

---

# **COSTUMER PROFILING DARI PRODUK ASURANSI DENGAN MENGGUNAKAN METODE K-MEANS MANHATTAN DISTANCE DI BPJS KETENAGAKERJAAN SEMARANG**

**Novy Amalia, Fajrian Nur Adnan. M.CS**

Jurusan Sistem Informasi, Ilmu Komputer, Universitas Dian Nuswantoro  
Fakultas Ilmu Komputer Universitas Dian Nuswantoro Semarang  
Jalan Nakula I No. 5-11 Semarang 50131  
Telp : (024) 3517261, Fax : (024) 3520165  
E-mail : 112201204599@mhs.dinus.ac.id<sup>1</sup>, fajrian@dsn.dinus.ac.id<sup>2</sup>

## **Abstrak**

*Pada dasarnya manusia tidak hanya membutuhkan kebutuhan primer yang harus mereka penuhi untuk keberlangsungan hidupnya. Akan tetapi manusia juga perlu untuk menjamin keamanan hidup mereka, kesehatan maupun permasalahan financial. Untuk mampu memenuhi kebutuhan tersebut. Karena itu manusia membutuhkan asuransi, oleh karena itu perusahaan perlu sebuah cara untuk mengelompokkan pelanggan yang dikenal juga dengan istilah customer segmentation atau customer profiling. Pada penelitian ini digunakan metode klastering (Clustering) dengan algoritma K-Means Manhattan. Sebelum dilakukan pengolahan data, dilakukan proses normalisasi data, kemudian data diolah menjadi beberapa klaster. Data yang telah diklasterisasi tersebut menghasilkan kategori prediksi yang nantinya digunakan untuk acuan dalam proses costumer profiling. Dari laporan tugas akhir ini dihasilkan sebuah pengklusteran costumer profiling terhadap produk asuransi. Dengan adanya penelitian tersebut, perusahaan asuransi dapat mengetahui hasil costumer profiling dan dapat memberikan tindakan preventif.*

*Kata kunci : Data Mining, Clustering, k-Means, Manhattan Distance, Asuransi, Costumer Profiling.*

## **Abstract**

*Basically, human beings not only need a primary requirement but their also need sekunder requirment to survival their life and need to secure their life, like to secure their health and financial problems. To fully and solve their problem about life secure, therefore need insurance. Therefore, companies need a way to categorize customer also known as customer segmentation or customer profiling. In this resorce used a method clustering (Clustering) with K-Means algorithm Manhattan distance. The clustered data generates predictions categories that will be use to process costumer profiling. Before processing the data, carried out the process of normalization of data, then the data is processed into several clusters. This essay has result many clustering which one can use for costumer profiling of insurance products and with this research also an insurance company can find out the results of customer profiling and can provide preventive measures.*

*Key words : Data Mining, Clustering, k-Means, Manhattan Distance, Insurance, Costumer Profiling.*

## **1. PENDAHULUAN**

Saat ini banyaknya perubahan pada dunia bisnis kian pesat salah satu latar belakang yang mendasari perubahan tersebut yaitu persaingan antar perusahaan. Hal ini tentu saja membuat berbagai perusahaan harus dapat bersaing dengan cara yang efektif agar tercapainya daya saing yang strategis agar nantinya memberikan keuntungan yang

maksimal. Salah satu persaingan yang sedang marak di sektor bisnis yaitu persaingan perusahaan asuransi. Semakin banyaknya persaingan di dunia asuransi membuat para *stakeholder* harus mampu melakukan sebuah terobosan-terobosan maupun inovasi yang nantinya dapat menjamin terus berlangsungnya bisnis asuransi mereka. Pada dasarnya manusia tidak hanya membutuhkan kebutuhan primer yang harus mereka penuhi untuk keberlangsungan hidupnya akan tetapi manusia juga perlu untuk menjamin keamanan hidup mereka, kesehatan maupun permasalahan financial. Untuk mampu memenuhi kebutuhan tersebut oleh karena itu manusia membutuhkan asuransi. Minat masyarakat mengenai asuransi pun semakin meningkat. Menurut Otoritas Jasa Keuangan (OJK) tercatat minat masyarakat terhadap asuransi meningkat sekitar 37% dalam kurun waktu 2010-2014 [1].

Perusahaan asuransi juga mulai gencar menawarkan produk-produk asuransi yang mereka tawarkan. Misalnya saja asuransi jiwa, asuransi kendaraan bermotor, asuransi pendidikan dan sebagainya. Pada kasus asuransi jiwa banyaknya produk asuransi jiwa yang ditawarkan seringkali membuat para konsumen bingung untuk memilih produk mana yang sesuai dengan kebutuhan dan sesuai dengan premi yang pas untuk mereka. Membahas permasalahan dalam perusahaan asuransi mengenai masalah yang berhubungan dengan konsumen, perusahaan perlu sebuah cara untuk mengelompokkan pelanggan yang dikenal juga dengan istilah *customer segmentation* atau *customer profiling*. Konsep ini pada dasarnya berusaha mengelompokkan pelanggan berdasarkan kesamaan karakteristik tertentu yang dimilikinya. Selain permasalahan konsumen juga tidak lepas bagaimana perusahaan dapat memasarkan produk yang ada ke pasar yang tepat. Inilah permasalahan yang nantinya akan melahirkan sebuah kebutuhan akan teknologi yang diharapkan dapat membentuk pengetahuan-pengetahuan baru maupun mampu membantu para *stakeholder* ataupun manajerial dalam mengatur strategi bisnis asuransi mereka. Melakukan prediksi akan minat suatu konsumen merupakan sebuah hal yang penting dalam perusahaan asuransi, dimana dengan melakukan sebuah prediksi minat konsumen perusahaan dapat mengambil sebuah strategi yang tepat maupun keputusan bagi konsumen mereka.

