



**MODEL *STACKING ENSEMBLE* UNTUK PREDIKSI
PENUNGGAKAN PEMBAYARAN LISTRIK DI KOTABARU**

SKRIPSI

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

**Oleh
SHOFIA AZKIA
NIM. 2111017220008**

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



**MODEL *STACKING ENSEMBLE* UNTUK PREDIKSI
PENUNGGAKAN PEMBAYARAN LISTRIK DI KOTABARU**

SKRIPSI

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

**Oleh
SHOFIA AZKIA
NIM. 2111017220008**

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

SKRIPSI

MODEL STACKING ENSEMBLE UNTUK PREDIKSI PENUNGGAKAN PEMBAYARAN LISTRIK DI KOTABARU

Oleh
Shofia Azkia
NIM. 2111017220008

Telah dipertahankan pada hari Rabu, tanggal 15 Januari 2025 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



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

Penguji I



Yeni Rahkmawati, S.Mat., M.Si.
NIP. 199404032022032014

Pembimbing II



Dewi Sri Susanti, S.Si., M.Si.
NIP. 197305161999032002

Penguji II



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

Banjarbaru, 17 Januari 2025
Mengetahui,
Koordinator Program Studi Statistika
FMIPA ULM



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

PERNYATAAN

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

Banjarbaru, 12 Januari 2025



Shofia Azkia
NIM. 2111017220008

PRODI STATISTIKA

ABSTRAK

Model Stacking Ensemble untuk Prediksi Penunggakan Pembayaran Listrik di Kotabaru (Oleh: Shofia Azkia; Pembimbing: Selvi Annisa dan Dewi Sri Susanti, 2025; 65 halaman)

Listrik merupakan kebutuhan dasar manusia yang sangat penting dalam kehidupan modern. Namun di balik pentingnya listrik, terdapat permasalahan yang seringkali muncul, yaitu tunggakan pembayaran listrik. Tingginya angka tunggakan tidak hanya berdampak negatif pada pendapatan perusahaan, tetapi juga berdampak luas bagi perekonomian negara, masyarakat, dan keberlangsungan layanan listrik. Perusahaan sangat perlu melakukan identifikasi lebih dini tentang potensi pelanggan yang menunggak. Hal ini dapat dilakukan dengan mengidentifikasi kelompok pelanggan yang menunggak menggunakan metode klasifikasi. Penelitian ini bertujuan untuk mengimplementasikan pemodelan klasifikasi *Stacking Ensemble* serta membandingkannya dengan *base model Decision Tree*, *Random Forest*, dan *Extra Tree* yang dilihat dari nilai *G-Mean*. *Stacking Ensemble* menggunakan ketiga *base model* tersebut dengan Regresi Logistik sebagai *meta model*. Proses pembentukan model *Stacking Ensemble* dimulai dengan melatih masing-masing *base model* secara terpisah menggunakan *data training* hingga menghasilkan prediksi. Hasil prediksi dari *base model* tersebut kemudian digabungkan untuk membentuk *dataset* baru yang digunakan sebagai input pada *meta model* Regresi Logistik. Dari pemodelan *Stacking Ensemble*, diperoleh nilai *G-Mean* pada *data testing* sebesar 0.5791. Dibandingkan model *Decision Tree*, *Random Forest*, dan *Extra Tree*, nilai *G-Mean* pada model *Stacking Ensemble* memiliki nilai yang paling tinggi sehingga model *Stacking Ensemble* menghasilkan performa yang lebih baik dibandingkan model-model pembangunnya.

Kata Kunci: *Stacking Ensemble*, Pemodelan, Klasifikasi, Tunggakan Pembayaran Listrik

ABSTRACT

Ensemble Stacking Model for Prediction of Electricity Payment Arrears in Kotabaru (By: Shofia Azkia; Supervisors: Selvi Annisa dan Dewi Sri Susanti, 2025; 65 page)

Electricity is a basic human need that is very important in modern life. But behind the importance of electricity, there are problems that often arise, namely arrears in electricity payments. The high number of arrears not only has a negative impact on the company's revenue, but also has a broad impact on the country's economy, society, and the sustainability of electricity services. Companies urgently need to make early identification of potential customers who are in arrears. This can be done by identifying groups of delinquent customers using classification methods. This research aims to implement Stacking Ensemble classification modeling and compare it with the base models of Decision Tree, Random Forest, and Extra Tree as seen from the G-Mean value. Stacking Ensemble uses the three base models with Logistic Regression as a meta model. The Stacking Ensemble model building process starts by training each base model separately using training data until it produces predictions. The prediction results from the base models are then combined to form a new dataset that is used as input to the Logistic Regression meta model. From the Stacking Ensemble modeling, the G-Mean value on the testing data is 0.5791. Compared to the Decision Tree, Random Forest, and Extra Tree models, the G-Mean value of the Stacking Ensemble model has the highest value so that the Stacking Ensemble model produces better performance than its building models.

