

---

## Customer Profiling dengan Menggunakan Metode *K-Means Euclidean Distance* di BPJS Ketenagakerjaan Tanjung Perak

Anastasya Febiyati Ayutrisula<sup>1</sup>, Aris Fanani<sup>2</sup>

<sup>1</sup>Universitas Islam Negeri Sunan Ampel Surabaya, [anastasyafebiyati53@gmail.com](mailto:anastasyafebiyati53@gmail.com)

<sup>2</sup>Universitas Islam Negeri Sunan Ampel Surabaya, [arisfa@uinsby.ac.id](mailto:arisfa@uinsby.ac.id)

**Abstrak:** Pada dasarnya manusia memerlukan jaminan sosial untuk menjamin keamanan hidup mereka dan mampu memenuhi kebutuhan hidup. Oleh karena itu, manusia membutuhkan asuransi berupa program jaminan dan perusahaan perlu melakukan suatu cara untuk mengelompokkan pelanggan berdasarkan karakteristik yang dimiliki atau yang dikenal dengan *customer profiling*. Pada penelitian ini digunakan metode *clustering* dengan algoritma *k-means euclidean distance*. Sebelum dilakukan pengolahan data, data perlu dinormalisasikan terlebih dahulu, kemudian data diolah menjadi beberapa *cluster*. Data yang telah di *clustering* akan menghasilkan suatu kategori yang akan digunakan dalam proses *customer profiling*. Dari penelitian ini dihasilkan *cluster* pertama termasuk dalam *customer* yang memiliki upah tinggi dengan kecenderungan memilih program JKM, JHT dan untuk *cluster* kedua termasuk dalam *customer* yang memiliki upah rendah dengan kecenderungan memilih program JKK, JKM. Dengan adanya penelitian ini, perusahaan dapat mengetahui hasil *customer profiling* untuk dapat melakukan tindakan selanjutnya.

**Kata kunci:** *Data mining, Clustering, K-Means, Euclidean Distance, Customer Profiling*

**Abstract:** Basically humans need social security to ensure the safety of their lives and be able to fulfill their daily needs. Therefore, humans need insurance in the form of a guarantee program and companies need to do a way to group customers based on the characteristics they have or what is known as *customer profiling*. In this study the clustering method using the *k-means euclidean distance* algorithm was used. Before processing data, data needs to be normalized first, then the data is processed into several clusters. Data that has been clustered will produce a category that will be used in the *customer profiling* process. From this research, the first cluster included in high-paying customers was the tendency to choose the JKM, JHT program and for the second cluster included in the low-wage customer with a tendency to choose the JKK program, JKM. With this research, companies can find out the results of *customer profiling* to be able to take further action.

**Keyword:** *Data Mining, Clustering, K-Means, Euclidean Distance, Customer Profiling*

## 1. Pendahuluan

Indonesia merupakan negara kepulauan yang sekitar 70% adalah wilayah laut dan telah tesebar 17.504 pulau. Kondisi demografis dan geografisnya menjadi peluang dan tantangan bagi pemerintah Indonesia untuk memenuhi amanat yang ada pada Pembukaan Undang-Undang Dasar (UUD) 1945 tentang memenuhi kebutuhan hidup masyarakat di Indonesia mencakup : mewujudkan suatu pemerintahan yang dapat melindungi segenap bangsa dan seluruh tumpah darah Indonesia, memajukan kesejahteraan umum, dan mencerdaskan kehidupan bangsa. Agar amanah tersebut bisa berjalan, pemerintah wajib memberikan jaminan sosial terutama jaminan sosial ketenagakerjaan bagi masyarakat [1].

Suatu bentuk perlindungan bagi tenaga kerja yang ada di Indonesia wajib dilaksanakan oleh sebuah perusahaan yang mempekerjakan seseorang untuk bekerja di perusahaan tersebut. Bentuk perlindungan tersebut yaitu dengan meningkatkan kesejahteraan dan melakukan pemeliharaan tenaga kerja dalam bentuk jaminan sosial. Jaminan sosial yang diberikan pada tenaga kerja harus dilakukan berdasarkan asas gotong-royong dan kekeluargaan yang telah tercantum dalam jiwa, Undang-Undang Dasar 1945, dan Pancasila [2].

