

## **BAB III**

### **METODOLOGI PENELITIAN**

Dalam penelitian ini metodologi berperan untuk mendapatkan data yang obyektif serta valid sehingga dapat digunakan untuk memecahkan permasalahan yang ada. Oleh sebab itu penulisan metodologi sangat diperlukan untuk mengumpulkan data sehingga dalam menyusun laporan penelitian ini dapat dipertanggungjawabkan.

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

##### **3.1.1. Alat Penelitian**

Dalam pelaksanaan penelitian ini dibutuhkan beberapa alat seperti perangkat lunak dan perangkat keras komputer. Adapun alat-alat pendukung yang digunakan pada penelitian ini adalah sebagai berikut :

##### 1. Perangkat Keras (*Hardware*)

Laptop atau Komputer

Demi kelancaran penelitian ini, penulis menggunakan laptop atau komputer untuk simulasi dan analisis hasil pengolahan data dengan spesifikasi sebagai berikut :

- 1) Laptop *Acer Aspire E 14*
- 2) RAM 8 Gb
- 3) *Hardisk* 500 Gb
- 4) *Processor Intel Core i5*
- 5) Sistem Operasi *Windows 10* 64-bit

## 2. Perangkat Lunak (*Software*)

### a. *Microsoft Office Excel 2016*

*Microsoft Office Excel 2016* adalah sebuah program aplikasi lembar kerja *spreadsheet* yang dibuat dan dikelola oleh *Microsoft Corporation* dan dapat dijalankan pada *Microsoft Windows* dan *Mac OS*.

*Microsoft Excel* merupakan *software* yang digunakan untuk membuat tabel dan juga *software* ini dapat digunakan untuk menghitung rumus-rumus matematika serta dapat menampilkan persyaratan suatu kondisi secara langsung tanpa harus menuliskannya satu persatu.

### b. *RapidMiner 5*

*RapidMiner* adalah sebuah perangkat lunak yang bersifat *open source*. *RapidMiner* digunakan untuk melakukan sebuah analisis terhadap *data mining*, analisis prediksi, dan *text mining*. *RapidMiner* menyediakan prosedur *machine learning* dan *data mining*. Dan di dalam aplikasi tersebut terdapat ETL, *data preprocessing*, visualisasi, *modelling*, dan evaluasi.

### c. *Microsoft Office Word 2016*

*Microsoft Office Word 2016* merupakan *software* yang digunakan untuk mengelola kata (*word processor*). Di dalamnya terdapat beberapa fitur unggulan yaitu *AutoCorrect*, *MailMerge*, *Autopage Index*, dan lain-lain. *Software* ini mendukung penyuntingan file *Adobe Reader (.pdf)* dan juga terintegrasi dengan *Skydrive* sehingga dapat menyimpan data di *Cloud*.

### 3.1.2. Bahan Penelitian

Dalam pelaksanaan penelitian ini dibutuhkan beberapa bahan penelitian. Adapun bahan-bahan pendukung yang digunakan pada penelitian ini adalah sebagai berikut:

#### a. Literatur

Literatur adalah jurnal, buku, dan artikel yang berhubungan dengan algoritma C4.5, *decision tree*, dan *data mining* yang relevan dengan penelitian. Semuanya digunakan sebagai bahan referensi untuk memperdalam pemahaman teori-teori dan penerapan data.

#### b. Data Alumni Universitas

Data berasal dari *database* BSI (Biro Sistem Informasi) Universitas Muhammadiyah Yogyakarta.

