

**ANALISIS *K-MODES CLUSTERING* DENGAN
METODE JARAK YUAN**

(Studi Kasus : Asuransi Perjalanan)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:

Sheilta Alphenia

18611109

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA**

2022

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**


Judul : Analisis K-Modes Clustering Dengan Metode
Jarak Yuan (Studi Kasus : Asuransi Perjalanan)
Nama Mahasiswa : Sheilta Alphenia
NIM : 18611109

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 25 Maret 2022

Pembimbing

Mengetahui,
Ketua Program Studi


(Kariyam, S.Si., M.Si.)


(Dr. Edy Widodo, S.Si., M.Si.)

HALAMAN PENGESAHAN
TUGAS AKHIR

ANALISIS *K-MODES CLUSTERING* DENGAN METODE JARAK YUAN
(Studi Kasus : Asuransi Perjalanan)

Nama Mahasiswa : Sheilta Alphenia

NIM : 18611109

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 29 Maret 2022

Nama Penguji

1. Muhammad Hasan Sidiq K, S.Si., M.Sc.
2. Achmad Fauzan, S.Pd., M.Si
3. Kariyam, S.Si., M.Si.

Tanda Tangan



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)

KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Puji syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat, anugerah serta hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir ini sebagaimana mestinya. Shalawat dan selamat tidak lupa tercurah kepada Nabi Muhammad SAW beserta keluarga dan para sahabat serta pengikutnya.

Selama penyusunan Tugas Akhir ini, penulis menyadari bahwa penulis mendapat banyak bimbingan dan bantuan dari berbagai pihak. Maka dari itu, pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Keluarga yang selalu memberi semangat dan doa.
2. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta.
3. Bapak Dr. Edy Widodo, S.Si., M.Si. selaku Ketua Jurusan Statistika beserta seluruh jajarannya.
4. Ibu Kariyam, S.Si., M.Si. yang telah memberi bimbingan selama penulisan Tugas Akhir ini.
5. Mas Achmad Kurniansyah telah meluangkan waktu untuk membantu saya dalam menyusun sintaks pada penelitian ini.
6. Nanda Lailatul Humairoh, Rahmi Novika Harahap, Afdelia Novianti yang selalu memberikan tempat untuk menuangkan segala emosi dan menjadi teman berproses selama empat tahun ini.
7. Aisyah Putri Zanuarizqi dan Putri Ermayani sebagai teman seperjuangan dalam mencoba tantangan baru seperti lomba dan Kampus Merdeka. Terimakasih telah memberikan pengalaman baru yang memang patut dicoba.
8. Teman seper bimbingan tugas akhir, Syafira Irsalina yang telah menemani dan bertukar pikiran selama proses bimbingan.
9. Erlambang sebagai pihak yang selalu memberikan arahan, bantuan berupa moril dan teman bertukar pikiran.

10. Terimakasih kepada diri sendiri karena sudah mampu bertahan berjuang melewati proses ini.

11. Serta semua pihak yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa laporan Tugas Akhir ini masih terdapat banyak kekurangan. Oleh karena itu segala kritik dan saran yang sifatnya membangun terbuka untuk penulis. Semoga laporan ini bermanfaat bagi penulis dan semua pihak yang membutuhkan. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Aamiin aamiin ya Robbal 'alamiin

Wassalamualaikum Wr.Wb

Yogyakarta, 5 Maret 2022



Sheilta Alphenia

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN TUGAS AKHIR	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xi
PERNYATAAN	xii
INTISARI	xiii
ABSTRACT	xiv
BAB 1 PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Jenis Penelitian dan Metode Analisis.....	4
1.5. Tujuan Penelitian	5
1.6. Manfaat Penelitian	5
BAB 2 TINJAUAN PUSTAKA.....	6
2.1. Penelitian Sebelumnya	6
BAB 3 LANDASAN TEORI	12
3.1. Data	12
3.1.1 Data Kategorik.....	12
3.1.2 <i>Data Mining</i>	13
3.2. Statistika Deskriptif.....	13
3.3. <i>Python</i>	14
3.4. Analisis <i>K-Modes Clustering</i>	14
3.4.1 Perhitungan Kedekatan Jarak Data Kategorik.....	14
3.4.2 Perhitungan Jarak Data Nominal.....	15
3.4.3 Perhitungan Jarak Data Ordinal.....	17
3.4.4 Pengukuran Ketidaksamaan Antara Dua Objek.....	19
3.4.5 Algoritma <i>Cluster K-Modes</i>	21
3.5. <i>Silhouette Coefficient</i>	23
3.6. Asuransi.....	25
3.7. Pendapatan Perseorangan.....	26
BAB 4 METODOLOGI PENELITIAN	27
4.1. Populasi Penelitian	27
4.2. Sumber Data.....	27
4.3. Variabel Penelitian	27
4.4. Alat dan Cara Organisir Data	30
4.5. Tahapan Penelitian	30
4.5.1 Tahapan <i>K-Modes Clustering</i>	31
4.5.2 Tahapan Perhitungan <i>Silhouette Coefficient</i>	33
BAB 5 HASIL DAN PEMBAHASAN	35
5.1. Analisis Deskriptif	35

5.2.	Sintaks Perhitungan Jarak dengan <i>Python</i>	39
5.2.1	Jarak Nominal	39
5.2.2	Jarak Ordinal	42
5.3.	<i>K-Modes Clustering</i>	43
5.3.1	<i>Initial Cluster</i>	44
5.3.2	Jarak Nominal	44
5.3.3	Jarak Ordinal	47
5.3.4	Jarak Gabungan	54
5.3.5	Penempatan Data Pada Setiap <i>Cluster</i>	56
5.3.6	Pemilihan <i>Initial Cluster</i> Baru	59
5.3.7	Iterasi	61
5.4.	<i>Silhouette Coefficient</i>	70
5.4.1	Jarak Nominal	70
5.4.2	Jarak Ordinal	71
5.4.3	Jarak Gabungan	72
5.4.4	Evaluasi <i>Cluster</i>	73
5.5.	Profilisasi	74
BAB 6 PENUTUP		78
6.1.	Kesimpulan	78
6.2.	Saran	79
DAFTAR PUSTAKA		80
LAMPIRAN		84

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Sebelumnya	7
Tabel 3.1 Data Percontohan	19
Tabel 3.2 Kriteria Nilai <i>Silhouette Coefficient</i>	24
Tabel 4.1 Deskripsi Variabel Penelitian.....	27
Tabel 5.1 Penjelasan Sintaks Perhitungan Jarak Nominal	39
Tabel 5.2 Penjelasan Sintaks Pencarian MSOA.....	42
Tabel 5.3 Penjelasan Sintaks Perhitungan <i>Dod₁</i>	43
Tabel 5.4 <i>Initial Cluster</i>	44
Tabel 5.5 Ukuran Domain Tiap Variabel.....	47
Tabel 5.6 Ukuran Domain Tiap Variabel.....	51
Tabel 5.7 Modus <i>Cluster</i> Berjumlah Dua Pertama	59
Tabel 5.8 Modus <i>Cluster</i> Berjumlah Tiga Pertama.....	59
Tabel 5.9 Modus <i>Cluster</i> Berjumlah Empat Pertama	60
Tabel 5.10 Modus <i>Cluster</i> Berjumlah Lima Pertama.....	60
Tabel 5.11 Modus <i>Cluster</i> Berjumlah Enam Pertama.....	60
Tabel 5.12 Modus <i>Cluster</i> Berjumlah Tujuh Pertama	61
Tabel 5.13 Modus <i>Cluster</i> Berjumlah Dua Kedua Tidak Asuransi	62
Tabel 5.14 Modus <i>Cluster</i> Berjumlah Dua Kedua Punya Asuransi.....	62
Tabel 5.15 Modus <i>Cluster</i> Berjumlah Dua Ketiga Punya Asuransi.....	62
Tabel 5.16 Modus <i>Cluster</i> Berjumlah Tiga Kedua Tidak Asuransi.....	63
Tabel 5.17 Modus <i>Cluster</i> Berjumlah Tiga Kedua Punya Asuransi	63
Tabel 5.18 Modus <i>Cluster</i> Berjumlah Empat Kedua Tidak Asuransi.....	64
Tabel 5.19 Modus <i>Cluster</i> Berjumlah Empat Ketiga Tidak Asuransi	64
Tabel 5.20 Modus <i>Cluster</i> Berjumlah Empat Kedua Punya Asuransi.....	64
Tabel 5.21 Modus <i>Cluster</i> Berjumlah Lima Kedua Tidak Asuransi.....	65
Tabel 5.22 Modus <i>Cluster</i> Berjumlah Lima Kedua Punya Asuransi	65
Tabel 5.23 Modus <i>Cluster</i> Berjumlah Enam Kedua Tidak Asuransi.....	66
Tabel 5.24 Modus <i>Cluster</i> Berjumlah Enam Ketiga Tidak Asuransi	66
Tabel 5.25 Modus <i>Cluster</i> Berjumlah Enam Keempat Tidak Asuransi.....	67
Tabel 5.26 Modus <i>Cluster</i> Berjumlah Enam Kedua Punya Asuransi	67

Tabel 5.27 Modus <i>Cluster</i> Berjumlah Enam Ketiga Punya Asuransi.....	68
Tabel 5.28 Modus <i>Cluster</i> Berjumlah Tujuh Kedua Tidak Asuransi	68
Tabel 5.29 Modus <i>Cluster</i> Berjumlah Tujuh Ketiga Tidak Asuransi	69
Tabel 5.30 Modus <i>Cluster</i> Berjumlah Tujuh Kedua Punya Asuransi.....	69
Tabel 5.31 Modus <i>Cluster</i> Berjumlah Tujuh Ketiga Punya Asuransi.....	69
Tabel 5.32 Modus <i>Cluster</i> Berjumlah Tujuh Keempat Punya Asuransi.....	70
Tabel 5.33 Modus <i>Cluster</i> Pertama.....	74
Tabel 5.34 Modus <i>Cluster</i> Kedua	74
Tabel 5.35 Modus <i>Cluster</i> Ketiga	75
Tabel 5.36 Modus <i>Cluster</i> Keempat	75
Tabel 5.37 Modus <i>Cluster</i> Pertama Punya Asuransi.....	76
Tabel 5.38 Modus <i>Cluster</i> Kedua Punya Asuransi	76

DAFTAR GAMBAR

Gambar 4.1 Diagram Alir Tahapan Penelitian	30
Gambar 4.2 Diagram Alir <i>K-Modes Clustering</i>	32
Gambar 4.3 Diagram Alir <i>Silhouette Coefficient</i>	33
Gambar 5.1 Umur Pengguna Jasa	35
Gambar 5.2 Tipe Pendapatan Terhadap Tipe Pekerjaan	36
Gambar 5.3 Frekuensi Penerbangan Pengguna Jasa	37
Gambar 5.4 Jumlah Keluarga dari Pengguna Jasa	37
Gambar 5.5 Pengguna Jasa Lulusan Perguruan Tinggi.....	38
Gambar 5.6 Penderita Penyakit Kronis Terhadap Kepemilikan Asuransi	38
Gambar 5.7 Nilai <i>Silhouette Score</i> Tiap <i>Cluster</i> Tidak Asuransi	73
Gambar 5.8 Nilai <i>Silhouette Score</i> Tiap <i>Cluster</i> Punya Asuransi.....	73

DAFTAR LAMPIRAN

Lampiran 1	84
Lampiran 2	85
Lampiran 3	95
Lampiran 4	101
Lampiran 5	102
Lampiran 6	103
Lampiran 7	104
Lampiran 8	105
Lampiran 9	106
Lampiran 10	106
Lampiran 11	108
Lampiran 12	109
Lampiran 13	110
Lampiran 14	111
Lampiran 15	112
Lampiran 16	113

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 5 Maret 2022



Sheilta Alphenia

INTISARI

ANALISIS *K-MODES CLUSTERING* DENGAN METODE JARAK YUAN)

(Studi Kasus : Asuransi Perjalanan)

Sheilta Alphenia

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Metode yang sering digunakan dalam pengumpulan informasi adalah analisis *cluster*. Namun analisis *cluster* yang banyak dijumpai hanya berfokus pada data numerik saja, padahal selain data numerik sering pula dijumpai data dalam bentuk kategorik. Analisis *cluster* dengan data kategorik sering disebut *K-Modes Clustering*, namun metode perhitungan jarak pada jenis analisis *cluster* ini masih banyak dijumpai tidak membedakan variabel nominal dan variabel ordinal. Pada penelitian ini dilakukan implementasi perhitungan jarak Yuan pada *K-Modes Clustering* agar informasi ordinal tersebut tidak dihilangkan. Implementasi dari metode yang diusulkan Yuan masih jarang ditemukan salah satunya bidang asuransi. Selaras dengan menghadapi *New Normal* maka akan mulai banyak destinasi wisata dan aktivitas perjalanan yang mulai diizinkan. Kebutuhan akan proteksi kesehatan selama melakukan aktivitas dapat diwujudkan dengan asuransi perjalanan. Perusahaan asuransi perlu memperhatikan kebutuhan pelancong seperti memberikan produk-produk asuransi menarik sesuai dengan kebutuhan para pelancong. Berdasarkan kebutuhan tersebut agar didapatkan produk yang cocok untuk pelancong, maka perlu diteliti lebih lanjut karakteristik pelancong dengan menggunakan metode analisis *cluster*. Karakteristik dari pelancong dapat dijadikan acuan untuk membuat produk-produk asuransi menarik sesuai dengan kebutuhan pelancong. Pada penelitian ini didapatkan 4 kelompok yang terbentuk dari data pengguna jasa yang tidak memiliki asuransi dari jasa asuransi perjalanan yang digunakan dengan didapatkan besaran *silhouette score* sebesar 0,00029477. Sementara pada data pengguna jasa yang memiliki asuransi didapatkan 2 kelompok dengan nilai *silhouette score* 0,0004494.

Kata Kunci : *K-Modes Clustering, Silhouette Score, Asuransi Perjalanan.*

ABSTRACT

K-MODES CLUSTERING ANALYSIS WITH YUAN'S DISTANCE METHOD

(Case Study : Travel Insurance)

Sheilta Alphenia

Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

The mainstream method to collect information from the data is clustering analysis. However, the majority clustering method only focused on numerical data even so there is more categorical data in the real case. Clustering analysis using categorical data which is called K-Modes Clustering, still used dissimilarity measurement that does not distinguish the differences between nominal and ordinal attributes. In this research uses dissimilarity measurement which has been proposed by Yuan so the information in ordinal attribute not be missed. The implementation of this dissimilarity measurement is seldom to be found especially in the insurance field. The need for health protection during activities can be realized with travel insurance. Insurance companies should pay attention to the needs of travelers such as providing attractive insurance products according to the needs of travelers. Based on these needs, in order to obtain products that are suitable for travelers, it is necessary to further investigate the characteristics of travelers. The characteristics of travelers can be used as a reference to make attractive insurance products according to the needs of travelers. In this study, using K-Modes Clustering Yuan's Dissimilarity there were 4 groups formed from the data of users of travel insurance services used who do not have insurance with the silhouette score of 0.00029477. Meanwhile, for service users who have insurance, there are 2 groups with a silhouette score of 0.0004494.

Keywords: *K-Modes Clustering, Silhouette Score, Travel Insurance.*

BAB 1

PENDAHULUAN

1.1. Latar Belakang Masalah

Pada era digitalisasi sekarang ini, data dapat ditemukan dimana saja. Banyaknya data tersebut seiring dengan kebutuhan data itu sendiri yang sekarang juga banyak digunakan. Data yang tersedia tidak hanya data berupa numerik namun juga kategorik. Data numerik sering dijumpai dalam bentuk angka dan data kategorik sering dijumpai dalam bentuk kata-kata. Dari data yang didapatkan digunakan untuk menghasilkan suatu informasi baru hingga membuat sebuah keputusan yang tepat berdasarkan kenyataan dengan data tersebut. Salah satu metode dalam pengumpulan informasi adalah analisis *cluster*. Analisis *cluster* merupakan suatu metode untuk mempartisi data menjadi beberapa kelompok dengan sifat homogen di dalam kelompok tersebut dan heterogen untuk antar kelompok. Pengelompokan data didasarkan dengan perhitungan jarak minimum data terhadap pusat *cluster*. Setiap kelompok akan memiliki karakteristik sendiri yang akan memberikan informasi baru. Analisis *cluster* banyak diterapkan di berbagai sektor seperti pemasaran, ekonomi, kesehatan, pemerintah, sosial dan budaya. Salah satu algoritma dari analisis *cluster* yang sering digunakan adalah algoritma untuk data numerik seperti *K-means clustering*.

Analisis *cluster* dengan data kategorik diperkenalkan oleh Huang (1998) yaitu *K-Modes Clustering*. Perlakuan analisis *cluster* data kategorik berbeda dengan perlakuan untuk data numerik. Pada *K-Means Clustering* penentuan pusat data didasari oleh pengukuran rata-rata sedangkan penentuan pusat *cluster* yang digagas oleh Huang untuk data kategorik berdasarkan dengan pengukuran modus. Namun perlakuan untuk perhitungan jarak data terhadap pusat *cluster* pada data kategorik berupa nominal dan ordinal dari algoritma yang digagas Huang masih sama. Data nominal berbeda dengan data ordinal karena data nominal hanya menghasilkan informasi berupa label saja sementara data ordinal menghasilkan informasi berupa label dan juga memiliki makna urutan sehingga perlakuan antara dua tipe skala data ini perlu dibedakan agar tidak mengurangi informasi yang terkandung dalam data. Yuan dkk. (2020) mengusulkan metode pengukuran jarak yang berbeda untuk data

ordinal sehingga informasi label urutan dari data ordinal tidak dihilangkan. Implementasi dari metode yang diusulkan Yuan masih jarang ditemukan dalam artian belum banyak penelitian analisis *K-Modes Clustering* dengan menggunakan metode jarak Yuan. Sedangkan data kategorik berupa nominal dan ordinal sering dijumpai dalam *dataset* berbagai bidang, salah satunya bidang asuransi.

Menurut berita dari (Kontan, 2021) selama pandemi, kebutuhan akan asuransi di bidang kesehatan terus tumbuh sejalan dengan kebutuhan proteksi selama pandemi. Masyarakat berupaya memitigasi risiko dengan membeli produk asuransi kesehatan. Kegiatan membeli produk asuransi kesehatan ini juga terjadi di negara India. Pandemi COVID-19 di India berhasil mengubah perilaku masyarakat terkait kesehatan. Masyarakat terutama generasi millennial di India gencar membeli asuransi jiwa. (Rahman, 2021) melalui media IDN Times mengabarkan bahwa perilaku masyarakat untuk membeli asuransi jiwa didasarkan tingginya tingkat kematian warga akibat COVID-19. Ironisnya pada tahun yang sama, di saat lebih dari 60 ribu kasus COVID-19 terjadi di India, pemerintah India kembali membuka destinasi wisata seperti Taj Mahal (Rabbi, 2021). Pembukaan destinasi wisata di tengah pandemi yang belum reda dapat dilakukan apabila didukung dengan kesadaran masyarakat dalam memproteksi diri pada bidang kesehatan. Selain dengan protokol kesehatan seperti menjaga jarak, memakai masker dan mencuci tangan, protokol kesehatan di kala berwisata dapat diupayakan dengan menggunakan asuransi jiwa perjalanan atau biasa disebut asuransi perjalanan.

Keadaan pembukaan destinasi wisata di tengah pandemi COVID-19 juga terjadi di Indonesia. Menurut Direktur Eksekutif Asosiasi Asuransi Umum Indonesia (AAUI) Dody Dalimunthe kepada (Kompas, 2021) bahwa prediksi ini didasarkan pada mulai banyak destinasi wisata dan aktivitas *travelling* yang diizinkan. Dody juga menambahkan agar perusahaan asuransi perlu memperhatikan kebutuhan pelancong seperti memberikan produk-produk asuransi menarik sesuai dengan kebutuhan para pelancong.

Berdasarkan kebutuhan tersebut agar didapatkan produk atau paket asuransi yang cocok untuk pelancong, maka perlu diteliti lebih lanjut karakteristik pelancong yang mendaftar (pelanggan) pada perusahaan asuransi tersebut. Maka dari itu perlu dilakukan studi terhadap data asuransi perjalanan dengan

menggunakan metode analisis *cluster* yang dapat dijadikan suatu metode untuk mengetahui karakteristik pelanggan tersebut. Karakteristik dari dapat dijadikan acuan untuk membuat produk-produk asuransi menarik sesuai dengan kebutuhan pelanggan. Dalam penelitian ini akan digunakan metode analisis *K-Modes Clustering* dengan menggunakan metode jarak Yuan dalam menentukan karakteristik pelanggan yang didapatkan berdasarkan profil dari masing-masing kelompok yang terbentuk.

Karakteristik pelanggan yang membeli asuransi telah diteliti oleh (Arizal & Agus, 2015) dan didapatkan perilaku dari konsumen di antaranya adalah faktor sosial yaitu keluarga; faktor pribadi seperti umur, pekerjaan, gaya hidup, dan keadaan ekonomi; faktor psikologis seperti motivasi, belajar, kepercayaan dan sikap. Gaya hidup dalam hal berwisata dapat dikaitkan dengan seberapa sering pelanggan tersebut berwisata. Motivasi pelanggan dapat dilihat dari faktor kesehatan seperti riwayat penyakit dari pelanggan tersebut.

Data terkait pengguna asuransi perjalanan di Indonesia masih minim ditemukan secara publik. Sehingga pada penelitian ini digunakan data pelanggan agensi perjalanan dari India yang terdiri dari pelanggan yang menggunakan asuransi dan tidak. Data tersebut diolah baik yang menggunakan asuransi atau tidak. Dari yang menggunakan asuransi perjalanan dapat dilihat karakteristik pelanggan yang mendaftar asuransi perjalanan dan dari data pelanggan yang tidak mendaftar asuransi perjalanan dapat dilihat kelompok pelanggan yang berpotensi untuk dipromosikan asuransi perjalanan. Data India tersebut dapat digunakan sebagai percontohan untuk data pengguna asuransi perjalanan di Indonesia karena terdapat kemiripan kejadian antara di Indonesia dan India. Kemiripan tersebut didasarkan pada kejadian pembukaan wisata di tengah pandemi COVID-19.

Alat dalam bentuk perangkat lunak dalam mengolah data bermacam-macam, salah satunya adalah *Python*. *Python* merupakan bahasa pemrograman yang bersifat *open source* atau dapat diakses secara gratis serta memiliki bahasa pemrograman yang cukup sederhana dan mudah dimengerti. Analisis data tersebut biasanya dihimpun dalam suatu *package*. *Package* merupakan sekumpulan modul dalam *Python*. *Package* tersebut merupakan sebuah cara untuk mengelola dan mengorganisir modul-modul *Python* dalam bentuk direktori.

Analisis *cluster* dengan metode *K-Modes Clustering* telah disediakan di dalam sebuah *package* dalam bahasa *Python*, namun pada *package* yang tersedia belum terdapat perhitungan jarak yang digagas oleh Yuan. Pada penelitian ini akan dibentuk sebuah pemrograman sederhana dalam bahasa pemrograman *Python* untuk menghitung jarak Yuan tersebut.

1.2. Rumusan Masalah

Dari latar belakang permasalahan tersebut dapat disusun beberapa rumusan masalah sebagai berikut :

1. Bagaimana implementasi metode analisis *K-Modes Clustering* dengan perhitungan jarak Yuan untuk data asuransi perjalanan?
2. Bagaimana profil *user* berdasarkan hasil metode analisis *K-Modes Clustering* perhitungan jarak Yuan untuk data asuransi perjalanan?
3. Bagaimana pengaplikasian algoritma *K-Modes Clustering* metode jarak Yuan dengan perangkat lunak *Python*?

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah :

1. Penelitian ini berfokus pada implementasi dari analisis *K-Modes Clustering*.
2. Metode pengukuran jarak yang digunakan adalah metode pengukuran jarak yang digagas oleh Yuan.
3. Data yang digunakan merupakan data sekunder yang diambil dari *website kaggle.com*.
4. Data yang digunakan merupakan data dari perusahaan agen perjalanan di India sebagai percontohan analisis *K-Modes Clustering* pada data asuransi.
5. Data diolah dengan menggunakan perangkat lunak *Python 3* dan *Microsoft Excel* untuk analisis *K-Modes Clustering*, serta *Tableau* untuk melakukan analisis statistika deskriptif.

1.4. Jenis Penelitian dan Metode Analisis

Jenis penelitian untuk tugas akhir ini adalah penelitian aplikatif. Metode analisis yang digunakan dalam tugas akhir ini yaitu statistika deskriptif dan analisis *cluster*.

1.5. Tujuan Penelitian

Tujuan dari dilakukannya penelitian ini adalah :

1. Mengetahui implementasi dari analisis *K-Modes Clustering* dengan metode pengukuran jarak yang digagas Yuan untuk data asuransi perjalanan.
2. Mengetahui profil *user* berdasarkan *cluster* yang terbentuk dari metode analisis *K-Modes Clustering* dengan metode pengukuran jarak yang digagas Yuan untuk data asuransi perjalanan.
3. Mengetahui pengaplikasian algoritma *K-Modes Clustering* metode jarak Yuan dengan perangkat lunak *Python*.

1.6. Manfaat Penelitian

Manfaat dari dilakukannya penelitian ini adalah :

1. Sebagai bahan informasi bagi peneliti maupun pembaca bahwa terdapat metode analisis *cluster* untuk data kategorik campuran, nominal dan ordinal, tanpa mengurangi informasi dari data ordinal.
2. Bahan referensi bagi peneliti lain untuk pengimplementasian analisis *cluster* pada bidang asuransi sehingga dapat diimplementasikan untuk studi kasus data asuransi riil dengan tipe data kategorik.
3. Sebagai bahan pembuatan gagasan paket asuransi untuk perusahaan bidang asuransi apabila diterapkan pada data asuransi riil.
4. Bagi agensi perjalanan di India dapat dilihat profil pengguna yang belum mendaftar asuransi perjalanan namun berpotensi mendaftar asuransi perjalanan berdasarkan karakteristik pengguna yang sudah mendaftarkan asuransi perjalanan.

BAB 2

TINJAUAN PUSTAKA

2.1. Penelitian Sebelumnya

Penelitian terkait *clustering* untuk data kategorik sudah banyak dilakukan. Huang (1998) memperkenalkan algoritma *K-Modes* dengan metode pencocokan ketidaksamaan sederhana untuk menangani pengelompokan khusus data yang bersifat kategorik. Algoritma ini menggantikan *means* pada algoritma *k-means* dengan modus dan menggunakan metode berbasis frekuensi untuk membantu memperbarui modus dalam proses *clustering* dalam proses meminimalkan fungsi biaya pada *clustering*. Pengukuran ketidaksamaan yang diperkenalkan oleh Huang untuk data kategorik adalah dengan mendefinisikannya sebagai total ketidakcocokan antara dua objek yang berhubungan. Semakin kecil jumlah ketidakcocokan maka semakin mirip dua objek tersebut. Huang mengaplikasikan *K-Modes algorithm* dengan pengukuran ketidaksamaan sederhana untuk data *soybean* dan dihasilkan kesimpulan bawah algoritma tersebut cocok digunakan untuk *dataset* yang besar. Selain itu *K-Modes algorithm* lebih cepat dalam peng-cluster-isasi data dibandingkan *k-means* dikarenakan hanya dibutuhkan lebih sedikit iterasi untuk pemusatan.

Cao dkk.(2012) menemukan keterbatasan dalam penggunaan pencocokan ketidaksamaan sederhana untuk *K-Modes Clustering* yang digunakan oleh Huang. Metode pencocokan ketidaksamaan sederhana seringkali menghasilkan kelompok dengan kesamaan intra lemah dan mengabaikan kesamaan tersembunyi antara nilai kategorik. Cao mengagas metode pengukuran jarak berdasarkan *rough set theory*. Karakteristik yang berbeda dari pengukuran jarak baru memperhitungkan distribusi dari nilai atribut pada seluruh semesta. Hasil dari pembelajaran konvergen dan kompleksitas waktu dari algoritma *K-Modes* berdasarkan pengukuran jarak Cao terindikasi efektif untuk kumpulan data besar.

Yuan dkk., (2020) mengagas perhitungan jarak untuk data kategorik campuran yaitu data nominal dan ordinal. Yuan melakukan gagasan ini dikarenakan beliau menemukan keterbatasan pada pengukuran jarak dari algoritma *K-Modes* yang sudah ada. Keterbatasan yang dimaksud adalah kegagalan penelitian terdahulu

dalam memperhatikan variabel ordinal dan membuat metrik persamaan eksplisit untuk variabel ordinal. Pada penelitiannya tersebut, Yuan mengatasi keterbatasan tersebut dengan membuat metrik pengukuran kesamaan variabel ordinal.

Tabel 2.1 merupakan penelitian terdahulu yang menyangkut *K-Modes Clustering* dan penerapan *cluster* dalam bidang asuransi.