## 2. METODE PENELITIAN

Objek penelitian pada tugas akhir ini adalah BPJS Ketenagakerjaan Cabang II Semarang, dimana pada penelitian ini membahas proses Costumer Profiling pada produk asuransi dengan perhitungan data mining yang menggunakan algoritma k-Means dengan Manhattan Distance.

Data mining yang sering disebut dengan knowledge discovery in database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam dataset besar [2]. Selain itu Data Mining adalah teknik untuk memanfaatkan dan mencari atau menggali hubungan-hubungan antar setiap pola yang tersembunyi dalam data melalui sebuah proses ekstrak [3]. Dapat disimpulkan data mining merupakan proses penggalian data tersembunyi dengan menggunakan sebuah metodologi analisis data untuk menemukan dan mendapatkan pola yang unik dan menarik dalam suatu dataset yang besar.

Menurut [2] KDD ada beberapa rangkain proses iteratif yaitu sebagai berikut ini:

1. Data Cleaning (menghilangkan noise dan datayang tidak konsisten)
2. Data Intergration (penggabungan dari beberapa sumber data)
3. Data Selection (pengambilan data yang relevan yang digunakan untuk analisis)

4. Data Transformation (data yang nantinya ditransformasikan untuk membentuk data yang sesuai yang nantinya diproses dalam data mining)
5. Data Mining (proses penting dimana proses ini mengaplikasikan metode yang digunakan untuk mengestrak pola data)
6. Pattern Evaluation (mengidentifikasi pola yang menarik dan yang dapat mewakili knowledge dari pola itu sendiri)
7. Knowledge Presentation (teknik representasi dan visualisasi knowledge yang digunakan untuk menampilkan sebuah knowledge kepada user)

k-Means adalah algoritma clustering yang berulang-ulang, dimulai dengan pemilihan k secara acak. k merupakan banyaknya kluster yang akan dibentuk [3]. Metode ini bertujuan untuk meminimkan jumlah jarak antar semua titik dengan pusat kluster. Langkah-langkah algoritma k-Means adalah sebagai berikut:

1. Menentukan beberapa kluster k
2. Menentukan nilai centroid. Nilai awal centroid di cari dengan secara acak. Sedangkan untuk nilai centroid saat iterasi, menggunakan rumus sebagai berikut.

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (2.1)$$

Dimana,

$v_{ij}$  adalah centroid atau rata-rata kluster ke-i untuk variable ke-j;

$N_i$  adalah jumlah data yang menjadi anggota dari kluster ke-i;

i,k adalah indeks dari kluster;

j adalah indeks dari variable;

$x_{kj}$  adalah nilai data ke-k yang terdapat pada kluster tersebut untuk variable ke-j.

3. Menghitung jarak antara titik centroid dengan setiap titik obyek. Menghitung dengan menggunakan euclidean distance, dengan rumus sebagai berikut

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2.2)$$

Dimana;

$D_e$  adalah euclidean distance;

$i$  adalah banyaknya objek;

$(x, y)$  adalah koordinat objek;

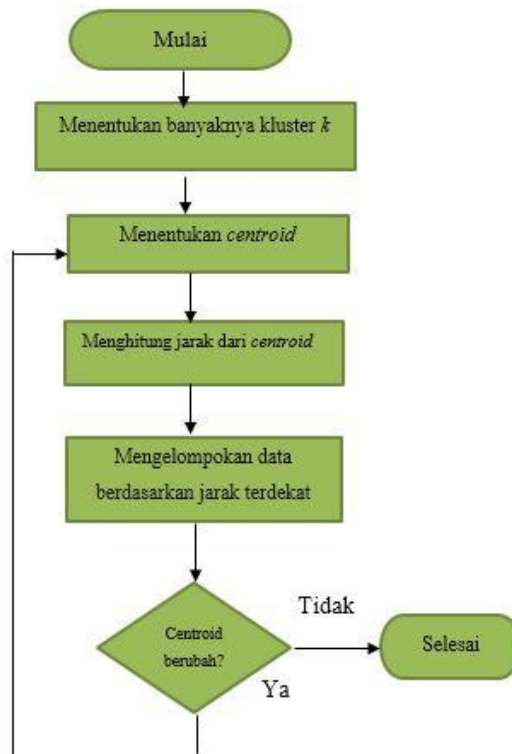
$(s, t)$  adalah koordinat centroid.

