

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Penelitian sebelumnya juga telah mengeksplorasi penggunaan model klasifikasi, seperti regresi logistik dan *Naïve Bayes*, untuk mengatasi atrisi karyawan. Misalnya, sebuah penelitian oleh Nesreen El-Rayes, Ming Fang, Michael Smith, dan Stephen M. Taylor Martin Tuchman School of Management, New Jersey Institute of Technology, Newark, New Jersey, Amerika Serikat dengan judul *Predicting employee attrition using tree-based models* [31]. Jurnal ini menunjukkan kepada kita bahwa memahami penyebab pergantian karyawan sangat penting bagi organisasi untuk mengembangkan strategi retensi yang efektif dan mengurangi atrisi atau pengunduran diri[11]. Mengidentifikasi faktor-faktor yang mempengaruhi atrisi, dapat membantu manajer mengambil tindakan untuk meningkatkan retensi karyawan dan mengurangi pergantian yang menyoroti potensi *Naïve Bayes* dalam menganalisis kumpulan data yang kompleks dan terkait dengan atrisi. Temuan ini menggarisbawahi relevansi dan penerapan model nomogram dalam konteks atrisi karyawan.

*Odds ratio* (rasio peluang) adalah suatu konsep yang penting dalam analisis regresi logistik. Interpretasi *odds ratio* dalam regresi logistik melibatkan pemahaman hubungan antara transformasi *logit* variabel hasil dan klasifikasinya. *Odds ratio* diperoleh dengan mengeksponenkan koefisien regresi, dan digunakan untuk membandingkan peluang relatif terjadinya hasil yang diinginkan dengan paparan terhadap variabel yang diminati [35]. Model regresi logistik sangat

berguna untuk menganalisis variabel hasil biner, dan interpretasi hasilnya, terutama rasio *odds*, sangat penting untuk memahami dampak dari klasifikasi terhadap hasil yang diinginkan [4].

Dalam upaya mengatasi permasalahan *employee attrition*, banyak perusahaan menggunakan berbagai pendekatan termasuk *People Analytics*, yang merupakan pendekatan analitik untuk memahami dan memanfaatkan data terkait karyawan. Salah satu alat yang umum digunakan dalam *people Analytics* adalah model klasifikasi. Dalam konteks ini, penelitian akan membahas pengembangan *A Nomogram Model for Employee Attrition Identification* menggunakan teknik *Logistic Regression* dan *Naive Bayes*. Model ini akan membantu perusahaan mengidentifikasi karyawan yang berpotensi meninggalkan perusahaan, sehingga langkah-langkah pencegahan dapat diambil dengan lebih efektif. Pendekatan yang dipakai pada penelitian ini, yaitu Regresi Logistik dan Naive Bayes. Tantangan utama dalam penelitian ini mencakup penanganan keberagaman faktor yang mempengaruhi pengunduran diri, manajemen data kompleks, dan perbandingan kinerja antara kedua metode klasifikasi. Desain nomogram juga diperhatikan agar informatif dan mudah diinterpretasikan.

Pemahaman mendalam terhadap karakteristik data, interpretasi hasil, serta penanganan bias dalam *dataset* menjadi kunci dalam memvalidasi dan memastikan keandalan model.

Sebagai penutup dari penelitian ini, kinerja model nomogram dapat dievaluasi dengan menggunakan metrik seperti akurasi, presisi, dan *recall*. Selain itu, implikasi praktis dari model ini untuk pengambilan keputusan SDM dan potensinya untuk diintegrasikan ke dalam sistem SDM yang sudah ada juga perlu

didiskusikan. Arah penelitian di masa depan, termasuk penyempurnaan model dan penerapannya dalam pengaturan organisasi yang berbeda, juga dapat diuraikan untuk memberikan kesimpulan yang komprehensif terhadap penelitian ini.

