ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI UNTUK MEMPREDIKSI TINGKAT GAJI PADA PEKERJAAN DATA SCIENCE

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika

disusun oleh

MUHAMMAD SACHIB FARHAN NAUVALDHI

21.11.4004

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI UNTUK MEMPREDIKSI TINGKAT GAJI PADA PEKERJAAN DATA SCIENCE

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika

disusun oleh

MUHAMMAD SACHIB FARHAN NAUVALDHI

21.11.4004

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI UNTUK MEMPREDIKSI TINGKAT GAJI PADA PEKERJAAN DATA SCIENCE

yang disusun dan diajukan oleh

Muhammad Sachib Farhan Nauvaldhi

21.11.4004

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Januari 2025

Dosen Pembimbing,

Ali Mustopa, S.Kom., M.Kom
NIK. 190302192

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI UNTUK MEMPREDIKSI TINGKAT GAJI PADA PEKERJAAN DATA SCIENCE

yang disusun dan diajukan oleh

Muhammad Sachib Farhan Nauvaldhi

21.11.4004

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Januari 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Nila Feby Puspitasari, S.Kom, M.Cs **NIK. 190302161**	
Yudi Sutanto, S,Kom., M. Kom **NIK. 190302039**	
Ali Mustopa, S.Kom., M.Kom **NIK. 190302192**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Sachib Farhan Nauvaldhi
NIM : 21.11.4004

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Perbandingan Algoritma Klasifikasi Untuk Memprediksi Tingkat Gaji Pada Pekerjaan Data Science

Dosen Pembimbing : Ali Mustopa, S.Kom., 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, 22 Januari 2025

Yang Menyatakan,

Muhammad Sachib Farhan Nauvaldhi

HALAMAN PERSEMBAHAN

Puji syukur saya panjatkan kepada Allah SWT atas rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan judul ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI UNTUK MEMPREDIKSI TINGKAT GAJI PADA PEKERJAAN DATA SCIENCE. Shalawat serta salam saya sampaikan kepada Nabi Muhammad SAW yang telah memberikan teladan hidup penuh kebijaksanaan.

Karya ini saya persembahkan dengan penuh rasa hormat dan terima kasih kepada:

1. Umi Siti Sholihah, S.E., M.M., ibu saya, yang selalu memberikan kasih sayang, doa, dan motivasi tiada henti.
2. Abah Yahya Fatihin, ayah saya, yang selalu memberikan dukungan moral, semangat, dan nasihat yang menguatkan saya.
3. M. Farid Mahendra, S.Kom., kakak saya, yang senantiasa memberi semangat dan dorongan positif dalam menghadapi tantangan.
4. Bapak Ali Mustopa, S.Kom., M.Kom., dosen pembimbing saya, yang telah memberikan bimbingan dan arahan yang sangat berarti.
5. Bapak-Ibu Dosen, yang telah memberikan ilmu dan bimbingan yang sangat berharga dalam perjalanan akademik saya.
6. Refi Naftila N, sahabat sejati till jannah, yang selalu menemani dan mendukung saya dalam penyusunan skripsi ini.
7. Teman-teman seperjuangan, khususnya Martonsky (Sultan Faaiz, Adi Dwi, dan Galih Dwi) dan Santuf (M. Rosyid dan Gita N Amalia), yang selalu memberi semangat dan dukungan dalam perjalanan ini.

Saya juga menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak langsung, dalam penyelesaian skripsi ini. Semoga segala bantuan dan kebaikan yang diberikan memperoleh balasan yang setimpal dan bermanfaat bagi kita semua.

KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT atas segala rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan judul ANALISIS PERBANDINGAN ALGORITMA KLASIFIKASI UNTUK MEMPREDIKSI TINGKAT GAJI PADA PEKERJAAN DATA SCIENCE. Penyusunan skripsi ini tidak terlepas dari dukungan dan bantuan banyak pihak yang telah memberikan kontribusi yang sangat berarti.

Pada kesempatan ini, saya ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta, yang telah memberikan kesempatan dan dukungan dalam menyelesaikan pendidikan saya.
2. Ibu Windha Mega Pradnya D, M.Kom., selaku Ketua Program Studi S1 Informatika, yang telah memberikan bimbingan dan arahan yang sangat bermanfaat dalam perjalanan akademik saya.
3. Bapak Ali Mustopa, M.Kom., selaku dosen pembimbing saya, yang telah dengan sabar dan bijaksana memberikan bimbingan dan arahan yang sangat berharga selama proses penelitian dan penulisan skripsi ini.
4. Umi Siti Sholihah, S.E., M.M., ibu saya, yang senantiasa memberikan kasih sayang, doa, dan motivasi yang tiada henti dalam setiap langkah saya.
5. Abah Yahya Fatihin, ayah saya, yang dengan penuh kasih dan dukungan moral selalu menguatkan saya dalam menghadapi segala tantangan hidup dan pendidikan.
6. M. Farid Mahendra, S.Kom., kakak saya, yang selalu memberikan semangat dan dorongan positif dalam setiap proses kehidupan saya.
7. Refi Naftila, sahabat till jannah, yang selalu mendampingi, memberi dukungan, dan memberikan kebersamaan yang sangat berarti dalam perjalanan ini.

8. BPC Amikom, atas segala dukungan, pengalaman, dan fasilitas yang telah diberikan, yang sangat membantu dalam penyelesaian skripsi ini.
9. Teman-teman seperjuangan, khususnya Martonsky dan Santuf, yang selalu memberikan semangat dan dukungan dalam menghadapi setiap rintangan dan kesulitan
10. Terima kasih juga saya sampaikan kepada diri saya sendiri, Muhammad Sachib Farhan Nauvaldhi, yang telah berusaha keras dalam menyelesaikan skripsi ini dengan tekad dan semangat yang tidak pernah padam.

