

BAB I

PENDAHULUAN

1.1 Latar Belakang

Data *science* adalah disiplin ilmu data yang khusus untuk mempelajari data, khususnya data kuantitatif (data numerik), baik data yang terstruktur maupun tidak terstruktur. Ilmu data telah muncul dalam menanggapi peningkatan jumlah data. Sebagian besar data telah tersedia untuk orang-orang di semua lapisan masyarakat, melalui jejaring sosial, perangkat seluler, dan berbagai perangkat sensor (yaitu "Internet Of Things"). Dengan berbagai kompleksitas data yang dihasilkan, profesi data *scientist* menjadi salah satu pekerjaan yang dicari di berbagai perusahaan untuk memberikan solusi akurat berbasis data[1].

Data *scientist* adalah profesi ilmu data dengan perkembangan yang pesat di berbagai bidang perusahaan. Meningkatnya profesi data *science*, terdapat faktor-faktor yang mempengaruhi gaji seorang data *scientist* yang menjadi permasalahan penting untuk dianalisis. Salah satu pertimbangan untuk menentukan besaran gaji profesi data *science* adalah pengalaman dan keahlian kerja yang dimiliki data *scientist*. Pengaruh variabel pengalaman dan keahlian kerja profesi data *science* sangat signifikan bagi perusahaan untuk menetapkan gaji yang adil dan kompetitif bagi seorang data *scientist*[2].

Pengalaman dan keahlian kerja suatu kecakapan yang harus dikuasai melalui pembelajaran dan bisa ditingkatkan melalui proses kerja sehari-hari. Pengalaman kerja didapatkan melalui proses panjang bagi seorang data *scientist* yang telah memiliki jam kerja lebih banyak[3]. Seiring dengan berjalannya waktu, seorang data *scientist* memiliki keahlian kerja masing-masing yang mempengaruhi efektivitas kerja dalam perusahaan. Pengalaman dan keahlian kerja yang didapatkan bisa mempengaruhi gaji data *science* yang diperoleh berdasarkan pengetahuan serta kemahiran kerja sesuai kemampuannya[4].

Analisis gaji data *science* berdasarkan pengaruh variabel pengalaman dan keahlian kerja dalam penelitian ini menggunakan algoritma *Extreme Gradient Boosting* (XGBoost). Algoritma XGBoost menggabungkan peningkatan dan pengoptimalan untuk mengatasi kesalahan pada model dengan mengurangi *error* dalam menangani masalah regresi dan klasifikasi dengan akurasi tinggi. Algoritma XGBoost dalam penelitian ini membangun model berdasarkan variabel data yang melibatkan pengalaman, keahlian, dan faktor relevan lainnya[5].

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat dirumuskan permasalahan sebagai berikut.

1. Bagaimana pengaruh pengalaman kerja terhadap besaran gaji profesi data *science*?
2. Seberapa besar pengaruh variabel keahlian kerja terhadap besaran gaji data *scientist*?
3. Berapa tingkat akurasi yang dihasilkan algoritma *Extreme Gradient Boosting* (XGBoost) dalam membantu menentukan gaji profesi data *science*?

1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan pada penelitian ini adalah sebagai berikut.

1. Dataset yang digunakan selama pembuatan skripsi diambil dari *website open source* yaitu *kaggle*.
2. Penelitian ini menekankan pada penggunaan algoritma *Extreme Gradient Boosting* (XGBoost) dan tidak membandingkan dengan algoritma lainnya.
3. Pembagian data *training* dan data *testing* dengan rasio 80:20.
4. Variabel utama yang dianalisis adalah pengalaman dan keahlian kerja terhadap gaji data *science*.
5. *Data science salaries* menggunakan data yang diambil pada tahun 2024.

6. Hasil penelitian ini menekankan pada pengaruh variabel pengalaman dan keahlian kerja terhadap gaji data *science* dengan menggunakan algoritma XGBoost.

1.4 Tujuan Penelitian

Adapun maksud dan tujuan dalam penelitian ini adalah.

1. Menganalisis pengaruh variabel pengalaman dan keahlian kerja terhadap gaji data *science* dengan menggunakan algoritma *Extreme Gradient Boosting*.
2. Mengetahui tingkat akurasi yang didapatkan dari algoritma *Extreme Gradient Boosting* untuk menganalisis gaji data *science* berdasarkan pengaruh variabel pengalaman dan keahlian kerja.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara langsung maupun tidak langsung. Adapun manfaat penelitian sebagai berikut.

a. Manfaat Teoritis

Penelitian ini diharapkan dapat mengembangkan ilmu pengetahuan di bidang teknologi informasi dan data analisis pada umumnya, dan khususnya tentang analisis gaji data *science* berdasarkan pengaruh variabel pengalaman dan keahlian kerja menggunakan algoritma XGBoost (*Extreme Gradient Boosting*).

b. Manfaat Praktis

Adapun manfaat praktis yang dapat diperoleh dalam penelitian ini sebagai berikut.

- 1) Bagi Profesional Data *Science*, penelitian ini dapat menambah wawasan dan keterampilan yang sangat berpengaruh terhadap besaran gaji data *science*.
- 2) Bagi *Human Resources* dan Rekruter, penelitian ini diharapkan dapat dijadikan sumber informasi atau referensi yang berkaitan dengan

analisis gaji data *science* sehingga dapat dengan mudah menilai calon karyawan bagi profesi data *scientist*.

1.6 Sistematika Penulisan

Untuk mengetahui pembahasan serta format penulisan laporan skripsi ini, maka peneliti membagi beberapa tahapan atau sistematika penulisan untuk dijadikan pedoman dalam melakukan penulisan sesuai dengan ruang lingkup yang terbagi atas (5) bab sebagai berikut:

BAB I Pendahuluan

Bab ini merupakan pendahuluan yang menjelaskan materinya sebagian besar menyempurnakan usulan penelitian tentang latar belakang, rumusan masalah, batasan masalah tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II Tinjauan Pustaka

Bab ini menjelaskan berbagai macam dasar teori dan metode yang digunakan sebagai landasan yang berkaitan dengan penelitian ini.

BAB III Metode Penelitian

Bab ini menjelaskan tentang metode yang digunakan dan kebutuhan untuk menganalisis data.

BAB IV Hasil dan Pembahasan

Bab ini menjelaskan tentang hasil dan pembahasan penelitian sesuai dengan rumusan masalah menggunakan metode *extreme gradient boosting*.

BAB V Penutup

Bab ini berisi kesimpulan hasil penelitian dan saran berkaitan dengan penelitian yang diusulkan untuk pengembangan lebih lanjut.

Daftar Pustaka

Berisi seluruh jurnal dan referensi acuan yang digunakan untuk membantu penelitian.