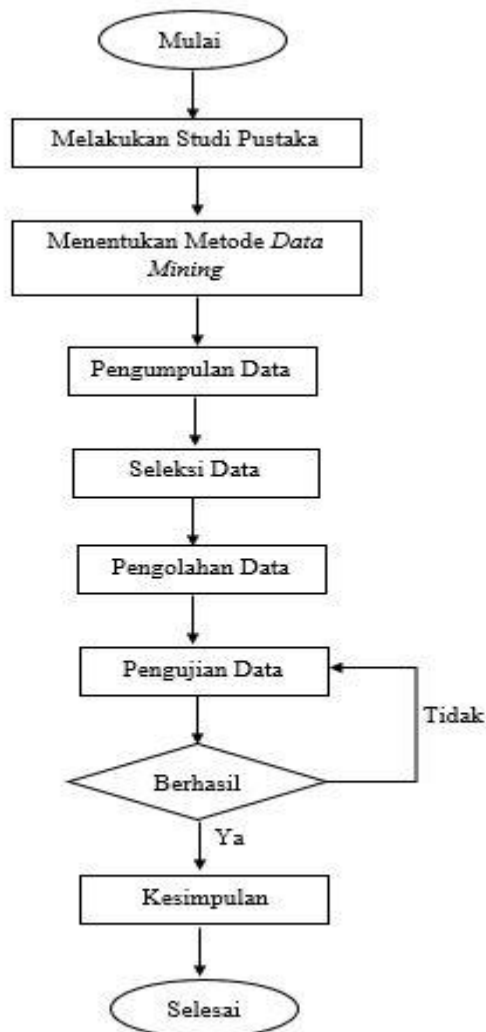
### 3.2 Lokasi Penelitian

Lokasi penelitian dilaksanakan di Biro Sistem Informasi (BSI) Universitas Muhammadiyah Yogyakarta yang beralamat di Jalan Brawijaya, Tamantirto, Kasihan, Bantul, Yogyakarta.

### 3.3. Prosedur Penelitian

Prosedur penelitian merupakan langkah-langkah yang digunakan untuk mengumpulkan data dan menjawab pertanyaan-pertanyaan dalam penelitian. Sehingga di dalam prosedur penelitian ini, penulis membahas tentang metode dan teknik pengumpulan data.

Penelitian ini menggunakan data valid dan menghasilkan kesimpulan yang mampu dibuktikan dengan pengamatan atau percobaan. Penelitian ini termasuk dalam penelitian eksperimen dengan menggunakan data jenis pekerjaan alumni Universitas Muhammadiyah Yogyakarta. Gambar 3.1 adalah diagram alir dari alur penelitian yang akan dilakukan sebagai berikut :



**Gambar 3. 1** Diagram Alir Penelitian

### 3.3.1. Melakukan Studi Pustaka

Pertama hal yang dilakukan adalah melakukan studi pustaka yang berkaitan dengan penelitian-penelitian sebelumnya tentang penggunaan algoritma C4.5, serta untuk menguatkan penelitian ini berdasarkan teori yang digunakan. Dalam kasus ini peneliti ingin melakukan analisis pada klasifikasi jenis pekerjaan alumni di Universitas Muhammadiyah Yogyakarta dengan menggunakan algoritma C4.5 sekaligus untuk mengetahui pola dan cara kerja klasifikasi dalam *data mining* tersebut.

### 3.3.2. Menentukan Metode *Data Mining*

Setelah melakukan studi pustaka tahap selanjutnya adalah menentukan metode yang sesuai untuk digunakan dalam teknik klasifikasi, setelah melakukan pengamatan dan observasi peneliti memilih metode klasifikasi menggunakan algoritma C4.5. Karena algoritma C4.5 dapat melakukan klasifikasi data dengan membentuk pohon keputusan dan dapat mengolah data dalam bentuk numerik serta kategorikal.

### 3.3.3. Melakukan Tahapan Metode Klasifikasi

Berikut tahapan metode klasifikasi menggunakan algoritma C4.5 yang akan dilakukan peneliti untuk memperoleh hasil penelitian antara lain :

#### A. Melakukan Pengumpulan Data

Pada tahap ini peneliti melakukan pengumpulan data dan ini merupakan tahapan yang penting karena dapat berpengaruh terhadap hasil penelitian, sehingga dalam mengumpulkan data harus dilakukan dengan benar. Peneliti mendapatkan data ini dari *database* BSI (Biro Sistem Informasi) Universitas Muhammadiyah Yogyakarta.

#### B. Melakukan Seleksi Data

Tahap ini dilakukan seleksi terhadap data *database*. Karena data yang diperoleh tidak semuanya digunakan, dipilih sesuai dengan atribut atau variabel yang dibutuhkan dalam penelitian dengan melakukan seleksi data sehingga menjadi *dataset*. Sebagai contoh pada penelitian ini atribut yang dipilih adalah fakultas, tahun lulus, ipk, angkatan, dan jenis pekerjaan. Pada tahap ini akan menghilangkan data yang *null*, data tidak valid, dan data yang ganda. karena data yang kosong ataupun tidak valid akan berpengaruh terhadap hasil yang diperoleh. Kemudian data akan diolah menggunakan *Microsoft Office Excel* 2016.