Keywords: Stacking Ensemble, Modelling, Classification, Electricity Payment Arrears

PRAKATA

Puji syukur kehadiran Allah *Subhanahu Wa Ta'ala*, yang telah memberikan rahmat serta karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yaitu skripsi yang berjudul "*Model Stacking Ensemble* untuk Prediksi Penunggakan Pembayaran Listrik di Kotabaru" dengan lancar. Proses penyusunan skripsi ini untuk memenuhi salah satu syarat dalam rangka menyelesaikan program sarjana di Program Studi Statistika Fakultas MIPA Universitas Lambung Mangkurat. Penulis ingin mengucapkan terima kasih sebesar-besarnya kepada pihak-pihak yang selalu mendukung dan memberikan perhatian serta bantuan selama proses penulisan Tugas Akhir ini, di antaranya:

1. Allah *Subhanahu Wa Ta'ala*, yang selalu memberikan saya kekuatan, kelancaran, dan kemudahan dalam proses pengerjaan Tugas Akhir.
2. Orang tua yang telah memberikan doa, dukungan, dan motivasi kepada penulis selama berkuliah sampai penyelesaian skripsi ini.
3. Ibu Selvi Annisa, S.Si., M.Si. dan Ibu Dewi Sri Susanti, S.Si., M.Si. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan dukungan dalam penyelesaian skripsi ini.
4. Ibu Yeni Rahkmawati, S.Mat., M.Si. dan Bapak Rifqi Aulya Rahman, S.Mat., M.Si. selaku dosen penguji yang telah memberikan masukan, saran, dan bantuan dalam perbaikan penulisan skripsi.
5. Bapak dan Ibu dosen pengajar serta staf Program Studi Statistika FMIPA ULM yang telah memberikan ilmu, motivasi, nasihat, kelancaran, dan dukungan selama masa perkuliahan.
6. Teman-teman seperjuangan yang selalu menemani, mendukung, dan memberikan semangat selama penulisan Tugas Akhir maupun selama masa perkuliahan, khususnya teman-teman "JST" dan "Potik Geng". Terima kasih telah menjadi partner yang sangat berharga dalam perjalanan perkuliahan ini.
7. Teman-teman Program Studi Statistika angkatan 2021 yang telah berjuang bersama dalam proses ini.
8. Seluruh pihak yang telah berkontribusi, baik secara langsung maupun tidak langsung.

Penulisan Tugas Akhir ini tentunya masih banyak kekurangan dalam penulisan maupun hasilnya, sehingga kritik dan saran membangun dari semua pihak sangat diharapkan agar bermanfaat bagi semua pihak. Akhir kata penulis berharap agar semua orang yang telah membantu dalam proses ini diberikan kebaikan serta hasil penelitian ini dapat berguna bagi pembacanya.

Banjarbaru, 12 Januari 2025

Shofia Azkia

DAFTAR ISI

PERNYATAAN	iii
ABSTRAK.....	iv
ABSTRACT	v
PRAKATA.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	viii
DAFTAR TABEL.....	ix
DAFTAR LAMPIRAN.....	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Penelitian Terdahulu	5
2.2 Kajian Teori.....	6
2.2.1 Ketidakseimbangan Data dan Proses Penanganannya	6
2.2.2 Variabel <i>Dummy</i>	8
2.2.3 <i>Stacking Ensemble</i>	8
2.2.4 <i>Decision Tree</i>	10
2.2.5 <i>Random Forest</i>	12
2.2.6 <i>Extra Tree Classifier</i>	13
2.2.7 Regresi Logistik.....	15
2.2.8 <i>Confusion Matrix</i>	15
BAB III METODE PENELITIAN	18
3.1 Sumber Data	18
3.2 Variabel Penelitian	18
3.3 Prosedur Penelitian	20
BAB IV HASIL DAN PEMBAHASAN	25
4.1 <i>Preprocessing Data</i>	25
4.2 Analisis Deskriptif.....	27
4.3 <i>Base Model</i>	32
4.3.1 Model <i>Decision Tree</i>	33
4.3.2 Model <i>Random Forest</i>	36
4.3.3 Model <i>Extra Tree</i>	39
4.4 Model <i>Stacking Ensemble</i>	41
BAB V PENUTUP	44
5.1 Kesimpulan	44
5.2 Saran.....	44
DAFTAR PUSTAKA.....	46
LAMPIRAN.....	49
RIWAYAT HIDUP	65