Tabel 2.1 Tabel Penelitian Sebelumnya

Tahun	Nama	Judul	Hasil Penelitian
2017	Tanti Yulianita dan Deden Istiawan	Implementasi Algoritma <i>K-Modes</i> untuk Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis	Peneliti menggunakan <i>K-Modes Clustering</i> untuk menentukan prioritas rehabilitasi daerah aliran sungai berdasarkan parameter lahan kritis. Hasil dari penelitian ini didapatkan tiga <i>cluster</i> dengan <i>cluster</i> pertama merupakan parameter lahan kritis berupa penutupan lahan sangat baik, kemiringan lereng yang datar dengan tingkat bahaya erosi yang berat serta manajemen yang buruk. <i>Cluster 2</i> memiliki kriteria lahan kritis berupa penutupan lahan sedang, kemiringan lereng landai, tingkat bahaya erosi sangat berat serta manajemen yang buruk. <i>Cluster 3</i> dengan kriteria agak kritis dengan penutupan lahan baik, kemiringan lereng landai, tingkat bahaya erosi yang sangat berat serta manajemen yang buruk.
2017	Nur Amah, Sri Wahyuningsih dan Fidia Deny Tisna Amijaya	Analisis <i>Cluster</i> Non-Hirarki Dengan Menggunakan Metode <i>K-Modes</i> pada Mahasiswa Program Studi Statistika Angkatan 2015 FMIPA Universitas Mulawarman	Analisis <i>cluster</i> pada penelitian ini digunakan untuk mengkaji pertimbangan pemilihan program studi Statistika pada mahasiswa Statistika angkatan 2015. Iterasi yang digunakan dalam metode ini adalah 3 iterasi dengan hasil 2 <i>cluster</i> . <i>Cluster</i> pertama adalah pilihan program studi pertama, jalur masuk studi adalah SNMPTN, IPK memuaskan, rutinitas belajar sebanyak 4 kali dalam seminggu dan rata-rata lama belajar mahasiswa per hari adalah antara 60 menit sampai

Tahun	Nama	Judul	Hasil Penelitian
			dengan 120 menit. <i>Cluster</i> kedua adalah pilihan program studi pertama jalur masuk studi adalah SNMPTN, IPK sangat memuaskan, rutinitas belajar sebanyak 6 kali dalam seminggu dan rata-rata lama belajar mahasiswa per hari adalah kurang dari sama dengan 60 menit.
2018	Ewin Karman Nduru, Efori Buulolo dan Pristiwanto	IMPLEMENTASI ALGORITMA <i>K-Modes</i> UNTUK MENENTUKAN STRATEGI MARKETING STMIK BUDI DARMA	Metode <i>K-Modes Clustering</i> digunakan untuk mengelompokkan karakteristik mahasiswa yang kemudian dapat membantu sebagai bahan referensi pihak marketing untuk menentukan strategi marketing STMIK Darma Medan. Didapatkan dua <i>cluster</i> dengan <i>centroid cluster</i> pertama dengan asal sekolah SMK N 7 Medan, asal kota Medan, media promosi <i>website</i> dan jurusan Teknik Informatika. Pada <i>cluster 2</i> adalah sekolah yaitu SMA N 14 Medan, asal kota Medan, media spanduk dan jurusan Teknik Informatika
2019	Rendy Novianto dan Leonard Goeirmanto	Penerapan Data Mining Menggunakan Algoritma K-Means <i>Clustering</i> untuk Menganalisa Bisnis Perusahaan Asuransi	Tujuan dari penelitian ini adalah memberikan kemudahan penilaian <i>agent</i> yang ada pada perusahaan asuransi. <i>Agent</i> asuransi adalah sumber bisnis yang berpengaruh terhadap produksi. Metode yang digunakan dalam penelitian ini adalah <i>K-Means Clustering</i> . Hasil dari pengelompokkan dengan menggunakan data nilai pertanggungan, premi dan <i>claim</i> terdapat 3 <i>cluster</i> . <i>Centroid cluster</i> pertama memiliki nilai pertanggungan 91272615275.197, nilai premi 285410948.72036 dan nilai <i>claim</i> 36184286.250554. <i>Centroid cluster</i> kedua memiliki nilai pertanggungan 5549940477779.2, nilai premi

Tahun	Nama	Judul	Hasil Penelitian
			8532317279 dan nilai <i>claim</i> 3255999387.6667. <i>Centroid cluster</i> ketiga memiliki nilai pertanggungangan 17725681470671, nilai premi 33442097141 dan nilai <i>claim</i> 10000000000.
2020	Fang Yuan, Youlong Yang, Tiantian Yuan	<i>A dissimilarity measure for mixed nominal and ordinal attribute data in K-Modes algorithm</i>	Gagasan baru untuk pengukuran jarak data kategorik terutama pada data berskala ordinal. Pada penelitian ini, dibedakan pengukuran kesamaan data ordinal dan data nominal sehingga tidak mengurangi informasi dari data ordinal.
2020	Affif Fauzi, Nurul Mega Saraswati dan Rito Cipta Sigitta Hariyono	PENERAPAN ALGORITMA <i>K-MODES</i> DAN C4.5 UNTUK PREDIKSI PEMILIHAN JURUSAN DI UNIVERSITAS PERADABAN PADA SISWA SMA (Studi Kasus: SMA Islam Ta'allumul Huda Bumiayu)	Peneliti mengangkat permasalahan dalam pemilihan jurusan kuliah. Pemilihan jurusan kuliah yang kurang tepat dapat mempengaruhi proses pembelajaran selama perkuliahan. Algoritma <i>K-Modes</i> untuk melakukan prediksi pemilihan jurusan menghasilkan nilai akurasi sebesar 87.5%
2020	Ichwanto Hadi, Leo Willyanto Santoso dan Alvin Nathaniel Tjondrowiguno	Sistem Rekomendasi Film menggunakan <i>User-based Collaborative Filtering</i> dan <i>K-Modes Clustering</i>	Analisis <i>K-Modes Clustering</i> dapat digunakan untuk membuat sistem rekomendasi film. Kegunaan metode ini adalah penambah ketepatan dari rekomendasi dengan mengelompokkan riwayat preferensi <i>user</i> . Jumlah <i>cluster</i> terbaik yang dihasilkan dengan metode <i>elbow</i> dan <i>silhouette</i> adalah 3 <i>cluster</i> .
2020	Rachmi Marsheilla Agus dan Eko Bagus Fahrizqi	Analisis Tingkat Kepercayaan Diri Saat Bertanding Atlet Pencak Silat Perguruan Satria Sejati	Teknik analisis data yang digunakan peneliti adalah teknik analisis deksriptif kuantitatif. Data yang digunakan oleh peneliti adalah data survei berkategori dengan skala data likert. Peneliti menggunakan diagram batang untuk membantu menggambarkan besarnya tingkat kepercayaan atlet putra dan putri dengan satuan persen.

Tahun	Nama	Judul	Hasil Penelitian
2021	Septyan Eka Prastya, Muhammad Zulfadhilah, Nurhaeni	<i>Implementation of K-Modes Clustering in Determining Traffic Accident Patterns</i>	Pengaplikasian klusterisasi dengan algoritma <i>K-Modes</i> menghasilkan pola dari kecelakaan kendaraan. Pola yang terbentuk adalah tiga dimana <i>cluster</i> pertama dengan ukuran data terbanyak diprofilisasi sebagai ketika cuaca cerah, terdapat kecelakaan ganda di jalan dengan jalan lurus yang melibatkan pengendara motor dengan pengendara motor.
2021	Ni Putu Mira Novita Dewi dan Ida Bagus Gede Dwidasmaras	<i>Implementation of K-Modes Algorithm for Clustering of Stress Causes in University Students</i>	Penelitian ini bertujuan untuk mengidentifikasi penyebab stres dalam kelompok mahasiswa dan hubungan faktor internal dan faktor eksternal. Algoritma yang digunakan adalah algoritma <i>K-Modes</i> dengan penentuan jumlah <i>cluster</i> dengan menggunakan metode <i>elbow</i> . Peneliti mendapatkan tiga <i>cluster</i> yang mana <i>cluster</i> 1 adalah grup yang level stress ringan dengan penyebab utama adalah permasalahan akademis dan mayoritas bukan dikarenakan masalah keuangan. <i>Cluster</i> 2 merupakan <i>cluster</i> dengan tingkat stress tinggi dengan penyebab stress bervariasi, namun lebih berfokus pada karir dan keterlibatan dalam tempat tinggal, kelompok dan masyarakat. <i>Cluster</i> 3 adalah <i>cluster</i> dengan tingkat stress medium dan didominasi oleh laki-laki. Penyebab utama stress dalam <i>cluster</i> ini adalah akademis dan finansial
2021	Yolanda Aprilia Putri Kartikasari, Yosep Agus Pranoto dan Deddy Rudhistiar	PENERAPAN METODE <i>K-MODES</i> UNTUK PROSES PENENTUAN PENERIMA BANTUAN LANGSUNG TUNAI (BLT)	Penelitian ini menerapkan metode <i>K-Modes Clustering</i> untuk membantu menentukan penerima BLT. Sistem penentuan penerima BLT diujikan dengan menggunakan data lama didapatkan tiga

Tahun	Nama	Judul	Hasil Penelitian
			<i>cluster</i> dengan nilai akurasi 96.875%.

Berdasarkan penelitian yang dilakukan oleh peneliti-peneliti sebelumnya, belum terdapat penelitian yang menggunakan metode *K-Modes Clustering* dengan perhitungan jarak Yuan pada asuransi perjalanan. Sehingga penelitian ini difokuskan pada implementasi dari metode *K-Modes Clustering* dengan perhitungan jarak Yuan pada asuransi perjalanan sebagai kebaruan dari penelitian sebelumnya.

BAB 3

LANDASAN TEORI

3.1. Data

Data merupakan sekumpulan fakta yang mengandung informasi. Menurut (Siyoto, 2015) data berdasarkan sifatnya dapat dibedakan menjadi dua yaitu data kualitatif (kategorik) dan data kuantitatif (numerik). Data numerik merupakan data yang berbentuk angka dengan dua bentuk informasi yaitu informasi diskrit dan informasi kontinu. Diskrit berarti dalam bentuk bilangan yang diperoleh dengan cara membilang atau mencacah, sedangkan sifat kontinu adalah data dalam bentuk angka atau bilangan yang diperoleh dari hasil pengukuran. Data kategorik merupakan data yang berupa kata-kata yang dapat diterjemahkan ke dalam bentuk angka. Data kategorik diperoleh dari teknik pengumpulan data seperti wawancara, analisis dokumen, diskusi terfokus atau transkrip. Bentuk lain dari data kategorik adalah dalam bentuk gambar maupun video. Penelitian ini akan berfokus pada data kategorik.

3.1.1 Data Kategorik

Data kategorik terbagi menjadi dua yaitu data nominal dan data ordinal. Menurut (Siyoto, 2015) data nominal merupakan data yang diperoleh melalui pengelompokan objek berdasarkan kategori tertentu. Perbedaan antar kategori pada objek hanya bersifat pembeda. Data nominal dapat dinyatakan dengan angka, namun angka tersebut bersifat hanya sebagai pembeda saja tidak bersifat sebagai urutan atau makna matematis sehingga tidak dapat dibandingkan. Contoh data nominal pada jenis kelamin yaitu misal laki-laki dilambangkan dengan angka 1 dan perempuan dilambangkan dengan angka 2. Angka 1 untuk laki-laki dan 2 untuk perempuan digunakan hanya untuk simbolisasi yang bersifat membedakan dua kategori jenis kelamin. Angka 2 tidak berarti lebih besar dari angka 1 karena laki-laki tidak memiliki mana lebih besar dari perempuan.

Menurut (Siyoto, 2015) data ordinal adalah data yang disusun secara berjenjang menurut besarnya yang makna sebagai pembeda dan juga memiliki arti pada tingkatannya. Data ordinal memiliki tingkatan tertentu yang dapat diurutkan mulai dari tingkatan terendah hingga tertinggi atau sebaliknya. Jarak atau rentang

antar jenjang tidak harus sama. Contoh dari data ordinal adalah tingkat pendidikan. Misal Taman Kanak-Kanak (TK) didefinisikan sebagai 1, Sekolah Dasar (SD) didefinisikan sebagai 2, Sekolah Menengah Pertama (SMP) didefinisikan sebagai 3, Sekolah Menengah Atas (SMA) sebagai 4, Diploma sebagai 5 dan Sarjana sebagai 6. SD memiliki tingkatan lebih tinggi bila dibandingkan dengan TK namun lebih rendah bila dibandingkan dengan SMP. Namun tingkatan ini tidak dapat dioperasikan matematika, misal SD (2) dijumlahkan dengan SMP (3) tidak berarti akan menghasilkan nilai (5) yaitu Diploma.

3.1.2 Data Mining

Data mining memiliki beberapa padanan kata seperti *knowledge discovery* atau *pattern recognition*. Istilah *knowledge discovery* memiliki arti bahwa tujuan *data mining* adalah mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. Istilah *pattern recognition* memiliki arti pengetahuan yang hendak digali berbentuk pola-pola yang masih perlu digali dari dalam bongkahan data yang tengah dihadapi. Definisi *data mining* sendiri merupakan suatu disiplin ilmu yang memiliki tujuan untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang telah dimiliki. (Susanto & Suryadi, 2010). Terdapat enam fungsi pokok dalam *data mining* yaitu fungsi deskripsi, estimasi, prediksi, klasifikasi, pengelompokan dan asosiasi. Dari enam fungsi tersebut dapat disederhanakan menjadi dua fungsi yaitu fungsi minor dan fungsi mayor. Fungsi minor meliputi tiga fungsi yang pertama yaitu deskripsi, estimasi dan prediksi. Sedangkan fungsi mayor meliputi tiga fungsi lainnya yaitu klasifikasi, pengelompokan dan asosiasi.

3.2. Statistika Deskriptif

Statistika deskriptif atau statistika deduktif merupakan suatu metode tentang bagaimana cara mengumpulkan angka-angka, melabelkan angka, menggambarannya, mengolah dan menganalisis angka-angka tersebut serta menginterpretasikannya dengan memberikan penafsiran. Statistika deskriptif biasa disajikan dalam bentuk grafik untuk dianalisis dan ditafsirkan dengan mengambil kesimpulan. Terdapat dua cara penyajian data yang sering digunakan, yaitu tabel dan diagram. Tabel terdiri dari berbagai jenis di antaranya tabel baris kolom, tabel kontingensi dan tabel distribusi frekuensi. Sedangkan diagram terdiri dari diagram

batang, diagram garis, diagram *ogive*, diagram *lorents* dan diagram lingkaran. (Silvia, 2020)

3.3. Python

Python merupakan bahasa pemrograman yang umum digunakan dalam berbagai tujuan penggunaan. Bahasa ini mirip dengan bahasa pemrograman lainnya seperti C++, *JavaScript* atau C# dan *Java Oracle*. Pada sekitar tahun 1980-an oleh Guido van Rossum, nama *Python* diambil dari program favorit beliau yaitu “*Monty Pythons Flying Circus*”. Bahasa pemrograman *Python* populer digunakan karena beberapa faktor berikut: (Hunt, 2019)

- a. Fleksibel dan sederhana penggunaannya sehingga mudah untuk dipelajari.
- b. Bahasa pemrograman ini digunakan oleh komunitas *Data Science* dimana *Python* memberikan bahasa pemrograman standar yang lebih jika dibandingkan dengan bahasa pemrograman lain seperti R.
- c. Bahasa pemrograman ini sesuai digunakan sebagai bahasa naskah pemrograman untuk tipe pekerjaan seperti DevOps dimana pekerjaan ini membutuhkan level pemrograman lebih tinggi dibandingkan bahasa pemrograman tradisional.
- d. *Python* mampu untuk dijalankan hampir di sistem operasi apapun, khususnya *Windows*, *MacOs* dan *Linux*.
- e. *Python* menyediakan modul yang jangkauannya luas dimana modul tersebut dapat digunakan untuk mengembangkan fitur dasar dari bahasa pemrograman tersebut.
- f. *Python* merupakan bahasa pemrograman gratis.

3.4. Analisis K-Modes Clustering

Analisis *cluster* merupakan metode untuk mengelompokkan data-data yang memiliki kemiripan pada variabel ke dalam salah satu dari sekian kelompok. Sedangkan data-data yang kurang memiliki kesamaan variabel akan ditempatkan pada kelompok yang berbeda (Susanto & Suryadi, 2010).

3.4.1 Perhitungan Kedekatan Jarak Data Kategorik

Permasalahan analisis data bersifat nominal telah banyak dilakukan di literatur, namun tidak demikian untuk data yang bersifat ordinal. Mayoritas literatur masih memperlakukan data ordinal sama dengan data nominal. Efek dari

penanganan yang salah dengan metode semacam ini yaitu mengakibatkan informasi berupa tingkatan dari data ordinal hilang.

Didefinisikan satu data *set* bertipe campuran $X = \{x_1, x_2, \dots, x_n\}$ dengan n titik data. x_i adalah sebuah titik data berdimensi- d yang digambarkan sebagai berikut :

$$x_i = (nom_i^1, nom_i^2, \dots, nom_i^p, ord_i^{p+1}, \dots, ord_i^{p+q}), \quad p + q = d \quad (3.1)$$

$\{nom_i^1, nom_i^2, \dots, nom_i^p\}$ adalah sebuah *subset* nominal dari atribut dari x_i tanpa urutan, dengan kata lain dengan p merupakan indeks untuk banyaknya variabel nominal, tidak terdapat hubungan urutan atau ranking antara nilai atribut. $\{ord_i^{p+1}, \dots, ord_i^{p+q}\}$ merepresentasikan sebuah *subset* ordinal dari atribut dari x_i , dengan q merupakan indeks banyaknya variabel ordinal, yang mana nilai atribut punya makna urutan atau ranking. Kemudian didefinisikan $\eta_k (k = 1, 2, \dots, p)$ sebagai ukuran dari domain dari nom^k dan $\delta_l (l = p + 1, \dots, p + q)$ sebagai ukuran dari domain dari ord^l .

3.4.2 Perhitungan Jarak Data Nominal

Definisi 3.1 Secara formal, sebuah sistem informasi nominal adalah sebuah lipat empat $IS = (U, A, V, f)$, dimana:

U : set tak kosong dari titik data, disebut *universe*

A : set tak kosong dari variabel

V : gabungan dari semua domain variabel, contoh, $V = \cup_{a \in A} V_a$, dimana V_a adalah domain dari atribut a dan bersifat terbatas dan tidak berurutan;

$f: U \times A \rightarrow V$: sebuah pemetaan disebut sebuah fungsi informasi sebagai contoh untuk sebarang $x \in U$ dan $a \in A$, $f(x, a) \in V_a$.

Dalam penelitian ini, U dan A dari sistem informasi nominal mengacu pada kumpulan data X dan $\{nom_i^1, nom_i^2, \dots, nom_i^p\}$ yang mana adalah *subset* nominal atribut dari set data X , secara berurut.

Definisi 3.2 Diberikan $IS = (U, A, V, f)$ sebagai sistem informasi nominal dan $P \subseteq A$, sebuah relasi biner $IND(P)$, disebut sebagai hubungan yang tidak dapat dibedakan (*indiscernibility relation*), didefinisikan sebagai :

$$IND(P) = \{(x, y) \in U \times U \mid \forall a \in P, f(x, a) = f(y, a)\} \quad (3.2)$$

Diberikan $U/IND(P) = \{[x]_p \mid x \in U\}$ sebagai bagian dari U yang dipaksakan oleh hubungan yang tidak dapat dibedakan $IND(P)$, dimana $[x]_p$ menunjukkan kelas ekuivalen (sama) yang ditentukan oleh x dengan memperhatikan P , sebagai contoh $[x]_p = \{y \in U \mid (x, y) \in IND(P)\}$. Kemudian dipunyai pengukuran kesamaan sebagai berikut.

Definisi 3.3 Diberikan $IS = (U, A, V, f)$ sebagai sistem informasi nominal dan $P \subseteq A$. Untuk sebarang $a \in P$ dan $x, y \in U$, sebuah pengukuran kemiripan antara objek x dan objek y dengan memperhatikan a didefinisikan sebagai

$$sim_a(x, y) = \mu_{\{y\}}^{\{a\}}(x) = \frac{|[x]_{\{a\}} \cap \{y\}|}{|[x]_{\{a\}}|} \quad (3.3)$$

Dalam definisi 3.3, $\mu_{\{y\}}^{\{a\}}(x)$ adalah fungsi keanggotaan *rough* adalah teori *rough set* dan memiliki nilai antara 0 dan 1. Domain dari fungsi keanggotaan *rough* adalah sebuah object y dari U . Kesamaan antara titik x dan titik y didefinisikan sebagai derajat dari hubungan tumpang tindih antara titik x dan titik y . Kesamaan $sim_a(x, y)$ dapat dideskripsikan sebagai

$$sim_{nom^k}(x, y) = \frac{f(x, nom^k) \equiv f(y, nom^k)}{\sum_{z \in U} f(x, nom^k) \equiv f(z, nom^k)} \quad (3.4)$$

dimana $x, y \in U$, dan $nom^k \in \{nom_1^1, nom_1^2, \dots, nom_1^p\}$, dan

$$f(x, nom^k) \equiv f(y, nom^k) = \begin{cases} 1 & \text{jika } f(x, nom^k) = f(y, nom^k) \\ 0 & \text{lainnya} \end{cases} \quad (3.5)$$

Sifat 3.1 Diberikan $IS = (U, A, V, f)$ sebagai sistem informasi nominal dan $P \subseteq A$. Untuk sebarang $a \in P$ dan $x, y \in U$, dipunyai:

1. Simetri $sim_a(x, y) = sim_a(y, x)$
2. Minimum $sim_a(x, y) = 0$ jika $f(x, a) \neq f(y, a)$
3. Maksimum $sim_a(x, y) = 1$ jika $f(x, a) = f(y, a)$ dan $|[x]_{\{a\}}| = 1$

Dalam algoritma pengelompokkan, kalkulasi kesamaan dengan prinsip yang telah disebutkan diatas membantu untuk menentukan lebih banyak kelompok yang padu dan erat dimana objek digrupkan ke dalam kelompok yang sama berkemungkinan besar membagikan nilai spesial dan berkarakter. Berdasarkan hal tersebut, definisi dari ketidaksamaan antara dua objek didefinisikan sebagai berikut

Definisi 3.4 Diberikan $IS = (U, A, V, f)$ sebagai sistem informasi nominal A mengacu pada $\{nom_i^1, nom_i^2, \dots, nom_i^p\}$. Untuk sebarang $a \in P$ dan $x, y \in U$ pengukuran ketidaksamaan dari atribut nominal antara x dan y dengan m,encau pada A didefinisikan sebagai

$$Ndis(x, y) = \sum_{nom^k \in U} Ndis_{nom^k}(x, y), \text{ dimana}$$

$$Ndis_{nom^k}(x, y) = 1 - sim_{nom^k}(x, y) \text{ dan } k = 1, 2, \dots, p \quad (3.6)$$

3.4.3 Perhitungan Jarak Data Ordinal

Seperti yang disebutkan sebelumnya $X = \{x_1, x_2, \dots, x_n\}$ dengan n titik data adalah satu set data bertipe campuran yang memiliki d -atribut.

$$x_i = (nom_i^1, nom_i^2, \dots, nom_i^p, ord_i^{p+1}, \dots, ord_i^{p+q}), \quad (3.7)$$

$$x_j = (nom_j^1, nom_j^2, \dots, nom_j^p, ord_j^{p+1}, \dots, ord_j^{p+q}),$$

Adalah dua titik data yang berubah-ubah, $\{ord_i^{p+1}, \dots, ord_i^{p+q}\}$ dan $\{ord_j^{p+1}, \dots, ord_j^{p+q}\}$ adalah bagian ordinal yang berhubungan dari dua titik. Jika ord^l sebagai variabel ordinal $ke - l$ dengan $l : (p + 1 \leq l \leq p + q)$ atribut ordinal dan δ_l sebagai ukuran yang berhubungan dengan domain dari ord^l .

Definisi 3.5 $MSOA(ord^l)$ adalah sebuah set mengandung atribut nominal nom^k yang mana menyesuaikan kondisi dari $|\eta_k - \delta_l| = \min\{|\eta_i - \delta_l|\}$ ($i = 1, 2, \dots, p$), dimana η_k adalah ukuran dari domain dari $nom^k, \eta_i (i = 1, 2, \dots, p)$ adalah ukuran dari domain yang atribut nominalnya berhubungan.

$MSOA(ord^l)$ biasanya mengandung satu atau dua elemen. *Set* ini adalah basis dari formula jarak atribut untuk klasifikasi urutan. Secara efektif dapat menyelesaikan dimensi dari pengukuran kesamaan dari atribut ordinal, sehingga tidak ada perbedaan dimensi yang besar antara pengukuran ketidaksamaan atribut ordinal dan ketidaksamaan dari atribut nominal. Setelah menyelesaikan kunci permasalahan yang disebutkan di atas, didapatkan formula jarak dari atribut ordinal.

Pengukuran ketidaksamaan $Odis_{ord^l}(x_i, x_j)$ dikalkulasikan sebagai berikut:

$$Odis_{ord^l}(x_i, x_j) = \begin{cases} 1 - sim_{ord^l}(x_i, x_j) & 1 \leq \delta_l \leq 2, \\ |ord_i^l - ord_j^l| Dod_1 & 2 < \delta_l \leq 2\eta_{max}, \\ |ord_i^l - ord_j^l| Dod_2 & \delta_l > 2\eta_{max} \end{cases} \quad (3.8)$$

dengan δ_l sebagai ukuran domain variabel ordinal.

Pengukuran ketidaksamaan di atas dari atribut ordinal adalah *piecewise function*. Parameter Dod_1 dan Dod_2 dalam setiap ekspresi aljabar adalah inti permasalahan dalam kalkulasi dari pengukuran ketidakmiripan dari atribut ordinal. Hal ini mengindikasikan perbedaan dari derajat antara dua kedekatan nilai atribut ordinal (Dod). Setiap parameter dibahas secara rinci sebagai berikut

- 1) $1 \leq \delta_l \leq 2$

Ketika $1 \leq \delta_l \leq 2$ bernilai benar, artinya ord^l hanya mempunyai dua nilai yang berbeda. Ketidaksamaan antara dua nilai yang berbeda mirip dengan atribut nominal. Sehingga digunakan pengukuran ketidaksamaan dari atribut nominal untuk mengkalkulasi ketidaksamaan dari dua nilai atribut dalam situasi ini. $sim_{ord^l}(x_i, x_j)$ dihitung berdasarkan definisi 3.4.

- 2) $2 < \delta_l \leq 2\eta_{max}$,

Seperti yang dijelaskan di atas, $\eta_k (k = 1, 2, \dots, p)$ adalah ukuran domain dari nom^k . $\eta_{max} = \max \{\eta_1, \eta_2, \dots, \eta_p\}$ merupakan nilai maksimum dari ukuran domain atribut nominal. Dalam rangka menghitung Dod_1 digunakan nilai $MSOA(ord^l)$.

Pertama, $MSOA(ord^l)$ dihitung dengan definisi 3.5. Kedua, setelah mendapatkan $MSOA(ord^l)$, dihitung semua $sim_{nom^t}(x, y)$ tak kosong untuk setiap atribut nominal nom^t dari $MSOA(ord^l)$ berdasarkan definisi 3.3.

$sim_{nom^t}(x, y) \neq 0$ jika $f(x, nom^t) = f(y, nom^t)$, perhitungan yang dibutuhkan adalah $sim_{nom^t}(x, y)$ dari sepasang nilai atribut yang sama untuk setiap $nom^t \in MSOA(ord^l)$. Kemudian $Dod_1 = \min \{sim_{nom^t}\}$ dimana $nom^t \in MSOA(ord^l)$.

- 3) $\delta_l > 2\eta_{max}$

Dalam kondisi ini, formula perhitungan Dod_2 dijelaskan sebagai berikut

$$Dod_2 = \frac{Dod_1}{2|\delta_l - 2\eta_{max}|} \quad (3.9)$$

dimana $\eta_{max} = \max \{\eta_1, \eta_2, \dots, \eta_p\}$ dan Dod_1 dihitung berdasarkan bagian kedua.

3.4.4 Pengukuran Ketidaksamaan Antara Dua Objek

Untuk sebarang dua objek x_i dan x_j ,

$\{nom_i^1, nom_i^2, \dots, nom_i^p, ord_i^{p+1}, \dots, ord_i^{p+q}\}$ dan

$\{nom_j^1, nom_j^2, \dots, nom_j^p, ord_j^{p+1}, \dots, ord_j^{p+q}\}$, merepresentasikan

atribut nominal dan ordinal dari x_i, x_j secara berurutan. Didefinisikan Dis sebagai pengukuran ketidaksamaan dari x_i dan x_j , didapatkan fungsi sebagai berikut

$$Dis(x_i, x_j) = Ndis(x_i, x_j) + \sum_{ord^l} \gamma_l Odis(x_i, x_j) \quad (3.10)$$

Dimana $\gamma_l \ l = p + 1, p + 2, \dots, p + q$ adalah sebuah parameter untuk menstabilkan bagian dari variabel ordinal, umumnya bernilai 1. Berdasarkan **Sifat 3.1** dan **Persamaan 3.8**, dapat dengan mudah dibuktikan **Sifat 3.2** berikut

Sifat 3.2 Pengukuran ketidaksamaan $Dis(x, y)$ dipunyai

1. Simetri $Dis(x, y) = Dis(y, x)$
2. Bernilai positif $Dis(x, y) \geq 0$
3. Ketidaksetaraan segitiga $Dis(x, y) + Dis(y, z) \geq Dis(x, z)$

Sifat ini menunjukkan pengukuran ketidaksamaan $Dis(x, y)$ adalah metrik jarak. Berlawanan dengan pengukuran ketidaksamaan Huang dan Cao, ketidaksamaan yang baru diilustrasikan sebagai berikut:

Contoh apabila dimiliki data dalam **Tabel 3.1** yang ditampilkan sebagai berikut.

Tabel 3.1 Data Percontohan

Objek	A1	A2	A3	A4	A5	A6
x_1	3	3	2	3	3	2
x_2	3	3	2	3	3	3
x_3	3	1	3	2	1	3
x_4	4	1	5	2	1	3
x_5	3	2	3	2	1	2

Objek	A1	A2	A3	A4	A5	A6
x_6	3	3	2	1	1	2
x_7	3	3	2	4	1	2

A1, A2, A3, A4, A5 dan A6 merupakan variabel dari data tersebut dengan variabel A4 merupakan variabel berskala ordinal sedangkan variabel lainnya merupakan variabel berskala nominal. Data pada **Tabel 3.1** digunakan untuk perhitungan ketidaksamaan yang baru antara dua objek :

(1) Pertama, dihitung δ_{a_4} yang mana ukuran dari domain dari atribut a_4 ;

$$\delta_{a_4} = 4$$

(2) Kedua, didefinisikan $MSOA(a_4)$ sebagai $\{a_1\}$ berdasarkan **Definisi 3.5** dan **Persamaan 3.8**.

