

Penerapan Data Mining Menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi

Rendy Novianto¹, Leonard Goeirmanto²

^{1,2}Universitas Mercubuana; Jln. Raya Meruya Selatan No 1 Kembangan, Jakarta Barat telp.
(021) 584 08 16

Jurusan Teknik Informatika, FASILKOM Mercubuana, Jakarta
e-mail: ¹41515110107@student.mercubuana.ac.id, ²leo@mercubuana.ac.id

Abstrak: Agent asuransi merupakan sumber bisnis yang memiliki peran yang berpengaruh terhadap produksi. Penilaian produksi dapat di lihat dari sisi harga pertanggungan, premi dan claim yang ada. Dengan demikian metode yang cocok untuk menentukan pembagian pengelompokan agent berdasarkan produksi dan nilai claim dengan menggunakan algoritma K-Means. Tujuan pengelompokan ini adalah untuk mempermudah memberikan penilaian kualitas agent yang ada pada perusahaan Asuransi. Hasil dari pengelompokan menggunakan algoritma K-Means adalah kondisi dengan data yang tidak berubah sampai dengan iterasi yang di tentukan. Proses yang dilakukan sangat cepat. Nilai yang di gunakan untuk penelitian adalah nilai claim.

Kata kunci: Data Mining, K-Means, Pengelompokan Data Asuransi, Prediksi Bisnis, Agent Asuransi.

Abstract: Insurance agent is business sources that have a role for affects production. Production assessment can be seem in terms of existing sum insured, premium and claim. This method suitable for determining the distribution of agent clustering based on production and value of claims by using the K-Means algorithm. The purpose of this grouping is to make it easier to provide an assessment quality of the agents in the insurance company. The results of clustering using the K-Means algorithm are conditions with data that does not change until the same value. The process is done very fast. The value used for research is the value of claim.

Keywords: Data Mining, K-Means, Insurance Data Clustering, Business Prediction, Insurance Agent.

1. PENDAHULUAN

Sebuah Perusahaan keuangan non Bank yang bergerak di bidang Jasa Asuransi Kerugian Nasional yang memiliki Agent cukup banyak. PT.XYZ berdasarkan data yang di terbitkan oleh AAUI (Asosiasi Asuransi Umum Indonesia) PT.XYZ memiliki anggota Agent sebanyak 42 yang terdaftar. Dari 42 Agent tersebut terdapat beberapa TOC (Type of Cover) yang mereka masukan ke Asuransi.

Dengan Sebanyak itu maka Divisi Marketing atau Divisi terkait memerlukan sistem yang memberikan informasi secara lebih cepat dan tepat. Pada saat ini reporting di lakukan dengan cara yang tidak mudah bagi seseorang yang mengerti aplikasi komputer sebagai contoh aplikasi

Microsoft Excel. Dimana proses pengolahan Data untuk mendapatkan informasi yang di inginkan membutuhkan waktu lama.

Akseptasi Policy yang membutuhkan waktu lama menjadi masalah pada PT.XYZ. Maka di perlukan metode atau tools yang dapat membantu menyelesaikan masalah tersebut agar produksi Policy tidak membutuhkan waktu lama. Tools tersebut bisa memprediksi sumber bisnis yang menguntungkan atau merugikan perusahaan.

Sehubungan dengan informasi yang di dapat bahwa Asosiasi Asuransi Umum Indonesia (AAUI) akan memperluas level sertifikasi agen Asuransi dari saat ini yang hanya satu level. Ke depan, agen asuransi umum akan menjual produk asuransi sesuai level sertifikasinya [1].

Sementara itu adanya reward yang di berikan oleh perusahaan asuransi yang di berikan kepada Agen. Terdapat 11 kategori penghargaan yang di akan di berikan AAJI. Sejumlah kategori penghargaan tersebut antara lain adalah Rookie Agent Top Premium, Rookie Top Policy, Leader Top Recruiter, dan Leader Top Premium [2].