## 1.2 Identifikasi Permasalahan

Identifikasi masalah pada penelitian ini adalah:

- a. Proses pemetaan atau identifikasi pengunduran diri karyawan menggunakan nomogram melibatkan penentuan variabel-variabel yang signifikan dan mempengaruhi keputusan pengunduran diri. Nomogram akan memberikan bobot atau skor relatif untuk setiap variabel yang terpilih, dan total skor ini digunakan untuk memperkirakan kemungkinan pengunduran diri. Identifikasi tersebut memerlukan pemahaman mendalam tentang hubungan antar variabel dan kontribusinya terhadap keputusan pengunduran diri.
- b. Performa nomogram berbasis Regresi Logistik (*logreg*) untuk mengidentifikasi pengunduran diri dapat dievaluasi dengan memeriksa akurasi, sensitivitas, spesifisitas, dan metrik evaluasi klasifikasi lainnya. Permasalahan yang mungkin muncul termasuk *overfitting*, seleksi fitur yang tidak optimal, atau ketidakseimbangan data.
- c. Evaluasi performa nomogram berbasis Naive Bayes juga melibatkan metrik evaluasi klasifikasi seperti akurasi, sensitivitas, dan spesifisitas. Tantangan yang mungkin timbul mencakup asumsi "naive" yang mendasari model ini, seperti asumsi independensi antar variabel.
- d. Dalam konteks *People Analytics*, penggunaan nomogram sebagai alat

untuk mengidentifikasi pengunduran diri dapat memberikan wawasan berharga kepada pemangku kepentingan. Namun, tantangan dalam mengimplementasikan nomogram sebagai alat *People Analytics* melibatkan kesesuaian model dengan konteks organisasi, interpretasi yang jelas terhadap hasil nomogram, dan kebijakan etis terkait penggunaan data karyawan. Identifikasi permasalahan potensial melibatkan penanganan aspek- aspek tersebut, seperti memastikan relevansi variabel dalam pemodelan, mengelola potensi bias, dan memahami batasan serta kekuatan dari model nomogram yang dikembangkan. Selain itu, memastikan bahwa nomogram memberikan informasi yang *actionable* dan dapat diinterpretasikan dengan mudah oleh praktisi *people analytics* juga menjadi perhatian penting.

### 1.3 Batasan Masalah

Batasan pada penelitian ini:

- a. **Sumber Data:** Penelitian ini akan menggunakan *dataset* yang disediakan oleh IBM Watson pada tahun 2015 sebagai sumber data utama. *Dataset* ini akan digunakan untuk melatih dan menguji model klasifikasi. Batasan ini mengacu pada ketersediaan data yang terbatas pada sumber data tertentu [32].
- b. **Alat Analisis:** Analisis data akan dilakukan menggunakan perangkat lunak *open source Orange Tree*. Penggunaan alat analisis ini dipilih berdasarkan ketersediaan dan kemampuan untuk mengimplementasikan model *logistic regression* dan *Naïve Bayes* dalam alat tersebut.
- c. **Waktu Penelitian:** Penelitian ini akan memfokuskan pada periode data

yang diberikan oleh *dataset* IBM Watson 2015. Oleh karena itu, perkembangan lebih lanjut setelah periode tersebut tidak akan menjadi bagian dari penelitian ini [32].

