PENERAPAN DATA MINING UNTUK KLASTERISASI JENIS PEKERJAAN MENGGUNAKAN METODE K-MEANS (STUDI KASUS: PENGGUNA APLIKASI SOSIAL MEDIA "X" DI KOTA SEMARANG)

SKRIPSI

Diajukan untuk Memenuhi Sebagian Syarat Guna Memperoleh
Gelar Sarjana Matematika
dalam Ilmu Matematika

Oleh : **MARCELA ANDARISTA MAHARANI**
NIM : 2008046035

FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI WALISONGO
SEMARANG
2024

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini :

Nama	:	Marcela Andarista Maharani
NIM	:	2008046035
Jurusan/Program Studi	:	Matematika/ Matematika

menyatakan bahwa skripsi yang berjudul :

PENERAPAN DATA MINING UNTUK KLASTERISASI JENIS PEKERJAAN MENGGUNAKAN METODE K-MEANS (STUDI KASUS: PENGGUNA APLIKASI SOSIAL MEDIA "X" DI KOTA SEMARANG)

secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 19 September 2024
Pembuat pernyataan,

Marcela Andarista Maharani
NIM : 2008046035

Scanned by TapScanner

KEMENTERIAN AGAMA R.I.
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Prof. Dr. Hamka (Kampus II) Ngaliyan Semarang
Telp. 024-7601295 Fax. 7615387

PENGESAHAN

Naskah skripsi berikut ini :

Judul : PENERAPAN DATA MINING UNTUK KLASTERISASI JENIS PEKERJAAN MENGGUNAKAN METODE K-MEANS (Studi Kasus: Pengguna Aplikasi Sosial Media "X" di Kota Semarang)
Penulis : Marcela Andarista Maharani
NIM : 2008046035
Jurusan : Matematika

Telah diujikan dalam sidang *tugas akhir* oleh Dewan Penguji Fakultas Sains dan Teknologi UIN Walisongo dan dapat diterima sebagai salah satu syarat memperoleh gelar sarjana dalam Ilmu Matematika.

Semarang, 16 Oktober 2024

DEWAN PENGUJI

Penguji I,

Emy Siswanah, M.Sc
NIP : 198702022011012014

Penguji II,

Ariska Kurnia Rachmawati, M.Sc
NIP : 198908112019032019

Penguji III,

Zulaikha, M.Si
NIP : 199204092019032021

Penguji IV,

Nur Khasanah, M.Si
NIP : 199112012019032017

Pembimbing,

Ariska Kurnia Rachmawati, M.Sc
NIP : 198908112019032019

Scanned by TapScanner

NOTA DINAS

Semarang, 19 September 2024

Yth. Ketua Program Studi Matematika
Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum warahmatullahi wabarakatuh

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan, arahan dan koreksi naskah skripsi dengan:

Judul	:	PENERAPAN DATA MINING UNTUK KLASTERISASI JENIS PEKERJAAN MENGGUNAKAN METODE K-MEANS (Studi Kasus: Pengguna Aplikasi Sosial Media "X" di Kota Semarang)
Nama	:	Marcela Andarista Maharani
NIM	:	2008046035
Jurusan	:	Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo untuk diujikan dalam Sidang Munaqasyah.

Wassalamu'alaikum warahmatullahi wabarakatuh

Pembimbing,

Ariska Kurnia Rachmawati, M.Sc
NIP : 198908112019032019

Scanned by TapScanner

ABSTRAK

Penelitian ini menerapkan teknik data mining, khususnya metode K-Means *clustering*, untuk mengelompokkan jenis pekerjaan di kalangan pengguna aplikasi media sosial "X" (*formerly* twitter) di Kota Semarang. Tujuan dari penelitian ini adalah membantu pencari kerja dalam menentukan rekomendasi jenis pekerjaan yang lebih tepat berdasarkan latar belakangnya, seperti pendidikan terakhir, kondisi ekonomi keluarga, hobi, dan lain-lain. Melalui klasterisasi, penelitian ini menganalisis pola dalam data untuk memberikan rekomendasi pekerjaan yang disesuaikan melalui simulasi sistem GUI berbasis MATLAB. Simulasi sistem ini mengelompokkan pengguna ke dalam *cluster* tertentu dan memberikan rekomendasi pekerjaan yang sesuai, bertujuan untuk memberikan perspektif alternatif bagi individu agar dapat mengejar karier yang selaras dengan keterampilan dan preferensi mereka. Data untuk penelitian ini dikumpulkan melalui kuesioner, banyak data yang dikumpulkan adalah sebanyak 199 data responden. Hasil klasterisasi didasarkan pada variabel seperti usia, pendidikan terakhir, dan jenis pekerjaan yang membagi data responden menjadi tiga *cluster* dengan karakteristik data yang sama dalam satu *cluster*. Penelitian ini menunjukkan efektivitas data mining dalam proses rekomendasi pekerjaan, dengan implikasi praktis bagi para pencari kerja.

Kata kunci : K-Means *clustering*, rekomendasi pekerjaan, simulasi sistem GUI MATLAB.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT. yang telah melimpahkan rahmat dan hidayah-Nya serta memberikan kelancaran dan kemudahan sehingga skripsi yang berjudul 'Penerapan Data Mining untuk Klasterisasi Jenis Pekerjaan Menggunakan Metode K-Means (Studi Kasus: Pengguna Aplikasi Sosial Media "X" di Kota Semarang)' dapat selesai dengan baik. Skripsi ini disusun dan diajukan untuk memperoleh gelar Sarjana S1 pada Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Walisongo Semarang.

Dalam penyusunan skripsi ini, banyak pihak yang telah memberikan bantuan, dukungan, hingga bimbingan. Oleh karena itu, penulis mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. Nizar, M.Ag. selaku Rektor Universitas Islam Negeri Walisongo Semarang.
2. Bapak Prof. Dr. H. Musahadi, M.Ag. selaku Dekan Fakultas Sains dan Teknologi.
3. Ibu Any Muanalifah, M.Si., Ph.D. selaku Ketua Program Studi Matematika.
4. Ibu Ariska Kurnia Rachmawati, M.Sc. selaku dosen pembimbing yang telah memberikan kritik, saran, serta arahan dalam penulisan skripsi ini.
5. Ibu Zulaikha, M.Si. selaku dosen wali yang telah mendampingi penulis selama perkuliahan ini.
6. Bapak dan Ibu Dosen Jurusan Matematika yang telah

memberikan ilmu kepada penulis, serta seluruh staf atas segala bantuannya.

7. Bapak, Ibu, dan Kakak tercinta yang selalu memberikan doa, dukungan, dan motivasi kepada penulis selama menyelesaikan skripsi ini.
8. Teman-teman S1 Matematika angkatan 2020 yang telah membersamai penulis selama menjalani proses perkuliahan ini.
9. Yolanda K. Nabilah, sahabat terkasih yang dengan setia selalu memberikan dukungan emosional dan mental selama proses penulisan skripsi ini.
10. Ajeng Sekar, Prita Arni, Silviana, dan Fuadah, sahabat seperjuangan yang senantiasa memberikan dukungan, baik dalam bentuk waktu maupun motivasi selama penulis menjalani masa studi dan penyusunan skripsi ini.
11. Huna, kucing peliharaan tercinta, yang dengan setia menemani penulis dalam setiap langkah pengerjaan skripsi ini, dari pagi hingga larut malam, menjadi sumber ketenangan di saat-saat sulit.
12. Bias idol dan karakter fiksi favorit penulis, yang selalu menjadi sumber hiburan ketika penulis merasa buntu dalam menyelesaikan skripsi ini, memberikan semangat baru untuk terus melanjutkan proses penulisan.
13. Diri sendiri, yang telah dengan gigih bertahan hingga detik terakhir penyusunan skripsi ini, dengan segala tantangan dan cobaan yang dihadapi, tetap tidak menyerah, dan

berhasil menyelesaikan tugas akademik ini dengan penuh ketekunan dan kesabaran.

14. Semua pihak yang tidak dapat penulis sebutkan satu persatu yang telah memberikan kontribusi hingga selesainya skripsi ini.

Akhir kata, penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Untuk itu, penulis mengharapkan saran dan kritik untuk kemajuan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi penulis khususnya, dan bagi para pembaca umum.

DAFTAR ISI

HALAMAN JUDUL ..... **i**
PERNYATAAN KEASLIAN ..... **ii**
PENGESAHAN ..... **iii**
NOTA PEMBIMBING ..... **iv**
KATA PENGANTAR ..... **viii**
DAFTAR ISI ..... **ix**
DAFTAR TABEL ..... **xi**
DAFTAR GAMBAR ..... **xiii**
DAFTAR LAMPIRAN ..... **xiv**
BAB 1 PENDAHULUAN **1**
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 6
1.3 Tujuan Penelitian 6
1.4 Manfaat Penelitian 6
1.5 Batasan Masalah 7
BAB 2 LANDASAN TEORI **9**
2.1 Data 9
2.2 Data Mining 11
2.2.1 Model Statistika pada Data Mining 11
2.2.2 Jenis Data Mining 12
2.2.3 Metode Data Mining 14
2.2.4 Pengelompokkan Data Mining 15
2.3 Teknik *Unsupervised Learning* 16
2.4 Metode Klasterisasi (*Clustering Method*) 16
2.5 Metode K-Means 20
2.6 Reduksi Dimensi Data dengan PCA 23
2.7 Metode *Silhouette* 26
2.8 Penelitian Terdahulu 27
BAB 3 METODE PENELITIAN **29**
3.1 Sumber Data dan Variabel Penelitian 29
3.2 Populasi dan Sampel Penelitian 30

3.3 Teknik Pengumpulan Data dan Instrumen Penelitian 32
3.4 Analisis Data 34
3.5 Rancangan Penelitian 37
BAB 4 HASIL PENELITIAN DAN PEMBAHASAN 39
4.1 Data 39
4.2 *Preprocessing Data Clustering* (Klasterisasi) 44
4.3 Penerapan Metode K-Means 45
4.3.1 Pemilihan Data 46
4.3.2 Normalisasi Data 51
4.3.3 Reduksi Dimensi Data dengan PCA 55
4.3.4 Penentuan Jumlah *Cluster* dan Posisi *Centroid* pada Iterasi Pertama 56
4.3.5 Klasterisasi dengan Metode K-Means 57
4.4 Pembuatan Simulasi Sistem Rekomendasi Pekerjaan 77
BAB 5 PENUTUP 82
5.1 Kesimpulan 82
5.2 Saran 83
DAFTAR PUSTAKA 84
Lampiran-lampiran 89

DAFTAR TABEL

Tabel	Judul	Halaman
Tabel 3.1	Variabel penelitian	30
Tabel 3.2	Jumlah Penduduk berdasarkan Kelompok Umur	31
Tabel 3.3	Jumlah Penduduk Kelompok Umur 20-34 Tahun	32
Tabel 3.4	Instrumen Penelitian	33
Tabel 4.1	Responden penelitian	39
Tabel 4.2	Responden berdasarkan usia	40
Tabel 4.3	Responden berdasarkan jenis kelamin	40
Tabel 4.4	Responden berdasarkan kondisi ekonomi keluarga	41
Tabel 4.5	Responden berdasarkan pendidikan terakhir	42
Tabel 4.6	Responden berdasarkan capaian penghargaan	42
Tabel 4.7	Responden berdasarkan status pekerjaan	42
Tabel 4.8	Responden berdasarkan golongan gaji	43
Tabel 4.9	Responden berdasarkan hubungan hobi dan pekerjaan	43
Tabel 4.10	Responden berdasarkan hubungan capaian penghargaan dengan pekerjaan	44
Tabel 4.11	Data dalam bentuk numerik	51
Tabel 4.12	Rata-rata (mean) untuk setiap variabel	52

Tabel 4.13	Standar deviasi populasi untuk setiap variabel	54
Tabel 4.14	Normalisasi data untuk setiap nilai data	55
Tabel 4.15	Reduksi dimensi data	56
Tabel 4.16	Data hasil klasterisasi	61
Tabel 4.17	Kecenderungan (dominasi) nilai data variabel usia	62
Tabel 4.18	Kecenderungan (dominasi) nilai data variabel jenis kelamin	63
Tabel 4.19	Kecenderungan (dominasi) nilai data variabel kondisi ekonomi keluarga	64
Tabel 4.20	Kecenderungan (dominasi) nilai data variabel pendidikan terakhir	65
Tabel 4.21	Kecenderungan (dominasi) nilai data variabel hobi	66
Tabel 4.22	Kecenderungan (dominasi) nilai data variabel capaian penghargaan	67
Tabel 4.23	Kecenderungan (dominasi) nilai data variabel jenis pekerjaan	68
Tabel 4.24	Kecenderungan (dominasi) nilai data variabel golongan gaji	69
Tabel 4.25	Kecenderungan (dominasi) nilai data variabel hubungan dengan hobi	70
Tabel 4.26	Kecenderungan (dominasi) nilai data variabel hubungan dengan penghargaan	71
Tabel 0.1	42 Kombinasi data inputan pengguna	121

DAFTAR GAMBAR

Gambar	Judul	Halaman
Gambar 3.1	*Flowchart* Metode K-Means	36
Gambar 3.2	Alur Rancangan Penelitian	37
Gambar 4.1	Grafik Data Hasil Klasterisasi	72
Gambar 4.2	Tampilan Awal	78
Gambar 4.3	Grafik Hasil Klasterisasi	79
Gambar 4.4	Tampilan Sebelum Data Di*input*kan	80
Gambar 4.5	Tampilan Setelah *Input* Data	81
Gambar 4.6	Tampilan Sistem Jika Data Tidak Termasuk dalam *Cluster*	81
Gambar 0.1	Data Responden	90
Gambar 0.2	Data Responden Bekerja	92
Gambar 0.3	Transformasi Data	93
Gambar 0.4	Normalisasi Data	94
Gambar 0.5	Reduksi Dimensi Data dengan PCA	97
Gambar 0.6	Iterasi-1 dengan Metode K-Means	100
Gambar 0.7	Iterasi-2 dengan Metode K-Means	101
Gambar 0.8	Hasil Klasterisasi dengan Metode K-Means	112

DAFTAR LAMPIRAN

Halaman

Lampiran 1	Data Hasil Jawaban Responden pada Kuesioner Penelitian	89
Lampiran 2	Data Responden Bekerja	91
Lampiran 3	Hasil Transformasi Data	93
Lampiran 4	Hasil Normalisasi Data	94
Lampiran 5	*Listing* Program Reduksi Dimensi dengan PCA	95
Lampiran 6	Hasil Reduksi Dimensi Data dengan PCA	97
Lampiran 7	*Listing* Program Penentuan *Cluster* Optimal dengan Metode Silhouette	98
Lampiran 8	Perhitungan Iterasi-1 untuk Klasterisasi dengan Metode K-Means	100
Lampiran 9	Perhitungan Iterasi-2 untuk Klasterisasi dengan Metode K-Means	101
Lampiran 10	*Listing* Program Klasterisasi Data dengan Metode K-Means	102
Lampiran 11	Hasil Klasterisasi dengan Metode K-Means	111
Lampiran 12	*Listing* Program Simulasi Sistem Rekomendasi Pekerjaan	113
Lampiran 13	Kombinasi Data Inputan untuk Setiap *Cluster* dalam Simulasi Sistem Rekomendasi Pekerjaan	121
Lampiran 14	Daftar Riwayat Hidup	123

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Setiap orang membutuhkan pekerjaan untuk memenuhi kebutuhan hidupnya. Banyaknya jenis pekerjaan yang ada membuat seseorang susah mengambil keputusan dalam memilih pekerjaan sesuai dengan kemampuan dan latar belakang yang dimiliki. Terdapat banyak faktor yang mempengaruhi pemilihan jenis pekerjaan yang sesuai. Seseorang yang telah menentukan pilihan pekerjaannya dengan tepat diharapkan dapat menggunakan keterampilannya dengan maksimal pada setiap bidang pekerjaan yang digeluti (Romaito, 2017). Seperti yang dituliskan pada Surah At-Taubah ayat 105:

وَقُلِ اعْمَلُوْا فَسَيَرَى اللّٰهُ عَمَلَكُمْ وَرَسُوْلُهٗ وَالْمُؤْمِنُوْنَۗ وَسَتُرَدُّوْنَ اِلٰى عٰلِمِ الْغَيْبِ وَالشَّهَادَةِ فَيُنَبِّئُكُمْ بِمَا كُنْتُمْ تَعْمَلُوْنَ

Artinya: Katakanlah (Nabi Muhammad), “Bekerjalah! Maka, Allah, rasul-Nya, dan orang-orang mukmin akan melihat pekerjaanmu. Kamu akan dikembalikan kepada (Zat) yang mengetahui yang gaib dan yang nyata. Lalu, Dia akan memberitakan kepada kamu apa yang selama ini kamu kerjakan.”

Ayat ini mengingatkan setiap orang untuk berbuat dan bekerja dengan baik. Dalam Islam, bekerja bukan hanya untuk mencari rezeki, tetapi juga merupakan wujud dari tanggung jawab manusia kepada Allah. Setiap usaha dan pekerjaan yang dilakukan dengan niat yang benar dan ikhlas merupakan bentuk ibadah.

Memasuki era *big data* yang semakin berkembang pesat, nilai-nilai tersebut dapat diaplikasikan melalui pemanfaatan teknologi yang mempermudah pengelolaan pekerjaan. Semua tingkatan pendidikan telah menerapkan berbagai sistem informasi canggih untuk mengelola informasi, tidak terkecuali dalam dunia pekerjaan. Namun, sebagian besar sistem manajemen pekerjaan yang saat ini digunakan hanya menyajikan beberapa fungsi sederhana, seperti entri informasi, modifikasi, dan permintaan (Yuhua *et al*, 2015). Kemudian masalah yang muncul adalah tentang menentukan cara yang efektif untuk menggali informasi dari sejumlah data pekerjaan berukuran besar.

Pengolahan data yang tepat diperlukan untuk memperoleh informasi. Banyaknya data yang harus diolah mengakibatkan terjadinya kesulitan dalam proses pengelompokkan data. Solusi untuk memecahkan masalah tersebut adalah dengan data mining. Data mining merupakan sebuah cara untuk menggali data dari *database* berukuran besar dengan tingkat kerumitan yang banyak digunakan pada aplikasi, seperti *mobile banking* dan komunikasi (Jananto, 2010). Larose (2005) menyatakan bahwa berdasarkan tugasnya, data mining dibagi menjadi beberapa kelompok , seperti deskripsi, estimasi, prediksi, klasifikasi, *clustering* (klasterisasi) dan *assosiation rule*.

Metode *clustering* (klasterisasi) adalah metode yang mengelompokkan objek-objek tertentu berdasarkan karakteristik dan kesamaannya. Macam-macam metode *clustering* (klasterisasi) meliputi K-Means, K-Medoid, *Hierarchical Clustering*, DBSCAN, Fuzzy C-Means, dan *Self-Organizing Map*. Salah satu metode *clustering* (klasterisasi) yang digunakan dalam penelitian ini adalah K-Means. Metode ini mempartisi data ke dalam satu

atau lebih *cluster*, di mana data dengan karakteristik yang mirip dikelompokkan dalam *cluster* yang sama. Kelebihan dari metode K-Means adalah mudah untuk diadaptasi/dijalankan dan merupakan salah satu metode yang umum dipelajari (Tan *et al*, 2005).

Alasan peneliti menggunakan aplikasi sosial media "X" (*formerly twitter*) sebagai tempat pengumpulan data adalah karena aplikasi tersebut memiliki pengguna yang aktif berdiskusi dan berbagi informasi yang lebih eksplisit secara *real-time*, sehingga responden lebih mudah diakses melalui penyebaran link atau tweet yang dapat dibagikan kembali oleh pengguna lain (*retweet* dan *quote-tweet*), hal ini meningkatkan peluang untuk menjangkau responden yang lebih luas dalam waktu singkat. Aplikasi ini juga memiliki fitur komunitas dengan berbagai karakteristik unik seperti akademik, hobi, hingga isu sosial, serta memiliki banyak akun yang mengelola pesan anonim dari pengguna (*base menfess*) yang memungkinkan pengguna dengan minat yang sama untuk berkumpul di satu ruang diskusi. Dengan memanfaatkan fitur dan akun pesan anonim tersebut, penyebaran kuesioner dapat dilakukan secara lebih terarah kepada responden sehingga meningkatkan validitas data yang diperoleh. Penyebaran kuesioner melalui *base menfess* memberikan rasa anonimitas bagi responden, yang sering kali meningkatkan keinginan mereka untuk berpartisipasi. Hal ini relevan dengan penelitian yang membutuhkan data jujur tanpa tekanan.

Pada penelitian yang dilakukan oleh Ginanjar (2016) berjudul '*Clustering* Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma K-Means', digunakan metode K-Means untuk melakukan pengelompokkan yang nantinya

akan membentuk kelompok belajar mahasiswa berdasarkan *cluster* nilai UTS. Dalam penelitian oleh Wulan dan Sudin (2021) yaitu 'Penerapan Data Mining *Clustering* untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci' bertujuan untuk mengelompokkan berbagai jenis dan merk mesin cuci dengan harapan bisa mempermudah bagi para ibu rumah tangga dalam memilih mesin cuci yang mereka yang dibutuhkan.

Pada penelitian berjudul '*Research on Evaluation of College Students' Professional Ability Based on K-means Clustering*' oleh Chen, Liu, & Ma (2018) dilakukan analisis kemampuan profesional siswa secara keseluruhan, membandingkan perbedaan yang ada di antara siswa sesuai dengan kemampuan profesional yang berbeda, dan memberikan hasil analisis yang memudahkan manajer pengajaran memahami distribusi kemampuan profesional siswa untuk mengembangkan rencana pengajaran yang tepat. Dalam penelitian yang ditulis oleh Maria dan Andi (2022), metode K-Means digunakan untuk mempartisi sekumpulan data dengan tujuan membantu pustakawan dalam mengelompokkan data peminjaman buku yang sangat disukai atau diminati oleh mahasiswa. Penelitian ini berjudul '*Implementation of Data Mining Clustering Using the K-Means Method in Grouping Library Books*'. Analisis lebih dalam dapat dilakukan dengan mengolah data transaksi peminjaman buku dengan hasil yang menjelaskan berbagai informasi mengenai sebaran intensitas peminjaman buku sehingga pihak perpustakaan dapat menentukan penambahan koleksi buku yang tepat sasaran.

Pada penelitian yang dilakukan oleh Liu dan Zhai (2017) dengan judul '*Research and Application of Data Mining in College Students Employment Based on Association Rules*', membahas

tentang informasi pekerjaan mahasiswa menggunakan algoritma Apriori untuk mengubah, menambang (Guohua, 2016), menarik aturan yang relevan, membuat panduan untuk pekerjaan mahasiswa dan tips bagi pendidik untuk mengatur jurusan dan kursus, sehingga dapat meningkatkan tingkat lapangan kerja mahasiswa. Penelitian tersebut mengambil data informasi ketenagakerjaan lulusan universitas dan mengembangkan sistem untuk menemukan beberapa faktor yang dapat mempengaruhi situasi ketenagakerjaan.

Dari penelitian terdahulu yang telah dijelaskan di atas, dapat disimpulkan bahwa ada kebutuhan akan sistem yang tidak hanya mengelompokkan data terkait pekerjaan, namun juga membantu individu, terutama para pencari kerja, dalam menentukan pilihan karier yang sesuai dengan latar belakang dan potensinya. Sebanyak 80% orang Indonesia dengan gelar sarjana, tidak bekerja sesuai dengan jenis pekerjaan yang diharapkan ketika lulus dari perguruan tinggi dikarenakan ketatnya persaingan dalam dunia kerja (Medkom 2023, diakses 11 Juni 2023). Tingginya persentase lulusan sarjana di Indonesia yang tidak bekerja sesuai bidang studinya di perguruan tinggi ini, menandakan adanya kesenjangan antara harapan dan kenyataan di dunia kerja. Inilah yang menjadi alasan bagi peneliti untuk mengembangkan simulasi sistem GUI yang *output*nya akan memberikan rekomendasi pekerjaan pada pengguna dengan maksud agar para pengguna sistem dapat melihat potensi lain yang dimiliki dan terdorong untuk bekerja di bidang lain yang sesuai dengan latar belakangnya.

Berdasarkan uraian di atas serta penelitian yang telah dilakukan sebelumnya, maka akan dikaji penelitian yang dapat menghasilkan sebuah program berbasis GUI yang menjadi

referensi bagi semua orang dalam menentukan jenis pekerjaan yang cocok dengan latar belakang yang dimiliki. Maka dari itu, peneliti melakukan penelitian yang berjudul **"PENERAPAN DATA MINING UNTUK KLASTERISASI JENIS PEKERJAAN MENGGUNAKAN METODE K-MEANS (Studi Kasus: Pengguna Aplikasi Sosial Media "X" di Kota Semarang)"**.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana model klasterisasi jenis pekerjaan menggunakan metode K-Means dan penerapannya sebagai simulasi sistem rekomendasi jenis pekerjaan?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui model klasterisasi jenis pekerjaan menggunakan metode K-Means dan penerapannya sebagai simulasi sistem rekomendasi jenis pekerjaan.

1.4 Manfaat Penelitian

Manfaat dari penelitian adalah sebagai berikut:

1. Bagi Peneliti

    (a) Peneliti dapat menerapkan ilmu yang dipelajari terutama mengenai metode K-Means dalam data mining.

(b) Membandingkan teori yang ada dengan masalah yang sebenarnya.

2. Bagi Akademik/Universitas

(a) Mengetahui kemampuan mahasiswa dalam penguasaan materi perkuliahan yang telah diperoleh.

(b) Mengetahui kemampuan mahasiswa dalam penerapan ilmu dan sebagai bahan evaluasi.

(c) Memberikan gambaran tentang kesiapan mahasiswa dalam menghadapi dunia kerja yang sesungguhnya.

(d) Sebagai sumber tambahan dalam membantu mahasiswa melakukan penelitian selanjutnya.

3. Bagi Masyarakat Umum

(a) Sebagai bahan referensi dalam membuat suatu penelitian.

(b) Memberikan rekomendasi pekerjaan sesuai dengan latar belakang pembaca.

1.5 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Responden merupakan pengguna aplikasi X (*formerly* twitter) di Kota Semarang dan memiliki pekerjaan baik tetap ataupun sampingan.

2. Batas usia responden yang dapat mengisi kuesioner antara usia 20-34 tahun.

3. Data yang diperlukan dalam penelitian ini adalah data latar belakang responden, seperti nama, usia, jenis kelamin (gender), hobi/kesenangan, kondisi ekonomi keluarga, capaian penghargaan, pendidikan terakhir, jenis pekerjaan, dan golongan gaji (bergaji tetap/tidak).

