

**STUDI KOMPARATIF KINERJA ALGORITMA ENSEMBLE
LEARNING DALAM ESTIMASI GAJI PADA PEKERJAAN
SEKTOR KECERDASAN BUATAN**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
Alwan Rizqlulloh
21.11.3921

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

**STUDI KOMPARATIF KINERJA ALGORITMA ENSEMBLE
LEARNING DALAM ESTIMASI GAJI PADA PEKERJAAN
SEKTOR KECERDASAN BUATAN**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
Alwan Rizqululloh
21.11.3921

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**STUDI KOMPARATIF KINERJA ALGORITMA ENSEMBLE
LEARNING DALAM ESTIMASI GAJI PADA PEKERJAAN SEKTOR
KECERDASAN BUATAN**

yang disusun dan diajukan oleh

Alwan Rizqiulloh

21.11.3921

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 8 Agustus 2025

Dosen Pembimbing,



Ferian Fauzi Abdulloh, S.Kom., M.Kom

NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
STUDI KOMPARATIF KINERJA ALGORITMA ENSEMBLE
LEARNING DALAM ESTIMASI GAJI PADA PEKERJAAN SEKTOR
KECERDASAN BUATAN

yang disusun dan diajukan oleh

Alwan Rizqiulloh

21.11.3921

Telah dipertahankan di depan Dewan Penguji
pada tanggal 25 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

Ferian Fauzi Abdullah, S.Kom., M.Kom.
NIK. 190302276

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 25 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Alwan Rizqiulloh

NIM : 21.11.3921

Menyatakan bahwa Skripsi dengan judul berikut:

**Studi Komparatif Kinerja Algoritma Ensemble Learning dalam
Estimasi Gaji pada Pekerjaan Sektor Kecerdasan Buatan**

Dosen Pembimbing : Ferian Fauzi Abdulloh, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 25 Agustus 2025

Yang Menyatakan,



Alwan Rizqiulloh

HALAMAN PERSEMBAHAN

Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, karunia, dan penyertaan-Nya yang tiada batas, sehingga penulisan skripsi ini dapat diselesaikan dengan sebaik mungkin.

Dengan penuh rasa syukur, karya ini penulis persembahkan kepada Allah SWT, yang telah menganugerahkan nikmat serta kemampuan sehingga proses penyusunan skripsi ini dapat terselesaikan.

Ucapan terima kasih yang mendalam penulis sampaikan kepada keluarga tercinta, khususnya kepada Ayah, Ibu, dan Adik, atas doa, pengertian, dan dukungan tulus yang menjadi fondasi utama dalam menyelesaikan studi hingga meraih gelar sarjana.

Tak lupa, apresiasi penulis berikan kepada semua pihak yang telah membantu dalam penyusunan skripsi ini, yang tidak dapat disebutkan satu per satu. Semoga Allah SWT senantiasa memberikan perlindungan, kesehatan, dan ketenteraman kepada kita semua.

KATA PENGANTAR

Segala puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga skripsi yang berjudul **"STUDI KOMPARATIF KINERJA ALGORITMA ENSEMBLE LEARNING DALAM ESTIMASI GAJI PADA PEKERJAAN SEKTOR KECERDASAN BUATAN"** dapat diselesaikan dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Proses penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, atas segala petunjuk, kemudahan, dan kekuatan yang diberikan.
2. Kedua orang tua tercinta, atas doa, kasih sayang, dan pengorbanan yang tiada terhingga.
3. Bapak Ferian Fauzi Abdulloh, M.Kom., selaku Dosen Pembimbing Skripsi, atas waktu, arahan, bimbingan, dan masukan berharga selama proses penelitian dan penulisan.
4. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta, atas dukungan dan fasilitas yang diberikan.
5. Bapak Majid Rahardi, S.Kom., M.Eng., Ibu Ike Verawati, S.Kom., M.Kom., dan Bapak Ferian Fauzi Abdulloh, M.Kom., selaku Dewan Penguji Skripsi, atas waktu, kritik, dan saran membangun demi kesempurnaan karya ini.
6. Seluruh dosen dan staf Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta, atas ilmu dan fasilitas yang telah diberikan selama perkuliahan.