(3) Ketiga, dikalkulasikan semua $sim_{nom}^t(x, y)$ untuk atribut nominal a_1 adalah $\{1, \frac{1}{4}\}$;

(4) Keempat, Dod_1 bernilai $\frac{1}{4}$;

(5) Kelima, kalkulasikan ketidaksamaan berdasarkan **Persamaan 3.10**:

$$Dis(x_1, x_6) = 1 - \frac{1}{4} + 1 - \frac{1}{2} + 1 - \frac{1}{2} + 2 * \frac{1}{4} + 1 + 1 - \frac{1}{2} = \frac{15}{4};$$

$$Dis(x_1, x_7) = 1 - \frac{1}{4} + 1 - \frac{1}{2} + 1 - \frac{1}{2} + 1 * \frac{1}{4} + 1 + 1 - \frac{1}{2} = \frac{14}{4};$$

$$Dis(x_2, x_6) = 1 - \frac{1}{4} + 1 - \frac{1}{2} + 1 - \frac{1}{2} + 2 * \frac{1}{4} + 1 + 1 = \frac{17}{4};$$

$$Dis(x_2, x_7) = 1 - \frac{1}{4} + 1 - \frac{1}{2} + 1 - \frac{1}{2} + 1 * \frac{1}{4} + 1 + 1 = \frac{16}{4};$$

$$Dis(x_3, x_6) = 1 - \frac{1}{4} + 1 + 1 + 1 * \frac{1}{4} + 1 - 1/3 + 1 = 56/12;$$

$$Dis(x_3, x_7) = 1 - \frac{1}{4} + 1 + 1 + 2 * \frac{1}{4} + 1 - 1/3 + 1 = 59/12;$$

$$Dis(x_4, x_6) = 1 + 1 + 1 + 1 * \frac{1}{4} + 1 - 1/3 + 1 = 59/12;$$

$$Dis(x_4, x_7) = 1 + 1 + 1 + 2 * \frac{1}{4} + 1 - 1/3 + 1 = 68/12;$$

$$Dis(x_5, x_6) = 1 - \frac{1}{4} + 1 + 1 + 1 * \frac{1}{4} + 1 - 1/3 + 1 - \frac{1}{2} = 50/12;$$

$$D_P(x_5, x_7) = 1 - \frac{1}{4} + 1 + 1 + 2 * \frac{1}{4} + 1 - 1/3 + 1 - \frac{1}{2} = 62/12;$$

Dimana $\gamma_{a4} = 1$

Parameter γ_i merupakan parameter untuk menstabilkan bagian dari variabel ordinal dengan nilai yang umum digunakan adalah 1.

Kemudian, objek x_3, x_4, x_5 dapat dimasukkan ke dalam *cluster* 1 dikarenakan jarak terkecil x_3, x_4, x_5 dengan x_6 dan x_7 adalah dengan x_6 sehingga data tersebut dikelompokkan satu klaster dengan x_6 lalu disebut *cluster* 1. Objek x_1, x_2 dapat dimasukkan ke dalam *cluster* 2 dengan pertimbangan yang sama dengan *cluster* 1.

3.4.5 Algoritma Cluster K-Modes

Algoritma *K-Modes* menggunakan paradigma *k-means* untuk mencari sebuah bagian dari X yang meminimumkan fungsi objektif $P(W, Z)$. Objektivitas pengelompokkan sebuah data set dari n titik data digunakan untuk mencari W dan Z yang meminimumkan fungsi $P(W, Z)$. Ketika formula(2) digunakan sebagai pengukuran ketidaksamaan untuk objek bertipe campuran, fungsi objektifnya menjadi :

$$F(W, Z) = \sum_{l=1}^k \sum_{i=1}^n \omega_{li} Dis(z_l, x_i) \quad (3.11)$$

Tergantung pada :

$$\omega_{li} \in \{0,1\}, \quad 1 \leq l \leq k, \quad 1 \leq i \leq n, \quad (3.12)$$

$$\sum_{l=1}^k \omega_{li} = 1, \quad 1 \leq i \leq n, \quad (3.13)$$

Dan

$$0 < \sum_{i=1}^n \omega_{li} < n, \quad 1 \leq l \leq k, \quad (3.14)$$

Dimana $k (< n)$ dikenal sebagai jumlah kelompok, $W = [\omega_{li}]$ adalah sebuah matriks k -oleh- n $[0,1]$, $Z = [z_1, z_2, \dots, z_k]$ dan z_l adalah modus *cluster* ke- l dengan set atribut P .

Peminimalan dari F di (3.11) dengan batasan di (3.12) ke (3.14) membentuk sebuah kelas dari batasan permasalahan pengoptimalan non-linier yang mana pemecahannya belum diketahui. Metode yang biasa mengacu pada optimalisasi dari

F di (3.11) digunakan optimalisasi sebagian untuk Z dan W . Proses ini dapat diformulasikan algoritma K -Modes seperti berikut.

- Tahap 1. Pilih k objek yang berbeda z_1, z_2, \dots, z_k dari X sebagai modus inisial $Z^{(1)} = z_1, z_2, \dots, z_k \in U^k$. Tentukan $W^{(1)}$ seperti bagaimana $F(W^{(1)}, Z^{(1)})$ diminimumkan. Ditentukan $t = 1$
- Tahap 2. Tentukan $Z^{(t+1)}$ seperti bagaimana $F(W^{(t)}, Z^{(t+1)})$ diminimumkan. Jika $F(W^{(t)}, Z^{(t+1)}) = F(W^{(t)}, Z^{(t)})$, lalu berhenti
- Tahap 3. Tentukan $W^{(t+1)}$ seperti bagaimana $F(W^{(t+1)}, Z^{(t+1)})$ diminimumkan. Jika $F(W^{(t+1)}, Z^{(t+1)}) = F(W^{(t)}, Z^{(t+1)})$, lalu berhenti; sebaliknya tentukan $t = t + 1$ lalu gunakan tahap 2.

Kode semu dari algoritma K -Modes dengan pengukuran ketidaksamaan usulan dideskripsikan di algoritma 1

Algoritma 1 Algoritma K -Modes dengan pengukuran ketidaksamaan yang diusulkan

1. Inisialisasi *oldmodes* sebagai sebuah *array* kosong berdimensi $k \times d$, $d = p + q$
2. Pilih secara acak k objek yang berbeda x_1, x_2, \dots, x_k dari X dengan n objek
3. dan masukkan x_1, x_2, \dots, x_k untuk *array* $k \times d$ variabel *newmodes*
4. inisialisasi $t=0$ dan $F_t(W, Z) = 0$
5. ketika $F_{t-1}(W, Z) \neq F_t(W, Z)$ lakukan
6. untuk $i = 1$ ke n lakukan
7. untuk $l = 1$ ke k lakukan
8. untuk $j = 1$ ke p lakukan
9. hitung $Ndis(x_i, z_l)$ berdasarkan persamaan 3.6
10. berhenti untuk $j=1$ ke p
11. untuk $j = p+1$ ke $p+q$ lakukan
12. hitung $Odis_{ord_j}(x_i, z_l)$ berdasarkan persamaan 3.8
13. berhenti untuk $j = p+1$ ke $p+q$
14. hitung ketidaksamaan antara objek ke- i dan modus ke- l berdasarkan formula (2) dan klasifikasi objek ke- i ke dalam kelompok yang modusnya paling dekat ke objek tersebut.
15. tambahkan ketidaksamaan untuk $F_t(W, Z)$
16. berhenti untuk $l = 1$ ke k
17. berhenti untuk $i = 1$ ke n
18. untuk $l = 1$ ke k lakukan
19. cari modus z_l dari setiap pengelompokkan dan masukkan ke *newmodes*;
20. berhenti untuk $l = 1$ ke k
21. $t+1$

22. jika $F_{t-1}(W, Z) = F_t(W, Z)$ lalu
23. jeda;
24. berhenti untuk jika $F_{t-1}(W, Z) = F_t(W, Z)$
25. berhenti untuk ketika $F_{t-1}(W, Z) \neq F_t(W, Z)$

3.5. *Silhouette Coefficient*

Kualitas dan kekuatan dari *cluster* dapat dilihat dengan menggunakan metode *Silhouette Coefficient*. Metode *silhouette* didasarkan pada perhitungan penuh matriks pasangan jarak dari seluruh data. (Dewi & Pramita, 2019). Nilai *silhouette coefficient* didapatkan dari rata-rata nilai *Silhouette Indeks Global*. Langkah-langkah untuk menghitung nilai *silhouette coefficient* adalah sebagai berikut.

1. Menghitung nilai kohesi atau dapat dilambangkan dengan a_i^j yaitu rata-rata jarak ke- i terhadap semua data lain yang masih dalam satu *cluster*.

$$a_i^j = \frac{1}{m_j - 1} \sum_{r=1; r \neq i}^{m_j} d(x_i^j, x_r^j) \quad (3.15)$$

Keterangan :

$j = \text{cluster}$

$i = \text{indeks dari data } (i = 1, 2, \dots, m_j)$

$a_i^j = \text{rata-rata jarak data ke-}i \text{ terhadap semua data dalam satu } \text{cluster}$

$m_j = \text{jumlah data dalam } \text{cluster ke-}j$

$d(x_i^j, x_r^j) = \text{jarak data ke-}i \text{ dengan data ke-}r \text{ dalam satu } \text{cluster } j$

2. Menghitung nilai separasi atau dapat dilambangkan dengan b_i^j yaitu nilai minimum dari rata-rata jarak data ke- i dalam *cluster* ke- j dengan semua data di *cluster* selain j .

$$b_i^j = \min_{n=1, 2, \dots, k; n \neq k} \left\{ \frac{1}{m_n} \sum_{r=1; r \neq i}^{m_n} d(x_i^j, x_r^n) \right\} \quad (3.16)$$

Keterangan :

$j = \text{cluster}$

$i = \text{indeks dari data } (i = 1, 2, \dots, m_j)$

$b_i^j = \text{rata-rata jarak data ke-}i \text{ terhadap semua data dalam yang tidak termasuk ke dalam } \text{cluster yang sama dengan data ke-}i$

$m_n = \text{jumlah data dalam } \text{cluster ke-}n$

$d(x_i^j, x_r^n)$ = jarak data ke- i dengan data ke- r dalam satu *cluster* n

3. Menghitung nilai *silhouette coefficient* dengan menggunakan persamaan 3.17

$$s_i^j = \frac{b_i^j - a_i^j}{\max \{a_i^j, b_i^j\}} \quad (3.17)$$

Keterangan :

j = *cluster*

i = indeks dari data ($i = 1, 2, \dots, m_j$)

a_i^j = rata-rata jarak data ke- i terhadap semua data dalam satu *cluster*

b_i^j = rata-rata jarak data ke- i terhadap semua data yang tidak termasuk ke dalam *cluster* yang sama dengan data ke- i

s_i^j = nilai *silhouette coefficient* data ke- i dalam satu *cluster*

Nilai rata-rata *silhouette coefficient* global (SC) dapat dihitung dengan persamaan 3.18.

$$SC = \frac{1}{N} \sum_{i=1}^{m_j} s_i^j \quad (3.18)$$

dengan N adalah banyaknya data.

Nilai *silhouette coefficient* memiliki rentang pada nilai -1 hingga 1 dengan kriteria nilai diperlihatkan pada **Tabel 3.2**.

Tabel 3.2 Kriteria Nilai *Silhouette Coefficient*

Nilai <i>Silhouette Coefficient</i>	Kriteria
$0,70 < SC \leq 1,00$	Struktur Kuat
$0,50 < SC \leq 0,70$	Struktur Baik
$0,26 < SC \leq 0,50$	Struktur Lemah
$\leq 0,25$	Struktur Buruk

Nilai rata-rata *silhouette coefficient* pada suatu *cluster* menunjukkan ukuran kedekatan kemiripan data dalam satu *cluster*. Semakin nilai rata-rata mendekati 1 artinya semakin baik pengelompokkan data pada *cluster* tersebut, semakin nilai rata-rata mendekati -1 artinya semakin buruk pengelompokkan data pada *cluster* tersebut.

3.6. Asuransi

Menurut Otoritas Jasa Keuangan (OJK, 2022), asuransi merupakan perjanjian antara perusahaan asuransi dan pemegang polis yang menjadi dasar bagi penerimaan premi oleh perusahaan asuransi sebagai imbalan untuk :

- a. memberikan penggantian kepada tertanggung atau pemegang polis karena kerugian, kerusakan, biaya yang timbul, kehilangan keuntungan, atau tanggung jawab hukum kepada pihak ketiga yang mungkin diderita tertanggung atau pemegang polis karena terjadinya suatu peristiwa yang tidak pasti; atau
- b. memberikan pembayaran yang didasarkan meninggal atau hidupnya tertanggung dengan manfaat yang besarnya telah ditetapkan dan/atau didasarkan pada hasil pengelolaan dana.

Perusahaan asuransi merupakan sebuah institusi kegiatan usaha yang bergerak di bidang :

- a. jasa pertanggungan atau pengelolaan risiko;
- b. pertanggungan ulang risiko;
- c. pemasaran dan distribusi produk asuransi atau produk asuransi syariah;
- d. konsultasi dan keberantaraan asuransi, asuransi syariah, reasuransi, atau reasuransi syariah; atau
- e. penilai kerugian asuransi atau asuransi syariah.

Secara umum produk asuransi di Indonesia terbagi menjadi dua yaitu asuransi jiwa dan asuransi umum. Asuransi jiwa merupakan jenis asuransi yang memberikan proteksi kerugian secara finansial akibat terjadinya kematian yang manfaatnya dapat dirasakan oleh ahli waris. Asuransi umum merupakan jenis asuransi yang manfaat proteksi kerugiannya diberikan pada si tertanggung itu sendiri. (Lifepal, 2021)

Asuransi umum sendiri terbagi dalam berbagai jenis, salah satunya adalah asuransi perjalanan. Asuransi perjalanan atau *travel insurance* merupakan jenis asuransi yang memberikan perlindungan selama tertanggung melakukan perjalanan baik dalam maupun luar negeri. (OJK, 2022)

3.7. Pendapatan Perseorangan

Pendapatan perseorangan adalah jumlah pendapatan yang diterima oleh setiap orang dalam masyarakat termasuk pendapatan yang diperoleh tanpa memberikan suatu kegiatan apapun (Ani, 2012)(Pendidikan & Dinamika, 2012). Pendapatan setiap orang dapat dikelompokkan ke dalam lima kelas. Menurut *World Bank and the Organization for the Economic Cooperation and Development (OECD)*, kelompok tersebut terdiri dari *poor live* dimana pendapatan seseorang 2 USD atau kurang setiap harinya, *low income* dengan pendapatan 2.01 USD – 10 USD, *middle income* dengan pendapatan 10.01 USD – 20 USD, *upper-middle income* dengan 20.01 USD – 50 USD dan *high income* dengan pendapatan lebih dari 50 USD. (Misra, 2021)

BAB 4

METODOLOGI PENELITIAN

4.1. Populasi Penelitian

Populasi yang digunakan dalam penelitian ini adalah pelanggan dari sebuah penyedia jasa pelayanan tur dan perjalanan di India sebanyak 1987 pelanggan yang terdaftar dari penyedia jasa tersebut. Sampel yang digunakan sebanyak 1277 pelanggan yang tidak mendaftar asuransi perjalanan dan 710 pelanggan yang mendaftar asuransi perjalanan.

4.2. Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder yang diambil dari *website* Kaggle, dengan alamat <https://www.kaggle.com/tejashvi14/travel-insurance-prediction-data> pada tanggal 02 Desember 2021.

4.3. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdapat sebanyak 9 variabel berjenis kategorik. Tujuh variabel di antaranya berskala nominal dan 2 variabel berskala ordinal. Pada **Tabel 4.1** dijelaskan keterangan dari masing-masing variabel yang digunakan dalam penelitian ini.

Tabel 4.1 Deskripsi Variabel Penelitian

No	Variabel	Notasi	Skala Data	Keterangan
1	Umur	<i>Age</i>	Ordinal	Umur dari pengguna jasa agen perjalanan. Rentang umur dan kategori mengacu pada Badan Pusat Statistik (BPS). - 0 : Umur \leq 30 tahun - 1 : Umur $>$ 30 tahun
2	Tipe pekerjaan	<i>Employment Type</i>	Nominal	Tipe pekerjaan dari pengguna jasa agen perjalanan. Proses <i>encoding</i> label menjadi angka didapatkan dari <i>library Python</i> . - 0 : <i>Government Sector</i> (pekerja di sektor pemerintahan)

No	Variabel	Notasi	Skala Data	Keterangan
				- 1 : <i>Private Sector/Self employed</i> (pekerja di sektor swasta atau bekerja sendiri)
3	Pendidikan	<i>GraduateOrNot</i>	Nominal	<p>Tipe pendidikan dari pengguna jasa agensi perjalanan. Apakah pengguna jasa memiliki latar belakang lulusan perguruan tinggi atau tidak. Proses <i>encoding</i> label menjadi angka didapatkan dari <i>library Python</i>.</p> <ul style="list-style-type: none"> - 0 : <i>No</i> (bukan lulusan perguruan tinggi) - 1 : <i>Yes</i> (lulusan perguruan tinggi)
4	Pendapatan Per Tahun	<i>AnnualIncome</i>	Ordinal	<p>Pendapatan pengguna jasa selama satu tahun dengan mata uang <i>Rupee</i>. Kategori <i>Annual income</i> dihitung dengan mengalikan batasan <i>daily income</i> menurut OECD dengan 260 hari kerja dan kemudian dikonversikan ke dalam mata uang <i>Rupee</i>.</p> <ul style="list-style-type: none"> - 0 : <i>Poor</i> (pendapatan \leq 38626.6 <i>Rupee</i>) - 1 : <i>Low Income</i> (38626.6 < pendapatan \leq 193122.8 <i>Rupee</i>) - 2 : <i>Middle Income</i> (193122.8 < pendapatan \leq 386243 <i>Rupee</i>) - 3 : <i>Upper-Middle Income</i> (386243 < pendapatan \leq 965608.8 <i>Rupee</i>)

No	Variabel	Notasi	Skala Data	Keterangan
				- 4 : <i>High Income</i> (pendapatan > 965608.8 <i>Rupee</i>)
5	Jumlah Keluarga	<i>FamilyMembers</i>	Nominal	<p>Kelompok jumlah anggota keluarga dari pengguna jasa. Proses <i>encoding</i> label menjadi angka didapatkan dari <i>library Python</i>.</p> <ul style="list-style-type: none"> - 0 : < 4 (keluarga kecil) - 1 : 4 (keluarga Keluarga Berencana) - 2 : Lainnya (lebih dari 4 anggota keluarga atau keluarga besar)
6	Riwayat Penyakit	<i>ChronicDisease</i>	Nominal	<p>Riwayat penyakit dari pengguna jasa agen perjalanan. Apakah pengguna jasa memiliki riwayat penyakit kronis seperti Diabetes, Asma atau tidak.</p> <ul style="list-style-type: none"> - 0 : <i>No</i> (tidak punya) - 1 : <i>Yes</i> (punya)
7	Seberapa sering melakukan penerbangan	<i>FrequentFlyer</i>	Nominal	<p>Variabel yang menunjukkan seberapa seringnya pengguna jasa melakukan penerbangan. Berdasarkan data riwayat pemesanan tiket pesawat dari pengguna jasa, setidaknya 4 tujuan yang berbeda dalam kurun waktu 2 tahun (2017-2019). Proses <i>encoding</i> label menjadi angka didapatkan dari <i>library Python</i>.</p> <ul style="list-style-type: none"> - 0 : <i>No</i> (penerbangan < 4 tujuan yang berbeda) - 1 : <i>Yes</i> (penerbangan \geq tujuan yang berbeda)

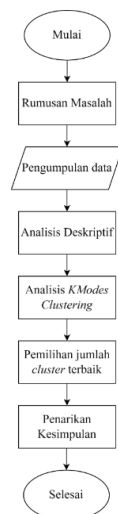
No	Variabel	Notasi	Skala Data	Keterangan
8	Penerbangan ke luar negeri	<i>EverTravelledAbroad</i>	Nominal	Variabel yang menunjukkan apakah pengguna jasa pernah melakukan perjalanan ke luar negeri. Proses <i>encoding</i> label menjadi angka didapatkan dari <i>library Python</i> . - 0 : <i>No</i> (belum pernah) - 1 : <i>Yes</i> (pernah)
9	Asuransi Perjalanan	<i>TravelInsurance</i>	Nominal	Variabel yang menjelaskan apakah pengguna jasa pernah membeli asuransi perjalanan atau belum. - 0 : <i>No</i> (belum pernah) - 1 : <i>Yes</i> (pernah)

4.4. Alat dan Cara Organisir Data

Alat yang digunakan dalam mengolah data pada penelitian ini adalah perangkat lunak *Microsoft Excel*, *Tableau* dan *Python*. Data diunduh dari *website* Kaggle dalam bentuk “*csv*” kemudian dilakukan visualisasi dengan menggunakan *Tableau*. Perangkat lunak *Python* digunakan untuk analisis *cluster K-Modes*.

4.5. Tahapan Penelitian

Pada penelitian ini dilakukan beberapa tahapan untuk mencapai hasil penelitian dengan alur sebagai berikut :

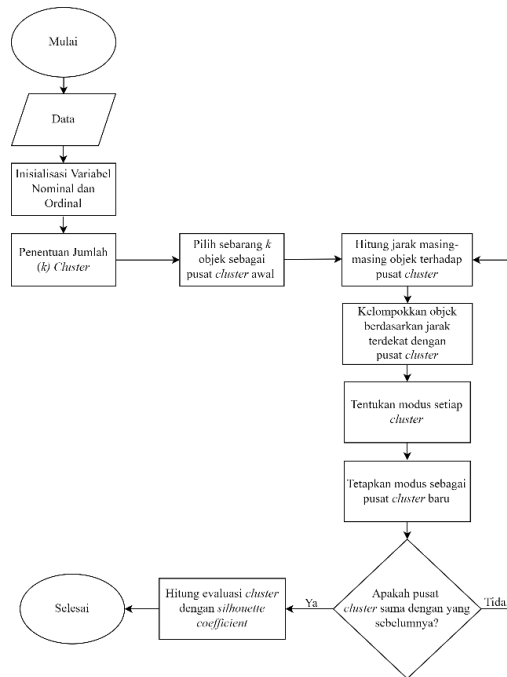


Gambar 4.1 Diagram Alir Tahapan Penelitian

Penelitian ini dimulai dari tahapan perumusan masalah yang diselaraskan dengan tujuan dari penelitian. Dalam penelitian ini peneliti ingin mengetahui bagaimana implementasi metode *K-Modes Clustering* dengan metode perhitungan jarak gagasan Yuan pada data asuransi perjalanan. Lalu bagaimana profil dari masing-masing *cluster* yang terbentuk. Setelah masalah dirumuskan, tahap berikutnya dari penelitian ini adalah pengumpulan data. Data yang dikumpulkan merupakan data pengguna jasa agen perjalanan dari *website* Kaggle.com. Data yang berhasil dikumpulkan kemudian dianalisis dengan menggunakan metode analisis deskriptif untuk diketahui bagaimana gambaran awal dari data tersebut. Tujuan dari analisis ini adalah agar data yang digunakan dapat lebih dikenali sebelum dilakukan analisis tahap lanjut. Kemudian setelah data dianalisis dengan analisis deskriptif, dilakukan analisis lebih lanjut dengan analisis pengelompokan *K-Modes Clustering* dengan perhitungan jarak gagasan Yuan. Pada tahapan ini dilakukan analisis pengelompokan dengan jumlah *cluster* yang berbeda yang kemudian dilakukan tahapan pemilihan jumlah *cluster* terbaik. Pemilihan jumlah *cluster* terbaik menggunakan metode *silhouette coefficient*. Kelompok yang telah terbentuk dari jumlah *cluster* terbaik kemudian dilakukan profilisasi agar didapatkan kesimpulan dan informasi terbaru dari data yang telah dikelompokkan.

4.5.1 Tahapan *K-Modes Clustering*

Pada tahapan *K-Modes Clustering* terjadi beberapa proses tahapan yang perlu dilakukan. Proses tersebut dijelaskan pada **Gambar 4.2**.



Gambar 4.2 Diagram Alir *K-Modes Clustering*

Tahapan *K-Modes Clustering* dijelaskan pada **Gambar 4.2** dengan rincian sebagai berikut :

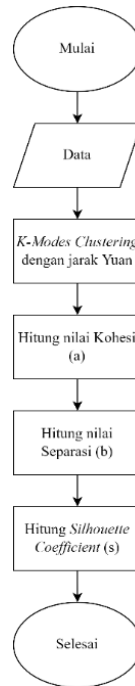
1. Penentuan variabel nominal dan ordinal.
2. Penentuan jumlah *cluster* sebanyak k .
3. Pemilihan pusat *cluster* awal secara acak dari objek sebanyak k .
4. Menghitung jarak nominal dan ordinal dari objek terhadap pusat *cluster* berdasarkan perhitungan jarak Yuan.
5. Mengelompokkan objek ke pusat *cluster* dengan jarak terdekat.
6. Mencari nilai modus masing-masing *cluster*.
7. Nilai modus kemudian digunakan sebagai pusat *cluster* baru.
8. Ulangi langkah 4 hingga 7 sampai pusat *cluster* tetap dan tidak ada objek yang berpindah *cluster*.
9. Menghitung nilai *silhouette coefficient* untuk mengevaluasi hasil *cluster*.

Metode *K-Modes Clustering* dengan perhitungan jarak Yuan digunakan pada penelitian ini dikarenakan perbedaan perlakuan perhitungan jarak pada variabel nominal dan ordinal. Perbedaan perhitungan jarak ini mengakibatkan informasi tingkatan pada variabel ordinal tidak berkurang. Variabel ordinal memiliki karakteristik yang berbeda dari variabel nominal. Pada variabel nominal, label digunakan hanya sebagai pembeda. Sedangkan, pada variabel ordinal, label tidak

hanya digunakan sebagai pembeda tetapi juga memiliki makna tingkatan. Sehingga perlakuan pada perhitungan jarak kedua variabel ini perlu dibedakan.

4.5.2 Tahapan Perhitungan *Silhouette Coefficient*

Silhouette Coefficient pada penelitian ini digunakan untuk menentukan jumlah *cluster* terbaik atau optimum. Tahap perhitungan dari *Silhouette Coefficient* digambarkan pada **Gambar 4.3**.



Gambar 4.3 Diagram Alir *Silhouette Coefficient*

Nilai *Silhouette Coefficient* didapatkan melalui tahapan :

1. Pertama melakukan pengelompokkan data. Pada penelitian ini digunakan metode pengelompokkan data *K-Modes Clustering* dengan perhitungan jarak Yuan.
2. Setelah data dikelompokkan kemudian dilakukan perhitungan nilai kohesi (a_i^j) yaitu rata-rata jarak ke- i terhadap semua data lain yang masih dalam satu *cluster*.
3. Kemudian dilakukan perhitungan untuk mencari nilai separasi (b_i^j) yaitu nilai minimum dari rata-rata jarak data ke- i dalam *cluster* ke- j dengan semua data di *cluster* selain j .

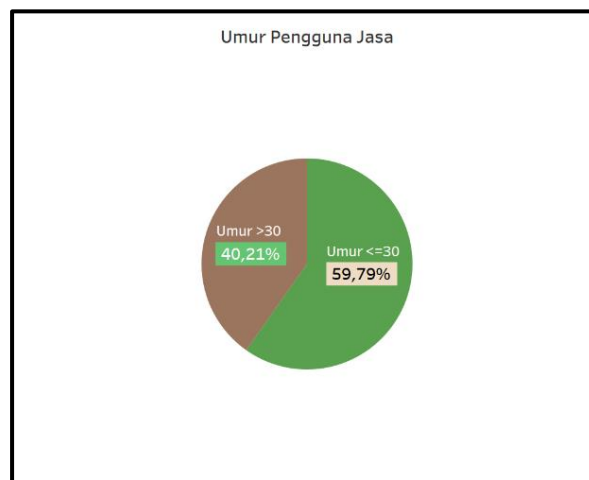
4. Nilai *silhouette coefficient* didapatkan dari pengurangan nilai separasi dengan nilai kohesi yang kemudian dibagi dengan nilai maksimum dari kohesi dan separasi.

BAB 5

HASIL DAN PEMBAHASAN

5.1. Analisis Deskriptif

Data yang digunakan merupakan data pengguna jasa agensi perjalanan dengan jumlah data sebanyak 1987 pengguna dengan sebanyak 1277 pengguna tidak menggunakan asuransi perjalanan dan sisanya yaitu sebanyak 710 pengguna menggunakan asuransi perjalanan. Berikut statistika deskriptif atau gambaran umum dari 1987 pengguna tersebut.

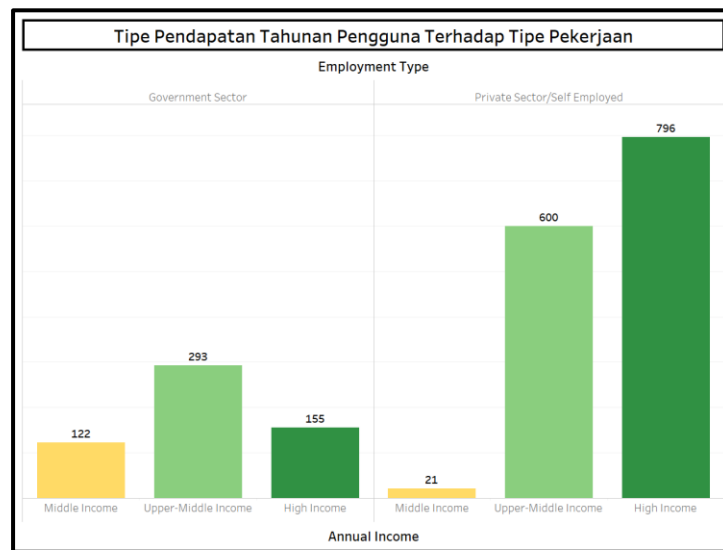


Gambar 5.1 Umur Pengguna Jasa

Umur pengguna jasa terbagi menjadi dua kategori yaitu umur pengguna di atas 30 tahun dan umur pengguna jasa maksimal 30 tahun. Dari **Gambar 5.1** didapatkan informasi bahwa pengguna jasa agensi mayoritas berumur kurang dari sama dengan 30 tahun. Dominasi umur pengguna muda ini dikarenakan sebanyak 66% tempat wisata di India dibuat untuk wisatawan muda. (PTI, 2016)

Tipe pekerjaan dari pengguna jasa agensi perjalanan dibagi menjadi dua tipe yaitu tipe pekerja pemerintahan dan pekerja sektor swasta maupun pribadi. Pengguna jasa agensi perjalanan ini, terbagi ke dalam tiga tipe pendapatan saja yaitu pendapatan menengah (*middle income*), pendapatan *upper-middle*, dan pendapatan tinggi. Pengguna jasa agensi perjalanan ini banyak didapati dari kalangan pekerja sektor swasta dan pribadi dengan pendapatan mayoritas kalangan ini adalah pendapatan tinggi (*high income*) yaitu sebanyak 796 orang. Pendapatan *upper-middle* pada kalangan pekerja sektor swasta dan pribadi terdiri dari sebanyak 600

orang, sedangkan pengguna dengan pendapatan menengah (*middle income*) hanya terdiri sebanyak 21 orang. Pada pengguna jasa dengan tipe pekerja pemerintahan, mayoritas terpusat pada pendapatan *upper-middle* yaitu sebanyak 293 orang. Pengguna dengan pendapatan tinggi dan menengah pada kalangan ini memiliki perbandingan jumlah yang relatif sama yaitu masing-masing sebanyak 155 orang dan 122 orang. **Gambar 5.2** merupakan visualisasi dari tipe pendapatan tahunan pengguna jasa terhadap tipe pekerjaan dari pengguna jasa itu sendiri.