4. Dalam proses pengembangan simulasi sistem, menggunakan bahasa pemograman MATLAB versi 2013a.

BAB 2

LANDASAN TEORI

2.1 Data

Data dianggap sebagai faktor kunci dalam proses pengambilan keputusan dan landasan utama untuk memastikan akuntabilitas. Tanpa data yang akurat dan berkualitas, menilai keputusan dengan efektif menjadi sangat sulit (Vassakis *et al*, 2018). Dalam definisi lainnya, data adalah himpunan fakta yang dinyatakan dalam bentuk karakter seperti huruf, angka, atau simbol lain, yang kemudian diolah menjadi informasi baru. Pengumpulan data dapat dilakukan melalui berbagai metode, seperti observasi, kuesioner, wawancara dengan narasumber, dan lain-lain (Nofriansyah, 2015).

Data yang akan ditambang berupa kumpulan data atau *database* yang nantinya akan diekstraksi menjadi sebuah informasi, seperti pola, *cluster*, pohon keputusan, dan lain sebagainya. Terdapat dua elemen dalam sebuah himpunan data, yaitu *attribute/feature/selection* dan *class*/label.

Attribute adalah karakteristik/sifat yang dimiliki sebuah objek data. Sementara itu, *class/label/target* merupakan atribut yang digunakan dalam pengambilan keputusan. Data mining memiliki tiga jenis data, yaitu:

1. *Numeric* adalah jenis data yang dapat dihitung atau diolah secara matematis, dan terbagi menjadi dua kategori, yaitu *ratio* dan interval.

    (a) *Ratio* adalah data yang diperoleh dengan mengukur jarak antara dua titik pada skala yang sudah diketahui.

Memiliki titik nol yang absolut. Contoh: jika A berusia 50 tahun dan B berusia 25 tahun, maka rasio perbandingan antara usia A dan B menunjukkan bahwa usia A dua kali lipat usia B.

(b) Interval adalah data yang dihasilkan dari pengukuran jarak antara dua titik pada skala yang sudah ditetapkan. Tipe data ini tidak memiliki titik nol yang mutlak. Contoh: jika data A, B, C, dan D secara berurutan memiliki skala 1, 2, 3, dan 4, maka interval antara A dan D adalah 4 - 1 = 3.

2. *Categorial* adalah jenis data yang tidak dapat dihitung, terbagi menjadi nominal dan ordinal.

   (a) Nominal adalah data yang diperoleh dengan cara mengkategorikan/mengklasifikasikan data ke dalam bentuk label. Menunjukkan beberapa objek yang berbeda. Contoh: jenis kelamin pria = 1 dan wanita = 2.

   (b) Ordinal adalah data yang diperoleh dengan cara mengkategorikan/mengklasifikasikan data ke dalam bentuk label. Di antara data tersebut terdapat keterkaitan atau urutan tertentu. Contoh: nilai A berkisar antara 81-100, nilai B antara 65-80, dan nilai C antara 56-64.

3. *Time series* adalah jenis data yang tidak dapat dihitung secara langsung. Data ini terdiri dari pengamatan terhadap satu atau lebih variabel dari waktu ke waktu. Dengan kata lain, data *time series* merupakan data yang dikumpulkan secara

berurutan dari periode ke periode pada satu atau lebih objek yang sama. Contoh: harga saham, indeks harga konsumen, dan lain-lain. (Fitri *et al*, 2021)

2.2 Data Mining

Data mining adalah proses menambang data berukuran besar dengan mencari pola atau aturan tertentu untuk menemukan informasi baru (Davies, 2004). Pola yang telah ditemukan harus memiliki arti dan memberikan keuntungan. Tujuan utama data mining adalah untuk mengekstraksi informasi berharga dan pola tersembunyi dari data yang besar dan kompleks. Data mining memiliki keterkaitan dengan bidang ilmu lain seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi. Tahap-tahap data mining terdiri dari seleksi, *preprocessing*, transformasi, data mining, dan evaluasi (Kusrini dan Luthfi, 2009).

Menurut Retno (2017), data mining dapat dimanfaatkan melalui sudut pandang komersil dan keilmuan. Dari perspektif komersil, data mining digunakan dalam penanganan meluapnya isi data menggunakan komputasi untuk mendapatkan informasi yang bersifat penting dan alasan yang dapat memicu meningkatnya daya saing. Sedangkan dari segi keilmuan, data mining digunakan untuk membuat analisis dan melakukan penyimpanan data yang *realtime* dan berukuran besar.

2.2.1 Model Statistika pada Data Mining

Model dari data mining terbagi menjadi empat metode, yaitu:

1. *Supervised Learning* adalah metode pendekatan yang memiliki data latih dan variabel yang menjadi target, bertujuan untuk mengelompokkan sebuah data ke dalam data yang sudah ada.

2. *Unsupervised Learning* adalah metode yang tidak memiliki data latih sehingga pengelompokkan dari data yang ada dibagi menjadi dua hingga tiga bagian.

3. *Regression Model* adalah metode untuk menganalisa hubungan antara variabel target berbentuk numerik dengan satu atau beberapa variabel prediktor berbentuk kategorik dan atau numerik.

4. *Association Rule* adalah metode untuk mengoptimalkan tata letak barang pada sebuah swalayan atau perpustakaan untuk mengetahui pola belanja konsumen.

2.2.2 Jenis Data Mining

Berikut akan dijelaskan jenis-jenis dari data mining.

1. *Market Basket Analysis (MBA)* adalah proses untuk menganalisa pola pelanggan dalam pemilihan *item* yang akan dibeli. MBA menggunakan data transaksi penjualan untuk menganalisis dan menemukan pola *item* yang sering dibeli secara bersamaan dalam satu kali transaksi. Pola yang diidentifikasi akan digunakan untuk merancang strategi penjualan atau pemasaran yang lebih efektif, seperti menempatkan *item* yang sering dibeli bersama di area yang berdekatan, merancang tampilan produk di katalog,

memberikan kupon diskon, dan lain sebagainya. MBA termasuk dalam kategori metode asosiasi.

2. *Memory Based Reasoning (MBR)* adalah integrasi antara metode klasifikasi dan penalaran yang berbasis memori. Proses ini menggunakan satu set data untuk membangun model yang dapat membuat prediksi atau asumsi mengenai objek baru. MBR terdiri dari dua komponen utama: fungsi kesamaan (yang menilai tingkat kesamaan antar pasangan objek) dan fungsi kombinasi (yang menggabungkan hasil dari himpunan tetangga terdekat untuk menghasilkan keputusan). MBR juga dapat dipahami sebagai proses yang mengidentifikasi kasus-kasus yang serupa dan menerapkan informasi dari kasus-kasus tersebut pada catatan baru. Di dalam *Enterprise Miner*, node MBR berfungsi sebagai alat pemodelan untuk mengklasifikasikan atau memprediksi observasi dengan memanfaatkan algoritma *K-Nearest Neighbor*.

3. *Cluster detection* termasuk ke dalam metode *clustering* (klasterisasi). Terdapat dua pendekatan dalam klasterisasi. Pertama, pendekatan yang menganggap bahwa sejumlah *cluster* telah ada dalam data dan bertujuan untuk membagi data tersebut ke dalam *cluster* yang ada. Kedua, pendekatan *clustering agglomerative*, yang menganggap bahwa setiap *item* pada awalnya merupakan *cluster* tersendiri, kemudian secara bertahap *item-item* ini digabungkan menjadi *cluster* yang lebih besar. Meskipun proses penggabungannya berbeda, komputasi yang dilakukan tetap serupa.

4. *Link analysis* adalah metode analisis data yang diterapkan dalam teori jaringan untuk menilai hubungan/koneksi antara node dalam jaringan. Node ini bisa berupa berbagai jenis entitas, seperti individu, organisasi, atau transaksi. Teknik ini sering digunakan dalam optimasi mesin pencari, serta di bidang intelijen, penelitian keamanan, dan analisis pasar maupun kesehatan.

5. *Rule induction* adalah sebuah proses dalam data mining yang bertujuan untuk menghasilkan aturan "jika-maka" dari sekumpulan data. Aturan ini, yang sering diperoleh melalui *decision tree*, menggambarkan hubungan antara atribut dan label kelas dalam data. Banyak kejadian di dunia nyata yang didasari pada induksi aturan yang secara alami mudah dipahami. Meskipun aturan-aturan ini tidak selalu berlaku secara mutlak, mereka umumnya didasarkan pada pola yang ditemukan dari pengalaman dan pengamatan sehari-hari. *Rule induction* menawarkan pendekatan klasifikasi yang kuat dan efektif dalam analisis data.

6. *Neural network* adalah model prediksi non-linear yang belajar melalui proses latihan dan meniru struktur jaringan saraf pada makhluk hidup. Model ini mampu memahami data yang kompleks dan ambigu, serta digunakan untuk mengekstraksi pola dan mendeteksi tren yang sangat rumit, yang sulit dijelaskan oleh manusia atau metode komputer lainnya.

2.2.3 Metode Data Mining

Metode dari data mining sendiri meliputi:

1. *Association* adalah tahap pencarian untuk mendapatkan *rule* antar *item*.
2. *Sequence* adalah proses pencarian hubungan *rule* antar kombinasi *item* dan mengimplementasikannya pada lebih dari satu periode.
3. *Clustering* (klasterisasi) adalah proses mengelompokkan data yang sama.
4. *Classification* (klasifikasi) adalah proses menemukan tipe yang membedakan skema dari suatu kelas data yang bertujuan untuk membuat suatu perkiraan dari kelas objek yang belum diketahui.
5. *Regression* adalah memetakan data yang ada pada sebuah perkiraan.
6. *Forecasting* adalah tahap estimasi perkiraan nilai berlandaskan pola yang ada pada kumpulan data.
7. *Deviation* adalah menemukan *basic* dari sebuah permasalahan dan pemecahannya yang dihadapkan sebagai informasi dalam proses pengambilan keputusan akhir.

2.2.4 Pengelompokkan Data Mining

Data mining dibagi menjadi dua kategori berdasarkan tugasnya (Tan *et al*, 2005) yaitu:

1. *Descriptive* mining adalah proses untuk menemukan karakteristik penting dalam data. Teknik-teknik yang termasuk dalam kategori ini meliputi *clustering* (klasterisasi), *association*, dan *sequential pattern analysis*.

2. *Predictive* mining adalah proses untuk menemukan pola dalam data yang dapat digunakan untuk memprediksi nilai atau kejadian di masa depan, berdasarkan variabel yang ada. Salah satu teknik dalam kategori ini adalah *classification* (klasifikasi).

2.3 Teknik *Unsupervised Learning*

Unsupervised learning merupakan tipe pembelajaran dalam *machine learning*. Algoritma ini tidak memerlukan data dengan label sebagai input. Tujuan dari algoritma ini adalah untuk menemukan informasi yang dikelompokkan berdasarkan kesamaan dan perbedaan tanpa menggunakan data *training*. Mesin diharapkan mampu menemukan pola dan struktur yang belum diketahui dalam data yang tidak berlabel.

Unsupervised learning dibagi menjadi dua kategori utama, yaitu *clustering* (klasterisasi) dan *association*. *Clustering* (klasterisasi) adalah proses pengelompokan sekumpulan objek ke dalam beberapa kelompok, di mana objek dalam satu kelompok memiliki karakteristik yang serupa. *Association*, di sisi lain, bertujuan untuk menemukan hubungan atau keterkaitan antara data dalam sebuah *database*.

2.4 Metode Klasterisasi (*Clustering Method*)

Metode klasterisasi dalam data mining bertujuan untuk mengidentifikasi kelompok objek yang memiliki kesamaan di dalam satu kelompok, namun berbeda dengan objek yang berada di dalam kelompok lain. Klasterisasi merupakan proses pengelompokan objek-objek berdasarkan kesamaan

dan karakteristik yang dimiliki. Dalam data mining, metode ini memisahkan data sesuai dengan analisis yang dibutuhkan menggunakan algoritma tertentu. Dalam partisi keras, sebuah objek hanya dapat tergabung dalam satu *cluster*. Namun, pada partisi halus, sebuah objek dapat menjadi anggota beberapa *cluster* dengan derajat tertentu.

Clustering (klasterisasi), yang juga dikenal sebagai segmentasi, bertujuan untuk mengidentifikasi kelompok alami dari suatu kasus berdasarkan kesamaan variabel. Metode ini mengelompokkan data yang memiliki kemiripan dalam variabel-variabel tertentu. *Clustering* (klasterisasi) merupakan bagian dari *unsupervised learning* dalam data mining karena semua variabel diperlakukan secara setara tanpa label. Sebagian besar metode *clustering* membangun model melalui serangkaian iterasi dan berhenti ketika model telah mencapai keseimbangan, atau ketika batasan segmentasi telah stabil.

Metode *clustering* (klasterisasi) dibagi menjadi dua, yaitu:

1. *Hierarchical clustering* biasanya digunakan ketika jumlah kelompok yang akan dipilih belum diketahui. Pengelompokan ini memiliki dua pendekatan, yaitu *divisive* (dari atas ke bawah) dan *agglomerative* (dari bawah ke atas). Metode ini digunakan untuk menghitung ukuran kesamaan antar objek. Beberapa teknik dalam *hierarchical clustering* termasuk *single linkage*, *complete linkage*, *average linkage*, dan *ward's minimum variance*, yang semuanya digunakan untuk mengukur jarak atau kesamaan antara *cluster*.

2. *Partitional clustering* menghasilkan berbagai partisi dan mengevaluasinya berdasarkan beberapa kriteria. Karena

output dari algoritma ini hanya berupa satu set *cluster*, maka pengguna harus memasukkan jumlah *cluster* (k) yang diinginkan. Salah satu algoritma metode ini yang sering digunakan adalah metode K-Means. Metode ini menetapkan anggota *cluster* berdasarkan pusat data (*centroid*) saat ini, kemudian memperbarui pusat data baru berdasarkan anggota *cluster* sebelumnya. Kedua langkah ini diulangi secara iteratif hingga mencapai optimalisasi fungsi objektivitas antar-*cluster*, yaitu ketika perbedaan antara *cluster* diminimalkan dan kesamaan dalam setiap *cluster* dimaksimalkan.

Tujuan utama dari metode *clustering* (klasterisasi) adalah untuk memahami struktur yang tersembunyi dalam data, mengidentifikasi pola-pola yang mungkin tidak terlihat secara langsung, serta membantu dalam analisis dan pengambilan keputusan lebih lanjut. Hal ini dapat membantu dalam segmentasi pelanggan, analisis pasar, pengelompokan data untuk prediksi, atau bahkan untuk reduksi dimensi dalam analisis data yang kompleks. Berikut adalah penjelasan mengenai tujuan-tujuan utama dilakukan klasterisasi:

1. Memahami struktur data yang tersembunyi
   Dalam banyak kasus, data yang dikumpulkan dapat memiliki struktur yang tidak langsung terlihat. Klasterisasi membantu dalam mengidentifikasi dan memvisualisasikan struktur ini, sehingga kita dapat memahami bagaimana data tersebut berinteraksi dan membentuk kelompok yang lebih homogen di dalamnya. Contoh: dalam analisis genomik, klasterisasi dapat mengungkapkan kelompok gen yang memiliki fungsi atau ekspresi serupa.

2. Identifikasi pola dalam data
   Klasterisasi dapat mengungkapkan pola-pola atau hubungan yang tidak jelas antara data. Dengan mengelompokkan data yang mirip, kita dapat mengidentifikasi tren atau pola umum yang mungkin tidak terdeteksi melalui analisis sederhana. Contoh: dalam analisis perilaku konsumen, klasterisasi bisa menunjukkan pola pembelian yang serupa di antara kelompok pelanggan yang berbeda.

3. Segmentasi pelanggan
   Dalam konteks bisnis, klasterisasi sering digunakan untuk segmentasi pelanggan, yaitu mengelompokkan pelanggan berdasarkan karakteristik tertentu seperti perilaku pembelian, demografi, atau preferensi produk. Dengan segmentasi, perusahaan dapat menargetkan kampanye pemasaran atau penawaran produk yang lebih tepat sesuai dengan kebutuhan dan keinginan tiap segmen. Contoh: mengidentifikasi pelanggan setia, pelanggan yang sensitif terhadap harga, atau pelanggan yang sering membeli produk premium.

4. Analisis pasar
   Klasterisasi membantu perusahaan atau peneliti dalam mengidentifikasi segmen pasar yang berbeda, sehingga mereka bisa mengembangkan strategi pemasaran atau produk yang sesuai untuk setiap segmen tersebut. Contoh: menentukan area geografis dengan permintaan produk yang tinggi atau rendah dan menyesuaikan strategi distribusi.

5. Pengelompokan data untuk prediksi
   Klasterisasi dapat digunakan sebagai langkah awal dalam

model prediksi, di mana data dikelompokkan terlebih dahulu sebelum digunakan untuk melatih model prediktif. Ini bisa meningkatkan akurasi model dengan memastikan bahwa model dilatih pada data yang lebih homogen. Contoh: mengelompokkan jenis pelanggan sebelum membangun model prediksi *churn* (berhenti berlangganan).

6. Reduksi dimensi dalam analisis data yang kompleks
   Klasterisasi juga dapat membantu dalam mengurangi kompleksitas data dengan menyederhanakan data menjadi kelompok-kelompok yang representatif. Ini berguna dalam analisis data besar atau multidimensional, di mana data mentah mungkin terlalu kompleks untuk dianalisis secara langsung. Contoh: dalam analisis teks, klasterisasi bisa digunakan untuk mengelompokkan dokumen ke dalam topik-topik utama sebelum analisis lebih lanjut.

2.5 Metode K-Means

K-Means adalah salah satu metode *clustering* (klasterisasi) bertipe partisional yang banyak digunakan untuk pengelompokan data. Data dengan karakteristik yang serupa dikelompokkan dalam satu *cluster*, sementara data dengan karakteristik berbeda dikelompokkan ke dalam *cluster* lain. Hal ini memastikan bahwa data dalam satu *cluster* memiliki variasi yang relatif kecil. Metode ini memecah sekumpulan data menjadi k-*cluster* untuk mendefinisikan pusat data (*centroid*) *cluster* sejumlah k dan setiap *cluster* memiliki satu pusat data.

Menurut MacQueen J.B (1967), K-Means dikenal sebagai metode *clustering* (klasterisasi) yang paling populer dan banyak

digunakan karena kesederhanaannya, kemudahan penerapan, serta kemampuannya untuk menangani data berukuran besar.

Langkah-langkah penerapan *clustering* (klasterisasi) dengan metode K-Means (Wulan dan Sudin, 2021) adalah:

1. Menentukan nilai k untuk jumlah *cluster* yang akan dibentuk.
2. Menentukan nilai untuk pusat *cluster* awal (*centroid*) sebanyak k. *Centroid* pertama dipilih secara acak.
3. Menghitung jarak setiap data ke masing-masing *centroid* menggunakan rumus jarak, hingga ditemukan jarak terdekat untuk setiap data. Ada berbagai cara untuk mengukur jarak antar suatu objek data dengan objek data yang lain, salah satunya adalah *euclidean distance*.

   Euclidean distance adalah menghitung akar pangkat dua dari perbandingan nilai dari sepasang objek. Rumus *euclidean distance* adalah

   $$D(i,j) = \sqrt{(X_{1_j} - X_{1_i})^2 + (X_{2_j} - X_{2_i})^2 + ... + (X_{k_j} - X_{k_i})^2} \tag{2.1}$$

   dengan

   $D(i,j) =$ jarak dari data ke $-\ i$ sampai pusat data (*centroid*) *cluster*-j
   $X_{k_i} =$ data ke $-\ i$ pada variabel/atribut data ke $-\ k$
   $X_{k_j} =$ data ke $-\ j$ pada variabel/atribut data ke $-\ k$

4. Mengelompokkan data ke dalam *cluster* sesuai dengan jarak terpendek atau terdekat antara data dengan *centroid*.

5. Menentukan *centroid* baru dengan menghitung rata-rata nilai data di setiap *cluster* berdasarkan pengelompokan sebelumnya.

6. Memeriksa apakah terjadi perubahan pada *centroid* atau apakah objek data berpindah *cluster*, jika iya, proses dilanjutkan ke iterasi berikutnya. Namun, jika tidak, *cluster* terakhir dicatat sebagai hasil akhir dari pengelompokan objek data yang dihasilkan.

Masalah yang akan muncul ketika menggunakan metode K-Means biasanya adalah dengan ditemukannya berbagai model klasterisasi yang berbeda tergantung pada pemilihan *centroid* awal, pemilihan jumlah *cluster*-k pada iterasi pertama yang kurang tepat dapat memengaruhi hasil akhir, algoritma mungkin tidak mencapai titik konvergensi jika *centroid* tidak diperbarui dengan baik, keberadaan data *outlier* dapat memengaruhi posisi *centroid* dan hasil *clustering*, K-Means cenderung menghasilkan *cluster* berbentuk bulat sehingga sulit menangani bentuk *cluster* yang lebih kompleks, dan data yang saling tumpang tindih dapat menyulitkan penentuan batas *cluster* yang jelas.

Beberapa kelebihan yang dimiliki oleh metode K-Means antara lain:

1. Mudah untuk diimplementasikan dan dijalankan: algoritma ini sederhana dan dapat dengan mudah diterapkan pada berbagai dataset.

2. Waktu yang dibutuhkan untuk menjalankan metode ini relatif cepat: proses eksekusi relatif cepat, sehingga efisien untuk digunakan pada data besar.

3. Mudah untuk diadaptasi: K-Means dapat dengan mudah disesuaikan untuk berbagai jenis data dan aplikasi.

4. Umum digunakan: metode ini banyak digunakan di berbagai bidang, menjadikannya sebagai salah satu metode *clustering* yang paling populer.

Sedangkan kekurangannya adalah sebagai berikut:

1. Sebelum menjalankan algoritma, nilai k dipilih secara acak, yang dapat menghasilkan pengelompokan data yang berbeda-beda. Jika nilai acak yang dipilih tidak optimal, hasil pengelompokkan pun akan kurang memuaskan.

2. Apabila data *training* memiliki dimensi tinggi, misalnya sebuah dataset dengan 20 variabel maka akan ada 20 dimensi. K-Means mencari jarak terdekat antara k dan titik lainnya, yang menjadi sulit pada data berdimensi tinggi. Masalah ini dikenal sebagai "*curse of dimensionality*".

3. Ketika terdapat banyak titik data, proses pencarian dan perhitungan jarak yang paling dekat dapat memakan waktu yang cukup lama, yang dapat mempengaruhi efisiensi algoritma.

2.6 Reduksi Dimensi Data dengan PCA

Principal Component Analysis (PCA) adalah teknik yang digunakan untuk mereduksi dimensi data. Metode ini bertujuan untuk menyederhanakan kumpulan data yang besar menjadi beberapa variabel yang lebih sedikit, tetapi tetap mempertahankan sebagian besar informasi penting dari data tersebut (Septa, 2016).

PCA bekerja dengan menganalisis hubungan antar fitur melalui matriks kovariansi, lalu mengekstrak pola atau kesamaan dalam data. Dari analisis ini, PCA menghasilkan komponen utama yang mewakili variabilitas data. Komponen-komponen ini kemudian digunakan untuk memproyeksikan data asli ke dalam ruang fitur baru yang lebih ringkas, dengan harapan sebagian besar informasi awal tetap terjaga meskipun jumlah fitur berkurang.

Prinsip kerja dalam PCA adalah sebagai berikut:

1. Standarisasi data sebelum dilakukan PCA. Data biasanya dinormalisasi atau distandarisasi agar setiap variabel memiliki rata-rata 0 dan variansi 1. Ini penting karena PCA sensitif terhadap skala variabel.

2. Kovarians atau matriks korelasi, yaitu PCA menghitung matriks kovarians atau korelasi untuk memahami bagaimana setiap variabel berhubungan satu sama lain. Jika variabel-variabel memiliki korelasi yang tinggi, mereka dapat direpresentasikan dengan satu komponen utama.

3. *Eigenvector* dan *eigenvalue*, yaitu PCA menghitung *eigenvector* dan *eigenvalue* dari matriks kovarians. *Eigenvector* menunjukkan arah komponen utama, sementara *eigenvalue* menunjukkan jumlah variasi atau informasi yang dijelaskan oleh setiap komponen utama. Komponen pertama akan memiliki variasi terbesar, diikuti oleh komponen berikutnya, yang memiliki variasi yang lebih sedikit.

4. Transformasi data, adalah data asli yang kemudian diproyeksikan ke dalam ruang baru yang ditentukan oleh *eigenvector*. Ruang baru ini disebut *principal components*

(komponen utama). Misalnya, dari sekumpulan variabel dengan dimensi tinggi, PCA dapat mereduksi data menjadi hanya dua atau tiga komponen utama yang paling signifikan, yang mewakili sebagian besar informasi dalam data.

5. Pemilihan jumlah komponen adalah salah satu keputusan penting dalam PCA. Untuk memilih jumlah komponen yang akan digunakan biasanya dilakukan dengan memeriksa *scree plot*, yang menunjukkan *eigenvalue* versus jumlah komponen. Titik lutut pada plot (tempat perubahan nilai *eigenvalue* melambat secara signifikan) menunjukkan jumlah komponen optimal yang dapat dipertahankan (Jolliffe, 2016).

Manfaat dari dilakukannya reduksi dimensi data dengan PCA adalah sebagai berikut:

1. Membantu mengurangi kompleksitas data tanpa kehilangan terlalu banyak informasi.
2. Dengan mereduksi data ke dalam dua atau tiga dimensi, PCA memudahkan visualisasi *cluster* atau pola dalam data.
3. PCA menghilangkan masalah multikolinearitas, karena komponen utama adalah ortogonal (tidak berkorelasi satu sama lain).
4. Dengan mengurangi jumlah fitur, PCA dapat mempercepat kinerja algoritma yang membutuhkan komputasi intensif (Abdi dan Williams, 2010).

2.7 Metode *Silhouette*

Metode *silhouette* adalah teknik evaluasi yang umum digunakan untuk mengukur kualitas dan kekuatan *cluster*. Metode ini membantu untuk menilai seberapa baik data ditempatkan dalam suatu *cluster* dengan membandingkan jarak antar data dalam *cluster* yang sama dengan jarak data ke *cluster* lain. Nilai *silhouette* yang tinggi menunjukkan bahwa data berada dalam *cluster* yang tepat, sedangkan nilai yang rendah mengindikasikan bahwa data mungkin tidak ditempatkan dengan benar (Rousseeuw, 1987).