Jaminan sosial merupakan suatu kebijaksanaan yang diberikan kepada tenaga kerja yang ada dilingkungan perusahaan dalam melakukan penyelenggaraan dan memberikan perlindungan dengan interaksi kerja yang bisa menguntungkan tenaga kerja dan pengusaha [3]. Jamsostek (Jaminan Sosial Tenaga Kerja) adalah suatu program yang dapat memberikan sebuah perlindungan kepada tenaga kerja untuk mengatasi permasalahan sosial ekonomi yang menggunakan sistem asuransi sosial dalam penyelenggaraannya. Perusahaan ini didirikan untuk memenuhi kebutuhan jaminan sosial tenaga kerja. Pada tahun 2014, jamsostek berganti nama menjadi BPJS (Badan Penyelenggara Jaminan Sosial) [4].

BPJS adalah suatu badan hukum publik yang telah dibentuk untuk menyelenggarakan jaminan sosial seperti : Jaminan Kecelakaan Kerja (JKK), Jaminan Kematian (JKM), Jaminan Hari Tua (JHT) dan Jaminan Pensiun (JP). Jaminan sosial diberikan kepada tenaga kerja untuk meningkatkan produktivitas kerja. Jaminan sosial merupakan suatu jaminan yang menjadi hak tenaga kerja yang berbentuk tunjangan berupa pelayanan, pengobatan dan uang sebagai pengganti pengasilan yang berkurang akibat peristiwa yang dialami oleh tenaga kerja seperti hamil, kecelakaan saat kerja, menganggur, dan meninggal dunia [5].

Hak yang akan diberikan kepada keluarga tenaga kerja ini yang dimaksud adalah dengan memberikan jaminan sosial berupa pelayanan jika anggota keluarga tersebut mengalami sakit dan membutuhkan bantuan tenaga medis. Selain itu, jika tenaga kerja meninggal dunia pihak keluarga juga akan diberikan biaya untuk pemakaman dan santunan kematian [5].

*Customer profiling* adalah metode untuk membuat sebuah potret pelanggan yang mencakup pribadi dan perincian transaksional. Hal ini dilakukan untuk membantu perusahaan membuat variasi keputusan yang berpusat pada pelanggan tentang bisnis mereka. Dengan menggunakan *customer profiling* pemasar dapat menentukan strategi pemasaran mana yang harus dilakukan [6].

*Clustering* adalah suatu proses pengelompokan data yang jaraknya paling dekat dan memiliki kemiripan dengan objek yang lain dan data yang akan dicluster dapat ditentukan secara random. Teknik *clustering* adalah teknik yang telah menunjukkan kesamaan karakteristik pada suatu data dan informasi yang akan dihasilkan dapat bermanfaat [7]. Tujuan dilakukan *clustering* adalah untuk menemukan set *cluster* yang optimal.

*K-Means* adalah suatu algoritma *clustering* yang dilakukan secara berulang-ulang. Algoritma *k-means* dapat dimulai dengan memilih data secara acak  $k$ ,  $k$  merupakan banyaknya jumlah *cluster* yang akan dibentuk [8].

Melihat dari penelitian sebelumnya yang dilakukan oleh Novy Amalia, dkk yang berjudul *Customer Profiling* Dari Produk Asuransi Dengan Menggunakan Metode *K-Means Manhattan Distance* di BPJS Ketenagakerjaan Semarang. Dalam penelitiannya digunakan metode *k-means* untuk mengelompokkan data *customer* berdasarkan kesamaan karakteristik yang dimiliki [9].

Dari penelitian yang telah dilakukan sebelumnya, metode algoritma *k-means* adalah salah satu metode yang telah berhasil digunakan dalam proses pengelompokkan. Di dalam kasus ini, peneliti akan melakukan penelitian pengelompokkan berdasarkan data upah di BPJS Ketenagakerjaan menjadi beberapa kelompok. Pengelompokkan tersebut menggunakan metode *k-means*. Dengan data yang telah dikelompokkan menggunakan *k-means* diharapkan dapat mempermudah dalam menghitung hasil *customer profiling* berdasarkan upah yang dimiliki.

Berdasarkan latar belakang masalah tersebut sehingga penulis mengangkat sebuah judul dari permasalahan ini adalah “*Customer Profiling* Dengan Menggunakan Metode *K-Means* Di BPJS Ketenagakerjaan Tanjung Perak”.

## 2. Tinjauan Pustaka

### A. BPJS Ketenagakerjaan

#### 2.1 Pengertian

BPJS Ketenagakerjaan merupakan badan hukum publik yang memiliki fungsi untuk melaksanakan suatu program seperti Jaminan Kematian (JKM), Jaminan Kecelakaan Kerja (JKK), Jaminan Pensiun (JP), dan Jaminan Hari Tua (JHT). Sebelum menjadi BPJS Ketenagakerjaan, dulunya bernama PT Jamsostek (Persero) dan kemudian berubah menjadi BPJS Ketenagakerjaan sejak tanggal 1 Januari 2014. Dalam memberikan suatu perlindungan sosial bagi tenaga kerja dan keluarganya, pemerintah Indonesia telah melakukan berbagai upaya untuk memberdayakan dan mendayagunakan pekerja secara optimal, mewujudkan kesempatan kerja, mewujudkan kesejahteraan tenaga kerja, memberikan suatu perlindungan dan meningkatkan kesejahteraan bagi pekerja dan keluarganya.

