

## **BAB I**

### **PENDAHULUAN**

#### 1.1 Latar Belakang Masalah

Salah satu elemen dalam perusahaan yang sangat penting adalah Sumber Daya Manusia (SDM). Pengelolaan SDM dari suatu perusahaan sangat mempengaruhi banyak aspek penentu keberhasilan kerja dari perusahaan tersebut. Jika SDM dapat diorganisir dengan baik, maka diharapkan perusahaan dapat menjalankan semua proses usahanya dengan baik [1]. Aset terbesar yang dimiliki sebuah perusahaan bukanlah miliknya bisnis atau tempat atau hal lain, tapi itu karyawan. Karyawan membuka jalan untuk keberhasilan perusahaan dan semua administrasi yang diperlukan untuk menjalankan perusahaan. Tetapi karena banyak alasan karyawan tidak senang dengan pekerjaan atau perusahaan tempat mereka bekerja dan ini kecenderungan atau hasil bagi mereka untuk meninggalkan perusahaan atau menemukan a perusahaan baru. Ini adalah salah satu kesulitan terpenting yang dihadapi pemilik perusahaan ketika organisasi mereka kehilangan personel paling cemerlang mereka. Karyawan yang baik adalah selalu menjadi aset berharga bagi perusahaan, dan kepergian mereka dapat mengakibatkan berbagai masalah, termasuk kerugian finansial, kinerja keseluruhan yang buruk, dan hilangnya akumulasi keahlian. Selanjutnya, dibandingkan dengan merekrut saat ini karyawan, mempekerjakan personel baru secara signifikan lebih banyak mahal, stres, dan memakan waktu [2].

Pada dasarnya perusahaan menginginkan karyawannya dapat bekerja secara maksimal sesuai dengan keahliannya masing-masing. Banyak sekali faktor-faktor yang mempengaruhi kinerja seseorang karyawan, seperti umur, jenis min, tingkat kedisiplinan, dll [3]. Menurut [4], kinerja

karyawan adalah dipengaruhi oleh psikologis, lingkungan, sosial, dan faktor fisik. Kebutuhan dari perusahaan untuk mengetahui faktor utama yang mempengaruhi loyalitas karyawan dengan menggunakan data mining.

Untuk bisa mendapatkan informasi dari data-data yang ada perlu dilakukan proses data mining seperti klasifikasi. Klasifikasi merupakan sebuah proses untuk menentukan model yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang kelasnya tidak diketahui, di dalam klasifikasi juga diberikan sejumlah record yang dinamakan training set, yang terdiri dari beberapa atribut, atribut dapat berupa kontinyu ataupun kategoris, salah satu atribut menunjukkan kelas untuk record [5].

Algoritma C4.5 adalah algoritma klasifikasi data dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihan ini misalnya dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Keakuratan prediksi yaitu kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun data ada nilai dari atribut yang hilang. Dan juga skalabilitas yaitu kemampuan untuk membangun model secara efisien untuk data berjumlah besar (aspek ini akan mendapatkan penekanan). Terakhir interpretabilitas yaitu model yang dihasilkan mudah dipahami [3]. Menurut [6], pohon keputusan atau *decision tree* adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Menurut [7], Algoritma C4.5 merupakan pengembangan dari ID3 yang mampu

mengatasi nilai yang hilang (*missing value*), mengatasi data bertipe kontinyu, dan melakukan pemangkasan pohon (*pruning tree*). Selain itu, dengan menggunakan Algoritma C4.5 dapat diketahui pulanilai akurasi klasifikasinya.