Gambar 5.2 Tipe Pendapatan Terhadap Tipe Pekerjaan

Tidak adanya tipe pendapatan *low income* dan *poor income* dalam data ini dikarenakan rata-rata pengeluaran yang dibutuhkan untuk berwisata di India adalah sekitar 2.597 Rupee per hari (Budget Your Trip, 2022). Sedangkan masyarakat dengan pendapatan *low income* adalah 152,50 Rupee hingga 762,51 Rupee per hari dan untuk *poor income* adalah kurang dari 152,50 Rupee per hari sehingga kedua kategori pendapatan ini tidak memungkinkan untuk mengadakan perjalanan menggunakan agen perjalanan pada studi kasus ini. Pengguna jasa didominasi oleh pegawai swasta dikarenakan cuti tahunan pegawai swasta lebih banyak dibandingkan dengan cuti tahunan pegawai negeri (Paycheck.in, 2021) (GOVERNMENT OF INDIA MINISTRY OF PERSONNEL, P.G. & PENSIONS, 2001)

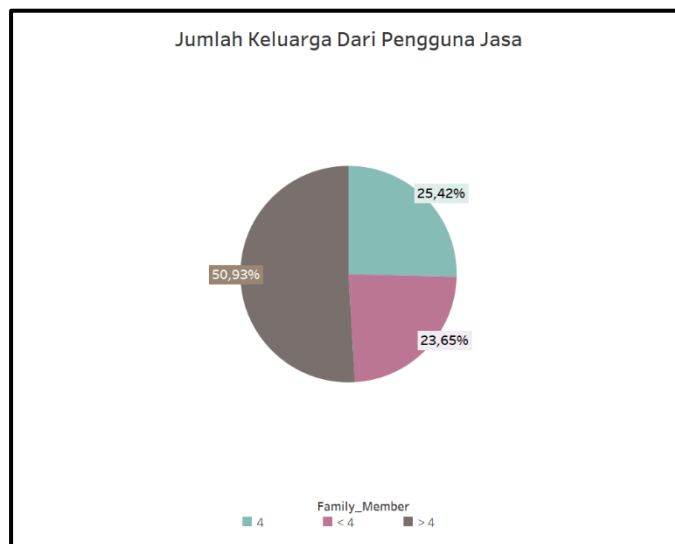
Pengguna jasa dari agensi pada penelitian ini mayoritas belum pernah bepergian ke luar negeri dan belum melakukan perjalanan sebanyak setidaknya 4 tujuan yang berbeda dalam kurun waktu 2 tahun (2017-2019). Sementara minoritas

pengguna pernah bepergian ke luar negeri dan sudah melakukan perjalanan sebanyak setidaknya 4 tujuan yang berbeda dalam kurun waktu 2 tahun (2017-2019). Perjalanan domestik lebih digemari masyarakat India karena masyarakat India cenderung menyukai *budget travel*. *Budget travel* adalah kegiatan berwisata yang murah dimana mencakup biaya akomodasi dan transportasi. Kegemaran akan *budget travel* ini dapat dicapai dengan perjalanan dalam negeri (Times Travel, 2019)

Ever Travelled Abroad	Frequent Flyer	
	No	Yes
No	1.358	249
Yes	212	168

Gambar 5.3 Frekuensi Penerbangan Pengguna Jasa

Pengguna jasa agensi perjalanan yang memiliki jumlah keluarga sebanyak 4 anggota berjumlah seperempat dari jumlah keseluruhan pengguna atau lebih tepatnya 25,42%. Pengguna jasa dengan jumlah anggota keluarga sebanyak < 4 hampir sebanyak jumlah keluarga sebanyak 4 yaitu sebesar 23,65% dari total keseluruhan pengguna. Sementara anggota keluarga lebih dari 4 merupakan kategori mayoritas yaitu sebesar 50,93%. Jumlah anggota keluarga yang lebih dari 4 ini selaras dengan data rata-rata jumlah keluarga di India sebesar 5.78. (Global Data Lab, 2019)



Gambar 5.4 Jumlah Keluarga dari Pengguna Jasa

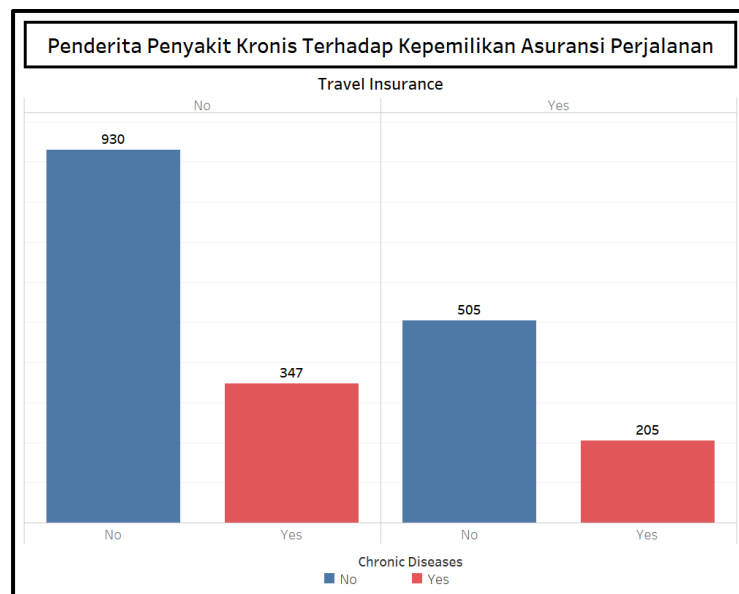
Pengguna jasa dari agensi perjalanan pada penelitian ini mayoritas merupakan lulusan perguruan tinggi yaitu sebesar 85,15% dari total pengguna jasa dari agensi

tersebut. Sementara hanya sebesar 14,85% pengguna jasa yang bukan merupakan lulusan dari perguruan tinggi.



Gambar 05.5 Pengguna Jasa Lulusan Perguruan Tinggi

Kebanyakan pengguna jasa agensi perjalanan tidak memiliki penyakit kronis sehingga mereka tidak mendaftar asuransi perjalanan. Namun pengguna yang memiliki penyakit kronis juga lebih banyak yang tidak mendaftar asuransi perjalanan dibandingkan yang mendaftar.



Gambar 5.6 Penderita Penyakit Kronis Terhadap Kepemilikan Asuransi

Data yang digunakan merupakan data pengguna jasa pada tahun 2019 dimana COVID-19 belum merebak di India. Sebelum virus COVID-19 merebak di India, masyarakat di India sebanyak 95% mengetahui kebutuhan akan berasuransi, namun

hanya 17% dari masyarakat India yang memiliki proteksi dengan cara berasuransi. (Nababan, 2014)

5.2. Sintaks Perhitungan Jarak dengan *Python*

Pada penelitian ini digunakan perangkat lunak *Python* untuk membentuk sintaks perhitungan jarak Yuan dengan cara menerjemahkan persamaan perhitungan jarak nominal dan ordinal ke dalam bahasa pemrograman *Python*.

5.2.1 Jarak Nominal

Tabel 5.1 Penjelasan Sintaks Perhitungan Jarak Nominal

Sintaks	Keterangan
<pre>def nominal (<i>cluster</i>) :</pre>	Pendefinisian fungsi nominal dengan masukan <i>cluster</i> yang merupakan masukan berupa indeks data yang akan dijadikan <i>initial cluster</i> .
<pre> new_val = []</pre>	Pembuatan <i>list</i> kosong yang akan diisi dengan hasil perhitungan jarak.
<pre> for i in data_nom.columns : pembilang = 0 penyebut = 0</pre>	<i>Looping</i> perhitungan untuk pembaruan nilai pembilang dan penyebut sepanjang banyaknya variabel nominal. Kemudian membuat nilai awal pembilang dan penyebut bernilai 0 agar nanti dapat dilakukan penjumlahan.
<pre> for j in range(len(data_nom[i])) : if data[i][j] == data[i][<i>cluster</i>]: pembilang = 1</pre>	<i>Looping</i> pembaruan nilai pembilang sebanyak panjang dari data variabel nominal ke- <i>i</i> . Jika nilai data ke- <i>j</i> variabel nominal ke- <i>i</i>

Sintaks	Keterangan
<pre> else : pembilang = 0 </pre>	<p>sama dengan nilai data pada <i>cluster</i> variabel nominal ke-i maka nilai pembilang adalah 1, jika tidak maka 0.</p>
<pre> for k in range(len(data_nom[i])) : if (k == j) (k == cluster) : continue elif data_nom[i][k] == data_nom[i][j] : penyebut += 1 </pre>	<p><i>Looping</i> pembaruan nilai penyebut sebanyak panjang data variabel nominal ke-i. Nilai penyebut tidak dihitung apabila indeks data ke-k sama dengan data ke-j atau indeks data ke-k sama dengan indeks <i>cluster</i>. Setiap nilai data ke-k variabel nominal ke-i sama dengan variabel nominal ke-i data ke-j maka penyebut bertambah nilainya dengan 1.</p>
<pre> try : z = 1 - (pembilang/penyebut) new_val.append(z) pembilang = 0 penyebut = 0 </pre>	<p>Kemudian dilakukan perintah perhitungan <i>dissimilaritas</i> dari variabel nominal yaitu $1 - \text{similaritas nominal}$ atau dalam hal ini $1 - (\text{pembilang/penyebut})$ yang disimpan dalam <i>z</i>. Kemudian nilai dalam <i>z</i> tersebut akan ditambahkan ke <i>list new_val</i>. Setelah ditambahkan maka nilai pembilang dan penyebut dikembalikan lagi ke 0 agar</p>

Sintaks	Keterangan
	perhitungan jarak variabel nominal sebelumnya tidak berpengaruh ke nilai jarak variabel nominal setelahnya.
<pre> except ZeroDivisionError : z = 1 - (pembilang/1) new_val.append(z) pembilang = 0 penyebut = 0 </pre>	Apabila didapati penyebut akhir bernilai 0 maka penyebut tersebut diganti nilainya 1 agar menghindari <i>error</i> pada perhitungan.
<pre> baru = np.array(new_val) </pre>	Perintah untuk mengembalikan nilai <code>new_val</code> berupa <i>list</i> menjadi <i>array</i> yang kemudian disebut <i>array</i> baru
<pre> new_nomi = pd.DataFrame(baru.reshape(data_nom.shape[1], data_nom.shape[0]).T, columns = data_nom.columns) </pre>	Perintah untuk membentuk ulang <i>array</i> baru yang disimpan dalam <i>data frame</i> <code>new_nomi</code>
<pre> new_nomi.iloc[cluster,:] = 0 </pre>	Perintah untuk memberikan nilai 0 pada jarak dari data yang dijadikan <i>initial cluster</i>
<pre> new_nomi['Jumlah'] = new_nomi.sum(axis=1) return new_nomi </pre>	Perintah untuk menjumlahkan data per baris untuk didapatkan nilai jarak semua variabel nominal. Kemudian perintah <code>return</code> digunakan untuk mengembalikan nilai <code>new_nomi</code>

5.2.2 Jarak Ordinal

Pada penelitian ini akan dibahas sintaks perhitungan jarak ordinal untuk syarat kedua dikarenakan syarat pertama memiliki persamaan perhitungan jarak yang sama dengan perhitungan jarak data nominal.

Tabel 5.2 Penjelasan Sintaks Pencarian MSOA

Sintaks	Keterangan
<pre>Nama_Kolom = data.columns ukuran_domain = []</pre>	Pembuatan <i>list</i> kosong yang akan diisi dengan ukuran domain dari masing-masing variabel nominal.
<pre>for i in range(len(data.columns)) : ud = int(input(f'Ukuran Domain {Nama_Kolom[i]} : ')) ukuran_domain.append(ud)</pre>	Perintah untuk memasukkan ukuran domain dari masing-masing variabel.
<pre>UK_DO = pd.DataFrame({ 'Variable' : Nama_Kolom, 'Ukuran Domain' : ukuran_domain })</pre>	Perintah untuk membuat <i>data frame</i> yang berisikan nama variabel dan ukuran domain.
<pre>UK_DO['Selisih'] = abs(UK_DO['Ukuran Domain'] - UK_DO[UK_DO['Variable'] == ordi[0]]['Ukuran Domain'])[UK_DO[UK_DO['Variable'] == ordi[0]].index[0]]) print ('Ukuran Domain Variabel Ordinal A4'.upper()) display(UK_DO)</pre>	Perintah untuk menghitung selisih ukuran domain dan menampilkan hasilnya.
<pre>MSOA = np.min(UK_DO['Selisih']) print ('MSOA') display(MSOA) ind_MSOA = UK_DO[(UK_DO['Selisih'] == MSOA) & (UK_DO['Variable'] != ordi[0])]['Variable'] print('Indeks MSOA')</pre>	Perintah untuk menghitung nilai selisih dari ukuran domain variabel nominal terhadap ordinal. Kemudian ditampilkan nama variabel, ukuran domain, selisih

Sintaks	Keterangan
ind_MSOA	ukuran domain dan indeks MSOA.

Kemudian dilakukan perhitungan similaritas data pada variabel di *set* MSOA dengan menggunakan perintah yang sama pada **Tabel 5.1** namun nilai z tidak bernilai $z=1-(\text{pembilang}/\text{pernyebut})$ atau $z = 1-(\text{pembilang}/1)$ namun nilai z adalah $z = (\text{pembilang}/\text{pernyebut})$ atau $z = (\text{pembilang}/1)$

Tabel 5.3 Penjelasan Sintaks Perhitungan Dod_l

Sintaks	Keterangan
<pre>for i in range(len(data)) : exec(f'jarak_ordinal_A4{i} syarat2({i})')</pre> <pre>sh1 = pd.concat([jarak_ordinal_A40, ..., jarak_ordinal_A4i, axis=1) sh1</pre>	Perintah untuk memanggil hasil similaritas data pada variabel MSOA.
<pre>sh1[sh1>0].min().min()</pre>	Perintah untuk mencari nilai Dod_l yang merupakan nilai minimum similaritas tak kosong.

Sementara pada perhitungan jarak data ordinal syarat kedua digunakan perintah $A4_cluster_i = \text{abs}((\text{data}[\text{ordi}[0]] - \text{data}[\text{ordi}[0]][\text{cluster_i}]) * \text{sh1}[\text{sh1}>0].\text{min}().\text{min}())$ dimana $cluster_i$ disesuaikan dengan jumlah kluster yang akan digunakan.

5.3. *K-Modes Clustering*

Pada penelitian ini digunakan analisis *K-Modes Clustering* dengan menggunakan data pengguna jasa yang sudah mendaftar yaitu sebanyak 710 pengguna dan belum mendaftar asuransi perjalanan yaitu sebanyak 1277 pengguna. Hal ini dikarenakan dengan mengetahui karakteristik dari pengguna jasa yang tidak mendaftar asuransi perjalanan maka diharapkan terdapat kelompok tertentu yang mempunyai peluang untuk lebih ditingkatkan perhatiannya agar ikut mendaftar program asuransi. Kelompok yang berpeluang tersebut didapatkan dengan cara

membandingkan kelompok tersebut dengan kelompok yang terbentuk pada data pengguna jasa yang telah mendaftar asuransi perjalanan.

5.3.1 *Initial Cluster*

Sebelum melakukan analisis pengelompokkan, terlebih dahulu ditentukan pusat kelompok yang pertama dengan memilih secara acak sejumlah k (banyaknya *cluster*) dari data. Pada penelitian ini akan dibuat pengelompokkan dengan *cluster* sebanyak 2, 3, 4, 5, 6, dan 7. **Tabel 5.4** merupakan *initial cluster* berdasarkan keluaran dari *Python*.

Tabel 5.4 *Initial Cluster*

Jumlah Cluster	Status	<i>Initial Cluster</i>
2	Belum Mendaftar Asuransi	{X166, X735}
	Sudah Mendaftar Asuransi	{X464, X264}
3	Belum Mendaftar Asuransi	{X362, X1148, X78}
	Sudah Mendaftar Asuransi	{X535, X581, X396}
4	Belum Mendaftar Asuransi	{X457, X437, X1007, X172}
	Sudah Mendaftar Asuransi	{X354, X304, X326, X606}
5	Belum Mendaftar Asuransi	{X432, X114, X72, X1133, X153}
	Sudah Mendaftar Asuransi	{X329, X546, X378, X77, X671}
6	Belum Mendaftar Asuransi	{X1170, X1125, X89, X1002, X289, X1135}
	Sudah Mendaftar Asuransi	{X679, X234, X584, X263, X195, X564}
7	Belum Mendaftar Asuransi	{X929, X1117, X775, X1063, X607, X293, X356}
	Sudah Mendaftar Asuransi	{X544, X13, X61, X140, X188, X387, X593}

5.3.2 **Jarak Nominal**

Pengukuran jarak pada variabel nominal antar data ke- i terhadap *initial cluster* merupakan suatu cara untuk mengetahui kemiripan antar data ke- i terhadap *initial cluster* pada variabel nominal. Semakin kecil jarak maka akan semakin mirip kedua data tersebut. Perhitungan jarak untuk variabel nominal dijelaskan pada persamaan 3.6 yaitu 1 dikurangi nilai similaritas, dengan nilai similaritas data ke- i dan ke- j dalam satu variabel

1. $\frac{1}{\text{jumlah data yang nilainya sama dengan } i \text{ selain } j}$, apabila data ke- i bernilai sama dengan data ke- j .

2. 0, apabila data ke- i nilainya tidak sama dengan data ke- j .

Keluaran perhitungan jarak nominal data terhadap *initial cluster* untuk *cluster* berjumlah 2, 3, 4, 5, 6, dan 7 adalah sebagai berikut.

1. *Cluster* berjumlah dua :

- Belum mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9973 & 5,995 \\ 5,9951 & 5,9963 \\ 5,9929 & 5,9941 \\ \vdots & \vdots \\ 5,997 & 5,9982 \\ 5,9961 & 5,9973 \\ 5,9951 & 5,9963 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9889 & 5,9942 \\ 5,9922 & 5,9870 \\ 5,9934 & 5,9866 \\ \vdots & \vdots \\ 5,9803 & 5,9964 \\ 5,9984 & 5,9846 \\ 5,9984 & 5,9846 \end{pmatrix}$$

2. *Cluster* berjumlah tiga :

- Belum mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9975 & 5,9957 & 5,9973 \\ 5,9952 & 5,9935 & 5,9951 \\ 5,9979 & 5,9962 & 5,9961 \\ \vdots & \vdots & \vdots \\ 5,9838 & 5,9954 & 5,997 \\ 5,9963 & 5,9945 & 5,9961 \\ 5,9968 & 5,9951 & 5,9922 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9937 & 5,9942 & 5,9868 \\ 5,9900 & 5,9870 & 5,9924 \\ 5,9982 & 5,9866 & 5,9851 \\ \vdots & \vdots & \vdots \\ 5,9872 & 5,9964 & 5,9959 \\ 5,9963 & 5,9846 & 5,99 \\ 5,9963 & 5,9846 & 5,99 \end{pmatrix}$$

3. *Cluster* berjumlah empat :

- Belum mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9938 & 5,9943 & 5,9905 & 5,9944 \\ 5,9946 & 5,9956 & 5,9957 & 5,9961 \\ 5,9973 & 5,9983 & 5,9944 & 5,9932 \\ \vdots & \vdots & \vdots & \vdots \\ 5,9966 & 5,9975 & 5,9937 & 5,9941 \\ 5,9957 & 5,9966 & 5,9928 & 5,9932 \\ 5,9962 & 5,9972 & 5,9973 & 5,9932 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9921 & 5,9896 & 5,9865 & 5,9945 \\ 5,9912 & 5,9946 & 5,9984 & 5,9912 \\ 5,9908 & 5,9942 & 5,9910 & 5,9932 \\ \vdots & \vdots & \vdots & \vdots \\ 5,9942 & 5,9918 & 5,9865 & 5,9942 \\ 5,9888 & 5,9922 & 5,996 & 5,9912 \\ 5,9888 & 5,9922 & 5,996 & 5,9912 \end{pmatrix}$$

4. *Cluster* berjumlah lima :

- Belum mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9933 & 5,9966 & 5,9957 & 5,995 & 5,9966 \\ 5,9946 & 5,9944 & 5,9935 & 5,9963 & 5,9944 \\ 5,9973 & 5,9971 & 5,9961 & 5,9973 & 5,9971 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 5,9966 & 5,9963 & 5,9954 & 5,9982 & 5,9963 \\ 5,9957 & 5,9955 & 5,9945 & 5,9973 & 5,9954 \\ 5,9963 & 5,996 & 5,9951 & 5,9933 & 5,996 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9920 & 5,9945 & 5,9918 & 5,9920 & 5,9893 \\ 5,9946 & 5,985 & 5,9904 & 5,9946 & 5,989 \\ 5,9966 & 5,9932 & 5,99 & 5,9966 & 5,9817 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 5,9918 & 5,988 & 5,9939 & 5,9918 & 5,9984 \\ 5,9946 & 5,9912 & 5,988 & 5,9946 & 5,9866 \\ 5,9946 & 5,9912 & 5,988 & 5,9946 & 5,9866 \end{pmatrix}$$

5. *Cluster* berjumlah enam :

- Belum mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9966 & 5,9928 & 5,9966 & 5,9976 & 5,9871 & 5,9957 \\ 5,9944 & 5,9945 & 5,9944 & 5,9953 & 5,9903 & 5,9935 \\ 5,9971 & 5,9932 & 5,9971 & 5,998 & 5,9917 & 5,9961 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 5,9963 & 5,9925 & 5,9963 & 5,9972 & 5,9875 & 5,9954 \\ 5,9955 & 5,9916 & 5,9954 & 5,9964 & 5,9966 & 5,9945 \\ 5,996 & 5,9961 & 5,996 & 5,9969 & 5,9966 & 5,9951 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9935 & 5,9865 & 5,9914 & 5,9966 & 5,9905 & 5,9920 \\ 5,9922 & 5,9984 & 5,9964 & 5,9808 & 5,9957 & 5,9884 \\ 5,9917 & 5,9910 & 5,996 & 5,989 & 5,9944 & 5,9966 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 5,9867 & 5,9865 & 5,9846 & 5,9901 & 5,9937 & 5,9855 \\ 5,9898 & 5,996 & 5,994 & 5,9870 & 5,9928 & 5,9946 \\ 5,9898 & 5,996 & 5,994 & 5,9870 & 5,9973 & 5,9946 \end{pmatrix}$$

6. Cluster berjumlah tujuh :

- Belum mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,995 & 5,9937 & 5,9937 & 5,9992 & 5,9905 & 5,9973 & 5,9928 \\ 5,9963 & 5,9955 & 5,9955 & 5,9969 & 5,9957 & 5,9951 & 5,9945 \\ 5,9941 & 5,9942 & 5,9942 & 5,998 & 5,9944 & 5,9961 & 5,9932 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 5,9982 & 5,9934 & 5,9934 & 5,9988 & 5,9937 & 5,997 & 5,9925 \\ 5,9973 & 5,9926 & 5,9926 & 5,998 & 5,9928 & 5,9961 & 5,9916 \\ 5,9963 & 5,9971 & 5,9971 & 5,994 & 5,9973 & 5,9922 & 5,9961 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Ndis(x_i, x_j) = \begin{pmatrix} 5,9896 & 5,9914 & 5,9918 & 5,9865 & 5,9921 & 5,9913 & 5,9896 \\ 5,9945 & 5,9964 & 5,9904 & 5,9984 & 5,9912 & 5,9963 & 5,9946 \\ 5,9942 & 5,996 & 5,99 & 5,9910 & 5,9908 & 5,9958 & 5,9942 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 5,9918 & 5,9894 & 5,9939 & 5,9865 & 5,9942 & 5,9934 & 5,9918 \\ 5,9922 & 5,994 & 5,988 & 5,996 & 5,9888 & 5,9939 & 5,9922 \\ 5,9922 & 5,994 & 5,988 & 5,996 & 5,9888 & 5,9939 & 5,9922 \end{pmatrix}$$

5.3.3 Jarak Ordinal

Pada data yang digunakan, terdapat dua variabel yang memiliki skala ordinal, yaitu variabel *Age* dan *AnnualIncome*. Sebelum melakukan perhitungan jarak ordinal pada kedua variabel tersebut, terlebih dahulu ditentukan ukuran domain dari seluruh variabel yang digunakan.

Tabel 5.5 Ukuran Domain Tiap Variabel

No	Variabel	Domain	Ukuran Domain
1	<i>Age</i>	{<=30, 30}	2
2	<i>Employment Type</i>	{ <i>Government Sector, Private Sector/Self employed</i> }	2
3	<i>GraduateOrNot</i>	{ <i>Yes, No.</i> }	2
4	<i>AnnualIncome</i>	{ <i>Middle Income, Upper-Middle Income, High Income</i> }	3
5	<i>FamilyMembers</i>	{<4, =4, >4}	3
6	<i>ChronicDiseases</i>	{ <i>Yes, No.</i> }	2
7	<i>FrequentFlyer</i>	{ <i>Yes, No.</i> }	2

No	Variabel	Domain	Ukuran Domain
8	<i>EverTravelledAbroad</i>	{Yes, No.}	2

Age, sebagai variabel ordinal pertama, memiliki ukuran domain sebanyak 2 maka perhitungan jarak *Age* memenuhi syarat pertama dari perhitungan jarak ordinal yaitu dengan persamaan berikut.

$$1 - sim_{ord}^l(x_i, x_j) \quad (0.1)$$

Sehingga jarak masing-masing data pada variabel ordinal *Age* dengan *initial cluster* untuk jumlah *cluster* 2, 3, 4, 5, 6, 7 dan 8 adalah sebagai berikut.

1. *Cluster* berjumlah dua :

- Belum mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 1 & 0,9978 \\ 1 & 0,9978 \\ 0,9988 & 1 \\ \vdots & \vdots \\ 0,9988 & 1 \\ 0,9988 & 1 \\ 1 & 0,9978 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 1 & 0,997 \\ 1 & 0,997 \\ 0,9973 & 1 \\ \vdots & \vdots \\ 0,9973 & 1 \\ 1 & 0,997 \\ 1 & 0,997 \end{pmatrix}$$

2. *Cluster* berjumlah tiga :

- Belum mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0,9978 & 0,9978 & 0,9978 \\ 0,9978 & 0,9978 & 0,9978 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0,9978 & 0,9978 & 0,9978 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0,9973 & 0,9973 & 0,9973 \\ \vdots & \vdots & \vdots \\ 0,9973 & 0,9973 & 0,9973 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

3. *Cluster* berjumlah empat :

- Belum mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ \vdots & \vdots & \vdots & \vdots \\ 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0,9973 & 0,9973 & 0,9973 & 0,9973 \\ \vdots & \vdots & \vdots & \vdots \\ 0,9973 & 0,9973 & 0,9973 & 0,9973 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

4. *Cluster* berjumlah lima :

- Belum mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0,9978 & 1 & 1 & 1 & 1 \\ 0,9978 & 1 & 1 & 1 & 1 \\ 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 0,9978 & 1 & 1 & 1 & 1 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0,997 & 1 & 0,997 & 0,997 & 0,997 \\ 0,997 & 1 & 0,997 & 0,997 & 0,997 \\ 1 & 0,9973 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0,9973 & 1 & 1 & 1 \\ 0,997 & 1 & 0,997 & 0,997 & 0,997 \\ 0,997 & 1 & 0,997 & 0,997 & 0,997 \end{pmatrix}$$

5. *Cluster* berjumlah enam :

- Belum mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0,9978 & 1 & 1 & 1 & 1 & 1 \\ 0,9978 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0,9978 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 1 & 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 0,9978 & 0,9978 & 1 & 1 & 1 & 1 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0,997 & 0,997 & 0,997 & 1 & 0,997 & 1 \\ 0,997 & 0,997 & 0,997 & 1 & 0,997 & 1 \\ 1 & 1 & 1 & 0,997 & 1 & 0,997 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 0,997 & 1 & 0,997 \\ 0,997 & 1 & 0,997 & 1 & 0,997 & 1 \\ 0,997 & 0,997 & 0,997 & 1 & 0,997 & 1 \end{pmatrix}$$

6. *Cluster* berjumlah tujuh :

- Belum mendaftarkan asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0,9978 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0,9978 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 1 & 0,9988 & 0,9988 & 0,9988 & 0,9988 & 0,9988 & 0,9988 \\ 0,9978 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

- Sudah mendaftarkan asuransi

$$Odis_1(x_i, x_j) = \begin{pmatrix} 1 & 1 & 0,997 & 0,997 & 0,997 & 0,997 & 1 \\ 1 & 1 & 0,997 & 0,997 & 0,997 & 0,997 & 1 \\ 0,9973 & 0,9973 & 1 & 1 & 1 & 1 & 0,9973 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,9973 & 0,9973 & 1 & 1 & 1 & 1 & 0,9973 \\ 1 & 1 & 0,997 & 0,997 & 0,997 & 0,997 & 1 \\ 1 & 1 & 0,997 & 0,997 & 0,997 & 0,997 & 1 \end{pmatrix}$$

AnnualIncome, sebagai variabel ordinal kedua, memiliki ukuran domain sebanyak 3 dengan ukuran domain maksimal untuk variabel nominal adalah 3 yaitu pada *FamilyMembers*. Variabel ordinal kedua ini memenuhi syarat kedua dari perhitungan jarak ordinal dikarenakan ukuran domain ordinal lebih dari 2 dan kurang dari 2 kali ukuran domain maksimal variabel nominal (sebanyak 2 x 3 atau 6). Persamaan yang digunakan pada syarat kedua yaitu sebagai berikut.

$$|ord_i^l - ord_j^l|Dod_1 \quad (0.2)$$

Dari persamaan tersebut, diperlukan nilai Dod_1 yang merupakan nilai minimum dari similaritas antar data pada variabel *MSOA*. *MSOA* merupakan variabel nominal yang memiliki nilai minimum pada selisih ukuran domain variabel nominal terhadap variabel ordinal yang memenuhi syarat kedua. Sehingga selisih ukuran domain antar variabel nominal terhadap variabel ordinal *AnnualIncome* dijelaskan pada **Tabel 5.6**.