7. Rekan-rekan seperjuangan di Program Studi Informatika, atas kebersamaan, motivasi, dan dukungan yang tulus.
8. Semua pihak yang tidak dapat disebutkan satu per satu, atas bantuan dan dukungan yang diberikan selama penyusunan skripsi ini.

Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari sempurna. Oleh sebab itu, kritik dan saran yang bersifat membangun akan diterima dengan terbuka demi penyempurnaan di masa mendatang. Besar harapan penulis agar skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang estimasi gaji dan penerapan algoritma ensemble learning.

Yogyakarta, 25 Agustus 2025

Penulis

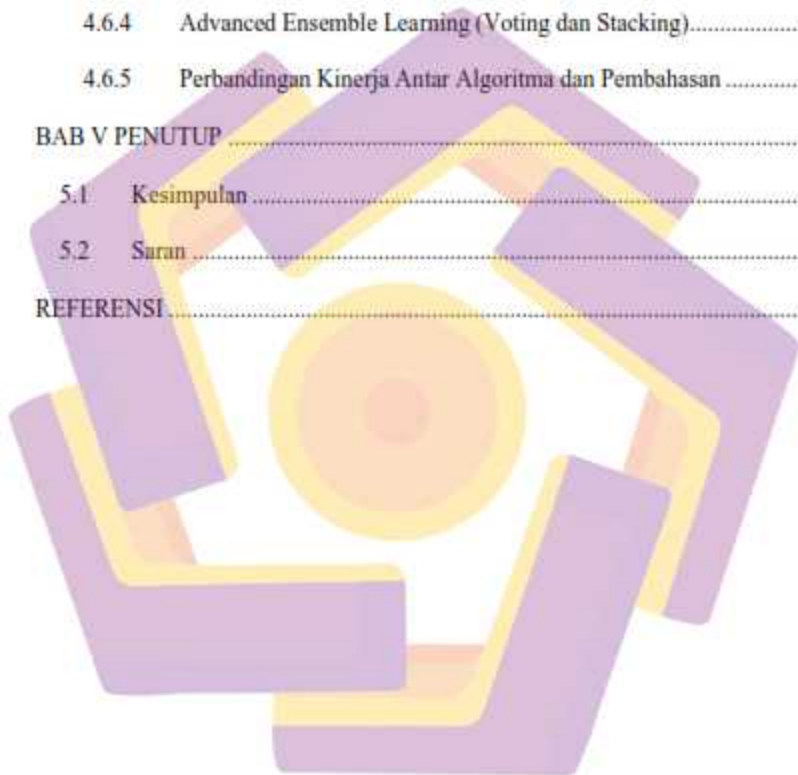


DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
ABSTRACT	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4

BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	15
2.2.1 Machine Learning dan Ensemble Learning	15
2.2.2 Analisis Regresi	16
2.2.3 Random Forest	16
2.2.4 Gradient Boosting	17
2.2.5 Optimalisasi Hyperparameter (GridSearchCV).....	19
2.2.6 Metrik Evaluasi Model	20
2.2.7 Advanced Ensembling	23
BAB III METODE PENELITIAN	25
3.1 Objek Penelitian.....	25
3.2 Alur Penelitian	25
3.3 Alat dan Bahan.....	27
3.3.1 Data Penelitian,	27
3.3.2 Alat/instrumen.....	29
BAB IV HASIL DAN PEMBAHASAN	32
4.1 Dataset.....	32
4.2 EDA	33
4.3 Filtering Data	46
4.4 Preprocessing Data	47
4.5 Regresi	50

4.6	Evaluasi Performa Model dan Pembahasan.....	55
4.6.1	Random Forest	55
4.6.2	XGBoost	57
4.6.3	CatBoost.....	59
4.6.4	Advanced Ensemble Learning (Voting dan Stacking).....	61
4.6.5	Perbandingan Kinerja Antar Algoritma dan Pembahasan	62
BAB V PENUTUP		65
5.1	Kesimpulan	65
5.2	Saran	65
REFERENSI		66