Pada permasalahan diatas mengenai karyawan maka untuk dapat mengatasi memprediksi loyalitas suatu karyawan dalam suatu perusahaan dengan mempertimbangkan berbagai faktor dan menggunakan algoritma untuk mendapatkan prediksi dengan akurasi yang besar salah satunya adalah *Decision Tree* C4.5 untuk mengklasifikasikan faktor paling penting untuk loyalitas karyawan. Pengujian sebelumnya oleh Sai Chandan P Redgg pada tahun 2021 menghasilkan akurasi yang cukup tinggi yaitu sebesar 86.39%. Merujuk hasil akurasi dari penelitian tersebut dapat dilihat bahwa *Decision Tree* menghasilkan akurasi yang tinggi namun hasil akurasi tersebut masih dapat ditingkatkan lagi dengan melakukan penelitian lanjutan untuk menghasilkan akurasi lebih tinggi dengan menambahkan *optimization feature weighting* (PSO). dan menggunakan dataset *kaggle*. Berdasarkan hasil penelitian yang dilakukan peneliti sebelumnya maka masih perlu dikembangkan lagi agar klasifikasi prediksi loyalitas karyawan mendapatkan tingkat akurasi yang lebih tinggi .

## 1.2 Identifikasi Masalah

Pada penelitian ini deperlukan adanya indentifikasi masalah, dimana masalah yang terdapat pada dataset ini adalah belum ter klasifikasinya atribut yang mempengaruhi loyalitas karyawan untuk bertahan pada sebuah perusahaan selama 2 (dua) tahun.

### 1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana cara mengkласifikasikan dan menghasilkan model algoritma yang memiliki akurasi yang sangat tinggi dengan menggabungkan *Particle Swarm Optimization* (PSO) dan algoritma *Decision Tree C4.5* pada data yang di dapat dari situs kaggle.com (<https://www.kaggle.com/datasets/tejashvi14/employee-future-prediction>).

### 1.4 Batasan Masalah

Batasan masalah pada penelitian ini yaitu membatasi ruang lingkup penelitian hanya pada klasifikasi prediksi masadepan karyawan pada perusahaan yang menggunakan metode *Particle Swarm Optimization* (PSO) dan algoritma *Decision Tree C4.5*. Dataset yang digunakan pada penelitian ini berasal dari situs Kaggle dataset yaitu data *employee future prediction* dengan atribut berjumlah 9, serta tools yang digunakan dalam penelitian ini adalah Rapid Miner 9.

### 1.5 Tujuan Penelitian

Berdasarkan permasalahan yang ada tujuan dari penelitian ini adalah untuk mengkласifikasikan atribut yang mempengaruhi masadepan karyawan dan meningkatkan akurasi dengan menggunakan metode seleksi fitur *Particle Swarm Optimization* (PSO) dan algoritma *Decision Tree C4.5*.

## 1.6 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah mempermudah dalam melakukan klasifikasi masadepan karyawan agar mendapatkan nilai akurasi yang tinggi dengan menggunakan data mining dan menguji beberapa variabel yang diperoleh dari dataset publik sehingga dapat dianalisis klasifikasi atribut yang berpengaruh dan tingkat akurasinya dengan menggunakan aplikasi rapidminer, serta dapat memberikan masukkan dalam ilmu pengetahuan untuk penelitian selanjutnya dan referensi ilmiah dalam penelitian penerapan data mining.

## 1.7 Sistematika Penulisan

Proposal ini terdiri dari tiga bagian yang secara garis besar sistematika penulisannya adalah sebagai berikut.

- a. Bab I Pendahuluan, berisi tentang latar belakang masalah, identifikasi masalah, rumusan masalah, identifikasi masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.
- b. Bab II Landasan Teori, berisi penelitian terkait dan landasan teori yang menjadi acuan dalam pelaksanaan penelitian ini.
- c. Bab III Metode Penelitian, berisi metode atau kerangka kerja yang digunakan dalam penelitian ini. Selain itu, pada bagian ini juga disajikan rencana penelitian berikut tahapan dan waktu pelaksanaannya.

- d. Bab IV Hasil Dan Pembahasan, berisi mengenai hasil, implementasi dan pembahasan penelitian. Hasil dari implementasi ini berupa gambar alat/program dan aplikasinya. Untuk penelitian lapangan hasil dapat berupa data (kualitatif maupun kuantitatif). Analisis dan pembahasan berupa hasil pengolahan data.
- e. Bab V Kesimpulan dan Saran, berisi simpulan dan saran dari isi penelitian yang sudah dibuat.