

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Evaluasi kinerja merupakan salah satu upaya dalam manajemen sumber daya manusia yang bertujuan untuk mengidentifikasi prestasi, keterbatasan dan kontribusi karyawan terhadap organisasi. Selain berperan penting dalam memberikan masukan bagi perkembangan karir karyawan, hasil evaluasi kinerja digunakan dalam berbagai pengambilan keputusan penting, seperti perubahan posisi, penentuan gaji, pemilihan kandidat promosi dan distribusi bonus (Aggarwal & Mitra Thakur, 2013). Keberhasilan evaluasi kinerja tentu menjadi prioritas utama organisasi karena memberikan manfaat dalam meningkatkan motivasi dan kepuasan kerja karyawan (Fakhimi & Raisy, 2013) sehingga berdampak positif pada produktivitas, kepuasan pelanggan, kualitas produk serta layanan jasa (Sageer, 2012), tapi keberhasilan ini seringkali terhambat akibat keterbatasan implementasi metode evaluasi secara tradisional di organisasi yang relatif subjektif, kompleks dan memakan waktu (Aggarwal & Mitra Thakur, 2013).

Sistem informasi atau evaluasi diusulkan oleh banyak organisasi untuk mengurangi kompleksitas penyelesaian tugas-tugas di bidang sumber daya manusia. Menurut (Zhu, 2021), sistem ini memiliki kemampuan dalam mengotomatisasi pengumpulan, penyimpanan data, serta kalkulasi statistik, tapi memiliki keterbatasan dalam melakukan prediksi kinerja. Keberadaan prediksi

diperlukan untuk menghasilkan proyeksi sementara kinerja karyawan, sehingga organisasi mampu mengetahui informasi kinerja lebih cepat (Gajbhiye & Gaikwad, 2020). Upaya alternatif yang dapat dilakukan untuk mengatasi keterbatasan prediksi tersebut ialah dengan menerapkan *supervised machine learning* (Cantoni, Bernazzani, & Piva, 2018). *Machine learning* atau algoritma pembelajaran mesin merupakan teknologi kecerdasan buatan yang mampu belajar melalui pola data lampau dan beradaptasi terhadap data baru. Adapun *supervised learning* merupakan jenis *machine learning* yang memerlukan data yang telah dilabeli oleh ahli untuk pelatihan dan membentuk model prediksi. Model merupakan algoritma yang telah melalui fase pelatihan serta evaluasi yang siap digunakan untuk prediksi. Menurut (Prasad, Choudhary, & B, 2020), penggunaan model berbasis *supervised learning* memungkinkan untuk mendapatkan hasil prediksi kinerja karyawan yang lebih objektif dan akurat.

*Random Forest* merupakan salah satu algoritma *supervised learning* yang didapat diandalkan untuk melakukan prediksi. Berdasarkan hasil prediksinya, algoritma *Random Forest* terbagi menjadi algoritma klasifikasi dan regresi. Algoritma klasifikasi digunakan untuk menghasilkan prediksi dengan nilai atau kelas yang dapat dihitung (diskrit) atau kategorik, sedangkan regresi digunakan untuk menghasilkan prediksi dengan nilai atau kelas yang dapat diukur (kontinu). Pada umumnya, kinerja karyawan dikelompokkan ke dalam kategori-kategori tertentu, sehingga lebih tepat menggunakan algoritma klasifikasi untuk prediksi kinerja karyawan. Penerapan *Random Forest* diyakini menghasilkan prediksi yang kompetitif dibandingkan algoritma lain karena memiliki jumlah parameter yang

sedikit, mampu menghasilkan kesalahan yang lebih rendah (Segal & Xiao, 2011), serta efektif dalam menangani derau (*noise*) (Biau & Scornet, 2016), tapi memiliki keterbatasan yang relatif mirip dengan algoritma *supervised learning* lain yakni kualitas prediksi yang bergantung pada kualitas data (Cantoni, Bernazzani, & Piva, 2018).