Tabel 5.6 Ukuran Domain Tiap Variabel

No	Variabel	Ukuran Domain	Selisih
1	<i>Employment Type</i>	2	1
2	<i>GraduateOrNot</i>	2	1
3	<i>FamilyMembers</i>	3	0
4	<i>ChronicDiseases</i>	2	1
5	<i>FrequentFlyer</i>	2	1
6	<i>EverTravelledAbroad</i>	2	1

Nilai minimum dari selisih ukuran domain adalah 0 yang terdapat pada variabel nominal *FamilyMembers*, sehingga variabel tersebut merupakan *MSOA*. Selanjutnya dilakukan perhitungan similaritas antar data pada variabel *FamilyMembers*. Matriks similaritas antar data pada variabel *FamilyMembers* pada data pengguna yang tidak mendaftar asuransi perjalanan menghasilkan matriks 1277×1277 dengan pembulatan sebagai berikut.

$$Sim_{FamilyMembers}(x_i, x_j) = \begin{pmatrix} 1 & 0,0016 & 0 & \dots & 0,0016 & 0,0016 & 0 \\ 0,0016 & 1 & 0 & \dots & 0,0016 & 0,0016 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0,0016 & 0,0016 & 0 & \dots & 1 & 0,0016 & 0 \\ 0,0016 & 0,0016 & 0 & \dots & 0,0016 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

Dod_I merupakan nilai minimum tak nol dari similaritas variabel pada *MSOA* sehingga nilai Dod_I adalah sebesar 0,0016 atau dengan tanpa pembulatan yaitu sebesar 0,0016077170418000643.

Sedangkan untuk pada data pengguna yang menggunakan asuransi perjalanan akan terbentuk matriks similaritas antar data pada variabel *FamilyMembers* sebesar 710×710 dengan pembulatan sebagai berikut.

$$Sim_{FamilyMembers}(x_i, x_j) = \begin{pmatrix} 1 & 0 & 0,0024 & \dots & 0 & 0,0024 & 0,0024 \\ 0 & 1 & 0 & \dots & 0,0063 & 0 & 0 \\ 0,0024 & 0 & 1 & \dots & 0 & 0,0024 & 0,0024 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0,0063 & 0 & \dots & 1 & 0 & 0 \\ 0,0024 & 0 & 0,0024 & \dots & 0 & 1 & 0,0024 \\ 0,0024 & 0 & 0,0024 & \dots & 0 & 0,0024 & 1 \end{pmatrix}$$

sehingga nilai Dod_I adalah sebesar 0,0024 atau dengan tanpa pembulatan yaitu sebesar 0,00239234449760765. Nilai ini bersifat konstan. Nilai Dod_I digunakan untuk menghitung nilai jarak pada variabel ordinal kedua yaitu variabel *AnnualIncome* dengan cara mengalikan Dod_I dengan selisih data ke- i dengan *initial*

cluster pada variabel ordinal kedua. Sehingga jarak masing-masing data pada variabel ordinal *Age* dengan *initial cluster* untuk jumlah *cluster* 2, 3, 4, 5, 6, 7 dan 8 adalah sebagai berikut.

1. *Cluster* berjumlah dua :

- Belum mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0,0016 & 0 \\ 0 & 0,0016 \\ 0,0016 & 0 \\ \vdots & \vdots \\ 0 & 0,0016 \\ 0 & 0,0016 \\ 0,0016 & 0 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0,0024 \\ 0,0024 & 0 \\ 0,0024 & 0 \\ \vdots & \vdots \\ 0 & 0,0024 \\ 0,0024 & 0 \\ 0,0024 & 0 \end{pmatrix}$$

2. *Cluster* berjumlah tiga :

- Belum mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0,0016 & 0 & 0 \\ 0 & 0,0016 & 0,0016 \\ 0,0016 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0,0016 & 0,0016 \\ 0 & 0,0016 & 0,0016 \\ 0,0016 & 0 & 0 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0,0024 & 0,0024 \\ 0,0024 & 0 & 0 \\ 0,0024 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0,0024 & 0,0024 \\ 0,0024 & 0 & 0 \\ 0,0024 & 0 & 0 \end{pmatrix}$$

3. *Cluster* berjumlah empat :

- Belum mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0,0016 & 0 & 0 \\ 0,0016 & 0,0032 & 0,0016 & 0,0016 \\ 0 & 0,0016 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0,0016 & 0,0032 & 0,0016 & 0,0016 \\ 0,0016 & 0,0032 & 0,0016 & 0,0016 \\ 0 & 0,0016 & 0 & 0 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0,0024 & 0,0024 & 0,0024 & 0,0024 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0,0024 & 0,0024 & 0,0024 & 0,0024 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

4. *Cluster* berjumlah lima :

- Belum mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0,0016 & 0 & 0 & 0 & 0 \\ 0 & 0,0016 & 0,0016 & 0,0016 & 0,0016 \\ 0,0016 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0,0016 & 0,0016 & 0,0016 & 0,0016 \\ 0 & 0 & 0,0016 & 0,0016 & 0,0016 \\ 0,0016 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0,0024 & 0 & 0,0024 & 0,0024 \\ 0,0024 & 0 & 0,0024 & 0 & 0 \\ 0,0024 & 0 & 0,0024 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0,0024 & 0 & 0,0024 & 0,0024 \\ 0,0024 & 0 & 0,0024 & 0 & 0 \\ 0,0024 & 0 & 0,0024 & 0 & 0 \end{pmatrix}$$

5. *Cluster* berjumlah enam :

- Belum mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0 & 0 & 0,0016 & 0,0016 & 0,0016 \\ 0,0016 & 0,0016 & 0,0016 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0016 & 0,0016 & 0,0016 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,0016 & 0,0016 & 0,0016 & 0 & 0 & 0 \\ 0,0016 & 0,0016 & 0,0016 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0016 & 0,0016 & 0,0016 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0,0024 & 0,0024 & 0,0024 & 0,0024 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,0024 & 0,0024 \\ 0 & 0 & 0 & 0 & 0,0024 & 0,0024 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,0024 & 0,0024 & 0,0024 & 0,0024 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,0024 & 0,0024 \\ 0 & 0 & 0 & 0 & 0,0024 & 0,0024 \end{pmatrix}$$

6. *Cluster* berjumlah tujuh :

- Belum mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0 & 0,0016 & 0,0016 & 0,0016 & 0 & 0 \\ 0,0016 & 0,0016 & 0 & 0 & 0 & 0,0016 & 0,0016 \\ 0 & 0 & 0,0016 & 0,0016 & 0,0016 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,0016 & 0,0016 & 0 & 0 & 0 & 0,0016 & 0,0016 \\ 0,0016 & 0,0016 & 0 & 0 & 0 & 0,0016 & 0,0016 \\ 0 & 0 & 0,0016 & 0,0016 & 0,0016 & 0 & 0 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0,0024 & 0 & 0,0024 \\ 0,0024 & 0,0024 & 0,0024 & 0,0024 & 0 & 0,0024 & 0 \\ 0,0024 & 0,0024 & 0,0024 & 0,0024 & 0 & 0,0024 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0,0024 & 0 & 0,0024 \\ 0,0024 & 0,0024 & 0,0024 & 0,0024 & 0 & 0,0024 & 0 \\ 0,0024 & 0,0024 & 0,0024 & 0,0024 & 0 & 0,0024 & 0 \end{pmatrix}$$

5.3.4 Jarak Gabungan

Jarak gabungan merupakan penjumlahan jarak pada variabel nominal dan variabel ordinal, sehingga keluaran dari matriks jarak gabungan adalah sebagai berikut.

1. *Cluster* berjumlah dua :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9989 & 6,9928 \\ 6,9951 & 6,9957 \\ 6,9932 & 6,9941 \\ \vdots & \vdots \\ 6,9957 & 6,9998 \\ 6,995 & 6,9989 \\ 6,9967 & 6,9941 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9889 & 6,9936 \\ 6,9945 & 6,984 \\ 6,9932 & 6,9866 \\ \vdots & \vdots \\ 6,9776 & 6,9988 \\ 7,0008 & 6,9816 \\ 7,0008 & 6,9816 \end{pmatrix}$$

2. *Cluster* berjumlah tiga :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,997 & 6,9936 & 6,9952 \\ 6,993 & 6,9929 & 6,9945 \\ 6,9995 & 6,9961 & 6,9961 \\ \vdots & \vdots & \vdots \\ 6,9838 & 6,997 & 6,9986 \\ 6,9963 & 6,9961 & 6,9977 \\ 6,9963 & 6,9929 & 6,989 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9937 & 6,9966 & 6,9989 \\ 6,9924 & 6,9870 & 6,9924 \\ 6,998 & 6,9839 & 6,9824 \\ \vdots & \vdots & \vdots \\ 6,9845 & 6,9961 & 6,9956 \\ 6,9986 & 6,9985 & 6,99 \\ 6,9986 & 6,9985 & 6,99 \end{pmatrix}$$

3. *Cluster* berjumlah empat :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9934 & 6,9959 & 6,9905 & 6,9944 \\ 6,9962 & 6,9988 & 6,9973 & 6,9977 \\ 6,9961 & 6,9986 & 6,9932 & 6,992 \\ \vdots & \vdots & \vdots & \vdots \\ 6,9969 & 6,9995 & 6,9940 & 6,9945 \\ 6,9961 & 6,9986 & 6,9932 & 6,9936 \\ 6,9962 & 6,9988 & 6,9973 & 6,9932 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & 6,9920 & 6,9889 & 6,9969 \\ 6,9912 & 6,9946 & 6,9984 & 6,9912 \\ 6,9882 & 6,9915 & 6,9884 & 6,9906 \\ \vdots & \vdots & \vdots & \vdots \\ 6,994 & 6,9915 & 6,9863 & 6,994 \\ 6,9888 & 6,9922 & 6,996 & 6,9912 \\ 6,9888 & 6,9922 & 6,996 & 6,9912 \end{pmatrix}$$

4. *Cluster* berjumlah lima :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9928 & 6,9967 & 6,9957 & 6,995 & 6,9966 \\ 6,9925 & 6,996 & 6,9951 & 6,9979 & 6,996 \\ 6,9989 & 6,9958 & 6,9949 & 6,9961 & 6,9958 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9966 & 6,9967 & 6,9957 & 6,9985 & 6,9967 \\ 6,9957 & 6,9958 & 6,9949 & 6,9977 & 6,9958 \\ 6,9957 & 6,996 & 6,9951 & 6,9933 & 6,996 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9890 & 6,9969 & 6,9887 & 6,9914 & 6,9887 \\ 6,994 & 6,985 & 6,99 & 6,9916 & 6,986 \\ 6,999 & 6,9906 & 6,9924 & 6,9966 & 6,9817 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9918 & 6,9877 & 6,9939 & 6,9942 & 7,0008 \\ 6,994 & 6,9912 & 6,9874 & 6,9916 & 6,9836 \\ 6,994 & 6,9912 & 6,9874 & 6,9916 & 6,9836 \end{pmatrix}$$

5. *Cluster* berjumlah enam :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & 6,9907 & 6,9966 & 6,9992 & 6,9921 & 6,9973 \\ 6,9938 & 6,994 & 6,996 & 6,9953 & 6,9957 & 6,9935 \\ 6,9971 & 6,9932 & 6,9958 & 6,9984 & 6,9948 & 6,9965 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9979 & 6,9941 & 6,9967 & 6,996 & 6,9924 & 6,9941 \\ 6,9971 & 6,9932 & 6,9958 & 6,9951 & 6,9916 & 6,9933 \\ 6,9938 & 6,994 & 6,996 & 6,9985 & 6,9989 & 6,9967 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9929 & 6,9858 & 6,9908 & 6,999 & 6,9841 & 6,9920 \\ 6,9891 & 6,9953 & 6,9934 & 6,9808 & 6,9897 & 6,9908 \\ 6,9918 & 6,9910 & 6,996 & 6,9863 & 6,9941 & 6,9963 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9891 & 6,9889 & 6,9869 & 6,9898 & 6,9875 & 6,9829 \\ 6,9867 & 6,9929 & 6,991 & 6,9870 & 6,996 & 6,9970 \\ 6,9867 & 6,9929 & 6,991 & 6,9870 & 6,996 & 6,9970 \end{pmatrix}$$

6. *Cluster* berjumlah tujuh :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9928 & 6,9937 & 6,9954 & 7,0001 & 6,9921 & 6,9973 & 6,9928 \\ 6,9957 & 6,9971 & 6,9955 & 6,9969 & 6,9957 & 6,9967 & 6,9961 \\ 6,9941 & 6,9929 & 6,9945 & 6,9984 & 6,9948 & 6,9949 & 6,9920 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9998 & 6,9938 & 6,9922 & 6,9976 & 6,9924 & 6,9974 & 6,9928 \\ 6,9989 & 6,9929 & 6,9913 & 6,9967 & 6,9916 & 6,9965 & 6,9920 \\ 6,9941 & 6,9971 & 6,9987 & 6,9956 & 6,9989 & 6,9922 & 6,9961 \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9896 & 6,9914 & 6,9887 & 6,9834 & 6,9914 & 6,9883 & 6,9920 \\ 6,9970 & 6,9988 & 6,9898 & 6,9977 & 6,9882 & 6,9956 & 6,9946 \\ 6,9939 & 6,9957 & 6,9924 & 6,9934 & 6,9908 & 6,9982 & 6,9915 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9891 & 6,9819 & 6,9939 & 6,9865 & 6,9966 & 6,9934 & 6,9915 \\ 6,9946 & 6,9964 & 6,9873 & 6,9953 & 6,9858 & 6,9932 & 6,9922 \\ 6,9946 & 6,9964 & 6,9873 & 6,9953 & 6,9858 & 6,9932 & 6,9922 \end{pmatrix}$$

5.3.5 Penempatan Data Pada Setiap *Cluster*

Jarak data terhadap masing-masing *initial cluster* telah dihitung sehingga langkah selanjutnya adalah penempatan data pada setiap *cluster*. Masing-masing

data ditempatkan satu *cluster* dengan *initial cluster* yang memiliki jarak paling minimum di antara *initial cluster* lain.

Berikut merupakan penempatan data pada setiap *cluster*. Nilai minimum jarak data terhadap *initial cluster* ditunjukkan dengan garis warna kuning.

1. *Cluster* berjumlah dua :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9989 & \mathbf{6,9928} \\ \mathbf{6,9951} & 6,9957 \\ \mathbf{6,9932} & 6,9941 \\ \vdots & \vdots \\ \mathbf{6,9957} & 6,9998 \\ \mathbf{6,995} & 6,9989 \\ 6,9967 & \mathbf{6,9941} \end{pmatrix} = \begin{pmatrix} \text{Cluster 2} \\ \text{Cluster 1} \\ \text{Cluster 1} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 1} \\ \text{Cluster 2} \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} \mathbf{6,9889} & 6,9936 \\ 6,9945 & \mathbf{6,984} \\ 6,9932 & \mathbf{6,9866} \\ \vdots & \vdots \\ \mathbf{6,9776} & 6,9988 \\ 7,0008 & \mathbf{6,9816} \\ 7,0008 & \mathbf{6,9816} \end{pmatrix} = \begin{pmatrix} \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \end{pmatrix}$$

2. *Cluster* berjumlah tiga :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,997 & \mathbf{6,9936} & 6,9952 \\ 6,993 & \mathbf{6,9929} & 6,9945 \\ 6,9995 & \mathbf{6,9961} & 6,9961 \\ \vdots & \vdots & \vdots \\ \mathbf{6,9838} & 6,997 & 6,9986 \\ 6,9963 & \mathbf{6,9961} & 6,9977 \\ 6,9963 & \mathbf{6,9929} & 6,989 \end{pmatrix} = \begin{pmatrix} \text{Cluster 2} \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} \mathbf{6,9937} & 6,9966 & 6,9989 \\ 6,9924 & \mathbf{6,9870} & 6,9924 \\ 6,998 & 6,9839 & \mathbf{6,9824} \\ \vdots & \vdots & \vdots \\ \mathbf{6,9845} & 6,9961 & 6,9956 \\ 6,9986 & \mathbf{6,9985} & 6,99 \\ 6,9986 & \mathbf{6,9985} & 6,99 \end{pmatrix} = \begin{pmatrix} \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \end{pmatrix}$$

3. *Cluster* berjumlah empat :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9934 & 6,9959 & \mathbf{6,9905} & 6,9944 \\ \mathbf{6,9962} & 6,9988 & 6,9973 & 6,9977 \\ 6,9961 & 6,9986 & 6,9932 & \mathbf{6,992} \\ \vdots & \vdots & \vdots & \vdots \\ 6,9969 & 6,9995 & \mathbf{6,9940} & 6,9945 \\ 6,9961 & 6,9986 & \mathbf{6,9932} & 6,9936 \\ 6,9962 & 6,9988 & 6,9973 & \mathbf{6,9932} \end{pmatrix} = \begin{pmatrix} \text{Cluster 3} \\ \text{Cluster 1} \\ \text{Cluster 4} \\ \vdots \\ \text{Cluster 3} \\ \text{Cluster 3} \\ \text{Cluster 4} \end{pmatrix}$$

- Sudah mendaftarkan asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & 6,9920 & \mathbf{6,9889} & 6,9969 \\ \mathbf{6,9912} & 6,9946 & 6,9984 & 6,9912 \\ \mathbf{6,9882} & 6,9915 & 6,9884 & 6,9906 \\ \vdots & \vdots & \vdots & \vdots \\ 6,994 & 6,9915 & \mathbf{6,9863} & 6,994 \\ \mathbf{6,9888} & 6,9922 & 6,996 & 6,9912 \\ \mathbf{6,9888} & 6,9922 & 6,996 & 6,9912 \end{pmatrix} = \begin{pmatrix} \text{Cluster 3} \\ \text{Cluster 1} \\ \text{Cluster 1} \\ \vdots \\ \text{Cluster 3} \\ \text{Cluster 1} \\ \text{Cluster 1} \end{pmatrix}$$

4. Cluster berjumlah lima :

- Belum mendaftarkan asuransi

$$d(x_i, x_j) = \begin{pmatrix} \mathbf{6,9928} & 6,9967 & 6,9957 & 6,995 & 6,9966 \\ \mathbf{6,9925} & 6,996 & 6,9951 & 6,9979 & 6,996 \\ 6,9989 & 6,9958 & \mathbf{6,9949} & 6,9961 & 6,9958 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{6,9966} & 6,9967 & 6,9957 & 6,9985 & 6,9967 \\ 6,9957 & 6,9958 & \mathbf{6,9949} & 6,9977 & 6,9958 \\ 6,9957 & 6,996 & 6,9951 & \mathbf{6,9933} & 6,996 \end{pmatrix} = \begin{pmatrix} \text{Cluster 1} \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 3} \\ \text{Cluster 4} \end{pmatrix}$$

- Sudah mendaftarkan asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9890 & 6,9969 & 6,9887 & 6,9914 & \mathbf{6,9887} \\ 6,994 & \mathbf{6,985} & 6,99 & 6,9916 & 6,986 \\ 6,999 & 6,9906 & 6,9924 & 6,9966 & \mathbf{6,9817} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9918 & \mathbf{6,9877} & 6,9939 & 6,9942 & 7,0008 \\ 6,994 & 6,9912 & 6,9874 & 6,9916 & \mathbf{6,9836} \\ 6,994 & 6,9912 & 6,9874 & 6,9916 & \mathbf{6,9836} \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 2} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 5} \\ \text{Cluster 5} \end{pmatrix}$$

5. Cluster berjumlah enam :

- Belum mendaftarkan asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & \mathbf{6,9907} & 6,9966 & 6,9992 & 6,9921 & 6,9973 \\ 6,9938 & 6,994 & 6,996 & 6,9953 & 6,9957 & \mathbf{6,9935} \\ 6,9971 & \mathbf{6,9932} & 6,9958 & 6,9984 & 6,9948 & 6,9965 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9979 & 6,9941 & 6,9967 & 6,996 & \mathbf{6,9924} & 6,9941 \\ 6,9971 & 6,9932 & 6,9958 & 6,9951 & \mathbf{6,9916} & 6,9933 \\ \mathbf{6,9938} & 6,994 & 6,996 & 6,9985 & 6,9989 & 6,9967 \end{pmatrix} = \begin{pmatrix} \text{Cluster 2} \\ \text{Cluster 6} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 5} \\ \text{Cluster 5} \\ \text{Cluster 1} \end{pmatrix}$$

- Sudah mendaftarkan asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9929 & 6,9858 & 6,9908 & 6,999 & 6,9841 & 6,9920 \\ 6,9891 & 6,9953 & 6,9934 & 6,9808 & 6,9897 & 6,9908 \\ 6,9918 & 6,9910 & 6,996 & 6,9863 & 6,9941 & 6,9963 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9891 & 6,9889 & 6,9869 & 6,9898 & 6,9875 & 6,9829 \\ 6,9867 & 6,9929 & 6,991 & 6,9870 & 6,996 & 6,9970 \\ 6,9867 & 6,9929 & 6,991 & 6,9870 & 6,996 & 6,9970 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 4} \\ \text{Cluster 4} \\ \vdots \\ \text{Cluster 6} \\ \text{Cluster 1} \\ \text{Cluster 1} \end{pmatrix}$$

6. Cluster berjumlah tujuh :

- Belum mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9928 & 6,9937 & 6,9954 & 7,0001 & 6,9921 & 6,9973 & 6,9928 \\ 6,9957 & 6,9971 & 6,9955 & 6,9969 & 6,9957 & 6,9967 & 6,9961 \\ 6,9941 & 6,9929 & 6,9945 & 6,9984 & 6,9948 & 6,9949 & 6,9920 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9998 & 6,9938 & 6,9922 & 6,9976 & 6,9924 & 6,9974 & 6,9928 \\ 6,9989 & 6,9929 & 6,9913 & 6,9967 & 6,9916 & 6,9965 & 6,9920 \\ 6,9941 & 6,9971 & 6,9987 & 6,9956 & 6,9989 & 6,9922 & 6,9961 \end{pmatrix} = \begin{pmatrix} \text{Cluster 1} \\ \text{Cluster 3} \\ \text{Cluster 7} \\ \vdots \\ \text{Cluster 3} \\ \text{Cluster 3} \\ \text{Cluster 6} \end{pmatrix}$$

- Sudah mendaftar asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9896 & 6,9914 & 6,9887 & 6,9834 & 6,9914 & 6,9883 & 6,9920 \\ 6,9970 & 6,9988 & 6,9898 & 6,9977 & 6,9882 & 6,9956 & 6,9946 \\ 6,9939 & 6,9957 & 6,9924 & 6,9934 & 6,9908 & 6,9982 & 6,9915 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9891 & 6,9819 & 6,9939 & 6,9865 & 6,9966 & 6,9934 & 6,9915 \\ 6,9946 & 6,9964 & 6,9873 & 6,9953 & 6,9858 & 6,9932 & 6,9922 \\ 6,9946 & 6,9964 & 6,9873 & 6,9953 & 6,9858 & 6,9932 & 6,9922 \end{pmatrix} = \begin{pmatrix} \text{Cluster 4} \\ \text{Cluster 5} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 5} \\ \text{Cluster 5} \end{pmatrix}$$

5.3.6 Pemilihan Initial Cluster Baru

Initial cluster baru ditentukan dari modus per cluster. Berikut merupakan modus data per cluster.

1. Cluster berjumlah dua :

Tabel 5.7 Modus Cluster Berjumlah Dua Pertama

Status Asuransi	Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
Belum Punya	1	0	1	1	>4	4	0	0	0
	2	1	0	1	>4	3	0	0	0
Sudah Punya	1	0	1	1	<4	3	0	0	0
	2	1	1	1	4	4	0	0	1

2. Cluster berjumlah tiga :

Tabel 5.8 Modus Cluster Berjumlah Tiga Pertama

Status Asuransi	Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
Belum Punya	1	0	1	1	>4	3	0	1	0
	2	1	0	1	>4	3	0	0	0
	3	0	1	1	4	3	0	0	0
Sudah Punya	1	0	1	1	4	3	0	0	0
	2	0	1	1	4	4	0	1	1

<i>Status Asuransi</i>	<i>Cluster</i>	<i>Age</i>	<i>Employment Type</i>	<i>Graduate Or Not</i>	<i>Family Members</i>	<i>Annual Income</i>	<i>Chronic Diseases</i>	<i>Frequent Flyer</i>	<i>Ever Travelled Abroad</i>
	3	1	1	1	4	4	1	0	0

3. *Cluster* berjumlah empat :

Tabel 5.9 Modus *Cluster* Berjumlah Empat Pertama

<i>Status Asuransi</i>	<i>Cluster</i>	<i>Age</i>	<i>Employment Type</i>	<i>Graduate Or Not</i>	<i>Family Members</i>	<i>Annual Income</i>	<i>Chronic Diseases</i>	<i>Frequent Flyer</i>	<i>Ever Travelled Abroad</i>
Belum Punya	1	0	1	1	>4	3	0	0	0
	2	0	1	0	>4	2	0	0	0
	3	0	1	1	>4	3	1	0	0
	4	0	1	1	4	3	0	0	0
Sudah Punya	1	0	1	1	4	4	0	0	1
	2	0	1	1	4	4	0	0	0
	3	1	0	1	4	3	1	0	0
	4	1	1	1	>4	4	0	0	0

4. *Cluster* berjumlah lima :

Tabel 5.10 Modus *Cluster* Berjumlah Lima Pertama

<i>Status Asuransi</i>	<i>Cluster</i>	<i>Age</i>	<i>Employment Type</i>	<i>Graduate Or Not</i>	<i>Family Members</i>	<i>Annual Income</i>	<i>Chronic Diseases</i>	<i>Frequent Flyer</i>	<i>Ever Travelled Abroad</i>
Belum Punya	1	1	0	1	>4	4	0	0	0
	2	0	1	0	>4	3	0	0	0
	3	0	1	1	>4	3	0	0	0
	4	0	0	1	4	3	0	0	0
	5	0	1	1	>4	3	0	1	0
Sudah Punya	1	1	1	1	>4	3	0	0	0
	2	0	1	1	<4	4	0	0	1
	3	0	1	1	4	3	0	0	0
	4	1	1	1	>4	4	0	0	0
	5	1	1	1	4	4	1	1	1

5. *Cluster* berjumlah enam :

Tabel 5.11 Modus *Cluster* Berjumlah Enam Pertama

<i>Status Asuransi</i>	<i>Cluster</i>	<i>Age</i>	<i>Employment Type</i>	<i>Graduate Or Not</i>	<i>Family Members</i>	<i>Annual Income</i>	<i>Chronic Diseases</i>	<i>Frequent Flyer</i>	<i>Ever Travelled Abroad</i>
Belum Punya	1	1	1	0	>4	3	0	0	0
	2	1	1	1	>4	3	1	0	0
	3	0	1	1	>4	3	0	0	0
	4	0	1	1	>4	4	0	1	0
	5	0	0	1	>4	4	0	0	0
	6	0	1	1	>4	4	0	0	0
Sudah	1	1	1	1	4	4	0	1	0

<i>Status Asuransi</i>	<i>Cluster</i>	<i>Age</i>	<i>Employment Type</i>	<i>Graduate Or Not</i>	<i>Family Members</i>	<i>Annual Income</i>	<i>Chronic Diseases</i>	<i>Frequent Flyer</i>	<i>Ever Travelled Abroad</i>
Punya	2	1	1	1	4	4	1	0	0
	3	1	1	1	4	4	0	0	0
	4	0	1	1	4	4	0	1	1
	5	1	1	1	<4	1	1	0	0
	6	0	1	1	0	1	0	0	0

7. *Cluster* berjumlah tujuh :

Tabel 5.12 Modus *Cluster* Berjumlah Tujuh Pertama

<i>Status Asuransi</i>	<i>Cluster</i>	<i>Age</i>	<i>Employment Type</i>	<i>Graduate Or Not</i>	<i>Family Members</i>	<i>Annual Income</i>	<i>Chronic Diseases</i>	<i>Frequent Flyer</i>	<i>Ever Travelled Abroad</i>
Belum Punya	1	1	0	1	<4	3	0	0	0
	2	0	1	0	>4	3	0	0	0
	3	0	1	1	>4	4	0	0	0
	4	0	1	1	4	3	0	1	0
	5	0	0	1	>4	3	1	0	0
	6	0	1	1	4	3	0	0	0
	7	0	1	1	>4	3	0	0	0
Sudah Punya	1	0	1	1	4	3	0	0	0
	2	0	0	1	4	3	0	0	0
	3	1	1	1	4	3	0	1	0
	4	1	0	1	4	3	1	0	0
	5	1	1	1	4	4	0	0	1
	6	0	1	0	4	3	0	0	0
	7	0	1	1	4	4	0	0	0

5.3.7 Iterasi

Modus masing-masing *cluster* yang telah didapatkan kemudian dijadikan *initial cluster* baru dan dihitung jarak data terhadap *initial cluster* baru tersebut. Jarak data terhadap masing-masing *initial cluster* baru dengan jarak minimum diberikan garis kuning dan ditunjukkan sebagai berikut.