Pada penelitian yang di lakukan sebelumnya oleh Asrul sani yaitu “Penerapan Metode K-Means Clustering pada Perusahaan”[3]. Yaitu membahas tentang seberapa hasil dari pengelompokan barang berpengaruh terhadap kebutuhan dari konsumen. Maka penulis melakukan pengujian lainnya yaitu aspek keuntungan dari suatu bisnis yang ada pada Perusahaan Asuransi dengan melakukan pengelompokan nilai Premi dan Claim dengan nilai pengelompokan yang sudah di tentukan. Maka dari hal tersebut penulis mengadopsi metode yang di gunakan dan di terapkan pada penelitian yang di lakukan penulis.

Tujuan dari pelaksanaan penelitian ini adalah untuk mempermudah pengelompokan data premi ataupun claiim yang ada sebagai landasan dasar pengambilan keputusan.

2. LANDASAN TEORI

Seperti yang di sebutkan oleh Sdri Irfiani, *Datamining* merupakan sebuah cara untuk menemukan pola yang menarik dan pengetahuan dalam data yang besar. Pada data mining menggunakan teknik matematika, statistik, artificial intelegent serta *machine learning* untuk mengidentifikasi dan mengekstrasi menjadi informasi atau pengetahuan yang bermanfaat yang di ambil dari berbagai database yang besar [4].

Dalam hal ini yang di sebut Data dan pengertian Data saya kutip dari penulis Sdri Sani. Dengan penjabaran Data adalah sesuatu yang belum mempunyai arti bagi penerimanya dan masih memerlukan adanya suatu pengolahan. Data bisa merujuk pada suatu keadaan, gambar, suara, huruf, angka, matematika, bahasa ataupun simbol-simbol lainnya yang bisa kita gunakan sebagai bahan untuk melihat lingkungan, objek, kejadian ataupun suatu konsep. Informasi merupakan hasil pengolahan dari sebuah model, formasi, organisasi, ataupun suatu perubahan berntuk dari data yang memiliki nilai tertentu dan bisa digunakan untuk menambah pengetahuan bagi yang menerimanya [3].

Cluster merupakan kumpulan objek data yang memiliki kemiripan antara satu dengan yang lain dalam kelompok lain. Clustering atau lebih di kenal dengan analisis cluster merupakan proses pengelompokan satu set benda fisik ataupun abstrak kedalam satu kelas objek yang sama [4].

Algoritma *K-Means* merupakan metode nonheirarchial yang pada awalnya mengambil sebagian dari banyak nya komponen dari populasi untuk dijadikan pusat *cluster* awal. Pada tahap ini pusat *cluster* di pilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat *cluster* yang telah di definisikan tergantung dari jaarak minimum antar komponen dengan tiap-tiap pusat cluster. Posisi pusat *cluster* akan di hitung kembali sampai

semua komponen data digolongkan ke dalam tiap-tiap pusat *cluster* dan terakhir terbentuk posisi pusat *cluster* baru. Beberapa alternatif penerapan K-Means dengan beberapa pengembangan teori-teori penghitungan terkait telah di usulkan. Hal ini termasuk pemilihan [5].

Selain itu adapun penelitian menyatakan bahwa kelebihan dengan menggunakan metode K-Means ini adalah kecepatan yang lebih tinggi dari dari pengelompokan objek. Kekurangannya adalah menentukan jumlah cluster sebelum percobaan [6].

Dengan berlandaskan teori pada penelitian sebelumnya atau bahkan sumber media lainnya. Peneliti berharap dapat memberikan hasil yang maksimal. Serta memberikan data yang validitasnya dapat di pertanggung jawabkan.

3. METODE PENELITIAN

Dalam penelitian ini menggunakan pendekatan algoritma K-Means. Algoritma K-Means merupakan salah satu algoritma dalam fungsi clustering atau pengelompokan data. Observasi atau kasus berdasarkan kemiripan objek yang diteliti. Sebuah cluster adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain. Clustering didefinisikan dengan membagi objek data dalam bentuk, entitas, contoh, ketaatan, unit ke dalam beberapa jumlah kelompok (grop, bagian atau kategori) [3].

