

# Klasifikasi Calon Pemegang Polis Untuk Pemilihan Produk Asuransi Dengan Perbandingan Metode K-Nearest Neighbor Dan Metode Naive Bayes

Ari Hidayatullah<sup>1</sup>, Irfan Nurdiansyah<sup>2</sup>

[irfan.nurdiansyah@dosen.undira.ac.id](mailto:irfan.nurdiansyah@dosen.undira.ac.id)<sup>1</sup>, [ari.hidayatullah@undira.ac.id](mailto:ari.hidayatullah@undira.ac.id)<sup>2</sup>

Universitas Dian Nusantara

Teknik dan Informatika

Jl. Tanjung Duren Barat II No. 1, Jakarta

## ARTICLE INFO

Received : 1 May 2024

Received in revised : 8 May 2024

Accepted : 18 May 2024

Available online : 1 July 2024

## ABSTRACT

The insurance business within an insurance company offers insurance products owned by the insurance company. In every insurance product there is a premium payment and the premium is the income of an insurance company at the rate of the amount insured. The problem that PT BNI Life Insurance has is that there are many stops in premium payments such as policy redemptions due to errors in the benefits received or incorrect selection of the insurance product, this can reduce the achievement of targets for an insurance company. The aim of this research is to find out the best classification algorithm compared between K-Nearest Neighbor and Naive Bayes to predict the type of insurance product that customers will choose. In this research, data mining methods are applied to compare two different methods, namely the K-Nearest Neighbor method and the Naive Bayes method. The level of accuracy results for the K-Nearest Neighbor method is 80% and the Naive Bayes method is 70.53%, which means that the K-Nearest Neighbor method is the best method to apply to an insurance product classification system based on the demographics of prospective customers.

**Keywords:** Data Mining, Classification, K-Nearest Neighbor, Naive Bayes, Insurance Product Selection.

## 1. Pendahuluan

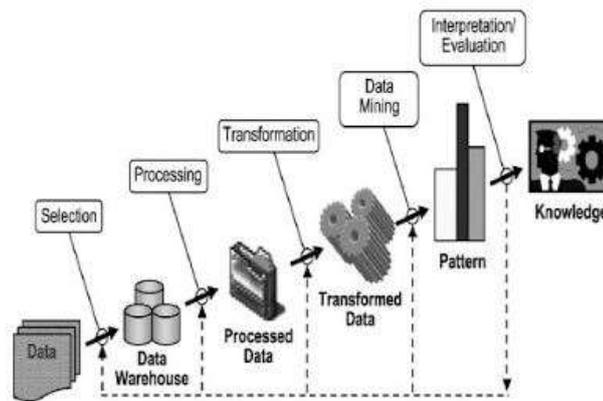
Bisnis asuransi pada perusahaan asuransi PT.BNI Life Insurance terhadap nasabah ialah menawarkan atau menjual suatu produk-produk asuransi. Permasalahan yang sering terjadi dalam perusahaan ialah kesalahan pemilihan produk asuransi pada nasabah. Oleh karena itu, diperlukan teknik data mining untuk menggali informasi yang berkaitan dengan produk-produk asuransi bagi calon nasabah. Dari beberapa teknik klasifikasi yang ada, penulis memilih teknik K-Nearest Neighbor dan Naive Bayes digunakan pada penelitian ini untuk memprediksi jenis produk asuransi yang akan dipilih oleh nasabah. karena dari berbagai literature pada penelitian perbandingan algoritma C.45 dan algoritma Naive Bayes hasil yang terbaik tersebut yaitu Naive Bayes dan pada studi literature lainnya terdapat perbandingan Naive Bayes dan KNN dan hasil tersebut lebih baik pada algoritma KNN yaitu untuk Naive Bayes accuracy sebesar 70.32% dan KNN sebesar 82.42%. Dengan adanya mengklasifikasikan produk asuransi dapat mempermudah pihak pemasaran produk asuransi dalam menawarkan produk asuransi terhadap calon pemegang polis atau nasabah. Untuk model klasifikasi produk asuransi terhadap calon pemegang polis pada pemilihan produk

asuransi diharapkan menjadi referensi perusahaan asuransi lainnya agar terhindarnya dari kesalahan pada pemilihan produk asuransi tersebut.

## 2. Metode penelitian

### 2.1 Pengertian Data Mining

Data mining adalah kombinasi secara logis antara pengetahuan data, dan analisa statistik yang dikembangkan dalam pengetahuan bisnis atau suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, tiruan dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat bagi pengetahuan yang terkait dari berbagai *database* besar (Widaningsih, 2019).



Gambar 2.1 Tahapan Data Mining

### 2.2 Metode Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Bustami, 2014).

Dasar pengukuran untuk mengukur kualitas dari penemuan teks, yaitu:

- *Precision*: tingkat ketepatan hasil klasifikasi terhadap suatu kejadian.
- *Recall*: tingkat keberhasilan mengenali suatu kejadian dari seluruh kejadian yang seharusnya dikenali.
- *F-Measure* adalah nilai yang didapatkan dari pengukuran *precision* dan *recall* antara *class* hasil *cluster* dengan *class* sebenarnya yang terdapat pada data masukan.

*Precision* dan *recall* bisa didapatkan dengan rumus sebagai berikut (Ode et al., 2019):

$$\text{Precision (i,j)} = \frac{n_{ij}}{n_j}$$

$$\text{Recall (i,j)} = \frac{n_{ij}}{n_i}$$

Rumus untuk menghitung nilai *F-Measure*:

$$F(i,j) = \frac{(b^2+1).(p(i,j).r(i,j))}{b^2.p(i,j)+r(i,j)}$$

Untuk menghitung akurasi digunakan formula :

$$\text{Akurasi} = \frac{\text{Jumlah data yang diprediksi secara benar}}{\text{Jumlah prediksi yang dilakukan}}$$

### 2.3 Metode K-Nearest Neighbor

*K-Nearest Neighbor* adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari *label class* pada *K-Nearest Neighbor*. *Euclidean Distance* berfungsi menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua obyek (Grassella et al., 2019). Rumus *Euclidean Distance* :

$$d(x_{ik}, x_{jk}) = \left( \sum_{k=1}^m (x_{ik} - x_{jk})^2 \right)$$

Keterangan :

$x_{ik}$  = Nilai x pada *training data*

$x_{jk}$  = Nilai x pada *testing data*

m = Batas jumlah banyaknya data

$d(x_{ik}, x_{jk})$  = Jarak antara *training data* dan *testing data*

### 2.4 Metode Naïve Bayes

*Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik. Klasifikasi *Naïve Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Chaira et al., 2016).