*Recursive Feature Elimination (RFE)* merupakan metode populer yang dapat digunakan untuk meningkatkan kualitas data dengan memilih atribut-atribut yang paling relevan. Cara kerja dari *Recursive Feature Elimination (RFE)* ialah dengan memilih kombinasi atribut terbaik sekaligus melakukan eliminasi berdasarkan peringkat fitur penting (*feature importance*) yang dihasilkan model secara rekursif sampai dengan jumlah atribut yang diinginkan. Menurut (Guyon, Weston, Barnhill, & Vapnik, 2002) dan (Cherrington, Thabtah, Lu, & Xu, 2019), hasil seleksi fitur menggunakan *Recursive Feature Elimination (RFE)* mampu mempercepat proses pelatihan dan meningkatkan kemampuan klasifikasi atau regresi. Terbukti, metode *Recursive Feature Elimination (RFE)* dapat diandalkan untuk meningkatkan kemampuan prediksi model *supervised learning* diberbagai bidang studi, seperti klasifikasi likuifikasi tanah (Demir & Şahin, 2021), klasifikasi untuk diagnosis penyakit ginjal (Senan, et al., 2021), klasifikasi kanker paru-paru (Rustam & Kharis, 2020), klasifikasi untuk prediksi penyakit parkinson (Sathiya, Reenadevi, & Sathiyabhama, 2021) dan klasifikasi kanker payudara (Ghani, Alam, & Jaskani, 2019).

Banyak peneliti yang telah mengusulkan penerapan metode *Recursive Feature Elimination (RFE)* untuk meningkatkan kemampuan klasifikasi *Random Forest*,

tapi bukan pada kasus klasifikasi kinerja. Klasifikasi kinerja karyawan yang menggunakan data dari bidang sumber daya manusia memiliki karakteristik tersendiri. Pertama, sedikitnya publikasi dataset terkait sumber daya manusia akibat faktor kerahasiaan organisasi sehingga menghambat proses pengumpulan data (Zhao, Hryniewicki, Cheng, Fu, & Zhu, 2018). Kedua, kinerja karyawan pada dataset tidak hanya terbatas pada kategori kinerja baik atau buruk, melainkan terdapat kemungkinan lebih dari dua kategori kinerja (multi kelas) yang dapat meningkatkan kompleksitas klasifikasi menggunakan *supervised learning*, seperti pada penelitian (Fadhil, 2021). Ketiga, seringkali dataset dari bidang sumber daya manusia memiliki derau (*noise*), tidak konsisten dan memiliki data hilang (*missing value*) yang berpotensi menghambat serta mengurangi kemampuan klasifikasi atau regresi *machine learning* (Zhao, Hryniewicki, Cheng, Fu, & Zhu, 2018).

Mengacu pada uraian tersebut, penelitian ini berfokus untuk mengidentifikasi hasil seleksi fitur menggunakan *Recursive Feature Elimination (RFE)* dan pengaruhnya terhadap kemampuan klasifikasi kinerja karyawan menggunakan *Random Forest*. Kemudian, penelitian ini melakukan komparasi antara kemampuan klasifikasi *Random Forest* sesudah penerapan *Recursive Feature Elimination (RF-RFE)* dengan *supervised learning* lain yaitu *Naïve Bayes*, *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, dan *Logistic Regression*. Penelitian ini menggunakan dataset publik bernama *Human Resource Dataset* (Patalano, 2020) dari *Kaggle* yang digunakan pada penelitian sebelumnya terkait klasifikasi kinerja karyawan oleh (Fadhil, 2021) dan (Jayadi R. , 2019) untuk mengatasi permasalahan keterbatasan akses data dibidang sumber daya manusia. *Human Resource Dataset*

memiliki 4 (empat) kategori kinerja, sehingga penggunaan algoritma klasifikasi *Random Forest* dalam penelitian ini berjenis multi kelas. Sementara itu, kemampuan klasifikasi model dalam penelitian ini diukur menggunakan metode *K-Fold Cross Validation* dengan metrik *Area Under The Receiver Operating Characteristics (AUC-ROC)*. Metode *K-Fold Cross Validation* merupakan metode yang dapat diandalkan untuk mengevaluasi model statistik dengan sedikit sampel (Bengio & Grandvalet, 2004). Adapun penelitian ini memilih metrik *AUC-ROC* yang lebih tepat digunakan untuk mengevaluasi kemampuan model dalam kasus klasifikasi tidak seimbang dibandingkan metrik *Accuracy*, *F1-Score*, *G-mean*, *MCC* dan *Kappa* (Wardhani, Rochayani, Iriany, Sulistyono, & Lestantyo, 2019).

