

PENERAPAN *K-MEANS CLUSTERING* UNTUK MEMPREDIKSI MINAT NASABAH PADA PT. ASURANSI JIWA BERSAMA 1912 BUMIPUTERA PRABUMULIH

Lhorend Mutiara Pratiwi¹, Diana², Eka Puji Agustini³

Dosen Universitas Bina Darma^{2,3}, Mahasiswa Universitas Bina Darma¹

Jalan Jendral Hmad Yani No. 12 Palembang

e-mail: lhorend.mutiara@gmail.com, diana@binadarma.ac.id,

ekapujiagustini@binadarma.ac.id

Abstract : *Data mining is a term used to find hidden knowledge idalam database. Clustering is the process of grouping data set into groups so that objects in a group have much in common and have many objects grouped perbedaandengan lain. K-means clustering is a method of non-hierarchical clustering of data is that the data in the form of one or more clusters / groups. The application of data mining using the stage Knowledge Discovery in Databases (KDD) consisting of Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Evaluation, Presentation and software which is used is Weka. Dari this study showed that the prospective customer who is as farmers, with an average income of 1.5 million, have interests that are high enough to become customers of insurance with the selected type of insurance is MitraBeasiswa, especially Gelumbang region. So with this research can help the marketing of PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih in the decision to promote to the public.*

Keywords : *Life Insurance, Data Mining, Clustering, K-Means, KDD, Weka.*

Abstrak : *Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi idalam database. Clustering merupakan proses pengelompokkan kumpulan data menjadi beberapa kelompok sehingga objek didalam suatu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaandengan objek dikelompok lain.K-means clustering adalah salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Penerapan data mining ini menggunakan tahapan Knowledge Discovery in Database (KDD) yang terdiri dari Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Evaluation, Presentation dan software yangdigunakan adalah Weka.Dari penelitian ini didapatkan hasil bahwa calon nasabah yang berprofesi sebagai petani, dengan pendapatan rata-rata 1.500.000, memiliki minat yang cukup tinggi untuk menjadi nasabah asuransi dengan jenis asuransi yang dipilih adalah MitraBeasiswa, khususnya diwilayah Gelumbang. Sehingga dengan adanya penelitian ini dapat membantu pihak marketing PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih dalam pengambilan keputusan untuk melakukan promosi kepada masyarakat.*

Kata Kunci : *Asuransi Jiwa, Data Mining, Clustering, K-Means, KDD, Weka.*

1. PENDAHULUAN

4.

5. Asuransi merupakan sarana finansial dalam tata kehidupan rumah tangga, baik dalam menghadapi resiko yang mendasar seperti resiko kematian atau dalam menghadapi resiko atas harta benda yang dimiliki. Usaha perasuransian sebagai

2. 1.1 Latar Belakang

3.

salah satu lembaga keuangan menjadi penting peranannya karena dari kegiatan perlindungan resiko, perusahaan asuransi menghimpun dana masyarakat dari penerimaan premi, yang kemudian menginvestasikan dana itu ke dalam berbagai kegiatan ekonomi perusahaan. Dengan peranan

asuransi tersebut dalam perkembangan pembangunan ekonomi yang semakin meningkat, maka semakin terasa kebutuhan akan hadirnya industri perusahaan asuransi yang kuat dan dapat diandalkan.

6. Oleh karena nasabah PT. AJB 1912 Prabumulih berasal dari berbagai daerah, maka dibutuhkan strategi khusus oleh bagian *marketing* perusahaan dalam melakukan pemasaran untuk mencari calon nasabah agar promosi yang dilakukan lebih efektif dan efisien. Karena sebelumnya bagian *marketing* perusahaan mengalami kesulitan dalam mempromosikan beberapa jenis asuransi yang ditawarkan kepada masyarakat.

7. Dari permasalahan diatas maka di buatlah suatu penelitian dengan memanfaatkan data-data nasabah pemegang polis dari tahun 2015 sampai dengan tahun 2016 seperti nama nasabah pemegang polis, daerah asal, jenis asuransi yang diambil dan jumlah pendapatan pekerjaan perbulan, dimana data-data tersebut akan diproses *data mining* dengan metode *clustering* yang nantinya akan menghasilkan informasi asuransi apa yang paling diminati masyarakat.

8. Untuk mendapatkan informasi yang berupa prediksi melalui proses *data mining* maka dalam penelitian ini menggunakan proses algoritma *k-means* dimana dalam proses ini menggunakan tahapan *data cleaning*, *data integration*, *data selection*, *data transformation*, *data mining*, *evaluation* dan *presentation* dan dalam proses *data mining* ini nantinya akan menggunakan *software Weka*. Selain itu *software Weka* juga memiliki kemampuan untuk mengelola data dengan metode *clustering* (pengelompokkan) yang dipakai dalam penelitian ini.

9. Penggunaan *data mining* sebagai bagian sebuah sistem informasi yang sangat penting untuk menjamin ketersediaan layanan bagi

penggunaannya. Aset atau sumber daya bagi instansi atau organisasi yang sangat berharga bahkan bisa dikatakan sangat penting, yaitu berupa data atau informasi, kerusakan terhadap data dapat mengancam kelangsungan hidup dari perusahaan tersebut. *Data mining* memiliki kelebihan dan kekurangan untuk mengetahui kedua hal tersebut perlu adanya rencana evaluasi pada sistem berjalan saat ini karena untuk menanggapi segala permasalahan yang ada agar dapat diketahui kelemahan yang akan diperbaiki menjadi lebih baik.

10. Dari uraian diatas penulis tertarik untuk mengangkat permasalahan ini ke dalam skripsi nya yang berjudul “**Penerapan K-Means Clustering untuk Memprediksi Minat Nasabah pada PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih**”.

10.2 Perumusan Masalah

11. Berdasarkan latar belakang diatas maka penulis merumuskan masalah dalam penelitian ini yaitu “Bagaimana cara menerapkan *k-means clustering* dalam memprediksi minat nasabah ?”

11.2 Batasan Masalah

12. Pada penelitian ini penulis hanya membatasi hanya pada pemanfaatan data nasabah asuransi pada tahun 2015 sampai dengan tahun 2016 berdasarkan data yang telah disediakan oleh PT. AJB Bumiputera 1912 Prabumulih untuk memprediksi minat nasabah selanjutnya.

12.2 Tujuan dan Manfaat

12.2.1 Tujuan

13. Penelitian ini bertujuan menerapkan teknik *k-means clustering* dalam *data mining* untuk menampilkan informasi dalam memprediksi minat nasabah.

13.2.1 Manfaat

14. Adapun manfaat dari penelitian ini adalah :

1. Memberikan gambaran dalam pengambilan keputusan.
2. Memberikan informasi yang belum diketahui yang sebelumnya masih tersembunyi di dalam gudang data sehingga menjadi informasi yang penting yang dapat membantu PT. AJB Bumiputera 1912 Prabumulih dalam melakukan promosi kepada masyarakat.

14.2 Metodologi Penelitian

15. 1.5.1 Tempat dan Waktu Penelitian

16. Waktu penelitian ini dilaksanakan selama 3 bulan yaitu mulai dari bulan November 2016 sampai bulan Januari 2017. Pada PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih.