DAFTAR TABEL

Tabel 3.1 Deskripsi Variabel Penelitian	28
Tabel 4.1 Tipe Data.....	32
Tabel 4.2 Tabel Kolom Kombinasi.....	47
Tabel 4.3 Tabel Label Encoding.....	48
Tabel 4.4 Hasil Split Data.....	50
Tabel 4.5 Parameter Sebelum Tuning.....	50
Tabel 4.6 Hasil Akurasi Sebelum Di Tuning.....	51
Tabel 4.7 Parameter Sesudah Tuning	52
Tabel 4.8 Hasil Akurasi Sesudah Di Tuning.....	53
Tabel 4.9 Parameter Advanced Ensembling.....	54
Tabel 4.10 Hasil Akurasi Advanced Ensembling	54
Tabel 4.11 Perbandingan Hasil Random Forest	56
Tabel 4.12 Perbandingan Hasil XGBoost	58
Tabel 4.13 Perbandingan Hasil CatBoost	60
Tabel 4.14 Perbandingan Semua Hasil Akurasi Ensemble Learning	64

DAFTAR GAMBAR

Gambar 2.1 Diagram Kerja Random Forest	17
Gambar 2.2 Diagram Kerja XGBoost.....	18
Gambar 2.3 Diagram Kerja CatBoost.....	19
Gambar 2.4 Rumus MAE	20
Gambar 2.5 Rumus MSE	21
Gambar 2.6 Rumus RMSE	21
Gambar 2.7 Rumus R^2	22
Gambar 2.8 Diagram Kerja Voting Regressor.....	23
Gambar 2.9 Diagram Kerja Stacking Regressor.....	24
Gambar 3.1 Alur Penelitian	26
Gambar 4.1 Distribusi Jenis Pekerjaan	33
Gambar 4.2 Distribusi Gaji.....	34
Gambar 4.3 Distribusi Mata Uang Gaji	35
Gambar 4.4 Distribusi Tingkat Pengalaman.....	35
Gambar 4.5 Distribusi Jenis Pekerjaan	36
Gambar 4.6 Distribusi Lokasi Perusahaan.....	37
Gambar 4.7 Distribusi Ukuran Perusahaan.....	38
Gambar 4.8 Distribusi Tempat Tinggal Karyawan.....	39
Gambar 4.9 Distribusi Rasio Kerja	39
Gambar 4.9 Distribusi Keterampilan Yang dibutuhkan.....	40

Gambar 4.10 Distribusi Pendidikan Yang dibutuhkan	41
Gambar 4.11 Distribusi Pengalaman Bertahun-tahun.....	41
Gambar 4.12 Distribusi Sektor Industri	42
Gambar 4.13 Distribusi Tanggal Posting.....	43
Gambar 4.14 Distribusi Batas Waktu Lamaran	44
Gambar 4.15 Distribusi Kualitas Keuntungan.....	45
Gambar 4.16 Distribusi Log Gaji.....	46
Gambar 4.17 Distribusi Bulan Untuk Melamar	46
Gambar 4.18 Scatter Plot Random Forest.....	57
Gambar 4.19 Scatter Plot XGBoost	59
Gambar 4.20 Scatter Plot CatBoost	61
Gambar 4.21 Scatter Plot Advanced Ensemble Learning.....	62

DAFTAR LAMBANG DAN SINGKATAN



AI	Artificial Intelligence (Kecerdasan Buatan)
ANN	Artificial Neural Network
AUC	Area Under the Curve
EDA	Exploratory Data Analysis
GBDT	Gradient Boosted Decision Trees
GPR	Gaussian Process Regression
KDD	Knowledge Discovery in Databases
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
R^2	R-squared (Koefisien Determinasi)
RMSE	Root Mean Squared Error
SVM	Support Vector Machine
SVR	Support Vector Regression
TF-IDF	Term Frequency-Inverse Document Frequency

