

**PENERAPAN *K-MEANS CLUSTERING* UNTUK  
MEMPREDIKSI MINAT NASABAH ASURANSI DI  
KOTA BATAM**

**SKRIPSI**



**Oleh:  
Sitinurdiani  
151510106**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
2019**

**PENERAPAN *K-MEANS CLUSTERING* UNTUK  
MEMPREDIKSI MINAT NASABAH ASURANSI DI  
KOTA BATAM**

**SKRIPSI**

**Untuk memenuhi salah satu syarat  
guna memperoleh gelar Sarjana**



**Oleh:  
Sitinurdiani  
151510106**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN KOMPUTER  
UNIVERSITAS PUTERA BATAM  
2019**

## PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, dan/atau magister), baik di Universitas Putera Batam maupun di perguruan tinggi lain.
2. Skripsi ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan pembimbing.
3. Dalam skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku diperguruan tinggi.

Batam, 02 Februari 2019  
Yang membuat pernyataan,

Sitinurdiani  
151510106

**PENERAPAN *K-MEANS CLUSTERING* UNTUK  
MEMPREDIKSI MINAT NASABAH ASURANSI DI  
KOTA BATAM**

**Oleh  
Sitinurdiani  
151510106**

**SKRIPSI  
Untuk memenuhi salah satu syarat  
guna memperoleh gelar Sarjana**

**Telah disetujui oleh Pembimbing pada tanggal  
seperti tertera dibawah ini**

**Batam, 02 Februari 2019**

**Andi Supriadi Chan, S.Kom., M.Kom.  
Pembimbing**

## KATA PENGANTAR

Segala puji bagi Allah Yang Maha Esa yang telah melimpahkan segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan tugas akhir yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada Program Studi Sistem Informasi Universitas Putera Batam.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati.

Dengan segala keterbatasan, penulis menyadari bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. Ibu Dr.Nur Elfi Husda, S.Kom., M.SI., selaku Rektor Universitas Putera Batam.
2. Ketua Program Studi Sistem Informasi Universitas Putera Batam, Bapak Muhammad Rasid Ridho, S.Kom., M.SI.
3. Bapak Andi Supriadi Chan, S.Kom., M.Kom. selaku pembimbing Skripsi pada Program Studi Sistem Informasi Universitas Putera Batam.
4. Dosen dan seluruh Staff Universitas Putera Batam.
5. Seluruh narasumber yang telah rela meluangkan banyak waktunya untuk mendukung penelitian ini.
6. Keluarga yang selalu memberikan doa dan motivasi yang baik.
7. Rekan-rekan mahasiswa/i Universitas Putera Batam yang turut memberikan doa dan dukungannya

8. Mitra kerja yang selalu memberikan masukan yang berguna untuk penelitian ini
9. Serta pihak-pihak lain yang tidak dapat disebutkan satu per satu.

Semoga Allah membalas kebaikan dan selalu mencurahkan hidayah serta taufikNya, Amin.

Batam, Februari 2019

Penulis

## ABSTRAK

PT. Asuransi Wahana Tata Batam Office merupakan salah satu perusahaan yang bergerak dalam bidang perasuransian. Persaingan yang terjadi dalam bisnis asuransi mengharuskan perusahaan untuk selalu memikirkan strategi dan terobosan yang dapat menjamin kelangsungan bisnis asuransi yang mereka jalankan. PT. Asuransi Wahana Tata Batam Office harus mampu bersaing dengan perusahaan lain untuk tetap mempertahankan nasabah yang sudah ada dan menarik perhatian calon nasabah baru. Data bisnis dalam jumlah yang besar merupakan salah satu aset berharga yang dimiliki oleh perusahaan. Jumlah data yang besar yang dimiliki oleh perusahaan dapat dimanfaatkan untuk menemukan beragam informasi yang tersembunyi dari data-data tersebut. Data mining adalah istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi didalam kumpulan data yang disimpan di basis data. Penelitian ini bertujuan untuk melakukan pengelompokan data nasabah. *Clustering* merupakan proses pengelompokan data menjadi beberapa kelompok, sehingga objek didalam suatu kelompok memiliki banyak kesamaan. Algoritma yang digunakan untuk pembentukan *cluster* adalah algoritma *k-means*. *K-means* adalah salah satu metode data *non-hierarchical clustering* yang dapat mengelompokkan data nasabah kedalam beberapa *cluster*/kelompok berdasarkan kemiripan dari data tersebut. Penerapan data mining dalam penelitian ini menggunakan tahapan KDD (*Knowledge Discovery In Database*) yang terdiri dari *Data Selection*, *Data Cleaning*, *Data Transformation*, *Data Mining*, *Interpretation/Evaluation*. Implementasi menggunakan software Rapid Miner yang dapat membantu menemukan nilai yang akurat. Atribut yang digunakan adalah pekerjaan, pendapatan dan alamat nasabah. *Cluster* nasabah terbentuk menjadi 3 kelompok, dengan cluster pertama 52 *items*, cluster kedua 30 *items* dan cluster ketiga 17 *items*. Hasil dari penelitian didapatkan bahwa calon nasabah yang bekerja dibidang wiraswasta, dengan pendapatan rata-rata 3.555.945, memiliki minat yang cukup tinggi untuk menjadi nasabah asuransi dengan jenis asuransi yang dipilih asuransi kecelakaan, khususnya didaerah Batam Center. Sehingga dengan adanya penelitian ini dapat membantu pihak marketing untuk menentukan strategi baru dalam mempromosikan beberapa produk asuransi yang ada di PT. Asuransi Wahana Tata Batam Office kepada masyarakat.