Persamaan dari *teorema Bayes* adalah :

$$P(H|X) = (P(H | X).P(H))/P(X)$$

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

$P(H|X)$  : Probabilitas hipotesis berdasarkan kondisi (posteriori probabilitas)

$P(H)$ : Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$  : Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$  : Probabilitas X

### 2.5 Evaluasi

Klasifikasi biner merupakan model statistik dan perhitungan yang membagi kumpulan data menjadi dua kelompok yaitu positif dan negatif. Tabel 2.1 berikut adalah *confusion matrix* untuk menjelaskan ukuran performansi klasifikasi.

Tabel 2.1 *Confusion Matrix*

| Aktual          | Prediksi              |                       |
|-----------------|-----------------------|-----------------------|
|                 | <i>Positive</i>       | <i>Negative</i>       |
| <i>Positive</i> | <i>True Positive</i>  | <i>False Negative</i> |
| <i>Negative</i> | <i>False Positive</i> | <i>True Negative</i>  |

Ukuran performansi termasuk ke dalam tahapan evaluasi. Beberapa ukuran performansi untuk teknik klasifikasi yaitu akurasi, error, dan *Area Under Receiver Operating Characteristics (ROC) curve (AUC)*. Akurasi adalah suatu ukuran rasio prediksi yang benar terhadap total jumlah sampel dievaluasi. *Error* adalah ukuran rasio prediksi yang salah terhadap total jumlah sampel yang dievaluasi. *AUC (the area under curve)* dihitung untuk mengukur perbedaan performansi.

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual dengan *false*

*positive rate (specificity)* sebagai garis horizontal dan *true positive rate (sensitivity)* sebagai garis vertikal. seakan-akan menggambarkan tawar-menawar antara sensitivitas (*benefit*) dan 1-spesifisitas (*cost*), yang tujuannya adalah untuk menentukan *cut off point* pada uji *diagnostic* yang bersifat kontinu.

Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok.

- a. 0.90-1.00 = Klasifikasi sangat baik
- b. 0.80-0.90 = Klasifikasi baik
- c. 0.70-0.80 = Klasifikasi cukup
- d. 0.60-0.70 = Klasifikasi buruk
- e. 0.50-0.60 = Klasifikasi salah

## 2.6 RapidMiner

Dalam pengolahan data mining umumnya digunakan *software* sebagai alat bantu. Beberapa *software* data mining diantaranya RapidMiner, weka, clementine, tanagra dan lain-lain. Menurut [www.rapidminer.com](http://www.rapidminer.com), *software* rapidminer digunakan untuk merancang aliran secara visual untuk menganalisis data *science* dan *machine learning* di dalam tim mulai dari analisis hingga pakar.

Rapidminer memiliki kemudahan dalam penggunaan, dapat mengumpulkan data dari semua sumber seperti basis data, *cloud*, dokumen, media sosial dan aplikasi bisnis. Selain itu dapat mengeksplorasi dan memvisualisasi data secara statistik. Tersedia beberapa model mesin pembelajaran dan model validasi.

## 2.7 Penelitian Terdahulu

Dari berdasarkan beberapa penelitian terkait klasifikasi calon nasabah untuk pemilihan produk yang berjudul “Analisis Dan Implementasi Perbandingan Algoritma C.45 Dengan *Naïve Bayes* Untuk Prediksi Penawaran Produk” dengan atribut Usia, Keanggotaan, Penggunaan Kartu Kredit, Pekerjaan, Penghasilan, Tanggungan, Produk menggunakan algoritma C.45 dan *Naïve Bayes* dengan akurasi 47,7% dengan C.45 dan untuk *Naïve Bayes* sebesar 57%.

Dari beberapa teknik klasifikasi yang ada, penulis memilih teknik *K-Nearest Neighbor* dan *Naïve Bayes* digunakan pada penelitian ini untuk memprediksi jenis produk asuransi yang akan dipilih oleh nasabah, karena dari berbagai literatur pada penelitian perbandingan algoritma C.45 dan algoritma *Naïve Bayes* hasil yang terbaik tersebut yaitu *Naïve bayes* dan pada studi literatur lainnya terdapat perbandingan *Naïve bayes* dan KNN dan hasil tersebut lebih baik pada algoritma KNN yaitu untuk *Naïve Bayes accuracy* sebesar 70.32% dan KNN sebesar 82.42%.