Dalam metode *silhouette*, klasterisasi yang terbaik adalah klasterisasi yang memiliki nilai rata-rata *silhouette* tertinggi di antara semua hasil klasterisasi yang dibandingkan. Nilai rata-rata *silhouette* menunjukkan seberapa baik objek data dikelompokkan:

1. Rata-rata *silhouette* tinggi (dekat +1), menunjukkan bahwa sebagian besar objek data ditempatkan dalam *cluster* yang tepat dan memiliki jarak yang jauh dari *cluster* lainnya. Ini berarti klasterisasi tersebut sangat baik.

2. Rata-rata *silhouette* rendah (dekat 0), menunjukkan bahwa objek data berada di batas antara dua *cluster* atau *cluster* yang terbentuk tidak jelas.

3. Rata-rata *silhouette* negatif (-1), menunjukkan bahwa banyak objek data lebih dekat ke *cluster* lain daripada *cluster* yang ditempati. Ini menandakan klasterisasi yang buruk.

Cara kerja dari metode *silhouette* adalah sebagai berikut:

1. Hitung jarak rata-rata intra-klaster ($a(i)$), yaitu rata-rata

jarak antara data poin dengan semua data poin lainnya di *cluster*nya sendiri.

2. Hitung jarak rata-rata antar-klaster ($b(i)$), yaitu rata-rata jarak antara data poin dengan semua data poin di *cluster* terdekat (yang bukan *cluster*nya sendiri).

3. Hitung nilai *silhouette* untuk setiap data poin dengan rumus berikut:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{2.2}$$

Di mana:

$S(i) =$ skor *silhouette* untuk data poin ke-i.
$a(i) =$ rata-rata jarak ke data poin dalam *cluster*nya sendiri.
$b(i) =$ rata-rata jarak ke data poin di *cluster* terdekat.

4. Setelah menghitung nilai *silhouette* untuk setiap poin, skor rata-rata seluruh dataset dapat digunakan untuk mengevaluasi kualitas klastering secara keseluruhan. Semakin tinggi rata-rata skor *silhouette*, semakin baik hasil klastering.

2.8 Penelitian Terdahulu

Penelitian utama yang dijadikan acuan bagi peneliti adalah penelitian berjudul '*Research and Application of Data Mining in College Students Employment Based on Association Rules*' karya Liu dan Zhai (2017). Penelitian tersebut mengambil data informasi pekerjaan lulusan perguruan tinggi sebagai objek

penelitian dan menggunakan algoritma Apriori sebagai gagasan utama, mengembangkan sistem data mining sebagai pedoman informasi pekerjaan siswa dan fungsi bimbingan guru serta pekerjaan, menemukan beberapa faktor yang mempengaruhi situasi pekerjaan, dan mewujudkan pedoman kepada siswa yang mungkin tidak dapat memperoleh pekerjaan dengan lancar dan guru yang mungkin tidak memiliki tingkat pekerjaan yang ideal. Sehingga mereka dapat memperoleh pengetahuan dan informasi yang lebih berharga, manajemen yang lebih baik, meningkatkan tingkat penyerapan tenaga kerja, dan meningkatkan daya saing perguruan tinggi.

Penelitian kedua berasal dari penelitian berjudul '*Predictive Analysis of Heart Disease Using K-Means and Apriori Algorithms*' karya Hadia, Devi, dan Nida (2019). Penelitian ini mengenai prediksi penyakit yang merupakan salah satu masalah penting yang sedang kita semua hadapi saat ini. Penelitian ini bertujuan untuk mengembangkan program berbasis data mining yang menganalisis data dalam jumlah besar dan mengekstraksi informasi yang dapat diubah menjadi pengetahuan yang berguna. Diperlukan sebuah alat analisis untuk mengembangkan sistem keputusan yang efektif yang bisa menggunakan data yang ada untuk memprediksi penyakit yang benar. Sehingga pada penelitian ini, peneliti memperkenalkan sistem otomatis yang meprediksi penyakit jantung menggunakan *clustering* (klasterisasi) K-Means dan Apriori yang dikombinasikan dengan *software* MATLAB. Kedua algoritma ini dapat digunakan untuk mendeteksi termasuk ke stadium mana penyakit yang diderita secara efektif. Tingkatan penyakit jantung yang diprediksikan adalah tingkat dengan resiko rendah, sedang, dan tinggi.

BAB 3

METODE PENELITIAN

3.1 Sumber Data dan Variabel Penelitian

1. Sumber data

   Data adalah sekumpulan fakta yang dikumpulkan dan disajikan untuk tujuan tertentu (Tika, 2006). Dalam penelitian ini, sumber data yang digunakan adalah sumber data primer, yang berarti data tersebut diambil langsung dari sumber pertama yang ada di lapangan. Termasuk sumber data primer adalah:

   (a) *Person* adalah seseorang yang dapat memberikan data berupa jawaban melalui kuesioner atau wawancara langsung.

   (b) *Place* adalah data yang diperoleh berdasarkan kondisi yang berlangsung berkaitan dengan masalah pada penelitian. Dalam hal ini, yang berkaitan dengan masalah pada penelitian yaitu mengetahui pengaruh latar belakang responden terhadap jenis pekerjaan yang didapat.

2. Variabel penelitian

   Variabel penelitian adalah ketentuan yang ditetapkan oleh peneliti yang menjadi acuan dalam memperoleh sebuah informasi untuk kemudian ditarik kesimpulannya. Secara teori, variabel diartikan sebagai atribut yang memiliki variasi antara satu objek dengan objek lain. Dalam penelitian ini, variabel yang akan digunakan ditunjukkan pada tabel

berikut ini.

Tabel 3.1. Variabel penelitian

No	Variabel	Jenis
1.	Nama	Kategorial
2.	Usia	Numerik
3.	Jenis kelamin	Kategorial
4.	Kondisi ekonomi keluarga	Kategorial
5.	Pendidikan terakhir	Kategorial
6.	Hobi	Kategorial
7.	Capaian penghargaan	Kategorial
8.	Jenis penghargaan	Kategorial
9.	Status pekerjaan	Kategorial
10.	Jenis pekerjaan	Kategorial
11.	Golongan gaji	Kategorial
12.	Hubungan hobi dan pekerjaan	Kategorial
13.	Hubungan capaian penghargaan dan pekerjaan	Kategorial

3.2 Populasi dan Sampel Penelitian

1. Populasi adalah keseluruhan sasaran penelitian yang akan diteliti, dan hasil penelitian diberlakukan pada populasi tersebut (Kasiran, 2010). Populasi dalam penelitian ini adalah pengguna aplikasi X (*formerly* twitter) di Kota Semarang dan berumur 20-34 tahun. Rentang usia tersebut karena usia bagi tenaga kerja berada diantara 20 hingga 40 tahun, usia ini dianggap sangat produktif bagi tenaga kerja karena apabila usia dibawah 20 tahun rata-rata individu masih belum memiliki kematangan *skill* yang cukup selain itu juga masih dalam proses pendidikan. Sedangkan pada usia diatas 40 tahun mulai terjadi penurunan kemampuan fisik bagi individu (Yasin dan Priyono, 2016).

2. Sampel adalah bagian kecil dari populasi yang digunakan sebagai sumber data untuk analisis, mewakili karakteristik populasi secara keseluruhan. Teknik pengambilan sampel menggunakan rumus Slovin (Sugiyono, 2015) dengan rumus:

$$n = \frac{N}{1 + (N * e^2)} \tag{3.1}$$

dengan

$n =$ Jumlah sampel minimal

$N =$ Populasi

$e =$ *Error* (dalam penelitian ini digunakan *error* sebesar 10% karena merupakan tingkat kesalahan maksimal yang dapat ditolerir pada penelitian ini)

Jumlah populasi ditentukan oleh data BPS Kota Semarang pada tahun 2023, berikut gambaran data penduduk di Kota Semarang berdasarkan kelompok umur.

Tabel 3.2. Jumlah Penduduk berdasarkan Kelompok Umur

Kelompok umur	Laki-laki	Perempuan	Jumlah
0-4	58.283	55.856	114.139
5-9	61.668	58.657	120.325
10-14	64.069	60.588	124.657
15-19	66.362	62.598	128.960
20-24	66.326	63.127	129.453
25-29	64.287	63.338	127.625
30-34	65.092	65.768	130.860
35-39	66.058	67.252	133.310
40-44	66.632	68.567	135.199
45-49	62.767	65.733	128.500
50-54	54.656	58.675	113.331
55-59	46.880	51.968	98.848
60-64	38.221	42.922	81.143
65+	57.136	71.257	128.393

Populasi yang dimaksud dalam penelitian ini adalah data jumlah penduduk di Kota Semarang berdasarkan kelompok umur 20-34 tahun pada tahun 2023.

Tabel 3.3. Jumlah Penduduk Kelompok Umur 20-34 Tahun

Kelompok umur	Jumlah
20-24	129.453
25-29	127.625
30-34	130.860
Total	387.938

Maka dapat ditentukan sampel penelitian sebagai berikut,

$$n = \frac{N}{1 + (N * e^2)} = \frac{387.938}{1 + (387.938 * (0,1)^2)} = 99,974.. \approx 100$$

Sehingga jumlah sampel minimal dalam penelitian adalah 100.

3.3 Teknik Pengumpulan Data dan Instrumen Penelitian

1. Teknik pengumpulan data yang akan digunakan adalah penyebaran angket/kuesioner. Peneliti akan memberikan daftar pertanyaan pada responden yang memiliki kriteria atau syarat sesuai dengan yang dibutuhkan.

2. Instrumen penelitian/alat yang digunakan adalah angket/kuesioner. Tujuannya adalah untuk mendapatkan keterangan dari sampel, sumber, maupun responden yang bermacam-macam dan tidak memungkinkan peneliti untuk menemui responden secara langsung satu persatu. Berikut merupakan instrumen penelitian yang akan digunakan:

Tabel 3.4. Instrumen Penelitian

No	Pertanyaan	Jawaban
1.	Menyebutkan nama/username X	Isian sigkat
2.	Memilih kategori usia	20-24, 25-29, 30-34
3.	Memilih jenis kelamin	Laki-laki, Perempuan
4.	Menentukan termasuk ke dalam golongan manakah kondisi ekonomi keluarga yang dimiliki	Menengah ke atas, Menengah, Menengah ke bawah
5.	Memilih pendidikan yang terakhir ditempuh	SMA/SMK sederajat, Diploma, Sarjana, Lainnya
6.	Menuliskan hobi	Isian singkat
7.	Menentukan apakah pernah mendapat penghargaan	Ya, Tidak
8.	Menuliskan penghargaan yang didapat apabila pernah mendapatkannya	Isian singkat
9.	Memilih status pekerjaan (bekerja/tidak)	Ya, Tidak
10.	Menuliskan jenis pekerjaan yang dijalani apabila bekerja	Isian singkat
11.	Menentukan golongan gaji yang didapatkan setiap bulan	Bergaji tetap, Bergaji tidak tetap, Tidak bekerja
12.	Menentukan apakah terdapat hubungan antara hobi dengan pekerjaan yang dimiliki	Ya, Tidak, Tidak bekerja
13.	Menentukan apakah terdapat hubungan antara penghargaan yang pernah didapat dengan pekerjaan yang dimiliki	Ya, Tidak, Tidak pernah mendapat penghargaan, Tidak bekerja

3.4 Analisis Data

Teknik analisis data yang akan digunakan dalam penelitian ini adalah sebagai berikut:

1. Mengumpulkan dan menyeleksi data
   Dalam hal ini, peneliti akan mengumpulkan data menggunakan metode penyebaran angket/kuesioner kepada responden dengan kriteria tertentu.

2. Olah data
   Data yang telah diseleksi kemudian akan diolah menggunakan metode K-Means. Langkah-langkah yang akan dilakukan untuk mengolah data dengan metode ini adalah:

   (a) Mentransformasi bentuk data kategorial ke numerik.

   (b) Menormalisasi data untuk menyamakan skala dalam data dan mereduksi dimensi data agar variabel berkorelasi tinggi tidak memberikan informasi berulang.

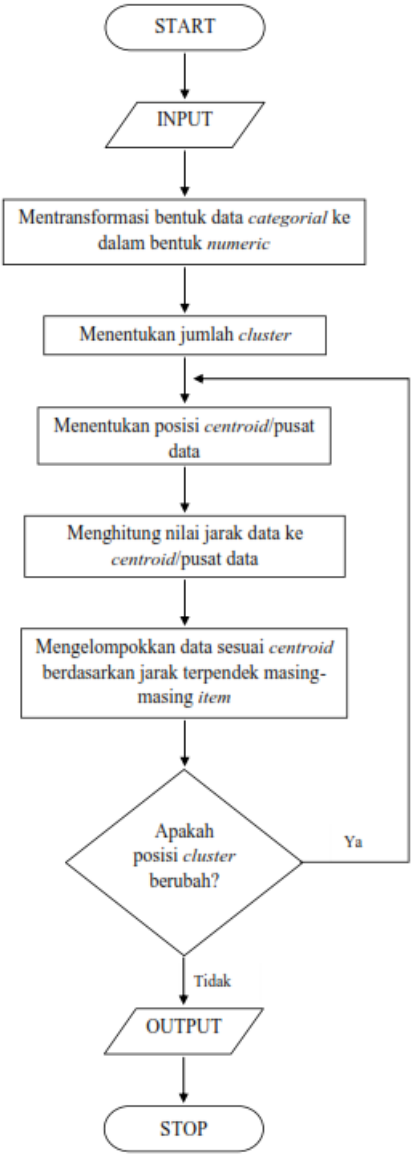
   (c) Menentukan jumlah *cluster* untuk iterasi pertama: jumlah *cluster* (dinotasikan sebagai k) dan posisi *centroid* (pusat data) pada iterasi awal akan ditentukan secara *random* (Marisa dan Ahmad, 2019). Untuk jumlah *cluster* akan ditentukan dengan metode *silhouette* agar tidak terjadi *overfitting* dalam proses menentukan jumlah *cluster*.

   (d) Menghitung nilai jarak data ke *centroid*/pusat data menggunakan metode *euclidean distances*.

(e) Mengelompokkan data sesuai *centroid*: setiap data dikelompokkan ke dalam *cluster* sesuai dengan jarak terdekat antara data dan *centroid*.

(f) Menentukan *centroid* untuk iterasi berikutnya: *centroid* diperbarui dengan mengkalkulasi rata-rata nilai data di setiap *cluster*, berdasarkan hasil pengelompokan sebelumnya.

(g) Mengulang kembali proses iterasi sebelumnya dengan data *centroid* yang berbeda. Iterasi akan terus dilakukan hingga tidak terjadi perubahan *centroid* atau posisi data dalam *cluster*.

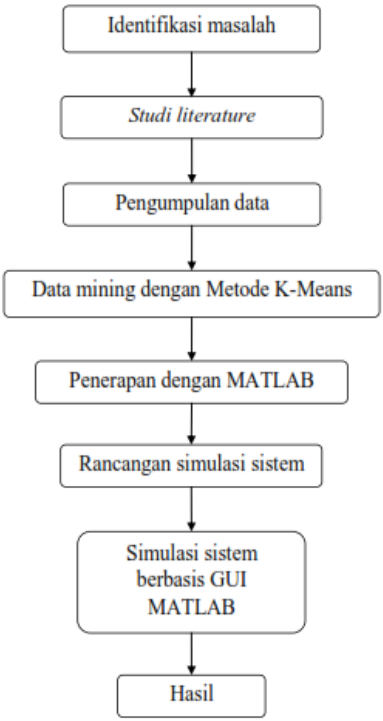


Gambar 3.1. *Flowchart* Metode K-Means

3. Pembentukan simulasi sistem berbasis GUI MATLAB
   Sistem akan dibuat sederhana, memiliki tampilan berbentuk kuesioner singkat dengan hasil akhir berupa penjelasan mengenai jenis pekerjaan yang direkomendasikan oleh sistem kepada pengguna.

3.5 Rancangan Penelitian

Rancangan penelitian akan direpresentasikan sebagai berikut:



Gambar 3.2. Alur Rancangan Penelitian

Tahapan dari susunan di atas akan dijelaskan sebagai berikut:

1. Identifikasi masalah, bertujuan untuk menentukan masalah yang akan diteliti sehubungan dengan judul penelitian yang diangkat, sehingga peneliti dapat mengetahui cara untuk menyelesaikan permasalahan tersebut.

2. *Studi literature* adalah tahap ketika peneliti membaca dan mempelajari buku, jurnal, internet, ataupun pernyataan berkaitan dengan masalah yang diteliti.

3. Pengumpulan data yang dilakukan adalah dengan menyebarkan kuesioner kepada responden yang memiliki kriteria sesuai dengan bentuk data yang dibutuhkan.

4. Data mining dengan metode K-Means. Pada tahap ini, peneliti akan menganalisa data menggunakan metode K-Means.

5. Penerapan dengan MATLAB versi 2013a untuk melakukan pengujian pada data.

6. Rancangan simulasi sistem, akan dirancang sebuah simulasi sistem mengenai rekomendasi pekerjaan berdasarkan data yang ada.

7. Sistem berbasis GUI MATLAB, rancangan simulasi sistem yang telah dibuat, kemudian akan direalisasikan menggunakan GUI berbasis MATLAB versi 2013a.

8. Hasil akan diperoleh setelah seluruh proses di atas dibuat. Maka terbentuklah sebuah *rule* dari data dan simulasi sistem GUI MATLAB yang dapat memberikan rekomendasi pekerjaan.

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Data

Data yang digunakan dalam penelitian ini telah dikumpulkan selama dua bulan, dari tanggal 12 Desember 2023 hingga 6 Februari 2024 di aplikasi sosial media "X" (*formerly* twitter) dengan menggunakan kuesioner online berupa Google Formulir. Banyak responden yang telah mengisi kuesioner tersebut adalah sebanyak 285 responden.

Tabel 4.1. Responden penelitian

No	Nama/*username*	Usia	...	Hubungan capaian penghargaan dengan pekerjaan
1.	thisisuraccount	21	...	Tidak ada hubungannya
2.	sya	27	...	Saya tidak bekerja
3.	sania	22	...	Saya tidak bekerja
4.	N	20	...	Saya tidak bekerja
5.	violet	23	...	Saya tidak bekerja
6.	Dwi Ayu N.	20	...	Saya tidak bekerja
7.	fasya	21	...	Saya tidak pernah mendapat penghargaan
8.	JZA	21	...	Saya tidak bekerja
9.	yolan	22	...	Saya tidak pernah mendapat penghargaan
...	...	...	...	...
285.	athereais	22	...	Ya, berhubungan

Berikut ini diuraikan gambaran umum responden dalam penelitian.

1. Responden berdasarkan usia

Berdasarkan usia, responden dalam penelitian ini dapat digambarkan seperti yang ditampilkan pada tabel berikut:

Tabel 4.2. Responden berdasarkan usia

Usia	Frekuensi	Persentase
20	32	11,22%
21	83	29,12%
22	65	22,8%
23	31	10,9%
24	31	10,9%
25	13	4,56%
26	7	2,45%
27	10	3,5%
28	5	1,75%
29	1	0,35%
30	2	0,7%
31	1	0,35%
32	2	0,7%
33	1	0,35%
34	1	0,35%

2. Responden berdasarkan jenis kelamin
   Berdasarkan jenis kelamin, responden dalam penelitian ini dapat digambarkan seperti yang ditampilkan pada tabel berikut:

   Tabel 4.3. Responden berdasarkan jenis kelamin

Jenis kelamin	Frekuensi	Persentase
Laki-laki	31	10,9%
Perempuan	254	89,1%

3. Responden berdasarkan kondisi ekonomi keluarga
   Kondisi ekonomi keluarga yang dimaksud dalam penelitian ini adalah perbedaan kelas sosial responden. Kelas sosial

mengacu pada perbedaan dalam masyarakat berdasarkan materi yang dimiliki. Menurut Arifin (2007), kelas sosial dibagi menjadi tiga golongan, yaitu:

(a) Kelas atas (*upper class)*. Kelompok ini terdiri dari konglomerat, eksekutif, dan individu dengan kekayaan yang signifikan.

(b) Kelas menengah (*middle class)*. Terdiri dari kaum profesional atau individu dengan profesi tetap, serta pemilik usaha kecil atau toko.

(c) Kelas bawah (*lower class)*. Kelompok ini terdiri dari mereka yang menerima pendapatan yang lebih sedikit daripada kebutuhan pokoknya, seringkali sebagai imbalan atas kerja yang dilakukan.

Jadi, dapat digambarkan sebagai berikut:

Tabel 4.4. Responden berdasarkan kondisi ekonomi keluarga

Kondisi ekonomi keluarga	Frekuensi	Persentase
Kelas atas	1	0,35%
Kelas menengah	237	83,15%
Kelas bawah	47	16,5%

4. Responden berdasarkan pendidikan terakhir

   Berdasarkan pendidikan terakhir, responden dalam penelitian ini dapat digambarkan seperti yang ditampilkan pada tabel berikut:

Tabel 4.5. Responden berdasarkan pendidikan terakhir

Pendidikan terakhir	Frekuensi	Persentase
SMP	2	0,7%
SMA/SMK Sederajat	164	57,6%
Diploma	14	4,9%
Sarjana	103	36,1%
Magister	2	0,7%

5. Responden berdasarkan capaian penghargaan
   Berdasarkan pernah atau tidaknya responden mendapatkan penghargaan digambarkan sebagaimana terlihat dalam tabel berikut:

Tabel 4.6. Responden berdasarkan capaian penghargaan

Capaian penghargaan	Frekuensi	Persentase
Pernah mendapat penghargaan	67	23,5%
Tidak pernah mendapat penghargaan	218	76,5%

6. Responden berdasarkan status pekerjaan
   Berdasarkan status pekerjaan, responden dalam penelitian ini dapat digambarkan seperti yang ditampilkan pada tabel berikut:

Tabel 4.7. Responden berdasarkan status pekerjaan

Status pekerjaan	Frekuensi	Persentase
Bekerja	199	69,8%
Tidak bekerja	86	30,2%

7. Responden berdasarkan golongan gaji
   Berdasarkan golongan gaji yang dimaksud di sini adalah responden memiliki gaji tetap (sama setiap bulan) atau tidak

tetap, responden dalam penelitian ini dapat digambarkan seperti yang ditampilkan pada tabel berikut:

Tabel 4.8. Responden berdasarkan golongan gaji

Golongan gaji	Frekuensi	Persentase
Bergaji tetap (sama setiap bulan)	110	38,6%
Bergaji tidak tetap	89	31,2%
Tidak bekerja	86	30,2%

8. Responden berdasarkan hubungan hobi dan pekerjaan

   Berdasarkan ada atau tidaknya hubungan antara hobi yang dimiliki responden dengan bidang pekerjaan yang dijalani dapat digambarkan sebagaimana terlihat dalam tabel berikut:

Tabel 4.9. Responden berdasarkan hubungan hobi dan pekerjaan

Hubungan hobi dan pekerjaan	Frekuensi	Persentase
Memiliki hubungan	61	21,4%
Tidak memiliki hubungan	138	48,4%
Tidak bekerja	86	30,2%

9. Responden berdasarkan hubungan capaian penghargaan dan pekerjaan

   Berdasarkan ada atau tidaknya hubungan antara penghargaan yang didapat responden dengan bidang pekerjaan yang dijalani dapat digambarkan sebagaimana terlihat dalam tabel berikut:

Tabel 4.10. Responden berdasarkan hubungan capaian penghargaan dengan pekerjaan

Hubungan capaian penghargaan dengan pekerjaan	Frekuensi	Persentase
Memiliki hubungan	22	7,7%
Tidak memiliki hubungan	27	9,5%
Tidak pernah mendapat penghargaan	150	52,6%
Tidak bekerja	86	30,2%

4.2 *Preprocessing Data Clustering* (Klasterisasi)

Preprocessing data adalah tahapan untuk menyiapkan data agar siap untuk diolah. Berikut tahap kerja *preprocessing data*:

1. *Data cleaning* adalah proses pembersihan data yang dilakukan melalui beberapa langkah, seperti mengisi nilai yang hilang, menghaluskan data yang memiliki *noise*, serta menyelesaikan inkonsistensi yang ditemukan dalam dataset. Proses ini penting untuk memastikan data yang digunakan dalam analisis berkualitas dan akurat. Contohnya dengan memperbaiki kesalahan format, seperti mengonversi semua tanggal ke format yang seragam.

2. *Data integration* yaitu data yang memiliki representasi berbeda akan disatukan dan semua masalahnya akan diselesaikan. Contohnya adalah menggabungkan data perilaku pengguna dari website dan aplikasi *mobile* dengan tujuan untuk mendapatkan pandangan holistik.

3. *Data transformation* yaitu data yang telah dikumpulkan akan dinormalisasi dan digeneralisasikan. Contohnya adalah dengan mengonversi data teks dari ulasan produk menjadi

format yang dapat diproses algoritma.

4. *Data reduction* yaitu mengubah data ke dalam bentuk pola, fokus, kategori sehingga data menjadi ringkas dan rinci. Contohnya adalah memilih fitur yang paling relevan dengan sistem rekomendasi, seperti riwayat pembelian terkini dan rating produk.

Tahap *preprocessing data* yang dilakukan adalah *cleaning* dan *transformation*. *Data cleaning* digunakan untuk membersihkan inkonsistensi pada data. Seperti pada pertanyaan "apakah Anda pernah mendapat penghargaan?", apabila responden menjawab "tidak pernah" maka pada pertanyaan "adakah hubungan antara penghargaan yang didapat dengan pekerjaan saat ini?" jawabannya adalah "Saya tidak pernah mendapat penghargaan". Sedangkan pada tahap *data transformation*, dilakukan transformasi data dari bentuk kategorial menjadi bentuk numerik agar lebih mudah diolah.

4.3 Penerapan Metode K-Means

Perlu dilakukan beberapa tahapan dalam menerapkan metode K-Means pada data. Data yang ada tidak bisa langsung diklasterisasi karena masih berupa data mentah, jika tetap dilakukan klasterisasi tanpa mengolah data terlebih dulu, maka hasil klasterisasinya tidak akan optimal. Tahapan dalam penerapan metode K-Means adalah sebagai berikut:

1. Untuk pengolahan data sebelum proses klasterisasi, akan dilakukan beberapa tahap seperti pemilihan data yang akan

diolah, normalisasi data, dan reduksi dimensi data dengan PCA.

2. Untuk klasterisasi data dengan K-Means, tahap yang akan dilakukan adalah sebagai berikut:

   (a) Penentuan *optimal cluster* dengan metode *silhouette*.