Karena pada penelitian ini menggunakan Manhattan distance maka rumus yang digunakan sebagai berikut ini :

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}|$$

4. Menggelompokkan objek. Yaitu untuk menentukan anggota kluster dengan menghitung jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada jarak matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke kluster dan nilai 0 untuk data yang dialokasikan ke kluster lain.

5. Kembali lagi ke langkah kedua. Melakukan iterasi hingga hasil centroid tetap dan anggota kluster tidak berpindah ke kluster lain. Gambar berikut ini merupakan diagram alir yang menjelaskan alur dari algoritma k-Means.



**Gambar 1** Gambar Diagram Alir Algoritma k-Means

Dalam data mining juga diperlukan sebuah validasi data, *Shilouette Index* (SI) yaitu metode yang digunakan untuk memvalidasi sebuah data, kluster tunggal, atau bahkan keseluruhan klaster. Metode ini juga banyak digunakan untuk melakukan validasi klaster dengan menggabungkan nilai kohesi dan separasi [10].

Nilai  $a_i$  digunakan untuk mengukur kemiripan sebuah data dengan kluster yang diikutinya. Nilai semakin kecil menunjukkan bahwa kluster tersebut tepat berada di kluster tersebut.  $b_i$  menunjukkan adanya tingkat keburukan data terhadap kluster yang lainnya. Semakin besarnya nilai  $b_i$  menunjukkan bahwa data tersebut semakin buruk. SI mempunyai nilai rentan -1 hingga +1. Jika nilai SI mendekati angka 1 menunjukkan bahwa data tersebut tidak tepat berada pada kluster tersebut. SI bernilai 0 atau mendekati 0 berarti bahwa tersebut ada pada posisi perbatasan dua kluster.

Berikut ini merupakan nilai ukuran *shilouette index* menurut Kaufman dan Rousseeuw [15]. Nilai shilouette coefficient (SC) :

1.  $0,7 < SC \leq 1$  *strong structure*
2.  $0,5 < SC \leq 0,7$  *medium structure*
3.  $0,25 < SC \leq 0,5$  *weak structure*
4.  $SC \leq 0,25$  *no structure*

### 3. HASIL DAN PEMBAHASAN

### 3.1 Data Pelanggan

Data pelanggan yang diperoleh dari BPJS Ketenagakerjaan Cabang II Semarang berupa data primer dalam bentuk excel. Data tersebut berisikan field-field yang menggambarkan ciri-ciri seorang customer yang terdiri dari no, nama, sex, alamat, lahir, perusahaan, upah dan program.

Data tersebut perlu diolah terlebih dahulu kedalam dataset SQLyog sebelum nantinya diolah dalam program perhitungan MATLAB untuk melakukan *customer profiling*.

**Tabel 3-1 Data Pelanggan BPJS**

NO	NAMA	SEX	ALAMAT	LAHIR	PERUSAHAAN	UPAH	PROGRAM
1	KUSNADI	L	JALIRT 01 RW 01 BONANG	18-02-1976	SUBUR MAKMUR MIGAS PRATAMA PT	1.909.000	JKK, JKM
2	SUROTO	L	PURWOREJO RT03 RW01 BONANG DEMAK	28-04-1988	MAKMUR JAYA MIGAS PRATAMA PT	1.941.000	JKK, JKM
3	SUPRAPTO	L	TEGALREJO RT 02 RW 09 KEL KEMIJEN KEC SEMARANG TIMUR KAB SEMARANG	28-07-1984	QUARTINDO SEJATI FURNITAMA PT	1.941.000	JKK, JHT
4	RINI NIRWATI	P	JOGO IV RT05 RW02 KEC SAYUNG	09-11-1988	ALTEK PRECISION MACHINERY PT	1.941.000	JKK, JKM
5	SITI MUNIFAH	P	DK KRAJAN LOR RT 001/003 JEPANGPAKIS JATI KUDUS	31-05-1981	LASINDO METAL UTAMA CV	1.941.000	JKK, JKM
6	FEBRI SULISTYANA	P	DS WONOLOPO RT04 RW01 KEC MIJEN KAB SEMARANG	12-02-1989	UNIVERSAL JIROLU CORPORATION CV	1.976.407	JKM
7	SUMARDIYANI	L	DS SARI RT02 RW04 KEL KALI SARI KEC GAJAH KAB DEMAK	16-07-1988	UPPKH KECAMATAN KRADENAN	1.981.000	JKK, JKM
8	ZENI AGGRAENI	P	TANGGUNG HARJO RT03 RW04 KELBALEROMO KEC DEMPET KAB DEMAK	01-07-1977	UPPKH KECAMATAN GABUS	1.981.000	JKK, JKM
9	MASRUM	L	BUARAN RT 09 RW 01 MAYONG JEPARA	29-11-1989	UPPKH KECAMATAN BRATI	2.013.000	JKK, JKM
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
1499	DYAH DEWI HARSANTI	P	NGELOWETAN, 001/001, MIJEN, DEMAK	29-08-1984	VICTORY APPAREL TKA PT	1.909.000	JKK, JHT, JKM

Dari data excel diatas hanya beberapa variable data yang akan digunakan untuk perhitungan *k-means*, oleh karena itu sebelum melakukan proses pengolahan data menggunakan algoritma *k-means* tersebut terlebih dahulu sebelum nantinya dilanjutkan ke proses pengolahan data kedalam SQLyog dengan menggunakan metode data mining.

