

BAB 1 PERSYARATAN PRODUK

Bab ini membahas mengenai hal umum dari produk yang dibuat, meliputi tujuan, ruang lingkup proyek, perspektif produk, fungsi produk dan hal umum yang lainnya.

1.1 Pendahuluan

Hal – hal umum yang dibahas pada sub bab ini adalah latar belakang pembuatan produk, tujuan pembuatan produk, ruang lingkup proyek, definisi-singkatan-akronim yang digunakan, dan sistematika pembahasan.

1.1.1 Latar Belakang

Fakultas Teknologi Informasi telah berdiri sejak 5 hingga 6 tahun yang lalu, saat pertama kali berdiri, Fakultas Teknologi Informasi masih tergabung dengan Fakultas Teknik Universitas Kristen Maranatha. Saat ini Fakultas Teknologi Informasi telah menjadi Fakultas yang terpisah dari Fakultas Teknik dan memiliki sekitar 1000 mahasiswa. Data mahasiswa yang dimiliki Fakultas Teknologi Informasi terus bertambah setiap tahunnya sedangkan data historis mahasiswa terus bertambah setiap semesternya.

Data-data mahasiswa yang dimiliki Fakultas Teknologi Informasi dapat dimanfaatkan untuk mengevaluasi kualitas lulusan yang dihasilkan Fakultas Teknologi Informasi. Evaluasi dilakukan dengan mengolah data-data mahasiswa tersebut dengan menggunakan *data mining*.

1.1.2 Tujuan

Secara khusus, pembuatan aplikasi ini bertujuan untuk :

1. Melakukan interpretasi terhadap hasil dari proses asosiasi dan klasifikasi yang dihasilkan WEKA.

2. Membandingkan hasil proses data mining yang dilakukan secara manual dengan hasil proses yang dilakukan oleh WEKA agar diketahui apakah WEKA dapat diandalkan untuk melakukan proses data mining atau tidak.
3. Membandingkan hasil proses data mining dengan jumlah data yang berbeda agar diketahui perbedaan hasil dari proses yang dilakukan terhadap data yang berbeda secara signifikan.

1.1.3 Ruang Lingkup Proyek

Aplikasi ini ditujukan untuk digunakan dalam penambangan data mahasiswa Fakultas Teknologi Informasi Universitas Kristen Maranatha. Pengguna mengoperasikan aplikasi ini pada suatu komputer dengan sistem operasi tertentu dan penampungan data yang tertentu juga.

Secara garis besar, aplikasi ini merupakan aplikasi yang digunakan untuk menambang data mahasiswa yang ditampung dalam suatu tempat penampungan data (*data repository*), yaitu *Microsoft Database File* (mdf) dan menyajikan hasil penambangan kepada pengguna aplikasi.

Data mahasiswa yang digunakan dapat berupa data mahasiswa secara keseluruhan (dalam satu fakultas) maupun secara lebih spesifik (per jurusan/program studi).

Aplikasi ini menggunakan *engine* WEKA versi 3.5.8 dalam pencarian pola yang terdapat dalam data mahasiswa dan pengklasifikasian data mahasiswa tersebut serta menampilkan hasil dari proses yang dilakukan ke dalam bentuk tabel, grafik, dan pohon keputusan sehingga dapat membantu fakultas / program studi dalam menganalisis data mahasiswanya.

1.1.4 Definisi, Akronim, dan Singkatan

- *Aggregation* : Pengelompokan atau penyimpulan yang digunakan pada data.
- *Association* : Pola / hubungan (*pattern*) antara variabel yang terbentuk karena variabel-variabel pembentuk pola sering muncul dalam satu rangkaian data (*tuples*) saat pembacaan data secara menyeluruh (*full scan*) pada suatu tempat penampungan data.
- *Attribute construction* : Proses penambahan atribut baru yang digunakan untuk membantu pemrosesan data.
- *Boolean association rules* : pola asosiasi yang bernilai diskrit (*boolean*), bukan yang bernilai *continuous*.
- *Characterization* : Rangkuman dari karakteristik atau fitur umum dari suatu *target class* dari data.
- *Classification* : Proses pencarian model (atau fungsi) yang menjelaskan kelas atau konsep data dan menggunakan model tersebut untuk memprediksi kelas dari suatu objek yang kelasnya tidak diketahui.
- *Cluster analysis* : Pengelompokan data berdasarkan kesamaan tertentu.
- *Confidence (certainty)* : Batas / nilai minimum dari suatu pola (*rule*) yang harus dipenuhi. Nilai minimum ini dibandingkan dengan nilai yang diperoleh dari hasil *full scan data repository* yang mencari *rule* yang menunjukkan persentase jumlah atribut yang berada di sebelah kiri dan kanan.
- *Data cleaning* : proses membuang data yang tidak lengkap dan tidak konsisten.
- *Database* : Kumpulan data yang terstruktur yang tersimpan dalam komputer.
- *Data integration* : menggabungkan beberapa sumber data yang akan digunakan untuk pengolahan data.

