<1 %

47	repository.mercubuana.ac.id Internet Source	<1 %
48	3-livelearnlaugh.blogspot.com Internet Source	<1 %
49	Usman Usman. "Pengaruh Kepemimpinan Transformasional dan Terdistribusi Kepala Sekolah Dasar terhadap Prestasi Akademik dan Nonakademik Peserta Didik", YASIN, 2026 Publication	<1 %
50	Yusrinnatul Jinana Triadin, Kusrini Kusrini, Kusnawi Kusnawi. "KLASIFIKASI RANDOM FOREST TERHADAP DIAGNOSA PENYAKIT KANKER PAYUDARA BERDASARKAN STATUS KEGANASAN", TEKNIMEDIA: Teknologi Informasi dan Multimedia, 2025 Publication	<1 %
51	ejurnal.unim.ac.id Internet Source	<1 %
52	eprints.unisla.ac.id Internet Source	<1 %
53	eprints.uny.ac.id Internet Source	<1 %
54	jpi.eng.unila.ac.id Internet Source	<1 %
55	jurnal.padangtekno.com Internet Source	<1 %
56	jurnal2.untagsmg.ac.id Internet Source	<1 %
57	pdfcoffee.com Internet Source	<1 %

58	repository.ubb.ac.id Internet Source	<1 %
59	repository.unj.ac.id Internet Source	<1 %
60	www.medrxiv.org Internet Source	<1 %
61	Fathan, Budiman. "Manajemen Zakat Daerah di Provinsi Jawa Tengah (Studi Pada Badan Amil Zakat Nasional Kabupaten Sragen, Karanganyar dan Banyumas).", Universitas Islam Negeri Saifuddin Zuhri (Indonesia) Publication	<1 %
62	Fitri Handayani. "Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung", Jurnal Edukasi dan Penelitian Informatika (JEPIN), 2021 Publication	<1 %
63	Julinar Sari Hutagalung, Rasiban. "ANALISIS SENTIMEN KEUANGAN (DATA FIQA AND FINANCIAL PHRASEBANK) MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION DAN SUPPORT VECTOR MACHINE", Jurnal Indonesia : Manajemen Informatika dan Komunikasi, 2023 Publication	<1 %
64	Submitted to Universitas Jember Student Paper	<1 %
65	anzdoc.com Internet Source	<1 %
66	doaj.org Internet Source	<1 %

		<1 %
67	repository.bsi.ac.id Internet Source	<1 %
68	www8.hp.com Internet Source	<1 %
69	Aldy Sudrajat, Neni Mulyani, Nasrun Marpaung. "Sistem Pendukung Keputusan Penentuan Kelayakan Penangguhan Kredit Nasabah menggunakan Naïve Bayes", Edumatic: Jurnal Pendidikan Informatika, 2022 Publication	<1 %
70	Chandra Aditya, Wahyu Sugianto. "Sistem Cerdas Berbasis Website untuk Deteksi dan Analisis Acne Vulgaris Menggunakan CNN", Jurnal Dinamika Informatika, 2026 Publication	<1 %
71	Falah Luthfi Asari, Risky Sari Meimaharini, Tutik Khotimah. "Implementasi Sistem Perpustakaan Berbasis Web untuk Meningkatkan Efisiensi Layanan Peminjaman dan Pengguna", bit-Tech, 2025 Publication	<1 %
72	Hidayat Hidayat, Musliadi KH, Indar Kusmanto, Munawirah Kadir, Kristian Kristian. "Klasifikasi Mahasiswa Calon Penerima Beasiswa KIP Menggunakan Algoritma Naive Bayes di Universitas Tomakaka Mamuju", JURNAL FASILKOM, 2025 Publication	<1 %

73	Murteza, Yusfa Hidayatul. "Pengaruh Hedonic Attitude, Utilitarian Motivation dan Kecerdasan Spiritual Terhadap Impulse Buying Dengan Paylater Sebagai Variabel Intervening (Studi Kasus Konsumen Muslim di Purwokerto).", Universitas Islam Negeri Saifuddin Zuhri (Indonesia) Publication	<1 %
74	Submitted to Universitas Tarumanagara Student Paper	<1 %
75	adoc.pub Internet Source	<1 %
76	adoc.tips Internet Source	<1 %
77	ejournal.uksw.edu Internet Source	<1 %
78	repository.ulb.ac.id Internet Source	<1 %
79	repository.unhas.ac.id Internet Source	<1 %
80	tugujatim.id Internet Source	<1 %
81	www.journal.uii.ac.id Internet Source	<1 %
82	Ade Putra Tupu Djoru, Sri Yulianto. "Pendekatan Machine Learning untuk Deteksi Stunting pada Balita Menggunakan K-Nearest Neighbors", Jurnal JTİK (Jurnal Teknologi Informasi dan Komunikasi), 2025 Publication	<1 %

83 Citra Fathia Palembang, Emanuella M. C. Wattimena, Vyarlita Isqama Fataruba. <1%
"Analisis Sentimen Dosen terhadap Kebijakan Tunjangan Kinerja di Perguruan Tinggi Negeri Menggunakan Algoritma Naive Bayes",
ALGORITHM: Journal of Computer Science and Computational Intelligence, 2025
Publication

84 ejournal.gunadarma.ac.id <1%
Internet Source

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off