Jaminan sosial pada BPJS Ketenagakerjaan dibagi menjadi dua yaitu bersifat jangka pendek seperti kecelakaan kerja dan jaminan untuk kesehatan dan jangka panjangnya seperti pensiun, Jaminan Hari Tua (JHT), dan kematian. Sejak diberlakukannya Undang-undang Nomor 40 Tahun 2004 Pasal 5 ayat (3) yaitu tentang Sistem Jaminan Sosial Nasional. Dalam mewujudkan harapan tersebut pemerintah harus melakukan tindakan sesuai dengan Undang-undang Nomor 40 Tahun 2004 tentang Sistem Jaminan Sosial Nasional untuk bagi masyarakat Indonesia [10].

#### 2.2 Program BPJS Ketenagakerjaan

##### 1. Jaminan Kecelakaan Kerja (JKK)

Kecelakaan tenaga kerja dalam pekerjaan adalah suatu resiko yang harus dihadapi. Dibutuhkan jaminan kecelakaan saat kerja untuk menanggulangi resiko yang terjadi seperti kehilangan penghasilan saat terjadi kecelakaan, mengalami cacat fisik atau mental, resiko kematian. Kecelakaan yang dialami oleh tenaga kerja adalah tanggung jawab dari pengusaha dengan membayar suatu iuran

jaminan kecelakaan kerja sesuai dengan pekerjaannya sekitar 0,24-1,74. Besarnya iuran yang harus dibayarkan, jika memiliki penghasilan sekitar 1.700.000-1.899.000 iuran yang harus dibayar adalah sekitar 18.000. Jaminan tersebut diberikan sebagai kompensasi untuk pekerja yang mengalami kecelakaan saat berangkat kerja hingga sampai dirumah.

2. Jaminan Kematian (JKM)

Jaminan kematian diberikan kepada ahli waris dari peserta yang mengikuti program jaminan BPJS Ketenagakerjaan yang telah meninggal dunia tetapi bukan karena mengalami kecelakaan saat kerja. Jaminan tersebut diberikan untuk meringankan beban keluarga seperti santunan berupa uang dan biaya pemakaman. Pada program ini tenaga kerja wajib membayar iuran sebesar 16.800.

3. Jaminan Hari Tua (JHT)

Jaminan Hari Tua (JHT) dilakukan dengan sistem Tabungan untuk hari tua yang ditanggung oleh pengusaha dan setiap bulannya di kredit pada rekening pekerja secara individu serta setiap tahunnya mendapatkan hasil dari pengembangan. Iuran yang diberikan sebesar 5,7% terdiri dari 3,7% dari yang memberikan pekerjaan dan 2% dari peserta. Besarnya iuran yang wajib dibayarkan oleh tenaga kerja pada program JHT jika memiliki upah sekitar 2.700.000-3.199.000 adalah sebesar 59.000.

4. Jaminan Pensiun (JP)

Jaminan Pensiun merupakan suatu jaminan sosial yang diberikan untuk menunjang kehidupan yang layak bagi tenaga kerja atau ahli waris dengan memberikan penghasilan kepada peserta ketika mengalami cacat, memasuki masa pensiun, atau meninggal dunia dan jaminan tersebut akan diberikan setiap bulannya.

## B. Customer Profiling

*Customer profiling* adalah suatu metode untuk membuat sebuah potret pelanggan yang mencakup pribadi dan perincian transaksional. Hal ini dilakukan untuk membantu perusahaan membuat variasi keputusan yang berpusat pada pelanggan tentang bisnis mereka. *Customer profiling* atau profil pelanggan dibuat dengan bantuan penelitian pelanggan [6].

*Customer profiling* atau profil pelanggan menggambarkan pelanggan berdasarkan atribut mereka seperti usia, jenis kelamin, pendapatan, dan program jaminan untuk mengidentifikasi karakteristik sekelompok orang dan menggambarkan seperti apa orang tersebut. Dengan menggunakan *customer profiling* pemasar dapat menentukan strategi pemasaran mana yang harus dilakukan.