17. 1.5.2 Metode Penelitian

18. Dalam melakukan penelitian ini, penulis menggunakan metode *deskriptif*. Metode *deskriptif* merupakan suatu metode yang meneliti status kelompok manusia, suatu objek, suatu kondisi, suatu pemikiran ataupun suatu kelas peristiwa pada masa sekarang. Maka dari itu penulis menggunakan metode *deskriptif* dikarenakan permasalahan yang sedang dilakukan penulis sekarang berdasarkan data yang sebenarnya atau yang bersifat fakta yaitu mengenai data nasabah pada tahun 2016 pada PT. AJB Bumiputera 1912 Prabumulih.

19. 1.5.3 Metode Pengumpulan Data

20. Metode pengumpulan data yang tepat yaitu dengan mempertimbangkan penggunaannya berdasarkan jenis data dan

sumbernya. Data yang objektif dan relevan dengan pokok permasalahan penelitian merupakan indikator keberhasilan suatu penelitian. Pengumpulan data penelitian ini dilakukan dengan cara sebagai berikut :

1. Observasi

21. Metode pengumpulan data dengan cara mengadakan pengamatan secara langsung kepada objek penelitian mengenai data-data penelitian yang dibutuhkan langsung ke kantor PT. Asuransi Jiwa Bersama Bumiputera 1912 cabang kota Prabumulih.

2. Studi Pustaka

22. Mengumpulkan data dengan cara mencari dan mempelajari data-data atau buku-buku ataupun dari referensi lain yang berhubungan dengan penulisan laporan penelitian.

23. 1.5.4 Data Penelitian

24. Dalam penelitian ini penulis menggunakan data primer dan data sekunder yang akan dijelaskan sebagai berikut :

1. Data Primer

25. Data primer adalah data yang langsung didapat dari tempat penelitian, yang dalam hal ini adalah PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih. Data tersebut berupa data nasabah pada 2015 sampai dengan tahun 2016 dan data sejarah PT. Asuransi Jiwa Bersama 1912. Data nasabah terdiri dari nama nasabah, nomor polis, jenis asuransi yang dipilih, pekerjaan, pendapatan pekerjaan perbulan, dan alamat.

2. Data Sekunder

26. Data Sekunder adalah data yang diperoleh selain dari PT. Asuransi Jiwa Bersama Bumiputera

itu sendiri juga data yang penulis peroleh dari buku-buku atau situs internet yang erat hubungannya dengan penulisan skripsi ini.

27. 1.5.5 Prosedur Penelitian

28. Adapun untuk menganalisis data dalam penerapan *data mining* ini menggunakan tahapan *Knowledge Discovery in Database (KDD)* yang terdiri dari beberapa tahapan (Sigit.P, 2013), yaitu *Data Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Evaluation, dan Preentation*.

29.

2.1. Landasan Teori

2.1.1 Pengertian *Data Mining*

30. *Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar (Turban et al : 2005).

31. 2.1.2 *Clustering*

32. Menurut Han dan Kamber (2011), *Clustering* adalah proses pengelompokan kumpulan data menjadi beberapa kelompok sehingga objek di dalam satu kelompok memiliki banyak kesamaan dan memiliki banyak perbedaan dengan objek dikelompok lain. Perbedaan dan persamaannya biasanya berdasarkan nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak. *Clustering* sendiri juga disebut *Unsupervised Classification*, karena *clustering*

lebih bersifat untuk dipelajari dan diperhatikan. *Cluster analysis* merupakan proses partisi satu set objek data ke dalam himpunan bagian. Setiap himpunan bagian adalah *cluster*, sehingga objek yang di dalam *cluster* mirip satu sama dengan yang lainnya, dan mempunyai perbedaan dengan objek dari *cluster* yang lain. Partisi tidak dilakukan dengan manual tetapi dengan algoritma *clustering*. Oleh karena itu, *Clustering* sangat berguna dan bisa menemukan *group* yang tidak dikenal dalam data.

33. 2.1.3 *K-Means*

34. *K-Means Clustering* merupakan salah satu metode data *clustering* non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil (Agusta : 2007).

35. Menurut Santosa (2007), langkah-langkah melakukan *clustering* dengan metode *k-means* adalah sebagai berikut :

1. Pilih 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 diberi nilai awal dengan angka-angka random.
3. Alokasikan semua data/objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan kedua jarak 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 :

$$36. \quad D(I,j) = \dots (1)$$

37. Dimana :

38. $D(I,j)$ = Jarak data ke I ke pusat *cluster* j

39. X_{ki} = Data ke I pada atribut data ke k

40. X_{kj} = Titik pusat ke j pada atribut ke 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-satu nya 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.

41. 2.1.4 Weka

42. *Weka (Waikato Environment for Knowledge Analysis)* adalah aplikasi *data mining open source* berbasis *Java*. Aplikasi ini dikembangkan pertama kali oleh Universitas

Waikato di Selandia Baru. *Weka* terdiri dari koleksi algoritma *machine learning* yang dapat digunakan untuk melakukan generalisasi/formulasi dari sekumpulan *data sampling*.

43. Algoritma ini bisa diterapkan secara langsung kedalam dataset atau bisa juga dipanggil dari kode *java* kita sendiri. *Weka* memiliki *tools* untuk data *pre-processing*, *classification*, *regression*, *clustering*, *association rules*, dan *visualication*. Pada *weka* ada beberapa metode pemilihan variabel dari suatu dataset, diantaranya *BestFirst*, *ExhaustiveSearch*, *FCBFSearch*, *GeneticSearch*, *GreedyStepwise*, *RaceSearch*, *RandomSearch*, *Rankerdan Ranker Search*.

44. 2.1.5 Pengertian Asuransi

45. Berikut ini adalah pengertian asuransi menurut para ahli asuransi dan sumber lain, di antaranya adalah :

1. Menurut H.M.N Purwosutjipto asuransi adalah perjanjian timbal balik antara penutup (pengambil) asuransi dengan penanggung, dimana penutup (pengambil) asuransi mengikatkan diri selama jalannya pertanggungan membayar uang premi kepada penanggung, sedangkan penanggung sebagai akibat langsung dari meninggalnya orang yang jiwanya dipertanggungkan atau telah lampaunya suatu jangka waktu yang diperjanjikan, mengikatkan diri untuk membayar sejumlah uang tertentu kepada orang yang ditunjuk oleh penutup (pengambil) asuransi sebagai penikmatnya.

2. Pengertian Asuransi Jiwa Menurut UU No. 2 Tahun 1992, Asuransi Jiwa adalah perjanjian antara dua pihak atau lebih, pihak penanggung

mengikatkan diri kepada tertanggung dengan menerima premi asuransi untuk memberikan suatu pembayaran yang didasarkan atas meninggal atau hidupnya seseorang yang diasuransikan.