- d. **Variabel-variabel yang Digunakan:** Penelitian ini akan menggunakan variabel-variabel yang telah tersedia dalam *dataset* IBM Watson 2015. Variabel tambahan yang mungkin relevan, tetapi tidak tersedia dalam *dataset* tersebut, tidak akan dimasukkan ke dalam analisis [32].
- e. **Generalisasi Batasan:** Hasil penelitian ini mungkin tidak dapat secara langsung diterapkan pada industri atau perusahaan tertentu selain dari lingkup *dataset* IBM Watson 2015. Oleh karena itu, hasil dan model yang dikembangkan dalam penelitian ini akan memiliki generalisasi yang terbatas [32].
- f. **Aspek Eksploratif:** Penelitian ini akan berfokus pada pengembangan model klasifikasi dengan menggunakan *Logistic Regression* dan *Naive Bayes* serta pembuatan nomogram untuk interpretasi. Aspek eksploratif yang lebih mendalam tentang faktor-faktor penyebab *employee attrition* mungkin tidak akan mendapatkan penekanan yang cukup dalam penelitian ini [13].
- g. **Keterbatasan Alat Analisis:** Penggunaan perangkat lunak *Orange Tree* mungkin memiliki keterbatasan dalam hal fungsionalitas dan analisis yang dapat dilakukan dibandingkan dengan perangkat lunak statistik lain.
- h. **Penggunaan Notasi Desimal:** Dalam penelitian ini, notasi desimal yang digunakan mengikuti standar internasional yang menggunakan

tanda titik sebagai pemisah antara bagian bulat dan pecahan. Misalnya, angka 3.14 akan digunakan untuk mewakili nilai pi, bukan 3,14 seperti yang umumnya digunakan dalam penulisan bahasa Indonesia.

Dengan mempertimbangkan batasan-batasan di atas, penelitian ini akan tetap memberikan wawasan berharga dalam mengembangkan model klasifikasi untuk *employee attrition* berdasarkan *dataset IBM Watson 2015* dan mengaplikasikan metode *Logistic Regression* dan *Naive Bayes* dengan alat analisis *Orange Tree*.

#### **1.4 Rumusan Masalah**

Pada bagian ini dijelaskan secara lebih terinci dari apa yang akan diselesaikan pada batasan masalah sehingga dikatakan masalah selesai dikerjakan.

- a. Bagaimana memetakan/mengidentifikasi *employee attrition* menggunakan nomogram?
- b. Bagaimana *performance* nomogram berbasis *logreg* untuk mengidentifikasi *employee attrition*?
- c. Bagaimana *performance* nomogram berbasis *Naive Bayes* untuk mengidentifikasi *employee attrition*?
- d. Bagaimana nomogram dapat digunakan menjadi *tools people analytics*?

#### **1.5 Tujuan Penelitian**

Mengembangkan model nomogram yang dapat digunakan dalam mengidentifikasi tingkat atrisi karyawan dengan membandingkan pendekatan *logistic regression* dan *Naive Bayes*.

## 1.6 Sistematika Penulisan

Penulisan dalam penelitian ini terbagi menjadi sedikitnya dalam lima bab, di mana setiap bab mempunyai bahasan mengenai tujuan dan isi yang berbeda-beda. Adapun sistematikanya sebagai berikut:

Bab I Pendahuluan. Bab ini membahas tentang gambaran secara singkat mengenai latar belakang masalah mengapa penelitian ini dilakukan sampai pada tujuan penelitian

Bab II Kajian Teori. Bab ini membahas tentang teori-teori yang akan digunakan atau penelitian yang sudah dilakukan terkait dengan rumusan permasalahan yang dibicarakan pada Bab 1. Bagian ini merupakan bagian kunci untuk menentukan metode yang akan dipakai pada bagian selanjutnya.

Bab III Metodologi Penelitian. Bab ini berisi tentang rancangan penelitian dan atau rancangan pengujian

Bab IV Hasil dan Pembahasan. Menguraikan hasil dari penelitian yang telah dilakukan dan melakukan argumentasi atas apa yang dihasilkan dengan melampirkan *paper* atau karya ilmiah yang sudah atau akan dipublikasi.

Bab V Kesimpulan dan Saran. Pada bab ini menjelaskan tentang kesimpulan berdasarkan hasil dari penelitian yang diperoleh, serta saran-saran konstruktif yang perlu dikembangkan untuk penelitian berikutnya sehingga penelitian berikutnya menjadi lebih baik.

Di bagian akhir dari penulisan ini dilampirkan Daftar Pustaka, Lampiran-lampiran serta daftar riwayat hidup penulis.