1. *Cluster* berjumlah dua :

- Pengguna yang tidak memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9973 & \mathbf{6,9912} \\ \mathbf{6,9935} & 6,9941 \\ \mathbf{6,9965} & 6,9973 \\ \vdots & \vdots \\ \mathbf{6,9941} & 6,9982 \\ \mathbf{6,9933} & 6,9973 \\ 6,9968 & \mathbf{6,9941} \end{pmatrix} = \begin{pmatrix} \text{Cluster 2} \\ \text{Cluster 1} \\ \text{Cluster 1} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 1} \\ \text{Cluster 2} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.13 Modus *Cluster* Berjumlah Dua Kedua Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	>4	4	0	0	0
2	1	0	1	>4	3	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

- Pengguna yang memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9920 & 6,9915 \\ 6,9908 & 6,9882 \\ 6,9963 & 6,9909 \\ \vdots & \vdots \\ 6,9829 & 6,9966 \\ 6,9970 & 6,9859 \\ 6,9970 & 6,9859 \end{pmatrix} = \begin{pmatrix} \text{Cluster 2} \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.14 Modus *Cluster* Berjumlah Dua Kedua Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	4	3	0	0	0
2	1	1	1	4	4	0	0	1

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9897 & 6,9915 \\ 6,9970 & 6,9882 \\ 6,9939 & 6,9909 \\ \vdots & \vdots \\ 6,9892 & 6,9966 \\ 6,9946 & 6,9859 \\ 6,9946 & 6,9859 \end{pmatrix} = \begin{pmatrix} \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.15 Modus *Cluster* Berjumlah Dua Ketiga Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	4	3	0	0	0
2	1	1	1	4	4	0	0	1

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

2. *Cluster* berjumlah tiga :

- Pengguna yang tidak memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9966 & 6,9957 & 6,9973 \\ 6,996 & 6,9951 & 6,9967 \\ 6,9958 & 6,9949 & 6,9949 \\ \vdots & \vdots & \vdots \\ 6,9967 & 6,9958 & 6,9974 \\ 6,9958 & 6,9949 & 6,9965 \\ 6,996 & 6,9951 & 6,9922 \end{pmatrix} = \begin{pmatrix} \text{Cluster 2} \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \text{Cluster 3} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.16 Modus *Cluster* Berjumlah Tiga Kedua Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	> 4	3	0	1	0
2	1	0	1	> 4	3	0	0	0
3	0	1	1	4	3	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

- Pengguna yang memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9897 & 6,9966 & 6,9841 \\ 6,9970 & 6,9871 & 6,9936 \\ 6,9939 & 6,984 & 6,9893 \\ \vdots & \vdots & \vdots \\ 6,9892 & 6,9961 & 6,9962 \\ 6,9946 & 6,9847 & 6,9912 \\ 6,9946 & 6,9847 & 6,9912 \end{pmatrix} = \begin{pmatrix} \text{Cluster 3} \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \text{Cluster 2} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.17 Modus *Cluster* Berjumlah Tiga Kedua Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	4	3	0	0	0
2	0	1	1	4	4	0	1	1
3	1	1	1	4	4	1	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

3. *Cluster* berjumlah empat :

- Pengguna yang tidak memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9957 & 6,9983 & 6,9928 & 6,9973 \\ 6,9951 & 6,9976 & 6,9962 & 6,9967 \\ 6,9949 & 6,9975 & 6,9920 & 6,9949 \\ \vdots & \vdots & \vdots & \vdots \\ 6,9958 & 6,9983 & 6,9929 & 6,9974 \\ 6,9949 & 6,9975 & 6,992 & 6,9965 \\ 6,9951 & 6,9976 & 6,9962 & 6,9922 \end{pmatrix} = \begin{pmatrix} \text{Cluster 3} \\ \text{Cluster 1} \\ \text{Cluster 3} \\ \vdots \\ \text{Cluster 3} \\ \text{Cluster 3} \\ \text{Cluster 4} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.18 Modus *Cluster* Berjumlah Empat Kedua Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	> 4	3	0	0	0
2	0	1	0	> 4	3	0	0	0
3	0	1	1	> 4	3	1	0	0
4	0	1	1	4	3	0	0	0

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9957 & 6,9967 & \mathbf{6,9928} & 6,9973 \\ \mathbf{6,9951} & 6,996 & 6,9962 & 6,9967 \\ 6,9949 & 6,9958 & \mathbf{6,9920} & 6,9949 \\ \vdots & \vdots & \vdots & \vdots \\ 6,9958 & 6,9967 & \mathbf{6,9929} & 6,9974 \\ 6,9949 & 6,9959 & \mathbf{6,992} & 6,9965 \\ 6,9951 & 6,996 & 6,9962 & \mathbf{6,9922} \end{pmatrix} = \begin{pmatrix} \text{Cluster 3} \\ \text{Cluster 1} \\ \text{Cluster 3} \\ \vdots \\ \text{Cluster 3} \\ \text{Cluster 3} \\ \text{Cluster 4} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.19 Modus *Cluster* Berjumlah Empat Ketiga Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	> 4	3	0	0	0
2	0	1	0	> 4	3	0	0	0
3	0	1	1	> 4	3	1	0	0
4	0	1	1	4	3	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

- Pengguna yang memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & 6,9920 & \mathbf{6,9835} & 6,9914 \\ \mathbf{6,9913} & 6,9946 & 6,9977 & 6,9916 \\ \mathbf{6,9882} & 6,9916 & 6,9935 & 6,9966 \\ \vdots & \vdots & \vdots & \vdots \\ 6,994 & 6,9915 & \mathbf{6,9866} & 6,9942 \\ \mathbf{6,9889} & 6,9922 & 6,9954 & 6,9916 \\ \mathbf{6,9889} & 6,9922 & 6,9954 & 6,9916 \end{pmatrix} = \begin{pmatrix} \text{Cluster 3} \\ \text{Cluster 1} \\ \text{Cluster 1} \\ \vdots \\ \text{Cluster 3} \\ \text{Cluster 1} \\ \text{Cluster 1} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.20 Modus *Cluster* Berjumlah Empat Kedua Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	4	4	0	0	1
2	0	1	1	4	4	0	0	0
3	1	0	1	4	3	1	0	0

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
4	1	1	1	>4	4	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

4. *Cluster* berjumlah lima :

- Pengguna yang tidak memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9928 & 6,9967 & 6,9957 & 6,995 & 6,9966 \\ 6,9925 & 6,996 & 6,9951 & 6,9979 & 6,996 \\ 6,9989 & 6,9958 & 6,9949 & 6,9961 & 6,9958 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9966 & 6,9967 & 6,9957 & 6,9985 & 6,9967 \\ 6,9957 & 6,9959 & 6,9949 & 6,9977 & 6,9958 \\ 6,9957 & 6,996 & 6,9951 & 6,9933 & 6,996 \end{pmatrix} = \begin{pmatrix} \text{Cluster 1} \\ \text{Cluster 1} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 3} \\ \text{Cluster 4} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.21 Modus *Cluster* Berjumlah Lima Kedua Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	0	1	>4	4	0	0	0
2	0	1	0	>4	3	0	0	0
3	0	1	1	>4	3	0	0	0
4	0	0	1	4	3	0	0	0
5	0	1	1	>4	3	0	1	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

- Pengguna yang memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9890 & 6,9969 & 6,9897 & 6,9914 & 6,9887 \\ 6,994 & 6,9850 & 6,9970 & 6,9916 & 6,9860 \\ 6,999 & 6,9906 & 6,9939 & 6,9966 & 6,9817 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9918 & 6,9967 & 6,9957 & 6,9985 & 6,9967 \\ 6,994 & 6,9913 & 6,9947 & 6,9916 & 6,9836 \\ 6,994 & 6,9913 & 6,9947 & 6,9916 & 6,9836 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 2} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 1} \\ \text{Cluster 5} \\ \text{Cluster 5} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.22 Modus *Cluster* Berjumlah Lima Kedua Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	1	1	>4	3	0	0	0
2	0	1	1	<4	4	0	0	1
3	0	1	1	4	3	0	0	0
4	1	1	1	>4	4	0	0	0
5	1	1	1	4	4	1	1	1

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

5. *Cluster* berjumlah enam :

- Pengguna yang tidak memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & \mathbf{6,9929} & 6,9957 & 6,9983 & 6,9934 & 6,9974 \\ 6,9939 & 6,9962 & 6,9951 & 6,9944 & 6,9963 & \mathbf{6,9935} \\ 6,9971 & \mathbf{6,9921} & 6,9949 & 6,9975 & 6,9961 & 6,9935 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9979 & \mathbf{6,9929} & 6,9958 & 6,9951 & 6,997 & 6,9942 \\ 6,9971 & \mathbf{6,9921} & 6,995 & 6,9943 & 6,9961 & 6,9933 \\ \mathbf{6,9939} & 6,9962 & 6,996 & 6,9976 & 6,9963 & 6,9967 \end{pmatrix} = \begin{pmatrix} \text{Cluster 2} \\ \text{Cluster 6} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \text{Cluster 1} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.23 Modus *Cluster* Berjumlah Enam Kedua Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	1	0	>4	3	0	0	0
2	0	1	1	>4	3	1	0	0
3	0	1	1	>4	3	0	0	0
4	0	1	1	>4	3	0	1	0
5	0	0	1	>4	3	0	0	0
6	0	1	1	>4	4	0	0	0

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & 6,9929 & 6,9957 & 6,9967 & \mathbf{6,9913} & 6,9974 \\ 6,9939 & 6,9962 & 6,9951 & 6,9961 & 6,9941 & \mathbf{6,9935} \\ 6,9971 & \mathbf{6,9921} & 6,9949 & 6,9958 & 6,9973 & 6,9965 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9979 & \mathbf{6,9929} & 6,9958 & 6,9967 & 6,9982 & 6,9942 \\ 6,9971 & \mathbf{6,9921} & 6,9949 & 6,9959 & 6,9974 & 6,9933 \\ \mathbf{6,9939} & 6,9962 & 6,9951 & 6,996 & 6,9941 & 6,9967 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 6} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \text{Cluster 1} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.24 Modus *Cluster* Berjumlah Enam Ketiga Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	1	0	>4	3	0	0	0
2	0	1	1	>4	3	1	0	0
3	0	1	1	>4	3	0	0	0
4	0	1	1	>4	3	0	1	0
5	1	0	1	>4	3	0	0	0
6	0	1	1	>4	4	0	0	0

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9945 & 6,9929 & 6,9957 & 6,9967 & 6,9913 & 6,9974 \\ 6,9939 & 6,9962 & 6,9951 & 6,9961 & 6,9941 & 6,9935 \\ 6,9971 & 6,9921 & 6,9949 & 6,9958 & 6,9973 & 6,9965 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9979 & 6,9929 & 6,9958 & 6,9967 & 6,9982 & 6,9942 \\ 6,9971 & 6,9921 & 6,9949 & 6,9959 & 6,9974 & 6,9933 \\ 6,9939 & 6,9962 & 6,9951 & 6,996 & 6,9941 & 6,9967 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 6} \\ \text{Cluster 2} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 2} \\ \text{Cluster 1} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.25 Modus *Cluster* Berjumlah Enam Keempat Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	1	0	> 4	3	0	0	0
2	0	1	1	> 4	3	1	0	0
3	0	1	1	> 4	3	0	0	0
4	0	1	1	> 4	3	0	1	0
5	1	0	1	> 4	3	0	0	0
6	0	1	1	> 4	4	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

- Pengguna yang memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9912 & 6,9841 & 6,9890 & 6,9966 & 6,9841 & 6,9920 \\ 6,9874 & 6,9926 & 6,9916 & 6,9871 & 6,99 & 6,9908 \\ 6,9900 & 6,9893 & 6,9942 & 6,984 & 6,9941 & 6,9963 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9963 & 6,9962 & 6,9958 & 6,9961 & 6,9876 & 6,9829 \\ 6,9850 & 6,9912 & 6,9892 & 6,9847 & 6,996 & 6,9970 \\ 6,9850 & 6,9912 & 6,9892 & 6,9847 & 6,996 & 6,9970 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 4} \\ \text{Cluster 4} \\ \vdots \\ \text{Cluster 6} \\ \text{Cluster 4} \\ \text{Cluster 4} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.26 Modus *Cluster* Berjumlah Enam Kedua Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	1	1	4	3	0	1	0
2	1	1	1	4	4	1	0	1
3	1	1	1	4	4	0	0	0
4	0	1	1	4	4	0	1	1
5	1	1	1	< 4	3	1	0	0
6	0	1	1	< 4	3	0	0	0

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9989 & 6,9866 & 6,9890 & 6,9966 & 6,9841 & 6,9920 \\ 6,9898 & 6,9902 & 6,9916 & 6,9871 & 6,9898 & 6,9908 \\ 6,9924 & 6,9859 & 6,9942 & 6,984 & 6,9941 & 6,9963 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9939 & 6,9986 & 6,9942 & 6,9961 & 6,9876 & 6,9829 \\ 6,9874 & 6,9878 & 6,9892 & 6,9847 & 6,996 & 6,9970 \\ 6,9874 & 6,9878 & 6,9892 & 6,9847 & 6,996 & 6,9970 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 4} \\ \text{Cluster 4} \\ \vdots \\ \text{Cluster 6} \\ \text{Cluster 4} \\ \text{Cluster 4} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.27 Modus *Cluster* Berjumlah Enam Ketiga Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	1	1	4	3	0	1	0
2	1	1	1	4	4	1	0	1
3	1	1	1	4	4	0	0	0
4	0	1	1	4	4	0	1	1
5	1	1	1	< 4	3	1	0	0
6	0	1	1	< 4	3	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

6. *Cluster* berjumlah tujuh :

- Pengguna yang tidak memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9929 & 6,9967 & 6,9973 & 6,9982 & \mathbf{6,9905} & 6,9973 & 6,9957 \\ 6,9957 & 6,9960 & \mathbf{6,9935} & 6,9976 & 6,9973 & 6,9967 & 6,9951 \\ 6,9941 & 6,9959 & 6,9965 & 6,9958 & \mathbf{6,9932} & 6,9949 & 6,9949 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9978 & 6,9967 & 6,9942 & \mathbf{6,9928} & 6,9941 & 6,9974 & 6,9958 \\ 6,999 & 6,9959 & 6,9933 & 6,9975 & \mathbf{6,9932} & 6,9965 & 6,995 \\ 6,9941 & 6,9960 & 6,9967 & 6,9931 & 6,9973 & \mathbf{6,9922} & 6,9951 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 3} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 4} \\ \text{Cluster 5} \\ \text{Cluster 6} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.28 Modus *Cluster* Berjumlah Tujuh Kedua Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	0	1	< 4	3	0	0	0
2	0	1	0	> 4	3	0	0	0
3	0	1	1	> 4	4	0	0	0
4	0	1	1	> 4	3	0	1	0
5	0	0	1	> 4	3	1	0	0
6	0	1	1	4	3	0	0	0
7	0	1	1	> 4	3	0	0	0

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9929 & 6,9967 & 6,9973 & 6,9967 & \mathbf{6,9905} & 6,9973 & 6,9957 \\ 6,9957 & 6,9960 & \mathbf{6,9935} & 6,9960 & 6,9973 & 6,9967 & 6,9951 \\ 6,9941 & 6,9959 & 6,9965 & 6,9958 & \mathbf{6,9932} & 6,9949 & 6,9949 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9978 & 6,9967 & 6,9942 & 6,9967 & \mathbf{6,9941} & 6,9974 & 6,9958 \\ 6,999 & 6,9959 & 6,9933 & 6,9959 & \mathbf{6,9932} & 6,9965 & 6,995 \\ 6,9941 & 6,9960 & 6,9967 & 6,996 & 6,9973 & \mathbf{6,9922} & 6,9951 \end{pmatrix} = \begin{pmatrix} \text{Cluster 5} \\ \text{Cluster 3} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 5} \\ \text{Cluster 5} \\ \text{Cluster 6} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.29 Modus *Cluster* Berjumlah Tujuh Ketiga Tidak Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	1	0	1	< 4	3	0	0	0
2	0	1	0	> 4	3	0	0	0
3	0	1	1	> 4	4	0	0	0
4	0	1	1	> 4	3	0	1	0
5	0	0	1	> 4	3	1	0	0
6	0	1	1	4	3	0	0	0
7	0	1	1	> 4	3	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

- Pengguna yang memiliki asuransi

$$d(x_i, x_j) = \begin{pmatrix} 6,9897 & 6,9914 & 6,9888 & \mathbf{6,9835} & 6,9915 & 6,9913 & 6,9920 \\ 6,9970 & 6,9988 & 6,9898 & 6,9977 & \mathbf{6,9882} & 6,9987 & 6,9946 \\ 6,994 & 6,9957 & 6,9924 & 6,9935 & \mathbf{6,9909} & 6,9956 & 6,9916 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9892 & \mathbf{6,982} & 6,994 & 6,9866 & 6,9966 & 6,9908 & 6,9915 \\ 6,9946 & 6,9964 & 6,9874 & 6,9954 & \mathbf{6,9859} & 6,9963 & 6,9922 \\ 6,9946 & 6,9964 & 6,9874 & 6,9954 & \mathbf{6,9859} & 6,9963 & 6,9922 \end{pmatrix} = \begin{pmatrix} \text{Cluster 4} \\ \text{Cluster 5} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 5} \\ \text{Cluster 5} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.30 Modus *Cluster* Berjumlah Tujuh Kedua Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	4	3	0	0	0
2	0	0	1	4	3	0	0	0
3	1	1	1	4	3	0	0	0
4	1	0	1	4	3	1	0	0
5	1	1	1	4	4	0	0	1
6	0	1	0	4	4	0	0	0
7	0	1	1	4	4	0	0	0

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9897 & 6,9914 & 6,9867 & \mathbf{6,9835} & 6,9915 & 6,9937 & 6,9920 \\ 6,9970 & 6,9988 & 6,994 & 6,9977 & \mathbf{6,9882} & 6,9963 & 6,9946 \\ 6,994 & 6,9966 & 6,9924 & 6,9935 & \mathbf{6,9909} & 6,9932 & 6,9916 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9892 & \mathbf{6,982} & 6,9918 & 6,9866 & 6,9966 & 6,9932 & 6,9915 \\ 6,9946 & 6,9964 & 6,9916 & 6,9954 & \mathbf{6,9859} & 6,9939 & 6,9922 \\ 6,9946 & 6,9964 & 6,9916 & 6,9954 & \mathbf{6,9859} & 6,9939 & 6,9922 \end{pmatrix} = \begin{pmatrix} \text{Cluster 4} \\ \text{Cluster 5} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 5} \\ \text{Cluster 5} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.31 Modus *Cluster* Berjumlah Tujuh Ketiga Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	4	3	0	0	0

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
2	0	0	1	4	3	0	0	0
3	1	1	1	4	4	0	0	0
4	1	0	1	4	3	1	0	0
5	1	1	1	4	4	0	0	1
6	0	1	0	4	4	0	0	0
7	0	1	1	4	4	0	0	0

Dilakukan perulangan perhitungan jarak data terhadap *initial cluster* baru dikarenakan masih terdapat perbedaan *initial cluster* baru dengan lama.

$$d(x_i, x_j) = \begin{pmatrix} 6,9897 & 6,9914 & 6,9890 & \mathbf{6,9835} & 6,9915 & 6,9937 & 6,9920 \\ 6,9970 & 6,9988 & 6,9916 & 6,9977 & \mathbf{6,9882} & 6,9963 & 6,9946 \\ 6,994 & 6,9966 & 6,9942 & 6,9935 & \mathbf{6,9909} & 6,9932 & 6,9916 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 6,9892 & \mathbf{6,982} & 6,9942 & 6,9866 & 6,9966 & 6,9932 & 6,9915 \\ 6,9946 & 6,9964 & 6,9892 & 6,9954 & \mathbf{6,9859} & 6,9939 & 6,9922 \\ 6,9946 & 6,9964 & 6,9892 & 6,9954 & \mathbf{6,9859} & 6,9939 & 6,9922 \end{pmatrix} = \begin{pmatrix} \text{Cluster 4} \\ \text{Cluster 5} \\ \text{Cluster 5} \\ \vdots \\ \text{Cluster 2} \\ \text{Cluster 5} \\ \text{Cluster 5} \end{pmatrix}$$

Didapatkan nilai modus dari masing-masing *cluster* tersebut adalah

Tabel 5.32 Modus *Cluster* Berjumlah Tujuh Keempat Punya Asuransi

Cluster	Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1	0	1	1	4	3	0	0	0
2	0	0	1	4	3	0	0	0
3	1	1	1	4	4	0	0	0
4	1	0	1	4	3	1	0	0
5	1	1	1	4	4	0	0	1
6	0	1	0	4	4	0	0	0
7	0	1	1	4	4	0	0	0

Dikarenakan modus dari *cluster* baru sama dengan modus *cluster* sebelumnya, maka iterasi dihentikan.

5.4. Silhouette Coefficient

5.4.1 Jarak Nominal

Pengukuran jarak pada variabel nominal antar data merupakan suatu cara untuk mengetahui kemiripan antar data pada variabel nominal. Semakin kecil jarak maka akan semakin mirip kedua data tersebut.

Jarak nominal antar data pengguna yang tidak memiliki asuransi dihasilkan berdimensi matriks 1277 x 1277 adalah sebagai berikut.

$$Ndis(x_i, x_j) = \begin{pmatrix} 0 & 5,9957 & 5,9944 & \dots & 5,9937 & 5,9928 & 5,9973 \\ 5,9957 & 0 & 5,9961 & \dots & 5,9954 & 5,9945 & 5,9951 \\ 5,9944 & 5,9961 & 0 & \dots & 5,9941 & 5,9932 & 5,9961 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 5,9937 & 5,9954 & 5,9941 & \dots & 0 & 5,9924 & 5,9970 \\ 5,9928 & 5,9945 & 5,9932 & \dots & 5,9925 & 0 & 5,9961 \\ 5,9973 & 5,9951 & 5,9961 & \dots & 5,9970 & 5,9961 & 0 \end{pmatrix}$$

Sedangkan untuk data pengguna yang memiliki asuransi dihasilkan matriks berdimensi 710 x 710 sebagai berikut.

$$Ndis(x_i, x_j) = \begin{pmatrix} 0 & 5,9966 & 5,9893 & \dots & 5,9938 & 5,9942 & 5,9942 \\ 5,9966 & 0 & 5,989 & \dots & 5,9901 & 5,9870 & 5,9870 \\ 5,9893 & 5,989 & 0 & \dots & 5,9984 & 5,9866 & 5,9866 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 5,9938 & 5,9901 & 5,9984 & \dots & 0 & 5,9964 & 5,9964 \\ 5,9942 & 5,9870 & 5,9866 & \dots & 5,9964 & 0 & 5,9846 \\ 5,9942 & 5,9870 & 5,9866 & \dots & 5,9964 & 5,9846 & 0 \end{pmatrix}$$

5.4.2 Jarak Ordinal

Pengukuran jarak pada variabel ordinal antar data merupakan suatu cara untuk mengetahui kemiripan antar data pada variabel ordinal. Variabel ordinal pada penelitian ini berjumlah dua sehingga masing-masing variabel ordinal dihitung jarak antar data berdasarkan syarat yang memenuhi variabel ordinal tersebut.

Variabel ordinal pertama yaitu *Age* memenuhi perhitungan jarak ordinal syarat 1 yang ditunjukkan pada persamaan 5.1. Perhitungan jarak ordinal ini memiliki perhitungan yang sama dengan perhitungan jarak pada variabel nominal. Jarak variabel ordinal pertama dari data pengguna jasa yang tidak memiliki asuransi dihasilkan adalah matriks 1277 x 1277 sebagai berikut.

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0 & 0,9987 & 1 & \dots & 1 & 1 & 0,9987 \\ 0,9987 & 0 & 1 & \dots & 1 & 1 & 0,9987 \\ 1 & 1 & 0 & \dots & 0,9988 & 0,9988 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 0 & 0,9988 & 1 \\ 1 & 1 & 0,9988 & \dots & 0,9988 & 0 & 1 \\ 0,9987 & 0,9987 & 1 & \dots & 1 & 1 & 0 \end{pmatrix}$$

sedangkan dari data pengguna jasa yang memiliki asuransi dihasilkan matriks 710 x 710 berikut.

$$Odis_1(x_i, x_j) = \begin{pmatrix} 0 & 0,997 & 1 & \dots & 1 & 0,997 & 0,997 \\ 0,997 & 0 & 1 & \dots & 1 & 0,997 & 0,997 \\ 1 & 1 & 0 & \dots & 0,9973 & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & 1 & 0,9973 & \dots & 0 & 1 & 1 \\ 0,997 & 0,997 & 1 & \dots & 1 & 0 & 0,997 \\ 0,997 & 0,997 & 1 & \dots & 1 & 0,997 & 0 \end{pmatrix}$$

Variabel ordinal pertama yaitu *AnnualIncome* memenuhi perhitungan jarak ordinal syarat 2 yang ditunjukkan pada persamaan 5.2. Pada data pengguna jasa yang tidak memiliki asuransi didapatkan nilai Dod_1 yang digunakan sebesar 0,0016077170418000643 didapatkan matriks jarak ordinal kedua adalah sebagai berikut.

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0,0016 & 0 & \dots & 0,0016 & 0,0016 & 0 \\ 0,0016 & 0 & 0,0016 & \dots & 0 & 0 & 0,0016 \\ 0 & 0,0016 & 0 & \dots & 0,0016 & 0,0016 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0,0016 & 0 & 0,0016 & \dots & 0 & 0 & 0,0016 \\ 0,0016 & 0 & 0,0016 & \dots & 0 & 0 & 0,0016 \\ 0 & 0,0016 & 0 & \dots & 0,0016 & 0,0016 & 0 \end{pmatrix}$$

Sedangkan pada data pengguna jasa yang memiliki asuransi didapatkan nilai Dod_1 sebesar 0.00239234449760765 sehingga didapatkan matriks jarak ordinal kedua sebagai berikut.

$$Odis_2(x_i, x_j) = \begin{pmatrix} 0 & 0,0024 & 0,0024 & \dots & 0 & 0,0024 & 0,0024 \\ 0,0024 & 0 & 0 & \dots & 0,0024 & 0 & 0 \\ 0,0024 & 0 & 0 & \dots & 0,0024 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0,0024 & 0,0024 & \dots & 0 & 0,0024 & 0,0024 \\ 0,0024 & 0 & 0 & \dots & 0,0024 & 0 & 0 \\ 0,0024 & 0 & 0 & \dots & 0,0024 & 0 & 0 \end{pmatrix}$$

5.4.3 Jarak Gabungan

Jarak gabungan merupakan penjumlahan jarak antar data pada variabel nominal dan variabel ordinal, sehingga keluaran dari matriks jarak gabungan pada data pengguna jasa yang tidak memiliki asuransi perjalanan adalah sebagai berikut.

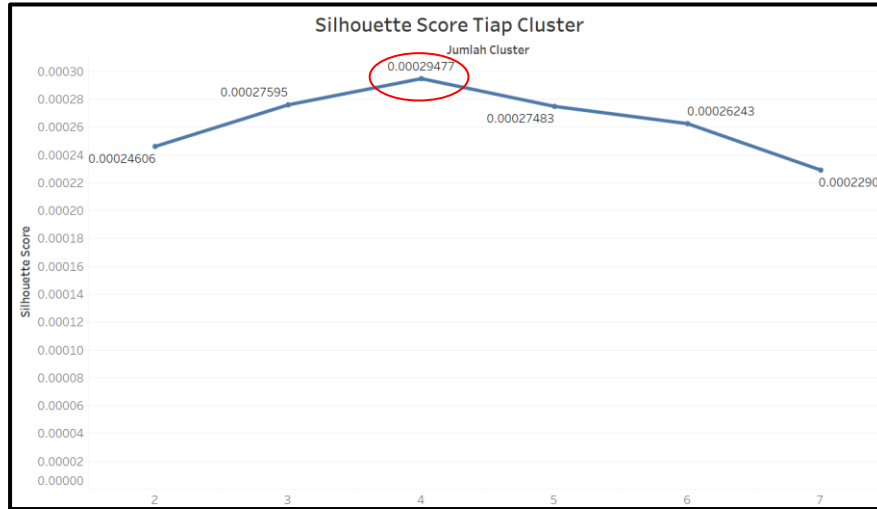
$$d(x_i, x_j) = \begin{pmatrix} 0 & 6,9952 & 6,9944 & \dots & 6,9953 & 6,9944 & 6,9952 \\ 6,9952 & 0 & 6,9978 & \dots & 6,9954 & 6,9945 & 6,9945 \\ 6,9944 & 6,9978 & 0 & \dots & 6,9945 & 6,9936 & 6,9961 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 6,9953 & 6,9954 & 6,9945 & \dots & 0 & 6,9912 & 6,9986 \\ 6,9944 & 6,9945 & 6,9936 & \dots & 6,9912 & 0 & 6,9977 \\ 6,9952 & 6,9945 & 6,9961 & \dots & 6,9986 & 6,9977 & 0 \end{pmatrix}$$

sedangkan matriks jarak pada data pengguna jasa yang memiliki asuransi perjalanan sebagai berikut.

$$d(x_i, x_j) = \begin{pmatrix} 0 & 6,996 & 6,9917 & \dots & 6,9938 & 6,9936 & 6,9936 \\ 6,996 & 0 & 6,989 & \dots & 6,9925 & 6,984 & 6,984 \\ 6,9917 & 6,989 & 0 & \dots & 6,9981 & 6,9866 & 6,9866 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 6,9938 & 6,9925 & 6,9981 & \dots & 0 & 6,9988 & 6,9988 \\ 6,9936 & 6,984 & 6,9866 & \dots & 6,9988 & 0 & 6,9816 \\ 6,9936 & 6,984 & 6,9866 & \dots & 6,9988 & 6,9816 & 0 \end{pmatrix}$$

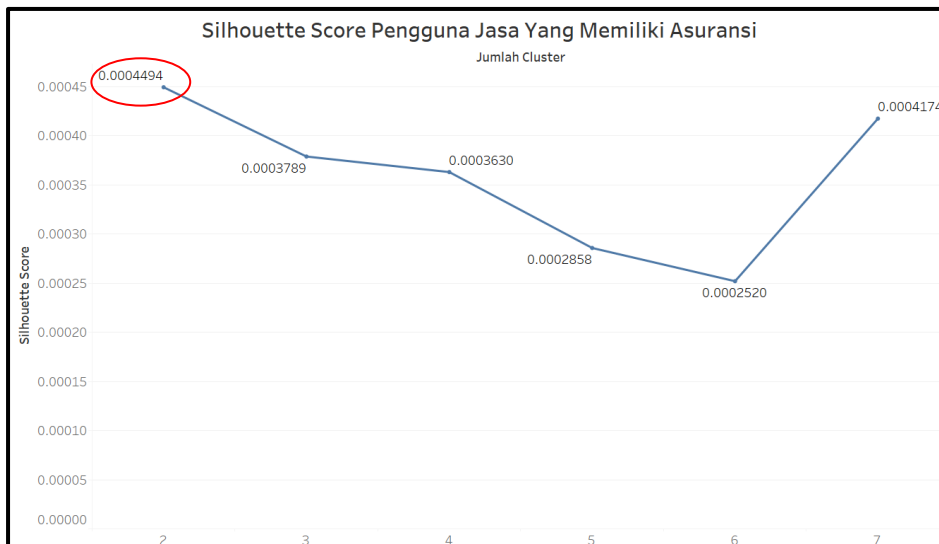
5.4.4 Evaluasi Cluster

Nilai *silhouette coefficient* dengan rujukan dari persamaan rumus 3.16, 3.17 dan 3.18 dari setiap *cluster* yang dihasilkan dijelaskan pada **Gambar 5.7** berikut.



Gambar 5.7 Nilai *Silhouette Score* Tiap *Cluster* Tidak Asuransi

Apabila dilihat dari nilai *silhouette cluster* maka digunakan nilai yang paling tinggi sehingga pada studi kasus ini digunakan *cluster* berjumlah 4 yaitu sebesar 0,00029477. Nilai *silhouette score* yang didapatkan pada studi kasus ini berada di bawah 0,25 yang dapat dikatakan bahwa *cluster* yang terbentuk berstruktur buruk. Sedangkan pada data dengan pengguna jasa yang memiliki asuransi didapatkan nilai *silhouette score* per *cluster* sebagai berikut.



Gambar 5.8 Nilai *Silhouette Score* Tiap *Cluster* Punya Asuransi

Didapatkan nilai *silhouette score* tertinggi pada *cluster* berjumlah dua yaitu sebesar 0,0004494 sehingga kemudian data pengguna jasa yang memiliki asuransi dikelompokkan sebanyak 2 *cluster*.