46. 4. ANALISIS DATA MINING

47. 4.1 Data Selection (Pemilihan Data)

48. *Data selection* merupakan langkah awal dalam melakukan proses *data mining*. Tahap ini dilakukan untuk mendapatkan kolom-kolom data yang tepat yang disebut dengan atribut. Langkah ini akan selalu dilakukan pertama kali dalam implementasi data mining. Data yang digunakan dalam penelitian ini berasal dari PT. Asuransi Jiwa Bersama Bumiputera yaitu data nasabah tahun 2016 kantor cabang Prabumulih. Format data yang digunakan adalah format .doc (*Word Document*). Atribut-atribut yang dipakai dalam proses *knowledge discovery in databases (KDD)* yaitu :

1. No Polis merupakan atribut yang berperan sebagai primary key, nomor resmi nasabah asuransi jika mereka sudah terdaftar secara sah.
2. Nama Pempol merupakan atribut yang menyatakan nama pemilik polis
3. Macas merupakan atribut yang menyatakan jenis asuransi yang dipilih
4. Pekerjaan merupakan atribut yang menyatakan pekerjaan yang dimiliki oleh pemegang polis
5. Pendapatan/Bulan merupakan atribut yang menyatakan pendapatan yang diperoleh oleh pemegang polis
6. Alamat merupakan atribut yang menyatakan lokasi atau tempat tinggal dimana pemegang polis

49. Dari atribut diatas untuk data nasabah pada tahun 2015-2016 sebelum dilakukan proses *data selection* adalah atribut no, atribut no polis, atribut nama pempol, atribut macas, atribut pekerjaan, atribut pendapatan/bulan, dan atribut alamat yang dapat dilihat seperti pada gambar 1 sebagai berikut

50.

No	No Polis	NamaPempol	Macas	Pekerjaan	Pendapatan /Bulan	Alamat
1	213103393310	SUDWANT	MitraMandiri	PNS	3.500.000	LUBAI ULU
2	207100189955	NGADIMAN WIDIK	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
3	213101544087	SUTERD	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
4	207102898835	ARWANI	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
5	208100689952	ABDUL KADIL	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
6	208100723438	ASMAKA BAHARIB	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
7	213101438472	RIYANTO	MitraMandiri	PNS	3.500.000	LUBAI ULU
8	214100131824	ADANI MALIK	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
9	213103383295	MARLENA	MitraMandiri	PNS	3.500.000	LUBAI ULU
10	213103391584	SITI FATHONAH	Mitra Guru	HONOREK	1.000.000	LUBAI ULU
11	207100313421	SINISA	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
12	213102883814	BARVONO	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
13	213102855523	SUDIRMAN	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
14	213102835498	SURKIRMAN	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
15	213102835860	SUDIRMAN	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
16	213102883829	WATINAH	Mitra Guru	HONOREK	1.000.000	LUBAI ULU
17	207100758651	SUDIRMAN	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
18	209102798075	SUPRATMAN	MitraCerdas	WIRASWASTA	2.500.000	LUBAI ULU
19	21310464488	IRIDI	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
20	213102835188	ROSDI JANUARDI	MitraCerdas	WIRASWASTA	2.500.000	LUBAI ULU
21	207100135227	EDU RIKARD	MitraCerdas	WIRASWASTA	2.500.000	LUBAI ULU
22	214102487813	PASOLILI	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
23	213103383295	TASMA KALSAFI	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
24	213103391697	LERIMARIANA W	MitraMandiri	PNS	3.500.000	LUBAI ULU

56.

57. Gambar 1. Data nasabah asuransi pada tahun 2015-2016 sebelum dilakukan proses *data selection*

58.

59. Setelah dilakukan proses *data selection* diatas, yaitu dengan mengurangi atribut yang tidak diperlukan untuk memudahkan pada saat melakukan tahap *data mining* selanjutnya. Dari proses tersebut maka didapatkan hasil nya menjadi atribut no, atribut macas, atribut pekerjaan, atribut pendapatan, dan atribut alamat yang dapat dilihat pada gambar 2 sebagai berikut:

No	Macas	Pekerjaan	Pendapatan /Bulan	Alamat
1	MitraMandiri	PNS	3.500.000	LUBAI ULU
2	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
3	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
4	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
5	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
6	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
7	MitraMandiri	PNS	3.500.000	LUBAI ULU
8	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
9	MitraMandiri	PNS	3.500.000	LUBAI ULU
10	Mitra Guru	HONOREK	1.000.000	LUBAI ULU
11	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
12	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
13	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
14	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
15	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
16	Mitra Guru	HONOREK	1.000.000	LUBAI ULU
17	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
18	MitraCerdas	WIRASWASTA	2.500.000	LUBAI ULU
19	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU
20	MitraCerdas	WIRASWASTA	2.500.000	LUBAI ULU
21	MitraCerdas	WIRASWASTA	2.500.000	LUBAI ULU
22	MitraBeasiswa	PETANI	1.500.000	LUBAI ULU

65. 4.2 Preprocessing

66. Tahap *pre-processing* meliputi tahapan integrasi atau penggabungan data seluruh nasabah asuransi dari semua agen yang berjumlah 9 dan juga pembersihan data untuk menghasilkan dataset yang bersih sehingga dapat digunakan

dalam tahap berikutnya yaitu *mining*. Berikut penjelasan dua proses diatas :

1. *Integrasi Data*, merupakan data yang diperoleh akan digabungkan menjadi satu tabel data yang besar sebagai fitur pelatihan nantinya untuk menjalankan proses algoritma *k-means*.

67. Hasil dari proses integrasi data tersebut dapat dilihat pada gambar 3 dibawah ini :

No	Nama	Pekerjaan	Pendapatan Bulanan	Alamat
1	MitraMandiri	PNS	3.500.000	LUBAI
2	MitraBeasiswa	PETANI	1.500.000	LUBAI
3	MitraBeasiswa	PETANI	1.500.000	LUBAI
4	MitraBeasiswa	PETANI	1.500.000	LUBAI
5	MitraBeasiswa	PETANI	1.500.000	LUBAI
6	MitraBeasiswa	PETANI	1.500.000	LUBAI
7	MitraMandiri	PNS	3.500.000	LUBAI
8	MitraBeasiswa	PETANI	1.500.000	LUBAI
9	MitraMandiri	PNS	3.500.000	LUBAI
10	Mitra Guru	HONORER	1.000.000	LUBAI
11	MitraBeasiswa	PETANI	1.500.000	LUBAI
12	MitraBeasiswa	PETANI	1.500.000	LUBAI
13	MitraBeasiswa	PETANI	1.500.000	LUBAI
14	MitraBeasiswa	PETANI	1.500.000	LUBAI
15	MitraBeasiswa	PETANI	1.500.000	LUBAI
16	Mitra Guru	HONORER	1.000.000	LUBAI
17	MitraBeasiswa	PETANI	1.500.000	LUBAI
18	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
19	MitraBeasiswa	PETANI	1.500.000	LUBAI
20	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
21	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
22	MitraBeasiswa	PETANI	1.500.000	LUBAI
23	MitraBeasiswa	PETANI	1.500.000	LUBAI
24	MitraMandiri	PNS	3.500.000	LUBAI

74. Gambar 3. Integrasi data agen menjadi satu data besar

75. 2. *Cleaning Data*, tahap *data cleaning* merupakan tahap awal dari proses *KDD*. Data yang telah digabung akan dilakukan pembersihan, membuang data yang kosong dan memastikan data tersebut relevan atau terkait satu sama lain. Juga pada tahap ini kita harus cermat terhadap adanya redundansi data jika ditemukan. Dan dipastikan tiap-tiap baris data harus bersifat unik.