Kata kunci: *Data Mining*, *Clustering*, *K-means*, *KDD*, Rapid Miner

## ***ABSTRACT***

*PT. Asuransi Wahana Tata Batam Office is one of the companies engaged in insurance. Competition that occurs in the insurance business requires companies to always think of strategies and breakthroughs that can guarantee the continuity of the insurance business they run. PT. Asuransi Wahana Tata Batam Office must be able to compete with other companies to maintain existing customers and attract the attention of new prospective customers. Large amounts of business data is one of the valuable assets owned by the company. The large amount of data that is owned by the company can be used to find a variety of information hidden from these data. Data mining is a term used to find hidden knowledge in a collection of data stored in a database. This study aims to group customer data. Clustering is the process of grouping data into several groups, so that objects in a group have many similarities. The algorithm used for cluster formation is the k-means algorithm. K-means is one of the non-hierarchical clustering data methods that can group customer data into several clusters / groups based on the similarity of the data. The application of data mining in this study uses the stages of KDD (Knowledge Discovery In Database) which consists of Data Selection, Data Cleaning, Data Transformation, Data Mining, Interpretation / Evaluation. Implementation using Rapid Miner software that can help find accurate values. The attributes used are the work, income and address of the customer. The customer cluster is formed into 3 groups, with the first cluster 52 items, the second cluster 30 items and the third cluster 17 items. The results of this study can be used to help determine a new strategy in promoting several insurance products at PT. Asuransi Wahana Tata Batam Office to the public.*

*Keyword: Data Mining, Clustering, K-means, KDD, Rapid Miner*



# DAFTAR ISI

<b>HALAMAN SAMPEL DEPAN</b>	
<b>HALAMAN JUDUL</b>	
<b>LEMBAR PERNYATAAN .....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>ABSTRAK .....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR TABEL .....</b>	<b>ix</b>
<b>DAFTAR GAMBAR.....</b>	<b>x</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Identifikasi Masalah .....	3
1.3. Batasan Masalah.....	4
1.4. Rumusan Masalah .....	5
1.5. Tujuan Penelitian.....	5
1.6. Manfaat Penelitian.....	6
1.6.1. Aspek Teoritis.....	6
1.6.2. Aspek Praktis .....	6
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>8</b>
2.1. Landasan Teori .....	8
2.1.1. Pengertian <i>Data Mining</i> .....	8
2.1.2. Tahapan <i>Data Mining</i> .....	9
2.1.3. Pengelompokkan <i>Data Mining</i> .....	10
2.1.4. <i>Clustering</i> .....	12
2.1.5. <i>K-means Clustering</i> .....	13
2.1.6. Pengertian Nasabah.....	15
2.1.7. Pengertian Asuransi .....	15
2.1.8. Rapid Miner .....	17

2.2.	Penelitian Terdahulu.....	17
2.3.	Kerangka Pemikiran .....	20
<b>BAB III METODE PENELITIAN .....</b>		<b>22</b>
3.1.	Desain Penelitian .....	22
3.2.	Populasi dan Sampel .....	25
3.2.1.	Populasi.....	25
3.2.2.	Sampel .....	25
3.3.	Teknik Pengumpulan Data .....	26
3.3.1.	Metode Pengumpulan Data.....	26
3.3.1.	Data Penelitian .....	27
3.4.	Metode Analisis Data .....	27
3.5.	Lokasi dan Jadwal Penelitian .....	28
3.5.1.	Lokasi Penelitian.....	28
3.5.2.	Waktu Penelitian.....	28
<b>BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....</b>		<b>30</b>
4.1.	Data Mining.....	30
4.2.	Analisis Data Mining.....	30
4.2.1.	<i>Data Selection</i> .....	30
4.2.2.	<i>Pre-processing (Data Cleaning)</i> .....	40
4.2.3.	<i>Transformation</i> .....	44
4.2.4.	Pengolahan Data .....	46
4.2.5.	Implementasi Software Rapid Miner .....	49
4.3.	Hasil dan Pembahasan.....	51
4.3.1.	Strategi Promosi.....	54
<b>BAB V SIMPULAN DAN SARAN.....</b>		<b>56</b>
5.1.	Simpulan.....	56
5.2.	Saran.....	57
<b>DAFTAR PUSTAKA .....</b>		<b>58</b>
<b>DAFTAR RIWAYAT HIDUP</b>		
<b>SURAT KETERANGAN PENELITIAN</b>		
<b>LAMPIRAN</b>		

## DAFTAR TABEL

<b>Tabel 2.1</b> Penelitian Terdahulu.....	18
<b>Tabel 3.1</b> Jadwal Penelitian.....	29
<b>Tabel 4.1</b> Data nasabah sebelum proses data selection .....	31
<b>Tabel 4.2</b> Data nasabah setelah proses data selection .....	36
<b>Tabel 4.3</b> Data nasabah setelah proses data cleaning .....	40
<b>Tabel 4.4</b> Inisialisasi data pekerjaan.....	45
<b>Tabel 4.5</b> Inisialisasi data alamat.....	46
<b>Tabel 4.6</b> Titik pusat awal setiap cluster .....	47
<b>Tabel 4.7</b> Titik pusat cluster setelah iterasi 1 .....	49
<b>Tabel 4.8</b> Titik pusat cluster pada iterasi 4.....	49
<b>Tabel 4.9</b> Hasil dari <i>cluster</i> 1 .....	52
<b>Tabel 4.10</b> Hasil dari cluster 2.....	53
<b>Tabel 4.11</b> Hasil dari cluster 3.....	54

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Tahapan Data Mining .....	9
<b>Gambar 2.2</b> Kerangka Pemikiran .....	21
<b>Gambar 3.3</b> Desain Penelitian .....	22
<b>Gambar 4.4</b> Konfigurasi Rapid Miner .....	50
<b>Gambar 4.5</b> Cluster Model .....	50
<b>Gambar 4.6</b> Grafik hubungan titik centroid dari tiap cluster.....	51
<b>Gambar 4.7</b> Titik pusat cluster pada Rapid Miner .....	51

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Setiap manusia tidak akan pernah lepas dari suatu masalah dan resiko. Dalam hidupnya, manusia akan selalu dihadapkan pada peristiwa yang tidak terduga yang akan terjadi, yang nantinya dapat menimbulkan kerugian bagi manusia itu sendiri. Contoh dari kejadian yang tidak terduga adalah kecelakaan, bencana alam, ataupun kematian. Mereka tidak akan pernah bisa menghindar dari resiko tersebut. Resiko merupakan kemungkinan terjadinya suatu kerugian yang tidak terduga dan yang tidak diinginkan. Untuk mengurangi kerugian dari kejadian-kejadian tidak terduga itu, seseorang harus pandai menyiasatinya. Salah satu langkah yang paling tepat untuk mengatasi permasalahan ini adalah dengan berinvestasi pada asuransi.