   (b) Menghitung jarak antara setiap data dan *centroid* (pusat data) menggunakan *euclidean distance*

   (c) Mengelompokkan data sesuai *centroid* sesuai jarak terpendek masing-masing data.

   (d) Menentukan *centroid* untuk iterasi-2: setelah data dikelompokkan berdasarkan jarak terdekat ke *centroid*, *centroid* diperbarui untuk iterasi berikutnya dengan mengkalkulasi rata-rata dari penjumlahan nilai data untuk setiap kelompok *cluster*.

   (e) Mengulang kembali proses iterasi sebelumnya dengan data *centroid* yang berbeda, hingga tidak terjadi perubahan posisi *cluster* atau data konvergen.

4.3.1 Pemilihan Data

Dalam pengolahan data, hanya data para pekerja yang digunakan yaitu sejumlah 199 data (berada dalam lampiran) karena penelitian ini bertujuan untuk menyusun simulasi sistem rekomendasi pekerjaan. Variabel pada data akan diubah sesuai dengan kebutuhan, seperti pada kolom "nama" akan diubah menjadi angka 1-199 sesuai dengan urutan data, karena hanya data pekerja yang akan digunakan, maka kolom "status pekerjaan" akan dihilangkan, dan pada kolom "jenis penghargaan yang diraih"

tidak akan digunakan karena hanya sebagai acuan peneliti untuk menentukan isi jawaban responden sesuai atau tidak.

Dilakukan transformasi data yang mengubah bentuk data kategorial menjadi numerik untuk memudahkan proses perhitungan. Berikut transformasi data dalam penelitian.

1. Kategori umur ditransformasikan menjadi tiga bagian, untuk umur 20-24 tahun akan diubah menjadi angka 1, untuk umur 25-29 tahun diubah menjadi angka 2, dan untuk umur 30-34 tahun menjadi angka 3.

2. Kategori jenis kelamin ditransformasi menjadi angka 1 untuk laki-laki dan 2 untuk perempuan.

3. Kategori kondisi ekonomi ditransformasi menjadi angka 1 untuk kelas bawah, 2 untuk kelas menengah, dan 3 untuk kelas atas.

4. Kategori pendidikan terakhir ditransformasi menjadi angka 1 untuk SMP, 2 untuk SMA/SMK Sederajat, 3 untuk Diploma, 4 untuk Sarjana, dan 5 untuk Magister.

5. Kategori capaian penghargaan ditransformasi menjadi angka 1 untuk pernah dan 2 untuk tidak pernah.

6. Kategori golongan gaji ditransformasi menjadi angka 1 untuk responden bergaji tetap dan 2 untuk bergaji tidak tetap.

7. Kategori hubungan hobi dengan pekerjaan ditransformasi menjadi angka 1 untuk berhubungan dan 2 untuk tidak berhubungan.

8. Kategori hubungan penghargaan dengan pekerjaan ditransformasi menjadi angka 1 untuk berhubungan, 2 untuk tidak berhubungan, dan 3 untuk tidak pernah mendapat penghargaan.

Selain bentuk transformasi data yang telah disebutkan di atas, terdapat transformasi tambahan pada variabel "jenis pekerjaan" dan "hobi". Peneliti telah membagi hasil jawaban responden ke dalam beberapa kategori sesuai jenis pekerjaan dan hobi yang telah disebutkan untuk memudahkan proses transformasi data dan perhitungannya. Jenis pekerjaan yang telah disebutkan oleh para responden dibagi menjadi 15 kategori, yaitu:

1. Wiraswasta adalah orang yang memiliki dan mengelola usaha sendiri.
2. *Freelance* adalah pekerjaan lepas yang dilakukan secara mandiri tanpa terikat pada satu perusahaan.
3. Staff perusahaan merupakan orang yang bekerja sama untuk membantu pimpinannya dalam mengelola sebuah perusahaan atau usaha, umumnya bekerja dalam kantor dan menyelesaikan urusan administrasi.
4. Pegawai Negeri Sipil (PNS) adalah Warga Negara Indonesia yang diangkat secara permanen oleh pejabat pembina kepegawaian untuk menjalankan tugas di posisi pemerintahan.
5. Tenaga kesehatan (nakes) adalah orang yang mengabdikan diri dalam bidang kesehatan, memiliki pengetahuan dan keterampilan melalui pendidikan di bidang kesehatan.

6. *Part time* adalah salah satu jenis pekerjaan yang memiliki waktu kerja fleksibel atau setengah dari jam kerja normal.

7. Pegawai magang adalah seseorang yang bekerja pada perusahaan dalam jangka waktu tertentu dengan tujuan untuk memperoleh pengalaman kerja dan keterampilan yang dapat diaplikasikan dalam kehidupan kerja di masa depan.

8. *Entertainer* adalah seseorang yang bekerja untuk menghibur orang-orang.

9. Karyawan adalah orang yang bekerja di suatu lembaga, kemudian bisa mendapatkan gaji dan mengerjakan pekerjaan fisik.

10. Asisten adalah seseorang yang membantu atau memberi pertolongan dan dukungan kepada seseorang yang memiliki jabatan di atasnya.

11. Tenaga pendidik adalah semua orang yang terlibat dalam tugas-tugas pendidikan, termasuk guru, dosen, manajer, administrator, *supervisor*, dan pegawai.

12. *Chef* adalah juru masak profesional dengan keterampilan tinggi dalam mengelola dapur dan kegiatan memasak.

13. Pekerja Harian Lepas (PHL) adalah orang yang bekerja untuk perusahaan tanpa ikatan kontrak kerja yang mengerjakan tugas-tugas sederhana berdasarkan proyek yang diberikan perusahaan, tidak terikat pada jam kerja perusahaan dan memiliki waktu kerja serta volume pekerjaan yang tidak menentu.

14. Tentara Nasional Indonesia (TNI) adalah angkatan bersenjata Indonesia yang bertugas menjaga keutuhan wilayah, kedaulatan negara, kehormatan bangsa, dan keselamatan rakyat.

15. Lainnya. Kategori ini dibuat untuk data pekerjaan yang tidak sesuai atau tidak diisi.

Jenis hobi yang telah disebutkan oleh para responden dibagi menjadi 5 kategori, yaitu:

1. *Making* atau membuat yaitu hobi dengan kegiatan yang dilakukan untuk menghasilkan sesuatu. Contoh hobi dengan kategori tersebut dalam data adalah fotografi, membuat konten, memasak, desain, *research*, dan merajut.

2. *Activity* atau aktifitas yaitu hobi dengan kegiatan fisik di luar ruangan terutama di alam. Contoh hobi dengan kategori tersebut dalam data adalah keliling kota dengan motor, jalan-jalan, *traveling*, kulineran, memancing, bersepeda, dan bersosialisasi dengan kelompok tertentu.

3. *Play* atau bermain yaitu hobi dengan aktifitas fisik yang tidak terbatas pada tempat. Contoh hobi dengan kategori tersebut dalam data adalah yoga, renang, olahraga, senam, sepakbola, dan bulutangkis.

4. *Arts* atau seni yaitu hobi yang berkaitan dengan kesenian. Contoh hobi dengan kategori tersebut dalam data adalah menyanyi, bermain musik, balet, membaca, menonton, menulis cerpen, menari, melukis, *modeling*, dan menggambar.

5. Lainnya. Kategori ini dibuat untuk data hobi yang tidak sesuai atau tidak diisi.

4.3.2 Normalisasi Data

Normalisasi data adalah salah satu tahapan dalam *preprocessing data* yaitu *data transformation*. Normalisasi data yang digunakan dalam penelitian ini adalah standardisasi (*Z-score normalization*) yaitu mengubah setiap nilai dalam data sehingga memiliki rata-rata (mean) nol dan standar deviasi satu. Tujuan dilakukan normalisasi data adalah untuk menyamakan skala nilai dalam data agar skala data yang lebih besar tidak mendominasi hasil klasterisasi. Langkah-langkah dalam proses normalisasi data dengan *Z-score normalization* adalah sebagai berikut:

1. Setelah dilakukan proses transformasi data yang mengubah bentuk data kategorial ke bentuk numerik, maka didapatkan bentuk data baru, yaitu

Tabel 4.11. Data dalam bentuk numerik

No. data	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1.	1	2	2	2	4	2	2	2	2	3
2.	1	1	1	4	4	1	3	1	2	2
3.	1	2	1	2	5	2	5	1	2	3
4.	1	2	1	2	4	2	2	1	2	3
5.	1	2	2	4	4	1	7	2	2	2
...	...	...	...	...	...	...	...	...	...	...
199.	1	2	2	4	4	1	9	1	1	1

dengan V1 = "usia", V2 = "jenis kelamin", V3 = "kondisi ekonomi keluarga", V4 = "pendidikan terakhir", V5 = "hobi", V6 = "capaian penghargaan", V7 = "jenis pekerjaan", V8 = "golongan gaji", V9 = "hubungan antara hobi

dengan pekerjaan", dan V10 = "hubungan antara capaian penghargaan dengan pekerjaan".

2. Hitung rata-rata (mean) untuk setiap variabel mulai dari data ke-1 hingga data ke-199.

$$mean = \frac{\sum \text{nilai data pada variabel}}{\text{banyak data}} \tag{4.1}$$

Contoh pada variabel "usia" atau V1:

$$mean = \frac{\sum \text{nilai data pada variabel}}{\text{banyak data}}$$

$$mean = \frac{1+1+1+1+1+...+1}{199} = \frac{239}{199} = 1,201005025$$

Didapatkan rata-rata untuk setiap variabel sebagai berikut:

Tabel 4.12. Rata-rata (mean) untuk setiap variabel

Variabel	Rata-rata (mean)
V1	1,201005025
V2	1,884422111
V3	1,824120603
V4	2,889447236
V5	3,64321608
V6	1,753768844
V7	5,231155779
V8	1,447236181
V9	1,693467337
V10	2,64321608

3. Hitung standar deviasi populasi untuk setiap variabel mulai

dari data ke-1 hingga data ke-199.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} \tag{4.2}$$

dengan

$$\begin{aligned} \sigma &= \text{Standar deviasi populasi} \\ N &= \text{Banyak data} \\ x_i &= \text{Nilai pada data ke} - i \\ \mu &= \text{Rata-rata (mean) pada setiap variabel} \end{aligned}$$

Contoh pada variabel "usia" atau V1:

$$\begin{aligned} \sigma &= \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}} \\ \sigma &= \sqrt{\frac{(1 - 1,201005025)^2 + ... + (1 - 1,201005025)^2}{199}} \\ &= \sqrt{\frac{45,95979899}{199}} = \sqrt{0,230953764} = 0,480576491 \end{aligned}$$

Didapatkan standar deviasi populasi untuk setiap variabel sebagai berikut:

Tabel 4.13. Standar deviasi populasi untuk setiap variabel

Variabel	Standar deviasi populasi
V1	0,480576491
V2	0,319718065
V3	0,393695423
V4	0,991338974
V5	1,325566802
V6	0,430814779
V7	3,815708628
V8	0,497208185
V9	0,461053565
V10	0,671263497

4. Normalisasi data untuk setiap nilai data mulai dari data ke-1 hingga data ke-199 dan dari variabel "usia" (V1) hingga "hubungan antara capaian penghargaan dengan pekerjaan" (V10).

$$normalisasi = \frac{x_i - \mu}{\sigma} \tag{4.3}$$

dengan

$x_i =$ Nilai pada data ke $- i$

$\mu =$ Rata-rata (mean) pada setiap variabel

$\sigma =$ Standar deviasi populasi

Contoh pada nilai data ke-1 untuk variabel "usia" atau V1 dan nilai data ke-1 untuk variabel "jenis kelamin" atau V2.

$$normalisasi = \frac{x_i - \mu}{\sigma}$$

Nilai data ke-1 untuk variabel "usia" atau V1:

$$normalisasi = \frac{1 - 1,201005025}{0,480576491} = -0,418258131$$

Nilai data ke-1 untuk variabel "jenis kelamin" atau V2:

$$normalisasi = \frac{2 - 1,884422111}{0,319718065} = 0,361499403$$

Didapatkan normalisasi data untuk setiap nilai data sebagai berikut:

Tabel 4.14. Normalisasi data untuk setiap nilai data

No. data	V1	V2	...	V10
1.	-0,418258131	0,361499403	...	0,531510981
2.	-0,418258131	-2,766256299	...	-0,95821698
3.	-0,418258131	0,361499403	...	0,531510981
4.	-0,418258131	0,361499403	...	0,531510981
5.	-0,418258131	0,361499403	...	-0,95821698
...	...	...	...	...
199.	-0,418258131	0,361499403	...	-2,447944941

4.3.3 Reduksi Dimensi Data dengan PCA

Reduksi dimensi dengan PCA dilakukan karena dalam data terdapat banyak variabel (kolom data) yang memiliki korelasi yang tinggi, misalnya pada "pendidikan terakhir" dan "jenis pekerjaan", jika orang dengan latar belakang pendidikan tinggi maka akan cenderung memiliki status pekerjaan yang lebih tinggi. Sehingga, perlu dilakukan reduksi untuk menghindari data yang berkorelasi tinggi tersebut memberikan informasi berulang. PCA juga membantu dalam menormalisasi dan menyeimbangkan (menjaga konsistensi) pengaruh berbagai variabel pada model klasterisasi,

terutama pada data dalam penelitian ini yang memiliki skala yang berbeda.

Untuk proses reduksi PCA akan dilakukan dengan bantuan MATLAB karena prosesnya lebih cepat dan tepat dibandingkan dengan perhitungan manual yang belum tentu sesuai serta membutuhkan banyak waktu. *Script* program MATLAB untuk reduksi dimensi data dengan PCA berada dalam lampiran. Berikut hasil reduksi dimensi data dengan PCA menggunakan program MATLAB.

Tabel 4.15. Reduksi dimensi data

No. data	PCA1	PCA2
1.	0,465908	-1,48426
2.	-1,00889	0,586971
3.	1,451489	-0,49489
4.	1,206988	-1,04487
5.	-1,60503	0,872521
...	...	...
199.	-2,79502	2,378843

4.3.4 Penentuan Jumlah *Cluster* dan Posisi *Centroid* pada Iterasi Pertama

Pada penelitian ini, penentuan jumlah *cluster* dilakukan dengan menggunakan metode *silhouette*. Untuk penentuan *cluster* terbaik dengan *silhouette* akan dilakukan dengan bantuan MATLAB karena data yang akan diolah merupakan data yang cukup besar. *Script* program MATLAB untuk proses penentuan *cluster* terbaik dengan *silhouette* berada dalam lampiran.

4.3.5 Klasterisasi dengan Metode K-Means

Setelah mendapatkan jumlah *cluster* optimal, kemudian akan dilakukan perhitungan K-Means untuk membagi data ke dalam beberapa *cluster* yang sesuai dengan metode *silhouette*.

1. Inisialisasi dan pembentukan *cluster*
   Pada hasil perhitungan metode *silhouette* sebelumnya pada data, didapatkan hasil *cluster* optimalnya adalah tiga, maka *centroid* (pusat data) awal berjumlah tiga. Pemilihan *centroid* ini dilakukan secara acak, yaitu data ke-137 untuk *cluster* 1, 30 untuk *cluster* 2, dan 45 untuk *cluster* 3. Pemilihan acak dilakukan untuk menghindari bias dalam proses klasterisasi. Algoritma akan mengeksplorasi berbagai kemungkinan klasterisasi.

No. data	PCA1	PCA2
30.	0,499676	-1,33483
45.	-2,60062	0,66823
137.	1,13974	0,785964

2. *Centroid* dan iterasi
   Setelah menentukan jumlah *cluster* dan data yang akan digunakan sebagai *centroid*, proses yang akan dilakukan selanjutnya adalah proses iteratif. Langkah-langkah dalam proses ini adalah sebagai berikut:

   (a) Setiap nilai data yang telah direduksi PCA akan dihitung jaraknya ke tiap *centroid* dengan jarak euclidean.

$$D(i,j) = \sqrt{(X_{1_j} - X_{1_i})^2 + (X_{2_j} - X_{2_i})^2}$$

Contoh perhitungan jarak ke *centroid* pada nilai data ke-1.

$$
\begin{aligned}
D(1,137) &= \sqrt{(0,4659-1,1397)^2+(-1,4842-0,7859)^2} \\
&= 2,368114411 \\
D(1,30) &= \sqrt{(0,4659-0,4996)^2+(-1,4842-1,3348)^2} \\
&= 0,153192675 \\
D(1,45) &= \sqrt{(0,4659-(-2,6006))^2+(-1,4842-0,6682)^2} \\
&= 3,746568719
\end{aligned}
$$

Dari hasil di atas dapat dilihat bahwa jarak data ke-1 terhadap *centroid* memiliki nilai yang berbeda. Jarak data ke-1 pada *centroid* data ke-30 memiliki nilai paling kecil (jarak terpendek) di antara ketiga jarak data ke *centroid*, artinya data ke-1 akan masuk ke dalam *cluster* 2. Lakukan hal yang sama untuk seluruh nilai data.

(b) Mengelompokkan data berdasarkan jarak terpendek terhadap *centroid*. *Cluster* 1 berisi data-data yang memiliki jarak terpendek ke centroid data ke-137, *cluster* 2 berisi data-data yang memiliki jarak terpendek ke centroid data ke-30, dan *cluster* 3 berisi data-data yang memiliki jarak terpendek ke centroid data ke-45. Berikut ditunjukkan beberapa data yang berada dalam *cluster* 1, 2, dan 3.

Cluster 1	*Cluster* 2	*Cluster* 3
9	1	2
17	3	5
24	4	7
26	6	8
31	14	10
...	...	...
195	198	199

(c) Membentuk *centroid* baru berdasarkan rata-rata nilai data pada tiap *cluster*. Misalnya data pada *cluster* 1 berikut.

No. data	PCA1	PCA2
9	0,066760755	0,155579633
17	1,231184322	1,106993286
24	1,170727212	0,76188329
26	0,819698614	0,471549395
31	1,313923289	0,863585371
...	...	...
195	1,152622331	1,127881552
Mean	1,06670738	0,925760341

Untuk cara yang sama dilakukan pula pada data yang berada di *cluster* 2 dan 3. Maka didapatkan *centroid* baru, yaitu

Mean *cluster*	1	2
Cluster 1	1,06670738	0,925760341
Cluster 2	0,447341771	-1,243210131
Cluster 3	-2,317422169	0,485892535

(d) Lakukan perhitungan jarak data ke *centroid* baru, kemudian kelompokkan data ke dalam *cluster* yang sama berdasarkan jarak terpendek data terhadap *centroid* baru. Jika data yang berada dalam

cluster-cluster sebelumnya tidak berpindah *cluster* (hasil konvergen), maka iterasi dihentikan. Namun, jika terdapat data yang berpindah *cluster*, maka iterasi terus dilanjutkan dengan cara yang sama.

3. Visualisasi hasil klasterisasi

   Setelah didapatkan hasil data yang terdapat pada *cluster* 1, 2, dan 3, langkah selanjutnya adalah mengubah bentuk data reduksi dimensi PCA ke dalam bentuk data transformasi. Untuk memvisualisasikan hasil klasterisasi perlu memperhatikan kecenderungan (dominasi) dalam nilai data pada tiap *cluster* yang mewakili sebuah kategori untuk tiap variabel. Untuk mempermudah, maka data hasil iterasi K-Means sebelumnya yang menggunakan data reduksi PCA akan dikembalikan bentuknya menjadi data transformasi sebelum dilakukan reduksi. Misalnya pada data ke-9 yang terdapat di *cluster* 1. Pada tabel dinyatakan "V1" dan "V2" pada data bernilai "1", yang ketika dikembalikan ke bentuk sebelum transformasi, maka akan didapatkan keterangan bahwa "pada data ke-9 untuk variabel usia (V1), kategorinya adalah 20-24 tahun (1) dan untuk variabel jenis kelamin (V2), kategorinya adalah laki-laki (1)".

Tabel 4.16. Data hasil klasterisasi

Cluster 1										
No. data	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
9.	1	1	2	2	2	2	9	1	1	3
17.	1	1	2	4	5	2	6	1	2	3
24.	2	2	2	4	3	2	4	1	2	3
26.	1	2	2	2	2	2	9	1	2	3
31.	2	2	2	4	4	2	4	1	2	3
...	...	...	...	...	...	...	...	...	...	...
195.	1	2	2	3	4	2	9	1	2	3
Cluster 2										
No. data	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1.	1	2	2	2	4	2	2	2	2	3
3.	1	2	1	2	5	2	5	1	2	3
4.	1	2	1	2	4	2	2	1	2	3
6.	1	2	2	2	4	2	2	2	2	3
14.	1	2	2	2	5	2	1	2	1	3
...	...	...	...	...	...	...	...	...	...	...
198.	2	1	2	2	2	2	1	2	1	3
Cluster 3										
No. data	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
2.	1	1	1	4	4	1	3	1	2	2
5.	1	2	2	4	4	1	7	2	2	2
7.	1	2	2	2	4	1	2	1	2	2
8.	1	1	2	4	4	1	2	2	1	2
10.	1	2	2	2	4	1	2	2	1	2
...	...	...	...	...	...	...	...	...	...	...
199.	1	2	2	4	4	1	9	1	1	1

Kecenderungan (dominasi) nilai data yang muncul ini dimaksudkan untuk mengetahui seberapa banyak kategori pada variabel tertentu muncul dalam satu *cluster*. Misalnya, jumlah kategori usia 20-24 tahun (1) dalam *cluster* 1 adalah sebanyak 58 dari 75 data yang ada dalam *cluster*, dan sebagainya. Perlu diperhatikan pula persentase untuk tiap kategori dengan cara membagi banyak data untuk

tiap kategori dengan banyak data dalam *cluster*. Hal ini dikarenakan jumlah data dalam *cluster* berbeda-beda, yaitu 75 data untuk *cluster* 1, 75 data untuk *cluster* 2, dan 49 data untuk *cluster* 3. Selengkapnya akan dijelaskan sebagai berikut.

(a) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel pertama (usia) adalah sebagai berikut.

Tabel 4.17. Kecenderungan (dominasi) nilai data variabel usia

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) 20-24 tahun	58	63	45	166
(2) 25-29 tahun	14	9	3	26
(3) 30-34 tahun	3	3	1	7
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) 20-24 tahun	77,3333%	84%	91,8367%	
(2) 25-29 tahun	18,6667%	12%	6,1224%	
(3) 30-34 tahun	4%	4%	2,0408%	
Jumlah	100%	100%	100%	

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa kategori 1 (20-24 tahun), mendominasi data yang ada dalam *cluster* 3. Untuk kategori 2 (25-29 tahun), mendominasi data yang ada dalam *cluster* 1. Sedangkan untuk kategori 3 (30-34 tahun), mendominasi dua *cluster* yaitu *cluster* 1 dan 2.

(b) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel kedua (jenis kelamin) adalah sebagai berikut.

Tabel 4.18. Kecenderungan (dominasi) nilai data variabel jenis kelamin

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) Laki-laki	8	9	6	23
(2) Perempuan	67	66	43	176
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) Laki-laki	10,6667%	12%	12,2449%	
(2) Perempuan	89,3333%	88%	87,7551%	
Jumlah	100%	100%	100%	

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa *cluster* 1 didominasi oleh data dengan kategori 2 (perempuan), sedangkan data pada *cluster* 2 dan 3, didominasi oleh data dengan kategori 1 (laki-laki).

(c) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel ketiga (kondisi ekonomi keluarga) adalah sebagai berikut.

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa *cluster* 2 didominasi oleh kategori 1 (kelas bawah), *cluster* 1 didominasi oleh dua kategori yaitu kategori 2 (kelas menengah) dan 3 (kelas atas), sedangkan *cluster* 3 didominasi oleh kategori 2 (kelas menengah).

Tabel 4.19. Kecenderungan (dominasi) nilai data variabel kondisi ekonomi keluarga

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) Kelas bawah	7	23	6	36
(2) Kelas menengah	67	52	43	162
(3) Kelas atas	1	0	0	1
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) Kelas bawah	9,3333%	30,6667%	12,2449%	
(2) Kelas menengah	89,3333%	69,3333%	87,7551%	
(3) Kelas atas	1,3333%	0	0	
Jumlah	100%	100%	100%	

(d) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel keempat (pendidikan terakhir) adalah sebagai berikut.

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa *cluster* 1 didominasi oleh kategori 1 (SMP), 3 (diploma), dan 4 (sarjana). *Cluster* 2 didominasi oleh kategori 1 (SMP) dan kategori 2 (SMA/SMK Sederajat). Untuk *cluster* 3 didominasi oleh kategori 5 (magister).

Tabel 4.20. Kecenderungan (dominasi) nilai data variabel pendidikan terakhir

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) SMP	1	1	0	2
(2) SMA/SMK Sederajat	20	56	26	102
(3) Diploma	7	2	4	13
(4) Sarjana	46	16	18	80
(5) Magister	1	0	1	2
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) SMP	1,3333%	1,3333%	0	
(2) SMA/SMK Sederajat	26,6667%	74,6667%	53,0612	
(3) Diploma	9,3333%	2,6667%	8,1633%	
(4) Sarjana	61,3333%	21,3333%	36,7347%	
(5) Magister	1,3333%	0	2,0408%	
Jumlah	100%	100%	100%	

(e) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel kelima (hobi) adalah sebagai berikut.

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa kategori 1 (*making*/membuat) dan 4 (*arts*/seni), mendominasi *cluster* 3. Untuk kategori 2 (*activity*/aktifitas) dan 3 (*play*/bermain), mendominasi *cluster* 1 dan 2. Terakhir untuk kategori 5 (lainnya), mendominasi *cluster* 1.

Tabel 4.21. Kecenderungan (dominasi) nilai data variabel hobi

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) *Making* (membuat)	5	12	8	25
(2) *Activity* (aktifitas)	8	8	4	20
(3) *Play* (bermain)	5	5	3	13
(4) *Arts* (seni)	27	30	27	84
(5) Lainnya	30	20	7	57
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) *Making* (membuat)	6,6667%	16%	16,3265%	
(2) *Activity* (aktifitas)	10,6667%	10,6667%	8,1633%	
(3) *Play* (bermain)	6,6667%	6,6667%	6,1224%	
(4) *Arts* (seni)	36%	40%	55,1020%	
(5) Lainnya	40%	26,6667%	14,2857%	
Jumlah	100%	100%	100%	

(f) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel keenam (capaian penghargaan) adalah sebagai berikut.

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa *cluster* 3 berisi data dengan responden yang pernah menerima penghargaan (1), sedangkan *cluster* 1 dan 2 berisi data responden yang tidak pernah menerima penghargaan (2).