76. Berikut adalah gambar dimana sebelum dilakukan proses *cleaning data* yang masih terdapat data yang tidak relavan dan dapat dilihat pada gambar 4 dibawah ini :

No	No Polis	NamaPempol	Macas	Pekerjaan	Pendapatan Bulanan	Alamat
77	209100417686	NenPericiArtini	MitraMandiri	PNS	3.500.000	LUBAI
78	209102264530	Reni Yunani, S.P	MitraMandiri	PNS	3.500.000	LUBAI
79	209102738738	JohnyElezer T	MitraBeasiswa	PETANI	1.500.000	LUBAI
80	209103258003	Fatimah	Mitra Guru	HONORER	1.000.000	LUBAI
81	211101075282	FitriSuryani	MitraBeasiswa	PETANI	1.500.000	LUBAI
82	211101075296	Yuni	MitraBeasiswa	PETANI	1.500.000	LUBAI
83	211101285804	Karyanto	Mitra Guru	HONORER	1.000.000	LUBAI
84	211101286211	Hendriwanto	MitraBeasiswa	PETANI	1.500.000	LUBAI
85	212100197565	Rusli	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
86	212100197769	Mukam	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
87	212101101176	Lela Diana	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
88	212101101244	HendroYulianto	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
89	21210268753	Nuhadi	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
90	212102739396	Zubaidah	Mitra Guru	HONORER	1.000.000	LUBAI
91	213100867564	DismalaDewi	MitraMandiri	PNS	3.500.000	LUBAI
92	213102237762	SyamsiAlam	MitraBeasiswa	PETANI	1.500.000	LUBAI
93	213102489017	Imata Takasari	MitraBeasiswa	PETANI	1.500.000	LUBAI
94	213102489167	Eva Ediza ZA	Mitra Guru	HONORER	1.000.000	LUBAI
95	213102489271	EmmySunilawati	Mitra Guru	HONORER	1.000.000	LUBAI
96	213102567105	StinMasitoh	MitraMandiri	PNS	3.500.000	LUBAI
97	21400077375	NurAfriyanti	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
98	214101501308	Benny Irawan	MitraBeasiswa	PETANI	1.500.000	LUBAI
99	215100782477	EllyNovitasari	MitraMandiri	PNS	3.500.000	LUBAI

86.

87. Gambar 4. Data Sebelum Dilakukan *Data Cleaning*

88. Setelah dilakukan proses *cleaning data*, maka dapat dilihat bahwa data yang tidak relevan seperti gambar diatas sudah tidak ada lagi. Hasil dari proses tersebut dapat dilihat pada gambar 5 dibawah ini :

No	No Polis	NamaPempol	Macas	Pekerjaan	Pendapatan Bulanan	Alamat
1	209100417686	NenPericiArtini	MitraMandiri	PNS	3.500.000	LUBAI
2	209102264530	Reni Yunani, S.P	MitraMandiri	PNS	3.500.000	LUBAI
3	209102738738	JohnyElezer T	MitraBeasiswa	PETANI	1.500.000	LUBAI
4	209103258003	Fatimah	Mitra Guru	HONORER	1.000.000	LUBAI
5	211101075282	FitriSuryani	MitraBeasiswa	PETANI	1.500.000	LUBAI
6	211101075296	Yuni	MitraBeasiswa	PETANI	1.500.000	LUBAI
7	211101285804	Karyanto	Mitra Guru	HONORER	1.000.000	LUBAI
8	211101286211	Hendriwanto	MitraBeasiswa	PETANI	1.500.000	LUBAI
9	212100197565	Rusli	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
10	212101101176	Lela Diana	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
11	212101101244	HendroYulianto	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
12	21210268753	Nuhadi	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
13	212102739396	Zubaidah	Mitra Guru	HONORER	1.000.000	LUBAI
14	213100867564	DismalaDewi	MitraMandiri	PNS	3.500.000	LUBAI
15	213102237762	SyamsiAlam	MitraBeasiswa	PETANI	1.500.000	LUBAI
16	213102489017	Imata Takasari	MitraBeasiswa	PETANI	1.500.000	LUBAI
17	213102489167	Eva Ediza ZA	Mitra Guru	HONORER	1.000.000	LUBAI
18	213102489271	EmmySunilawati	Mitra Guru	HONORER	1.000.000	LUBAI
19	213102567105	StinMasitoh	MitraMandiri	PNS	3.500.000	LUBAI
20	21400077375	NurAfriyanti	MitraCerdas	WIRASWASTA	2.500.000	LUBAI
21	214101501308	Benny Irawan	MitraBeasiswa	PETANI	1.500.000	LUBAI
22	215100782477	EllyNovitasari	MitraMandiri	PNS	3.500.000	LUBAI

96. Gambar 5. Data Setelah Dilakukan *Data Cleaning*

97. 4.3 *Data Transformation*

98. Tahap *transformation* data merupakan tahap merubah data kedalam bentuk yang sesuai untuk di *mining* . Tahap ini juga berguna untuk membentuk format data yang diterima di perangkat lunak *data mining* yang akan memprosesnya. Biasanya perangkat lunak *data mining* menggunakan format data .csv atau excel. Berhubung data yang diperoleh berbentuk dokumen kata (.doc), maka terlebih dahulu kita akan mengubahnya ke dalam format excel atau csv. Hal ini dapat dilihat pada gambar 6 dibawah ini :

99.

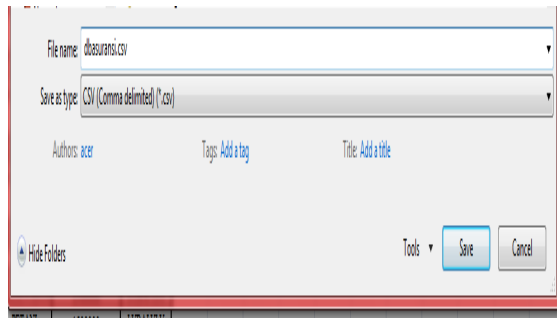
100.

101.

108.

109. Setelah data diubah ke dalam format *excel*, maka selanjutnya mengubah data dari format *excel* kedalam format *.csv* dan disimpan dan hasil nya dapat dilihat gambar 7 dibawah ini :

110.



111. **Gambar 7. Mengubah *dataset* dari format *excel* ke format *.csv***

112.

4.4. Analisa *K-Means*

113. Agar data nasabah dapat diolah dengan metode *k-means clustering* , maka data yang berjenis data nominal seperti macas, jenis pekerjaan, dan alamat harus di inisialisasikan terlebih dahulu dalam bentuk angka.

114. Untuk melakukan inisialisasi macas atau jenis asuransi yang dipilih nasabah dilakukan dengan langkah-langkah sebagai berikut :

1. Macas tersebut diurutkan dari yang terbesar berdasarkan frekuensi jenis asuransi yang dipilih nasabah.
2. Kemudian macas yang memiliki frekuensi terbesar diberi inisial dengan angka 1, dan jenis asuransi yang memiliki frekuensi terbesar kedua diberi inisial dengan angka 2, begitu seterusnya hingga jenis asuransi dengan

frekuensi paling sedikit. Hasil dari inisialisasi macas dapat dilihat pada tabel 1.

115. Maca s	116. Frek	117. Inisial
118. Mitra Beasiswa	119. 78	120. 1
121. Mitra Mandiri	122. 64	123. 2
124. Mitra Cerdas	125. 39	126. 3
127. Mitra Guru	128. 28	129. 4
130. Mitra Melati	131. 9	132. 5
133. Mitra Mandiri	134. 1	135. 6

136.

3. Kemudian jenis pekerjaan juga perlu di inisialisasikan kedalam bentuk angka. Seperti pada macas, pada jenis pekerjaan juga diberi inisialisasikan berdasarkan frekuensi nasabah pada jenis pekerjaan tersebut.