Asuransi atau pertanggungan merupakan sesuatu yang tidak asing lagi bagi masyarakat Indonesia. PT. Asuransi Wahana Tata Batam Office merupakan salah satu perusahaan yang bergerak dalam bidang perasuransian yang memiliki cabang di beberapa kota besar di Indonesia, salah satunya terletak di kota Batam. Bisnis asuransi yang dijalankan oleh PT. Asuransi Wahana Tata Batam Office dibagi menjadi dua jenis yaitu, bisnis asuransi korporasi dan bisnis asuransi ritel. Untuk produk yang ditawarkan dalam bisnis asuransi korporasi terdiri dari asuransi rekayasa *engineering*, asuransi *marine cargo* dan asuransi *custom bond*. Bisnis ritel

terdiri dari asuransi mobil, asuransi rumah serta asuransi kecelakaan (<http://www.aswata.co.id>).

PT. Asuransi Wahana Tata Batam Office bukan satu-satunya perusahaan asuransi yang ada di Batam. Persaingan yang terjadi dalam bisnis asuransi mengharuskan perusahaan untuk selalu memikirkan strategi dan terobosan yang dapat menjamin kelangsungan bisnis asuransi yang mereka jalankan. PT. Asuransi Wahana Tata Batam Office harus mampu bersaing dengan perusahaan lain untuk tetap mempertahankan nasabah yang sudah ada dan menarik perhatian calon nasabah baru. Tetapi, bagian *marketing* PT. Asuransi Wahana Tata Batam Office saat ini merasa kesulitan dalam memasarkan beberapa produk asuransi kepada calon nasabah baru. Strategi yang diterapkan bagian *marketing* perusahaan dalam melakukan pemasaran untuk mencari calon nasabah baru belum maksimal dan sesuai dengan yang diharapkan. Untuk dapat melakukan promosi yang lebih efektif dan efisien, maka perusahaan dapat memanfaatkan data nasabah terdahulu yang tersimpan di basis data.

Data bisnis dalam jumlah yang besar merupakan salah satu aset berharga yang dimiliki oleh perusahaan. Perusahaan telah mengumpulkan data nasabah asuransi sekian tahun lamanya. Kumpulan data tersebut disimpan dan dibiarkan menumpuk begitu saja di basis data, yang seolah-olah data tersebut hanya berguna sebagai arsip dan tidak ada manfaat lainnya. Jumlah data yang besar yang dimiliki oleh perusahaan dapat dimanfaatkan untuk menemukan beragam informasi yang tersembunyi dari data-data tersebut. Hal ini melahirkan suatu kebutuhan terhadap teknologi yang dapat memanfaatkannya dalam menggali pengetahuan-pengetahuan

baru, yang dapat membantu menentukan dan menerapkan strategi bisnis baru yang akan diambil oleh perusahaan. Sebuah teknologi baru yang dapat dimanfaatkan untuk menggali dan mengekstrasi pengetahuan dari gudang data adalah *data mining*. *Data mining* atau penambangan data adalah teknik yang relatif cepat dan mudah untuk menemukan pengetahuan, pola dan relasi antar data secara otomatis. Salah satu teknik dalam pengolahan data mining adalah dengan menggunakan metode *k-means clustering*. Dengan menggunakan metode ini, data yang telah didapatkan dapat dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan dari data-data tersebut. Dengan adanya pengelompokan data seperti ini, diharapkan bagian *marketing* PT. Asuransi Wahana Tata Batam Office dapat membuat keputusan dan melakukan pemasaran produk asuransi dengan strategi yang tepat untuk mendapatkan calon nasabah baru.

Terkait dengan uraian yang telah dipaparkan diatas, maka penulis tertarik untuk melakukan penelitian dengan judul “PENERAPAN *K-MEANS CLUSTERING* UNTUK MEMPREDIKSI MINAT NASABAH ASURANSI DI KOTA BATAM”.

## **1.2. Identifikasi Masalah**

Berdasarkan latar belakang diatas, maka penulis mencoba mengidentifikasi masalah yang ada yaitu:

1. Bagian marketing PT. Asuransi Wahana Tata Batam Office mengalami kesulitan dalam mempromosikan beberapa produk asuransi yang mereka miliki.
2. Strategi pemasaran yang dilakukan oleh perusahaan belum maksimal, sehingga menyulitkan pihak perusahaan untuk mendapatkan calon nasabah baru.
3. Data bisnis dalam jumlah yang besar yang dimiliki oleh perusahaan asuransi belum dimanfaatkan dengan optimal.

### **1.3. Batasan Masalah**

Agar tercapai tujuan sesuai dengan apa yang diharapkan, maka diperlukan suatu batasan masalah yang membatasi masalah-masalah yang akan dicoba untuk mendapatkan solusinya. Adapun batasan masalah dari topik penelitian ini yaitu:

1. Penelitian dilakukan pada PT. Asuransi Wahana Tata Batam Office
2. Jenis produk asuransi yang dijadikan penelitian adalah asuransi ritel
3. Teknik *data mining* yang digunakan adalah teknik klasterisasi menggunakan algoritma *k-means*
4. Pengujian *data mining* menggunakan *software data mining* Rapid Miner



#### **1.4. Rumusan Masalah**

Berdasarkan latar belakang masalah diatas, maka dapat dirumuskan permasalahan dalam penelitian ini yaitu:

1. Bagaimana cara menerapkan algoritma *k-means clustering* dalam memprediksi minat nasabah asuransi?
2. Bagaimana implementasi hasil *k-means clustering* dalam menentukan strategi baru untuk memprediksi minat nasabah asuransi?

