

**PENERAPAN ANALISIS KOMPONEN UTAMA UNTUK
MENINGKATKAN AKURASI KLASIFIKASI PADA ALGORITMA DECISION
TREE C4.5 DALAM MENDIAGNOSA PENYAKIT DIABETES MELITUS**

Michael Yulius Sitanggang
NIM: 4183230026

ABSTRAK

Pemisahan atribut merupakan proses utama algoritma klasifikasi Decision Tree C4.5. Proses pemisahan atribut C4.5 masih belum optimal terhadap pembentukan pohon keputusan untuk memindahkan fitur yang tidak diperlukan. Fitur yang tidak diperlukan dapat memicu adanya data noise berupa outlier sehingga terjadi overfitting dan menyebabkan kesalahan klasifikasi dan ketidakseimbangan data. Banyak algoritma telah diusulkan untuk mengurangi kesalahan klasifikasi pada Decision Tree C4.5. Reduksi fitur diharapkan dapat mengurangi data noise sehingga mampu meningkatkan akurasi model. Kerangka reduksi fitur dilakukan guna menyederhanakan data berdimensi tinggi dengan atribut yang saling berkorelasi menjadi data berdimensi rendah dan atribut yang tidak berkorelasi. Pada penelitian ini, metode reduksi fitur yang diusulkan adalah analisis komponen utama (PCA) dan kemudian dilakukan klasifikasi menggunakan algoritma Decision Tree C4.5. Pengujian model C4.5 dengan PCA dilakukan menggunakan dataset dari UCI Machine Learning yang dikumpulkan berdasarkan tiga sumber yaitu: Cleveland Clinic Foundation; V.A. Medical Center Long Beach, CA; University Hospital Zurich, Switzerland dan terdiri dari 419 data rekam medis pasien dengan 15 atribut observasi. Pengukuran kinerja model klasifikasi dilihat berdasarkan akurasi, presisi dan sensitivitas. Hasil pengujian menunjukkan bahwa model usulan ini dapat meningkatkan akurasi algoritma klasifikasi yang semula sebesar 64.29% meningkat menjadi 70.84%.

Kata kunci: Decision Tree C4.5, Klasifikasi, Principal Component Analysis, Data Noise, Reduksi Fitur