Tabel 4.22. Kecenderungan (dominasi) nilai data variabel capaian penghargaan

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) Pernah menerima	0	0	49	49
(2) Tidak pernah menerima	75	75	0	150
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) Pernah menerima	0	0	100%	
(2) Tidak pernah menerima	100%	100%	0	
Jumlah	100%	100%	100%	

(g) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel ketujuh (jenis pekerjaan) adalah sebagai berikut.

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa *cluster* 1 didominasi oleh kategori 4 (PNS), 5 (nakes), 9 (karyawan), 10 (asisten), 11 (tenaga pendidik), 12 (*chef*), dan 13 (PHL). Untuk *cluster* 2 didominasi oleh kategori 1 (wiraswasta) dan 2 (*freelance*). Sedangkan untuk *cluster* 3, didominasi oleh kategori 3 (staff perusahaan), 6 (*part time*), 7 (pegawai magang), 8 (*entertainer*), 14 (TNI), dan 15 (lainnya).

Tabel 4.23. Kecenderungan (dominasi) nilai data variabel jenis pekerjaan

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) Wiraswasta	1	27	6	34
(2) *Freelance*	0	22	10	32
(3) Staff perusahaan	10	15	16	41
(4) Pegawai Negeri Sipil (PNS)	6	0	0	6
(5) Tenaga kesehatan (nakes)	3	2	0	5
(6) *Part time*	1	1	1	3
(7) Pegawai magang	1	1	3	5
(8) *Entertainer*	1	1	1	3
(9) Karyawan	35	5	3	43
(10) Asisten	6	1	3	10
(11) Tenaga pendidik	7	0	3	10
(12) *Chef*	1	0	0	1
(13) Pekerja Harian Lepas (PHL)	2	0	0	2
(14) Tentara Nasional Indonesia (TNI)	0	0	1	1
(15) Lainnya	1	0	2	3
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) Wiraswasta	1,3333%	36%	12,2449%	
(2) *Freelance*	0	29,3333%	20,4082%	
(3) Staff perusahaan	13,3333%	20%	32,6531%	
(4) Pegawai Negeri Sipil (PNS)	8%	0	0	
(5) Tenaga kesehatan (nakes)	4%	2,6667%	0	
(6) *Part time*	1,3333%	1,3333%	2,0408%	
(7) Pegawai magang	1,3333%	1,3333%	6,1224%	
(8) *Entertainer*	1,3333%	1,3333%	2,0408%	
(9) Karyawan	46,6667%	6,6667%	6,1224%	
(10) Asisten	8%	1,3333%	6,1224%	
(11) Tenaga pendidik	9,3333%	0	6,1224%	
(12) *Chef*	1,3333%	0	0	
(13) Pekerja Harian Lepas (PHL)	2,6667%	0	0	
(14) Tentara Nasional Indonesia (TNI)	0	0	2,0408%	
(15) Lainnya	1,3333%	0	4,0816%	
Jumlah	100%	100%	100%	

(h) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel kedelapan (golongan gaji) adalah sebagai berikut.

Tabel 4.24. Kecenderungan (dominasi) nilai data variabel golongan gaji

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) Gaji tetap	68	18	24	110
(2) Gaji tidak tetap	7	57	25	89
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) Gaji tetap	90,6667%	24%	48,9796%	
(2) Gaji tidak tetap	9,3333%	76%	51,0204%	
Jumlah	100%	100%	100%	

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa kategori 1 (bergaji tetap) mendominasi data pada *cluster* 1 dan kategori 2 (bergaji tidak tetap) mendominasi data pada *cluster* 2 dan 3.

(i) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel kesembilan (hubungan dengan hobi) adalah sebagai berikut.

Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa kategori 1 (berhubungan dengan hobi) mendominasi data pada *cluster* 2 dan 3 dan kategori 2 (tidak berhubungan dengan hobi) mendominasi data pada *cluster* 1.

Tabel 4.25. Kecenderungan (dominasi) nilai data variabel hubungan dengan hobi

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) Berhubungan	12	25	24	61
(2) Tidak berhubungan	63	50	25	138
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) Berhubungan	16%	33,3333%	48,9796%	
(2) Tidak berhubungan	84%	66,6667%	51,0204%	
Jumlah	100%	100%	100%	

(j) Kecenderungan (dominasi) nilai data untuk tiap *cluster* pada variabel kesepuluh (hubungan dengan penghargaan) adalah sebagai berikut.

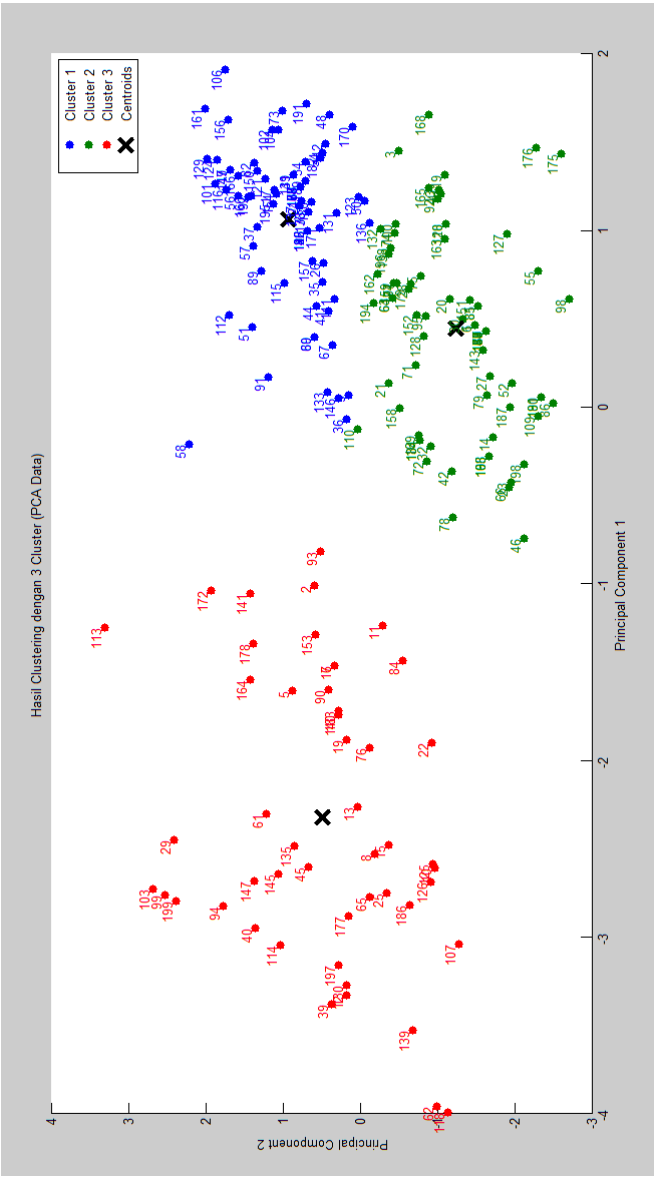
Jika dilihat dari persentase untuk setiap kategori pada masing-masing *cluster*, maka dapat disimpulkan bahwa *cluster* 3 berisi data dengan responden yang memiliki pekerjaan berhubungan (1) maupun tidak berhubungan (2) dengan capaian penghargaan yang diraih dan *cluster* 1 dan 2 berisi data responden yang tidak pernah menerima penghargaan (3).

Tabel 4.26. Kecenderungan (dominasi) nilai data variabel hubungan dengan penghargaan

Banyak data				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	Jumlah
(1) Berhubungan	0	0	22	22
(2) Tidak berhubungan	0	0	27	27
(3) Tidak pernah menerima penghargaan	75	75	0	150
Jumlah	75	75	49	199
Persentase				
Kategori	*Cluster* 1	*Cluster* 2	*Cluster* 3	
(1) Berhubungan	0	0	44,8980%	
(2) Tidak berhubungan	0	0	55,1020%	
(3) Tidak pernah menerima penghargaan	100%	100%	0	
Jumlah	100%	100%	100%	

4. Grafik klasterisasi

   Setelah melalui proses klasterisasi dengan metode K-Means, berikutnya dibuatlah grafik data hasil klasterisasi. Grafik tersebut merupakan hasil klasterisasi data menggunakan PCA yang membentuk tiga *cluster* dan untuk setiap *cluster* ditampilkan berdasarkan *principal component* 1 dan 2. Kedua komponen ini mewakili sebagian besar variasi dalam data, sehingga setiap titik mewakili data berdasarkan kombinasi linier dari variabel-variabel aslinya. Titik yang dekat satu sama lain menunjukkan data yang memiliki karakteristik serupa, sementara titik yang berjauhan menunjukkan perbedaan yang lebih besar di antara data tersebut. Angka-angka di setiap titik menunjukkan indeks/nomor data. Jadi, setiap nomor merepresentasikan satu data tertentu dalam dataset.



Gambar 4.1. Grafik Data Hasil Klasterisasi

Data yang ada dalam *cluster* 1 merupakan data untuk responden dengan karakteristik sebagai berikut:

(a) Berusia antara 25-34 tahun atau termasuk ke dalam kategori 2 dan 3 dalam variabel "usia".

(b) Berjenis kelamin perempuan atau termasuk ke dalam kategori 2 dalam variabel "jenis kelamin".

(c) Memiliki latar belakang ekonomi pada kelas menengah dan atas atau termasuk ke dalam kategori 2 dan 3 dalam variabel "kondisi ekonomi keluarga".

(d) Memiliki riwayat pendidikan terakhir yaitu setingkat SMP, diploma, dan sarjana atau termasuk ke dalam kategori 1, 3, dan 4 dalam variabel "pendidikan terakhir".

(e) Memiliki hobi *activity*/aktifitas, *play*/bermain, dan lainnya atau termasuk ke dalam kategori 2, 3, dan 5 dalam variabel "hobi".

(f) Tidak pernah mendapat penghargaan apapun yang bisa digunakan dalam proses pencarian kerja atau termasuk ke dalam kategori 2 dalam variabel "capaian penghargaan".

(g) Memiliki jenis pekerjaan yang beragam yaitu PNS (4), nakes (5), karyawan (9), asisten (10), tenaga pendidik (11), *chef* (12), dan PHL (13).

(h) Mendapatkan gaji tetap (sama setiap bulan) untuk pekerjaannya atau termasuk ke dalam kategori 1 dalam variabel "golongan gaji".

(i) Pekerjaan yang didapatkan tidak ada hubungannya

dengan hobi atau termasuk ke dalam kategori 2 dalam variabel "hubungan dengan hobi".

(j) Pekerjaan yang didapat tidak berhubungan dengan penghargaan karena responden tidak pernah mendapat penghargaan atau termasuk ke dalam kategori 3 dalam variabel "hubungan dengan penghargaan".

Data yang ada dalam *cluster* 2 merupakan data untuk responden dengan karakteristik sebagai berikut:

(a) Berusia antara 30-34 tahun atau termasuk ke dalam kategori 3 dalam variabel "usia".

(b) Berjenis kelamin laki-laki atau termasuk ke dalam kategori 1 dalam variabel "jenis kelamin".

(c) Memiliki latar belakang ekonomi pada kelas bawah atau termasuk ke dalam kategori 1 dalam variabel "kondisi ekonomi keluarga".

(d) Memiliki riwayat pendidikan terakhir yaitu setingkat SMP dan SMA/SMK Sederajat atau termasuk ke dalam kategori 1 dan 2 dalam variabel "pendidikan terakhir".

(e) Memiliki hobi *activity*/aktifitas dan *play*/bermain atau termasuk ke dalam kategori 2 dan 3 dalam variabel "hobi".

(f) Tidak pernah mendapat penghargaan apapun yang bisa digunakan dalam proses pencarian kerja atau termasuk ke dalam kategori 2 dalam variabel "capaian penghargaan".

(g) Memiliki jenis pekerjaan yaitu wiraswasta (1) dan *freelance* (2).

(h) Mendapatkan gaji tidak tetap untuk pekerjaannya atau termasuk ke dalam kategori 2 dalam variabel "golongan gaji".

(i) Pekerjaan yang didapatkan berhubungan dengan hobi atau termasuk ke dalam kategori 1 dalam variabel "hubungan dengan hobi".

(j) Pekerjaan yang didapat tidak berhubungan dengan penghargaan karena responden tidak pernah mendapat penghargaan atau termasuk ke dalam kategori 3 dalam variabel "hubungan dengan penghargaan".

Data yang ada dalam *cluster* 3 merupakan data untuk responden dengan karakteristik sebagai berikut:

(a) Berusia antara 20-24 tahun atau termasuk ke dalam kategori 1 dalam variabel "usia".

(b) Berjenis kelamin laki-laki atau termasuk ke dalam kategori 1 dalam variabel "jenis kelamin".

(c) Memiliki latar belakang ekonomi pada kelas menengah atau termasuk ke dalam kategori 2 dalam variabel "kondisi ekonomi keluarga".

(d) Memiliki riwayat pendidikan terakhir yaitu setingkat magister atau termasuk ke dalam kategori 5 dalam variabel "pendidikan terakhir".

(e) Memiliki hobi *making*/membuat dan *arts*/seni atau termasuk ke dalam kategori 1 dan 4 dalam variabel "hobi".

(f) Pernah mendapat penghargaan yang bisa digunakan dalam proses pencarian kerja atau termasuk ke dalam

kategori 1 dalam variabel "capaian penghargaan".

(g) Memiliki jenis pekerjaan yang beragam yaitu staff perusahaan (3), *part time* (6), pegawai magang (7), *entertainer* (8), TNI (14), dan lainnya (15).

(h) Mendapatkan gaji tidak tetap untuk pekerjaannya atau termasuk ke dalam kategori 2 dalam variabel "golongan gaji".

(i) Pekerjaan yang didapatkan berhubungan dengan hobi atau termasuk ke dalam kategori 1 dalam variabel "hubungan dengan hobi".

(j) Pekerjaan yang didapat ada yang berhubungan dan ada yang tidak berhubungan dengan penghargaan karena responden pernah mendapat penghargaan atau termasuk ke dalam kategori 1 dan 2 dalam variabel "hubungan dengan penghargaan".

5. Tingkat Akurasi Hasil Klasterisasi Data dengan Metode K-Means

   Dalam pengukuran tingkat akurasi pada hasil klasterisasi data, digunakan *silhouette score* sebagi metode pengukuran akurasi karena memberikan informasi tentang seberapa baik data dikelompokkan (Tan *et al*, 2005). Berikut dijelaskan interpretasi *silhouette score*:

   (a) *Score* 0.71 – 1.0 artinya pengelompokan yang sangat baik. Data dalam setiap *cluster* sangat kompak dan terpisah jauh dari *cluster* lain.

   (b) *Score* 0.51 – 0.7 artinya pengelompokan yang baik. Data dalam *cluster* cukup kompak, tetapi masih ada sedikit tumpang tindih antara *cluster*.

(c) *Score* 0.26 – 0.5 artinya pengelompokan yang agak baik. Terdapat beberapa *overlap* antara *cluster*, sehingga beberapa data mungkin tidak diklasifikasikan dengan sempurna.

(d) *Score* < 0.25 artinya pengelompokan yang buruk. Banyak data tidak cocok dengan *cluster*nya, dan *cluster* tidak terdefinisi dengan baik.

Untuk hasil klasterisasi pada penelitian ini, didapatkan tingkat akurasi berdasarkan rata-rata *silhouette score* sebesar 0,5874 atau 58,74% yang berarti pengelompokan data baik, tetapi masih ada tumpang tindih antara data dalam *cluster*.

4.4 Pembuatan Simulasi Sistem Rekomendasi Pekerjaan

Berdasarkan hasil visualisasi data hasil klasterisasi, maka dibuatlah simulasi sistem rekomendasi pekerjaan sesuai dengan hasil klasterisasi untuk setiap *cluster*. Dari seluruh hasil didapatkan 42 kombinasi data yang dapat di*input*kan dalam sistem oleh pengguna, yaitu:

1. Untuk *cluster* 1 terdapat 36 kombinasi data: 2 kategori "usia", 1 kategori "jenis kelamin", 2 kategori "kondisi ekonomi keluarga", 3 kategori "pendidikan terakhir", 3 kategori "hobi", dan 1 kategori "capaian penghargaan").

2. Untuk *cluster* 2 terdapat 4 kombinasi data: 1 kategori "usia", 1 kategori "jenis kelamin", 1 kategori "kondisi ekonomi keluarga", 2 kategori "pendidikan terakhir", 2 kategori "hobi", dan 1 kategori "capaian penghargaan").

3. Untuk *cluster* 3 terdapat 2 kombinasi data: 1 kategori "usia", 1 kategori "jenis kelamin", 1 kategori "kondisi ekonomi keluarga", 1 kategori "pendidikan terakhir", 2 kategori "hobi", dan 1 kategori "capaian penghargaan").

Berikut dijelaskan isi dari program simulasi sistem rekomendasi pekerjaan.

1. Tampilan awal

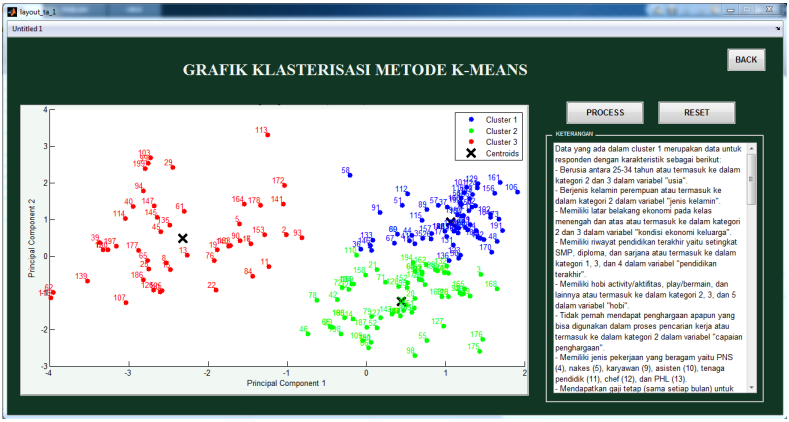
   Pada tampilan awal, ditunjukkan detail penelitian. Terdapat dua pilihan menu yang tersedia yaitu grafik hasil klasterisasi dan simulasi sistem rekomendasi pekerjaan.

Gambar 4.2. Tampilan Awal

2. Grafik hasil klasterisasi

   Pada bagian ini, akan ditunjukkkan hasil klasterisasi pada data yang telah digunakan dalam penelitian ini.

Gunanya adalah untuk menunjukkan pada pengguna pengelompokkan data dalam penelitian yang nantinya akan digunakan sebagai acuan pembuatan simulasi sistem.



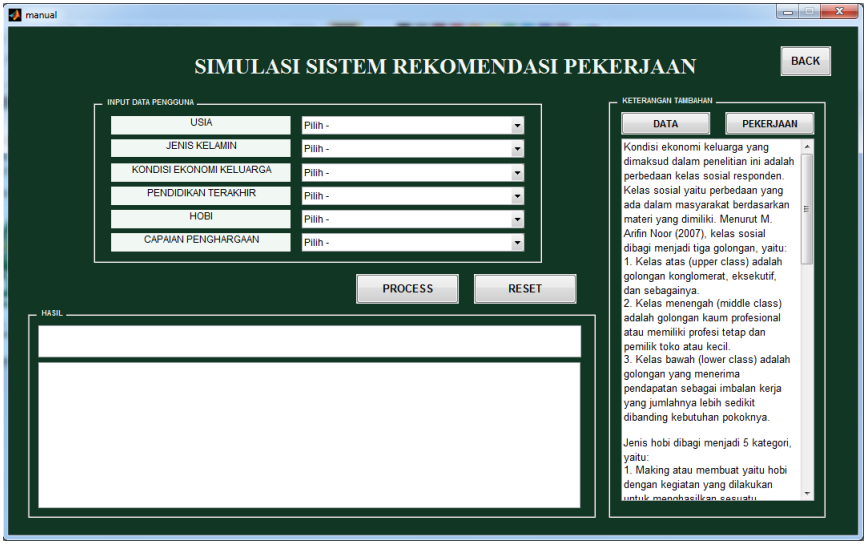
Gambar 4.3. Grafik Hasil Klasterisasi

Grafik tersebut merupakan hasil klasterisasi data menggunakan PCA (*Principal Component Analysis)* yang membentuk tiga *cluster*. Titik yang dekat satu sama lain menunjukkan data yang memiliki karakteristik serupa, sementara titik yang berjauhan menunjukkan perbedaan yang lebih besar di antara data tersebut. Angka-angka yang tertera di setiap titik menunjukkan indeks atau nomor data.

3. Simulasi sistem rekomendasi pekerjaan

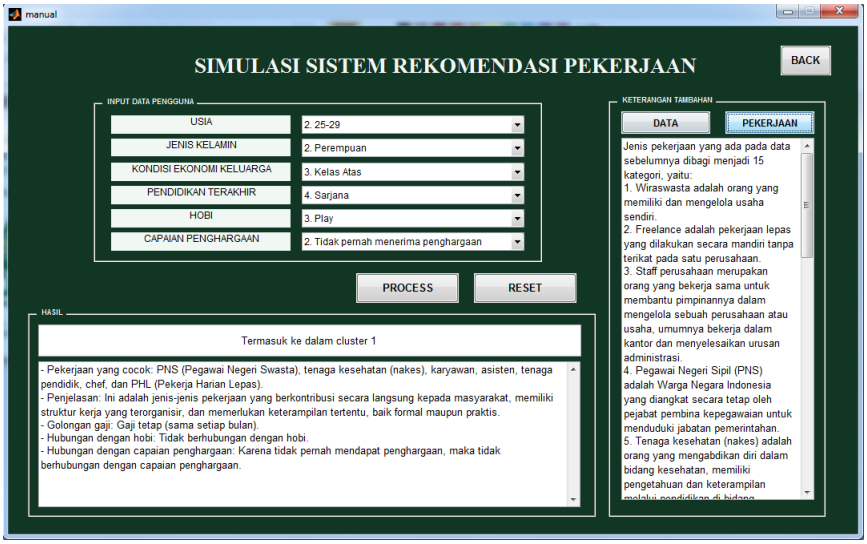
   Untuk bagian ini, pengguna diharuskan memilih latar belakang yang sesuai dengan dirinya, seperti usia, jenis kelamin, kondisi ekonomi keluarga, pendidikan terakhir, hobi, dan capaian penghargaan. Pada kolom "keterangan tambahan" diberikan penjelasan lebih lanjut mengenai

data dan pekerjaan yang akan membantu pengguna untuk memahami arti dari tiap kategori pada variabel pertanyaan dan jenis pekerjaan yang didapatkan.

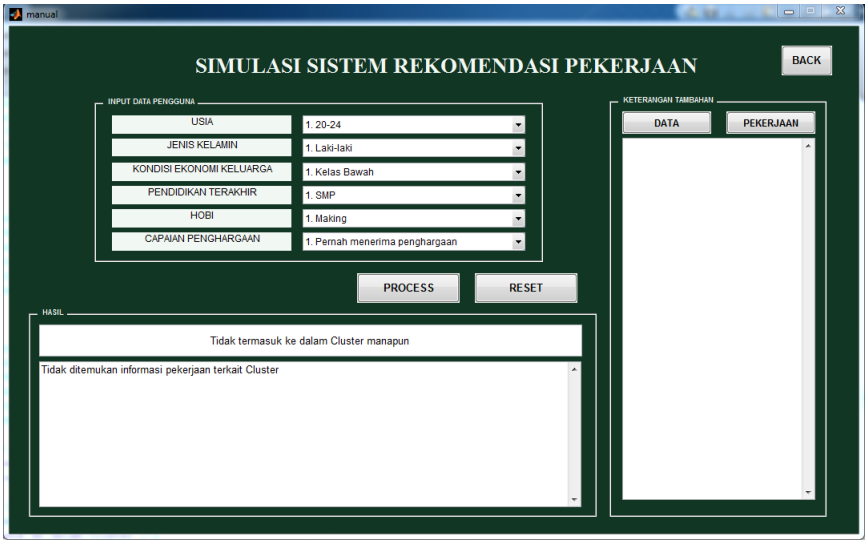


Gambar 4.4. Tampilan Sebelum Data Di*input*kan

Terdapat dua bentuk *output* yang dihasilkan dari proses *input* data oleh pengguna, yaitu untuk data yang termasuk dalam salah satu *cluster* dan yang tidak masuk ke dalam *cluster* manapun (tidak termasuk dalam 42 kombinasi data hasil klasterisasi).



Gambar 4.5. Tampilan Setelah *Input* Data



Gambar 4.6. Tampilan Sistem Jika Data Tidak Termasuk dalam *Cluster*

BAB 5
PENUTUP

5.1 Kesimpulan

Model klasterisasi jenis pekerjaan dengan 199 data yang diperoleh dari responden bekerja pengguna sosial media "X" (*formerly* twitter) menggunakan metode K-Means, menghasilkan tiga *cluster* dengan tingkat akurasi sebesar 58,74%. Karakteristik setiap *cluster* adalah sebagai berikut:

1. *Cluster* 1: terdiri dari responden berusia antara 25-34 tahun, didominasi perempuan dengan latar belakang ekonomi pada kelas menengah ke atas, pendidikan terakhir yaitu setingkat SMP, diploma, dan sarjana. Memiliki hobi *activity*/aktifitas, *play*/bermain, dan lainnya. Umumnya bekerja sebagai PNS, nakes, karyawan, asisten, tenaga pendidik, *chef*, dan PHL dengan gaji tetap (sama setiap bulan), tetapi pekerjaan yang didapat tidak berhubungan dengan hobi maupun penghargaan karena responden tidak pernah mendapat penghargaan.

2. *Cluster* 2: terdiri dari responden berusia antara 30-34 tahun, didominasi laki-laki dengan latar belakang ekonomi pada kelas bawah, pendidikan terakhir yaitu setingkat SMP dan SMA/SMK sederajat. Memiliki hobi *activity*/aktifitas dan *play*/bermain. Umumnya bekerja sebagai wiraswasta dan *freelance* dengan gaji tidak tetap, pekerjaan yang didapat berhubungan dengan hobi, namun tidak berhubungan dengan penghargaan karena responden tidak pernah

mendapat penghargaan.

3. *Cluster* 3: terdiri dari responden berusia antara 20-24 tahun, didominasi laki-laki dengan latar belakang ekonomi pada kelas menengah, pendidikan terakhir yaitu setingkat magister. Memiliki hobi *making*/membuat dan *arts*/seni. Umumnya bekerja sebagai staff perusahaan, *part time*, pegawai magang, *entertainer*, TNI, dan lainnya dengan gaji tidak tetap, pekerjaan yang didapat berhubungan dengan hobi maupun penghargaan yang pernah diraih.

