

BAB I PENDAHULUAN

1.1 Latar Belakang

Sektor kecerdasan buatan (Artificial Intelligence/AI) telah mengalami pertumbuhan eksponensial dalam beberapa tahun terakhir, mengubah lanskap industri dan menciptakan berbagai peluang karir baru. Dengan pesatnya inovasi di bidang ini, kebutuhan akan profesional AI yang terampil semakin meningkat, memicu persaingan ketat di pasar kerja. Salah satu aspek krusial dalam pengambilan keputusan karir dan perencanaan strategis bagi individu maupun organisasi adalah estimasi gaji yang akurat. Gaji dalam sektor AI dapat bervariasi secara signifikan tergantung pada sejumlah faktor seperti spesialisasi, tingkat pengalaman, lokasi geografis, dan jenis teknologi yang dikuasai, sebagaimana yang dilaporkan secara luas dalam survei industri tahunan [1].

Estimasi gaji yang tepat tidak hanya membantu calon karyawan menetapkan ekspektasi yang realistis, tetapi juga memungkinkan perusahaan merumuskan strategi kompensasi yang kompetitif untuk menarik dan mempertahankan talenta terbaik. Namun, kompleksitas data gaji yang melibatkan banyak variabel non-linear dan interaksi antar variabel membuat estimasi ini menjadi tantangan. Model statistik klasik seperti regresi linier seringkali tidak cukup mampu menangkap pola-pola kompleks ini, sehingga menghasilkan prediksi yang kurang akurat, sebuah kesimpulan yang didukung oleh studi-studi komparatif yang menunjukkan keunggulan model non-linear [2].

Untuk mengatasi keterbatasan model konvensional, *machine learning* telah muncul sebagai pendekatan yang menjanjikan. Khususnya, algoritma *ensemble learning* menawarkan solusi yang sangat andal. Metode *ensemble* bekerja dengan menggabungkan prediksi dari beberapa model dasar (*base learners*) untuk menghasilkan satu prediksi akhir yang lebih akurat dan stabil. Pendekatan ini secara konsisten terbukti lebih unggul karena kemampuannya dalam mengurangi *overfitting* dan menangani hubungan data yang kompleks.

sebuah temuan yang terus divalidasi dalam berbagai studi komparatif modern [3].

Beberapa teknik ensemble learning yang paling populer dan menunjukkan kinerja terbaik adalah Random Forest dan berbagai implementasi gradient boosting seperti XGBoost dan CatBoost. Algoritma-algoritma ini unggul karena efisiensi dan akurasi dalam menangani data yang kompleks. Selain itu, pendekatan advanced ensembling seperti Voting Regressor dan Stacking Regressor semakin mendapat perhatian karena kemampuannya menggabungkan kekuatan beberapa model untuk menghasilkan prediksi yang lebih stabil dan akurat. Berbagai penelitian terapan telah menunjukkan bahwa model berbasis *boosting* seperti XGBoost secara konsisten mencapai kinerja prediktif teratas dibandingkan model lainnya dalam tugas serupa [4]. Meskipun demikian, studi komparatif yang mendalam mengenai kinerja spesifik algoritma ensemble learning yang berbeda dalam konteks estimasi gaji di sektor kecerdasan buatan masih relatif terbatas, terutama dengan fokus pada data yang lebih mutakhir pasca-2020 yang merefleksikan tren pasar kerja AI terkini.

Oleh karena itu, penelitian ini bertujuan untuk melakukan studi komparatif terhadap kinerja beberapa algoritma ensemble learning, seperti Random Forest, XGBoost, dan CatBoost, serta penerapan metode voting dan stacking sebagai advanced ensembling dalam mengestimasi gaji pada pekerjaan sektor kecerdasan buatan. Dengan menganalisis keunggulan dan kelemahan masing-masing pendekatan, diharapkan dapat diidentifikasi metode yang paling efektif untuk memprediksi gaji secara akurat, sehingga memberikan kontribusi signifikan bagi individu dan organisasi dalam membuat keputusan yang lebih informatif di pasar kerja AI yang terus berkembang.