### 3.2 Data Cleaning

Data yang di dapatkan dari BPJS Ketenagakerjaan lebih dahulu harus diperiksa lebih lanjut. Setelah diteliti dan dilihat lebih lanjut ternyata data yang di dapatkan perlu adanya data yang dibersihkan karena data tersebut tidak relevant untuk melakukan proses perhitungan. Dari data BPJS Ketenagakerjaan tersebut ada beberapa yang harus di *cleaning* yaitu NAMA, ALAMAT dan PERUSAHAAN. Hal ini dikarenakan variable nama, alamat dan perusahaan tidak digunakan dalam proses data mining karena memiliki keinkonsistenan data.

NO	NAMA	SEX	ALAMAT	LAHIR	PERUSAHAAN	UPAH	PROGRAM
1	KUSNADI	L	JALIRT 01 RW 01 BONANG	18-02-1976	SUBUR MAKMUR MIGAS PRATAMA PT	1.909.000	JKK, JKM
2	SUROTO	L	PURWOREJO RT03 RW01 BONANG DEMAK	28-04-1988	MAKMUR JAYA MIGAS PRATAMA PT	1.941.000	JKK, JKM
3	SUPRAPTO	L	TEGALREJO RT 02 RW 09 KEL KEMIJEN KEC SEMARANG TIMUR KAB SEMARANG	28-07-1984	QUARTINDO SEJATI FURNITAMA PT	1.941.000	JKK, JHT
4	RINI NIRWATI	P	JOGO IV RT05 RW02 KEC SAYUNG	09-11-1988	ALTEK PRECISION MACHINERY PT	1.941.000	JKK, JKM
5	SITI MUNIFAH	P	DK KRAJAN LOR RT 001/003 JEPANGPAKIS JATI KUDUS	31-05-1981	LASINDO METAL UTAMA CV	1.941.000	JKK, JKM

↓

NO	SEX	LAHIR	UPAH	PROGRAM
1	L	1976	1.909.000	'JKK, JKM'
2	L	1988	1.941.000	'JKK, JKM'
3	L	1984	1.941.000	'JKK, JHT'
4	P	1988	1.941.000	'JKK, JKM'
5	P	1981	1.941.000	'JKK, JKM'

**Gambar 2 Hasil Data Cleaning**

### 3.3 Data Integration

Data *Integration* atau integrasi data merupakan proses penggabungan data dari berbagai *database* ke dalam suatu *database* baru. Integrasi data ini bisanya dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Dalam data yang di peroleh dari BPJS Ketenagakerjaan proses data *integration* tidak dilakukan dikarenakan data yang di dapat sudah saling terintegrasi satu sama lain dan tidak membutuhkan adanya proses integrasi data lagi.

### 3.4 Data Selection

Data *selection* berfungsi untuk melakukan pemilihan (seleksi) data dari sekumpulan sebuah data operasional. Proses ini perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data BPJS Ketenagakerjaan yang telah didapatkan untuk penelitian ini kemudian dilakukanlah data *selection* yang mana proses ini dilakukan untuk penyeleksian data yang digunakan untuk proses penelitian, setelah diteliti dan dianalisa lebih lanjut dari beberapa data yang ada, ternyata data yang digunakan sesuai dan isinya mencakup semua kebutuhan yang dibutuhkan dalam penelitian yang akan dilakukan.

### 3.5 Data Transformation

Proses *transformation* merupakan proses untuk menggabungkan atau mengubah data kedalam format yang sesuai untuk diproses dalam data mining. Ada beberapa metode dalam data mining yang membutuhkan format data yang khusus sebelum bisa diaplikasikan dalam proses data mining tersebut. Sebagai contoh ada beberapa metode standar seperti analisis asosiasi dan clustering, metode tersebut hanya bisa menerima input data berjenis data katagorikal maupun numerik. Proses inilah yang disebut *transformasi* data, dalam penelitian ini dilakukan sebuah *transformasi* data untuk

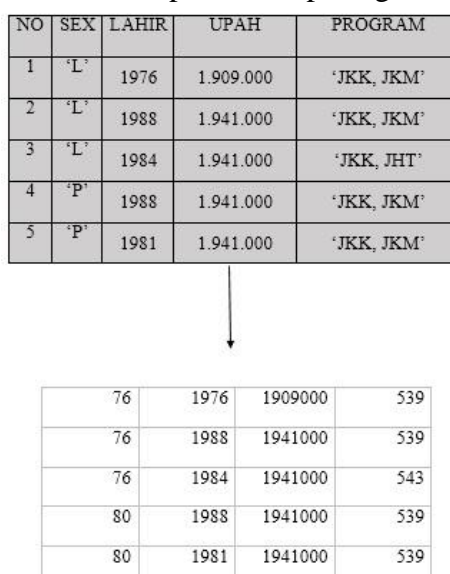
mengubah variable “Program” yang berupa data berjenis *character* menjadi data berjenis *numerical*.