Hasil simulasi sistem rekomendasi jenis pekerjaan menunjukkan bahwa penerapan 42 aturan kombinasi data yang diperoleh dari karakteristik setiap *cluster* dapat mengidentifikasi jenis pekerjaan yang sesuai dengan latar belakang pengguna. Simulasi ini memperlihatkan bahwa sistem dapat memberikan rekomendasi sesuai dengan hasil olah data dalam penelitian yang sesuai dengan karakteristik tertentu dari setiap *cluster*, sehingga meningkatkan kecocokan antara latar belakang pengguna dengan jenis pekerjaan yang direkomendasikan.

5.2 Saran

Untuk meningkatkan keragaman jenis pekerjaan yang dianalisis dalam penelitian ini, disarankan untuk memperluas jumlah sampel data yang digunakan. Selain itu, penambahan variabel baru seperti "jumlah jam kerja" dapat memberikan dimensi tambahan pada analisis, sehingga hasil klasterisasi dapat menjadi lebih komprehensif dan mendetail.

DAFTAR PUSTAKA

Abdi, H., Williams, L. J. Principal Component Analysis. 2010. *WIREs Comput. Stat.*, No. 2, hal. 433-459

Abdurrahman, Ginanjar. 2016. Clustering Data Ujian Tengah Semester (UTS) Data Mining Menggunakan Algoritma Apriori. *Jurnal Sistem & Teknologi Informasi Indonesia*, Vol. 1, No. 2.

Amin, Hadia., Devi, Abita., dan Ul Amin, Nadia. 2019. 'Predictive Analysis of Heart Disease Using K-Means and Apriori Algorithms'. *JASC: Journal of Applied Science and Computations*, ISSN NO: 2076-5131.

Arifin, Noor. 2007. *Ilmu Sosial Dasar*. Bandung: CV. Pustaka Setia.

Chen, ChuanShi., Liu, Kun., & Ma, Kun. 2018. 'Research on Evaluation of College Students' Professional Ability Based on K-means Clustering'. *MATEC Web of Conferences*.

Davies., Beynon, P. 2004. *Database Systems Third Edition*. New York: Palgrave Mcmillan.

Fajri, Ahmad Fikri. 2016. Implementasi Algoritma Apriori dalam Menentukan Program Studi yang Diambil Mahasiswa. *Jurnal Ipteks Terapan, Research of Applied Science and Education*, Vol. 10(2), hal. 81-85.

Fauzi, Saleh, & Asror. 2016. Penerapan Metode Association Rule Menggunakan Algoritma Apriori pada Simulasi Prediksi Hujan Wilayah Kota Bandung. *JITTER: Jurnal Ilmiah Teknologi Informasi Terapan*, Vol. 2(3).

Jananto, Arief. 2010. *Memprediksi Kinerja Mahasiswa Menggunakan Teknik Data Mining (Studi kasus data akademik mahasiswa UNISBANK)*. Yogyakarta: Universitas Gajah Mada.

Jolliffe, I. T., Cadima, J. Principal Component Analysis: A review and recent developments. 2016. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*

Kasiran, Moh. 2010. *Metodologi Penelitian Kualitatif-Kuantitatif*. Malang: UIN Maliki Press.

Kusrini., Luthfi, Taufiq., & Emha. 2009. *Algoritma Data Mining*. Yogyakarta: CV. Andi Offset.

Larose, D.T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.

Lijuan, L., Changliang, Z. 2017. 'Research and Application of Data Mining in College Students Employment Based on Association Rules'. *Computer and Information Science*, Vol. 10(3), hal. 54-59.

Liu, Y. H., Chen, J. G., & Zhang, C. Y. 2015. 'Two-way Recommendation System of Domestic College Students' Employment Information Based on Data Mining'. *Journal of Shenyang University (Natural Science Edition)*, Vol. 27(3), hal. 226-232.

MacQueen, J. B. 1967. *Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.

Marisa, F., Maukar, A. I., & Akhriza, T. M. 2021. *Data Mining: Konsep dan Penerapannya*. Sleman: Deepublish.

Medkom. 2023. *Mencapai Kesuksesan Meski Karier Tak Sesuai Jurusan Kuliah*. Diunduh di https://news.maranatha.edu/ tanggal 11 Juni 2023.

Murti, D. H., Suciati, N., & Nanjaya, D. J. 2005. Clustering Data Non-Numerik dengan Pendekatan Algoritma K-Means dan Hamming Distance Studi Kasus Biro Jodoh. *JUTI*, Vol. 4, No. 1, hal. 46-53.

Nofriansyah, D. 2015. *Modul: Data Mining, Algoritma Data Mining, dan Pengujian*. Sleman: Deepublish.

Putra, Septa Firmansyah. 2016. Feature Selection pada Dataset Faktor Kesiapan Bencana pada Provinsi di Indonesia Menggunakan metode PCA(Principal Componen Analysis). *Jurnal Teknik ITS*, Vol. 5, No. 2.

Putria, Narti Eka. 2018. Implementasi Data Mining Penjualan Tiket Pesawat Menggunakan Algoritma Apriori pada Terminal Tiket Batam Tour & Travel. *CBIS: Computer Based Information System Journal*, Vol. 6(1).

Rachmawati, Ariska Kurnia. Analisis Klasifikasi Status Gizi Anak dan Balita Menggunakan Algoritma K-Means.

Rachmawati, Ariska K., Saleh, Minhayati., & Ramdani, Muhammad N. 2021. Comparison and Pradiction of Data Mining Models to Determine the Classification of Family Planning Program User Status. *Indonesian Journal of Mathematics Education*, Vol. 4, No. 2.

Romaito, Jennyfer. 2017. Penggunaan Analisis Regresi Logistik Multinomial Untuk Klasifikasi Pemilihan Lapangan Pekerjaan di DKI Jakarta. *Sarjana Thesis*. Universitas Negeri Jakarta.

Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, No. 20, hal. 53-65.

Sugiyono. 2015. *Metode Penelitian Pendidikan: Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Tan, P. N., Stainbach, M., & Kumar, V. 2005. *Introduction to Data Mining*. New York: Addison Wesley Publisher.

Tika, Moh. Pabundu. 2006. *Metodologi Riset Bisnis*. Jakarta: PT. Bumi Aksara.

Ulfah, Maria., Irtawty, Andi Sri. 2022. Implementation of Data Mining Clustering Using the K-Means Method in Grouping Library Books. *Proceeding International Applied Business and Engineering Conference*.

Vassakis, K., Petrakis, E., & Kopanakis, I. 2018. 'Big Data Analytics: Applications, Prospects, and Challenges'. *Lecture Notes on Data Engineering and Communications Technologies*, hal. 3-20.

Vulandari, Retno Tri. 2017. *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: Gava Media.

Wahyuni, W. A., Saepudin, S. 2021. Penerapan Data Mining Clustering untuk Mengelompokkan Berbagai Jenis Merk Mesin Cuci. *SISMATIK: Seminar Nasional Sistem Informasi dan Manajemen Informatika*.

Yasin, Muhammad., Priyono, Joko. 2016. Analisis Faktor Usia, Gaji dan Beban Tanggungan Terhadap Produksi *Home* Industri Sepatu di Sidoarjo (Studi Kasus di Kecamatan Krian). *Jurnal Ekonomi & Bisnis*, Vol. 1(1), hal. 95-120.

Zhang, G. H. 2016. 'Application and Practice of Data Mining in Early Warning of Independent College Curriculum'. *Modern Electronic Technology*, Vol. 39(17), hal. 136-139.

Lampiran 1. Data Hasil Jawaban Responden pada Kuesioner Penelitian

Keterangan (pertanyaan dalam kuesioner):

Q1 : Tuliskan nama Anda (nama lengkap/*username* X/inisial/panggilan)

Q2 : Tuliskan usia Anda dalam tahun (mis. 21)

Q3 : Apa jenis kelamin Anda?

Q4 : Termasuk ke dalam golongan manakah kondisi ekonomi keluarga Anda?

Q5 : Apa pendidikan terakhir Anda?

Q6 : Apa hobi Anda? (silakan ketik "-" apabila tidak ingin menjawab) (mis. berolahraga)

Q7 : Apakah Anda pernah mendapat sebuah penghargaan? (penghargaan yang dimaksud di sini adalah jenis penghargaan yang dapat dipertimbangkan dalam pemilihan kerja/dapat dituliskan ke dalam CV)

Q8 : Jika pernah, silakan tulis penghargaan apa yang Anda dapat (silakan ketik "-" jika tidak pernah)

Q9 : Apakah saat ini Anda bekerja?

Q10 : Jika ya, tuliskan jenis pekerjaan yang Anda jalani saat ini (silakan ketik "-" jika tidak bekerja) (mis. guru olahraga)

Q11 : Termasuk ke dalam golongan manakah gaji yang Anda dapat?

Q12 : Apakah ada hubungan antara hobi dengan pekerjaan Anda saat ini?

Q13 : Apakah ada hubungan antara penghargaan yang Anda sebutkan di atas dengan pekerjaan Anda saat ini?

Gambar 0.1. Data Responden

NO	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
1	thisisuraccount	21	Perempuan	Kelas menengah	SMA/SMK sederajat	menyanyi	tidak pernah	-	Ya, saya seorang pekerja	joki	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
2	sya	27	Perempuan	Kelas bawah	SMA/SMK sederajat	membaca buku	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
3	sania	22	Perempuan	Kelas menengah	Sarjana	menulis	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Tidak ada hubungannya	Ya, berhubungan
4	N	20	Perempuan	Kelas menengah	SMA/SMK sederajat	Nonton anime	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
5	violet	23	Perempuan	Kelas menengah	Sarjana	nonton film	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
6	Dwi	20	Perempuan	Kelas menengah	SMA/SMK sederajat	Menonton film/drama	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
7	fasya	21	Perempuan	Kelas bawah	SMA/SMK sederajat	menyanyi	tidak pernah	-	Ya, saya seorang pekerja	admin olshop	Bergaji tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
8	JZA	21	Perempuan	Kelas menengah	SMA/SMK sederajat	olahraga	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Ya, berhubungan	Ya, berhubungan
9	yolan	22	Perempuan	Kelas bawah	SMA/SMK sederajat	-	tidak pernah	-	Ya, saya seorang pekerja	therapis	Bergaji tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
10	ariana	20	Perempuan	Kelas menengah	SMA/SMK sederajat	menonton film	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
11	anindia	20	Perempuan	Kelas menengah	SMA/SMK sederajat	-	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
12	Rara	22	Perempuan	Kelas menengah	SMA/SMK sederajat	Membaca	tidak pernah	-	Ya, saya seorang pekerja	Video Editor	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
13	Isna	21	Perempuan	Kelas menengah	SMA/SMK sederajat	mendengarkan musik	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
14	Adisa	22	Perempuan	Kelas menengah	SMA/SMK sederajat	menonton drama dan film	pernah	meraih medali emas softball se sumbagsel	Ya, saya seorang pekerja	freelancer	Bergaji tetap	Tidak ada hubungannya	Tidak ada hubungannya
15	Sajad	21	Laki-laki	Kelas menengah	SMA/SMK sederajat	Motoran	tidak pernah	-	Ya, saya seorang pekerja	Karyawan	Bergaji tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
16	BS	22	Laki-laki	Kelas menengah	Sarjana	-	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
17	N	24	Perempuan	Kelas menengah	Diploma	mendengarkan musik	tidak pernah	-	Tidak, saya tidak bekerja	Freelance editor & proofreader	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
18	htwtver	20	Perempuan	Kelas menengah	SMA/SMK sederajat	Membaca buku	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
19	Maulana	20	Laki-laki	Kelas menengah	SMA/SMK sederajat	-	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Tidak ada hubungannya	Tidak ada hubungannya
20	aurora	21	Perempuan	Kelas menengah	SMA/SMK sederajat	menulis	pernah	juara 2 videography se-PTKIN Indonesia	Ya, saya seorang pekerja	Freelancer Tester	Bergaji tidak tetap	Ya, berhubungan	Tidak ada hubungannya
21	abel	21	Perempuan	Kelas menengah	Sarjana	membaca	pernah	juara 3 olimpiade pajak tingkat provinsi	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Tidak ada hubungannya	Tidak ada hubungannya
22	GT	21	Perempuan	Kelas bawah	SMA/SMK sederajat	yoga	pernah	menang lomba english speech	Ya, saya seorang pekerja	rekruter	Bergaji tetap	Tidak ada hubungannya	Tidak ada hubungannya
23	P	21	Perempuan	Kelas menengah	Sarjana	-	pernah	Kejuaraan karya ilmiah	Ya, saya seorang pekerja	Pemilik online shop	Bergaji tidak tetap	Ya, berhubungan	Ya, berhubungan
24	amp	21	Perempuan	Kelas bawah	SMA/SMK sederajat	menari, menyanyi	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
25	RJP	21	Perempuan	Kelas menengah	SMA/SMK sederajat	-	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
26	Dilah	21	Perempuan	Kelas menengah	SMA/SMK sederajat	-	pernah	Mtq	Ya, saya seorang pekerja	Konten kreator	Bergaji tidak tetap	Ya, berhubungan	Tidak ada hubungannya
27	J	21	Perempuan	Kelas bawah	SMA/SMK sederajat	-	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
28	S	21	Perempuan	Kelas menengah	SMA/SMK sederajat	-	tidak pernah	-	Ya, saya seorang pekerja	Jualan Pulsa	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
29	Rachel	20	Perempuan	Kelas menengah	SMA/SMK sederajat	Seni	pernah	-	Ya, saya seorang pekerja	Part time	Bergaji tidak tetap	Ya, berhubungan	Tidak ada hubungannya
30	Arzenit	22	Perempuan	Kelas menengah	SMA/SMK sederajat	Membaca	tidak pernah	*_*	Tidak, saya tidak bekerja	*_*	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
31	AWP	20	Perempuan	Kelas menengah	SMA/SMK sederajat	musik	pernah	emas atletik porprov	Ya, saya seorang pekerja	surveyor	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak ada hubungannya
32	Farah Fazira	25	Perempuan	Kelas menengah	Sarjana	Olahraga	pernah	Best Participant Exchange Program	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
33	Fizky	23	Laki-laki	Kelas menengah	Sarjana	-	tidak pernah	-	Ya, saya seorang pekerja	Part time pramuniaga	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
34	Dehasyifa	22	Perempuan	Kelas bawah	SMA/SMK sederajat	Bernyanyi	tidak pernah	-	Ya, saya seorang pekerja	Jasa Ekspedisi	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
35	Agustin	28	Perempuan	Kelas bawah	Sarjana	-	tidak pernah	-	Ya, saya seorang pekerja	Admin Swalayan	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
36	Ai	24	Perempuan	Kelas menengah	SMA/SMK sederajat	Memasak	tidak pernah	-	Tidak, saya tidak bekerja	-	Saya tidak bekerja	Saya tidak bekerja	Saya tidak bekerja
37	Ileeknos	22	Perempuan	Kelas menengah	SMA/SMK sederajat	-	tidak pernah	-	Ya, saya seorang pekerja	asisten MUA	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
...	...	...	...	...	...	...	...	...	...	...	...	...	...
285	atheresis	22	Perempuan	Kelas menengah	Sarjana	nonton film	pernah	diterima magang msib	Ya, saya seorang pekerja	karyawan swasta	Bergaji tetap	Ya, berhubungan	Ya, berhubungan

Lampiran 2. Data Responden Bekerja

Keterangan (variabel data):

V1 : Variabel usia

V2 : Variabel jenis kelamin

V3 : Variabel kondisi ekonomi keluarga

V4 : Variabel pendidikan terakhir

V5 : Variabel hobi

Kategori V5 : Variabel hobi setelah dikategorikan menjadi 5 jenis

V6 : Variabel capaian penghargaan

V7 : Variabel jenis pekerjaan

Kategori V7 : Variabel jenis pekerjaan setelah dikategorikan menjadi 15 jenis

V8 : Variabel golongan gaji

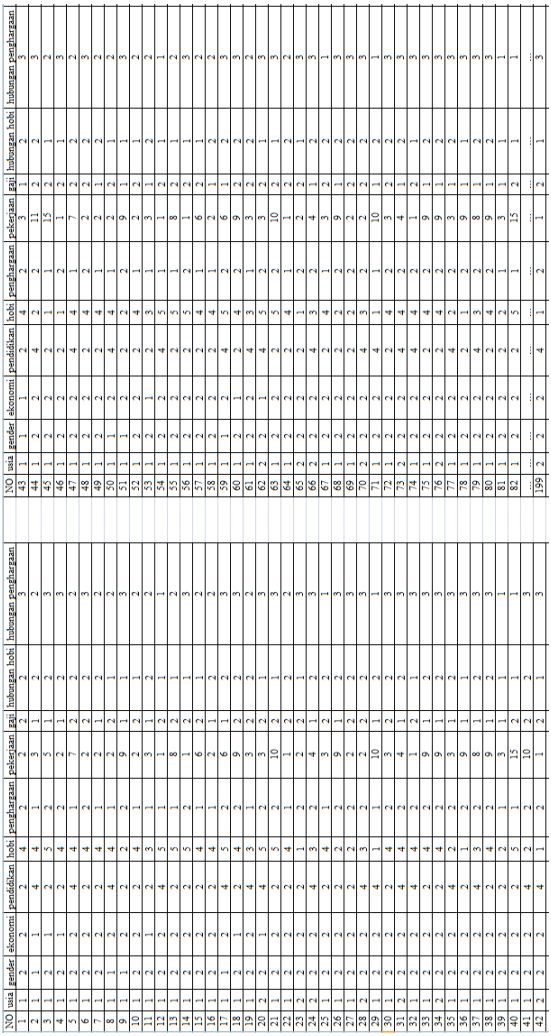
V9 : Variabel hubungan antara hobi dengan pekerjaan

V10 : Variabel hubungan antara capaian penghargaan dengan pekerjaan

Gambar 0.2. Data Responden Bekerja

NO	V1	V2	V3	V4	V5	Kategori V5	V6	V7	Kategori V7	V8	V9	V10
1	21	Perempuan	Kelas menengah	SMA/SMK sederajat	menyanyi	arts	tidak pernah	joki	freelance	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
2	24	Laki-laki	Kelas bawah	Sarjana	Bermain musik	arts	pernah	Bekerja di perusahaan tambang nikel di Morowali	staff perusahaan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak ada hubungannya
3	22	Perempuan	Kelas bawah	SMA/SMK sederajat	-	lainnya	tidak pernah	therapis	nakes	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
4	21	Perempuan	Kelas bawah	SMA/SMK sederajat	menyanyi	arts	tidak pernah	admin olshop	freelance	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
5	21	Perempuan	Kelas menengah	Sarjana	ballet	arts	pernah	magang	magang	Bergaji tidak tetap	Tidak ada hubungannya	Tidak ada hubungannya
6	22	Perempuan	Kelas menengah	SMA/SMK sederajat	Membaca	arts	tidak pernah	Video Editor	freelance	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
7	22	Perempuan	Kelas menengah	SMA/SMK sederajat	menonton drama dan film	arts	pernah	freelancer	freelance	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak ada hubungannya
8	23	Laki-laki	Kelas menengah	Sarjana	Bikin cerpen	arts	pernah	Freelance penulis, jasa ketik, dan design	freelance	Bergaji tidak tetap	Ya, berhubungan	Tidak ada hubungannya
9	21	Laki-laki	Kelas menengah	SMA/SMK sederajat	Motoran	activity	tidak pernah	Karyawan	karyawan	Bergaji tetap (sama setiap bulan)	Ya, berhubungan	Tidak pernah dapat penghargaan
10	21	Perempuan	Kelas menengah	SMA/SMK sederajat	menulis	arts	pernah	Freelencer Tester	freelance	Bergaji tidak tetap	Ya, berhubungan	Tidak ada hubungannya
11	21	Perempuan	Kelas bawah	SMA/SMK sederajat	yoga	play	pernah	rekruter	staff perusahaan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak ada hubungannya
12	21	Perempuan	Kelas menengah	Sarjana	-	lainnya	pernah	Pemilik online shop	wiraswasta	Bergaji tidak tetap	Ya, berhubungan	Ya, berhubungan
13	21	Perempuan	Kelas menengah	SMA/SMK sederajat	-	lainnya	pernah	Konten kreator	entertain	Bergaji tidak tetap	Ya, berhubungan	Tidak ada hubungannya
14	21	Perempuan	Kelas menengah	SMA/SMK sederajat	-	lainnya	tidak pernah	Jualan Pulsa	wiraswasta	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
15	20	Perempuan	Kelas menengah	SMA/SMK sederajat	Seni	arts	pernah	Part time	part time	Bergaji tidak tetap	Ya, berhubungan	Tidak ada hubungannya
16	20	Perempuan	Kelas menengah	SMA/SMK sederajat	musik	arts	pernah	surveyor	freelance	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak ada hubungannya
17	23	Laki-laki	Kelas menengah	Sarjana	-	lainnya	tidak pernah	Part time pramusiaga	part time	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
18	22	Perempuan	Kelas bawah	SMA/SMK sederajat	Bernyanyi	arts	tidak pernah	Jasa Ekspedisi	karyawan	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
19	24	Perempuan	Kelas menengah	Sarjana	Renang	play	pernah	Marketing and Social Media Specialist	staff perusahaan	Bergaji tidak tetap	Tidak ada hubungannya	Tidak ada hubungannya
20	28	Perempuan	Kelas bawah	Sarjana	-	lainnya	tidak pernah	Admin Swalayan	staff perusahaan	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
21	22	Perempuan	Kelas menengah	SMA/SMK sederajat	-	lainnya	tidak pernah	asisten MUA	asisten	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
22	21	Perempuan	Kelas menengah	SMA/SMK sederajat	Bermusik	arts	pernah	Jualan cincin manik manik	wiraswasta	Bergaji tidak tetap	Tidak ada hubungannya	Tidak ada hubungannya
23	27	Perempuan	Kelas menengah	SMA/SMK sederajat	Fotografi	making	tidak pernah	Fotografer	freelance	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
24	26	Perempuan	Kelas menengah	Sarjana	olahraga	play	tidak pernah	PNS	PNS	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
25	21	Perempuan	Kelas menengah	SMA/SMK sederajat	Mendengarkan musik	arts	pernah	Ads Specialist	staff perusahaan	Bergaji tidak tetap	Tidak ada hubungannya	Ya, berhubungan
26	22	Perempuan	Kelas menengah	SMA/SMK sederajat	Jalan-jalan	activity	tidak pernah	Barista	karyawan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
27	22	Perempuan	Kelas menengah	SMA/SMK sederajat	traveling	activity	tidak pernah	Freelance	freelance	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
28	26	Perempuan	Kelas menengah	Sarjana	Olahraga, musik	play	tidak pernah	Freelance project	freelance	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
29	22	Perempuan	Kelas menengah	Sarjana	Jalan jalan membuat konten	making	pernah	Asisten Dosen	asisten	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Ya, berhubungan
30	23	Perempuan	Kelas menengah	SMA/SMK sederajat	watching movie	arts	tidak pernah	admin	staff perusahaan	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
31	25	Perempuan	Kelas menengah	Sarjana	Menyanyi	arts	tidak pernah	PNS	PNS	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
32	22	Perempuan	Kelas menengah	Sarjana	Membaca	arts	tidak pernah	Wirausaha	wiraswasta	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
33	21	Perempuan	Kelas menengah	SMA/SMK sederajat	Dance	arts	tidak pernah	Karyawan Swasta	karyawan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
34	27	Perempuan	Kelas menengah	SMA/SMK sederajat	Bernyanyi	arts	tidak pernah	Karyawan Swasta	karyawan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
35	21	Perempuan	Kelas menengah	Sarjana	Jalan jalan	activity	tidak pernah	Admin	staff perusahaan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
36	21	Perempuan	Kelas menengah	SMA/SMK sederajat	memasak	making	tidak pernah	waiters di salah satu tempat makan	karyawan	Bergaji tetap (sama setiap bulan)	Ya, berhubungan	Tidak pernah dapat penghargaan
37	22	Perempuan	Kelas menengah	Sarjana	Olahraga	play	tidak pernah	content creator	entertain	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
38	21	Perempuan	Kelas menengah	SMA/SMK sederajat	Membaca novel	arts	tidak pernah	Kasir	karyawan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
39	20	Perempuan	Kelas menengah	SMA/SMK sederajat	Travelling	activity	pernah	Admin	staff perusahaan	Bergaji tetap (sama setiap bulan)	Ya, berhubungan	Ya, berhubungan
40	22	Perempuan	Kelas menengah	SMA/SMK sederajat	-	lainnya	pernah	-	lainnya	Bergaji tidak tetap	Ya, berhubungan	Ya, berhubungan
41	22	Perempuan	Kelas menengah	Sarjana	traveling	activity	tidak pernah	research assistant	asisten	Bergaji tidak tetap	Tidak ada hubungannya	Tidak pernah dapat penghargaan
42	25	Perempuan	Kelas menengah	Sarjana	memasak	making	tidak pernah	wiraswasta	wiraswasta	Bergaji tidak tetap	Ya, berhubungan	Tidak pernah dapat penghargaan
43	22	Laki-laki	Kelas bawah	SMA/SMK sederajat	bernyanyi	arts	tidak pernah	recruitment	staff perusahaan	Bergaji tetap (sama setiap bulan)	Tidak ada hubungannya	Tidak pernah dapat penghargaan
...	...	...	...	...	...	...	...	...	...	...	...	...
199	22	Perempuan	Kelas menengah	Sarjana	nonton film	arts	pernah	karyawan swasta	karyawan	Bergaji tetap (sama setiap bulan)	Ya, berhubungan	Ya, berhubungan