K-Means merupakan metode penganalisaan data pada Data Mining dimana proses pemodelan tanpa supervisi dan merupakan salah satu metode yang mengelompokan data secara partisi. Pada metode K-Means data di kelompokkan menjadi beberapa kelompok dimana setiap kelompok mempunyai karakteristik yang mirip atau sama dengan lainnya namun dengan kelompok lainnya memiliki karakteristik yang berbeda. Metode ini meminimalisasi perbedaan antar data didalam satu Cluster. Serta memaksimalkan perbedaan dengan cluster yang lain [4].

Secara umum metode menggunakan algoritma K-Means sebagai berikut [4] [7]:

- a. Tentukan k sebagai jumlah cluster yang di bentuk. Penentuan banyaknya jumlah cluster k dilakukan dengan beberapa faktor seperti pertimbangan teoritis dan konseptual yang di usulkan untuk menentukan berapa banyaknya cluster.
- b. Bangkitkan k Centroid (titik pusat cluster) awal secara random. Untuk menentukan centroid awal dilakukan secara acak dari beberapa objek yang tersedia sebanyak k cluster, untuk menghitung centroid cluster ke-i berikutnya.

$$v = \sum_{k=1}^n (X_i) / N$$
$$i = 1, 2, 3 \dots n$$

Rumus 1. Menentukan Centroid Awal

Penjelasan adalah sebagai berikut:

Dimana nilai v adalah nilai centroid pada cluster,

X_i : objek ke-i, dimana objek tersebut adalah nilai yang berubah sesuai iterasi

n : banyaknya objek atau jumlah objek yang menjadi anggota cluster.

- a. Hitung jarak setiap objek ke masing- masing centroid dari masing-masing cluster. Kemudian hitung jarak antara objek dengan centroid, dalam penelitian ini menggunakan Euclidian Distance.

$$d(x, y) = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Rumus 2. Menghitung Jarak Antar Centroid

Penjelasan adalah sebagai berikut:

Dimana

D adalah nilai jarak nya,

xi: objek x ke i,

y : daya y ke-i,

n : banyaknya objek.

- b. Alokasikan masing-masing objek ke dalam centroid yang paling terdekat.
- c. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan.
- d. Ulangi langkah 3 jika posisi centroid baru tidak sama.

Selain itu dari penelitian sebelumnya K-Means dinamis yang menghitung nilai ambang sebagai centroid K-Means dan berdasarkan nilai jumlah yang terbentuk. Pada setiap literasi K-Means jika jarak euklidian antara dua titik kurang dari atau sama dengan nilai ambang. Maka dua titik data ini akan berada dalam kelompok yang sama. Jika tidak, metode yang di usulkan akan membuat cluster baru dengan titik data yang berbeda. Hasil menunjukkan bahwa metode yang di usulkan mengungguli metode K-Means asli [8].

Dalam pengembangan atau pembuatan sistem terdapat beberapa tahap yang dilakukan yaitu :

1. Tahapan pengumpulan Data

Melakukan pengumpulan data produksi dan claim sebagai landasan penelitian. Dengan cara meminta kepada pihak yang berkewenangan mempunyai data tersebut.

2. Tahap perencanaan

Memahami permasalahan dan tujuan yang di inginkan dengan data yang ada.

3. Tahap analisis

Melakukan langkah penyelesaian dan mencari solusi yang paling tepat untuk mendapatkan hasil yang maksimal dan yang ingin di hasilkan.