5.5. Profilisasi

Kelompok yang terbentuk dari data pengguna jasa yang tidak memiliki asuransi adalah sejumlah empat kelompok dengan rincian kelompok dengan persentase variabel untuk setiap *cluster* sebagai berikut:

1. Kelompok pertama dengan modus *cluster*

Tabel 5.33 Modus *Cluster* Pertama

Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
0 (63,3%)	1 (63,68%)	1 (100%)	>4 (67,19%)	3 (53,16%)	0 (100%)	0 (84,91%)	0 (94,21%)

merupakan kelompok yang beranggotakan mayoritas kelompok umur 30 tahun ke bawah atau kelompok pekerja awal dan tidak memiliki penyakit kronis. Tipe pekerjaan pada kelompok ini adalah pekerja di sektor swasta dengan latar belakang lulusan perguruan tinggi. Kelompok ini mayoritas memiliki jumlah anggota keluarga berjumlah lebih dari 4 anggota. Pada kelompok ini kebanyakan tidak pernah ke luar negeri dan bukan pengguna jasa yang sering melakukan penerbangan. Pengguna jasa pada kelompok ini kebanyakan merupakan masyarakat dengan pendapatan menengah ke atas. Kelompok ini terdiri dari sebanyak 570 pengguna.

2. Kelompok kedua dengan modus *cluster*

Tabel 5.34 Modus *Cluster* Kedua

Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
0 (63,78%)	1 (77,55%)	0 (100%)	>4 (42,86%)	3 (56,63%)	0 (73,98%)	0 (84,69%)	0 (93,37%)

merupakan kelompok yang beranggotakan mayoritas kelompok umur 30 tahun ke bawah atau kelompok pekerja awal dan tidak memiliki penyakit kronis. Tipe pekerjaan pada kelompok ini adalah pekerja di sektor swasta dengan latar belakang bukan lulusan perguruan tinggi. Kelompok ini mayoritas memiliki jumlah anggota keluarga berjumlah lebih dari 4 anggota. Pada kelompok ini kebanyakan tidak pernah ke luar negeri dan bukan pengguna jasa yang sering melakukan penerbangan. Pengguna jasa

pada kelompok ini kebanyakan merupakan masyarakat dengan pendapatan menengah ke atas. Kelompok ini terdiri dari sebanyak 196 pengguna.

3. Kelompok ketiga dengan modus *cluster*

Tabel 5.35 Modus *Cluster* Ketiga

Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
0 (63,8%)	1 (62,90%)	1 (100%)	>4 (71,04%)	3 (49,32%)	1 (100%)	0 (88,69%)	0 (92,31%)

merupakan kelompok yang beranggotakan mayoritas kelompok umur 30 tahun ke bawah atau kelompok pekerja awal dan memiliki penyakit kronis. Tipe pekerjaan pada kelompok ini adalah pekerja di sektor swasta dengan latar belakang lulusan perguruan tinggi. Kelompok ini mayoritas memiliki jumlah anggota keluarga berjumlah lebih 4 anggota. Pada kelompok ini kebanyakan tidak pernah ke luar negeri dan bukan pengguna jasa yang sering melakukan penerbangan. Pengguna jasa pada kelompok ini kebanyakan merupakan masyarakat dengan pendapatan menengah ke atas. Kelompok ini terdiri dari sebanyak 221 pengguna.

4. Kelompok keempat dengan modus *cluster*

Tabel 5.36 Modus *Cluster* Keempat

Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
0 (63,45%)	1 (66,55%)	1 (100%)	4 (100%)	3 (51,03%)	0 (74,14%)	0 (87,24%)	0 (93,45%)

merupakan kelompok yang beranggotakan mayoritas kelompok umur 30 tahun ke bawah atau kelompok pekerja awal dan tidak memiliki penyakit kronis. Tipe pekerjaan pada kelompok ini adalah pekerja di sektor swasta dengan latar belakang lulusan perguruan tinggi. Kelompok ini mayoritas memiliki jumlah anggota keluarga berjumlah 4 anggota. Pada kelompok ini kebanyakan tidak pernah ke luar negeri dan bukan pengguna jasa yang sering melakukan penerbangan. Pengguna jasa pada kelompok ini kebanyakan merupakan masyarakat dengan pendapatan menengah ke atas. Kelompok ini terdiri dari sebanyak 290 pengguna.

Sedangkan pada data pengguna jasa yang memiliki asuransi dikelompokkan menjadi dua kelompok dengan rincian sebagai berikut.

1. Kelompok pertama dengan modus *cluster*

Tabel 5.37 Modus *Cluster* Pertama Punya Asuransi

Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
0 (68,98%)	1 (75%)	1 (78,92%)	4 (57,23%)	3 (64,46%)	0 (70,18%)	0 (75,90%)	0 (97,89%)

merupakan kelompok yang beranggotakan sebanyak 332 yang mana berusia maksimal 30 tahun dan merupakan pekerja swasta. Pada kelompok ini pengguna jasa mayoritas merupakan lulus perguruan tinggi dengan jumlah keluarga 4 dan pendapatan *upper-middle*. Pengguna pada kelompok ini jarang melakukan bepergian maupun belum pernah ke luar negeri. Pada kelompok ini mayoritas pengguna tidak memiliki riwayat penyakit kronis.

2. Kelompok kedua dengan modus *cluster*

Tabel 5.38 Modus *Cluster* Kedua Punya Asuransi

Age	Employment Type	Graduate Or Not	Family Members	Annual Income	Chronic Diseases	Frequent Flyer	Ever Travelled Abroad
1 (60,85%)	1 (84,92%)	1 (92,33%)	4 (60,85%)	4 (97,62%)	0 (71,96%)	0 (57,94%)	1 (76,98%)

merupakan kelompok yang beranggotakan sebanyak 378 yang mana berusia lebih dari 30 tahun dan merupakan pekerja swasta. Pada kelompok ini pengguna jasa mayoritas merupakan lulus perguruan tinggi dengan jumlah keluarga 4 dan pendapatan tinggi (*high income*). Pengguna pada kelompok ini jarang melakukan penerbangan yang berbeda namun pernah ke luar negeri. Pada kelompok ini mayoritas pengguna tidak memiliki riwayat penyakit kronis.

Apabila dibandingkan antara karakteristik *cluster* yang terbentuk pada kelompok keempat pengguna jasa yang tidak memiliki asuransi sama dengan karakteristik kelompok pertama pengguna jasa yang memiliki asuransi. Kesamaan karakteristik tersebut ditandai dengan garis warna kuning pada **Tabel 5.33** dan **5.34**. Kesamaan karakteristik ini dapat memungkinkan untuk pengguna jasa yang tidak memiliki asuransi pada kelompok keempat dapat memiliki asuransi. Hal tersebut dapat diupayakan dengan memberikan penawaran spesial.

Penawaran spesial yang diberikan dapat berupa memberikan produk menarik seperti manfaat perawatan anak, pembatalan karena penundaan perjalanan atau kehilangan pendapatan akibat kecelakaan. Pada kelompok kedua pengguna jasa yang memiliki asuransi merupakan pengguna jasa yang pernah keluar negeri (tanda

warna kuning pada **Tabel 5.35**), aspek tersebut memungkinkan untuk pengguna melakukan bepergian ke luar negeri lagi. Pada kelompok tersebut dapat diberikan penawaran produk seperti produk pemulangan jenazah, agar tidak terdapat kendala pengurusan jenazah apabila terdapat kecelakaan hingga kematian di luar negeri.

BAB 6

PENUTUP

6.1. Kesimpulan

Dari penelitian yang telah dilakukan dapat dirangkum ke dalam kesimpulan berikut :

1. Metode *K-Modes Clustering* dengan menggunakan perhitungan jarak Yuan dapat diimplementasikan ke dalam data asuransi perjalanan dengan pengguna jasa yang tidak memiliki asuransi membentuk kelompok sebanyak 4 kelompok dan nilai *silhouette coefficient* sebesar 0,00029477. Sedangkan pada data pengguna jasa yang memiliki asuransi membentuk kelompok sebanyak 2 kelompok dan nilai *silhouette coefficient* sebesar 0,000494.
2. Keempat kelompok yang terbentuk didominasi dari pengguna jasa yang berusia 30 tahun ke bawah dengan tipe pekerjaan pekerja swasta. Kelompok yang terbentuk juga didominasi oleh pengguna jasa yang sering bepergian dan belum pernah ke luar negeri dengan pendapatan menengah ke atas. Kelompok pertama merupakan kelompok yang mayoritas memiliki jumlah anggota keluarga berjumlah lebih dari 4 anggota dan latar belakang lulusan perguruan tinggi. Pengguna jasa pada kelompok ini tidak memiliki riwayat penyakit kronis. Kelompok kedua merupakan kelompok yang mayoritas memiliki jumlah anggota keluarga berjumlah lebih dari 4 anggota dan latar belakang bukan lulusan perguruan tinggi. Pengguna jasa pada kelompok ini tidak memiliki riwayat penyakit kronis. Kelompok ketiga merupakan kelompok yang mayoritas memiliki jumlah anggota keluarga berjumlah lebih dari 4 anggota dan latar belakang lulusan perguruan tinggi. Pengguna jasa pada kelompok ini memiliki riwayat penyakit kronis. Kelompok keempat merupakan kelompok yang mayoritas memiliki jumlah anggota keluarga berjumlah 4 anggota dan latar belakang lulusan perguruan tinggi. Pengguna jasa pada kelompok ini tidak memiliki riwayat penyakit kronis. Pada pengguna jasa yang memiliki asuransi terdapat dua kelompok yang

mana keduanya merupakan pengguna jasa lulusan perguruan tinggi, pegawai swasta, memiliki jumlah keluarga empat, tidak melakukan bepergian dengan lebih dari empat penerbangan yang berbeda dan tidak memiliki penyakit kronis. Kelompok pertama pengguna jasa yang memiliki asuransi sama dengan kelompok keempat pengguna jasa yang tidak memiliki asuransi sehingga dapat dilakukan upaya untuk meningkatkan minat berasuransi pada kelompok tersebut. Kelompok kedua merupakan kelompok pendapatan tinggi yang pernah ke luar negeri.

3. Didapatkan sintaks perhitungan jarak data nominal dan data ordinal sesuai dengan perhitungan jarak Yuan namun belum didapatkan sintaks untuk mengkombinasikan perhitungan jarak Yuan tersebut ke dalam analisis *K-Modes Clustering* secara efisien seperti menggunakan *package*.

6.2. Saran

Saran yang dapat diajukan setelah dilakukan penelitian ini adalah:

1. Data yang digunakan dapat menggunakan data asuransi perjalanan di Indonesia.
2. Pada penelitian ini hanya digunakan satu metode evaluasi pemilihan *cluster* terbaik sehingga pada penelitian selanjutnya dapat digunakan lebih dari satu metode evaluasi agar dapat dilakukan perbandingan.
3. Pada penelitian selanjutnya dapat dikembangkan *package* perhitungan analisis *K-Modes Clustering* dengan jarak Yuan agar lebih efisien untuk digunakan.
4. Agen perjalanan yang digunakan pada studi kasus dapat mengupayakan penawaran seperti pada bagian pembahasan pada kelompok pengguna jasa yang tidak memiliki asuransi karena terdapat kelompok yang sama antara kelompok pengguna jasa yang tidak memiliki asuransi dan memiliki asuransi. Sedangkan pada kelompok pengguna jasa yang memiliki asuransi dapat diberikan penawaran produk yang sesuai dengan karakteristik kelompok tersebut.

DAFTAR PUSTAKA

- Agus, R. M., & Fahrizqi, E. B. (2020). ANALISIS TINGKAT KEPERCAYAAN DIRI SAAT BERTANDING ATLET PENCAK SILAT PERGURUAN SATRIA SEJATI. *MULTILATERAL: Jurnal Pendidikan Jasmani dan Olahraga*, 87-98.
- Amah, N., Wahyuningsih, S., Deny, F., & Amijaya, T. (2017). Analisis *Cluster* Non-Hirarki Dengan Menggunakan Metode *K-Modes* pada Mahasiswa Program Studi Statistika Angkatan 2015 FMIPA Universitas Mulawarman Non-Hierarchical *Cluster* Analysis Using *K-Modes* Method. *Eksponensial*, 8, 9–16.
- Ani, R. A. (2012). PENINGKATAN KEMAMPUAN BELAJAR PENDAPATAN NASIONAL DENGAN KOMIK EKONOMI PADA SISWA KELAS X SMA N 1 MRANGGEN. *JURNAL PENDIDIKAN EKONOMI DINAMIKA PENDIDIKAN*, 67 - 78.
- Arizal, M. N., & Agus, S. (2015). FAKTOR-FAKTOR YANG MEMPENGARUHI PERILAKU KONSUMEN DALAM MEMBELI POLIS ASURANSI JIWA PADA PT. ASURANSI JIWASRAYA (Persero) PEKANBARU(Studi Kasus). *Jurnal Ilmiah Ekonomi dan Bisnis*, 157-165.
- Budget Your Trip. (2022, April 21). *India Travel Budget*. Retrieved from Budget Your Trip: <https://www.budgetyourtrip.com/india>
- Cao, F., Liang, J., Li, D., Bai, L., & Dang, C. (2012). A dissimilarity measure for the *K-Modes* clustering algorithm. *Knowledge-Based Systems*, 26, 120–127. <https://doi.org/10.1016/j.knosys.2011.07.011>
- Dewi, D., & Pramita, D. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix : Jurnal Manajemen Teknologi dan Informatika*, 102-109.
- Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient. *Journal of Applied Informatics and Computing (JAIC)*, 109-116.

- Fauzi, A., Saraswati, N. M., & Hariyono, R. C. (2020). PENERAPAN ALGORITMA *K-MODES* DAN C4.5 UNTUK PREDIKSI PEMILIHAN JURUSAN DI UNIVERSITAS PERADABAN PADA SISWA SMA (Studi Kasus: SMA Islam Ta'allumul Huda Bumiayu). *IJIR, 1*, 57-64.
- Global Data Lab. (2019). *Average Household Size India (Asia/Pacific)*. Global Data Lab.
- GOVERNMENT OF INDIA MINISTRY OF PERSONNEL, P.G. & PENSIONS. (2001, July 10). Sub: Holiday to be observed in Government Offices during the year 2002. *Holiday to be observed in Government Offices during the year 2002*. New Delhi.
- Huang, Z. (1998). Extensions to the k-Means Algorithm for *Clustering* Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery 2*, 283-304. *Data Mining and Knowledge Discovery, 2*(3), 283–304. https://www.researchgate.net/publication/220451944_Huang_Z_Extensions_to_the_k-Means_Algorithm_for_Clustering_Large_Data_Sets_with_Categorical_Values_Data_Mining_and_Knowledge_Discovery_2_283-304
- Hunt, J. (2019). *A Beginners Guide to Python 3 Programming*. New York City: Springer International Publishing.
- Kompas. (2021, Oktober 27). *Bisnis Asuransi Perjalanan Diprediksi Bakal Mulai Cerah*. Retrieved from <https://money.kompas.com/https://money.kompas.com/read/2021/10/27/203000826/bisnis-asuransi-perjalanan-diprediksi-bakal-mulai-cerah>
- Kontan. (2021, Agustus 30). *Ada pandemi, bisnis asuransi kesehatan terus melaju*. Retrieved from Kontan.co.id: <https://keuangan.kontan.co.id/news/ada-pandemi-bisnis-asuransi-kesehatan-terus-melaju>
- Lifepal. (2021, Mei 21). *Pengertian Asuransi Umum dan Bedanya dengan Asuransi Jiwa*. Retrieved from Lifepal: <https://lifepal.co.id/media/asuransi-umum/>
- Misra, U. (2021, April 22). *ExplainSpeaking: Why saving the middle class is critical for India's economy and its democracy*. Retrieved from The Indian Express: <https://indianexpress.com/article/explained/why-saving-the-middle-class-is-critical-for-indias-economy-and-its-democracy-7279620/>

- Nababan, C. N. (2014, Mei 20). *Masyarakat Indonesia sadar asuransi tapi ogah beli*. Retrieved from Keuangan Kontan: <https://keuangan.kontan.co.id/news/masyarakat-indonesia-sadar-asuransi-tapi-ogah-beli>
- NOVIANTO, R. (2019). Penerapan Data Mining menggunakan Algoritma K-Means *Clustering* untuk Menganalisa Bisnis Perusahaan Asuransi. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 6(1), 85–95. <https://doi.org/10.35957/jatisi.v6i1.150>
- OJK. (2022, Januari 18). *Asuransi*. Retrieved from Otoritas Jasa Keuangan: <https://www.ojk.go.id/id/kanal/iknb/pages/asuransi.aspx>
- OJK. (2022, Januari 18). *YANG HOBI JALAN-JALAN HARUS TAU ASURANSI PERJALANAN*. Retrieved from SIKAPI: <https://sikapiuangmu.ojk.go.id/FrontEnd/CMS/Article/20551>
- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi *Clustering* Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48. https://doi.org/10.51557/pt_jiit.v6i2.659
- Paycheck.in. (2021, April 2). *Annual Leave and Holidays*. Retrieved from Paycheck.in: <https://paycheck.in/labour-law-india/leave-and-holidays#:~:text=An%20adult%20worker%20is%20entitled,adult%20and%20young%20workers%20respectively.>
- PTI. (2016, September 9). *Young travellers driving Indian travel landscape, reveals Study*. Retrieved from Financial Express: [https://www.financialexpress.com/lifestyle/young-travellers-driving-indian-travel-landscape-reveals-study/372560/#:~:text=Reuters\)-,Travellers%20in%20the%20age%20group%20of%202018%2D35%20years%20are,travel%20landscape%20in%20the%20country.](https://www.financialexpress.com/lifestyle/young-travellers-driving-indian-travel-landscape-reveals-study/372560/#:~:text=Reuters)-,Travellers%20in%20the%20age%20group%20of%202018%2D35%20years%20are,travel%20landscape%20in%20the%20country.)
- Putri Kartikasari, Y. A., Agus Pranoto, Y., & Rudhistiar, D. (2021). Penerapan Metode *K-Modes* Untuk Proses Penentuan Penerima Bantuan Langsung Tunai (Blt). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(1), 389–397. <https://doi.org/10.36040/jati.v5i1.3300>
- Rabbi, C. P. (2021). *Belum Pulih dari Pandemi Covid-19, India Buka Taj Mahal bagi Turis*. India: katadata.

- Rahman, V. E. (2021). *Permintaan Asuransi Millennial India Meroket gegara COVID*. Jakarta: IDN Times.
- Silvia, V. (2020). *Statistika Deskriptif*. Sleman: Penerbit Andi.
- Siyoto, S. (2015). *Dasar Metodologi Penelitian*. Sleman: Literasi Media Publishing.
- Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: ANDI.
- Times Travel. (2019, Mei 2). *Indians choose domestic destinations over international ones this summer*. Retrieved from Times of India: <https://timesofindia.indiatimes.com/travel/destinations/indians-choose-domestic-destinations-over-international-ones-this-summer/as69143691.cms>
- Yuan, F., Yang, Y., & Yuan, T. (2020). A dissimilarity measure for mixed nominal and ordinal attribute data in *K-Modes* algorithm. *Applied Intelligence*, 50(5), 1498–1509. <https://doi.org/10.1007/s10489-019-01583-5>
- Yulianita, T., & Istiawan, D. (2017). Implementasi Algoritma *K-Modes* untuk Penentuan Prioritas Rehabilitasi Daerah Aliran Sungai Berdasarkan Parameter Lahan Kritis. The 6th University

LAMPIRAN

Lampiran 1 Data Penelitian

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravelInsurance
0	1	Government Sector	Yes	3	>4	1	No	No	0
1	1	Private Sector/Self Employed	Yes	4	>4	0	No	No	0
2	1	Private Sector/Self Employed	Yes	3	4	1	No	No	1
3	0	Private Sector/Self Employed	Yes	3	<4	1	No	No	0
4	0	Private Sector/Self Employed	Yes	3	>4	1	Yes	No	0
5	0	Private Sector/Self Employed	No	4	4	0	No	No	0
6	1	Government Sector	Yes	4	4	0	No	No	0
7	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	1
8	0	Private Sector/Self Employed	Yes	4	>4	1	Yes	Yes	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1978	1	Private Sector/Self Employed	Yes	3	>4	0	No	No	1
1979	0	Private Sector/Self Employed	Yes	4	4	1	No	No	0
1980	0	Private Sector/Self Employed	No	4	>4	1	No	No	0
1981	0	Government Sector	Yes	3	<4	0	No	No	1
1982	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	1
1983	0	Private Sector/Self Employed	Yes	4	>4	1	No	Yes	0
1984	0	Private Sector/Self Employed	Yes	4	>4	1	No	No	0
1985	1	Private Sector/Self Employed	Yes	4	>4	0	Yes	Yes	1
1986	1	Private Sector/Self Employed	Yes	3	4	0	No	No	0

Lampiran 2 Sintaks Perhitungan Jarak

```
# Load Initial Module
import pandas as pd
import numpy as np
import random
from sklearn.metrics.pairwise import cosine_similarity

# Input Data
travel = pd.read_csv('TravelInsurancePrediction.csv',
sep=";")
travel.head()
data= travel[['Age', 'Employment
Type', 'GraduateOrNot', 'FamilyMembers', 'AnnualIncome', 'ChronicDi
seases', 'FrequentFlyer', 'EverTravelledAbroad']]
data

#Preprocessing Data
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
data = data.apply(le.fit_transform)
data.head()

# Inisialisasi Variable Ordinal dan Nominal

ordi = []
banyak_ordi = int(input('Banyak Variable Ordinal : '))

for i in range (banyak_ordi) :
    x = input(f'Masukan Variable Yang Berskala Ordinal Ke-
{i+1} : ')
    ordi.append(x)

data_nom = data.drop(ordi, axis=1)
data_ordi = data[ordi]
```

```

# Input Jumlah Cluster

init_cluster = int(input('Masukan Jumlah Cluster : '))
print (f'Cluster Yang Ingin Dibuat Sebanyak {init_cluster}
Cluster')

# Perhitungan Nominal
def nominal (cluster) :
    new_val = []

    for i in data_nom.columns :
        pembilang = 0
        penyebut = 0

        for j in range(len(data_nom[i])) :
            if data[i][j] == data[i][cluster]:
                pembilang = 1
            else :
                pembilang = 0

        for k in range(len(data_nom[i])) :
            if (k == j) | (k == cluster) :
                continue
            elif data_nom[i][k] == data_nom[i][j] :
                penyebut += 1
        try :
            z = 1 - (pembilang/penyebut)
            new_val.append(z)
            pembilang = 0
            penyebut = 0

        except ZeroDivisionError :
            z = 1 - (pembilang/1)
            new_val.append(z)
            pembilang = 0
            penyebut = 0
    baru = np.array(new_val)
    new_nomi = pd.DataFrame (baru.reshape (data_nom.shape[1],

```

```

        data_nom.shape[0]).T, columns = data_nom.columns)
        new_nomi.iloc[cluster,:] = 0
        new_nomi['Jumlah'] = new_nomi.sum(axis=1)
        return new_nomi

# Hasil Jarak Nominal
daftar_cluster = []

for i in range(init_cluster) :
    exec(f'daftar_cluster.append("cluster_{i+1}")')

for i in daftar_cluster :
    exec(f'jarak_nominal_{i} = nominal({i})')

# Fungsi Perhitungan Variable Ordinal (Syarat 1)
ordil= data_ordi.drop(['AnnualIncome'], axis=1)
ordil
def syarat1 (cluster) :
    new_vall = []
    for i in ordil.columns :
        pembilang = 0
        penyebut = 0

        for j in range(len(ordil[i])) :
            if data[i][j] == data[i][cluster]:
                pembilang = 1
            else :
                pembilang = 0

        for k in range(len(ordil[i])) :
            if (k == j) | (k == cluster) :
                continue
            elif ordil[i][k] == ordil[i][j] :
                penyebut += 1

    try :
        z = 1 - (pembilang/penyebut)
        new_vall.append(z)

```

```

        pembilang = 0
        penyebut = 0

    except ZeroDivisionError :
        z = 1 - (pembilang/1)
        new_val1.append(z)
        pembilang = 0
        penyebut = 0
    baru1 = np.array(new_val1)
    new_ordil = pd.DataFrame(baru1.reshape(ordil.shape[1],
ordil.shape[0]).T, columns = ordil.columns)
    new_ordil.iloc[cluster,:] = 0
    new_ordil['Jumlah'] = new_ordil.sum(axis=1)
    return new_ordil

# Hasil Jarak Ordinal ke-1
daftar_cluster = []

for i in range(init_cluster) :
    exec(f'daftar_cluster.append("cluster_{i+1}")')

for i in daftar_cluster :
    exec(f'jarak_ordinal1_{i} = syarat1({i})')

##Syarat Kedua Variabel Odinal Kedua

# Input Ukuran Domain

Nama_Kolom = data.columns
ukuran_domain = []

for i in range(len(data.columns)) :
    ud = int(input(f'Ukuran Domain {Nama_Kolom[i]} : '))
    ukuran_domain.append(ud)

UK_DO = pd.DataFrame({
    'Variable' : Nama_Kolom,
    'Ukuran Domain' : ukuran_domain

```

```

    })
    UK_DO['Selisih'] = abs(UK_DO['Ukuran Domain'] -
UK_DO[UK_DO['Variable']] == ordi[1]['Ukuran
Domain'][UK_DO[UK_DO['Variable']] == ordi[1]].index[0])
    UK_DO.drop(UK_DO[UK_DO['Variable']] == ordi[0]].index[0],
inplace=True)
    print ('Ukuran Domain Variabel Ordinal Annual Income'.upper())
    display(UK_DO)

    MSOA = np.min(UK_DO['Selisih'])
    print ('MSOA')
    display(MSOA)

    ind_MSOA = UK_DO[(UK_DO['Selisih'] == MSOA) &
(UK_DO['Variable'] != ordi[1])]['Variable']
    print('Indeks MSOA')
    ind_MSOA

    # Hitung Jarak Antar Data AnnualIncome

    def syarat2 (cluster) :
        new_val = []

        for i in ind_MSOA :
            pembilang = 0
            penyebut = 0

            for j in range(len(data[i])) :
                if data[i].index[j] == cluster :
                    pembilang == 0
                else :
                    if data[i][j] == data[i][cluster]:
                        pembilang = 1
                    else :
                        pembilang = 0
            for k in range(len(data[i])) :
                if (k == j) | (k == cluster) :
                    continue

```

```

        elif data[i][k] == data[i][j] :
            penyebut += 1

    try :
        z = (pembilang/penyebut)
        new_val.append(z)
        pembilang = 0
        penyebut = 0
    except ZeroDivisionError :
        z = (pembilang/1)
        new_val.append(z)
        pembilang = 0
        penyebut = 0

    baru = np.array(new_val)
    new_ordi = pd.DataFrame({
        ind_MSOA[3] : baru
    })
    return new_ordi

sh1 = pd.concat([jarak_ordinal_20, jarak_ordinal_21,
jarak_ordinal_22, jarak_ordinal_23, jarak_ordinal_24,
jarak_ordinal_25,
jarak_ordinal_26, jarak_ordinal_27, jarak_ordinal_28, jarak_ordinal_29,
jarak_ordinal_210, jarak_ordinal_211,
jarak_ordinal_212, jarak_ordinal_213, jarak_ordinal_214,
jarak_ordinal_215,
jarak_ordinal_216, jarak_ordinal_217, jarak_ordinal_218, jarak_ordinal_219,
jarak_ordinal_220, jarak_ordinal_221,
jarak_ordinal_222, jarak_ordinal_223, jarak_ordinal_224,
jarak_ordinal_225,
jarak_ordinal_226, jarak_ordinal_227, jarak_ordinal_228, jarak_ordinal_229,
jarak_ordinal_230, jarak_ordinal_231,
jarak_ordinal_232, jarak_ordinal_233, jarak_ordinal_234,
jarak_ordinal_235,
jarak_ordinal_236, jarak_ordinal_237, jarak_ordinal_238, jarak_ordinal_239,
jarak_ordinal_240, jarak_ordinal_241, jarak_ordinal_242, jarak_ordinal_243,
jarak_ordinal_244, jarak_ordinal_245, jarak_ordinal_246, jarak_ordinal_247,
jarak_ordinal_248, jarak_ordinal_249, jarak_ordinal_250, jarak_ordinal_251,
jarak_ordinal_252, jarak_ordinal_253, jarak_ordinal_254, jarak_ordinal_255,
jarak_ordinal_256, jarak_ordinal_257, jarak_ordinal_258, jarak_ordinal_259,
jarak_ordinal_260, jarak_ordinal_261, jarak_ordinal_262, jarak_ordinal_263,
jarak_ordinal_264, jarak_ordinal_265, jarak_ordinal_266, jarak_ordinal_267,
jarak_ordinal_268, jarak_ordinal_269, jarak_ordinal_270, jarak_ordinal_271,
jarak_ordinal_272, jarak_ordinal_273, jarak_ordinal_274, jarak_ordinal_275,
jarak_ordinal_276, jarak_ordinal_277, jarak_ordinal_278, jarak_ordinal_279,
jarak_ordinal_280, jarak_ordinal_281, jarak_ordinal_282, jarak_ordinal_283,
jarak_ordinal_284, jarak_ordinal_285, jarak_ordinal_286, jarak_ordinal_287,
jarak_ordinal_288, jarak_ordinal_289, jarak_ordinal_290, jarak_ordinal_291,
jarak_ordinal_292, jarak_ordinal_293, jarak_ordinal_294, jarak_ordinal_295,
jarak_ordinal_296, jarak_ordinal_297, jarak_ordinal_298, jarak_ordinal_299,
jarak_ordinal_300])