DAFTAR ISTILAH

Dataset:	Kumpulan data terstruktur yang digunakan untuk melatih dan menguji model machine learning
Ensemble Learning	Teknik machine learning yang menggabungkan beberapa model individual untuk menghasilkan satu prediksi akhir yang lebih kuat dan akurat.
Evaluasi Model	Proses mengukur kinerja model prediksi menggunakan metrik statistik untuk menilai akurasi dan keandalannya.
Feature Engineering	Proses membuat fitur-fitur baru dari data yang sudah ada untuk meningkatkan performa model.
GridSearchCV	Teknik optimisasi yang secara sistematis menguji semua kombinasi hyperparameter untuk menemukan konfigurasi model terbaik.
Hyperparameter	Parameter eksternal dari sebuah model yang nilainya diatur sebelum proses pembelajaran dimulai untuk mengontrol perilaku algoritma.
Label Encoding	Teknik pra-pemrosesan untuk mengubah data kategorikal (teks) menjadi representasi numerik.
Overfitting	Kondisi di mana model machine learning berkinerja sangat baik pada data latih tetapi buruk pada data baru (data uji), karena terlalu menangkap noise dari data latih.
Regresi	Tugas dalam supervised learning yang bertujuan untuk memprediksi nilai output yang bersifat kontinu atau numerik.
Tuning	Proses menyesuaikan atau mengoptimalkan hyperparameter sebuah model untuk meningkatkan kinerjanya.

INTISARI

Estimasi gaji di sektor kecerdasan buatan (AI) menjadi tantangan akibat kompleksitas faktor-faktor seperti pengalaman, lokasi, dan spesialisasi, yang sulit dimodelkan oleh metode statistik konvensional. Penelitian ini bertujuan untuk melakukan studi komparatif kinerja algoritma ensemble learning—Random Forest, XGBoost, dan CatBoost—serta teknik advanced ensembling (Voting dan Stacking) dalam mengestimasi gaji di sektor AI. Menggunakan dataset publik dari Kaggle dengan 15.000 sampel, model-model tersebut dilatih dan dioptimalkan melalui penyetelan hyperparameter. Kinerja model dievaluasi menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan R^2 Score. Hasil evaluasi menunjukkan bahwa proses tuning berhasil meningkatkan kinerja semua model, dengan CatBoost menjadi model tunggal terbaik (R^2 0.9233). Namun, performa puncak dicapai oleh Stacking Regressor, yang menggabungkan model-model yang telah dioptimalkan dan mencapai R^2 Score tertinggi sebesar 0.9235. Hasil penelitian ini memberikan model estimasi gaji yang sangat akurat, yang bermanfaat bagi para pencari kerja sebagai acuan negosiasi dan bagi perusahaan dalam merancang strategi kompensasi yang kompetitif.

Kata kunci: estimasi gaji, ensemble learning, Stacking Regressor, CatBoost, XGBoost.

ABSTRACT

Salary estimation in the artificial intelligence (AI) sector is challenging due to the complexity of factors such as experience, location, and specialization, which are difficult to model using conventional statistical methods. This study aims to conduct a comparative performance analysis of ensemble learning algorithms—Random Forest, XGBoost, and CatBoost—as well as advanced ensembling techniques (Voting and Stacking) for salary estimation in the AI sector. Using a public dataset from Kaggle with 15,000 samples, these models were trained and optimized through hyperparameter tuning. Model performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 Score metrics. The evaluation results demonstrate that the tuning process successfully improved the performance of all models, with CatBoost emerging as the best single model (R^2 0.9233). However, peak performance was achieved by the Stacking Regressor, which combined the optimized models to reach the highest R^2 Score of 0.9235. This research provides a highly accurate salary estimation model, which is beneficial for job seekers as a negotiation benchmark and for companies in designing competitive compensation strategies.

Keyword: salary estimation, ensemble learning, Stacking Regressor, CatBoost, XGBoost.