Lampiran 3. Hasil Transformasi Data



NO	usia	gender	ekonomi	pendidikan	hobi	pendapatan	pekerjaan	jasa	hubungan hobi	hubungan pengharapan
1	1	1	2	4	2	1	3	2	2	3
2	1	1	2	4	2	1	4	2	2	3
3	1	2	1	4	5	2	5	1	2	3
4	1	2	1	2	4	2	7	2	2	3
5	1	2	2	4	4	1	7	2	2	2
6	1	2	2	2	4	2	2	2	2	3
7	1	2	2	2	4	1	2	2	2	2
8	1	1	2	4	4	1	2	2	1	2
9	1	1	2	2	2	2	9	1	1	3
10	1	2	2	2	4	1	3	2	1	2
11	1	2	1	2	5	1	1	2	1	1
12	1	2	1	4	5	1	8	2	1	2
13	1	2	2	2	5	2	1	2	2	3
14	1	2	2	2	5	2	1	2	1	3
15	1	2	2	4	4	1	6	2	1	2
16	1	2	2	2	4	1	2	2	2	2
17	1	1	2	4	5	2	6	1	2	3
18	1	2	1	2	4	2	9	2	2	3
19	1	2	2	4	3	1	3	2	2	2
20	2	2	1	4	5	2	3	2	1	3
21	2	2	2	2	4	1	10	2	2	3
22	1	2	2	4	4	1	1	2	2	3
23	2	2	2	2	1	2	2	4	1	2
24	2	2	2	4	3	2	4	1	2	3
25	1	2	2	2	4	1	3	2	1	2
26	1	2	2	2	2	2	9	1	2	3
27	1	2	2	2	2	2	2	2	2	3
28	2	2	2	4	3	2	2	2	2	3
29	1	2	2	4	1	1	10	1	2	1
30	1	2	2	2	4	2	3	2	2	3
31	1	2	2	4	4	2	4	1	2	3
32	1	2	2	4	4	2	1	2	1	2
33	1	2	2	4	4	2	9	1	2	3
34	2	2	2	4	2	2	9	1	2	3
35	1	2	2	4	2	2	3	1	2	3
36	1	2	2	2	1	2	9	1	1	3
37	1	2	2	4	3	2	8	1	2	3
38	1	2	2	2	4	2	9	1	2	3
39	1	2	2	2	2	1	3	2	1	2
40	1	2	2	2	5	1	15	2	1	1
41	1	2	2	2	5	2	10	2	2	3
42	1	2	2	4	1	2	1	2	1	3
43	1	1	2	4	2	2	3	2	2	3
44	1	2	2	2	2	2	11	2	2	3
45	1	2	2	2	1	2	15	2	1	2
46	1	2	2	4	1	2	7	2	1	3
47	1	2	2	4	4	2	2	2	2	2
48	1	2	2	2	4	2	2	2	2	3
49	1	2	2	2	4	1	2	2	1	2
50	1	1	2	2	4	1	2	2	1	2
51	1	1	2	2	2	2	9	2	1	3
52	1	1	2	2	4	1	3	2	1	2
53	1	2	2	2	4	1	5	2	1	2
54	1	2	2	2	5	1	1	2	1	1
55	1	2	2	4	2	1	8	2	1	2
56	1	2	2	2	5	2	1	2	1	3
57	1	2	2	2	4	1	6	2	1	2
58	1	2	2	2	4	1	2	2	1	2
59	1	1	2	4	5	2	6	1	2	3
60	1	2	1	2	4	2	9	2	2	3
61	1	2	2	4	3	1	3	2	2	2
62	2	2	1	4	5	2	10	2	1	3
63	2	2	2	2	4	1	1	2	2	3
64	1	2	2	4	4	2	1	2	2	3
65	2	2	2	2	1	2	2	4	1	2
66	2	2	2	4	3	2	4	1	2	3
67	1	2	2	2	4	1	3	2	1	2
68	1	2	2	2	4	1	9	2	1	3
69	1	2	2	2	2	2	2	2	2	3
70	2	2	2	2	4	2	2	2	2	3
71	1	2	2	4	1	1	10	1	2	1
72	1	2	2	2	4	2	3	2	2	3
73	2	2	2	4	4	2	4	1	2	3
74	2	2	2	4	4	2	1	2	1	2
75	1	2	2	4	4	2	9	1	2	3
76	2	2	2	2	4	2	9	1	2	3
77	1	2	2	4	2	2	3	1	2	3
78	1	2	2	2	1	2	9	1	1	3
79	1	2	2	4	3	2	8	1	2	3
80	1	2	2	4	4	2	9	1	2	3
81	1	2	2	2	4	2	3	1	2	1
82	1	2	2	2	5	1	15	2	1	1
...	...	...	...	...	...	...	...	...	...	...
199	2	2	2	4	1	2	3	2	1	3

Gambar 0.3. Transformasi Data

Lampiran 4. Hasil Normalisasi Data

NO	usia	gender	ekonomi	pendidikan	hobi	penghargaan	pekerjaan	gaji	hubungan hobi	hubungan penghargaan
1	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	0,571547607	-0,846803594	1,111735156	0,664852604	0,531510981
2	-0,418258131	-2,766256299	-2,093294855	1,120255325	0,269155745	-1,749635531	-0,584729076	-0,899494808	0,664852604	-0,95821698
3	-0,418258131	0,361499403	-2,093294855	-0,897218065	1,023550015	0,571547607	-0,060580039	-0,899494808	0,664852604	0,531510981
4	-0,418258131	0,361499403	-2,093294855	-0,897218065	0,269155745	-1,749635531	0,463568997	1,111735156	0,664852604	-0,95821698
5	-0,418258131	0,361499403	0,446739756	1,120255325	0,269155745	0,571547607	-0,846803594	1,111735156	0,664852604	0,531510981
6	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	0,571547607	-0,846803594	1,111735156	0,664852604	0,531510981
7	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	-1,749635531	-0,846803594	-0,899494808	0,664852604	-0,95821698
8	-0,418258131	-2,766256299	0,446739756	1,120255325	0,269155745	-1,749635531	-0,846803594	1,111735156	-1,504092777	-0,95821698
9	-0,418258131	-2,766256299	0,446739756	-0,897218065	-1,239632796	0,571547607	0,987718033	-0,899494808	-1,504092777	0,531510981
10	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	-1,749635531	-0,846803594	1,111735156	-1,504092777	-0,95821698
11	-0,418258131	0,361499403	-2,093294855	-0,897218065	-0,485238526	-1,749635531	-0,584729076	-0,899494808	0,664852604	-0,95821698
12	-0,418258131	0,361499403	0,446739756	1,120255325	1,023550015	-1,749635531	-1,108878112	1,111735156	-1,504092777	-2,447944941
13	-0,418258131	0,361499403	0,446739756	-0,897218065	1,023550015	-1,749635531	0,725643515	1,111735156	-1,504092777	-0,95821698
14	-0,418258131	0,361499403	0,446739756	-0,897218065	1,023550015	0,571547607	-1,108878112	1,111735156	-1,504092777	0,531510981
15	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	-1,749635531	0,201494479	1,111735156	-1,504092777	-0,95821698
16	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	-1,749635531	-0,846803594	-0,899494808	0,664852604	-0,95821698
17	-0,418258131	-2,766256299	0,446739756	1,120255325	1,023550015	0,571547607	0,201494479	-0,899494808	0,664852604	0,531510981
18	-0,418258131	0,361499403	-2,093294855	-0,897218065	0,269155745	0,571547607	0,987718033	1,111735156	0,664852604	0,531510981
19	-0,418258131	0,361499403	0,446739756	1,120255325	-0,485238526	-1,749635531	-0,584729076	1,111735156	0,664852604	-0,95821698
20	1,662576072	0,361499403	-2,093294855	1,120255325	1,023550015	0,571547607	-0,584729076	1,111735156	-1,504092777	0,531510981
21	-0,418258131	0,361499403	0,446739756	-0,897218065	1,023550015	0,571547607	1,249792551	1,111735156	-1,504092777	0,531510981
22	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	-1,749635531	-1,108878112	1,111735156	0,664852604	-0,95821698
23	1,662576072	0,361499403	0,446739756	-0,897218065	-1,994027066	0,571547607	-0,846803594	1,111735156	-1,504092777	0,531510981
24	1,662576072	0,361499403	0,446739756	1,120255325	-0,485238526	0,571547607	-0,322654558	-0,899494808	0,664852604	0,531510981
25	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	-1,749635531	-0,584729076	1,111735156	0,664852604	-2,447944941
26	-0,418258131	0,361499403	0,446739756	-0,897218065	-1,239632796	0,571547607	0,987718033	-0,899494808	0,664852604	0,531510981
27	-0,418258131	0,361499403	0,446739756	-0,897218065	-1,239632796	0,571547607	-0,846803594	1,111735156	0,664852604	0,531510981
28	1,662576072	0,361499403	0,446739756	1,120255325	-0,485238526	0,571547607	-0,846803594	1,111735156	0,664852604	0,531510981
29	-0,418258131	0,361499403	0,446739756	1,120255325	-1,994027066	-1,749635531	1,249792551	-0,899494808	0,664852604	-2,447944941
30	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	0,571547607	-0,584729076	1,111735156	0,664852604	0,531510981
31	1,662576072	0,361499403	0,446739756	1,120255325	0,269155745	0,571547607	-0,322654558	-0,899494808	0,664852604	0,531510981
32	-0,418258131	0,361499403	0,446739756	1,120255325	0,269155745	0,571547607	-1,108878112	1,111735156	-1,504092777	0,531510981
33	-0,418258131	0,361499403	0,446739756	-0,897218065	0,269155745	0,571547607	0,987718033	-0,899494808	0,664852604	0,531510981
34	1,662576072	0,361499403	0,446739756	-0,897218065	0,269155745	0,571547607	0,987718033	-0,899494808	0,664852604	0,531510981
35	-0,418258131	0,361499403	0,446739756	1,120255325	-1,239632796	0,571547607	-0,584729076	-0,899494808	0,664852604	0,531510981
...	...	...	...	...	...	...	...	...	...	...
199	-0,418258131	0,361499403	0,446739756	1,120255325	0,269155745	-1,749635531	0,987718033	-0,899494808	-1,504092777	-2,447944941

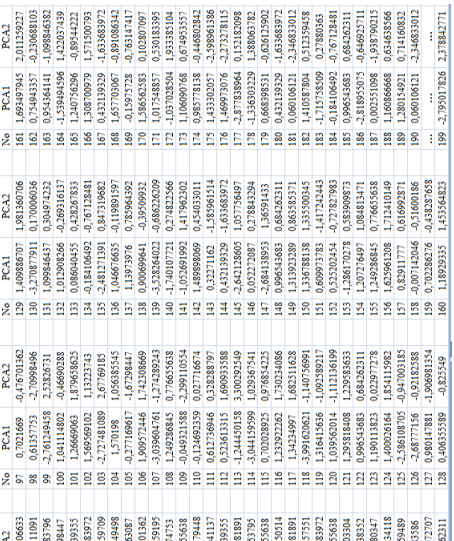
Gambar 0.4. Normalisasi Data

Lampiran 5. *Listing* Program Reduksi Dimensi dengan PCA

```
clc;
clear all;
% Menghapus file hasil yang ada sebelumnya
result_file = 'PCA_Reduced_Data.xlsx';
if exist(result_file, 'file') == 2
delete(result_file);
end
% Baca data dari file excel
filename = 'kmeans_manual_pca.xlsx';
% File excel data yang telah dinormalisasi
sheet = 'Sheet2';
xlRange = 'A1:K200';
[num, txt, raw] = xlsread(filename, sheet, xlRange);
% Ekstrak angka-angka di kolom pertama untuk identifikasi
numbers = num(:, 1);
% Mengambil kolom pertama dari data numerik
numericData = num(:, 2:end);
% Mengambil data mulai dari kolom ke-2 hingga terakhir
% Lakukan PCA pada data yang sudah dinormalisasi
[coeff, score, ~] = pca(numericData);
% Tampilkan hasil reduksi PCA
disp('Hasil reduksi PCA (2 komponen utama):');
for i = 1:size(score, 1)
% Loop berdasarkan jumlah baris pada 'score'
fprintf('Data %d (PCA1, PCA2): ', numbers(i));
fprintf('%f %f\n', score(i, 1), score(i, 2));
disp('========================================');
end
% Simpan hasil normalisasi dan PCA ke dalam file Excel
pca_result = [numbers, score(:, 1:2)];
% Gabungkan nomor data dengan hasil PCA
normalized_and_pca_labels = [{'No'}, {'PCA1'}, {'PCA2'}];
```

```
% Tulis hasil normalisasi dan PCA ke sheet baru
normalized_pca_sheet_name = 'NormalizedAndPCA';
xlswrite(result_file, normalized_and_pca_labels,
 normalized_pca_sheet_name, 'A1');
xlswrite(result_file, pca_result,
 normalized_pca_sheet_name, 'A2');
disp(['Hasil normalisasi dan PCA telah disimpan di sheet "',
 normalized_pca_sheet_name, '" pada file ', result_file]);
```

Lampiran 6. Hasil Reduksi Dimensi Data dengan PCA



No	PCA1	PCA2	No	PCA1	PCA2	No	PCA1	PCA2	No	PCA1	PCA2	No	PCA1	PCA2	No	PCA1	PCA2
1	0,465907698	-1,484259433	33	1,106090768	0,674953557	65	-2,76957719	-0,117306633	97	0,7021669	-0,476701362	129	1,409886707	1,981360706	161	1,693497945	2,011259227
2	-1,008888035	0,386970316	34	1,389702006	0,704852078	66	-0,454134063	-1,924111091	98	0,61357753	-2,70998496	130	-3,270877911	0,170006036	162	0,754943357	-0,230688103
3	1,451489109	-0,49489106	35	0,710151529	0,480838149	67	0,35315291	0,358985796	99	-2,761249458	2,5282731	131	1,099844437	0,304974232	163	0,954364141	-1,098846382
4	1,260987927	-1,044868759	36	-0,066574778	0,181122277	68	-0,277169611	-1,67298447	100	1,044114802	-0,46690288	132	1,012508266	-0,269316137	164	-1,539494596	1,42203739
5	-1,605034247	0,872230723	37	1,022189448	1,329682926	69	0,396662445	0,597239355	101	1,266669663	1,879659625	133	0,086040455	0,428267833	165	1,240756296	-0,89544222
6	0,465907698	-1,484259433	38	1,106090768	0,674953557	70	0,432139529	-1,653683972	102	1,569569102	1,13223743	134	-0,184106492	-0,767128481	166	1,308700979	1,571500795
7	-1,465134723	0,332782323	39	-3,377083316	0,737033384	71	0,241832327	-0,780559709	103	-2,727481089	2,6776185	135	-2,481211391	0,847319682	167	0,432139529	-1,653683972
8	-2,280139448	-0,190371734	40	-2,946079153	1,357994483	72	-0,302953557	-0,864484498	104	1,5703198	1,056383545	136	1,046676655	-0,119899597	168	1,657703067	-0,859106542
9	0,066760755	0,155578633	41	0,542726456	0,419388709	73	1,679827308	1,008663087	105	-0,277169417	-1,672984417	137	1,139739976	0,783964592	169	-0,139573728	-0,783147417
10	-2,810106629	-0,9969183	42	-0,365815854	-1,197867043	74	0,70215699	-0,478701502	106	1,569922446	1,712308669	138	0,906896451	-0,39509912	170	1,398358357	0,110010987
11	-1,325775188	0,417907323	43	0,576459984	-1,565013248	75	0,284691433	-0,31174715	107	-3,009236145	0,774659543	139	-3,140107722	0,274823566	171	1,017028554	1,593385161
12	-2,246210142	0,020966617	45	-2,600615971	0,668223767	77	1,249286845	0,776655638	109	-0,049231388	-2,299110554	141	-1,052691992	1,417962302	173	1,106090768	0,674953557
14	-0,167741909	-1,720706929	46	-0,740326217	-2,127515253	78	-0,623355267	-1,094179448	110	-0,124692539	0,027716674	142	1,485888869	0,454603011	174	0,98577818	-0,446803842
15	-2,474942056	-0,371484843	47	1,34234997	1,685211628	79	0,070087853	-1,630941137	111	0,612736946	0,328288797	143	0,32271162	-1,583965154	175	1,433303057	-2,598961386
16	-1,465134723	0,332782323	48	1,6540992	0,401656676	80	0,396662445	0,597239355	112	0,523613315	1,690933588	144	0,432139529	-1,653683972	176	1,469973076	-2,273278115
17	1,231184322	1,106992286	49	1,34234997	1,685211628	81	0,573335406	-1,531981891	113	-1,244450158	3,300292549	145	-2,642128605	1,057756497	177	-2,877838964	0,153182098
18	1,059362014	-1,112136199	50	1,166595508	-0,056855671	82	1,384360319	1,374353795	114	-3,044159599	1,029367541	146	0,052272087	0,278842594	178	-1,136303229	1,380663782
19	-1,883303798	0,173120485	51	0,456076578	1,392084599	83	1,249286845	0,776655638	115	0,702028925	0,976854225	147	-2,684138953	1,36591433	179	0,668398531	-0,636125902
20	0,613744929	-1,159951881	52	0,153886631	-1,964332859	84	-1,436600351	-0,548450514	116	1,232922262	1,730234086	148	0,996543683	0,684262311	180	0,432139529	-1,653683972
21	0,136173407	-0,375886075	53	0,432139529	-1,653683972	85	0,573335406	-1,531981891	117	1,34234997	1,685211628	149	1,313923289	0,865585371	181	0,060106121	-2,346833012
22	-1,900707583	-0,929882502	54	1,441575256	0,488278355	86	0,026337752	-2,496257551	118	-3,991620621	-1,140756991	150	1,336788138	1,335500345	182	1,410587804	0,512359458
23	-0,423146611	-1,948192193	55	0,769415067	-2,207532514	87	0,432139529	-1,653683972	119	1,316415456	-1,092589217	151	0,609973783	-1,417242443	183	-1,715758509	0,27880363
24	1,170727212	0,76188329	56	1,199153893	1,580809547	88	1,249286845	0,776655638	120	1,039562014	-1,112136199	152	0,525202454	-0,727827983	184	-0,184106492	-0,767128481
25	-2,751418338	-0,348076632	57	0,912761739	1,377405585	89	0,769565662	1,275703304	121	1,295818408	1,229583633	153	-1,286170278	0,583909873	185	0,996543683	0,684262311
26	0,819698614	0,471549395	58	-0,209482286	2,209100425	90	-1,596792267	0,414938552	122	0,996543683	0,684262311	154	1,207276497	1,08481347	186	-2,818959075	-0,646925711
27	0,179515543	-1,687663595	59	0,432139529	-1,653683972	91	0,169684424	1,188680347	123	1,190113823	0,02397728	155	1,249286845	0,776655638	187	0,002551098	-1,958790215
28	0,699385984	-0,650207004	60	1,249286845	0,776655638	92	1,182758092	-1,010434118	124	1,400026164	1,854115982	156	1,625961208	1,712401149	188	1,160866668	0,634635866
29	-2,447760374	2,411886104	61	-2,301082284	1,218092578	93	-0,813821066	0,507759489	125	-2,586108705	-0,947003185	157	0,82911777	0,616992871	189	1,280154921	0,71416832
30	0,499676066	-1,334834894	62	-3,957852253	-0,991332452	94	-2,820544214	1,77183586	126	-2,68777156	-0,921825588	158	-0,007142046	-0,516090186	190	0,060106121	-2,346833012
31	1,319923289	0,865585371	63	0,617088403	-0,424997841	95	0,515341911	-0,835072707	127	0,980147881	-1,906981354	159	0,702386276	-0,438287658	...	...	...
32	-0,21787486	-0,91655302	64	0,617088403	-0,424997841	96	0,996543683	0,684262311	128	0,406355589	-0,825549	160	1,18929935	1,453564823	199	-2,795017826	2,378842771

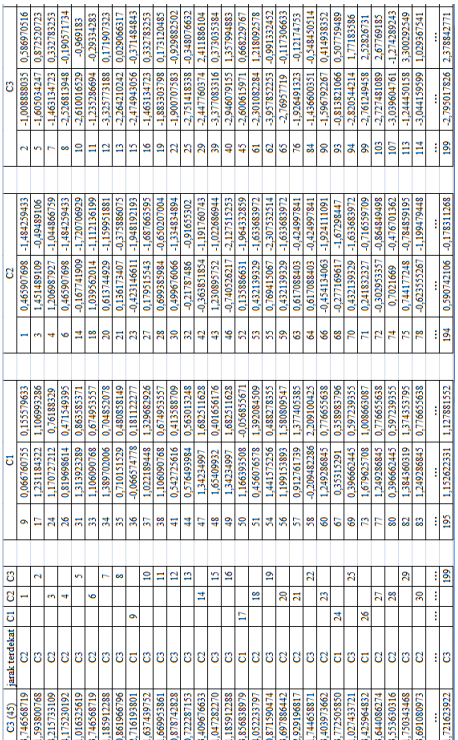
Gambar 0.5. Reduksi Dimensi Data dengan PCA

Lampiran 7. *Listing* Program Penentuan *Cluster* Optimal dengan Metode Silhouette

```
clc;
clear all;
% Baca data dari file excel
filename = 'PCA_Reduced_Data.xlsx';
% File excel data yang telah direduksi PCA
sheet = 'NormalizedAndPCA';
xlRange = 'A1:C200';
[num, txt, raw] = xlsread(filename, sheet, xlRange);
% Ekstrak angka-angka di kolom pertama untuk identifikasi
numbers = num(:, 1);
% Mengambil kolom pertama dari data numerik
numericData = num(:, 2:end);
% Mengambil data mulai dari kolom ke-2 hingga terakhir
% Tetapkan nilai acak agar hasil K-means tetap konsisten
rng(6);
% Tentukan rentang jumlah cluster yang akan dievaluasi
max_clusters = 15;
mean_silhouette_vals = zeros(max_clusters, 1);
% Loop untuk menghitung nilai Silhouette untuk
 setiap jumlah cluster
for k = 2:max_clusters
disp(['Jumlah cluster: ', num2str(k)]);
% Jalankan K-means pada data hasil PCA
[idx, C, sumd, D] = kmeans(numericData, k, 'Replicates',
 5, 'Display', 'final');
% Menghitung nilai Silhouette
silhouette_vals = silhouette(numericData, idx);
% Simpan nilai Silhouette rata-rata untuk
 jumlah cluster ini
mean_silhouette_vals(k) = mean(silhouette_vals);
% Tampilkan hasil dari setiap iterasi k-means
```

```
disp(['Jumlah sum of distances: ', num2str(sum(sumd))]);
disp(['Nilai rata-rata: ', num2str
 (mean_silhouette_vals(k))]);
disp(['Jarak ke centroid: ', num2str(mean(D, 1))]);
disp('Centroid untuk setiap cluster:');
disp(C);
disp('==========================================');
end
% Plot nilai Silhouette rata-rata untuk
 setiap jumlah cluster
figure;
plot(2:max_clusters, mean_silhouette_vals(2:end), 'b-o');
title('Metode Silhouette untuk Menentukan
 Jumlah Cluster Optimal');
xlabel('Jumlah Cluster');
ylabel('Nilai Silhouette Rata-rata');
grid on;
% Cari jumlah cluster optimal berdasarkan
 nilai Silhouette tertinggi
[~, optimal_clusters] = max(mean_silhouette_vals);
disp(['Jumlah cluster optimal berdasarkan metode
Silhouette: ', num2str(optimal_clusters)]);
disp('==========================================');
```

Lampiran 8. Perhitungan Iterasi-1 untuk Klasterisasi dengan Metode K-Means

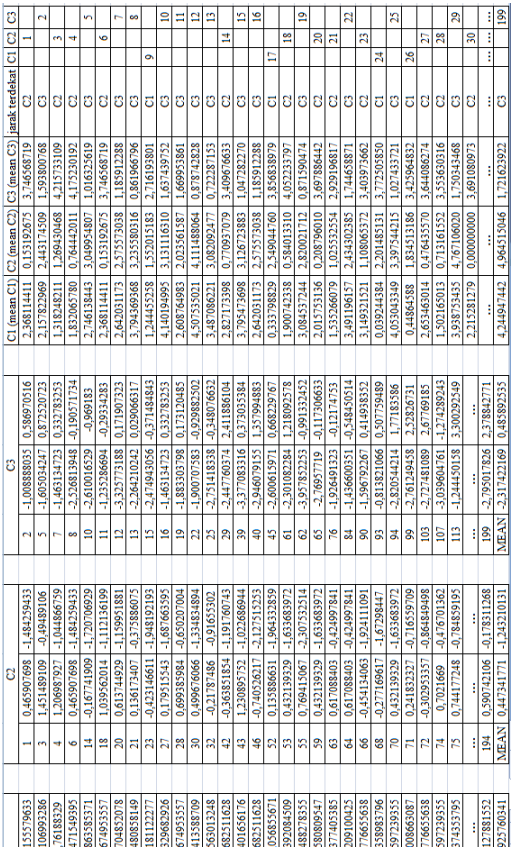


NO	PCA1	PCA2	C1 (137)	C2 (30)	C3 (45)	jarak terdekat	C1	C2	C3
1	0,465907698	-1,484259433	2,368114411	0,153192675	3,74568719	C2		1	
2	-1,008888035	0,586970516	2,157822969	2,445174599	1,595800768	C3			2
3	1,451489109	-0,494891061	1,318248211	1,269430468	4,215733109	C2		3	
4	1,206987927	-1,044866759	1,832065780	0,764442011	4,175230192	C2		4	
5	-1,607034247	0,872520723	2,746138443	3,040934807	1,016635619	C3			5
6	0,465907698	-1,484259433	2,368114411	0,153192675	3,74568719	C2			6
7	-1,463134723	0,332783253	2,642031173	2,575573008	1,185912308	C3			7
8	-2,528813948	-0,190571734	3,794369368	3,235580316	0,861066796	C3			8
9	0,006760755	0,155579633	1,244453258	1,525063183	2,761039801	C1	9		
10	-2,610016529	-0,9698185	4,140194995	3,131116310	1,617439752	C3			10
11	-1,252386994	-0,295342385	2,608749983	2,026569187	1,699953861	C3			11
12	-3,252573188	0,171907233	4,507535021	4,111488064	0,878742828	C3			12
13	-2,264210242	0,020960517	3,487086221	3,082092477	0,722387155	C3			13
14	-0,167174099	-1,720796929	2,382713398	0,709539789	3,540966653	C2		14	
15	-2,467132756	-0,917184845	3,762456998	3,179523883	1,147892370	C3			15
16	-1,251184325	1,106991286	0,933798829	2,549044790	3,856819979	C1	17		
17	1,039562014	-1,112156199	1,900742338	0,384013310	4,052233797	C2		18	
18	-1,883303798	0,173120485	3,068457244	2,820021712	0,871590474	C3			19
19	0,613774929	-1,159951881	2,017755156	0,208796010	3,697886442	C2		20	
20	0,136617407	-0,375886075	1,535366079	1,025532554	2,929196817	C2		21	
21	-1,900707583	-0,929882502	3,491196157	2,434302385	1,144658871	C3			22
22	-0,421146611	1,948192195	1,314921521	1,108065372	3,403973662	C2		23	
23	1,170727112	0,76188329	0,039244384	2,201485131	3,772505850	C1	24		
24	-2,751418338	-0,340076632	4,053043349	3,397544215	1,027443721	C3			25
25	0,819698614	0,471549395	0,448645588	1,834513186	3,425964832	C1	26		
26	-2,657463014	-1,687465595	4,064355570	3,644086274	1,042586447	C2		27	
27	0,179515543	-0,699285984	1,502165013	0,713161552	3,553630316	C2		28	
28	-2,447760374	2,411888104	3,938755435	4,767106020	1,750343468	C3			29
29	0,499676066	-1,334834894	2,215281279	0,000000000	3,691080973	C2		30	
30	-2,795017826	2,378842771	4,240467442	4,964519946	1,721623922	C3			199
...	...	...	...	...	...	...	...	...	...