Semoga karya ini dapat bermanfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang informatika, dan memberi kontribusi positif bagi masyarakat luas.

Yogyakarta, 15 Januari 2025

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR LAMBANG DAN SINGKATAN

KNN	K-Nearest Neighbors
SVM	Support Vector Machine
RF	Random Forest
XGBoost	Extreme Gradient Boosting
EDA	Exploratory Data Analysis
AUC	Area Under Curve
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
ROC	Receiver Operating Characteristic
F1	F1 Score
Log Loss	Logarithmic Loss
ML	Machine Learning
CV	Cross Validation
USD	United States Dollar

DAFTAR ISTILAH

K-Nearest Neighbors	Algoritma klasifikasi yang menggunakan kedekatan titik data untuk menentukan kelas.
Support Vector Machine	Algoritma klasifikasi yang mencari hyperplane terbaik untuk memisahkan kelas.
Random Forest	Algoritma ensemble yang menggunakan banyak pohon keputusan untuk meningkatkan akurasi.
Extreme Gradient Boosting	Algoritma boosting yang menggabungkan banyak pohon keputusan untuk prediksi yang lebih akurat.
Exploratory Data Analysis	Proses analisis data untuk memahami pola dan hubungan antar variabel.
Area Under Curve	Ukuran kinerja model klasifikasi yang menunjukkan kemampuan membedakan antara kelas positif dan negatif.
True Positive	Jumlah prediksi positif yang benar.
True Negative	Jumlah prediksi negatif yang benar.
False Positive	Jumlah prediksi negatif yang salah.
False Negative	Jumlah prediksi positif yang salah.
Receiver Operating Characteristic	Grafik yang menunjukkan trade-off antara sensitivitas dan spesifisitas model.
F1 Score	Rata-rata harmonis antara presisi dan recall, digunakan untuk menilai keseimbangan model.
Logarithmic Loss	Metrik yang mengukur seberapa baik model memprediksi probabilitas setiap kelas.
Cross Validation	Metode evaluasi model yang membagi data menjadi beberapa subset untuk mengurangi variabilitas.
United States Dollar	Mata uang yang digunakan dalam konteks gaji.

INTISARI

Penelitian ini bertujuan untuk mengevaluasi efektivitas empat model klasifikasi dalam memprediksi gaji di sektor data science, menggunakan dataset "Data Science Salaries 2023." Dataset ini mencakup atribut-atribut penting, seperti jabatan, tingkat pengalaman, jenis pekerjaan, ukuran perusahaan, dan informasi relevan lainnya yang berpotensi memengaruhi struktur gaji. Penelitian ini menguji empat model klasifikasi, yaitu Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, dan XGBoost, dengan tujuan untuk menentukan model yang paling optimal dalam memprediksi kategori gaji berdasarkan kinerja metrik tertentu. Proses penelitian mencakup tahapan yang sistematis, dimulai dengan pembersihan data untuk menghapus duplikasi dan menangani data yang hilang, dilanjutkan dengan rekayasa fitur untuk meningkatkan relevansi data dengan model yang diuji. Data kemudian dibagi menjadi dua subset, yakni data latih dan data uji, dengan proporsi 80:20, memastikan model dapat dievaluasi pada data yang belum pernah dilihat sebelumnya. Setiap model dilatih menggunakan data latih dan dievaluasi menggunakan data uji berdasarkan metrik seperti akurasi, presisi, recall, F1-score, dan log loss.

Hasil penelitian menunjukkan bahwa model XGBoost memberikan kinerja terbaik dengan akurasi mencapai 98,93% dan log loss sebesar 0,038. Model ini secara konsisten unggul dalam semua metrik evaluasi, menjadikannya pilihan yang sangat baik untuk analisis data yang kompleks. Random Forest menempati posisi kedua dengan akurasi 96,67%, sementara Logistic Regression dan KNN menunjukkan kinerja yang cukup memadai namun tidak sekompetitif dua model lainnya.

Kata kunci: Pekerjaan Ilmu Data, Algoritma Pembelajaran Mesin, Klasifikasi Pekerjaan

ABSTRACT

This research aims to evaluate the effectiveness of four classification models in predicting salaries in the data science sector, using the dataset "Data Science Salaries 2023." This dataset includes important attributes, such as job title, experience level, job type, company size, and other relevant information that could potentially affect salary structure. This research tested four classification models, namely Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and XGBoost, with the aim of determining the most optimal model in predicting salary categories based on the performance of certain metrics. The research process includes systematic stages, starting with data cleaning to remove duplicates and deal with missing data, followed by feature engineering to improve the relevance of the data to the tested models. The data was then divided into two subsets, training and test data, in a proportion of 80:20, ensuring the models could be evaluated on data that had not been seen before. Each model was trained using the training data and evaluated using the test data based on metrics such as accuracy, precision, recall, F1-score, and log loss.

The results showed that the XGBoost model provided the best performance with an accuracy of 98.93% and a log loss of 0.038. This model consistently excelled in all evaluation metrics, making it an excellent choice for complex data analysis. Random Forest came in second with 96.67% accuracy, while Logistic Regression and KNN showed adequate performance but were not as competitive as the other two models..

Keyword: *Data Science Jobs, Machine Learning Algorithms, Job Classification*