

ABSTRACT

EXPLORATION AND CLASSIFICATION OF EMPLOYEE ATTRITION USING THE DECISION TREE METHOD WITH THE APPLICATION OF THE C4.5 ALGORITHM

By

Risma Nurul Hidayati

Employee attrition is a significant problem for companies because it can have a negative impact on productivity and company sustainability. To find the information contained in the employee attrition data statistically and visualize one of them using the Exploratory Data Analysis (EDA) method. EDA provides an initial description of the data used, and can assist in the classification stages of data mining. The decision tree method is a data mining technique with the aim of analyzing and modeling the relationship between variables that influence a decision, in this case the problem of employee attrition. One of the algorithms that is often used in the decision tree method is the C4.5 algorithm. One of the advantages of the C4.5 algorithm is that it can handle categorical and numeric data types and uses the gain ratio as the root determination of the model. In this study, the main factor influencing employee attrition decisions is performance ratings. The application of the C4.5 algorithm to the available employee attrition data obtains an accuracy of 88%, a recall of 99% and a precision of 88%, which means that the decision tree method is quite good at classifying data.

Keywords: Employee Attrition, Exploratory Data Analysis, Decision Tree, Algorithm C4.5.

ABSTRAK

EKSPLORASI DAN KLASIFIKASI ATRISI KARYAWAN MENGUNAKAN METODE *DECISION TREE* DENGAN PENERAPAN ALGORITMA C4.5

Oleh

Risma Nurul Hidayati

Atrisi karyawan menjadi masalah yang signifikan bagi perusahaan karena dapat berdampak negatif pada produktivitas dan keberlanjutan perusahaan. Untuk menemukan informasi yang terkandung pada data atrisi karyawan secara statistik dan visualisasi salah satunya menggunakan metode *Exploratory Data Analysis* (EDA). EDA memberikan gambaran awal tentang data yang digunakan, serta dapat membantu dalam tahapan klasifikasi data mining. Metode *decision tree* merupakan salah satu teknik data mining dengan tujuan menganalisis dan memodelkan hubungan antara variabel yang mempengaruhi sebuah keputusan dalam hal ini masalah atrisi karyawan. Salah satu algoritma yang sering digunakan pada *metode decision tree* adalah algoritma C4.5. Kelebihan algoritma C4.5 salah satunya dapat menangani data bertipe kategorik dan numerik serta menggunakan gain ratio sebagai penentuan akar pada model. Pada penelitian ini, faktor utama yang mempengaruhi keputusan atrisi karyawan adalah peringkat kinerja. Penerapan algoritma C4.5 pada data atrisi karyawan yang tersedia mendapatkan akurasi sebesar 88%, recall sebesar 99% dan precision sebesar 88% yang artinya metode *decision tree* cukup baik dalam mengklasifikasi data.

Kata Kunci: Atrisi Karyawan, *Exploratory Data Analysis*, *Decision Tree*, Algoritma C4.5.