Berikut ini tabel proses data selection menggunakan ASCII.

**Tabel Error! No text of specified style in document.-2 Tabel Konversi ke ASCII**

JKK, JKM	224+44+32+226+13=539
JKK, JHT	224+44+32+230+13=543
JKM	226+13=239
JHT, JKM	230+44+32+226+13=545
JKK, JHT, JKM	224+44+32+230+44+32+226+13=845
JKK, JHT, JKM,JP	224+44+32+230+44+32+226+44+154+13=1043
JKK, JKM,JP	224+44+32+226+44+154+13=737
JKK, JHT, JP	224+44+32+230+44+154+13=773
JHT, JKM,JP	230+44+32+226+44+154+13=743

Hal ini dilakukan guna memudahkan dalam proses pengimplementasian algoritma. Hasil dari proses perubahan tersebut dapat dilihat pada gambar 3



**Gambar 3 Hasil Perubahan Variabel Sex dan Program**

### 3.6 Data Mining k-Means

Pada proses data mining ini akan membahas bagaimana perhitungan manual yang terjadi di dalam program yang dibuat. Perhitungan manual ini berfungsi sebagai tolak ukur untuk mengoreksi apakah program yang dibuat apakah sudah memiliki hasil yang sama dengan perhitungan manualnya atau belum.

Berikut ini langkah-langkah perhitungan *k-means* :

1. Menentukan pusat kluster awal secara acak dari *dataset*. Banyaknya kluster  $k=2$  dan didapatkan *cluster* awal sebagai berikut ini

**Tabel 3-3 Centroid Awal**

SEX	LAHIR	UPAH	PROGRAM
80	1984	2628000	845
76	1984	2428000	743

2. Menghitung jarak antara data terhadap setiap pusat klaster. Misalkan, untuk menghitung jarak data jenis kelamin pertama dengan pusat klaster pertama. Perhitungan jarak yang digunakan dengan metode perhitungan jarak *manhattan distance*

Jarak antara data ke-1 dengan centroid pertama

	SEX	LAHIR	UPAH	PROGRAM
Data ke -1	76	1976	1909000	539
Centroid ke - 1	80	1984	2628000	845
Jarak =  data- centroid	4	8	719000	306
Total Jarak	719318			

Lakukan perhitungan ke seluruh data, berikut ini tabel hasil perhitungan.

**Tabel 3-4 Hasil Perhitungan jarak tiap data**

Data ke-	SEX	LAHIR	UPAH	PROGRAM	C1	C2	Cluster
1	76	1976	1909000	539	719318	519212	C2
2	76	1988	1941000	539	687314	487208	C2
3	76	1984	1941000	543	687306	487200	C2
...	...	...	...	...	...	...	...
1421	76	1994	2528000	1043	100212	100310	C1
1422	76	1994	2528000	1043	100212	100310	C1
1423	80	1982	2528000	845	100002	100108	C1
...	...	...	...	...	...	...	...
1444	76	1975	2778000	1043	150211	350309	C1
...	...	...	...	...	...	...	...
1449	80	1984	1909000	845	719000	519106	C2

3. Selanjutnya kelompokkan hasil perhitungan jarak tersebut menurut pembagian clusternya. suatu data akan menjadi anggota klaster dengan jarak terkecil dari pusat klaster. Misalkan untuk data pertama, jarak terkecil terdapat pada cluster pertama sehingga data pertama akan menjadi anggota dari cluster pertama. Begitu pula untuk data ketiga, jarak terkecil terdapat pada klaster kedua sehingga data tersebut masuk sebagai anggota klaster kedua.



- Setelah melalui tahap pengkelompokan kluster dengan menggabungkan kluster awal kemudian proses selanjutnya yaitu melakukan pengelompokan tiap cluster dan pembuatan atau penentuan centroid baru. Berikut ini tabel perhitunganya.

**Tabel 4-5 Pengelompokan Data dengan Cluster 1 dan 2**

Data ke-	SEX	LAHIR	UPAH	PROGRAM	Cluster
1	76	1976	1909000	539	2
2	76	1988	1941000	539	2
3	76	1984	1941000	543	2
4	80	1988	1941000	539	2
5	80	1981	1941000	539	2
...	...	...	...	...	...
1449	80	1984	1909000	845	2
Centroid baru	76,642	1984,178	3392910,714	908	

Data ke-	SEX	LAHIR	UPAH	PROGRAM	Cluster
1421	76	1994	2528000	1043	1
1422	76	1994	2528000	1043	1
1423	80	1982	2528000	845	1
1424	76	1985	2528000	845	1
1425	80	1989	2578000	845	1
...	...	...	...	...	...
1476	76	1979	12928000	845	1
Centroid baru	77,061	1986,400	2101315,05	759,7636868	

- Setelah melakukan pengelompokan dan perhitungan data dengan kluster maka dihasilkan sebuah centroid baru seperti pada tabel berikut.