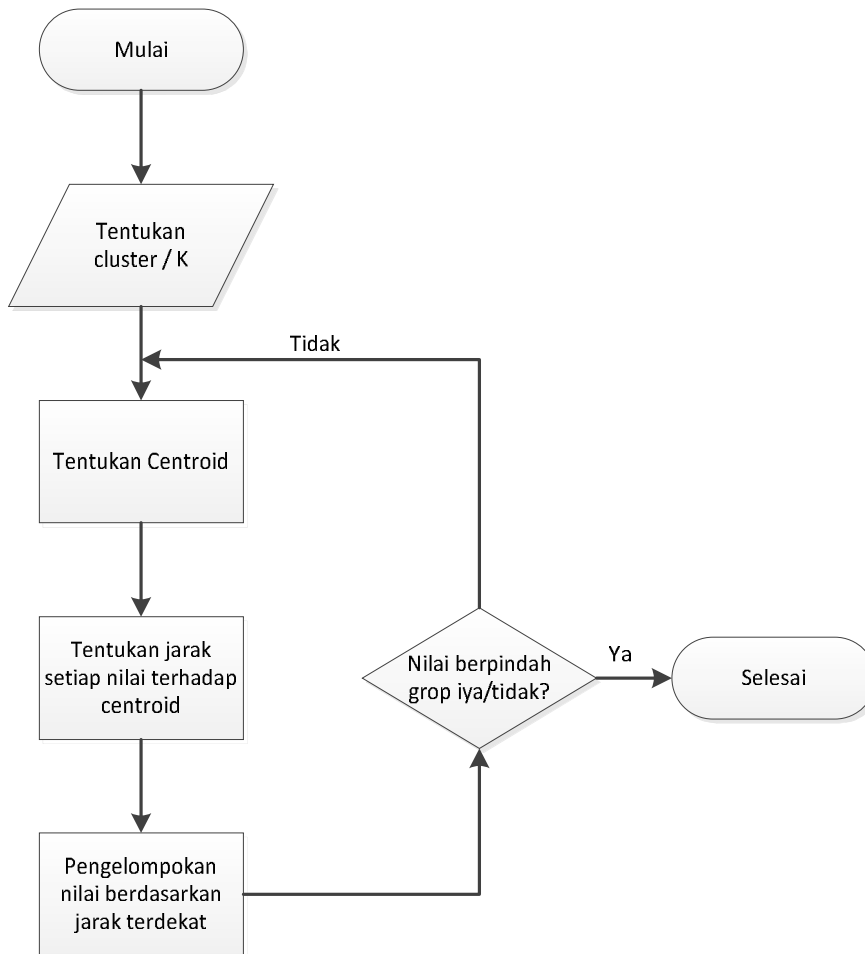
4. Tahap perancangan

Membuat *interfacedan* basis data untuk menopang kinerja sistem.

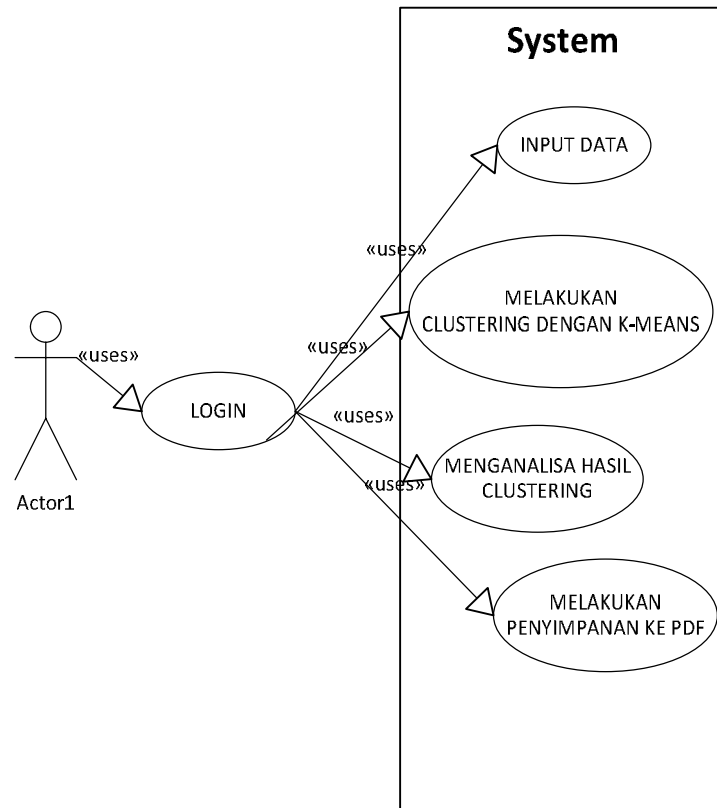
Adapun kendala dari peneliatan clustering ini yang telah di lakukan sebelumnya. *Hypercube* kendala di definisikan untuk setiap centroid dan nilai yang di peroleh untuk setiap atribut dari setiap kelas, untuk penggunaan jarak euclidean sebagai kriteria kesamaan dalam prosedur pengelompokan. Penelitian dilakukan dengan 42 dataset [9].