137. Hasil inisialisasi jurusan tersebut dapat dilihat pada

138. tabel 2 dibawah ini :

139. Pekerjaan	140. Frek	141. Inisial
142. Petani	143. 77	144. 1
145. PNS	146. 57	147. 2
148. Wiraswasta	149. 38	150. 3
151. Honorer	152. 29	153. 4
154. Wirasusaha	155. 6	156. 5
157. SPG	158. 5	159. 6
160. IRT	161. 2	162. 7
163. Perawat	164. 2	165. 8
166. PNS Dokter	167. 1	168. 9

169. Pol isi	170. 1	171. 1 0
172. Wi raswast a	173. 1	174. 1 1

175.

176.

177.

178.

179.

180.

181.

182.

183.

184.

185.

186. Tabel 2. Inisialisasi Jenis Pekerjaan

Nasabah

4. Kemudian menginisialisasikan alamat kedalam bentuk angka. Seperti pada jenis pekerjaan, pada alamat juga diberi inisialisasi berdasarkan frekuensi alamat nasabah tersebut.

187. Hasil inisialisasi alamat tersebut dapat dilihat pada tabel 3 dibawah ini :

188.

189.

190.

191.

192. Alamat	193. Frek	194. Inisial
195. Gelumba ng	196. 79	197. 1
198. Kelekar	199. 33	200. 2
201. Lubai Ulu	202. 30	203. 3
204. Tan jung Raman	205. 29	206. 4
207. Lubai	208. 22	209. 5
210. Lembak	211. 12	212. 6
213. Ramban g Lubai	214. 9	215. 7
216. Lembak	217. 1	218. 0

Timur	5	8
-------	---	---

219.

220. Tabel 3. Inisialisasi Alamat Nasabah

221. Setelah semua data nasabah pada tahun 2016 di inisialisasi kedalam bentuk angka, maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *k-means clustering*. Untuk dapat melakukan pengelompokkan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah yaitu :

1. Menentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data-data nasabah yang ada akan dikelompokkan menjadi empat *cluster*.

2. Tentukan titik pusat awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan secara *random* dan didapat titik pusat dari setiap *cluster* dapat dilihat pada tabel 4 sebagai berikut :

222. Centroid	223. Macas	224. Pekerjaan	225. Pend/bl n	226. Alamat
227. cluster 0	228. 2	229. 2	230. 3500000	231. 3
232. cluster 1	233. 1	234. 1	235. 1500000	236. 3
237. cluster 2	238. 2	239. 2	240. 3500000	241. 5
242. cluster 3	243. 3	244. 3	245. 2500000	246. 7

247.

248. Tabel 4. Titik Pusat Awal Setiap Cluster

3. Tempatkan setiap data pada *cluster*. Dalam penelitian ini digunakan metode *simple k-means* untuk mengalokasikan setiap data kedalam suatu *cluster*, sehingga data akan dimasukkan dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung

jarak setiap data dengan titik pusat setiap *cluster*:

249. Sebagai contoh akan dihitung jarak dari data nasabah pertama kepusat *cluster* pertama :

$$250. D(1,0)=$$

$$251. = 0$$

252. Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data nasabah pertama dengan pusat *cluster* pertama adalah 0.

253. Jarak dari data nasabah pertama ke pusat *cluster* kedua :

$$254. D(1,1)= 2$$

255. Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data nasabah pertama dengan pusat *cluster* kedua adalah 2.

256. Jarak dari data nasabah pertama ke pusat *cluster* ketiga :

$$257. D(1,2)= 1000000$$

258. Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data nasabah pertama dengan pusat *cluster* ketiga adalah 1000000.

259. Jarak dari data nasabah pertama ke pusat *cluster* keempat :

$$260. D(1,1)= 2000000$$

261. Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data nasabah pertama dengan pusat *cluster* ketiga adalah 2000000.

262. Berdasarkan hasil perhitungan diatas dapat disimpulkan bahwa jarak data nasabah pertama yang paling dekat adalah dengan *cluster* 0, sehingga data nasabah pertama dimasukkan ke

dalam *cluster* 0. Berikut adalah tampilan sebagian hasil perhitungan untuk 219 data nasabah dapat dilihat pada gambar 8 dibawah ini :

No	Macas	Pekerjaan	Pend/Bln	Alamat	C0	C1	C2	C3	Terkecil	Cluste
1	2	2	3500000	3	0	2	1000000	2000000	C0	0
2	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
3	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
4	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
5	1	1	1500000	3	1500000	2000000	1000000	2	C3	3
6	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
7	2	2	3500000	3	0	2	1000000	2000000	C0	0
8	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
9	2	2	3500000	3	0	2	1000000	2000000	C0	0
10	4	4	1000000	3	2500000	2500000	1500000	500000	C3	3
11	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
12	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
13	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
14	1	1	1500000	3	2000000	1000000	1000000	2	C3	3
15	1	1	1500000	3	2000000	1000000	1000000	2	C3	3
16	4	4	1000000	3	2500000	2500000	1500000	500000	C3	3
17	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
18	3	3	2500000	3	1000000	1000000	4	1000000	C2	2
19	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
20	3	3	2500000	3	1000000	1000000	4	1000000	C2	2
21	3	3	2500000	3	1000000	1000000	4	1000000	C2	2
22	1	1	1500000	3	2000000	2000000	1000000	2	C3	3
23	1	1	1500000	3	2000000	2000000	1000000	2	C3	3

270. **Gambar 8. Hasil perhitungan setiap data ke setiap *cluster***

271.

4. Setelah semua data diletakkan kedalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut.

5. Setelah didapatkan titik pusat yang baru dari setiap *cluster*, lakukan kembali dari langkah ketiga hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari suatu *cluster* ke *cluster* yang lain. Dalam penelitian ini iterasi *clustering* data nasabah terjadi sebanyak 3 kali iterasi. Pada iterasi ketiga titik pusat dari setiap *cluster* sudah tidak berubah dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* lain.

272. Setelah dilakukan perhitungan seperti yang diuraikan diatas, maka dapat diketahui hasil dari analisis perhitungan *k-means* menggunakan cara manual. Dan hasil dari perhitungan manual tersebut adalah sebagai berikut ini :

1. *Cluster* 0

273. *Cluster* 0 terdiri dari 52 orang,

No	Macas	Pekerjaan	Pend/bulan	Alamat/banyaknya
1.	MitraMandiri	PNS	3.500.000	Lubai Ulu = 6 orang
2.	MitraMandiri	PNS	3.500.000	Lubai = 6 orang
3.	MitraMandiri	PNS	3.500.000	Rambang Lubai = 2 orang
4.	MitraMandiri	PNS	3.500.000	Lembak = 2 orang
5.	MitraMandiri	PNS	3.500.000	Lembak Timur = 2 orang
6.	MitraMandiri	PNS	3.500.000	Gelumbang = 19 orang
7.	MitraMandiri	Polisi	5.000.000	Gelumbang = 1 orang
8.	MitraMandiri	Wirausaha	3.000.000	Gelumbang = 3 orang
9.	MitraMandiri	PNS	3.500.000	Kelekar = 11 orang

275.

276.

277.

278.

279. Gambar 9. Hasil analisis clustering pada cluster 0

280. Dari hasil perhitungan diatas pada *cluster 0* dapat disimpulkan bahwa karakteristik nasabah pada *cluster 0* didominasi oleh nasabah yang memilih macas Mitra Mandiri dengan jenis pekerjaan PNS, berpenghasilan Rp. 3.500.000 perbulan dan beralamat di Gelumbang.