DAFTAR GAMBAR

Gambar 1.1 Persentase Pelanggan Menunggak per Subunit	1
Gambar 2.1 Cara Kerja SMOTE	7
Gambar 2.2 Skema Model <i>Stacking</i>	10
Gambar 2.3 Struktur <i>Decision Tree</i>	11
Gambar 2.4 Struktur <i>Random Forest</i>	13
Gambar 2.5 Tahap Pelatihan <i>Extra Tree</i>	14
Gambar 3.1 Prosedur Penelitian.....	24
Gambar 4.1 Proporsi Pelanggan Menunggak Pembayaran Listrik.....	27
Gambar 4.2 Proporsi Pelanggan Menunggak Berdasarkan Jenis Tarif	28
Gambar 4.3 Distribusi Frekuensi Pelanggan Menunggak Berdasarkan Jenis Tarif.....	28
Gambar 4.4 Proporsi Pelanggan Menunggak Berdasarkan Sub-Subunit.....	29
Gambar 4.5 Proporsi Pelanggan Menunggak Berdasarkan Bulan	29
Gambar 4.6 Distribusi Frekuensi Pelanggan Menunggak Berdasarkan Daya Listrik.....	30
Gambar 4.7 Distribusi Pembayaran untuk Pelanggan Menunggak Berdasarkan Daya Listrik 450 VA, 900 VA, 1300 VA, dan 2200 VA	31
Gambar 4.8 Struktur <i>Decision Tree</i> yang Terbentuk	33

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	5
Tabel 2.2 <i>Confusion Matrix</i>	16
Tabel 3.1 Variabel Penelitian	18
Tabel 4.1 Sampel Data Duplikat	25
Tabel 4.2 Sampel Pembuatan Kolom Penunggakan Transaksi dari kolom TglTransaksi	26
Tabel 4.3 Sampel Pembuatan Kolom Bulan dari Kolom TglTransaksi	26
Tabel 4.4 Sampel Penggabungan Kategori Jenis Tarif	26
Tabel 4.5 Jumlah Baris Data pada <i>Data Training</i> dan <i>Data Testing</i>	32
Tabel 4.6 Komposisi <i>Data Training</i> Sebelum dan Setelah SMOTE	32
Tabel 4.7 Hasil <i>Confusion Matrix</i> dari <i>Decision Tree</i> pada <i>Data Training</i>	34
Tabel 4.8 Hasil <i>Confusion Matrix</i> dari <i>Decision Tree</i> pada <i>Data Testing</i>	35
Tabel 4.9 Perbandingan Evaluasi Model <i>Decision Tree</i> pada <i>Data Training</i> dan <i>Data Testing</i>	36
Tabel 4.10 Hasil <i>Confusion Matrix</i> dari <i>Random Forest</i> pada <i>Data Training</i>	37
Tabel 4.11 Hasil <i>Confusion Matrix</i> dari <i>Random Forest</i> pada <i>Data Testing</i>	38
Tabel 4.12 Perbandingan Evaluasi Model <i>Random Forest</i> pada <i>Data Training</i> dan <i>Data Testing</i>	38
Tabel 4.13 Hasil <i>Confusion Matrix</i> dari <i>Extra Tree</i> pada <i>Data Training</i>	39
Tabel 4.14 Hasil <i>Confusion Matrix</i> dari <i>Extra Tree</i> pada <i>Data Testing</i>	40
Tabel 4.15 Perbandingan Evaluasi Model <i>Extra Tree</i> pada <i>Data Training</i> dan <i>Data Testing</i>	41
Tabel 4.16 Sampel <i>Dataset</i> Baru untuk Melatih <i>Meta model</i>	41
Tabel 4.17 Sampel <i>Dataset</i> Baru untuk Mengevaluasi Model <i>Stacking</i> <i>Ensemble</i>	42
Tabel 4.18 Hasil <i>Confusion Matrix</i> dari <i>Stacking Ensemble</i>	42
Tabel 4.19 Perbandingan Nilai <i>G-Mean</i> dari Masing-Masing Model	43

DAFTAR LAMPIRAN

Lampiran 1. Pemilihan Variabel pada <i>Dataset</i>	49
Lampiran 2. <i>Dataset</i> yang sudah Bersih setelah <i>Preprocessing</i>	50
Lampiran 3. Nilai <i>G-Mean</i> dari <i>Tuning Parameter</i> Model <i>Decision Tree</i>	51
Lampiran 4. Nilai <i>G-Mean</i> dari <i>Tuning Parameter</i> Model <i>Random Forest</i>	52
Lampiran 5. Nilai <i>G-Mean</i> dari <i>Tuning Parameter</i> Model <i>Extra Tree</i>	53
Lampiran 6. <i>Syntax Python</i>	54

PRODI STATISTIKA