

ABSTRAK

Cuaca adalah keadaan udara pada saat tertentu dan pada wilayah tertentu yang relatif sempit dan pada jangka waktu yang singkat. Permasalahan yang terdapat pada prakiraan cuaca yaitu banyaknya atribut dan parameter penentu cuaca yang akan mengurangi kinerja dari algoritma dan menyebabkan akurasi menjadi rendah. Hal ini disebut dengan *curse of dimensionality*. Hal lain yang dapat menyebabkan akurasi sebuah algoritma tidak maksimal yaitu adanya ketidakseimbangan kelas antara kelas mayoritas dan kelas minoritas. Hal ini disebut dengan *imbalance class*. Data penelitian yang digunakan didapat dari situs *Iowa Environmental Mesonet* untuk data historis wilayah Surabaya dari tahun 2021 sampai dengan tahun 2024 yang diambil setiap 24 jam sekali setiap harinya. Datasets terdiri dari 18 atribut dan 13 kelas dengan jumlah data sebanyak 1388 record data. Algoritma klasifikasi data mining yang digunakan pada penelitian yaitu algoritma *decision tree C4.5*. Didapatkan pola klasifikasi cuaca. Yakni metode *Particle Swarm Optimization* (PSO) digunakan untuk mengoptimalkan fitur yang ada pada datasets dan penggunaan metode. Pengujian model algoritma menggunakan model validasi *10-fold cross validation* menghasilkan akurasi algoritma *decision tree C4.5* yang cukup baik yaitu 84.72%, setelah dioptimasikan meningkat menjadi kurang lebih sebesar 87.24 %.

Kata kunci: *Decision Tree C4.5, RapidMiner 5, Data mining, Particle Swarm Optimazion (PSO)*

ABSTRACT

Weather refers to atmospheric conditions at a specific time and within a relatively small, defined area over a short period. A key challenge in weather forecasting is the large number of weather-determining attributes and parameters, which can degrade algorithm performance and lead to low accuracy. This phenomenon is known as the curse of dimensionality. Another factor that can prevent an algorithm from achieving maximum accuracy is class imbalance between the majority and minority classes. This is known as class imbalance. The research data used was obtained from the Iowa Environmental Mesonet website for historical data on the Surabaya region from 2021 to 2024, collected once every 24 hours daily. The dataset consists of 18 attributes and 13 classes, with a total of 1,388 data records. The data mining classification algorithm used in this study is the C4.5 decision tree algorithm. A weather classification pattern was obtained. Specifically, the Particle Swarm Optimization (PSO) method was used to optimize the features in the dataset and the algorithm's performance. Testing the algorithm model using a 10-fold cross-validation model yielded a fairly good accuracy for the C4.5 decision tree algorithm of 84.72%; after optimization, this increased to approximately 87.24%.

Keywords : C4.5 Decision Tree, RapidMiner 5, Data Mining, Particle Swarm Optimization (PSO)