2. Cluster 1

281. *Cluster 1* terdiri dari 109 orang, dapat dilihat pada gambar 9 dibawah ini :

282.

1.	MitraBeasiswa	Petani	1.500.000	Lubai Ulu = 22 orang
2.	MitraGuru	Honoror	1.000.000	Lubai Ulu = 2 orang
3.	MitraBeasiswa	Petani	1.500.000	Lubai = 6 orang
4.	MitraGuru	Honoror	1.000.000	Lubai =6 orang
5.	MitraBeasiswa	Petani	1.500.000	RambangLubai = 6 orang
6.	Mitra Guru	Honoror	1.000.000	RambangLubai = 1 orang
7.	MitraBeasiswa	Petani	1.500.000	Gelumbang = 20 orang
8.	MitraBeasiswa	Honoror	1.000.000	Gelumbang = 9 orang
9.	Mitra Melati	IRT	800.000	Gelumbang= 2 orang
10.	Mitra Melati	SPG	1.200.000	Gelumbang = 2 orang
11.	MitraBeasiswa	Petani	1.500.000	Tanjung Raman = 10 orang
12.	MitraBeasiswa	Petani	1.500.000	Kelekar = 9 orang
13.	MitraMelati	SPG	1.200.000	Kelekar = 1 orang
14.	MitraGuru	Honoror	1.000.000	Kelekar = 13 orang

283. Gambar 9.Hasil analisis clustering pada cluster 1

284. Dari hasil perhitungan diatas pada *cluster 1* dapat disimpulkan bahwa karakteristik nasabah pada *cluster 1* didominasi oleh nasabah yang memilih macas Mitra Beasiswa

dengan jenis pekerjaan Petani, berpenghasilan Rp. 1.500.000 perbulan dan beralamat di Lubai Ulu.

3. Cluster 2

285. *Cluster 2* terdiri dari 16 orang, dapat dilihat pada gambar 10 dibawah ini :

1.	MitraMandiri	PNS	3.500.000	TanjungRaman = 9 orang
2.	MitraGuru	Honoror	1.000.000	TanjungRaman = 3 orang
3.	MitraMandiri	Wirausaha	3.000.000	TanjungRaman = 2 orang
4.	MitraMelati	SPG	1.200.000	TanjungRaman = 1 orang
5.	MitraMandiri	Wirausaha	3.000.000	Kelekar = 1 orang

291. Gambar 10. Hasil analisis clustering pada cluster 2

292. Dari hasil perhitungan diatas pada *cluster 2* dapat disimpulkan bahwa karakteristik nasabah pada *cluster 2* didominasi oleh nasabah yang memilih macas Mitra Mandiri dengan jenis pekerjaan PNS, berpenghasilan Rp. 3.500.000 perbulan dan beralamat di Tanjung Raman.

4. Cluster 3

293. *Cluster 3* terdiri dari 42 orang, dapat dilihat pada gambar 11 dibawah ini :

No	Macas	Pekerjaan	Penghasilan	Alamat/Umurnya
1.	MitraCerdas	wiraswasta	2.500.000	Lubai Ulu = 3 orang
2.	MitraCerdas	wiraswasta	2.500.000	Lubai = 5 orang
3.	MitraCerdas	wiraswasta	2.500.000	Rambang Lubai = 3 orang
4.	MitraCerdas	wiraswasta	2.500.000	Lembak = 7 orang
5.	MitraCerdas	wiraswasta	2.500.000	Lembak Timur = 1 orang
6.	MitraCerdas	wiraswasta	2.500.000	Gelumbang = 14 orang
7.	MitraMelati	Perawat	2.000.000	Gelumbang = 2 orang
8.	MitraCerdas	wiraswasta	2.500.000	Tanjung Raman = 4 orang
9.	MitraCerdas	wiraswasta	2.500.000	Kelekar = 3 orang

298. **Gambar 11. Hasil analisis *clustering* pada *cluster 3***

299. Sedangkan hasil dari perhitungan diatas pada *cluster 3* dapat disimpulkan bahwa karakteristik nasabah pada *cluster 3* didominasi oleh nasabah yang memilih macas Mitra Cerdas dengan jenis pekerjaan Wiraswasta, berpenghasilan Rp. 2.500.000 perbulan dan beralamat di Gelumbang.

300. **5. HASIL DAN PEMBAHASAN**

301. **5.1. Data Mining**

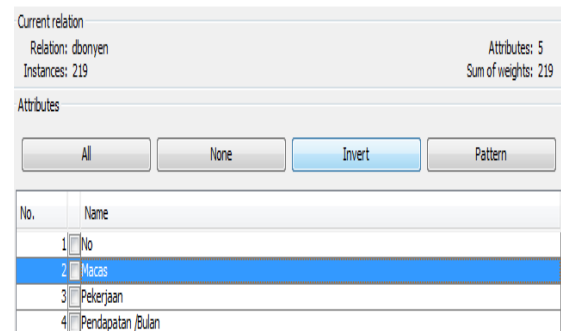
302. Setelah melakukan transformasi, tahap berikutnya adalah melakukan proses *mining data*. Proses dilakukan bertujuan untuk mencari pola dari data atau mencari nilai emas dari suatu data. Untuk menambang data, kita memerlukan teknik tertentu. Cara yang bisa kita terapkan untuk mengambil pengetahuan dari sekumpulan data yang amat besar. Teknik yang akan kita gunakan adalah *clustering* dengan menggunakan algoritma *k-means*, apabila di *Weka* dikenal sebagai algoritma *SimpleKMeans*. Pengetahuan yang akan diambil adalah berapa banyak atau kecenderungan minat nasabah terhadap jenis asuransi tersebut. Dengan mengetahui tingkatan minat mereka, kemudian kita dapat memprediksi kemungkinan mereka untuk menjadi nasabah. Tahapan ini adalah inti dari tahapan *KDD (Knowledge discovery in databases)*, melakukan evaluasi terhadap data yang sudah dimodelkan sebelumnya.

5.1.1 Penerapan Algoritma *K-Means* menggunakan aplikasi *Weka*

303. Selanjutnya yaitu memilih file yang berisi data nasabah untuk di inputkan kedalam aplikasi *Weka* agar dapat dilakukan tahap evaluasi. Setelah file berisi data nasabah dipilih maka akan

mendapatkan hasil atribut yang terdeteksi oleh *Weka* dan hasilnya dapat dilihat pada gambar 12 sebagai berikut :

304.



305. **Gambar 12. Atribut yang terdeteksi oleh *Weka* setelah *file* dipilih**

306. Ringkasan data yang berhasil dibaca adalah atribut macas, atribut pekerjaan, atribut pendapatan perbulan, atribut alamat beserta grafik ringkasan jumlah atribut, secara menyeluruh dapat dilihat pada gambar sebagai berikut :

a. Macas

307. Isi dari atribut macas seperti dijelaskan diatas dapat dilihat pada gambar 13

Selected attribute			
Name: Macas		Type: Nominal	
Missing: 0 (0%)		Unique: 1 (0%)	
		Distinct: 6	
No.	Label	Count	Weight
1	MitraMandiri	64	64.0
2	MitraBeasiswa	78	78.0
3	Mitra Guru	28	28.0
4	MitraCerdas	39	39.0
5	MitraMandiri	1	1.0
6	MitraMelati	9	9.0

312.