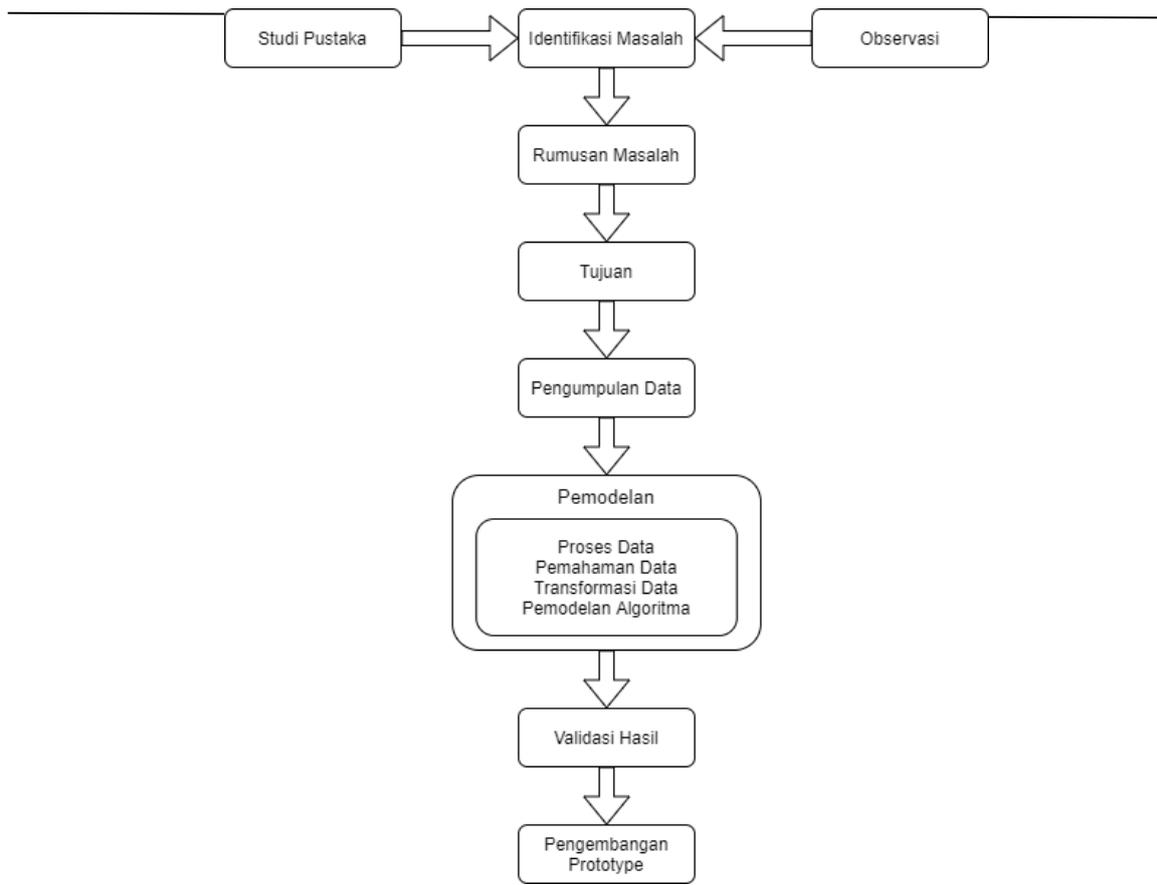
## 3. Metode Penelitian

### 3.1 Metode Penelitian

Metode penelitian ini merupakan cara ilmiah yang digunakan untuk menemukan masalah dengan tujuan tertentu. Metode yang akan digunakan ialah dengan cara mengumpulkan data yang dibutuhkan, pemrosesan data, tahapan training data, menganalisa klasifikasi data nasabah menggunakan metode *K-Nearest Neighbor* dan metode *Naïve Bayes*.

### 3.2 Tahapan Metode Penelitian

Berikut ini tahapan metode penelitian yang akan dilakukan pada penelitian ini dan dapat dilihat sebagai berikut:



Gambar 3.1 Tahapan Penelitian

**3.3 Pemodelan Sistem**

Pada penelitian ini pada memodelkan system yang dibuat yaitu bertahap dari mengumpulkan data yang akan diolah dan melakukan *pre-processing* data seperti cleansing data atau memilih atribut atau transformasi data setelah data diproses maka dilanjutkan dengan melakukan pembagian data yaitu sebagai data *training* dan *testing* pada pembagian data tersebut masing-masing akan di proses dalam algortima pada penelitian ini yaitu *K-Nearest Neighbor* dan *Naïve Bayes* dan hasil tersebut akan di implementasikan dalam bentuk system berbasis website untuk kedepannya.

**4. Hasil dan Pembahasan**

**4.1 Pengumpulan Data**

Pengumpulan data dilakukan meliputi data nasabah dalam hal ini pemegang polis, data asuransi dan data persistensi dari polis yang terdapat pada sistem asuransi jiwa individu. Periode data yang akan diambil berdasarkan tanggal terbit polis pada 1 Januari 2016 sampai dengan 31 Agustus 2023, baik untuk data inforce maupun data lapse, yang akan dijadikan data training (80%) dan data testing(20%).

Tabel 4.1 Sampel data training

| Jenis_kel | Pendapatan | Usia_masuk_a | Lama_as | Lama_pe | Jumlah_nak | Nama_cara_bayar | Status_perokok | Penyakit | Kdproduksi | Pilih produk |
|-----------|------------|--------------|---------|---------|------------|-----------------|----------------|----------|------------|--------------|
|-----------|------------|--------------|---------|---------|------------|-----------------|----------------|----------|------------|--------------|

| ami<br>n |                       | suran<br>si | urans<br>i | mbay<br>aran |   |                |   |                       |           |       |
|----------|-----------------------|-------------|------------|--------------|---|----------------|---|-----------------------|-----------|-------|
| L        | > 20<br>Jt <<br>40 Jt | 33          | 47         | 10           | 1 | Tahunan        | Y | Throat                | BO<br>M   | Ya    |
| P        | > 60<br>Jt            | 31          | 50         | 10           | 1 | Bulanan        | T | Anus                  | BO<br>M   | Tidak |
|          | < 10<br>Jt            | 23          | 58         |              | 1 |                | T | Ulna<br>and<br>radius | BLS<br>P  | Tidak |
| P        |                       |             | 58         | 10           | 1 | Bulanan        | T | Heart                 | BLM<br>PP | Ya    |
| L        | < 10<br>Jt            | 43          |            | 10           | 2 | Tahunan        | Y | Testes                | BO<br>M   | Ya    |
| L        | > 60<br>Jt            | 36          | 45         | 5            | 3 |                | T |                       |           | Tidak |
|          | < 10<br>Jt            | 33          | 48         | 11           | 1 | Semeste<br>ran | Y | Pancre<br>as          | BSW<br>P  | Tidak |
| L        | > 20<br>Jt <<br>40 Jt | 29          | 52         | 52           | 2 |                | Y | Benig<br>n<br>lesions | BLS<br>P  | Tidak |
| L        | < 10<br>Jt            | 43          |            | 10           | 2 | Tahunan        | Y | Testes                | BO<br>M   | Ya    |