### C. Melakukan Pengolahan Data

Setelah semua data yang diperlukan telah dipilih, maka tahap penelitian selanjutnya adalah pengolahan data. Pada tahap ini akan dilakukan transformasi atau mengubah nilai atribut data ke dalam bentuk data yang sesuai agar data dapat diproses menggunakan algoritma C4.5. Sehingga akan diperoleh *dataset* utuh yang digunakan untuk proses ke tahapan selanjutnya.

### D. Melakukan Pengujian Data

Pada tahap pengujian hasil akan dilakukan pengujian data baik secara manual dengan algoritma C4.5 dan menggunakan *software RapidMiner*. Berikut tahapan dalam menggunakan *software RapidMiner* sebagai berikut :

- a) Tahapan pengujian menggunakan *RapidMiner* pertama membuka aplikasi *RapidMiner*, selanjutnya klik *New Process*, setelah itu akan muncul tampilan *Main Process* yang digunakan untuk melakukan sebuah proses baru.
- b) Selanjutnya klik *Import* pada *Panel Operators* dan pilih Data klik *Read Excel*, setelah itu akan muncul tampilan *Read Excel* di *Main Process*.
- c) Setelah itu untuk memasukkan *dataset Excel* yang telah dibuat klik *Import Configuration Wizard*, kemudian akan muncul 4 tahapan yang ada di dalam *Configuration Wizard* tadi, tahap pertama memilih data yang akan digunakan, tahap kedua memilih lembar kerja yang berisi *dataset*, tahap ke tiga mengisi *Annotation* jika diperlukan, dan tahap ke empat mengatur atribut sesuai dengan kebutuhan. Setelah itu klik *Finish*.
- d) Selanjutnya klik *Modeling* pada *Panel Operators* dan pilih *Classification and Regression* dapat disesuaikan dengan kebutuhan karena dalam penelitian ini akan melakukan klasifikasi maka metode ini yang dipilih, *new operator*, kemudian pilih *Tree Induction* dan klik *Decision Tree* dipilih metode tersebut karena akan menentukan hasil akhirnya berupa pohon keputusan.

- e) Setelah itu hubungan *port* pada *Read Excel* dengan *port Decision Tree* setelah semua terhubung, klik *run* untuk mengetahui hasil dari semua proses yang telah dilakukan. Jika pengujian belum menghasilkan luaran yang sesuai dengan kebutuhan, maka tahapan penelitian kembali ke pengolahan data.

Dibawah ini cara pengujian menggunakan metode algoritma C4.5 adalah sebagai berikut:

- a) Hal yang pertama dilakukan adalah menyiapkan *dataset* yang akan digunakan sebagai bahan pengujian.
- b) Setelah itu mencari nilai *entropy* dan *information gain* total yang diperoleh dari jumlah seluruh data tadi, dengan menggunakan rumus seperti dibawah ini :

$Entropy (S) = \sum_{i=0}^n -p_i * \log_2(p_i) \dots\dots(1)$  digunakan untuk menghitung nilai *entropy* dari atribut yang ada di *dataset*.

$Gain (S, A) = Entropy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S) \dots\dots(2)$  digunakan untuk menghitung nilai *gain* pada atribut yang ada di *dataset*.

$Split Info(S, A) = -\sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \dots\dots\dots(3)$  digunakan untuk menghitung nilai *split* yang ada di *dataset*.

$GainRatio (S, A) = \frac{Gain(S,A)}{SpiltInfo(S,A)} \dots\dots\dots(4)$  digunakan untuk menentukan *gain ratio* pada sebuah *dataset*.

- c) Selanjutnya adalah membuat pohon keputusan dengan melihat nilai *gain* tertinggi dari semua atribut yang ada, setelah itu nilai *gain ratio* tertinggi dari atribut tadi akan dijadikan sebagai akar dalam pohon keputusan.
- d) Kemudian membuat cabang dari masing-masing nilai atribut yang tersisa sesuai dengan jumlah nilai variabel *gain ratio* tertinggi. menggunakan rumus (1) dan (2).

- e) Tahap terakhir adalah mengulangi proses pada setiap cabang hingga hingga semua kasus pada cabang sampai memiliki kelas yang sama dan sampai tidak bisa dilakukan sebuah proses lagi.

#### **3.3.4. Membuat Kesimpulan Penelitian**

Berdasarkan hasil pengujian maka dapat ditarik kesimpulan yang mengacu pada rumusan masalah dan tujuan penelitian. Saran yang digunakan untuk mengembangkan penelitian selanjutnya serta dimasukkan untuk meningkatkan kualitas penelitian.