#### **1.5. Tujuan Penelitian**

Setelah merumuskan masalah, maka penulis menguraikan tujuan dari penulisan penelitian ini yaitu:

1. Untuk mengetahui bagaimana cara penerapan algoritma *k-means clustering* dalam memprediksi minat nasabah asuransi
2. Untuk mengetahui bagaimana implementasi hasil *k-means clustering* dalam menentukan strategi baru untuk memprediksi minat nasabah asuransi

## 1.6. Manfaat Penelitian

### 1.6.1. Aspek Teoritis

Secara teoritis, adapun hasil dari penelitian ini diharapkan dapat digunakan sebagai masukan atau referensi dalam penerapan teknik klusterisasi menggunakan metode *k-means clustering* yang dapat diterapkan pada berbagai bidang termasuk bidang bisnis.

### 1.6.2. Aspek Praktis

Manfaat praktis dari penelitian ini adalah sebagai berikut:

1. Bagi penulis
  - a. Memberikan kesempatan bagi penulis mempraktekkan ilmu secara langsung
  - b. Sebagai klarifikasi untuk mendapatkan gelar sarjana
2. Bagi instansi terkait

Penelitian penerapan *data mining* untuk memprediksi minat nasabah asuransi pada PT. Asuransi Wahana Tata Batam Office ini dapat membantu menyediakan pengetahuan dan informasi yang mendukung untuk pengambilan keputusan yang tepat dalam menentukan calon nasabah baru dan strategi pemasaran yang dilakukan berdasarkan data nasabah terdahulu.

### 3. Bagi akademis

Sebagai salah satu referensi bagi peneliti-peneliti khususnya bagi mahasiswa-mahasiswi di Universitas Putera Batam, terutama yang melakukan penelitian dengan memanfaatkan teknologi data mining.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Landasan Teori**

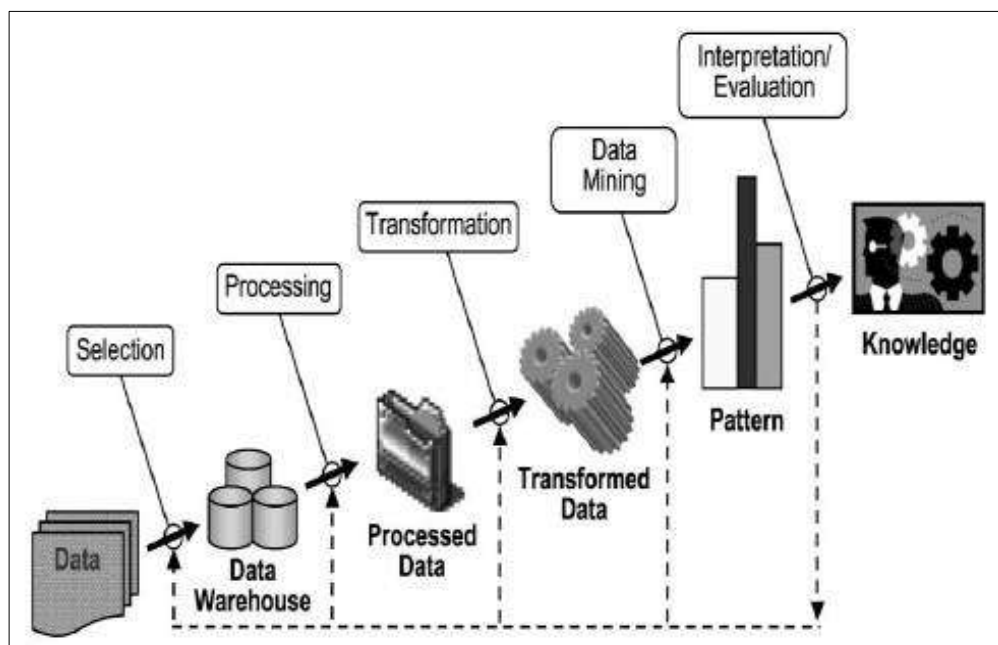
##### **2.1.1. Pengertian *Data Mining***

*Data mining* merupakan istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban & dkk, 2005)

Hampir senada dengan Turban, Beni (2014) mendefinisikan *data mining* sebagai kegiatan menemukan pola yang menarik dari data yang berjumlah besar, data dapat disimpan di *database*, *data warehouse*, atau penyimpanan informasi x lainnya. Pola yang ditemukan merupakan pola yang menarik, berharga dan penting dalam memberikan keuntungan kepada pemilik data. Pendapat yang mirip juga dikemukakan oleh Geetha dan Manimekalai (2013) bahwa *data mining* adalah proses menggunakan berbagai alat analisa data untuk menemukan pola dan hubungan dalam data yang dapat digunakan untuk membuat prediksi yang *valid*. Dari berbagai pendapat diatas, dapat disimpulkan bahwa *data mining* adalah serangkaian proses dalam menemukan pengetahuan tersembunyi dan berharga dari sekumpulan data yang dapat digunakan untuk membuat prediksi yang *valid*.

### 2.1.2. Tahapan *Data Mining*

*Data mining* merupakan rangkaian beberapa proses, maka *data mining* dapat dibagi menjadi beberapa tahap. Tahap-tahap *data mining* dapat diilustrasikan pada gambar 2.1



**Gambar 2.1** Tahapan *Data Mining*

(Sumber: <https://www.kajianpustaka.com/2017/09/data-mining.html>)

Berdasarkan gambar diatas, tahap-tahap *data mining* adalah:

1. *Data selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dimulai. Pada proses ini data diseleksi, hasil

dari data yang sudah diseleksi akan digunakan untuk proses data mining, misalnya data apa saja yang dibutuhkan untuk proses selanjutnya.