## C. Clustering

*Clustering* adalah metode yang biasanya digunakan untuk mengelompokkan sekumpulan data sesuai dengan kesamaan karakteristik yang dimiliki antara data satu dengan data yang lain. *Clustering* sendiri merupakan salah satu metode algoritma data mining yang telah diterapkan tanpa adanya latihan terlebih dahulu dan tidak diperlukan target outputnya. Tujuan dilakukan *clustering* adalah untuk menemukan set *cluster* yang optimal [7].

#### D. Algoritma K-Means

*K-means* adalah salah satu algoritma yang ada pada data mining digunakan untuk mengelompokkan suatu data. Metode *k-means* merupakan salah satu metode yang ada pada algoritma *clustering* berbasis jarak yang dapat membagi data menjadi beberapa *cluster* dan algoritma *k-means* hanya bisa bekerja pada simbol yang numerik. Langkah-langkah menggunakan algoritma k-Means [11]:

1. Tentukan jumlah *cluster*  $k$ .
2. Tentukan nilai dari suatu *centroid*, pada iterasi 0 yaitu menentukan titik pusat (*centroid*) dari tiap-tiap *cluster* dengan cara melakukan pengambilan secara random (acak), sedangkan pada iterasi selanjutnya menghitung rata-rata dengan menggunakan rumus seperti pada Persamaan 1 berikut:

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

Keterangan :

$v_{ij}$  = rata-rata dari suatu *cluster*  $i$  untuk variable ke- $j$  atau *centroid*

$N_i$  = jumlah dari suatu data yang akan menjadi anggota dari *cluster* ke- $i$

$i, k$  = indeks dari suatu *cluster*

$j$  = indeks dari suatu variable

$x_{kj}$  = nilai dari data ke- $k$  yang ada pada *cluster* tersebut untuk variable ke- $j$ .

3. Menghitung jarak antara titik objek dengan titik *centroid* menggunakan rumus euclidean distance Persamaan 2:

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

Keterangan :

$D_e$  = euclidean distance

$i$  = banyaknya suatu objek

$(x, y)$  = koordinat pada objek

$(s, t)$  = koordinat pada *centroid*.

4. Menghitung jarak terdekat dari objek dengan menentukan anggota *clusternya*. Nilai yang akan diperoleh nantinya dalam data keanggotaan jarak pada matriks yaitu 0 atau 1, dimana nilai 0 merupakan data yang akan dialokasikan ke suatu *cluster* lain dan nilai 1 dialokasikan ke *cluster*.
5. Melakukan iterasi sampai anggota *cluster* tidak berpindah atau tetap dan hasil *centroid* tersebut tidak berubah.

### 3. Metode Penelitian

Pada bab ini akan membahas tentang jenis dan sumber data, metode pengumpulan data, metode analisis data, dan tahap penelitian.

#### A. Jenis dan Sumber Data

Dalam penulisan laporan ini, pengumpulan data yang diperoleh secara langsung dari BPJS Ketenagakerjaan berupa data primer yang diperlukan dalam bentuk excel.

#### B. Metode Penelitian

Data yang akan digunakan dalam penelitian ini membahas tentang suatu proses *Customer Profiling* dengan menggunakan metode *k-means* yang dihitung dengan data

mining. Data mining merupakan suatu proses kegiatan seperti pengumpulan data untuk menemukan keteraturan pola hubungan yang ada di *database* besar. Ada beberapa tahap yang digunakan untuk menganalisis suatu data dalam penerapan data mining :

1. Data *Cleaning* (menghilangkan data yang tidak dibutuhkan)
2. Data *Integration* (penggabungan data dari beberapa sumber)
3. Data *Selection* (pemilihan data yang penting untuk di analisis)
4. Data *Transformation* (transformasi dari suatu data)
5. Data Mining (proses terpenting yang akan diaplikasikan menggunakan metode untuk mengestraksi pola pada data)
6. *Pattern Evaluation* (mengidentifikasi suatu pola yang unik dan menarik)
7. *Knowledge Presentation* (teknik untuk merepresentasi dan visualisasi pengetahuan yang akan digunakan untuk menampilkan suatu pengetahuan kepada pengguna)

*K-means* merupakan metode yang digunakan untuk mengelompokkan suatu data dalam bentuk satu dan dilakukan secara berulang-ulang, sedangkan *k* adalah banyaknya jumlah *cluster* yang akan digunakan [12].

### C. Metode Analisis Data

Data yang digunakan dalam penelitian ini adalah data profil pelanggan dari BPJS Ketenagakerjaan Tanjung Perak Surabaya. Selanjutnya akan dianalisis untuk mengetahui hasil *centroid* yang optimal dalam melakukan pelabelan.