Metode lainnya yang di gunakan untuk pengelompokan adalah metode Fuzzy C-Means. Metode Fuzzy C-Means adalah teknik pengelompokan data dimana keberadaan tiap-tiap titik data dalam suatu cluster di tentukan oleh derajat keanggotaan. Penerapan metode tersebut dapat di terapkan dalam pengelompokan skor TOFL berdasarkan jenis soal dengan skor tertinggi dengan masukan nilai *skor Listening Comprehensif*, *skor Structure and written expression* dan *skor Reading Comprehension* [10].

Berikut adalah flowchart dan UML Rancangan aplikasi Datamining yang digunakan.



Gambar 1. Flow Aplikasi



Gambar 2. Use Case Aplikasi

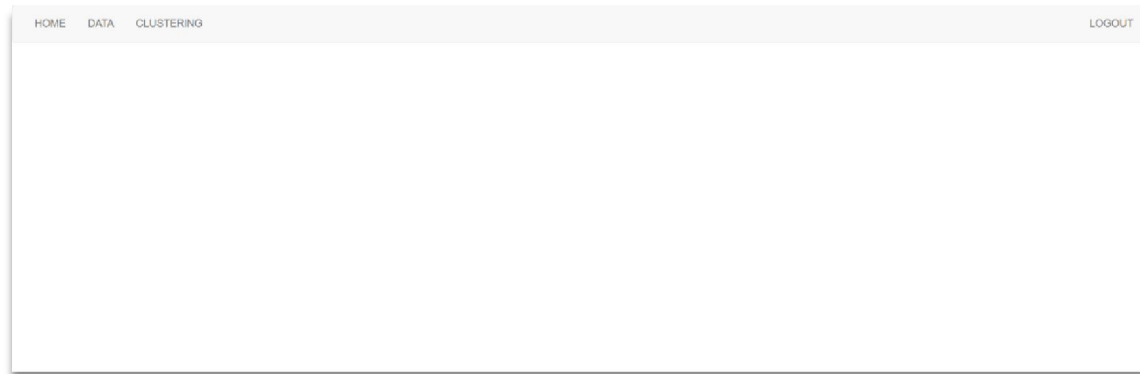
3. HASIL DAN PEMBAHASAN

Aplikasi yang di gunakan yaitu berbasis web. Pembuatan nya sendiri menggunakan bahasa program PHP dengan frame word Code Igniter di dukung database MySQL.