**Tabel 3-6 Centroid Baru**

SEX	LAHIR	UPAH	PROGRAM
76,64286	1984,179	3392911	908
77,06168	1986,401	2101315	759,7637

- Kemudian setelah terbentuknya centroid baru langkah berikutnya yaitu kembali lagi pada langkah perhitungan jarak data terhadap pusat kluster dengan menggunakan centroid baru. Lakukan perhitungan hingga tidak ada cluster yang berubah.

**Tabel 3-7 Tabel Cluster iterasi 1-4**

Data ke-	Cluster				Posisi
	Iterasi ke-1	Iterasi ke-2	Iterasi ke-3	Iterasi ke-4	
1	2	2	2	2	Tidak Berubah
2	2	2	2	2	Tidak Berubah
3	2	2	2	2	Tidak Berubah
...	...	...	...	...	...
1421	1	2	2	2	Tidak Berubah
1422	1	2	2	2	Tidak Berubah
1423	1	2	2	2	Tidak Berubah

...	...	...	...	...	...
1444	1	1	2	2	Tidak Berubah
...	...	...	...	...	...
...	...	...	...	...	...
1499	2	2	2	1499	Tidak Berubah

Setelah melakukan empat kali iterasi ternyata pada iterasi ke-4 tidak ada lagi cluster yang berpindah ke cluster lain. Oleh karena proses dihentikan dikarenakan sudah tidak ada lagi cluster yang berpindah.

7. Kemudian kelompokan data menurut centroid dan hasil cluster lakukan perbandingan dengan nilai k= 3 dan 4. Dan di dapatkan hasil sebagai berikut ini.

**Tabel 3-8 Tabel Perbandingan Hasil Cluster**

Cluster	K=2		K=3			K=4			
	1	2	1	2	3	1	2	3	4
Data ke-									
1	-	2	-	-	3	-	-	3	-
2	-	2	-	-	3	-	-	3	-
...	...	...	...	...	...	...	...	...	...
941	-	2	-	2	-	-	2	-	-
...	...	...	...	...	...	...	...	...	...
1421	-	2	1	-	-	1	-	-	-
1422	-	2	1	-	-	1	-	-	-
...	...	...	...	...	...	...	...	...	...
1444	-	2	1	-	-	1	-	-	-
...	...	...	...	...	...	...	...	...	...
1499	-	2	-	-	3	-	-	3	-

### 3.7 Evaluasi Hasil Clustering dengan Silhouette Index-Manhattan

Berikut ini validasi data hasil klastering dengan menggunakan Silhouette Index sebagai suatu standar validasi dalam tahap evaluasi. Hasil perhitungan validasi berdasarkan hasil klastering Setelah melihat tabel (3.8) lakukan perlakuan yang sama saat melakukan perhitungan SI yaitu melakukan juga perhitungan dengan K3&K4 dengan hasil sebagai berikut ini.

**Tabel 3-9 Hasil Evaluasi Clustering dengan Silhouette Index**

	K2		K3			K4			
	C1	C2	C1	C2	C3	C1	C2	C3	C4
SI	-0,54	0,58	0,87	0,37	0,36	0,88	0,61	0,23	-0,32
SI Global	0,422858		0,3229			0,351304			

Tabel di atas merupakan hasil hitung Silhouette Index (SI) yang berisi SI awal dimana perhitungan ini menggunakan metode manhattan distance, kemudian didapatkan nilai SI global di setiap cluster. Untuk ukuran nilai SI menurut Kaufman dan Rousseeuw,

pada kolom SI cluster, dapat disimpulkan bahwa cluster pertama sebesar -0.532881847 (C1) termasuk dalam *weak structure*, sedangkan klaster kedua (C2) sebesar 0,579442988 termasuk dalam *medium structure*. Nilai SI global sebesar 0.422858 juga termasuk dalam *weak structure*. Akan tetapi nilai k=2 memiliki nilai SI global yang lebih besar dibandingkan nilai SI global yang lain. Sehingga dapat disimpulkan bahwa k optimal pada k=2 dengan nilai SI global yang terbesar diantara k yang lain. Nilai SI yang mendekati 1 menandakan bahwa data tersebut makin tepat di dalam cluster. Sedangkan SI bernilai 0 (atau mendekati 0) berarti data tersebut posisinya berada di perbatasan di antara dua cluster.

### 3.8 Analisa Hasil Cluster

Setelah melakukan empat kali iterasi dan perhitungan nilai SI global, melakukan perhitungan distance dan mencari hasil cluster di setiap data hingga tidak adanya perubahan cluster. Diperolehlah titik pusat akhir (centroid akhir) sebagai berikut.