### D. Tahap Penelitian

Penelitian ini memiliki beberapa tahap. Dalam kasus ini, metode yang akan digunakan adalah algoritma *k-means*. Ada beberapa tahap dalam penelitian ini: Tahap pertama menentukan banyaknya jumlah *cluster* *k*, tahap kedua menentukan *centroid* dengan cara melakukan pengambilan *random* (acak), tahap ketiga menghitung jarak dari tiap *centroid*, tahap keempat mengelompokkan data berdasarkan jarak terdekat, dan tahap terakhir melakukan iterasi sampai anggota *cluster* tidak berubah.

## 4. Hasil dan Pembahasan

### A. Data Pelanggan

Data pelanggan yang diperoleh dari BPJS Ketenagakerjaan Tanjung Perak Surabaya berupa data primer dalam bentuk excel. Data tersebut menggambarkan ciri-ciri dari *customer* yang terdiri dari no, nama, jenis kelamin, alamat, lahir, perusahaan, upah, dan program. Data ini perlu diolah untuk melakukan *customer profiling*. Data pelanggan ditunjukkan oleh Tabel 1.

### B. Data *Cleaning*

Data yang telah didapatkan dari BPJS Ketenagakerjaan harus diperiksa kembali sebelum dilakukan proses selanjutnya. Setelah dilihat dan diteliti, ternyata data yang diperoleh tersebut perlu dilakukan pembersihan karena ada data yang tidak relevan untuk dilakukan dalam proses perhitungan. Data dari BPJS Ketenagakerjaan beberapa diantaranya harus di *cleaning* seperti nama, alamat, dan perusahaan karena variabel tersebut tidak akan digunakan dalam data mining karena memiliki keinkonsistenan data.

**Tabel 1.** Data Pelanggan BPJS

No	Nama	Sex	Alamat	Lahir	Perusahaan	Upah	Program
1	Adam	L	Surabaya	09-08-1997	A	2.710.000	JKK, JKM
2	Andi	L	Surabaya	13-06-1984	B	2.710.000	JKK, JKM, JHT, JP
3	Beby	P	Surabaya	14-11-1990	C	2.950.000	JKK, JKM
4	Endang	P	Surabaya	13-08-1980	D	2.710.000	JKK, JKM, JHT, JP
5	Burhan	L	Surabaya	13-08-1993	E	2.710.000	JKK, JKM, JHT
6	Bella	P	Surabaya	13-12-1970	F	3.000.000	JKK, JKM
7	Beni	L	Surabaya	13-12-1986	G	3.000.000	JKK, JKM, JHT, JP
8	Reno	L	Surabaya	14-03-1972	H	2.710.000	JKK, JKM
9	Jamal	L	Surabaya	14-08-1974	I	2.950.000	JKK, JKM
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-
200	Kasan	L	Surabaya	17-09-1987	J	3.000.000	JKK, JKM

**Tabel 2.** Hasil Data Cleaning

No	Nama	Sex	Alamat	Lahir	Perusahaan	Upah	Program
1	Adam	L	Surabaya	09-08-1997	A	2.710.000	JKK,JKM
2	Andi	L	Surabaya	13-06-1984	B	2.710.000	JKK,JKM,JHT,JP
3	Beby	P	Surabaya	14-11-1990	C	2.950.000	JKK,JKM
4	Endang	P	Surabaya	13-08-1980	D	2.710.000	JKK,JKM,JHT,JP
5	Burhan	L	Surabaya	13-08-1993	E	2.710.000	JKK,JKM,JHT



No	Sex	Lahir	Upah	Program
1	L	1997	2.710.000	JKK,JKM
2	L	1984	2.710.000	JKK,JKM,JHT,JP
3	P	1990	2.950.000	JKK,JKM
4	P	1980	2.710.000	JKK,JKM,JHT,JP
5	L	1993	2.710.000	JKK,JKM,JHT

### C. Data Integration

Data *integration* adalah suatu proses untuk menggabungkan sebuah data dari berbagai dataset ke dalam dataset yang baru. Pada data *integration* ini bisa dilakukan pada suatu atribut yang memiliki keunikan seperti jenis program, nama, dan lainnya. data yang telah diperoleh dari BPJS Ketenagakerjaan Tanjung Perak Surabaya dalam proses integrasi data tidak perlu dilakukan karena data tersebut tidak membutuhkan proses ini lagi dan sudah saling terintegrasi.

### D. Data Selection

Data *Selection* memiliki fungsi untuk melakukan suatu pemilihan pada sebuah data dari sekumpulan data operasional. Proses data *selection* perlu dilakukan sebelum menuju tahap berikutnya dimulai untuk menggali informasi dalam KDD. Data yang diperoleh dari BPJS Ketenagakerjaan untuk melakukan penelitian selanjutnya akan dilakukan proses data *selection* untuk menyeleksi sebuah data yang akan digunakan untuk proses penelitian, setelah diteliti dan dianalisa kembali ada beberapa data yang akan digunakan telah mencakup kebutuhan dan sesuai dengan yang dibutuhkan dalam proses penelitian ini.