#### 2. *Pre-processing (Cleaning)*

Proses pembersihan data untuk membuang data yang tidak konsisten dan *noise*.

#### 3. *Transformation*

Data diubah menjadi bentuk yang sesuai dan diperlukan untuk *di-mining*. Transformasi dan pemilihan data ini juga menemukan kualitas dari hasil *data mining* nantinya karena ada beberapa karakteristik dari teknik-teknik *data mining* tertentu yang tergantung pada tahap ini.

#### 4. *Data Mining*

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi.

#### 5. *Interpretation (Evaluation)*

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### **2.1.3. Pengelompokan *Data Mining***

Menurut Larose dalam (Setiawan & Tes, 2016), data mining dapat dikelompokkan menjadi 6, yaitu:

1. Deskripsi

Deskripsi merupakan cara untuk mendeskripsikan pola dan kecenderungan yang terdapat dalam data yang dimiliki.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, dalam estimasi variabel target lebih kearah numerik dari pada kearah numerik. Model yang dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediksi.

3. Prediksi

Berbeda dengan klasifikasi dan estimasi, pada prediksi nilai dari hasil akan ada dimasa mendatang

4. Klasifikasi

Dalam klasifikasi terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokkan *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan *record-record* dalam kluster lain.

6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

#### **2.1.4. Clustering**

Tujuan utama dari analisis *cluster* adalah untuk mengelompokkan objek- objek berdasarkan karakteristik yang dimilikinya. Analisis cluster mengklasifikasikan objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam cluster yang sama (Sri Mulyati1), 2015).

Klasterisasi banyak digunakan dalam berbagai bidang dengan beragam aplikasi vital yang sangat penting, diantaranya adalah: riset pasar, dimana klasterisasi digunakan untuk segmentasi dan *profiling* pelanggan yang membantu dalam merancang strategi-strategi produk, harga, tempat, dan promosi. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical* (Ong, 2013).

*Hierarchical clustering* adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya



akan membentuk sebuah cluster. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Ong, 2013).

Berbeda dengan metode *hierarchical clustering*, metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua *cluster*, tiga *cluster*, atau lain sebagainya). Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means Clustering* (Santoso, 2010) dalam (Ong, 2013).

#### **2.1.5. *K-means Clustering***

*K-means* merupakan algoritma *clustering* yang berulang-ulang. Langkah pertama algoritma *K-means* dimulai dengan pemilihan secara acak  $K$ ,  $K$  merupakan banyaknya cluster yang ingin dibentuk. Kemudian secara random tetapkan nilai  $K$ . Nilai  $K$  yang sudah ditetapkan untuk sementara menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidean hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid (*distance space*). Beberapa cara yang telah diimplementasikan dalam menghitung jarak (*distance*) antara data dan centroid termasuk diantaranya  $L1$  (*Mahanttan/City Block distance space*),  $L2$  (*Euclidean distance space*), dan  $Lp$  (*Minkowski distance space*) (Studi et al., 2016).

Menurut Santosa (2007) dalam (Ong, 2013), langkah-langkah melakukan *clustering* dengan metode *K-Means* adalah sebagai berikut:

1. Tetapkan jumlah cluster  $k$ .
2. Inisialisasi  $k$  pusat cluster ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberiduberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \dots (1)$$

Dimana:

$D(i,j)$  = jarak data ke  $i$  ke pusat *cluster*  $j$

$X_{ki}$  = data ke  $i$  pada atribut data ke  $k$

$X_{kj}$  = titik pusat ke  $j$  pada atribut  $k$

4. Hitung kembali pusat cluster dengan keanggotaan cluster yang sekarang. Pusat cluster adalah rata-rata dari semua data/ objek dalam cluster tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.

5. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi

### **2.1.6. Pengertian Nasabah**

Nasabah adalah aset atau kekayaan utama perusahaan karena tanpa pelanggan perusahaan tidak berarti apa-apa. Ada istilah “pelanggan adalah raja” yang semua kebutuhan dan keinginan pelanggan harus dipenuhi, dalam banyak hal perusahaan memang harus mengikuti kebutuhan dan keinginan pelanggannya jika ingin produk yang ditawarkan laku dipasaran.

### **2.1.7. Pengertian Asuransi**

Menurut Ketentuan Pasal 246 KUHD dalam (Guntara, 2016), Asuransi atau Pertanggungan adalah Perjanjian dengan mana penanggung mengikatkan diri kepada tertanggung dengan menerima premi untuk memberikan penggantian kepadanya karena kerugian, kerusakan atau kehilangan keuntungan yang diharapkan yang mungkin dideritanya akibat dari suatu evenemen (peristiwa tidak pasti).

Asuransi atau pertanggungan adalah *perjanjian antara dua pihak atau lebih dengan mana pihak penanggung mengikatkan diri kepada tertanggung dengan menerima premi asuransi untuk memberikan penggantian kepada tertanggung*

*karena kerugian, kerusakan atau kehilangan keuntungan yang diharapkan, atau tanggung jawab hukum kepada pihak ketiga yang mungkin akan diderita tertanggung yang timbul dari suatu peristiwa yang tidak pasti, atau untuk memberikan suatu pembayaran yang didasarkan atas meninggal atau hidupnya seseorang yang dipertanggungkan, menurut Ketentuan Undang-undang No.2 tahun 1992 tertanggal 11 Februari 1992 tentang Usaha Perasuransian (“UU Asuransi”) yang sudah dicabut oleh Undang-undang No. 40 tahun 2014 tertanggal 17 Oktober 2014 tentang Perasuransian yang memuat pengertian asuransi sebagai berikut : Asuransi adalah perjanjian antara dua pihak, yaitu perusahaan asuransi dan pemegang polis, yang menjadi dasar bagi penerimaan premi oleh perusahaan asuransi sebagai imbalan untuk:*

- a. memberikan penggantian kepada tertanggung atau pemegang polis karena kerugian, kerusakan, biaya yang timbul, kehilangan keuntungan, atau tanggung jawab hukum kepada pihak ketiga yang mungkin diderita tertanggung atau pemegang polis karena terjadinya suatu peristiwa yang tidak pasti; atau
- b. memberikan pembayaran yang didasarkan pada meninggalnya tertanggung atau pembayaran yang didasarkan pada hidupnya tertanggung dengan manfaat yang besarnya telah ditetapkan dan/atau didasarkan pada hasil pengelolaan dana.