313. **Gambar 13. Ringkasan jumlah masing-masing jenis asuransi**

b. Pekerjaan

314. Isi dari atribut pekerjaan seperti dijelaskan diatas dapat dilihat pada gambar 14

Selected attribute

Name: Pekerjaan	Distinct: 11	Type: Nominal
Missing: 0 (0%)		Unique: 3 (1%)

No.	Label	Count	Weight
1	PNS	57	57.0
2	PETANI	77	77.0
3	HONORER	29	29.0
4	WIRASWASTA	38	38.0
5	PNS Dokter	1	1.0
6	POLISI	1	1.0
7	IRT	2	2.0
8	WISASWASTA	1	1.0
9	PERAWAT	2	2.0
10	SPG	5	5.0
11	WIRAUUSAHA	6	6.0

319.

320. **Gambar 14. Ringkasan jumlah masing-masing jenis pekerjaan nasabah**

c. Pendapatan Per Bulan

321. isi dari atribut pendapatan perbulan tersebut dapat dilihat pada gambar 15

Selected attribute

Name: Pendapatan /Bulan	Distinct: 11	Type: Numeric
Missing: 0 (0%)		Unique: 2 (1%)

Statistic	Value
Minimum	800000
Maximum	10000000
Mean	2234703.196
StdDev	1088337.373

326.

327. **Gambar 15. Penyajian statistik dari atribut pendapatan perbulan nasabah**

d. Alamat

328. Isi dari atribut alamat nasabah tersebut dapat dilihat pada gambar 16 dibawah ini

Selected attribute

Name: Alamat	Distinct: 8	Type: Nominal
Missing: 0 (0%)		Unique: 0 (0%)

No.	Label	Count	Weight
1	LUBAI ULU	30	30.0
2	LUBAI	22	22.0
3	RAMBANG LUBAI	9	9.0
4	LEBAK	12	12.0
5	LEBAK TIMUR	5	5.0
6	GELUMBANG	79	79.0
7	TANJUNG RAMAN	29	29.0
8	KELEKAR	33	33.0

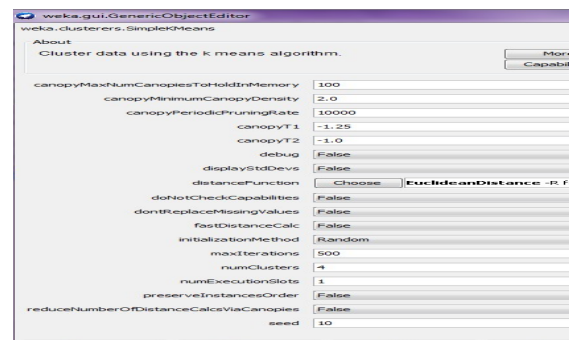
332.

333. **Gambar 16. Ringkasan Jumlah dari alamat nasabah**

334. **5.2 Evaluation (Data Mining Result)**

335. Untuk menjalankan *clustering* dengan algoritma *k-means* ada parameter yang bisa kita atur terlebih dahulu. Parameter yang paling penting adalah jumlah *cluster*, maksimum iterasi, fungsi jarak, dan metode inisial *centroid*. Tampilan dari pengaturan algoritma *k-means* pada aplikasi *Weka* dapat dilihat pada gambar 17 sebagai berikut :

336.



337. **Gambar 17. Pengaturan algoritma *k-means* pada aplikasi *Weka***

338. Dalam penelitian ini , *cluster* yang akan dibuat berjumlah 4 *cluster*, dengan fungsi jarak yang digunakan adalah fungsi jarak *euclidean*, maksimum iterasi yang ditempuh sebanyak 500 iterasi dengan pembangkitan *centroid* mula-mula secara *random*.

339. Hasil dari melakukan proses algoritma *k-means* menggunakan fungsi jarak *euclidean* dapat dilihat pada gambar 18 di bawah

```

$ java -cp weka.jar weka.gui.GenericObjectEditor -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 4
Relation:
  @dataset=weka.filters.unsupervised.attribute.Remove-21
Instances: 219
Attributes: 4
  Nominal
  Pekerjaan
  Pendapatan /Bulan
  Alamat
Test mode: evaluate on training data

=== Clustering model (full training set) ===

KMeans
=====

Number of iterations: 5
Within cluster sum of squared errors: 210.8465550037926

Initial starting points (random):
Cluster 0: MitreMandiri, PNS, 3500000, LUBAI
Cluster 1: MitreMandiri, PETANI, 1500000, LUBAI
Cluster 2: MitreMandiri, PNS, 3500000, "TANJUNG RAMAN"
Cluster 3: MitreMandiri, PNS, 3500000, GELUMBANG

```


341.

342.

343. **Gambar 18. Inisial *centroid* dari proses perhitungan *k-means***

344. Pada tahap awal, *k-means* akan membentuk *centroid* terlebih dahulu. *Centroid* merupakan nilai pusat atau nilai acuan bagi data lain agar dapat berkelompok atau membentuk *cluster*. Jumlah *centroid* akan mengikuti jumlah *cluster* dan dimensi data yang diproses berjumlah 4 yaitu atribut macas, atribut pekerjaan, atribut penghasilan perbulan, dan atribut alamat. Sehingga struktur dimensi *centroid* pun akan bernilai 4.

345. Pada gambar kita dapat lihat bahwa *cluster* 0 atau yang pertama, mengambil pola atribut macas MitraMandiri, atribut pekerjaan PNS, dengan penghasilan 3500000 dan atribut alamat di LUBAI. Begitu juga selanjutnya, yang mana ini menjelaskan bahwa aplikasi *Weka* otomatis memilih secara acak atau *random*, nilai pusat atau nilai acuan untuk mengelompokkan data nasabah ke dalam *cluster* yang telah ditentukan agar dapat dilihat pola datanya. *Centroid* akan terus diperbarui untuk dapat menyesuaikan dengan data nasabah yang ada, dan dikelompokkan berdasarkan fungsi jarak yang telah ditentukan sampai mencapai maksimum iterasi.

346. Hasil dari proses pelatihan sekaligus evaluasi yang telah dilakukan pada aplikasi *Weka* dapat dilihat pada gambar 19 sebagai berikut:

```
Final cluster centroids:
Attribute      Full Data      Cluster#
              (219.0)      0
              (53.0)      1
              (109.0)      2
              (16.0)      3
              (41.0)      4

Macas          MitraBeasiswa MitraMandiri MitraBeasiswa MitraMandiri MitraCerdas
Pekerjaan      PETANI        PNS          PETANI        PNS          WIRASWASTA
Pendapatan /Bulan 2294703.1963 3622641.5094 1391743.1193 2762500    2475609.7561
Alamat         GELUMBANG     GELUMBANG     GELUMBANG     TANJUNG RANAU GELUMBANG

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances
0      53 ( 24%)
1      109 ( 49%)
2      16 (  7%)
3      41 ( 19%)
```

350.

351.

352. **Gambar 19. Hasil *training* dan evaluasi *k-means***

353. Hasil dari *training* akan membentuk *centroid* baru yang jelas berbeda dari sebelumnya yaitu dari tahap inialisasi. Pada akhir pelatihan *Centroid* 0 , mewakili data yang mana memiliki acuan bahwa nasabah yang punya penghasilan di >1.300.000 , berprofesi sebagai Petani dan memilih produk Mitra Beasiswa berjumlah 50% dan banyak diminati di daerah Gelumbang. Namun profesi yang sama dengan penghasilan >2.000.000 memperoleh data sedikit dalam kelompoknya.