### E. Data Transformation

Data *Transformation* adalah proses penggabungan dari sebuah data ke dalam suatu format yang telah disesuaikan untuk dilakukan proses data mining. Di dalam data mining ada beberapa metode yang khusus dibutuhkan pada format data sebelum data tersebut diaplikasikan ke dalam data mining. Proses ini disebut data *transformation* dan pada penelitian ini akan dilakukan proses tersebut untuk mengubah suatu variabel “Program” menjadi data dengan jenis numerik. Berikut adalah tabel proses seleksi data dengan menggunakan ASCII.

**Tabel 3.** Konversi ke ASCII

JKK, JHT, JKM, JP	224+44+32+230+44+32+226+44+32+154+13=1075
JKM	226+13=239
JKK, JHT	224+44+32+230+13=543
JKM, JHT	226+44+32+230+13=545
JKK, JKM, JHT	224+44+32+226+44+32+230+13=845
JKK, JKM, JP	224+44+32+226+44+32+154+13=769
JKK, JKM	224+44+32+226+13=539
JHT, JKM, JP	230+44+32+226+44+32+154+13=775
JKK, JHT, JP	224+44+32+230+44+32+154+13=773

Proses ini dilakukan untuk memudahkan dalam melakukan proses mengimplementasikan pada algoritma. Hasil yang diperoleh dari proses perubahan tersebut dapat dilihat pada Tabel 4.

**Tabel 4.** Hasil Perubahan Variabel Sex dan Program.

No	Sex	Lahir	Upah	Program
1	L	1997	2.710.000	JKK, JKM
2	L	1984	2.710.000	JKK, JKM, JHT, JP
3	P	1990	2.950.000	JKK, JKM
4	P	1980	2.710.000	JKK, JKM, JHT, JP
5	L	1993	2.710.000	JKK, JKM, JHT

↓

No	Sex	Lahir	Upah	Program
1	L	1997	2.710.000	539
2	L	1984	2.710.000	1075
3	P	1990	2.950.000	539
4	P	1980	2.710.000	1075
5	L	1993	2.710.000	845

### F. Data Mining *K-Means*

Pada proses ini akan membahas tentang bagaimana melakukan perhitungan secara manual yang ada pada program yang telah dibuat. Perhitungan ini dilakukan untuk mengoreksi apakah hasil yang ada sesuai dengan perhitungannya atau tidak.

Langkah-langkah perhitungan menggunakan *k-means*:

1. Langkah awal adalah menentukan jumlah *cluster*. Jumlah *cluster*  $k = 2$



2. Menentukan pusat awal *cluster* dengan pengambilan secara random (acak). Titik pusat dalam penelitian ini adalah *cluster* untuk upah tinggi (C1) dan upah rendah (C2) dapat dilihat pada tabel

**Tabel 5. Centroid Awal**

Titik Pusat Awal	Sex	Lahir	Upah	Program
<i>Cluster 1</i>	80	1989	2.800.000	845
<i>Cluster 2</i>	76	1979	2.710.000	539

3. Setelah melakukan titik pusat (*centroid*), langkah selanjutnya adalah menghitung jarak objek ke titik pusat (*centroid*) dengan menggunakan rumus perhitungan jarak *euclidean distance*. Di bawah ini terdapat contoh untuk perhitungan jarak data dengan titik pusat.

Contoh:

Perhitungan jarak data ke-1 dengan pusat *cluster* pertama adalah:

$$D_{11} = \sqrt{(76-80)^2 + (1997-1989)^2 + (2.710.000-2.800.000)^2 + (539-845)^2}$$

$$= 90000.52064$$

Perhitungan jarak data ke-1 dengan pusat *cluster* kedua adalah:

$$D_{12} = \sqrt{(76-76)^2 + (1997-1979)^2 + (2.710.000-2.710.000)^2 + (539-539)^2}$$

$$= 18.24828759$$

Perhitungan dilakukan ke seluruh data, berikut adalah tabel hasil dari perhitungannya.

