

**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 Muhammat 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

|                                                |     |
|------------------------------------------------|-----|
| HALAMAN SAMPUL DEPAN                           |     |
| HALAMAN JUDUL                                  |     |
| LEMBAR PERNYATAAN .....                        | i   |
| LEMBAR PENGESAHAN .....                        | ii  |
| KATA PENGANTAR.....                            | iii |
| ABSTRAK .....                                  | v   |
| ABSTRACT .....                                 | vi  |
| DAFTAR ISI.....                                | vii |
| DAFTAR TABEL .....                             | ix  |
| DAFTAR GAMBAR.....                             | x   |
| BAB I PENDAHULUAN.....                         | 1   |
| 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   |
| BAB II TINJAUAN PUSTAKA.....                   | 8   |
| 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 |