LOGIN

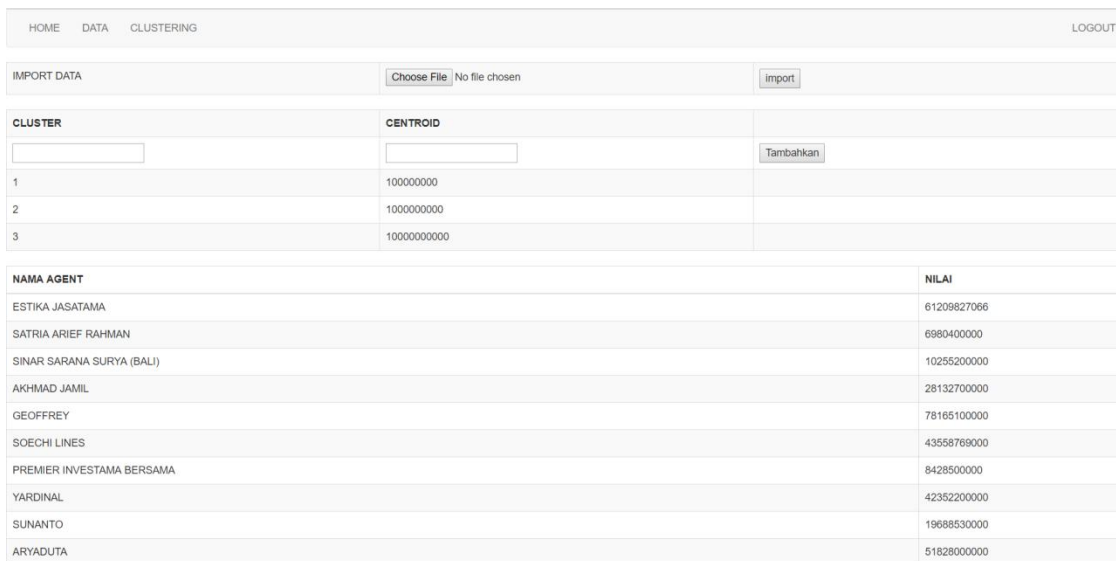
Gambar 3. Login Aplikasi

Halaman login untuk masuk ke dalam aplikasi untuk menginputkan jumlah *cluster* dan data yang akan di kelompokkan. Login dengan menggunakan *username* dan *password* yang sebelumnya sudah di daftarkan.



Gambar 4. Halaman Utama

Pada halaman utama terdapat 4 menu yaitu *Home*, *Data*, *Clustering* dan *Logout*. *Home* yaitu sebagai menu halaman utama. *Data* yaitu sebagai menu pengimputan *Clustering* dan pengimputan data. *Clustering* yaitu menu untuk memproses pengkelompokan. *Logout* yaitu menu untuk menutup aplikasi.



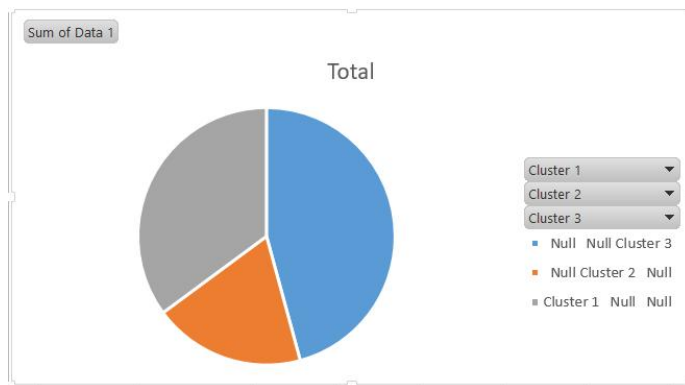
Gambar 5. Halaman Pengimputan *Cluster* dan Pengimputan Data.

Pada Gambar 5 diatas adalah halaman yang digunakan untuk menentukan nilai cluster dan jumlah cluster. Serta sebagai halaman pengimputan data dengan cara mengimpor data menggunakan format Microsoft Excel. Penentuan cluster mengambil dari target kejadian nilai yang terendah hingga di anggap memiliki nilai tertinggi sesuai dengan pengelompokannya.

