

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kebutuhan akan profesional sains data telah meningkat akhir-akhir ini, didorong oleh ketergantungan yang semakin besar pada pengambilan keputusan berbasis data di beberapa industri. Seiring dengan upaya perusahaan untuk menggunakan data, memahami tren kompensasi di sektor sains data menjadi sangat penting bagi pemberi kerja dan pelamar kerja. Perkiraan gaji yang tepat memfasilitasi akuisisi dan retensi talenta, sekaligus memungkinkan individu untuk membuat keputusan karier yang terdidik [1].

Penelitian ini bermaksud untuk mengevaluasi secara cermat beberapa teknik klasifikasi biner yang digunakan pada dataset “Gaji Ilmu Data 2023”. Peneliti ingin menentukan model terbaik yang efisien untuk meramalkan kategori gaji untuk menghasilkan wawasan yang signifikan tentang tren kompensasi dalam sektor sains data [2]. Dataset ini terdiri dari beberapa variabel, termasuk jabatan, tingkat pengalaman, jenis pekerjaan, dan ukuran perusahaan, sehingga memberikan dasar yang kuat untuk studi [3].

Penelitian ini akan menilai beberapa metodologi machine learning, termasuk Regresi Logistik, K-Nearest Neighbors (KNN), Random Forest, dan XGBoost (XGB) Classifier. Setiap model akan dievaluasi berdasarkan parameter kinerja yang penting termasuk akurasi, presisi, recall, skor F1, dan log loss [4],[5]. Penelitian ini berusaha untuk menggarisbawahi kemajuan dalam pendekatan klasifikasi dan relevansinya dalam aplikasi praktis, sehingga meningkatkan pemahaman tentang dinamika gaji dalam bidang ilmu data [6].

Seiring dengan kemajuan bidang ilmu data, penilaian berkelanjutan terhadap teknik kategorisasi sangat penting untuk menjaga ketepatan dan keefektifannya [7]. Diharapkan untuk memberikan kontribusi dalam pengembangan model prediksi upah yang akurat dengan memanfaatkan teknik pembelajaran mesin, guna mendukung pengambilan keputusan yang lebih tepat

oleh pembuat kebijakan, perusahaan, dan individu. Studi ini bertujuan meningkatkan pemahaman tentang faktor-faktor yang memengaruhi upah dan memperkaya literatur ilmu data untuk perbaikan keputusan ekonomi dan social [8].

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, berikut adalah rumusan masalah yang akan diselesaikan dalam penelitian ini adalah seberapa akurat model XGB Classifier, Random Forest Classifier, Logistic Regression, dan K-Neighbors Classifier dalam memprediksi gaji?

1.3 Batasan Masalah

1. Penelitian ini hanya akan menggunakan dataset "Data Science Salaries 2023" yang mencakup variabel seperti jabatan, tingkat pengalaman, jenis pekerjaan, dan ukuran perusahaan.
2. Algoritma klasifikasi yang akan dievaluasi terbatas pada Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, dan XGBoost (XGB) Classifier.
3. Penelitian ini tidak akan membahas faktor eksternal yang mempengaruhi gaji, seperti kondisi ekonomi atau lokasi geografis.
4. Hasil penelitian ini hanya berlaku untuk data yang diambil pada tahun 2023 dan tidak mencakup tren gaji di tahun-tahun sebelumnya atau mendatang.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menganalisis efektivitas algoritma klasifikasi dalam memprediksi gaji di sektor data science, dengan fokus utama pada perbandingan kinerja antara model XGBoost, Random Forest, Logistic Regression, dan K-Nearest Neighbors (KNN). Selain itu, penelitian ini diharapkan dapat menyampaikan rekomendasi yang konstruktif bagi peneliti selanjutnya, baik dalam hal pemilihan metode analisis yang lebih efektif, pengembangan model prediksi yang lebih akurat, maupun penerapan temuan penelitian ini dalam berbagai konteks.

1.5 Manfaat Penelitian

a. Manfaat Teoritis

Hasil penelitian dapat menjadi referensi bagi penelitian selanjutnya yang ingin mengeksplorasi lebih dalam tentang algoritma klasifikasi dan aplikasinya dalam prediksi gaji.

b. Manfaat Praktis

Menghasilkan beberapa model klasifikasi untuk memprediksi tingkat gaji di bidang data science bertujuan mengeksplorasi dan membandingkan berbagai algoritma guna memperoleh model yang paling akurat dan efisien.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini menguraikan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian (teoritis dan praktis), serta sistematika penulisan. Bagian ini memberikan gambaran menyeluruh mengenai konteks dan arah penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini membahas landasan teori, kajian literatur, dan penelitian terdahulu yang relevan. Kerangka konseptual juga disusun untuk memberikan dasar analisis yang mendukung penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode penelitian, termasuk jenis penelitian, objek, populasi, sampel, teknik pengumpulan data, instrumen penelitian, dan analisis data. Penjelasan diberikan secara rinci untuk memastikan transparansi metodologis.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil penelitian berdasarkan analisis data yang dilakukan, disertai pembahasan yang menghubungkan temuan dengan teori dan kajian sebelumnya.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian dan saran untuk pengembangan lebih lanjut, baik bagi penelitian berikutnya maupun penerapannya dalam konteks praktis.