**Tabel 6. Hasil Perhitungan Jarak Tiap Data**

Data ke-	Sex	Lahir	Upah	Program	C1	C2
1	76	1997	2.710.000	539	90000,52	18,24
2	76	1984	2.710.000	1075	90000,29	539,02
3	80	1990	2.950.000	539	150000.31	240000.03
4	80	1980	2.710.000	1075	90000,29	539,01
5	76	1993	2.710.000	845	90000.018	309.316
6	80	1970	3.000.000	539	200000.23	290000.02
7	76	1986	3.000.000	1075	200000.13	290000.50
8	76	1972	2.710.000	539	90000,52	7.6157
9	76	1974	2.950.000	539	150000.31	240000.01
10	80	1989	2.950.000	539	150000.31	240000.03

4. Selanjutnya hasil perhitungan jarak pada setiap data dikelompokkan berdasarkan *clusternya*. Dari pusat *cluster*, data tersebut akan menjadi anggota *cluster* dengan jarak yang terdekat. Dengan membandingkan hasil antara *cluster* dan diambil yang paling kecil. Berikut ini akan ditampilkan data matriks dengan inisialisasi 1 berarti data berada di *cluster* tersebut. Berikut adalah tabel hasil pengelompokkan dari iterasi pertama dapat dilihat pada Tabel 7.
5. Setelah melakukan pengelompokkan *cluster* menggunakan *cluster* awal, tahap selanjutnya yaitu mengelompokkan setiap *cluster* dengan membuat atau menentukan *centroid* terbarunya. Perhitungan ditunjukkan oleh Tabel 8.
6. Setelah mengelompokkan dan melakukan perhitungan data dengan *cluster* maka menghasilkan sebuah *centroid* yang baru. *Centroid* baru ditunjukkan oleh Tabel 9.
7. Setelah *centroid* baru telah terbentuk, proses selanjutnya yaitu kembali ke tahap sebelumnya pada proses perhitungan jarak pada data terhadap suatu pusat *cluster* menggunakan *centroid* yang baru. Perhitungan tersebut dilakukan sampai *cluster* tidak berubah.

**Tabel 7.** Matriks Hasil Iterasi Pertama

No	C1	C2	C3
1		1	-
2		1	-
3	1		-
4		1	-
5		1	-
6	1		-
7	1		-
8		1	-
9	1		-
10	1		-

**Tabel 8.** Pengelompokan Data Pada *Cluster 1* dan *Cluster 2*

Data ke-	SEX	LAHIR	UPAH	PROGRAM	Cluster
1	76	1997	2.710.000	539	2
2	76	1984	2.710.000	1075	2
4	80	1980	2.710.000	1075	2
5	76	1993	2.710.000	845	2
8	76	1972	2.710.000	539	2
...	...	...	...	...	...
197	80	1980	2.710.000	539	2
Centroid baru	78,019	1983,761	2.539.523,81	736,5809524	

Data ke-	SEX	LAHIR	UPAH	PROGRAM	Cluster
3	80	1990	2.950.000	539	1
6	80	1970	3.000.000	539	1
7	76	1986	3.000.000	1075	1
9	76	1974	2.950.000	539	1
10	80	1989	2.950.000	539	1
...	...	...	...	...	...
200	76	1989	3.000.000	539	1
Centroid baru	77,347	1980,905	2.961.052,632	712,2421053	

**Tabel 9.** Centroid Baru

	SEX	LAHIR	UPAH	PROGRAM
C1	77,019	1980.905263	2961052,632	712,5809524
C2	78,347	1983.761905	2539523.81	736,2421053

**Tabel 10.** Tabel *Cluster* Iterasi 1-2

Data ke -	Cluster		Posisi
	Iterasi ke - 1	Iterasi ke - 2	
1	2	2	Tidak Berubah
2	2	2	Tidak Berubah
3	1	1	Tidak Berubah
...	...	...	...
170	2	2	Tidak Berubah
171	2	2	Tidak Berubah
172	2	2	Tidak Berubah
...	...	...	...
198	1	1	Tidak Berubah
...	...	...	...
200	1	1	Tidak Berubah

Setelah melakukan dua kali iterasi dan ternyata pada iterasi ke-2 tidak ada *cluster* yang berubah ke *cluster* yang lain. Oleh karena itu, proses telah dihentikan karena *cluster* tidak ada yang berubah.