- *Data mining* : Proses menemukan informasi (pengetahuan) dari data – data yang tersimpan dalam *database*, *data warehouses* atau jenis penyimpanan data (*data repositories*) yang lainnya.
- *Data preprocessing* : Langkah pembersihan data dari semua data yang tidak diperlukan. Langkah – langkah tersebut meliputi, *data cleaning*, *data integration*, *data selection*, dan *data transformation*.
- *Data repositories* : Jenis penyimpanan data / penampungan data.
- *Data selection* : pemilihan data yang relevan yang akan diolah.
- *Data transformation* : Proses mengubah format data dengan tujuan mempercepat pengolahan data.
- *Data warehouses* : Salah satu jenis data repository, data yang ditampung sangat kompleks, datanya dapat berupa gabungan dari beberapa *database*.
- *Discrimination* : Perbandingan karakter atau fitur umum dari dua target class yang berbeda.
- *Evolution analysis* : proses yang menggambarkan dan memodelkan kebiasaan atau trend untuk objek yang kelakuannya berubah sepanjang waktu.
- *Frequent itemset* : Sekumpulan *item* (atribut) yang sering muncul bersama pada *transactional dataset*.
- *Frequent Patterns* : Pola yang sering muncul pada suatu kumpulan data.
- *Frequent subsequence* : Sekumpulan *item* (atribut) yang sering muncul bersama yang dilihat berdasarkan urutan kemunculannya (*sequential pattern*).
- *Generalization* : Proses mengganti data yang memiliki *level* rendah (*primitive*) menjadi data yang memiliki *level* yang lebih tinggi.

- *Normalization* : Proses membuat skala data menjadi skala yang lebih kecil.
- *Outlier analysis* : Metode yang memproses data – data yang memiliki karakteristik yang berbeda dengan karakteristik pada umumnya.
- *Query language* : Bahasa yang digunakan untuk menambang *database*.
- *Relational database* : *Database* yang terstruktur berdasarkan tabel – tabel yang berelasi.
- *Sequential pattern* : Pola yang terbentuk berdasarkan urutan kemunculan *itemset* yang terdapat pada pola tersebut.
- *Smoothing* : Proses membuang data yang tidak lengkap, menggunakan teknik seperti *binning*, *regression*, dan *clustering*.
- *Support* : Batas / nilai minimum dari suatu pola (*rule*) yang harus dipenuhi. Nilai minimum ini dibandingkan dengan nilai yang diperoleh dari hasil *full scan data repository* yang mencari persentase jumlah *rule* yang terdapat pada data repository.
- *Transactional databases* : *Database* yang berisi satu *file* dimana tiap *tuple* berisi keterangan transaksi.
- *Tuple* : rangkaian data yang terdapat dalam *dataset*. Rangkaian data tersebut berupa satu baris data (*row / itemset*) dalam tabel.

1.1.5 Sistematika Pembahasan

Struktur laporan dibuat berdasarkan format yang telah ditentukan, dimulai dengan bab satu membahas gambaran umum mengenai produk yang dibuat, gambaran fungsionalitas produk, pengguna, batasan dan persyaratan. Bab dua membahas gambaran pengembangan produk secara lebih

detail, termasuk informasi penggunaan bahasa pemrograman, *tools* yang digunakan, gambaran rancangan awal hingga fitur – fitur dari produk. Bab tiga membahas landasan teori, perancangan dari produk, meliputi identifikasi sistem, diagram hingga desain arsitektur yang digunakan. Bab empat membahas modul implementasi, perjalanan implementasi, dan realisasi *user interface design*. Bab lima membahas pengujian yang dilakukan, yang meliputi rencana pengujian dan perjalanan pengujian. Bab enam membahas kesimpulan dan saran.

1.2 Gambaran Keseluruhan

1.2.1 Perspektif Produk

Aplikasi ini dikembangkan dengan menggunakan bahasa pemrograman C# yang dibuat dengan bantuan aplikasi Microsoft *Visual Studio 2005*. *Library data mining* yang digunakan adalah WEKA 3.5.8.

Aplikasi ini tidak memiliki *data repository* yang terintegrasi, user dapat memilih sendiri jenis *data repository* yang akan diolah. *Data repository* yang dapat digunakan antara lain, *database file*, *CSV file* atau *arff file*. Hasil pengolahan data ditampilkan dalam bentuk tulisan (*text*) maupun pohon bercabang (*tree*) yang telah dimodifikasi sehingga pengguna awam dapat mengerti hasil yang diperoleh.

Aplikasi ini dibuat dengan tampilan yang *user friendly* sehingga pengguna dapat dengan mudah menggunakan aplikasi ini. Penggunaan *Object Oriented Programming* pada aplikasi ini juga dapat memudahkan *developer* dalam membaca maupun melanjutkan pengembangan aplikasi ini.

1.2.2 Fungsi Produk

Fungsi-fungsi aplikasi ini diantaranya :

1. Memilih *data repository* yang akan digunakan
2. Memilih data pada *data repository* yang digunakan.
3. Memilih Atribut yang akan digunakan.
4. Penyaringan data yang digunakan.
5. Mencari pola dari data.
6. Mengklasifikasi data.
7. Memprediksi data.
8. Menampilkan hasil kepada pengguna.
9. *Import* data.

1.2.3 Karakteristik Pengguna

Aplikasi ini memerlukan pengguna yang dapat mengoperasikan perangkat komputer. Pengguna yang memiliki pengetahuan umum tentang *database* dan *data mining* lebih diutamakan.

1.2.4 Batasan-Batasan

Batasan – batasan pada aplikasi ini adalah :

- *Data repository* yang digunakan adalah Microsoft *database file* (mdf).
- Aplikasi hanya menampilkan hasil dari proses yang diminta.
- Proses yang dapat dipilih hanya proses asosiasi dan klasifikasi.
- Teknik asosiasi yang digunakan adalah Apriori, sedangkan teknik klasifikasi yang digunakan adalah *treeJ48* dan *naïve bayes*.
- Teknik penyaringan data yang digunakan adalah *NumericToNominal* dari *filter unsupervised*.

1.2.5 Asumsi dan Ketergantungan

- Aplikasi ini dikembangkan dengan menggunakan bahasa pemrograman C# dan *library* yang

digunakan adalah WEKA 3.5.8 dalam JAVA yang dikonversi menggunakan IKVM.

- WEKA 3.5.8 yang bersifat *open source*.
- Data yang digunakan adalah data mahasiswa fakultas teknologi informasi angkatan 2003 - 2007.