

BAB 3
METODE PENELITIAN

3.1 Metode Pengumpulan Data

Penelitian menggunakan data-data nasabah yang diperoleh dari sebuah perusahaan asuransi yang memiliki kevalidan. Data-data tersebut dijabarkan dalam tabel 3.1 dibawah ini :

Tabel 3.1 Metode Pengumpulan Data

No.	Data	Sumber Data	Teknik Pengumpulan Data	Keterangan	Jenis Data
1.	Data Nasabah	BPJS Ketenagakerjaan	Studi Pustaka	Data nasabah asuransi yang berisi data diri nasabah asuransi	Data Primer
2.	Data Produk Asuransi	BPJS Ketenagakerjaan	Studi Pustaka	Data jenis asuransi berisi data jenis-jenis produk asuransi jiwa.	Data Primer
3.	Data Angsuran Premi	BPJS Ketenagakerjaan	Studi Pustaka	Data berisi data angsuran premi para nasabah asuransi.	Data Primer

4.	Insurance Policy	BPJS Ketenagakerjaan	Wawancara	Berisikan data tentang kebijakan masalah asuransi.	Data Sekunder
5.	Data Perkembangan Asuransi	BPJS Ketenagakerjaan	Observasi dan Survei	Berisikan data perkembangan asuransi dan minat konsumen terhadap asuransi.	Data Sekunder

3.2 Metode Analisis

Dalam memprediksi minat konsumen terhadap produk asuransi penelitian ini menggunakan algoritma K-Nearest Neighbor dengan menggunakan Manhattan Distance. Tahapan-tahapannya yaitu sebagai berikut:

1. Pengembangan Algoritma

Pengembangan algoritma yang diterapkan dalam penelitian ini menggunakan metode metode classification K-Nearest Neighbor yang akan digunakan dalam memprediksi minat konsumen terhadap produk asuransi.

a. Knowledge Discovery in Database (KDD)

Dalam KDD terdapat 5 tahapan penting yang digunakan untuk mengidentifikasi pola tersembunyi dalam dataset yang besar.

1) *Data Selection*

Proses pemilihan (seleksi) data dari sekumpulan data yang telah diperoleh dari BPJS Ketenagakerjaan Cabang II Semarang. Data yang tidak diperlukan akan dihilangkan. Data hasil seleksi yang telah melalui proses selection dan

akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2) *Pre – processing / Cleaning*

Data yang telah melalui proses selection akan dihilangkan duplikasi data data inkonsisten data. Data yang semula memiliki 8 atribut kemudian dihilangkan 3 atribut menjadi 5 atribut. Hal ini dilakukan karena ketiga atribut tersebut tidak diperlukan dalam proses data mining.

3) *Transformation*

Dari data yang telah melalui proses cleaning kemudian dilakukan proses transformation. Dimana data tersebut disesuaikan untuk kebutuhan proses data mining. Dari 5 atribut yang telah dipilih terdapat dua atribut yang membutuhkan proses transformation. Dua atribut tersebut masih berupa *character* sehingga perlu dilakukan transformasi data dari *character* menjadi *numeric*.

4) *Data Mining*

Dari data yang telah melalui 3 tahapan awal kemudian akan dicari pola dari data tersebut dengan menggunakan metode K-Nearest Neighbor Manhattan Distance. Langkah-langkahnya adalah sebagai berikut:

- a) Menentukan parameter K (jumlah tetangga paling dekat).
- b) Menghitung kuadrat jarak (query instance) masing-masing objek terhadap data sampel yang diberikan.
- c) Kemudian mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak Euclid terkecil.
- d) Mengumpulkan kategori Y (Klasifikasi nearest neighbor).
- e) Dengan menggunakan kategori nearest neighbor yang paling mayoritas maka dapat diprediksikan nilai query instance yang telah dihitung.

5) *Interpretation / Evaluation*

Tahap ini akan diuji apakah hasil dari perhitungan yang dilakukan telah sesuai dengan fakta yang ada.

6) *Knowledge Presentation*

Merupakan tahap visualisasi dan penyajian pengetahuan dari perhitungan yang telah dilakukan sehingga dapat mengkomunikasikan hasil dari data mining. Dari perhitungan yang dilakukan kemudian divisualisasikan ke dalam program penentuan produk dengan menggunakan MATLAB sehingga hasil dari perhitungannya dapat dengan mudah dipahami.