Untuk data setelah proses data cleansing, data akan yang digunakan dalam penelitian ini menjadi sebesar 474 data yang terdapat pada tabel 4.2

Tabel 4.2 Setelah Proses *Data Cleansing*

| N<br>o      | Jen<br>is_<br>kel<br>ami<br>n | Pend<br>apata<br>n    | Usia<br>_mas<br>uk_a<br>suran<br>si | Lam<br>a_as<br>urans<br>i | Lam<br>a_pe<br>mbay<br>aran | Juml<br>ah_a<br>nak | Nama_c<br>ara_bay<br>ar | Statu<br>s_per<br>okok | Peny<br>akit              | Kdpr<br>oduk | Pilih<br>prod<br>uk |
|-------------|-------------------------------|-----------------------|-------------------------------------|---------------------------|-----------------------------|---------------------|-------------------------|------------------------|---------------------------|--------------|---------------------|
| 1           | L                             | > 20<br>Jt <<br>40 Jt | 33                                  | 47                        | 10                          | 1                   | tahunan                 | Y                      | throa<br>t                | BO<br>M      | Ya                  |
| 2           | P                             | > 60<br>Jt            | 31                                  | 50                        | 10                          | 1                   | bulanan                 | T                      | anus                      | BO<br>M      | Tida<br>k           |
| 3           | P                             | < 10<br>Jt            | 23                                  | 58                        | 3                           | 1                   | tahunan                 | T                      | ulna<br>and<br>radiu<br>s | BLS<br>P     | Tida<br>k           |
| 4           | P                             | > 60<br>Jt            | 48                                  | 58                        | 10                          | 1                   | bulanan                 | T                      | heart                     | BLM<br>PP    | Tida<br>k           |
| 5           | L                             | < 10<br>Jt            | 43                                  | 37                        | 10                          | 2                   | tahunan                 | Y                      | testes                    | BO<br>M      | Ya                  |
| ...         |                               |                       |                                     |                           |                             |                     |                         |                        |                           |              |                     |
| 4<br>7<br>4 | L                             | >60<br>Jt             | 25                                  | 56                        | 5                           | 1                   | bulanan                 | Y                      | diare                     | BLM<br>S     | Ya                  |

#### 4.2 Analisa dan Pemodelan Data

Data akan dipilih beberapa atribut yang akan digunakan dalam penelitian ini yang terdiri dari jenis kelamin, pendapatan, usia masuk, lama asuransi, lama pembayaran premi, jumlah anak, cara bayar asuransi, status perokok, produk asuransi, metode bayar, premi dan status polis. Pada atribut status polis atau tingkat persistensi akan dijadikan sebagai label atau variable yang untuk memprediksi tingkat persistensi pada proses data mining tersebut.

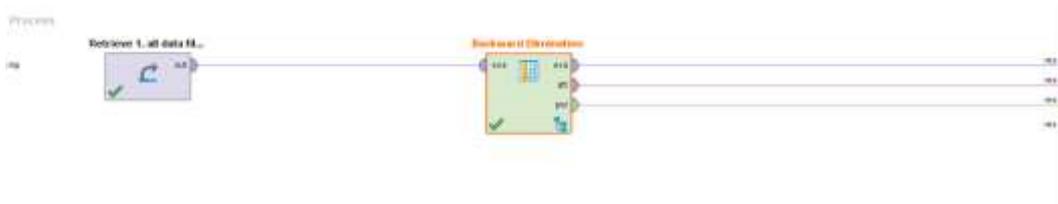
Hasil pemilihan atribut maka data-data penelitian akan diproses tahap transformasi data, tahap transformasi data ini yaitu mengelompokan data dengan kriteria tertentu, tahap ini untuk mempermudah perhitungan pada algoritma pada data mining. Pada tahap akhir pada pemodelan ialah tahap perhitungan pada setiap algoritma pada yang digunakan dalam penelitian ini yaitu algoritma *K-Nearest Neighbor* dan Algoritma *Naïve Bayes* dengan menggunakan rumus pada setiap algoritma tersebut. Atribut-atribut yang menjadi parameter seperti yang ditampilkan pada tabel 4.1.

**Tabel 4.3 Parameter**

| No | Atribut               | Nilai  |
|----|-----------------------|--|
| 1  | Jenis_Kelamin         | Laki-laki<br>Perempuan   |
| 2  | Pendapatan            | < 10 Jt<br>> 10 Jt < 20 Jt<br>>20 Jt < 40 Jt<br>> 40 Jt < 60 Jt<br>> 60 Jt |
| 3  | Usia_Masuk_Asuransi   | 1..99 Tahun  |
| 4  | Lama_Asuransi         | 1..99 Tahun  |
| 5  | Lama_Pembayaran_Premi | 1..99 Tahun  |
| 6  | Jumlah_Anak           | 1..10  |
| 7  | Nama_Cara_Bayar       | Bulanan<br>Triwulanan<br>Semesteran<br>Tahunan<br>Sekaligus                |
| 8  | Status_Perokok        | Ya<br>Tidak  |
| 9  | Penyakit              | Area Kepala<br>Area Badan<br>Area Kaki                                     |
| 10 | Kode_Produk           | BLMPP<br>BLMS<br>BSWP<br>BOM<br>BLSP                                       |
| 11 | Pilih Produk          | Ya<br>Tidak  |