354. Setelah dilakukan *training* pembentukan *centroid* maka dapat dilihat hasil dari pengelompokkan *cluster* 0, *cluster* 1, *cluster* 2, dan *cluster* 3 seperti pada gambar 20 dibawah ini :

```
0,MitraMandiri,PNS,3500000,'LUBAI ULU',cluster0
1,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
2,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
3,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
4,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
5,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
6,MitraMandiri,PNS,3500000,'LUBAI ULU',cluster0
7,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
8,MitraMandiri,PNS,3500000,'LUBAI ULU',cluster0
9,'Mitra Guru',HONORER,1000000,'LUBAI ULU',cluster1
10,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
11,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
12,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
13,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
14,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
15,'Mitra Guru',HONORER,1000000,'LUBAI ULU',cluster1
16,MitraBeasiswa,PETANI,1500000,'LUBAI ULU',cluster1
17,MitraCerdas,WIRASWASTA,2500000,'LUBAI ULU',cluster3
```

360. **Gambar 20. Hasil pengelompokkan *cluster* 0, 1, 2, dan 3.**

361. Pola persebaran (*dispersion*) cluster dari data nasabah yang digunakan mempunyai hubungan yang erat. Persebaran pola membicarakan hal dimana terdapat pola data nasabah dan dimana tidak terdapat pola di suatu daerah pola. Dengan kata lain persebaran pola berbicara tentang lokasi pola. Persebaran *cluster*

dan jumlah data nasabah pada kelompok nya dapat disimpulkan pada gambar 21 di bawah ini :

362.

363.

364.

365.

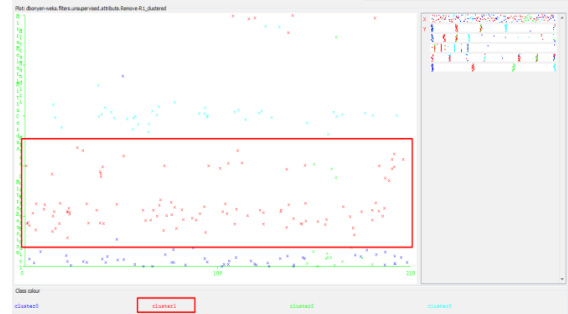
366. **Gambar 21. Persebaran Cluster dan jumlah data dalam kelompoknya**

367. Berdasarkan gambar diatas menjelaskan bahwa mayoritas nasabah berada pada *cluster* 1 atau *cluster* nomor 2 dengan jumlah warna titik merah yang paling banyak. Dilihat juga dari data gambar sebelumnya yang menyatakan bahwa jumlah anggota data yang terkelompok pada *cluster* 1 berjumlah 50% setengah dari kemungkinan yang dapat menjadikan nasabah baru. Sehingga didapatkan hasil dari proses *data mining* menggunakan teknik *clustering* dan perhitungan algoritma *k-means* adalah bahwa calon nasabah yang berprofesi sebagai Petani, dengan pendapatan rata-rata 1.500.000 , memiliki minat yang cukup tinggi untuk menjadi nasabah asuransi dengan jenis asuransi MitraBeasiswa, khususnya di wilayah Gelumbang.

6. KESIMPULAN DAN SARAN

6.2 Kesimpulan

368. Berdasarkan hasil penelitian dan pembahasan dari bab-bab sebelumnya, didapatkan beberapa kesimpulan yang dapat diambil dari penelitian mengenai penerapan *k-means clustering* untuk memprediksi minat nasabah pada PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih adalah sebagai berikut :



mengenai minat nasabah berdasarkan jenis asuransi yang dipilih, pendapatan perbulan, dan alamat nasabah.

2. Dari penelitian yang dilakukan didapatkan hasil dari proses *data mining* menggunakan teknik *clustering* dan perhitungan algoritma *k-means* adalah bahwa calon nasabah yang berprofesi sebagai Petani, dengan pendapatan rata-rata 1.500.000 , memiliki minat yang cukup tinggi untuk menjadi nasabah asuransi dengan jenis asuransi yang dipilih adalah MitraBeasiswa, khususnya di wilayah Gelumbang.
3. Dengan adanya penelitian ini dapat membantu pihak *marketing* PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih dalam pengambilan keputusan untuk melakukan promosi kepada masyarakat.

6.2 Saran

369. Setelah melakukan penelitian mengenai penerapan *k-means clustering* untuk memprediksi minat nasabah pada PT. Asuransi Jiwa Bersama Bumiputera 1912 Prabumulih penulis memiliki saran sebagai berikut :

1. Dengan adanya penelitian ini diharapkan dapat dimanfaatkan untuk pencarian nasabah baru agar proses yang dilakukan lebih efisien.

370. Disarankan agar penelitian ini dapat digunakan dan dikembangkan dalam melakukan penelitian-penelitian selanjutnya.

371. DAFTAR PUSTAKA

372. Pramudiono, 2006. *Pengertian data mining menurut para ahli* (<http://hariannetral.com/2014/09/Pengertian-data-mining-apa-itu-data-mining.html>). Diakses pada tanggal 20 November 2016 pukul 10.00 Wib.
373. Sejarah Asuransi Jiwa Bersama Bumiputera 1912 (<http://www.bumiputera.com/>). Diakses pada tanggal 20 November 2016 pukul 10.30 Wib.
374. Turban et al, 2005. *Pengertian data mining*
375. (<http://globallavebookx.blogspot.co.id/2015/04.pengertian-data-mining-menurut-ahli.html>). Diakses pada tanggal 21 November 2016 pada pukul 09.00 Wib.
376. Cahy, Suryana, 2010. *jenis-jenis data* (<https://csuryana.wordpress.com/2010/03/25/data-dan-jenis-data-penelitian/>). Diakses pada tanggal 21 November 2016 pada pukul 13.00 Wib.
377. Agusta,Y.2007. Dikutip oleh Johan Oscar Ong. *K-Means – Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol. 3 (Februari 2007) : 47-60.
378. Santosa, B. 2007. Dikutip oleh Johan Oscar Ong. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
379. Sigit, P. 2013. Tahapan KDD (*Knowledge Discovery in Database*). (<http://sigitprabowo.blogspot.co.id/2013/04/data-mining-tahap-tahapan-knowledge.html/>). Diakses pada tanggal 15 Desember 2016 pada pukul 10.00 Wib.
380. Larose, 2005. *Pengertian data mining menurut para ahli*.
381. (<http://pengertian-menurut.blogspot.co.id/2016/02/pengertian-dan-pengelompokan-data.html>). Diakses pada tanggal 16 Desember 2016 pada pukul 10.00 Wib.
382. Han dan Kamber, 2011. *Pengelompokan Clustering*.
383. (<http://library.binus.ac.id/eColls/eThesisdoc/Bab2DOC/2012-1-00011-SI%20Bab2001.doc>). Diakses pada tanggal 10 Januari 2017 pada pukul 09.00 Wib.
384. Susanto, Erdi, 2012. *Data mining menggunakan Weka*.
385. (<http://www.erdisusanto.com/2012/06/data-mining-menggunakan-weka.html>). Diakses pada tanggal 10 Januari 2017 pada pukul 09.30 Wib.
386. Purwosutjipto, 2015. *Pengertian Asuransi Jiwa Menurut Para Ahli*.
387. (<http://www.pengertianpakar.com/2015/03/pengertian-asuransi-jiwa.html>). Diakses pada tanggal 11 Januari 2017 pada pukul 10.00 Wib.