

## **BAB III**

### **METODE PENELITIAN**

#### **3.1. Bahan dan Alat Penelitian**

Bahan yang digunakan dalam penelitian ini adalah data hasil ujian berbasis kertas yang dibandingkan dengan data referensi peserta ujian dan data kode soal yang terdaftar. Komponen data peserta ujian yang digunakan adalah nomor tes dan tanggal lahir. Komponen data lembar jawab ujian yang digunakan yaitu isian nomor tes, isian tanggal lahir, isian kode soal ujian, dan informasi kevalidan lembar jawab. Sedangkan komponen data soal ujian yang digunakan hanya kode soal.

Adapun alat penelitian terdiri atas perangkat keras dan perangkat lunak. Perangkat keras terdiri atas sebuah *notebook* dengan spesifikasi prosesor Intel Core i3 1,8 GHz, kapasitas memori 4 GB, dan kapasitas *harddisk* 500 GB. Sedangkan perangkat lunak yang digunakan berupa sistem operasi Microsoft Windows 10, pemrograman berbasis web PHP 5.5 dengan server web Apache 2.4, dan *script* tampilan web HTML 4. Adapun untuk keperluan analisis data, digunakan perangkat lunak WEKA versi 3.8.1.

#### **3.2. Prosedur Penelitian**

Prosedur penelitian ditunjukkan dalam Gambar 3.1. Keterangan dalam tanda kurung merepresentasikan tahapan *data mining* yang diterapkan. Metode yang dilakukan dalam penelitian ini terdiri atas studi literatur, observasi, pengembangan sistem, operasional sistem dan pengambilan data, analisis data, dan penulisan laporan. Studi literatur dilakukan dengan mengumpulkan informasi dari buku, artikel jurnal, tesis, dan sumber ilmiah lain yang terkait dengan penelitian ini. Observasi dilakukan dengan pengamatan pada sistem yang ada. Observasi tersebut setara dengan dua tahap awal *data mining* yaitu pemahaman bisnis dan pemahaman data.



Gambar 3.1 Prosedur Penelitian

Dari tahap pemahaman bisnis dan data diperoleh informasi sebagai berikut. Dalam pemrosesan awal lembar jawab ujian berbasis kertas, dilakukan digitalisasi isian serta validasi komponen data peserta untuk memasok proses selanjutnya yaitu koreksi jawaban dan penilaian. Secara sederhana, setiap *record* lembar jawab diberi label status valid jika isian lembar jawab sesuai dengan data referensi peserta serta isian kode soal terdaftar dalam kode soal yang ditentukan saat ujian. Jika isian tidak sesuai dengan data referensi peserta dan kode soal,

maka lembar jawab diberi label status tidak valid. Kemudian dibentuk file teks yang berisi data isian peserta dan status lembar jawab. *Record* dalam file tersebut tidak dapat langsung dijadikan masukan untuk piranti analisis data.

Pengembangan sistem *preprocessing* dilakukan berdasarkan pada spesifikasi *requirement* saat observasi. Diperlukan sistem untuk pra-pemrosesan data berbasis web (PHP dan HTML) dalam rangka mengimplementasikan tahapan ketiga *data mining* yaitu persiapan data. Pada tahap tersebut dilakukan transformasi data sehingga hasilnya siap untuk dianalisis. Setelah sistem pra-pemrosesan data berhasil dikembangkan dan diuji, dilakukan penerapan sistem dan pengambilan data hasil pra-pemrosesan.

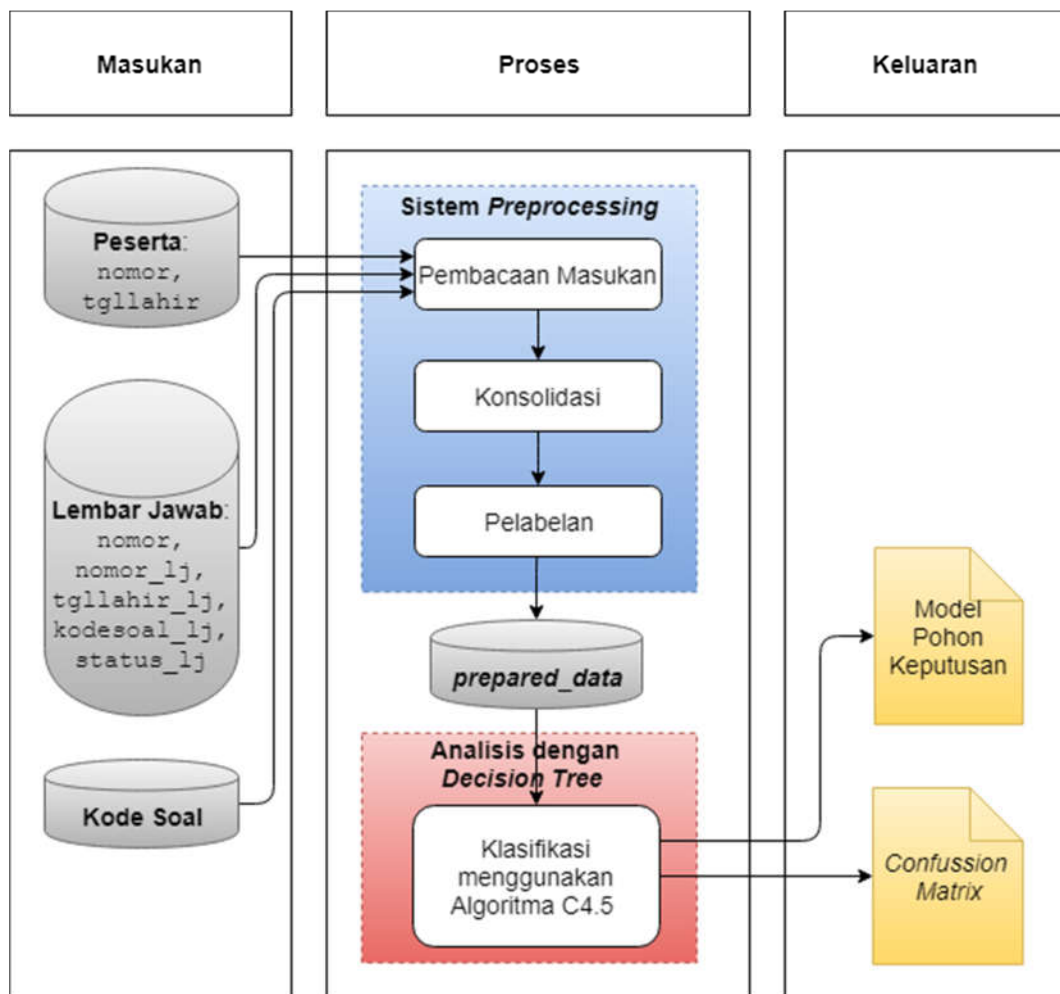
Selanjutnya dilakukan analisis data yang meliputi penyusunan asumsi, pengolahan data untuk membangun model, dan penarikan kesimpulan. Penyusunan asumsi dilakukan dengan merumuskan asumsi pertama yang berisi pernyataan yang akan diselidiki, dan merumuskan asumsi kedua yang merupakan pernyataan tandingan dari asumsi pertama. Pengolahan data menggunakan piranti analisis data WEKA di mana proses ini setara dengan tahap pemodelan dan evaluasi dalam *data mining*. Kesimpulan yang diambil berdasarkan hasil perhitungan berupa menerima atau menolak asumsi pertama. Adapun langkah terakhir adalah penyusunan laporan penelitian dalam bentuk tesis.