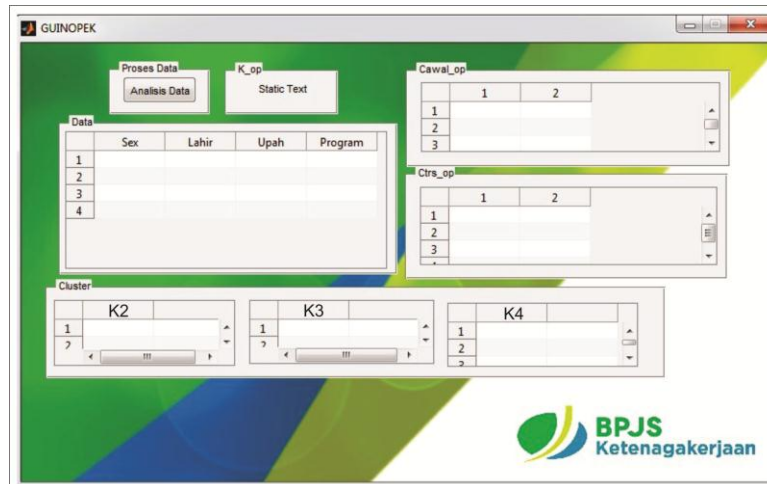
**Tabel 3-10 Tabel titik pusat cluster**

	SEX	LAHIR	UPAH	PROGRAM
C1	76,25	1983	4984875	919,25
C2	77,05	1986,35	2118976,8	763,6405934

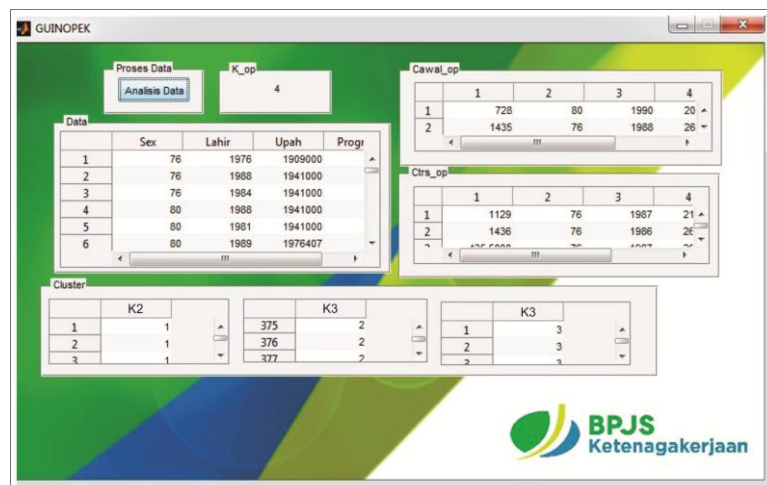
Tabel 3.10 di atas merupakan hasil titik pusat akhir (centroid). Titik pusat akhir ada dua, yaitu C1 dan C2 yang mewakili masing-masing klaster. Klaster-klaster yang merupakan anggota dari C1 dan C2 adalah klaster yang memiliki jarak terdekat terhadap titik pusat. Mengacu pada kolom titik pusat pada tabel 3.10, prediksi dapat dikatakan jika C2 merupakan centroid yang optimal. Dan dengan mengetahui hasil dari centroid tersebut maka diharapkan nantinya dalam penelitian berikutnya centroid ini dapat digunakan untuk mengklasifikasikan maupun melakukan pelabelan sesuai dengan hasil analisa diatas. Menurut hasil dari C2 dapat di labelkan menjadi customer dengan jenis kelamin laki-laki yang lahir pada tahun berkisar tahun 1986 dengan gaji berkisar 2118976 dan program yang biasanya dipilih “JHT,JKM,JP”.

### 3.9 Hasil Program

Gambar dibawah ini merupakan tampilan GUI (Graphical User Interface) menggunakan program MATLAB. program ini hanya untuk menunjang proses perhitungan apakah proses perhitungan manualnya sudah sesuai atau belum. Dimana program tersebut menampilkan hasil centroid awal optimal, cluster optimal dan nilai centroid.



**Gambar 4 Tampilan Awal Program**



**Gambar 5 Gambar Hasil Program**

Gambar 5 diatas merupakan tampilan program setelah menekan button Analisa Data. Gambar diatas berisi hasil dari program matlab yang digunakan untuk mengolah data asuransi BPJS Ketenagakerjaan. Program diatas menampilkan hasil dari centroid awal optimal, cluster optimal dan nilai centroid.

#### 4. KESIMPULAN

Kesimpulan yang didapatkan berdasarkan penelitian ini adalah:

1. Dengan data pelanggan yang diolah menggunakan metode klustering k-Means tersebut, maka dapat dilakukan pengambilan keputusan sebagai acuan tindakan preventif. Selain itu hasil klasterisasi diatas dapat digunakan untuk mengklasifikasi. Dan dapat disimpulkan C2 merupakan cluster optimal. Menurut hasil dari C2 dapat di labelkan menjadi costumer dengan jenis kelamin laki-laki yang lahir pada tahun berkisar tahun 1986 dengan gaji berkisar 2118976 dan program yang bisanya dipilih "JHT,JKM,JP".
2. Hasil dari klasterisasi diuji dengan perhitungan validasi menggunakan Silhouette Index (SI) menghasilkan nilai SI klaster, klaster pertama sebesar -0,5329 (C1) termasuk dalam *non structure*, sedangkan klaster kedua (C2) sebesar 0,5794

---

termasuk dalam *medium structure*. Nilai SI global sebesar 0,4229 termasuk dalam *weak structure*.