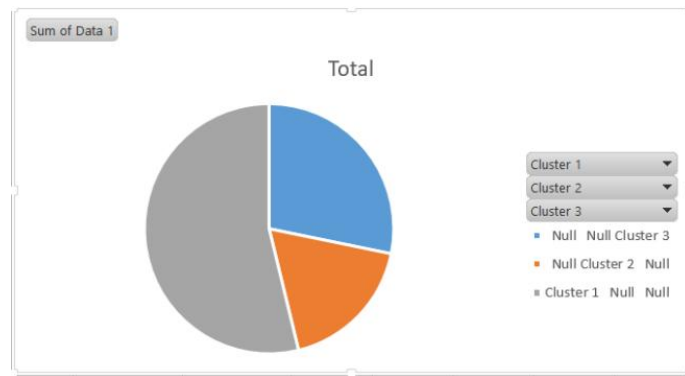
ITERASI 1					
Agent	Data 1	Cluster 1	Cluster 2	Cluster 3	
Agent 1	61209827066	Null	Null	Null	OK
Agent 2	6980400000	Null	Null	Null	OK
Agent 3	10255200000	Null	Null	Null	OK
Agent 4	28132700000	Null	Null	Null	OK
Agent 5	78165100000	Null	Null	Null	OK
Agent 6	4358769000	Null	Null	Null	OK
Agent 7	8428500000	Null	Null	Null	OK
Agent 8	4235200000	Null	Null	Null	OK
Agent 9	19688530000	Null	Null	Null	OK
Agent 10	51828000000	Null	Null	Null	OK
Agent 11	22082689282	Null	Null	Null	OK
Agent 12	112115687500	Null	Null	Null	OK
Agent 13	42921700000	Null	Null	Null	OK
Agent 14	51982030000	Null	Null	Null	OK
Agent 15	69192852201	Null	Null	Null	OK
Agent 16	24322116429	Null	Null	Null	OK
Agent 17	4364887400	Null	Null	Null	OK
Agent 18	36103515627	Null	Null	Null	OK
Agent 19	219908674732	Null	Null	Null	OK
Agent 20	46934775000	Null	Null	Null	OK
Agent 21	97628650000	Null	Null	Null	OK
Agent 22	6645500000	Null	Null	Null	OK
Agent 23	36823578350	Null	Null	Null	OK
Agent 24	78614500000	Null	Null	Null	OK
Agent 25	10391000000	Null	Null	Null	OK
Agent 26	47704000000	Null	Null	Null	OK
Agent 27	10252991000	Null	Null	Null	OK
Agent 28	88458929996	Null	Null	Null	OK
Agent 29	78713820000	Null	Null	Null	OK
Agent 30	10039400000	Null	Null	Null	OK
Agent 31	17299432228	Null	Null	Null	OK
Agent 32	74845632000	Null	Null	Null	OK
Agent 33	123335000000	Null	Null	Null	OK
Agent 34	559920692785	Null	Null	Null	OK
Agent 35	74616500000	Null	Null	Null	OK
Agent 36	1660357424161	Null	Null	Null	OK
Agent 37	5143900000000	Null	Null	Null	OK

Gambar 6. Hasil Proses Clustering

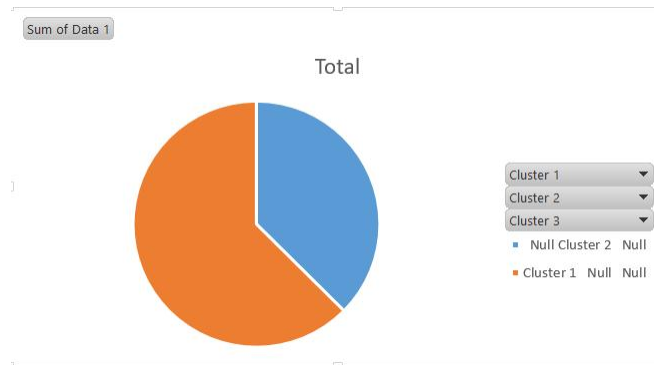
Pada Gambar 6 menunjukkan hasil proses *Clustering* data dimana pada kolom *Cluster* di tandai dengan kata OK maka data tersebut masuk pada *Clusternya*. Proses Clustering terdapat 3 Atribut yaitu Nilai Pertanggung, Nilai Premi dan Nilai Claim.



Gambar 7. Grafik Clustering Nilai Pertanggung



Gambar 8. Grafik Clustering Nilai Premi



Gambar 9. Grafik Clustering Nilai Claim

Pada Gambar 7,8,9 adalah adalah grafik hasil dari proses pengelompokan menggunakan algoritma K-Means. Terdapat 3 cluster pada proses tersebut. Pembagian cluster tersebut membagi nilai dengan perbedaan nominal yang di tentukan oleh marketing. Maka bisa di sebutkan penelitian ini berhasil. Berikut adalah jarak rata-rata nilai yang di kelompokkan

1. Nilai Pertanggungan

Cluster 1	91272615275.197
Cluster 2	5549940477779.2
Cluster 3	17725681470671

2. Nilai Premi

Cluster 1	285410948.72036
Cluster 2	8532317279
Cluster 3	33442097141

3. Nilai Claim

Cluster 1	36184286.250554
Cluster 2	3255999387.6667
Cluster 3	10000000000

Proses pada saat pengelompokan sangat cepat dan tidak membutuhkan waktu yang lama.