### **3.3. Kerangka Sistem Informasi**

Kerangka sistem informasi yang dibangun terdiri atas unsur masukan, proses, dan keluaran. Kerangka tersebut ditunjukkan pada Gambar 3.2. Tahapan *data mining* yang diterapkan dalam sistem tersebut adalah persiapan data, pemodelan, dan evaluasi. Persiapan data dilakukan dengan membangun dan menjalankan program berbasis web, sedangkan pemodelan dan evaluasi dilakukan menggunakan piranti analisis data.

Dalam proses *preprocessing* data terdapat tahap pembacaan file masukan, pencocokan data, dan pelabelan isian komponen data. Proses ini dilakukan menggunakan program berbasis web yang dikembangkan sebelumnya. Pertama kali dilakukan tiga hal yaitu pembacaan file data peserta dan hasilnya dimuat

dalam tabel *ap* (*array of peserta*), pembacaan masukan daftar kode soal dari *keyboard*, dan pembacaan file hasil digitalisasi lembar jawab yang hasilnya disimpan dalam tabel *aj* (*array of jawaban*). File lembar jawab juga berisi informasi nomor tes yang seharusnya diisi oleh peserta, dan status lembar jawab hasil proses eksternal.



Gambar 3.2 Kerangka Sistem Informasi

Selanjutnya dilakukan proses konsolidasi isi tabel *ap* dan tabel *aj* untuk mendapatkan data yang bersesuaian antara nomor tes yang seharusnya (*nomor*), nomor yang diisi di lembar jawab (*nomor\_lj*), tanggal lahir yang seharusnya

(*tgllahir*), tanggal lahir yang diisikan di lembar jawab (*tgllahir\_lj*), kode soal yang diisikan di lembar jawab (*kodesoal\_lj*), dan status lembar jawab (*status\_lj*). Tahapan ini menggunakan nomor tes yang terdapat pada file lembar jawab sebagai *foreign key* yang mengacu ke nomor tes di data peserta. Dengan demikian, nilai tanggal lahir juga dapat diperoleh dari data peserta berdasarkan *key* tersebut.

Kemudian dilakukan proses pelabelan isian komponen data. Setiap nilai isian di lembar jawab yang sama dengan nilai komponen data peserta diberi label benar, sedangkan isian yang berbeda dilabeli salah. Khusus kode soal, dilakukan pencarian berdasarkan nilai kode soal yang telah dimasukkan. Apabila kode soal di lembar jawab terdaftar dalam daftar kode soal, maka status isian kode soal tersebut dilabeli benar, sedangkan apabila tidak ditemukan dalam daftar kode soal, maka isian kode soal tersebut diberi label salah. Adapun nilai status lembar jawab diberi label benar untuk nilai 0 atau diberi nilai salah untuk nilai 1. Hasil pelabelan disimpan dalam sebuah file teks berekstensi CSV untuk selanjutnya menjadi masukan proses analisis data.

Dalam proses analisis data, dilakukan klasifikasi menggunakan algoritma C4.5 sehingga menghasilkan keluaran sistem. Keluaran sistem tersebut berupa model pohon keputusan yang mencerminkan aturan pengambilan keputusan berdasarkan kondisi data. Selain itu, keluaran sistem yang lain berupa *confussion matrix* menggambarkan tingkat ketepatan klasifikasi data. Opsi pengujian klasifikasi dilakukan pada dua jenis, yaitu *Cross Validation 10 folds* dan *Percentage Split 66%*. Kemudian dilakukan analisis dan interpretasi atas model pohon keputusan yang terbentuk, validasi hasil dibandingkan dengan teori, dan evaluasi hasil berdasarkan *confussion matrix*. Pertimbangan yang dilakukan dapat menjadi dasar untuk memutuskan asumsi mana yang dapat diterima.