Hasil penelitian ini diharapkan dapat menjadi referensi bagi organisasi, praktisi, akademisi dan pihak lain dalam perancangan basis data dan sistem evaluasi kinerja cerdas (*Intelligence Evaluation System*). Dengan demikian, penelitian ini berjudul **“Pengaruh Seleksi Fitur *Recursive Feature Elimination (RFE)* Terhadap Klasifikasi Multi Kelas Kinerja Karyawan Menggunakan *Random Forest*”**.

## **1.2 Rumusan Masalah**

Rumusan masalah yang diuraikan pada penelitian ini adalah:

1. Bagaimana perbedaan rata-rata nilai *AUC-ROC* model *Random Forest* untuk klasifikasi kinerja karyawan sebelum dan sesudah penerapan *Recursive Feature Elimination (RFE)*?
2. Atribut-atribut optimal apa yang dihasilkan seleksi fitur menggunakan *Recursive Feature Elimination (RFE)* pada *Human Resource Dataset*?

3. Bagaimana perbandingan rata-rata nilai *AUC-ROC* model *Naïve Bayes*, *K-Nearest Neighbors*, *SVM* dan *Logistic Regression* dengan model *Random Forest* sesudah penerapan *Recursive Feature Elimination (RF-RFE)* untuk klasifikasi kinerja karyawan?

### 1.3 Batasan Masalah

Terdapat batasan-batasan pada penelitian ini diantara lain:

1. Studi kasus penelitian ini merupakan dataset publik dari *Kaggle* bernama *Human Resource Dataset* yang diunggah oleh (Patalano, C., 2020).
2. Algoritma *supervised machine learning* yang digunakan dalam penelitian ini adalah *Random Forest* dengan parameter *default*. Adapun pada algoritma lain yang dibandingkan dalam penelitian ini adalah *Naïve Bayes*, *Support Vector Machine*, *K-Nearest Neighbors* dan *Logistic Regression* dengan parameter *default*.
3. Metode seleksi fitur menggunakan *Recursive Feature Elimination (RFE)*. Penelitian ini tidak membahas metode seleksi fitur lain.
4. Evaluasi *machine learning* menggunakan metode *K-Fold Cross Validation* dengan nilai  $K=5$ .
5. Metrik yang digunakan untuk mengukur kemampuan klasifikasi model pada penelitian ini adalah *AUC-ROC* dengan parameter *one-vs-rest* dan *one-vs-one*.
6. Menggunakan bahasa pemrograman *python* di *Google Colab*.

#### 1.4 Tujuan Penelitian

Penelitian ini memiliki bertujuan untuk mengukur perbedaan rata-rata *AUC-ROC* model *Random Forest* sebelum dan sesudah penerapan *Recursive Feature Elimination (RFE)* untuk klasifikasi kinerja karyawan, mengetahui hasil seleksi fitur berbasis *RFE* pada *Human Resource Dataset*, dan membandingkan rata-rata *AUC-ROC* model *Random Forest* sesudah penerapan *Recursive Feature Elimination (RF-RFE)* dengan rata-rata *AUC-ROC* model *supervised learning* lain seperti *Naïve Bayes*, *KNN*, *SVM* dan *Logistic Regression* untuk klasifikasi kinerja karyawan.

#### 1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Mengetahui pengaruh penerapan metode seleksi fitur menggunakan *RFE* terhadap kemampuan klasifikasi *Random Forest* sebagai referensi pengembangan sistem evaluasi cerdas berbasis *machine learning* bagi organisasi atau pihak lain yang memiliki kepentingan serupa.
2. Mengetahui cara memilih atribut-atribut penting menggunakan seleksi fitur berbasis *Recursive Feature Elimination (RFE)* sebagai referensi pengembangan basis data terkait kinerja karyawan bagi organisasi atau pihak lain yang memiliki kepentingan serupa.
3. Mengetahui perbandingan kemampuan klasifikasi *Random Forest* sesudah penerapan *RFE* dengan kemampuan klasifikasi model *Naïve Bayes*, *KNN*, *SVM* dan *Logistic Regression* sebagai referensi pengembangan sistem

evaluasi cerdas berbasis *machine learning* bagi organisasi atau pihak lain yang memiliki kepentingan serupa.

## 1.6 Metodologi Penelitian

Metodologi dalam penelitian ini mendeskripsikan tentang jenis penelitian, alat penelitian, sumber dan metode pengumpulan data, serta tahapan penelitian. Adapun tahapan pada penelitian ini meliputi studi pustaka, pengumpulan data, proses awal pengolahan data, eksperimen, serta penarikan kesimpulan.