Setelah tahap hasil seleksi atribut, untuk proses *feature selection* merupakan tahap pemilihan atribut yang relevan yang untuk data yang akan diuji, jika atribut tersebut tidak signifikan maka

atribut-atribut akan dihilangkan dari atribut yang sebelumnya. Pada penelitian ini menggunakan *feature selection backward elimination* yang berfungsi mengoptimalkan kinerja suatu model dengan sistem kinerja mundur. Dan untuk tahap proses *feature selection* penelitian ini menggunakan *software RapidMiner*, proses *feature selection* dengan menggunakan *software RapidMiner* terdapat pada gambar 4.1



Gambar 4.1 Proses *Feature Selection*

Setelah menjalankan proses tahap *feature selection*, maka akan menghasilkan suatu pencarian atribut yang dianggap tidak signifikan dalam proses data mining, hasil seleksi atribut dengan proses *feature selection* terdapat pada gambar 4.2.

| #Attribute          | weight |
|---------------------|--------|
| JENIS_KELAMIN       | 1      |
| PENDAPATAN          | 1      |
| USIA_MASUK_ASURANSI | 1      |
| LAMA_ASURANSI       | 1      |
| LAMA PEMBAYARAN     | 1      |
| JUMLAH ANAK         | 0      |
| NAMA_CARA_BAYAR     | 1      |
| STATUS_PEROKOK      | 1      |
| PENYAKIT            | 1      |
| KODEPRODUK          | 1      |

Gambar 4.2 Hasil *Feature Selection Backward Elimination*

Pada gambar 4.2 diatas dapat disimpulkan bahwa pada *weight* dengan nilai 0 pada atribut jumlah\_anak maka atribut tersebut dapat tidak digunakan dalam proses data mining yang variabel-variabel tersebut dianggap tidak signifikan. Pada atribut-atribut yang setelah menggunakan *feature selection* terdapat pada table 4.4.

Tabel 4.4 Atribut Setelah *Feature Selection*

| No | Atribut               |
|----|-----------------------|
| 1  | Jenis_Kelamin         |
| 2  | Pendapatan            |
| 3  | Usia_Masuk_Asuransi   |
| 4  | Lama_Asuransi         |
| 5  | Lama_Pembayaran_Premi |
| 6  | Nama_Cara_Bayar       |
| 7  | Status_Perokok        |
| 8  | Penyakit              |
| 9  | Kode_Produk           |
| 10 | Pilih_Produk          |

### 4.3 Transformasi

Pada tahap transformasi data merupakan merubah data menjadi bentuk tabel yang sudah dikelompokkan berdasarkan target yang akan di prediksi yaitu menentukan status polis pada nasabah. Setelah proses data transformasi menjadi hasil numerik, penelitian ini akan memproses perhitungan algoritma *K-Nearest Neighbor* dan algoritma *Naïve Bayes*, fungsi pada proses transformasi ini agar dapat mempermudah proses *data mining* dalam perhitungan dan data tersebut di *grouping* berdasarkan kriteria.

Tabel 4.5 Transformasi Atribut Jenis Kelamin

| Jenis_Kelamin | Transformasi |
|---------------|--------------|
| L             | 1            |
| P             | 2            |

Untuk atribut pendapatan pada calon nasabah akan dikelompokkan berdasarkan parameter pada tabel 4.6.

Tabel 4.6 Transformasi Atribut Pendapatan

| Pendapatan      | Transformasi |
|-----------------|--------------|
| < 10 Jt         | 1            |
| > 10 Jt < 20 Jt | 2            |
| > 20 Jt < 40 Jt | 3            |
| > 40 Jt < 60 Jt | 4            |
| > 60 Jt         | 5            |

Untuk atribut usia masuk asuransi pada calon nasabah akan dikelompokkan berdasarkan parameter pada tabel 4.7.

Tabel 4.7 Transformasi Atribut Usia Masuk Asuransi

| Usia Masuk Asuransi | Transformasi |
|---------------------|--------------|
| < 20 Tahun          | 1            |
| 20 – 50 Tahun       | 2            |
| > 50 Tahun          | 3            |

Untuk atribut lama asuransi pada calon nasabah akan dikelompokkan berdasarkan parameter pada tabel 4.8.

Tabel 4.8 Transformasi Lama Asuransi

| Lama Asuransi | Transformasi |
|---------------|--------------|
| < 20 Tahun    | 1            |
| 20 – 50 Tahun | 2            |
| > 50 Tahun    | 3            |

Untuk atribut lama pembayaran premi pada calon nasabah akan dikelompokkan berdasarkan parameter pada tabel 4.9.

Tabel 4.9 Transformasi Lama Pembayaran Premi

| Lama Pembayaran Premi | Transformasi |
|-----------------------|--------------|
| < 20 Tahun            | 1            |
| 20 – 50 Tahun         | 2            |
| > 50 Tahun            | 3            |

Untuk atribut jumlah anak pada calon nasabah akan dikelompokkan berdasarkan parameter pada table 4.10.

Tabel 4.10 Transformasi Jumlah Anak

| Jumlah Anak | Transformasi |
|-------------|--------------|
| 0           | 1            |
| 1 - 3       | 2            |
| > 3         | 3            |

Untuk atribut nama cara bayar pada calon nasabah akan dikelompokan berdasarkan parameter pada table 4.11.

Tabel 4.11 Transformasi Nama Cara Bayar

| Nama Cara Bayar | Transformasi |
|-----------------|--------------|
| Bulanan         | 1            |
| Triwulan        | 2            |
| Semesteran      | 3            |
| Tahunan         | 4            |
| Sekaligus       | 5            |

Untuk atribut status perokok pada calon nasabah akan dikelompokan berdasarkan parameter pada table 4.12.

Tabel 4.12 Transformasi Status Perokok

| Status Perokok | Transformasi |
|----------------|--------------|
| Ya             | 1            |
| Tidak          | 2            |

Untuk atribut penyakit pada calon nasabah akan dikelompokan berdasarkan parameter pada table 4.13.