```


inal_239,
 jarak_ordinal_240, jarak_ordinal_241,
jarak_ordinal_242, jarak_ordinal_243, jarak_ordinal_244,
jarak_ordinal_245,
jarak_ordinal_246,jarak_ordinal_247,jarak_ordinal_248,jarak_ordinal_249,
 jarak_ordinal_250, jarak_ordinal_251,
jarak_ordinal_252, jarak_ordinal_253, jarak_ordinal_254,
jarak_ordinal_255,
jarak_ordinal_256,jarak_ordinal_257,jarak_ordinal_258,jarak_ordinal_259,
 jarak_ordinal_260, jarak_ordinal_261,
jarak_ordinal_262, jarak_ordinal_263, jarak_ordinal_264,
jarak_ordinal_265,
jarak_ordinal_266,jarak_ordinal_267,jarak_ordinal_268,jarak_ordinal_269,
 jarak_ordinal_270, jarak_ordinal_271,
jarak_ordinal_272, jarak_ordinal_273, jarak_ordinal_274,
jarak_ordinal_275,
jarak_ordinal_276,jarak_ordinal_277,jarak_ordinal_278,jarak_ordinal_279,
 jarak_ordinal_280, jarak_ordinal_281,
jarak_ordinal_282, jarak_ordinal_283, jarak_ordinal_284,
jarak_ordinal_285,
jarak_ordinal_286,jarak_ordinal_287,jarak_ordinal_288,jarak_ordinal_289,
 jarak_ordinal_290, jarak_ordinal_291,
jarak_ordinal_292, jarak_ordinal_293, jarak_ordinal_294,
jarak_ordinal_295,
jarak_ordinal_296,jarak_ordinal_297,jarak_ordinal_298,jarak_ordinal_299,
 jarak_ordinal_2100, jarak_ordinal_2101,
jarak_ordinal_2102, jarak_ordinal_2103, jarak_ordinal_2104,
jarak_ordinal_2105,
jarak_ordinal_2106,jarak_ordinal_2107,jarak_ordinal_2108,jarak_ordinal_2109,
 jarak_ordinal_2110, jarak_ordinal_2111,
jarak_ordinal_2112, jarak_ordinal_2113, jarak_ordinal_2114,
jarak_ordinal_2115,

jarak_ordinal_2116, jarak_ordinal_2117, jarak_ordinal_2118, jarak_ordinal_2119,
jarak_ordinal_2120, jarak_ordinal_2121,
jarak_ordinal_2122, jarak_ordinal_2123, jarak_ordinal_2124,
jarak_ordinal_2125,
jarak_ordinal_2126, jarak_ordinal_2127, jarak_ordinal_2128, jarak_ordinal_2129,
jarak_ordinal_2130, jarak_ordinal_2131,
jarak_ordinal_2132, jarak_ordinal_2133, jarak_ordinal_2134,
jarak_ordinal_2135,
jarak_ordinal_2136, jarak_ordinal_2137, jarak_ordinal_2138, jarak_ordinal_2139,
jarak_ordinal_2140, jarak_ordinal_2141,
jarak_ordinal_2142, jarak_ordinal_2143, jarak_ordinal_2144,
jarak_ordinal_2145,
jarak_ordinal_2146, jarak_ordinal_2147, jarak_ordinal_2148, jarak_ordinal_2149,
jarak_ordinal_2150, jarak_ordinal_2151,
jarak_ordinal_2152, jarak_ordinal_2153, jarak_ordinal_2154,
jarak_ordinal_2155,
jarak_ordinal_2156, jarak_ordinal_2157, jarak_ordinal_2158, jarak_ordinal_2159,
jarak_ordinal_2160, jarak_ordinal_2161,
jarak_ordinal_2162, jarak_ordinal_2163, jarak_ordinal_2164,
jarak_ordinal_2165,
jarak_ordinal_2166, jarak_ordinal_2167, jarak_ordinal_2168, jarak_ordinal_2169,
jarak_ordinal_2170, jarak_ordinal_2171,
jarak_ordinal_2172, jarak_ordinal_2173, jarak_ordinal_2174,
jarak_ordinal_2175,
jarak_ordinal_2176, jarak_ordinal_2177, jarak_ordinal_2178, jarak_ordinal_2179,
jarak_ordinal_2180, jarak_ordinal_2181,
jarak_ordinal_2182, jarak_ordinal_2183, jarak_ordinal_2184,
jarak_ordinal_2185,
jarak_ordinal_2186, jarak_ordinal_2187, jarak_ordinal_2188, jarak_ordinal_2189,
jarak_ordinal_2190, jarak_ordinal_2191,
jarak_ordinal_2192, jarak_ordinal_2193, jarak_ordinal_2194,

jarak_ordinal_21195,
jarak_ordinal_21196, jarak_ordinal_21197, jarak_ordinal_21198, jarak_ordinal_21199,
jarak_ordinal_21200, jarak_ordinal_21201,
jarak_ordinal_21202, jarak_ordinal_21203, jarak_ordinal_21204,
jarak_ordinal_21205,
jarak_ordinal_21206, jarak_ordinal_21207, jarak_ordinal_21208, jarak_ordinal_21209,
jarak_ordinal_21210, jarak_ordinal_21211,
jarak_ordinal_21212, jarak_ordinal_21213, jarak_ordinal_21214,
jarak_ordinal_21215,
jarak_ordinal_21216, jarak_ordinal_21217, jarak_ordinal_21218, jarak_ordinal_21219,
jarak_ordinal_21220, jarak_ordinal_21221,
jarak_ordinal_21222, jarak_ordinal_21223, jarak_ordinal_21224,
jarak_ordinal_21225,
jarak_ordinal_21226, jarak_ordinal_21227, jarak_ordinal_21228, jarak_ordinal_21229,
jarak_ordinal_21230, jarak_ordinal_21231,
jarak_ordinal_21232, jarak_ordinal_21233, jarak_ordinal_21234,
jarak_ordinal_21235,
jarak_ordinal_21236, jarak_ordinal_21237, jarak_ordinal_21238, jarak_ordinal_21239,
jarak_ordinal_21240, jarak_ordinal_21241,
jarak_ordinal_21242, jarak_ordinal_21243, jarak_ordinal_21244,
jarak_ordinal_21245,
jarak_ordinal_21246, jarak_ordinal_21247, jarak_ordinal_21248, jarak_ordinal_21249,
jarak_ordinal_21250, jarak_ordinal_21251,
jarak_ordinal_21252, jarak_ordinal_21253, jarak_ordinal_21254,
jarak_ordinal_21255,
jarak_ordinal_21256, jarak_ordinal_21257, jarak_ordinal_21258, jarak_ordinal_21259,
jarak_ordinal_21260, jarak_ordinal_21261,
jarak_ordinal_21262, jarak_ordinal_21263, jarak_ordinal_21264,
jarak_ordinal_21265,
jarak_ordinal_21266, jarak_ordinal_21267, jarak_ordinal_21268, jarak_ordinal_21269,
jarak_ordinal_21270, jarak_ordinal_21271,

```
    jarak_ordinal_21272,                jarak_ordinal_21273,  
    jarak_ordinal_21274, jarak_ordinal_21275, jarak_ordinal_21276],  
axis=1)  
  
# Nilai Minimum Selain 0 atau dot1  
  
sh1[sh1>0].min().min()  
  
ordi_cluster_i      =      abs((data[ordi[1]]      -  
data[ordi[1]][cluster_i])*sh1[sh1>0].min().min())
```

Lampiran 3 Sintaks *K-Modes Clustering*

```
# Load Initial Module
import pandas as pd
import numpy as np
import random
from sklearn.metrics.pairwise import cosine_similarity

# Input Data
travel = pd.read_csv('TravelInsurancePrediction.csv',
sep=";")
travel.head()
data= travel[['Age', 'Employment
Type', 'GraduateOrNot', 'FamilyMembers', 'AnnualIncome', 'ChronicDi
seases', 'FrequentFlyer', 'EverTravelledAbroad']]
data

#Preprocessing Data
from sklearn import preprocessing
le = preprocessing.LabelEncoder()
data = data.apply(le.fit_transform)
data.head()

# Inisialisasi Variable Ordinal dan Nominal

ordi = []
banyak_ordi = int(input('Banyak Variable Ordinal : '))

for i in range (banyak_ordi) :
    x = input(f'Masukan Variable Yang Berskala Ordinal Ke-
{i+1} : ')
    ordi.append(x)

data_nom = data.drop(ordi, axis=1)
data_ordi = data[ordi]
```

```

# Jumlah Cluster

init_cluster = int(input('Masukan Jumlah Cluster : '))
print (f'Cluster Yang Ingin Dibuat Sebanyak {init_cluster}
Cluster')

# Menentukan Initial Cluster

for i in range(init_cluster) :
    exec(f'cluster_{i+1} = random.randint(0, (len(data))-
1)')

cluster_1
cluster_2

jarakgabungan = pd.read_csv('cluster2_1.csv', sep=";")
jarakgabungan.head()

#Penempatan Cluster baru
data['Label'] = jarakgabungan.idxmin(axis=1)
data

cluster_1 = data[data['Label'] == 'Terhadap Cluster
1'].drop('Label', axis = 1).mode().dropna()
cluster_1

cluster_2 = data[data['Label'] == 'Terhadap Cluster
2'].drop('Label', axis = 1).mode().dropna()
cluster_2

datanya = pd.read_csv("datanya.csv", sep=";")
data = datanya.drop("Unnamed: 0", axis=1)
data

# Variable Ordinal dan Nominal

ordi = []
banyak_ordi = int(input('Banyak Variable Ordinal : '))

```

```

for i in range (banyak_ordi) :
    x = input(f'Masukan Variable Yang Berskala Ordinal Ke-
{i+1} : ')
    ordi.append(x)

data_nom = data.drop(ordi, axis=1)
data_ordi = data[ordi]

# Jumlah Cluster

init_cluster = int(input('Masukan Jumlah Cluster : '))
print (f'Cluster Yang Ingin Dibuat Sebanyak {init_cluster}
Cluster')

cluster_1 = 1277
cluster_2 = 1278

# Perhitungan Nominal
def nominal (cluster) :
    new_val = []

    for i in data_nom.columns :
        pembilang = 0
        penyebut = 0

        for j in range(len(data_nom[i])) :
            if data[i][j] == data[i][cluster]:
                pembilang = 1
            else:
                pembilang = 0

            for k in range(len(data_nom[i])) :
                if (k == j) | (k == cluster) :
                    continue
                elif data_nom[i][k] == data_nom[i][j] :
                    penyebut += 1

        try :
            z = 1 - (pembilang/penyebut)

```

```

        new_val.append(z)
        pembilang = 0
        penyebut = 0

    except ZeroDivisionError :
        z = 1 - (pembilang/1)
        new_val.append(z)
        pembilang = 0
        penyebut = 0
    baru = np.array(new_val)
    new_nomi = pd.DataFrame(baru.reshape(data_nom.shape[1],
data_nom.shape[0]).T, columns = data_nom.columns)
    new_nomi.iloc[cluster,:] = 0
    new_nomi['Jumlah'] = new_nomi.sum(axis=1)
    return new_nomi

# Hasil Jarak Nominal
daftar_cluster = []

for i in range(init_cluster) :
    exec(f'daftar_cluster.append("cluster_{i+1}")')

for i in daftar_cluster :
    exec(f'jarak_nominal_{i} = nominal({i})')

# Fungsi Perhitungan Variable Ordinal (Syarat 1)
ordil= data_ordi.drop(['AnnualIncome'], axis=1)
ordil
def syarat1 (cluster) :
    new_val1 = []
    for i in ordil.columns :
        pembilang = 0
        penyebut = 0

        for j in range(len(ordil[i])) :
            if data[i][j] == data[i][cluster]:
                pembilang = 1

```



```

        else :
            pembilang = 0

        for k in range(len(ordil[i])) :
            if (k == j) | (k == cluster) :
                continue
            elif ordil[i][k] == ordil[i][j] :
                penyebut += 1

        try :
            z = 1 - (pembilang/penyebut)
            new_vall.append(z)
            pembilang = 0
            penyebut = 0

        except ZeroDivisionError :
            z = 1 - (pembilang/1)
            new_vall.append(z)
            pembilang = 0
            penyebut = 0

        barul = np.array(new_vall)
        new_ordil = pd.DataFrame(barul.reshape(ordil.shape[1],
ordil.shape[0]).T, columns = ordil.columns)
        new_ordil.iloc[cluster,:] = 0
        new_ordil['Jumlah'] = new_ordil.sum(axis=1)
        return new_ordil

# Hasil Jarak Ordinal ke-1
daftar_cluster = []

for i in range(init_cluster) :
    exec(f'daftar_cluster.append("cluster_{i+1}")')

for i in daftar_cluster :
    exec(f'jarak_ordinal1_{i} = syarat1({i})')

```

```

sh = pd.read_csv('msoa.csv', sep=";")
sh1= sh.drop("Unnamed: 0", axis=1)
sh1

# Jarak Ordinal ke-2

AI_cluster_1      =      abs((data[ordi[1]]      -
data[ordi[1]][cluster_1])*sh1[sh1>0].min().min())
AI_cluster_2      =      abs((data[ordi[1]]      -
data[ordi[1]][cluster_2])*sh1[sh1>0].min().min())

# Jarak Gabungan
jarakgabungan_cluster_1 = jarak_nominal_cluster_1['Jumlah']
+ jarak_ordinal1_cluster_1['Jumlah']+AI_cluster_1
jarakgabungan_cluster_2 = jarak_nominal_cluster_2['Jumlah']
+ jarak_ordinal1_cluster_2['Jumlah']+AI_cluster_2
jarakgabungan1 = pd.DataFrame()
jarakgabungan1['Terhadap      Cluster      1']      =
jarakgabungan_cluster_1
jarakgabungan1['Terhadap      Cluster      2']      =
jarakgabungan_cluster_2

jarakgabungan=jarakgabungan1.drop([1277,1278])
jarakgabungan

jarakgabungan.to_csv("jarakgabungancluster2_1.csv")

#Penempatan Cluster baru
data['Label'] = jarakgabungan.idxmin(axis=1)

data.dropna()

data[data['Label'] == 'Terhadap Cluster 1'].drop('Label',
axis = 1).mode().dropna()
data[data['Label'] == 'Terhadap Cluster 2'].drop('Label',
axis = 1).mode().dropna()
data.dropna().to_csv("hasilclusterdua_1.csv")

```

Lampiran 4 Sintaks Perhitungan *Silhouette Score*

```
sil=pd.read_csv("jarak.csv",sep=";")
sil
label = sil["Label"]
from sklearn.metrics import silhouette_score
silhouette_score(sil.drop("Label",axis=1), label,
metric='precomputed')
```

Lampiran 5 Hasil Pengelompokkan 2 *Cluster* Data Tidak Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Government Sector	Yes	3	>4	1	No	No	Terhadap <i>Cluster</i> 2
1	1	Private Sector/Self Employed	Yes	4	>4	0	No	No	Terhadap <i>Cluster</i> 1
2	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap <i>Cluster</i> 1
3	0	Private Sector/Self Employed	Yes	3	<4	1	No	No	Terhadap <i>Cluster</i> 1
4	0	Private Sector/Self Employed	Yes	3	>4	1	Yes	No	Terhadap <i>Cluster</i> 1
5	0	Private Sector/Self Employed	No	4	4	0	No	No	Terhadap <i>Cluster</i> 2
6	1	Government Sector	Yes	4	4	0	No	No	Terhadap <i>Cluster</i> 2
7	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap <i>Cluster</i> 2
8	0	Private Sector/Self Employed	Yes	4	>4	1	Yes	Yes	Terhadap <i>Cluster</i> 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1268	1	Private Sector/Self Employed	Yes	4	4	0	No	No	Terhadap <i>Cluster</i> 2
1269	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap <i>Cluster</i> 1
1270	0	Private Sector/Self Employed	Yes	4	>4	0	Yes	No	Terhadap <i>Cluster</i> 1
1271	1	Government Sector	Yes	4	4	0	No	No	Terhadap <i>Cluster</i> 2
1272	0	Private Sector/Self Employed	Yes	3	>4	1	No	No	Terhadap <i>Cluster</i> 1
1273	0	Private Sector/Self Employed	No	4	>4	1	No	No	Terhadap <i>Cluster</i> 1
1274	0	Private Sector/Self Employed	Yes	4	>4	1	No	Yes	Terhadap <i>Cluster</i> 1
1275	0	Private Sector/Self Employed	Yes	4	>4	1	No	No	Terhadap <i>Cluster</i> 1
1276	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap <i>Cluster</i> 2

Lampiran 6 Hasil Pengelompokan 2 *Cluster* Data Dengan Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap <i>Cluster</i> 1
1	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap <i>Cluster</i> 2
2	0	Private Sector/Self Employed	Yes	4	4	1	Yes	Yes	Terhadap <i>Cluster</i> 2
3	0	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap <i>Cluster</i> 2
4	1	Government Sector	Yes	3	4	0	No	No	Terhadap <i>Cluster</i> 1
5	1	Government Sector	Yes	4	4	0	Yes	Yes	Terhadap <i>Cluster</i> 2
6	0	Private Sector/Self Employed	Yes	3	>4	0	No	No	Terhadap <i>Cluster</i> 1
7	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap <i>Cluster</i> 1
8	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap <i>Cluster</i> 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
701	0	Private Sector/Self Employed	No	3	<4	0	No	No	Terhadap <i>Cluster</i> 1
702	0	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap <i>Cluster</i> 1
703	1	Private Sector/Self Employed	Yes	4	<4	0	No	Yes	Terhadap <i>Cluster</i> 2
704	0	Private Sector/Self Employed	Yes	4	>4	0	No	Yes	Terhadap <i>Cluster</i> 2
705	0	Private Sector/Self Employed	No	4	4	0	No	Yes	Terhadap <i>Cluster</i> 2
706	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap <i>Cluster</i> 1
707	0	Government Sector	Yes	3	<4	0	No	No	Terhadap <i>Cluster</i> 1
708	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap <i>Cluster</i> 2
709	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap <i>Cluster</i> 2

Lampiran 7 Hasil Pengelompokan 3 Cluster Data Tidak Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Government Sector	Yes	3	>4	1	No	No	Terhadap Cluster 2
1	1	Private Sector/Self Employed	Yes	4	>4	0	No	No	Terhadap Cluster 2
2	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 2
3	0	Private Sector/Self Employed	Yes	3	<4	1	No	No	Terhadap Cluster 1
4	0	Private Sector/Self Employed	Yes	3	>4	1	Yes	No	Terhadap Cluster 3
5	0	Private Sector/Self Employed	No	4	4	0	No	No	Terhadap Cluster 3
6	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 1
7	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 2
8	0	Private Sector/Self Employed	Yes	4	>4	1	Yes	Yes	Terhadap Cluster 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1268	1	Private Sector/Self Employed	Yes	4	4	0	No	No	Terhadap Cluster 2
1269	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 2
1270	0	Private Sector/Self Employed	Yes	4	>4	0	Yes	No	Terhadap Cluster 1
1271	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 2
1272	0	Private Sector/Self Employed	Yes	3	>4	1	No	No	Terhadap Cluster 3
1273	0	Private Sector/Self Employed	No	4	>4	1	No	No	Terhadap Cluster 2
1274	0	Private Sector/Self Employed	Yes	4	>4	1	No	Yes	Terhadap Cluster 2
1275	0	Private Sector/Self Employed	Yes	4	>4	1	No	No	Terhadap Cluster 2
1276	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 3

Lampiran 8 Hasil Pengelompokan 3 Cluster Data Dengan Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 3
1	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 2
2	0	Private Sector/Self Employed	Yes	4	4	1	Yes	Yes	Terhadap Cluster 2
3	0	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 2
4	1	Government Sector	Yes	3	4	0	No	No	Terhadap Cluster 1
5	1	Government Sector	Yes	4	4	0	Yes	Yes	Terhadap Cluster 2
6	0	Private Sector/Self Employed	Yes	3	>4	0	No	No	Terhadap Cluster 1
7	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 3
8	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
701	0	Private Sector/Self Employed	No	3	<4	0	No	No	Terhadap Cluster 1
702	0	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 1
703	1	Private Sector/Self Employed	Yes	4	<4	0	No	Yes	Terhadap Cluster 2
704	0	Private Sector/Self Employed	Yes	4	>4	0	No	Yes	Terhadap Cluster 2
705	0	Private Sector/Self Employed	No	4	4	0	No	Yes	Terhadap Cluster 2
706	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 1
707	0	Government Sector	Yes	3	<4	0	No	No	Terhadap Cluster 1
708	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 2
709	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 2

Lampiran 9 Hasil Pengelompokan 4 Cluster Data Tidak Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Government Sector	Yes	3	>4	1	No	No	Terhadap Cluster 3
1	1	Private Sector/Self Employed	Yes	4	>4	0	No	No	Terhadap Cluster 1
2	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 3
3	0	Private Sector/Self Employed	Yes	3	<4	1	No	No	Terhadap Cluster 3
4	0	Private Sector/Self Employed	Yes	3	>4	1	Yes	No	Terhadap Cluster 2
5	0	Private Sector/Self Employed	No	4	4	0	No	No	Terhadap Cluster 4
6	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 1
7	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 3
8	0	Private Sector/Self Employed	Yes	4	>4	1	Yes	Yes	Terhadap Cluster 1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1268	1	Private Sector/Self Employed	Yes	4	4	0	No	No	Terhadap Cluster 1
1269	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 3
1270	0	Private Sector/Self Employed	Yes	4	>4	0	Yes	No	Terhadap Cluster 1
1271	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 1
1272	0	Private Sector/Self Employed	Yes	3	>4	1	No	No	Terhadap Cluster 4
1273	0	Private Sector/Self Employed	No	4	>4	1	No	No	Terhadap Cluster 2
1274	0	Private Sector/Self Employed	Yes	4	>4	1	No	Yes	Terhadap Cluster 3
1275	0	Private Sector/Self Employed	Yes	4	>4	1	No	No	Terhadap Cluster 3
1276	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 4

Lampiran 10 Hasil Pengelompokan 4 Cluster Data Dengan Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 3
1	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 1
2	0	Private Sector/Self Employed	Yes	4	4	1	Yes	Yes	Terhadap Cluster 1
3	0	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 1
4	1	Government Sector	Yes	3	4	0	No	No	Terhadap Cluster 3
5	1	Government Sector	Yes	4	4	0	Yes	Yes	Terhadap Cluster 3
6	0	Private Sector/Self Employed	Yes	3	>4	0	No	No	Terhadap Cluster 4
7	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 2
8	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
701	0	Private Sector/Self Employed	No	3	<4	0	No	No	Terhadap Cluster 2
702	0	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 2
703	1	Private Sector/Self Employed	Yes	4	<4	0	No	Yes	Terhadap Cluster 1
704	0	Private Sector/Self Employed	Yes	4	>4	0	No	Yes	Terhadap Cluster 4
705	0	Private Sector/Self Employed	No	4	4	0	No	Yes	Terhadap Cluster 1
706	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 3
707	0	Government Sector	Yes	3	<4	0	No	No	Terhadap Cluster 3
708	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 1
709	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 1

Lampiran 11 Hasil Pengelompokan 5 Cluster Data Tidak Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Government Sector	Yes	3	>4	1	No	No	Terhadap Cluster 1
1	1	Private Sector/Self Employed	Yes	4	>4	0	No	No	Terhadap Cluster 1
2	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 3
3	0	Private Sector/Self Employed	Yes	3	<4	1	No	No	Terhadap Cluster 5
4	0	Private Sector/Self Employed	Yes	3	>4	1	Yes	No	Terhadap Cluster 2
5	0	Private Sector/Self Employed	No	4	4	0	No	No	Terhadap Cluster 1
6	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 5
7	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 1
8	0	Private Sector/Self Employed	Yes	4	>4	1	Yes	Yes	Terhadap Cluster 1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1268	1	Private Sector/Self Employed	Yes	4	4	0	No	No	Terhadap Cluster 3
1269	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 3
1270	0	Private Sector/Self Employed	Yes	4	>4	0	Yes	No	Terhadap Cluster 5
1271	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 1
1272	0	Private Sector/Self Employed	Yes	3	>4	1	No	No	Terhadap Cluster 4
1273	0	Private Sector/Self Employed	No	4	>4	1	No	No	Terhadap Cluster 2
1274	0	Private Sector/Self Employed	Yes	4	>4	1	No	Yes	Terhadap Cluster 3
1275	0	Private Sector/Self Employed	Yes	4	>4	1	No	No	Terhadap Cluster 3
1276	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 4

Lampiran 12 Hasil Pengelompokan 5 Cluster Data Dengan Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 5
1	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 2
2	0	Private Sector/Self Employed	Yes	4	4	1	Yes	Yes	Terhadap Cluster 5
3	0	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 5
4	1	Government Sector	Yes	3	4	0	No	No	Terhadap Cluster 1
5	1	Government Sector	Yes	4	4	0	Yes	Yes	Terhadap Cluster 5
6	0	Private Sector/Self Employed	Yes	3	>4	0	No	No	Terhadap Cluster 1
7	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 5
8	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
701	0	Private Sector/Self Employed	No	3	<4	0	No	No	Terhadap Cluster 3
702	0	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 2
703	1	Private Sector/Self Employed	Yes	4	<4	0	No	Yes	Terhadap Cluster 4
704	0	Private Sector/Self Employed	Yes	4	>4	0	No	Yes	Terhadap Cluster 2
705	0	Private Sector/Self Employed	No	4	4	0	No	Yes	Terhadap Cluster 1
706	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 2
707	0	Government Sector	Yes	3	<4	0	No	No	Terhadap Cluster 5
708	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 5
709	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 1

Lampiran 13 Hasil Pengelompokan 6 Cluster Data Tidak Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Government Sector	Yes	3	>4	1	No	No	Terhadap Cluster 5
1	1	Private Sector/Self Employed	Yes	4	>4	0	No	No	Terhadap Cluster 6
2	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 2
3	0	Private Sector/Self Employed	Yes	3	<4	1	No	No	Terhadap Cluster 4
4	0	Private Sector/Self Employed	Yes	3	>4	1	Yes	No	Terhadap Cluster 1
5	0	Private Sector/Self Employed	No	4	4	0	No	No	Terhadap Cluster 5
6	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 4
7	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 5
8	0	Private Sector/Self Employed	Yes	4	>4	1	Yes	Yes	Terhadap Cluster 5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1268	1	Private Sector/Self Employed	Yes	4	4	0	No	No	Terhadap Cluster 1
1269	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 2
1270	0	Private Sector/Self Employed	Yes	4	>4	0	Yes	No	Terhadap Cluster 4
1271	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 5
1272	0	Private Sector/Self Employed	Yes	3	>4	1	No	No	Terhadap Cluster 2
1273	0	Private Sector/Self Employed	No	4	>4	1	No	No	Terhadap Cluster 1
1274	0	Private Sector/Self Employed	Yes	4	>4	1	No	Yes	Terhadap Cluster 2
1275	0	Private Sector/Self Employed	Yes	4	>4	1	No	No	Terhadap Cluster 2
1276	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 1

Lampiran 14 Hasil Pengelompokan 6 Cluster Data Dengan Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 5
1	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 4
2	0	Private Sector/Self Employed	Yes	4	4	1	Yes	Yes	Terhadap Cluster 4
3	0	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 4
4	1	Government Sector	Yes	3	4	0	No	No	Terhadap Cluster 1
5	1	Government Sector	Yes	4	4	0	Yes	Yes	Terhadap Cluster 4
6	0	Private Sector/Self Employed	Yes	3	>4	0	No	No	Terhadap Cluster 6
7	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 2
8	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
701	0	Private Sector/Self Employed	No	3	<4	0	No	No	Terhadap Cluster 6
702	0	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 6
703	1	Private Sector/Self Employed	Yes	4	<4	0	No	Yes	Terhadap Cluster 5
704	0	Private Sector/Self Employed	Yes	4	>4	0	No	Yes	Terhadap Cluster 4
705	0	Private Sector/Self Employed	No	4	4	0	No	Yes	Terhadap Cluster 4
706	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 1
707	0	Government Sector	Yes	3	<4	0	No	No	Terhadap Cluster 6
708	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 4
709	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 4

Lampiran 15 Hasil Pengelompokan 7 Cluster Data Tidak Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Government Sector	Yes	3	>4	1	No	No	Terhadap Cluster 5
1	1	Private Sector/Self Employed	Yes	4	>4	0	No	No	Terhadap Cluster 3
2	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 5
3	0	Private Sector/Self Employed	Yes	3	<4	1	No	No	Terhadap Cluster 4
4	0	Private Sector/Self Employed	Yes	3	>4	1	Yes	No	Terhadap Cluster 2
5	0	Private Sector/Self Employed	No	4	4	0	No	No	Terhadap Cluster 1
6	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 1
7	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 5
8	0	Private Sector/Self Employed	Yes	4	>4	1	Yes	Yes	Terhadap Cluster 1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1268	1	Private Sector/Self Employed	Yes	4	4	0	No	No	Terhadap Cluster 7
1269	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 5
1270	0	Private Sector/Self Employed	Yes	4	>4	0	Yes	No	Terhadap Cluster 4
1271	1	Government Sector	Yes	4	4	0	No	No	Terhadap Cluster 1
1272	0	Private Sector/Self Employed	Yes	3	>4	1	No	No	Terhadap Cluster 6
1273	0	Private Sector/Self Employed	No	4	>4	1	No	No	Terhadap Cluster 2
1274	0	Private Sector/Self Employed	Yes	4	>4	1	No	Yes	Terhadap Cluster 5
1275	0	Private Sector/Self Employed	Yes	4	>4	1	No	No	Terhadap Cluster 5
1276	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 6

Lampiran 16 Hasil Pengelompokkan 7 Cluster Data Dengan Asuransi

No.	Age	Employment Type	GraduateOrNot	AnnualIncome	Family Members	Chronic Diseases	Frequent Flyer	EverTravelled Abroad	Label
0	1	Private Sector/Self Employed	Yes	3	4	1	No	No	Terhadap Cluster 4
1	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 5
2	0	Private Sector/Self Employed	Yes	4	4	1	Yes	Yes	Terhadap Cluster 5
3	0	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 5
4	1	Government Sector	Yes	3	4	0	No	No	Terhadap Cluster 4
5	1	Government Sector	Yes	4	4	0	Yes	Yes	Terhadap Cluster 5
6	0	Private Sector/Self Employed	Yes	3	>4	0	No	No	Terhadap Cluster 1
7	0	Private Sector/Self Employed	Yes	4	4	1	No	No	Terhadap Cluster 7
8	1	Private Sector/Self Employed	Yes	4	<4	0	Yes	Yes	Terhadap Cluster 5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
701	0	Private Sector/Self Employed	No	3	<4	0	No	No	Terhadap Cluster 6
702	0	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 1
703	1	Private Sector/Self Employed	Yes	4	<4	0	No	Yes	Terhadap Cluster 5
704	0	Private Sector/Self Employed	Yes	4	>4	0	No	Yes	Terhadap Cluster 5
705	0	Private Sector/Self Employed	No	4	4	0	No	Yes	Terhadap Cluster 6
706	1	Private Sector/Self Employed	Yes	3	4	0	No	No	Terhadap Cluster 3
707	0	Government Sector	Yes	3	<4	0	No	No	Terhadap Cluster 2
708	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 5
709	1	Private Sector/Self Employed	Yes	4	4	0	Yes	Yes	Terhadap Cluster 5