

INFORMATION GAIN UNTUK PEMILIHAN FITUR PADA KLASIFIKASI HEREGISTRASI CALON MAHASISWA DENGAN MENGGUNAKAN K-NN

M. ADIB AL KAROMI

Program Studi Magister Teknik Informatika

Fakultas Ilmu Komputer

Universitas Dian Nuswantoro

Semarang 50131, Indonesia

ABSTRAKSI

STMIK Widya Pratama merupakan salah satu perguruan tinggi swasta yang ada di Pekalongan. Setiap tahun STMIK membuka pendaftaran untuk mahasiswa dengan dibantu tim Penerimaan Mahasiswa Baru beserta tim marketing. Data lampau menunjukkan bahwa jumlah mahasiswa yang melakukan heregistrasi kurang dari 80% dari jumlah pendaftar. Bahkan pada tahun 2013 ada 37,42% calon mahasiswa tidak melakukan heregistrasi. Sedikit banyaknya jumlah mahasiswa akan mempengaruhi kemajuan dan keuntungan sebuah Perguruan Tinggi Swasta. Data mining merupakan sebuah proses ekstraksi untuk mendapatkan suatu informasi yang sebelumnya tidak diketahui dari sebuah data. Data mining dapat menganalisa kasus lama untuk menemukan pola dari data dengan menggunakan teknik pengenalan pola seperti statistik dan matematika. Salah satu peran utama data mining adalah klasifikasi. Terdapat banyak teknik klasifikasi dengan karakteristiknya masing masing dan salah satunya adalah K-NN. Tingkat akurasi suatu algoritma sangat dipengaruhi oleh tipe data dan banyaknya atribut dalam sebuah data. K-NN digunakan untuk data berdimensi tinggi dan banyak dipakai dalam proses klasifikasi. Salah satu kelemahan K-NN adalah penentuan k yang paling optimal serta rentan terhadap atribut yang tidak relevan. Setiap atribut memiliki peran yang sama dalam klasifikasi menggunakan K-NN. Seleksi fitur merupakan salah satu tahap pre processing data dalam klasifikasi untuk meningkatkan akurasi. Dalam seleksi fitur semua fitur yang dianggap berpengaruh terhadap hasil klasifikasi akan dipertahankan dan sisanya tidak akan dipakai dalam proses klasifikasi. Dalam tulisan ini akan disajikan information gain untuk seleksi fitur klasifikasi heregistrasi calon mahasiswa dengan menggunakan algoritma K-NN.