Tabel 4.13 Transformasi Penyakit

| Penyakit    | Transformasi |
|-------------|--------------|
| Area Kepala | 1            |
| Area Badan  | 2            |
| Area Kaki   | 3            |

Untuk atribut kode produk pada calon nasabah akan dikelompokan berdasarkan parameter pada table 4.14.

Tabel 4.14 Transformasi Kode Produk

| Kode Produk | Transformasi |
|-------------|--------------|
| BLMPP       | 1            |
| BLMS        | 2            |
| BSWP        | 3            |
| BOM         | 4            |
| BLSP        | 5            |

Setelah proses data transformasi menjadi hasil numerik, penelitian ini akan memproses perhitungan metode *K-Nearest Neighbor* dan metode *Naïve Bayes*, fungsi pada proses transformasi ini agar dapat mempermudah proses *data mining* dalam perhitungan dan data tersebut di *grouping* berdasarkan kriteria, maka hasil data transformasi tersebut pada tabel 4.15.

Tabel 4.15 Sampel Data Hasil Transformasi

| Jenis_kelamin | Pendapatan | Usia_masuk_asuransi | Lama_asuransi | Lama_pembayaran | Jumlah_anak | Nama_cara_bayar | Status_perokok | Penyakit | Kdproduk | Pilih_produk |
|---------------|------------|---------------------|---------------|-----------------|-------------|-----------------|----------------|----------|----------|--------------|
| 1             | 3          | 2                   | 2             | 1               | 2           | 4               | 1              | 1        | 4        | Ya           |
| 2             | 5          | 2                   | 2             | 1               | 2           | 1               | 2              | 1        | 4        | Tidak        |
| 2             | 1          | 2                   | 3             | 1               | 2           | 4               | 2              | 1        | 5        | Tidak        |
| 2             | 5          | 2                   | 3             | 1               | 2           | 1               | 2              | 2        | 1        | Tidak        |
| 1             | 1          | 2                   | 2             | 1               | 2           | 4               | 1              | 3        | 4        | Ya           |
| 1             | 5          | 2                   | 2             | 1               | 2           | 3               | 2              | 2        | 4        | Tidak        |
| 2             | 1          | 2                   | 2             | 1               | 2           | 3               | 1              | 3        | 3        | Tidak        |
| 1             | 3          | 2                   | 3             | 3               | 2           | 4               | 1              | 1        | 5        | Tidak        |
| 1             | 5          | 2                   | 3             | 1               | 2           | 1               | 1              | 2        | 3        | Ya           |
| 2             | 4          | 2                   | 2             | 1               | 2           | 4               | 2              | 1        | 4        | Tidak        |
| 1             | 1          | 2                   | 2             | 1               | 2           | 1               | 2              | 1        | 3        | Ya           |
| 2             | 4          | 2                   | 2             | 2               | 3           | 4               | 2              | 1        | 5        | Tidak        |
| 2             | 1          | 2                   | 2             | 1               | 2           | 1               | 1              | 3        | 4        | Tidak        |
| 1             | 3          | 2                   | 2             | 1               | 2           | 4               | 1              | 2        | 3        | Ya           |

#### 4.4 Pemodelan Data Testing dan Data Training

Setelah proses transformasi data yang terdapat 474 record yang digunakan untuk *data training* dan *data testing*. *Data training* adalah data yang digunakan untuk di-mining yang telah melewati pada tahap-tahap sebelumnya. Sedangkan *data testing* adalah data yang akan digunakan untuk menguji rule klasifikasi yang diperoleh dari data training. Pada penelitian ini menggunakan 80:20 yang terdapat 345 data yang tidak jadi memilih produk asuransi dan 129 data yang jadi memilih produk asuransi tersebut. Dalam penelitian ini penulis menguji data pada *selection* dan *feature selection backward elimination* pada data transformasi dan belum di transformasikan, data tersebut diuji dengan aplikasi rapid miner dengan menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*. Berikut ini presentase pada pembagian data training dan data testing pada tabel 4.16.

Tabel 4.16 Presentase Data Training dan Data Testing

| Pembagian            | Presentase | Total Data |
|----------------------|------------|------------|
| <i>Data Training</i> | 80%        | 380        |
| <i>Data Testing</i>  | 20%        | 94         |
| Total                | 100%       | 474        |

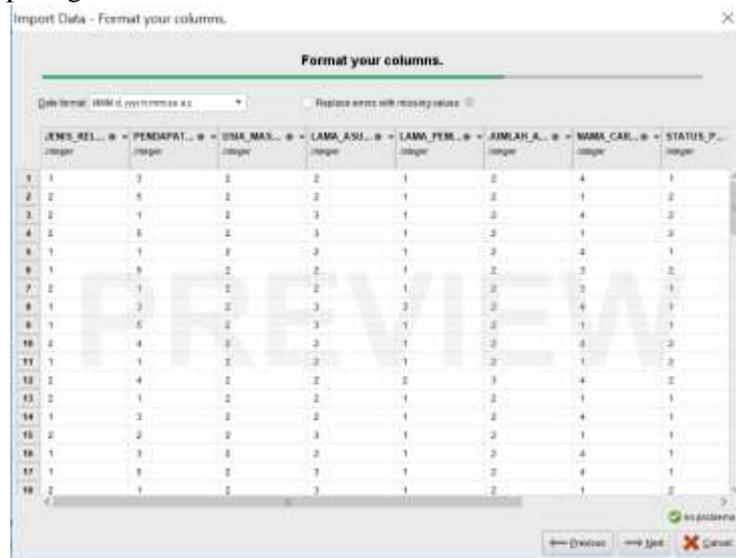
Berdasarkan pada tabel 4.16 dapat diketahui bahwa *presentase* 80:20 yaitu 80% data untuk *training* dan 20% data untuk *testing*, hal ini dikarenakan algoritma pembelajaran dapat lebih terlatih ketika model didapatkan dari hasil pelatihan diujikan dapat memperoleh hasil yang lebih akurat.