### G. Analisa Hasil Cluster

Dalam penelitian ini peneliti mengambil 2 *cluster*, karena setelah dilakukan percobaan dengan 3 *cluster* ternyata hasilnya tidak berpengaruh. Sehingga penelitian ini hanya mengambil 2 *cluster* saja. Kemudian setelah melakukan dua kali iterasi, perhitungan jarak data dan mencari hasil dari *cluster* pada setiap data sampai *cluster* tidak ada yang berubah. Diperoleh hasil dari titik pusat yang terakhir (*centroid* terakhir):

**Tabel 11.** Titik Pusat Cluster

	SEX	LAHIR	UPAH	PROGRAM
C1	78,019	1983,761	2.961.052,63	736,5809524
C2	77,347	1980,905	2.539.523.81	712,2421053

Tabel 11 menunjukkan hasil dari titik pusat akhir. Titik pusat akhirnya ada dua yaitu C1 dan C2 yang mewakili dari masing-masing *cluster*. *Cluster* tersebut adalah anggota dari C1 dan C2 yang mempunyai jarak terdekat dari titik pusat *cluster*. Setelah titik pusat akhir terbentuk, tahap selanjutnya yaitu menginterpretasikan hasil dari masing-masing *cluster*. Jika dilihat dari kolom titik pusat pada tabel di atas, masing-masing *cluster* memiliki karakteristik. Untuk *cluster* yang pertama termasuk dalam *customer* yang memiliki upah tinggi dengan kecenderungan memilih program JKM, JHT. Untuk *cluster* kedua termasuk dalam *customer* yang memiliki upah rendah dengan kecenderungan memilih program JKK, JKM. Setelah mengetahui hasil yang diperoleh dari *centroid*, diharapkan pada penelitian selanjutnya *centroid* tersebut dapat diklasifikasikan ataupun dilakukan pelabelan sesuai hasil analisa. Diketahui hasil dari C1 dapat dilabelkan menjadi *customer* berjenis kelamin laki-laki yang lahir pada tahun 1983 dengan gaji sekitar 2.961.052 dapat memilih program jaminan “JKM, JHT” dan C2 dapat dilabelkan menjadi *customer* yang berjenis kelamin laki-laki yang lahir tahun 1980 dengan gaji sekitar 2.539.523 dapat memilih program jaminan “JKK, JKM”.

## 5. Penutup

### A. Kesimpulan

Kesimpulan dari penelitian *customer profiling* yaitu *cluster* pertama termasuk dalam *customer* yang memiliki upah tinggi dengan kecenderungan memilih program jaminan “JKM, JHT” dan untuk *cluster* kedua termasuk dalam *customer* yang memiliki upah rendah dengan kecenderungan memilih program jaminan “JKK, JKM”.

### Daftar Pustaka

- [1] P. Perkembangan, M. Ekonomi, I. Terhadap, P. Perkembangan, M. Ekonomi, and I. Terhadap, “No Title.”
- [2] A. Zulfa and I. Dwimawanti, “Analisis Kinerja BPJS Ketenagakerjaan Kantor Cabang Semarang Pemuda Untuk Meningkatkan Perluasan Kepesertaan Sektor Informal,” pp. 1–13, 2014.
- [3] I. Hukum and H. Ekonomi, “Fakultas hukum universitas atma jaya yogyakarta 2016,” 2016.

- [4] A. V. Septiarani, N. W. Setyanto, and R. Y. Efranto, “Ketenagakerjaan Dengan Metode Fuzzy-Servqual Dan ( Studi Kasus BPJS Ketenagakerjaan Cabang Malang ) Service Quality Improvements Of BPJS Ketenagakerjaan Using Fuzzy-Servqual And PGCV Index ( Study Case : BPJS Ketenagakerjaan Cabang Malang ),” pp. 859–870.
- [5] “Jurnal pelaksanaan bpjs ketenagakerjaan pada pekerja gardena department store dan supermarket di kota yogyakarta,” 2017.
- [6] P. T. Upadhyay, “Customer Profiling and Segmentation using Data Mining Techniques.”
- [7] A. Bastian *et al.*, “No Title,” no. 1, pp. 26–32.
- [8] Y. Darmi and A. Setiawan, “Penerapan Metode Clustering K-Means Dalam,” vol. 12, no. 2, pp. 148–157, 2016.
- [9] M. Distance, “Customer Profiling Dari Produk Asuransi Dengan Menggunakan Metode K-Means Manhattan Distance,” no. x, pp. 1–15, 1978.
- [10] S. Ketenagakerjaan and D. I. Indonesia, “Implikasi Perubahan PT. Jamsostek (Persero) Menjadi Badan Penyelenggara Jaminan Sosial (BPJS) Terhadap Jaminan Sosial Ketenagakerjaan Di Indonesia Hj. Suzanalisa 1,” vol. 15, no. 3, pp. 119–128, 2015.
- [11] J. Informatika, W. Mega, and P. Dhuhita, “Clustering Menggunakan Metode K-Means Untuk,” vol. 15, no. 2, 2015.
- [12] D. B. Atyanto, M. Er, and R. Soelaiman, “Customer Profiling Dengan Menggunakan K-Means Clustering Untuk Mendukung Pengambilan Keputusan Strategis Di PT Pelabuhan Indonesia III,” 2011.