## 5. SARAN

Untuk meningkatkan kinerja dan menyempurnakan penelitian ini, berikut ini beberapa saran yang dapat digunakan dalam penelitian selanjutnya adalah:

1. Mengembangkan GUI (Graphical User Interface) agar tampilan lebih menarik dan user friendly.
2. Diharapkan dapat membandingkan dengan algoritma data mining yang lain agar mengetahui algoritma yang lebih baik dalam melakukan customer profiling.

## DAFTAR PUSTAKA

- [1] N. Ain, “<http://www.jiwasaraya.co.id>,” replubika , 10 Oktober 2014. [Online]. Available: <http://www.jiwasaraya.co.id>. [Diakses 4 November 2015].
- [2] L. Ulfah, “Penerapan data Mining untuk memprediksi premi ideal menggunakan metode K-Means cluster guna meningkatkan minat konsumen asuransi (studi kasus : Asuransi Bumiputera cabang Pamulang),” p. 01, 2013.
- [3] D. B. Atyanto, “Customer Profiling dengan Menggunakan k-Means Clustering untuk Mendukung Pengambilan Keputusan Strategis di PT. Pelabuhan Indonesia III (Persero),” p. 1, 2011.
- [4] D. A.B dan D. Kulkarni, “ Application of Data Mining Techniques in Life Insurance,” vol. II, pp. 31-40, 2012.
- [5] samsudin, “OJK,” 6 febuari 2013. [Online]. Available: <http://www.ojk.go.id/asuransi>. [Diakses 7 november 2015].
- [6] M. G. A, Dasar-dasar Asuransi Jiwa dan Asuransi Kesehatan, Jakarta: Intermedia, 2007.
- [7] H. Rahim, “Optimisme Pertumbuhan Asuransi Indonesia,” p. 20, 2013.
- [8] J. Han dan M. Kamber, Data Mining : Concept and Techniques Third Edition, San Frasco: Morgan Kaufmann Publishers, 2008.
- [9] E. Prasetyo, DATA MINING : Mengolah Data menjadi Informasi Menggunakan

Matlab, Yogyakarta: Andi, 2014.

- [10] P. E, Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab, Yogyakarta: ANDI, 2014.
- [11] .. J. Oyelade, O. O. Oladipupo dan I. C. Obagbuwa, ““Application of k-Means Clustering algorithm for prediction of Students Academic Performance”,” *International Journal of Computer Science and Information Security*, vol. 07, pp. 292-295, 2010.
- [12] R. T. d. K. S, ““Aplikasi k-Means untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) & Ukuran Kerangka”,” *Seminar Nasional Aplikasi Teknologi Informasi*, pp. 43-48, 2008.
- [13] W. I. H, F. Eibe dan H. M. A, DATA MINING "Practical Machine Learning Tools and Techniques" 3rd Edition, United State : Morgan Kauffman , 2011.
- [14] N. S. Yaniar, “Perbandingan Ukuran Jarak pada Proses Pengenalan Wajah berbasis PCA,” p. 6, 2008.
- [15] K. L dan R. P. J, Finding Groups in Data, New York: John Wiley & Sons, 2007.
- [16] Anhar, Panduan Menguasai PHP dan Mysql, Jakarta: Media Kita, 2010.
- [17] R. A. M, Pemograman Web Dinamis Menggunakan Php dan Mysql, Yogyakarta: ANDI, 2011.
- [18] W. Y, MEbangun Bisnis Online dengan Mambo, Jakarta: PT.Elex Media Komputindo, 2008.
- [19] T. B. R dan Misinem, Pemrograman Pada Port Printer, Yogyakarta: Ardana Media, 2008.
- [20] BPJS , “Website Resmi BPJS Ketenagakerjaan,” [Online]. Available: <http://www.bpjsketenagakerjaan.go.id/page/profil/Sejarah.html>. [Diakses 10 Mei 2016].
- [21] R. S.Pressman, Rekayasa Perangkat Lunak, Jakarta: Andi Offset, 2012.

- [22] R. Mcleod dan Schell, Sistem Informasi Manajemen . Edisi 9, Jakarta : PT.Index, 2007.
- [23] R. Soetam, Konsep Dasar Rekasyas Perangkat Lunak, Pustaka : Jakarta, 2010.
- [24] M. Shalahudin dan R. A. S, Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek, Bandung: Informatika, 2013.
- [25] A. Nugroho, Analisis dan Perancangan Sistem Informasi, Bandung: Informatika, 2008.
- [26] M. E. Khan, Different Forms of Software Testing for Finding Error, 2010.
- [27] Anhar, Panduan Menguasai PH dan Mysql, Jakarta: Media Kita, 2010.
- [28] M. Arief, Pemrograman Web Dinamis Menggunakan Php dan Mysql, Yogyakarta: ANDI, 2011.
- [29] Y. Wicaksono, Membangun Bisnis Online dengan Mambo, Jakarta: PT. Elex Media Komputindo, 2008.