#### 4.5 Pemodelan Dengan Algoritma *K-Nearest Neighbor* dan *Naïve Bayes*

Pada pengujian data mining yang digunakan pada penelitian ini menggunakan *tools Rapid Miner*, pada tahapan proses pengujian data dengan *Rapid Miner* terdiri dari beberapa tahap:

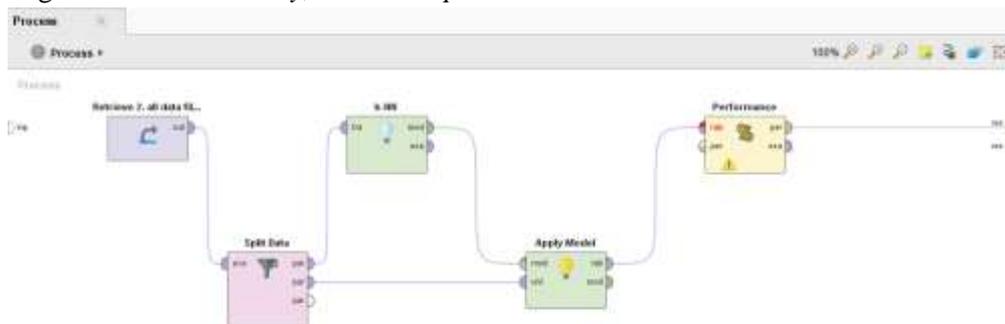
1. *Import Data* merupakan proses tahap memasukan data dari excel ke dalam *Rapid Miner*, data yang di akan dimasukan kedalam *Rapid Miner* tersebut hasil dari transformasi data yang di proses sebelumnya.

2. Proses pemilihan data pada format excel yang telah di proses transformasi pada proses sebelumnya yang akan diproses *data mining* dalam *Rapid Miner*,
3. Setelah memilih data maka selanjutnya memilih tipe data yang akan digunakan dalam proses *data mining*, pada *Rapid Miner* terdiri dari *polynomial*, *binominal*, *real*, *integer*, *date\_time*, *date*, *time* seperti gambar 4.5.



Gambar 4.3 Memilih Tipe Data

4. Selanjutnya pengujian metode pada data yang telah diimport ke *Rapid Miner*, pada penelitian membandingkan metode *K-Nearest Neighbor* dan metode *Naïve Bayes*. Pada gambar 4.6 proses metode *K-Neares Neighbor* dengan cara meletakkan data yang telah di import dan menggunakan *split data tools* dan pada *split data* membagi menjadi 80:20 dan mencari metode *K-Nearest Neighbor* dan menggunakan k-9, memilih operator *apply model* untuk menghubungkan antara data *training* dan data *testing* pada operator *performance* untuk mengetahui hasil *accuracy*, *recall* dan *precision*.

Gambar 4.4 Proses Data Dengan Metode *K-Nearest Neighbor*

5. Hasil pengujian merupakan proses melihat hasil dari proses pengujian data dari metode *K-Neares Neighbor* dengan melihat di *result history*. Pada tab *performance vector* terdapat hasil *accuracy*, *precision* dan *recall*, untuk melihat hasil tersebut terdapat pada gambar 4.7.

data tesing k-nn results.Process result  
 Completed: Jan 30, 2021 8:16:53 AM (execution time: 0 s)

```

Performance Vector (Performance)
Result not stored in repository.

PerformanceVector:
accuracy: 80.00%
ConfusionMatrix:
True:  Ya      Tidak
Ya:    3       1
Tidak: 18      73
AUC (optimistic): 0.712 (positive class: Tidak)
AUC: 0.604 (positive class: Tidak)
AUC (pessimistic): 0.495 (positive class: Tidak)
precision: 80.22% (positive class: Tidak)
ConfusionMatrix:
True:  Ya      Tidak
Ya:    3       1
Tidak: 18      73
    
```

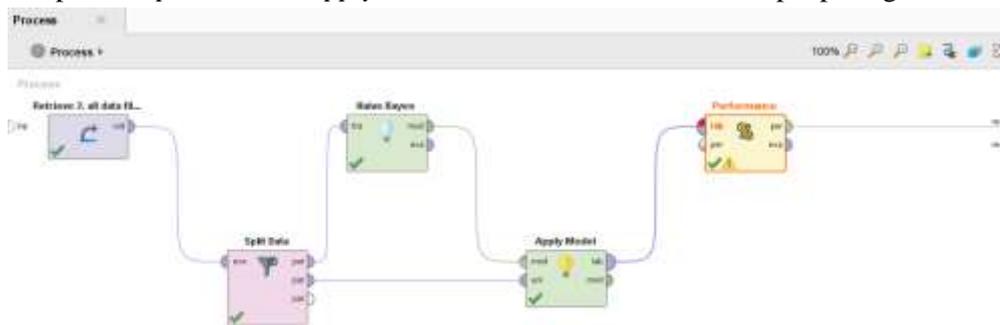
Gambar 4.5 Hasil Result Metode *K-Neares Neighbor*

6. Dan untuk hasil pada accuracy pengujian data dengan metode *K-Neares Neighbor* terdapat pada tab *performance vector*, untuk melihat hasil tersebut terdapat pada gambar 4.8.

|            | Sub To | Sub Tidak | total precision |
|------------|--------|-----------|-----------------|
| pred Ya    | 3      | 1         | 75.00%          |
| pred Tidak | 18     | 73        | 80.22%          |
| data real  | 14.29% | 85.71%    |                 |

Gambar 4.6 Hasil Accuracy Metode *K-Neares Neighbor*

7. Selanjutnya pengujian metode *Naïve Bayes* dengan data yang sudah diimport ke *Rapid Miner*. Untuk proses ini mencari operator metode *Naïve Bayes* lalu menghubungkan kembali dari operator *split* data dan *apply* model tersebut. Untuk hasil terdapat pada gambar 4.9



Gambar 4.7 Proses Data Metode *Naïve Bayes*

8. Untuk hasil pengujian metode *Naïve Bayes* dari proses pengujian data maka dapat dilihat di hasil *result history*, untuk hasil tersebut terdapat pada gambar 4.10.