### **2.1.8. Rapid Miner**

Software yang digunakan dalam penelitian ini adalah Rapid Miner. Tujuan dari penggunaan software ini adalah membandingkan hasil dengan perhitungan secara teoritis dengan hasil yang didapatkan dengan proses di Rapid Miner.

Rapid Miner adalah perangkat lunak yang digunakan untuk mengolah data. Rapid miner mengekstrak pola-pola dari dataset yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan serta database (Dumai et al., 2017). Dengan menggunakan Rapid Miner, pengguna dapat mengolah data dengan jumlah yang banyak dengan menggunakan operator-operator. Modifikasi data dilakukan dengan memanfaatkan operator-operator tersebut. Mekanismenya untuk melihat hasil, data tinggal hanya dihubungkan dengan node hasil. Rapid Miner mampu menampilkan hasil secara visual dan grafik.

## **2.2. Penelitian Terdahulu**

Beberapa hasil penelitian terdahulu yang memanfaatkan teknik *clustering* dalam berbagai bidang yaitu:

**Tabel 2.1** Penelitian Terdahulu

Judul	Nama Pengarang	ISSN	Hasil
Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University	Johan Oscar Ong (2013)	1412-6869	Penelitian ini digunakan untuk menentukan strategi maketing President University dengan mengelompokkan data mahasiswa yang telah lulus mulai dari nama mahasiswa hingga kota asal
Application of k-Means Clustering algorithm for prediction of Students' Academic Performance	Oyelade, O. J Oladipupo, O. O Obagbuwa, I. C	1947-5500	Lembaga swasta di Nigeria mengimplementasikan algoritma k-means untuk memprediksi dan memantau perkembangan kinerja akademik siswa di insitusi yang lebih tinggi yang bertujuan untuk membuat keputusan yang lebih efektifk dalam membuat perencanaan akademik dimasa yang akan datang.

**Tabel 2.1** Lanjutan

<p>Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)</p>	<p>Fitri Yunita</p>	<p>2302-8149</p>	<p>Algoritma K-means clustering digunakan untuk membantu menentukan strategi mempromosikan masing – masing program studi yang ada di Universitas Islam Indragiri.</p>
<p>Penerapan Data Mining Dengan Metode Interpolasi Untuk Memprediksi Minta Konsumen Asuransi (Studi Kasus Asuransi Metlife)</p>	<p>Sandy Kurniawan, Taufiq Hidayat</p>	<p>0854-4743</p>	<p>Penelitian ini menggali informasi dari basis data nasabah dan transaksi pembayaran premi asuransi sehingga dapat memprediksi seberapa besar angsuran premi asuransi yang terbaik sesuai dengan kondisi calon nasabah.</p>

