

---

## 1. Pendahuluan

### Latar Belakang

Persaingan antar perusahaan sangat bergantung pada produktivitas karyawan [1]. Kamus Besar Bahasa Indonesia (KBBI) mendefinisikan karyawan sebagai orang yang bekerja pada suatu lembaga (kantor, perusahaan, dan sebagainya) dengan mendapatkan gaji (upah). Karyawan direkrut dengan tujuan untuk mewujudkan visi dan misi perusahaan. Setiap perusahaan pasti menginginkan karyawan yang berkualitas tinggi, sehingga manajemen sumber daya manusia (SDM) yang efektif sangat diperlukan. Salah satu tujuan manajemen SDM adalah mempertahankan karyawan yang kompeten dalam pekerjaannya. Namun, dalam pengelolaan SDM terdapat tantangan berupa *employee attrition* [2]. *Employee attrition* adalah karyawan produktif yang memutuskan untuk meninggalkan perusahaan karena berbagai alasan, seperti tekanan kerja, lingkungan yang tidak sesuai, atau gaji yang tidak memuaskan [3]. Tingkat *employee attrition* yang tinggi menimbulkan masalah signifikan bagi perusahaan karena tingginya biaya pemutusan hubungan kerja, lowongan kerja, rekrutmen, pelatihan, dan penggantian karyawan [4]. Oleh karena itu, perusahaan harus dapat mengurangi tingkat *attrition* dan mengambil tindakan pencegahan yang diperlukan untuk meminimalkan kerugian yang mungkin dialami.

Penelitian ini memprediksi *employee attrition* menggunakan seleksi fitur *Chi-square* untuk memilih fitur-fitur yang memiliki hubungan signifikan dengan fitur target. Metode klasifikasi yang digunakan adalah *Decision Tree* dan *Extreme Gradient Boosting (XGBoost)*. *Decision Tree* sangat populer digunakan untuk membangun model klasifikasi dibandingkan dengan metode lainnya karena kemudahannya dalam interpretasi dan implementasi [5]. *Decision Tree* dapat menangani data numerik dan kategorikal, memerlukan sedikit pemrosesan data, dan memiliki kinerja baik dengan dataset yang kompleks dalam waktu yang relatif singkat [6]. Sementara itu, *XGBoost* efektif dalam menangani hubungan yang kompleks dalam data, memiliki ketahanan terhadap *overfitting*, dan memberikan kinerja prediksi yang sangat baik [7].

### Topik dan Batasannya

Rumusan masalah dalam penelitian ini adalah penerapan metode *Decision Tree* dan *XGBoost* untuk memprediksi *employee attrition*, serta menganalisis pengaruh penerapan seleksi fitur *Chi-square* terhadap performa prediksi. Seleksi fitur ini bertujuan untuk meningkatkan akurasi model dengan memilih fitur-fitur yang paling relevan. Untuk menguji pengaruh seleksi fitur terhadap performa model, digunakan lima skenario model. Skenario pertama adalah model *baseline*, yaitu model yang dibangun tanpa menggunakan seleksi fitur. Skenario berikutnya adalah model dengan seleksi fitur berdasarkan 10, 15, 20, dan 25 fitur terbaik yang dipilih menggunakan seleksi fitur *Chi-square*. Setiap skenario bertujuan untuk melihat sejauh mana jumlah fitur yang dipilih dapat mempengaruhi hasil prediksi. Selain itu, penelitian ini juga membandingkan dan mengevaluasi performa masing-masing model dari kedua metode dalam hal akurasi prediksi *employee attrition*.

Adapun batasan masalah dalam penelitian ini adalah penggunaan dataset *Employee Attrition for Healthcare* yang diperoleh dari repositori Kaggle. Dataset tersebut terdiri dari 35 fitur, yang mencakup 34 fitur input dan 1 fitur target, dengan total 1.676 sampel. Data disajikan dalam format CSV dan menggunakan Bahasa Inggris.

### Tujuan

Tujuan dari penelitian ini adalah untuk mengimplementasikan metode *Decision Tree* dan *XGBoost* dalam memprediksi *employee attrition*. Penelitian ini juga bertujuan untuk menganalisis pengaruh penerapan seleksi fitur *Chi-square* terhadap hasil prediksi yang dihasilkan oleh kedua metode tersebut. Selain itu, penelitian ini bertujuan untuk membandingkan dan mengevaluasi performa kedua metode dalam memprediksi *employee attrition*.

### Organisasi Tulisan

Bagian selanjutnya dalam penelitian ini adalah bagian 2 yang membahas studi terkait dengan mengulas penelitian sebelumnya yang relevan dengan topik ini. Bagian 3 menjelaskan sistem yang dibangun. Bagian 4 memaparkan evaluasi terhadap hasil pengujian model. Terakhir, bagian 5 menyimpulkan hasil penelitian serta memberikan saran untuk penelitian lebih lanjut.