1.2 Rumusan Masalah

Bagaimana kinerja masing-masing algoritma ensemble learning (seperti Random Forest, XGBoost, dan CatBoost) serta metode advanced ensembling (Voting Regressor dan Stacking Regressor) dalam melakukan estimasi gaji pada pekerjaan di sektor kecerdasan buatan?

1.3 Batasan Masalah

Dalam penelitian ini memiliki beberapa batasan untuk diteliti, berikut batasan masalah penelitian yang dilakukan yaitu :

1. Penelitian ini hanya berfokus pada perbandingan tiga algoritma ensemble learning populer untuk tugas regresi, yaitu Random Forest, XGBoost, dan CatBoost, serta metode advanced ensembling berupa Voting Regressor dan Stacking Regressor. Metode lain di luar lingkup ini tidak dibahas.
2. Penelitian ini akan menggunakan dataset yang didapat dari website Kaggle untuk estimasi gaji pada pekerjaan sektor kecerdasan buatan. Namun, penelitian ini tidak akan mencakup pengumpulan data baru.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah untuk:

1. Tujuan utama penelitian ini adalah untuk menganalisis dan membandingkan secara komprehensif kinerja dari tiga algoritma ensemble learning (Random Forest, XGBoost, dan CatBoost) serta metode advanced ensembling (Voting Regressor dan Stacking Regressor) untuk mengidentifikasi model yang paling akurat, efektif, dan andal dalam melakukan estimasi gaji pada berbagai pekerjaan di sektor kecerdasan buatan (AI).
2. Meningkatkan akurasi model regresi melalui proses hyperparameter tuning dan penerapan metode advanced ensembling untuk menghasilkan prediksi estimasi gaji yang lebih optimal.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini berdasarkan tujuan diatas adalah sebagai berikut :

1. Maanfaat Teoritis

Penelitian ini akan memberikan pemahaman yang lebih mendalam

tentang keefektifan dan perbandingan kinerja dari berbagai algoritma ensemble learning yaitu Random Forest, XGBoost, CatBoost, serta metode advanced ensembling seperti Voting Regressor dan Stacking Regressor dalam konteks estimasi gaji di sektor Kecerdasan Buatan (AI). Temuan ini diharapkan dapat menjadi landasan teoritis dalam pengembangan dan pemilihan metode prediksi yang lebih tepat, adaptif, dan andal untuk data pasar kerja di masa depan.

2. Manfaat Praktis bagi Objek Penelitian

Penelitian ini dapat memberikan wawasan berbasis data mengenai faktor-faktor kunci yang memengaruhi tingkat gaji di sektor AI. Bagi pencari kerja dan profesional, wawasan ini dapat menjadi referensi dalam negosiasi gaji dan perencanaan karir yang strategis. Bagi perusahaan dan departemen SDM, hasil penelitian dapat menjadi acuan untuk menetapkan standar gaji yang kompetitif dan adil, sekaligus memanfaatkan potensi advanced ensembling untuk meningkatkan akurasi estimasi, sehingga membantu strategi rekrutmen dan retensi talenta.

3. Manfaat bagi Peneliti Selanjutnya

Hasil penelitian ini dapat menjadi studi dasar (baseline) sekaligus benchmark bagi penelitian selanjutnya. Peneliti lain dapat menggunakan temuan ini untuk mengembangkan model estimasi yang lebih canggih, menguji efektivitas algoritma baru, atau mengeksplorasi kombinasi advanced ensembling yang lebih optimal dalam konteks estimasi kompensasi profesional.

1.6 Sistematika Penulisan

Sistematika penulisan yang disusun pada skripsi ini terdapat 5 bab dan pokok-pokok permasalahan sebagai berikut:

1. BAB I PENDAHULUAN, berisi Latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta

sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA, berisi studi literatur dan dasar teori yang digunakan pada penelitian
3. BAB III METODE PENELITIAN, berisi tentang alur penelitian, metode yang digunakan dan alat dan bahan
4. BAB IV HASIL DAN PEMBAHASAN, berisi tentang hasil dan pembahasan dari penelitian.
5. BAB V PENUTUP, berisi Kesimpulan pada penelitian dan memuat saran guna untuk penelitian selanjutnya