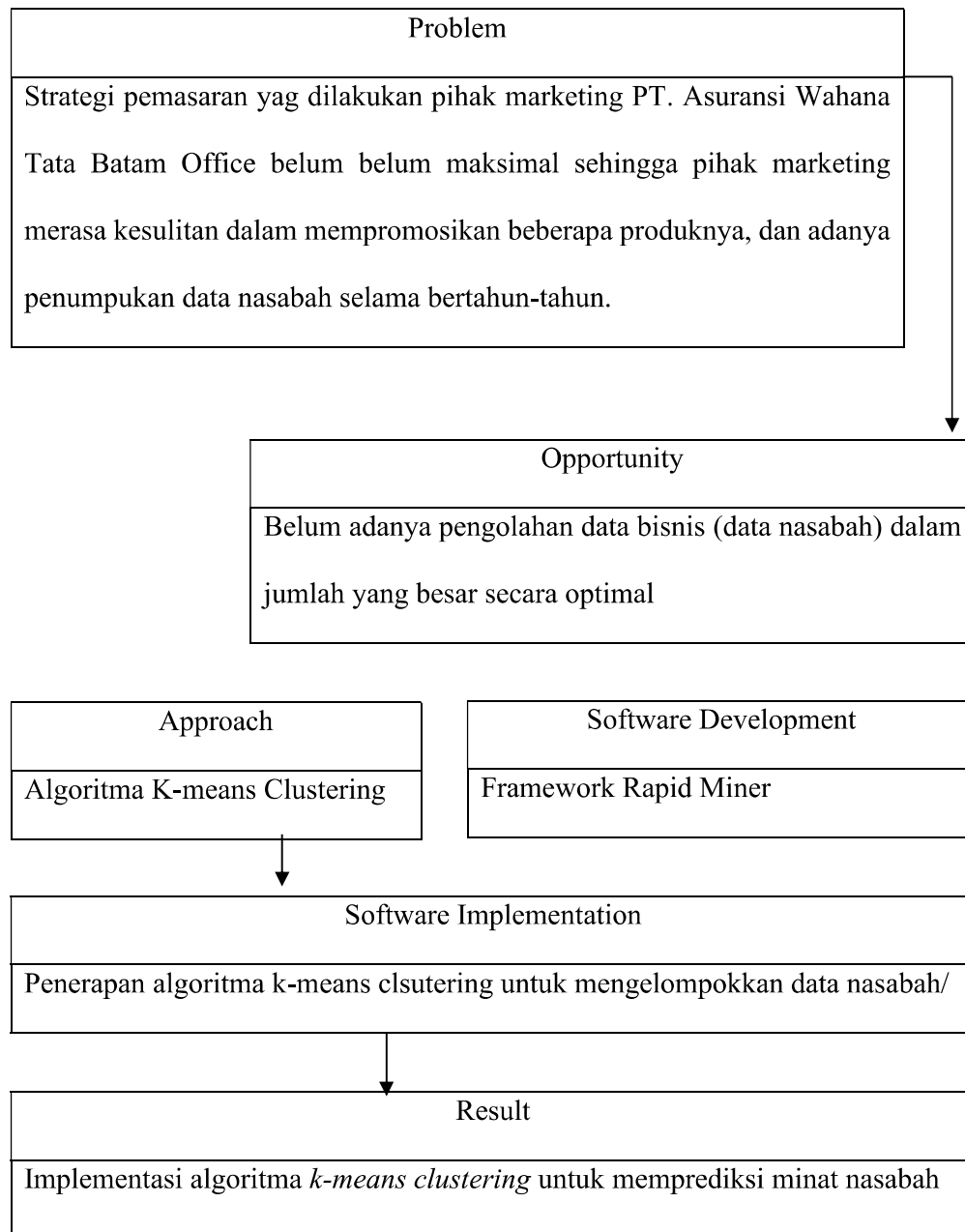
**Tabel 2.1 Lanjutan**

Implementasi Data Mining Pelanggan Potensial Menggunakan Algoritma K-Means	Randi Rian Putra, Cendra Wadisman	2614-1574	K-means diimplementasikan dalam pemilihan pelanggan potensial di MC Laundry.
--	-----------------------------------	-----------	--

### 2.3. Kerangka Pemikiran

Kerangka pemikiran adalah narasi (uraian) atau pernyataan (proposisi) tentang kerangka konsep pemecahan masalah yang telah diidentifikasi atau dirumuskan. Melalui uraian dalam kerangka berpikir, peneliti dapat menjelaskan secara komprehensif variabel-variabel apa saja yang diteliti dan dari teori apa variabel-variabel itu diturunkan, serta mengapa variabel-variabel itu saja yang diteliti. Uraian dalam kerangka berpikir harus mampu menjelaskan dan menegaskan secara komprehensif asal-usul variabel yang diteliti, sehingga variabel-variabel yang tercantum di dalam rumusan masalah dan identifikasi masalah semakin jelas asal-usulnya (<https://id.scribd.com/id>, n.d.).





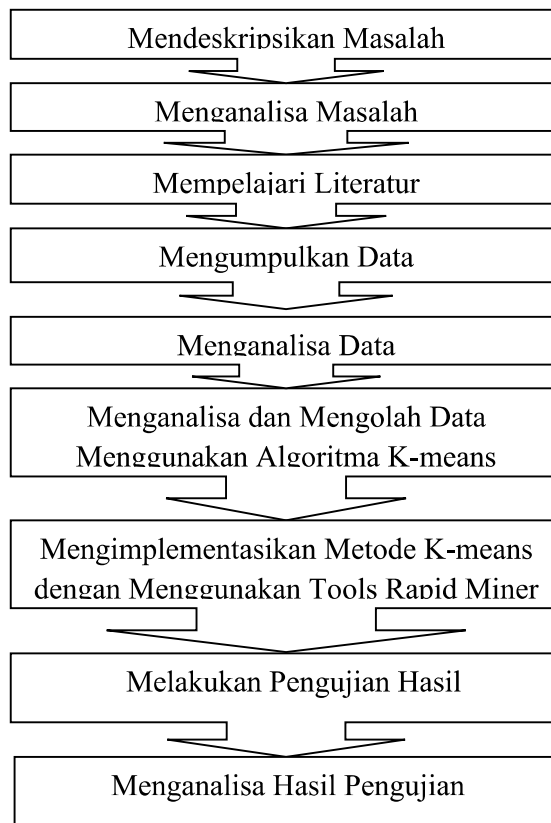
**Gambar 2.2** Kerangka Pemikiran

(Sumber: Data Penelitian, 2018)

## BAB III METODE PENELITIAN

### 3.1. Desain Penelitian

Dalam melaksanakan penelitian dibutuhkan desain penelitian agar memudahkan pembaca untuk melihat alur penelitian yang dilakukan. Berikut tahapan penelitian yang akan dilakukan dan ditetapkan pada proses penelitian ini adalah:



**Gambar 3.3** Desain Penelitian

(Sumber: Penelitian, 2018)

Tahapan penelitian dapat dijelaskan sebagai berikut:

1. Mendeskripsikan masalah

Mendeskripsikan masalah yang akan diteliti perlu ditentukan terlebih dahulu. Mendeskripsikan masalah dalam penelitian dengan menentukan dan mendefinisikan batasan masalah yang akan diteliti, sehingga membantu dalam mendapatkan solusi yang terbaik dari masalah tersebut.

2. Menganalisa masalah

Analisa masalah merupakan langkah untuk dapat memahami masalah yang telah ditentukan batasannya. Dengan menganalisa masalah yang telah ditentukan tersebut, diharapkan masalah dapat dipahami dengan baik.

3. Mempelajari literatur

Penulis melakukan pengumpulan pengetahuan dari berbagai sumber literatur berupa buku-buku, jurnal, dan karya ilmiah yang berkaitan dengan topik yang penulis angkat

4. Mengumpulkan data

Dalam mengumpulkan data, penulis menggunakan beberapa metode yaitu:

- a. Metode wawancara, penulis melakukan tanya jawab dengan staf kantor PT. Asuransi Wahana Tata Batam Office mengenai data nasabah asuransi.
- b. Metode observasi, penulis melakukan pengamatan langsung terhadap objek penelitian yaitu pada PT. Asuransi Wahana Tata Batam Office.

5. Menganalisa data

Penulis melakukan analisa data menggunakan tahapan data mining yang meliputi: seleksi data, *pre-processing* dan pembersihan data, transformasi data, *data mining*, *evaluasi/interpretation*.