Studi pustaka dilakukan di awal penelitian untuk merangkum teori-teori, kajian, literatur serta penelitian-penelitian sebelumnya yang berkaitan dengan kinerja karyawan, *Human Resource Dataset*, evaluasi teknik klasifikasi multi kelas, algoritma *Random Forest* dan metode seleksi fitur berbasis *Recursive Feature Elimination (RFE)*. Studi pustaka dalam penelitian ini diuraikan di bab 2.

Tahap selanjutnya dalam penelitian ini ialah mengumpulkan data dari sumber data. *Human Resource Dataset* didapatkan dari situs *Kaggle*. Tahap ini mendeskripsikan secara lengkap prosedur dan implementasi pengumpulan serta deskripsi dari dataset. Prosedur pengumpulan data dalam penelitian ini diuraikan pada bab 3, sedangkan implementasi pengumpulan data diuraikan di bab 4.

Proses awal pengolahan data (*preprocessing*) memuat prosedur penanganan kelemahan dataset sehingga dapat meminimalisir adanya kegagalan eksperimen di tahap selanjutnya. Prosedur proses pengolahan awal data dimuat di bab 3, sedangkan implementasi dari proses awal pengolahan data ditampilkan di bab 4.

Data siap digunakan untuk eksperimen setelah melewati tahap *preprocessing*. Terdapat 2 jenis rancangan model yang akan dievaluasi kemampuan klasifikasinya menggunakan metode *10-Fold Cross Validation* dengan metrik *AUC-ROC* pada tahap eksperimen. Pertama, rancangan model *Random Forest (RF)*, *Naïve Bayes*, *K-Nearest Neighbors (KNN)*, *Support Vector Machine (SVM)* dan *Logistic Regression* tanpa *Recursive Feature Elimination (RFE)*. Kedua, rancangan model *Random Forest* dengan *Recursive Feature Elimination (RF+RFE)*. Hasil dari evaluasi didapatkan sampel *AUC-ROC* model-model *supervised learning* tersebut. Adapun hasil lain dari eksperimen, didapatkan atribut-atribut relevan berdasarkan penggunaan seleksi fitur berbasis *RFE* pada *Human Resource Dataset*. Kemudian sampel *AUC-ROC* dari kedua model dianalisis menggunakan metode statistik deskriptif dan visualisasi data sebagai bahan kesimpulan serta saran pada tahap selanjutnya. Secara keseluruhan prosedur dan implementasi tahapan eksperimen ini diuraikan di bab 3 dan 4.

Penarikan kesimpulan merupakan tahapan akhir dari penelitian ini. Penelitian ini dapat dikatakan berhasil apabila kesimpulan mampu menjawab rumusan masalah serta tujuan penelitian. Adapun keterbatasan dari penelitian dimuat sebagai saran untuk penelitian-penelitian selanjutnya. Kesimpulan dan saran dalam penelitian ini diuraikan di bab 5.

## 1.7 Sistematika Penulisan

Penulisan dalam penelitian ini memakai sistematika sebagai berikut:

### BAB I PENDAHULUAN

Bab ini berisi tentang uraian garis besar penelitian yang terdiri atas latar belakang, rumusan, batasan masalah, manfaat, metodologi penelitian dan sistematika penulisan.

### BAB II LANDASAN TEORI

Bab ini berisi tentang uraian materi mengenai teori dan metode yang digunakan pada penelitian serta studi-studi literatur penelitian terkait kecerdasan buatan dalam kinerja, teknik klasifikasi, *RFE*, algoritma *Random Forest*, *Naïve Bayes*, *KNN*, *SVM* dan *Logistic Regression*.

### BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang jenis, alat, objek dan variabel penelitian serta tahapan penelitian. Tahapan penelitian dalam penelitian ini meliputi studi pustaka, pengumpulan data, proses awal pengolahan data, eksperimen dan penarikan kesimpulan.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang uraian hasil dan pembahasan dari pelaksanaan eksperimen.

### BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan berdasarkan hasil dari eksperimen sebagai penyelesaian masalah penelitian serta saran-saran berupa

keterbatasan penelitian sebagai evaluasi untuk penelitian selanjutnya.