data tesing naive bayes (1 results. Process results)  
Completed: Jan 30, 2021 8:23:14 AM (execution time: 0 s)

```

Performance Vector (Performance)
Result not stored in repository.

PerformanceVector:
accuracy: 70.53%
ConfusionMatrix:
True:  Ya      Tidak
Ya:    6       13
Tidak: 15      61
precision: 80.26% (positive class: Tidak)
ConfusionMatrix:
True:  Ya      Tidak
Ya:    6       13
Tidak: 15      61
recall: 82.43% (positive class: Tidak)
ConfusionMatrix:
True:  Ya      Tidak

```

Gambar 4.8 Hasil Result History Metode *Naïve Bayes*

9. Dan untuk *accuracy* pada metode *Naïve Bayes* terdapat pada tab *performance vector* untuk melihat hasil *accuracy* metode *Naïve Bayes* terdapat pada gambar 4.11.

|              | pred Ya | pred Tidak | class precision |
|--------------|---------|------------|-----------------|
| true Ya      | 6       | 13         | 31.58%          |
| true Tidak   | 15      | 61         | 80.26%          |
| class recall | 28.57%  | 82.43%     |                 |

Gambar 4.9 Hasil *Accuracy* Pada Metode *Naïve Bayes*

#### 4.6 Evaluasi dan Validasi

Pengujian *Confusion Matrix* untuk dataset yang diolah menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes* yaitu 80% data *training* dan 20% data *testing* yang dapat dilihat pada tabel

Tabel 4.17 *Confusion Matrix K-Neares Neighbor* dan *Naïve Bayes*

| Metode                    | Accuracy | Precision | Recall | F - Measure |
|---------------------------|----------|-----------|--------|-------------|
| <i>K-Nearest Neighbor</i> | 80%      | 80.22%    | 96.65% | 0.388       |
| <i>Naïve Bayes</i>        | 70.53%   | 80.26%    | 82.43% | 0.2267      |

## 5. Kesimpulan dan Saran

### 5.1 Kesimpulan

Berdasarkan pembahasan yang telah dijelaskan pada sebelumnya, maka dalam penelitian pemilihan produk asuransi dari berdasarkan demografi dengan menggunakan metode *K-Nearest Neighbor* dan metode *Naive Bayes* dapat disimpulkan untuk metode *K-Nearest Neighbor* menghasilkan *accuracy* sebesar 80% dan *precision* sebesar 80.22% dan untuk *recall* sebesar 96.65%, sedangkan untuk metode *Naive Bayes* untuk *accuracy* sebesar 70.53% dan *precision* sebesar 80.26% dan untuk *recall* sebesar 82.43% yang artinya untuk metode *K-Nearest Neighbor* lebih baik dari metode *Naive Bayes*. Sehingga pada penelitian ini akan menggunakan metode *K-Nearest Neighbor* untuk melakukan penerapan pada sistem untuk pemasaran produk asuransi tersebut terhadap calon nasabah. Dan diharapkan dengan adanya mengklasifikasikan produk asuransi dapat mempermudah pihak pemasaran produk asuransi dalam menawarkan

---

produk asuransi terhadap calon pemegang polis atau nasabah. Untuk model klasifikasi produk asuransi terhadap calon pemegang polis pada pemilihan produk asuransi diharapkan menjadi referensi perusahaan asuransi lainnya agar terhindarnya dari kesalahan pada pemilihan produk asuransi tersebut.

## DAFTAR PUSTAKA

- [1] Widaningsih, S. (2019) 'Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm', *Jurnal Tekno Insentif*, 13(1), pp. 16–25.
- [2] Bustami (2014) 'Penerapan Algoritma Naive Bayes Untuk Nasabah Asuransi', *Jurnal Informatika*, 8(1), pp. 884–898.
- [3] Ode, W., Kadir, N. and Pramono, B. S. (2019) 'Penerapan Data Mining Dengan Metode Knearest Neighbor (Knn) Untuk Mengelompokan Minat Konsumen Asuransi (PT. JASARAHARJA PUTERA)', 5(1), pp. 97–104.
- [4] Chaira, L. M. et al. (2016) 'Pemilihan Jenis Asuransi Berdasarkan Demografi Calon Pemegang Polis Dengan Metode Naïve Bayes Classifier', *Limits: Journal of Mathematics and Its Applications*, 13(2), p. 11.
- [5] Betrisandi (2017) 'Klasifikasi Nasabah Asuransi Jiwa Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination', 9(1), pp. 96–101.
- [6] Grassella, Purnamasari, I. and Amijaya, F. D. T. (2019) 'Klasifikasi Status Pembayaran Premi Menggunakan Algoritma Neighbor Weighted K-Nearest Neighbor (Nwknn) (Studi Kasus: Pt. Bumiputera Kota Samarinda)', *VARIANCE: Journal of Statistics and Its Applications*, 1(2), pp. 56–63.
- [7] Hidayatullah, Ari, E. Mudiawati dan M. Syafrullah, 2021, "Penerapan Data Mining untuk Memprediksi Minat Nasabah Terhadap Produk Asuransi Meninggal Dunia dengan Metode Naive Bayes (Studi Kasus : PT. BNI Life Insurance)", *Jurnal Teknologi Informasi*, vol. 16, no 2, pp. 103-115.
- [8] Putri, H. Y. (2018) 'Komparasi Data Mining Untuk Memprediksi Minat Klien Dalam Memilih Produk Asuransi Pendidikan ( Studi pada PT . AJB Bumiputera 1912 Karanganyar )'.
- [9] Khoerunnisa, A., Irawan, B. and Rumani, M. R. (2016) 'Analisis dan implementasi perbandingan algoritma c.45 dengan naïve bayes untuk prediksi penawaran produk', *E-Proceeding of Engineering*, 3(3), pp. 5029–5035.