6. Menganalisa dan mengolah data menggunakan algoritma *k-means*

Data yang diperoleh dari tempat penelitian yang sudah dianalisa, selanjutnya dilakukan pengolahan menggunakan algoritma *k-means*.

7. Mengimplementasikan Metode *K-means* dengan Menggunakan Tool Rapid Miner

Dalam pengimplementasian metode ini menggunakan spesifikasi hardware & software sebagai berikut:

- a. Laptop HP *processor* Intel Celeron N3060 1.6 Ghz *Memory* 4GB.
- b. Sistem operasi windows 10, sebagai perangkat lunak sistem operasi laptop.
- c. Perangkat lunak microsoft office excel 2016 sebagai *database* pada aplikasi Rapid Miner.
- d. Perangkat lunak Rapid Miner 8.1.003 sebagai tool untuk pengujian Algoritma *K-means*.

8. Melakukan pengujian hasil

Pada tahap ini, penulis melakukan pengujian menggunakan software data mining open source Rapid Miner.

9. Menganalisa hasil pengujian

Hasil yang didapat dalam pengujian perlu dianalisa, dibahas dan dievaluasi, apakah sudah menjawab permasalahan yang dipaparkan pada Bab I atau belum.

## **3.2. Populasi dan Sampel**

### **3.2.1. Populasi**

Populasi merupakan kumpulan yang lengkap dari seluruh elemen yang sejenis dan dapat dibedakan menjadi obyek penelitian. Populasi dapat dibedakan menjadi 2 antara lain: populasi tak terhingga dan populasi terhingga (Heridiansyah, 2012). Populasi yang akan digunakan sebagai penelitian adalah data nasabah asuransi PT. Asuransi Wahana Tata Batam Office.

### **3.2.2. Sampel**

Jika populasi besar, dan peneliti tidak mungkin mempelajari semua yang ada pada populasi, misalnya karena keterbatasan dana, tenaga, dan waktu, maka peneliti dapat menggunakan sampel yang diambil dari populasi itu. Apa yang dipelajari dan diperoleh dari sampel itu kesimpulannya akan diberlakukan untuk populasi. Untuk itu sampel yang ada dalam penelitian harus benar-benar bisa mewakili populasi yang ada. Sampel yang akan digunakan peneliti adalah 101 data set nasabah asuransi yang mengambil jenis asuransi ritel.

### **3.3. Teknik Pengumpulan Data**

Pengumpulan data dilakukan untuk memperoleh informasi yang penulis butuhkan dalam rangka mencapai tujuan penelitian. Teknik pengumpulan data, tergantung pada faktor utama dan jenis data.

#### **3.3.1. Metode Pengumpulan Data**

Metode pengumpulan data pada penelitian ini dilakukan dengan cara sebagai berikut:

1. Wawancara

Penulis melakukan tanya jawab dengan staf kantor PT. Asuransi Wahana Tata Batam Office mengenai data nasabah asuransi.

2. Observasi

Merupakan teknik pengumpulan data dimana peneliti melakukan pengamatan langsung ke objek penelitian mengenai data-data penelitian yang dibutuhkan ke PT. Asuransi Wahana Tata Batam Office.

3. Studi Pustaka

Peneliti mengumpulkan data dengan cara mencari dan mempelajari data-data dan buku-buku serta refensi lain yang berhubungan dengan penulisan laporan penelitian.

### 3.3.1. Data Penelitian

Dalam penelitian ini, penulis menggunakan data yang bersumber dari data primer dan data sekunder.

#### 1. Data Primer

Data primer adalah data yang diperoleh peneliti secara langsung, yang dalam hal ini adalah data yang di dapat dari PT. Asuransi Wahana Tata Batam Office. Data tersebut berupa data nasabah PT. Asuransi Wahana Tata Batam Office pada tahun 2016 sampai 2017. Data nasabah asuransi yang diperlukan oleh peneliti terdiri dari nama nasabah, nomor polis, jenis asuransi yang dipilih, pekerjaan, pendapatan perbulan, pekerjaan dan alamat nasabah.

#### 2. Data Sekunder

Data yang diperoleh dari sumber yang sudah ada, seperti dari buku-buku, situs internet yang memiliki hubungan erat dengan penulisan skripsi ini.

### 3.4. Metode Analisis Data

Metode analisis data yang dilakukan dalam penelitian ini untuk penerapan *data mining* adalah menggunakan tahapan *Knowledge Discovery in Database* (KDD) atau tahapan *data mining* yang terdiri dari:

1. Data selection
2. Pre-processing (*data cleaning*)

3. Transformation
4. Data mining
5. Interpretation (evaluation)

### **3.5. Lokasi dan Jadwal Penelitian**

#### **3.5.1. Lokasi Penelitian**

Lokasi dalam penelitian ini dilakukan pada PT. Asuransi Wahana Tata Batam Office yang terletak di Komplek Baloi Office Park No.6, Lubuk Baja. Alasan memilih PT. Asuransi Wahan Tata Batam Office berdasarkan pertimbangan:

1. PT. Asuransi Wahan Tata Batam Office memiliki data bisnis yang besar, yang belum dimanfaatkan dengan optimal.
2. Rendahnya minat atau ketertarikan calon nasabah terhadap beberapa produk asuransi yang ditawarkan

#### **3.5.2. Waktu Penelitian**

Waktu pelaksanaan penelitian akan dilaksanakan selama 3 bulan, yaitu mulai bulan Oktober sampai bulan Desember 2018.



**Tabel 3.1** Jadwal Penelitian

No	Uraian	Oktober				November				Desember			
		Minggu Ke											
		1	2	3	4	1	2	3	4	1	2	3	4
1	Persiapan penelitian	■											
2	Pengumpulan data		■	■									
3	Menganalisis dan mengolah data				■	■	■	■					
4	Mengimplementasi k-means dengan Weka Tool								■	■	■		
5	Pengujian dan analisa hasil												■

(Sumber: Data Penelitian, 2018)