4. KESIMPULAN

Berdasarkan hasil pengujian dengan record 454 data yang di lakukan. Metode algoritma K-Means dapat di terapkan untuk mengelompokan data nilai pertanggungan, premi dan claim berdasarkan clustering dengan nilai terendah sedang dan tertinggi berdasarkan ketentuan perusahaan. Dengan pengelompokan sebanyak 3 cluster dimana memiliki batasan nilai yang berbeda.

5. SARAN

Berdasarkan kesimpulan penelitian berharap dapat mengembangkan kembali aplikasi tersebut dengan data yang lebih banyak. Sehingga pengolahan data dapat lebih banyak dan luas cakupannya.

UCAPAN TERIMA KASIH

Dengan penuh rasa syukur kehadirat Allah SWT dan setelah itu tanpa menghilangkan rasa hormat yang mendalam, saya selaku penyusun dan penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak-pihak yang telah membantu penulis untuk menyelesaikan penelitian ini, penulis terima kasih kepada:

1. Allah SWT, yang telah memberikan segala kemudahan, kelancaran dan kesehatan hingga sekarang.
2. Ibu Rina Rosdiana, Selaku Manager tempat penulis berkerja.
3. Ibu Desi Ramayanti S.Kom, MT, Selaku Ketua Prodi FASILKOM.
4. Bapak Dicky Firdaus, S.Kom, MM, Selaku Wakil Ketua Prodi FASILKOM.
5. Dwi Prastya Baskoro, Selaku teman seperjuangan dalam penulisan dan penelitian

DAFTAR PUSTAKA

- [1] M. Tobing 2016, *Sertifikasi Agen Dibuat Tiga Level*, *kompas.com*, [Online], Available: <https://keuangan.kontan.co.id/news/sertifikasi-agen-dibuat-tiga-level>. [Accessed: 17-May-2019].
- [2] F. Sartina Dewi. 2015, *AAJI Ganjar Agen-agen Asuransi Belasan Kategori Penghargaan*, *Bisnis.com*, [Online]. Available: <https://finansial.bisnis.com/read/20150821/215/464280/aaji-ganjar-agen-agen-asuransi-belasan-kategori-penghargaan>. [Accessed: 17-May-2019].
- [3] A. Sani, 2018, *Penerapan Metode K-Means Clustering pada Perusahaan Penerapan Metode K-Means Clustering pada Perusahaan*, Diakses dari <https://www.researchgate.net>
- [4] E. Irfiani and S. S. Rani. 2018, *Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita*, Vol. 6, No. 4, pp. 161–168, Diakses dari jurnal.untan.ac.id
- [5] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana, and F. Khusnu, *Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means, Clust. K-Means*, pp. 1–7, 2012, Diakses dari yudistira.lecture.ub.ac.id
- [6] N. Litvinenko, O. Mamyrbayev, A. Shayakhmetova, and M. Turdalyuly. 2019, Clusterization by The K-means Method when K is Unknown, *ITM Web Conf.*, Vol. 24, p. 01013, Diakses dari <https://www.mdpi.com>
- [7] M. PIETRZYKOWSKI, Mini-Model Method Based on K-means Clustering, *Przełqd Elektrotechniczny*, Vol. 1, No. 1, pp. 75–78, 2017.
- [8] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman. 2019, A Dynamic K-means Clustering for Data Mining,” *Indones. J. Electr. Eng. Comput. Sci.*, Vol. 13, No. 2, pp. 521–526, Diakses dari <https://pdfs.semanticscholar.org>

- [9] P. N. Smyrlis. 2018, *Constrained K-Means Classification*, Diakses dari <https://www.researchgate.net>
- [10] R. F. Kemala, I. F. Astuti, and S. Maharani, 2019, *Penerapan Metode Fuzzy C-Means pada Aplikasi Simulasi TOEFL (Test Of English As A Foreign Language) Berbasis Web (Studi Kasus: Fakultas MIPA Universitas Mulawarman)*, *Inform. Mulawarman J. Ilm. Ilmu Komput*, Vol. 14, No. 1, pp. 17.