	C1			C2			C3	
9	0,006760755	0,155579633	1	0,465907698	-1,484259433	2	-1,008888035	0,586970516
17	1,231184322	1,106991286	3	1,451489109	-0,49489106	5	-1,607034247	0,872520723
24	1,170727212	0,76188329	4	1,206987927	-1,044866759	7	-1,463134723	0,332783253
26	0,819698614	0,471549395	6	0,465907698	-1,484259433	8	-2,528813948	-0,190571734
31	1,319323289	0,865385371	14	-0,167174099	-1,720796929	10	-2,610016529	-0,9698183
33	1,106990768	0,671953557	18	1,039562014	-1,112156199	11	-1,252386994	-0,295342385
34	1,389702006	0,70485208	20	0,613744929	-1,159951881	12	-3,252573188	0,171907233
35	0,710151529	0,480858149	21	0,136617407	-0,375886075	13	-2,264210242	0,020960517
36	-0,066174778	0,181122277	23	-0,421146611	-1,948192195	15	-2,467132756	-0,917184845
37	1,022189448	1,529682926	27	0,179515543	-1,687465595	16	-1,463134723	0,332783253
38	1,106990768	0,671953557	28	0,699285984	-0,653020704	19	-1,883303798	0,173120485
41	0,542726616	0,413588709	30	0,499676066	-1,334834894	22	-1,900707583	-0,929882502
44	0,518445984	0,565015348	32	-0,21787486	-0,916575582	25	-2,751418338	-0,340076632
47	1,34234997	1,682114828	42	-0,556851854	-1,197196743	29	-2,447760374	2,411888104
48	1,64259997	1,940166076	43	-1,200535212	-1,022665544	39	-3,378093166	0,137904884
50	1,146493508	0,056855671	45	-0,115886611	-1,364433859	40	-2,960615975	0,668329767
51	0,458076578	1,392084359	55	0,432139329	-1,653683972	61	-2,300182284	1,218092578
54	1,441572256	0,488278355	55	0,260415967	-2,307525514	62	-3,957852255	0,998133452
56	1,199153893	1,580809547	59	0,432139329	-1,653683972	65	-2,76957719	-0,117306633
57	0,912761739	1,377405585	63	0,617088403	-0,42499784	76	-1,929649133	-0,12174753
58	-0,209482366	2,269100425	64	0,617088403	-0,42499784	84	-1,436600351	-0,548450514
60	1,249286845	0,776655638	66	-0,454134063	-1,934111091	90	-1,596792267	0,41498352
67	0,335157591	0,358983796	68	-0,277496617	-1,67298447	93	-0,813821066	0,507759489
69	0,596662345	0,597239555	70	0,432139329	-1,653683972	94	-2,820544214	1,77185586
73	1,679625708	1,008663087	71	0,241832327	-0,716559709	99	-2,761249458	2,52826731
77	1,249286845	0,776655638	72	-0,302953357	-0,864849498	103	-2,727481089	2,67769185
80	0,596662345	0,597239555	74	0,7021669	-0,476701362	107	-3,039604761	1,274289243
82	1,384560319	1,374357795	75	0,744177248	-0,784859195	113	-1,244459158	3,390292549
83	1,249286845	0,776655638	78	-0,623555267	-1,199479448	114	-3,044159599	1,029567541
...	...	...	...	...	...	...	...	...
195	1,152622331	1,127881552	194	0,590742106	-0,17831126	199	-2,795017826	2,378842771

Gambar 0.6. Iterasi-1 dengan Metode K-Means

Lampiran 9. Perhitungan Iterasi-2 untuk Klasterisasi dengan Metode K-Means



	C1			C2			C3		C1 (mean C1)	C2 (mean C2)	C3 (mean C3)	jarak terdekat	C1	C2	C3
9	0,066760755	0,155579633	1	0,465907698	-1,484259433	2	-1,008888035	0,586970516	2,368114411	0,153192675	3,746568719	C2		1	
17	1,231184322	1,106993286	3	1,45148909	-0,49489106	5	-1,605034247	0,872520723	2,157822969	2,443174509	1,593800768	C3			2
24	1,170727212	0,76188329	4	1,206987927	-1,04886759	7	-1,463134723	0,332783253	1,318248211	1,269430468	4,215733109	C2		3	
26	0,819698614	0,471549395	6	0,465907698	-1,484259433	8	-2,526813948	-0,190571734	1,832065780	0,764442011	4,175230192	C2		4	
31	1,313932289	0,865585371	14	0,167741909	-1,720706029	10	-2,610016539	-0,9669183	2,746138443	3,040954807	1,016225619	C3			5
33	1,106090768	0,671955357	18	1,009562014	-1,112136199	11	-1,255386694	-0,293934283	2,368114411	0,153192675	3,746568719	C2		6	
34	1,389702006	0,70485208	20	0,613744929	-1,159951881	12	-3,525773188	0,171907323	2,642091173	2,575573038	1,185912288	C3			7
35	0,710151529	0,480858149	21	0,136173407	-0,375886075	13	-2,264210242	0,029066317	3,794369368	3,235580316	0,861966796	C3			8
36	-0,066574778	0,181122277	23	-0,423146611	-1,948192193	15	-2,474943056	-0,371484843	1,244455258	1,552015183	2,716193801	C1	9		
37	1,022189448	1,329682926	27	0,179515543	-1,687665595	16	-1,463134723	0,332783253	4,140194995	3,153116510	1,657439752	C3			10
38	1,106090768	0,671955357	28	0,699285984	-0,650207004	19	-1,883303798	0,173120485	2,608764983	2,023561387	1,669953861	C3			11
41	0,542725616	0,415388709	30	0,499676066	-1,334834894	22	-1,900707583	-0,929882502	4,507535021	4,111488064	0,878742828	C3			12
44	0,576495984	0,565013248	32	-0,21787486	-0,916555302	25	-2,751418338	-0,348076652	3,487086221	3,082092477	0,722287153	C3			13
47	1,34234997	1,682511628	42	-0,363851854	-1,191760743	29	-2,447760374	2,411886104	2,827117398	0,770937079	3,409676633	C2		14	
48	1,65400932	0,401656176	43	1,230895752	-1,022686944	39	-3,377083316	0,373055384	3,795473698	3,126723883	1,04728270	C3			15
49	1,34234997	1,682511628	46	-0,740526017	-2,127515253	40	-2,946079155	1,557994983	2,642091173	2,575573038	1,185912288	C3			16
50	1,166393508	-0,008556671	52	0,153886631	-1,964332859	45	-2,606015971	0,668229767	0,333798829	2,549044760	3,858833979	C1	17		
51	0,456076578	1,392084509	53	0,432139329	-1,633683972	61	-2,30108284	1,218092578	1,900742338	0,384013310	4,052233797	C2		18	
54	1,441575256	0,488278355	55	0,769415067	-2,307532514	62	-3,957852255	-0,991332452	3,084557344	2,820021712	0,871590474	C3			19
56	1,199153893	1,580809547	59	0,432139329	-1,633683972	65	-2,76957719	-0,117306633	2,015735136	0,208796010	3,697886442	C2		20	
57	0,912761739	1,377405365	63	0,617088403	-0,424997841	76	-1,926491323	-0,12174753	1,535366079	1,035532554	2,929196817	C2		21	
58	-0,209482286	2,209100425	64	0,617088403	-0,424997841	84	-1,436600351	-0,548450314	3,491196157	2,434302385	1,744658871	C3			22
60	1,249286845	0,776655638	66	-0,454134063	-1,924111091	90	-1,596792267	0,414938352	3,149321521	1,108065372	3,403973662	C2		23	
67	0,35315291	0,358983796	68	-0,277169617	-1,67298447	93	-0,813821066	0,507759489	0,039244384	2,201485131	3,772505850	C1	24		
69	0,396662445	0,597239355	70	0,432139329	-1,633683972	94	-2,820544214	1,77185586	4,053043349	3,397544215	1,027437221	C3			25
73	1,679625708	1,008665887	71	0,241852327	-0,716559709	99	-2,761249458	2,52826731	0,44864588	1,854513186	3,425964852	C1	26		
77	1,249286845	0,776655638	72	-0,302953357	-0,864849498	103	-2,727481089	2,67769185	2,653465914	0,476435570	3,644086274	C2		27	
80	0,396662445	0,597239355	74	0,7021669	-0,476701362	107	-3,039604761	-1,274289243	1,501265013	0,713161552	3,553603216	C2		28	
82	1,384360319	1,374355795	75	0,744177248	-0,784559195	113	-1,244459158	3,300292549	3,958753435	4,767106020	1,790343468	C3			29
...	...	...	...	...	...	...	...	...	2,212381279	0,000000000	3,691080973	C2		30	
195	1,155622331	1,127881552	194	0,590742106	-0,178311268	199	-2,795017826	2,378842771	...	...	...	C1	...	...	...
MEAN	1,06670738	0,925760341	MEAN	0,447341771	-1,243210131	MEAN	-2,317422169	0,485892555	4,244947442	4,964515046	1,721623922	C3			199

Gambar 0.7. Iterasi-2 dengan Metode K-Means

Lampiran 10. *Listing* Program Klasterisasi Data dengan Metode K-Means

```
clc;
clear all;
% Menghapus file Excel lama jika ada
result_file = 'hasil_kmeans.xlsx';
if exist(result_file, 'file') == 2
delete(result_file);
end
% Membaca data dari file Excel
filename = 'data_kmeans.xlsx';
sheet = 'Sheet2';
xlRange = 'A1:K200';
[num, txt, raw] = xlsread(filename, sheet, xlRange);
% Ekstrak angka-angka di kolom pertama untuk identifikasi
numbers = num(:, 1);
% Membaca label kriteria dari baris pertama
labels = raw(1, 2:end);
% Ubah data menjadi nilai numerik, abaikan data non-numerik
% Lewati baris pertama (header) dan kolom pertama (indeks)
numericData = cellfun(@(x) str2double(x), raw(2:end, 2:end));
% Cek jika data tidak kosong dan memiliki lebih dari
 satu kolom
if isempty(numericData)
error('Data kosong. Mohon periksa kembali file dan sheet.');
elseif size(numericData, 2) <= 1
error('Data tidak memiliki kolom yang cukup.
 Pastikan data dimuat dengan benar.');
end
% Normalisasi data menggunakan mean dan standar deviasi
data_mean = mean(numericData, 1);
data_std = std(numericData, 1);
numericData = bsxfun(@rdivide, bsxfun(@minus, numericData,
 data_mean), data_std);
```

```
% Tampilkan hasil normalisasi
disp('Hasil normalisasi data:');
for i = 1:length(numbers)
fprintf('Data %d: ', numbers(i));
fprintf('%f ', numericData(i, :));
fprintf('\n');
end
disp('========================================');
% Lakukan PCA untuk mengurangi dimensi
[coeff, score, ~] = pca(numericData);
% Tampilkan hasil reduksi PCA
disp('Hasil reduksi PCA (2 komponen utama):');
for i = 1:length(numbers)
fprintf('Data %d (PCA1, PCA2): ', numbers(i));
fprintf('%f ', score(i, 1:2));
fprintf('\n');
end
disp('========================================');
% Tetapkan nilai acak agar hasil K-means tetap konsisten
rng(6);
% Tentukan rentang jumlah cluster yang akan dievaluasi
max_clusters = 15;
mean_silhouette_vals = zeros(max_clusters, 1);
% Loop untuk menghitung nilai Silhouette untuk
 setiap jumlah cluster
for k = 2:max_clusters
disp(['Jumlah cluster: ', num2str(k)]);
% Jalankan K-means pada data hasil PCA
[idx, C, sumd, D] = kmeans(score(:, 1:2), k, 'Replicates',
 5, 'Display', 'final');
% Menghitung nilai Silhouette
silhouette_vals = silhouette(score(:, 1:2), idx);
% Simpan nilai Silhouette rata-rata untuk
 jumlah cluster ini
mean_silhouette_vals(k) = mean(silhouette_vals);
% Tampilkan hasil dari setiap iterasi k-means
```

```
disp(['Jumlah sum of distances: ', num2str(sum(sumd))]);
disp(['Nilai rata-rata: ', num2str
 (mean_silhouette_vals(k))]);
disp('======================================');
end
% Plot nilai Silhouette rata-rata untuk
 setiap jumlah cluster
figure;
plot(2:max_clusters, mean_silhouette_vals(2:end), 'b-o');
title('Metode Silhouette untuk Menentukan
 Jumlah Cluster Optimal');
xlabel('Jumlah Cluster');
ylabel('Nilai Silhouette Rata-rata');
grid on;
% Cari jumlah cluster optimal berdasarkan
 nilai Silhouette tertinggi
[~, optimal_clusters] = max(mean_silhouette_vals);
disp(['Jumlah cluster optimal berdasarkan metode Silhouette:
 ', num2str(optimal_clusters)]);
disp('================================================');
% Iterasi K-means dengan data PCA yang dinormalisasi
converged = false;
prev_idx = [];
iter = 0;
while ~converged && iter < max_clusters
iter = iter + 1;
[idx_iter, C_iter] = kmeans(score(:, 1:2), optimal_clusters,
 'Replicates', 5, 'Display', 'final');
disp('================================================');
disp(['Proses iterasi untuk ', num2str(optimal_clusters),
 ' cluster, iterasi ke-', num2str(iter)]);
disp(['Indeks cluster: ', num2str(unique(idx_iter)')]);
disp(['Centroid untuk setiap cluster:']);
disp(C_iter);
% Menampilkan perhitungan jarak ke centroid
 untuk 5 data teratas
```

```
disp('Perhitungan jarak Euclidean ke centroid
 untuk 5 data teratas:');
for j = 1:min(5, size(score, 1))  % Batasi ke 5 data
 teratas atau sebanyak data yang ada
disp(['Data ke-', num2str(numbers(j))]);
% Menentukan data mana yang paling dekat dengan centroid
for i = 1:optimal_clusters
% Menghitung jarak Euclidean dari data ke centroid
distance = sqrt(sum((score(j, 1:2) - C_iter(i, :)).^2, 2));
disp(['  Jarak ke Centroid ', num2str(i), ': ',
 num2str(distance)]);
% Ambil data yang berada di cluster ke-i
cluster_points = score(idx_iter == i, 1:2);
% Menghitung jarak Euclidean dari setiap titik ke centroid
centroid = C_iter(i, :);  % Centroid untuk cluster i
% Menghitung jarak Euclidean menggunakan bsxfun
distances = sqrt(sum(bsxfun(@minus, cluster_points,
 centroid).^2, 2));
% Temukan indeks data yang paling dekat dengan centroid
[~, minIndex] = min(distances);
% Temukan indeks asli dari data tersebut
centroidDataIndex = find(idx_iter == i);
% Menampilkan data yang paling dekat dengan centroid
disp(['Centroid ', num2str(i), ' paling dekat dengan data ke-',
 num2str(numbers(centroidDataIndex(minIndex)))]);
end
disp('=============================================');
end
% Tampilkan perhitungan jarak ke centroid
 untuk 5 data terbawah
disp('Perhitungan jarak Euclidean ke centroid
 untuk 5 data terbawah:');
for j = max(1, size(score, 1)-4):size(score, 1)
 % Ambil 5 data terbawah atau sesuai dengan jumlah data
disp(['Data ke-', num2str(numbers(j))]);
% Menentukan data mana yang paling dekat dengan centroid
```

```
for i = 1:optimal_clusters
% Menghitung jarak Euclidean dari data ke centroid
distance = sqrt(sum((score(j, 1:2) - C_iter(i, :)).^2, 2));
disp([’  Jarak ke Centroid ’, num2str(i), ’: ’,
 num2str(distance)]);
% Ambil data yang berada di cluster ke-i
cluster_points = score(idx_iter == i, 1:2);
% Menghitung jarak Euclidean dari setiap titik ke centroid
centroid = C_iter(i, :);  % Centroid untuk cluster i
distances = sqrt(sum(bsxfun(@minus, cluster_points,
 centroid).^2, 2));
% Temukan indeks data yang paling dekat dengan centroid
[~, minIndex] = min(distances);
% Temukan indeks asli dari data tersebut
centroidDataIndex = find(idx_iter == i);
% Menampilkan data yang paling dekat dengan centroid
disp([’Centroid ’, num2str(i), ’ paling dekat dengan data ke-’,
 num2str(numbers(centroidDataIndex(minIndex)))]);
end
disp(’======================================’);
end
% Cek konvergensi
if isempty(prev_idx)
prev_idx = idx_iter;
else
if isequal(prev_idx, idx_iter)
converged = true;
else
prev_idx = idx_iter;
end
end
% Tampilkan data yang berada di setiap cluster
for i = 1:optimal_clusters
clusterData = [numbers(idx_iter == i),
 numericData(idx_iter == i, :)];
disp([’Data yang berada di cluster ’, num2str(i), ’:’]);
```

```
disp(clusterData);
end
disp('================================================');
end
% Run k-means clustering again with normalized data
 on PCA scores
[idx_optimal, C_optimal] = kmeans(score(:, 1:2),
 optimal_clusters, 'Replicates', 5, 'Display', 'final');
% Plot hasil clustering pada data PCA
figure;
hold on;
colors = lines(optimal_clusters);
for i = 1:optimal_clusters
% Plot setiap cluster dengan warna berbeda
scatter(score(idx_optimal == i, 1), score(idx_optimal == i, 2),
 36, colors(i, :), 'filled');
end
% Plot centroids
plot(C_optimal(:, 1), C_optimal(:, 2), 'kx', 'MarkerSize',
 15, 'LineWidth', 3);
% Tambahkan nomor data ke plot clustering dengan warna
 sesuai cluster
no = raw(2:end, 1); % Ekstrak kolom 'NO' dari data asli
for i = 1:optimal_clusters
cluster_points = find(idx_optimal == i);
text(score(cluster_points, 1), score(cluster_points, 2),
 num2str(cell2mat(no(cluster_points))), ...
'VerticalAlignment', 'bottom', 'HorizontalAlignment',
 'right', 'Color', colors(i, :));
end
legend(arrayfun(@(x) ['Cluster ' num2str(x)],
 1:optimal_clusters, 'UniformOutput', false), 'Centroids');
title(['Hasil Clustering dengan ', num2str(optimal_clusters),
 ' Cluster (PCA Data)']);
xlabel('Principal Component 1');
ylabel('Principal Component 2');
```

```
hold off;
% Tambahkan nomor dan hasil clustering ke data asli
result = [numbers, numericData, idx_optimal];
extendedLabels = [{'No'}, labels, {'Cluster'}];
% Menampilkan hasil sebelum disimpan
disp('Hasil clustering untuk semua data:');
disp(extendedLabels);
disp(result);
% Menampilkan hasil untuk setiap cluster
for i = 1:optimal_clusters
clusterData = result(result(:, end) == i, :);
% Menampilkan data cluster sebelum disimpan
disp(['Hasil clustering untuk Cluster ' num2str(i) ':']);
disp(extendedLabels);
disp(clusterData);
end
% Perhitungan Silhouette Score
silhouette_vals = silhouette(score(:, 1:2), idx);
 % 'score(:, 1:2)' adalah data, 'idx' adalah label hasil
 clustering
% Menghitung Silhouette Score rata-rata
mean_silhouette_score = mean(silhouette_vals);
% Tampilkan hasil Silhouette Score
fprintf('Nilai rata-rata Silhouette untuk hasil klasterisasi:
 %.4f\n', mean_silhouette_score);
% Interpretasi hasil
if mean_silhouette_score > 0.7
    disp('Tingkat akurasi pada proses klasterisasi
    menunjukkan hasil sangat baik.');
elseif mean_silhouette_score > 0.5
    disp('Tingkat akurasi pada proses klasterisasi
    menunjukkan hasil baik.');
elseif mean_silhouette_score > 0.25
    disp('Tingkat akurasi pada proses klasterisasi
    menunjukkan hasil agak baik.');
else
```

```
    disp('Tingkat akurasi pada proses klasterisasi
    menunjukkan hasil buruk.');
end
% Mengubah nilai silhouette menjadi persentase
silhouette_percentage = mean_silhouette_score * 100;
% Menampilkan hasil dalam bentuk persentase
fprintf('Tingkat akurasi pada proses klasterisasi menunjukkan
 hasil %.2f%%\n', silhouette_percentage);
disp('=============================================');
% Menambahkan header untuk hasil normalisasi
normalized_labels = [{'No'}, labels];
normalized_data_with_header = [normalized_labels;
 num2cell([numbers, numericData])];
% Menulis hasil normalisasi data ke dalam sheet baru
 di file Excel
normalized_sheet_name = 'NormalizedData';
xlswrite(result_file, normalized_labels,
 normalized_sheet_name, 'A1');
xlswrite(result_file, normalized_data_with_header(2:end, :),
 normalized_sheet_name, 'A2');
disp(['Hasil normalisasi telah disimpan di sheet "',
 normalized_sheet_name, '" pada file ', result_file]);
% Simpan hasil normalisasi dan PCA ke dalam file Excel
pca_result = [numbers, score(:, 1:2)];
 % Gabungkan nomor data dengan hasil PCA
normalized_and_pca_labels = [{'No'}, {'PCA1'}, {'PCA2'}];
% Tulis hasil normalisasi dan PCA ke sheet baru
normalized_pca_sheet_name = 'NormalizedAndPCA';
xlswrite(result_file, normalized_and_pca_labels,
normalized_pca_sheet_name, 'A1');
xlswrite(result_file, pca_result,
 normalized_pca_sheet_name, 'A2');
disp(['Hasil normalisasi dan PCA telah disimpan di sheet "',
 normalized_pca_sheet_name, '" pada file ', result_file]);
% Simpan hasil clustering ke file Excel
xlswrite(result_file, extendedLabels, 'Sheet1', 'A1');
```

```
xlswrite(result_file, result, 'Sheet1', 'A2');
% Simpan data asli untuk Cluster 1 dan Cluster 2 ke dalam
 sheet Excel yang berbeda
cluster1_data = raw([false; idx_optimal == 1], :);
cluster2_data = raw([false; idx_optimal == 2], :);
cluster3_data = raw([false; idx_optimal == 3], :);
xlswrite(result_file, raw(1, :), 'Cluster1', 'A1');
xlswrite(result_file, raw(1, :), 'Cluster2', 'A1');
xlswrite(result_file, raw(1, :), 'Cluster3', 'A1');
 % Tulis header
xlswrite(result_file, cluster1_data, 'Cluster1', 'A2');
xlswrite(result_file, cluster2_data, 'Cluster2', 'A2');
xlswrite(result_file, cluster3_data, 'Cluster3', 'A2');
 % Tulis data cluster
disp(['Hasil clustering telah disimpan di ', result_file]);
```

Lampiran 11. Hasil Klasterisasi dengan Metode K-Means

Keterangan (variabel data):

V1 : Variabel usia
V2 : Variabel jenis kelamin
V3 : Variabel kondisi ekonomi keluarga
V4 : Variabel pendidikan terakhir
V5 : Variabel hobi setelah dikategorikan menjadi 5 jenis
V6 : Variabel capaian penghargaan
V7 : Variabel jenis pekerjaan setelah dikategorikan menjadi 15 jenis
V8 : Variabel golongan gaji
V9 : Variabel hubungan antara hobi dengan pekerjaan
V10 : Variabel hubungan antara capaian penghargaan dengan pekerjaan

Cluster 1										
NO	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
9	1	1	2	2	2	2	9	1	1	3
17	1	1	2	4	5	2	6	1	2	3
24	2	2	2	4	3	2	4	1	2	3
26	1	2	2	2	2	2	9	1	2	3
31	2	2	2	4	4	2	4	1	2	3
33	1	2	2	2	4	2	9	1	2	3
34	2	2	2	2	4	2	9	1	2	3
35	1	2	2	4	2	2	3	1	2	3
36	1	2	2	2	1	2	9	1	1	3
37	1	2	2	2	4	3	2	8	1	2
38	1	2	2	2	4	2	2	9	1	2
41	1	2	2	2	4	2	2	10	2	2
44	1	2	2	2	4	2	2	11	2	2
47	1	2	2	2	4	2	5	2	9	1
48	2	2	1	2	2	2	5	2	11	1
49	1	2	2	2	4	2	5	2	9	1
50	1	2	2	1	4	4	2	2	10	2
51	1	2	2	2	4	4	2	2	9	1
54	3	2	2	2	3	3	2	2	5	1
56	1	2	2	2	4	4	2	2	9	1
57	1	2	2	2	4	2	2	2	9	1
58	1	2	2	3	4	1	2	12	1	1
60	1	2	2	2	2	5	2	2	9	1
67	1	1	2	2	2	4	2	2	9	1
69	1	2	2	2	4	5	2	2	3	1
73	1	2	2	1	4	5	2	2	9	1
77	1	2	2	2	2	5	2	2	9	1
80	1	2	2	2	4	5	2	2	3	1
82	1	2	2	2	2	5	2	13	1	1
...	...	...	...	...	...	...	...	...	...	...
195	1	2	2	3	4	2	9	1	2	3

Cluster 2										
NO	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	1	2	2	2	4	2	2	2	2	3
3	1	2	1	2	5	2	5	1	2	3
4	1	2	1	2	4	2	2	1	2	3
6	1	2	2	2	4	2	2	2	2	3
14	1	2	2	2	5	2	1	2	1	3
18	1	2	1	2	4	2	9	2	2	3
20	2	2	1	4	5	2	3	2	1	3
21	1	2	2	2	5	2	10	2	1	3
23	2	2	2	2	1	2	2	2	1	3
27	1	2	2	2	2	2	2	2	2	3
28	2	2	2	4	3	2	2	2	2	3
30	1	2	2	2	4	2	3	2	2	3
32	1	2	2	4	4	2	1	2	1	3
42	2	2	2	4	1	2	3	1	2	3
43	1	1	1	2	4	2	3	1	2	3
46	1	2	2	2	1	2	1	2	1	3
52	1	1	2	2	2	2	1	2	2	3
53	1	2	2	2	4	2	1	2	2	3
55	1	2	1	2	4	2	1	2	2	3
59	1	2	2	2	4	2	1	2	2	3
63	1	2	2	2	2	2	3	1	2	3
64	1	2	2	2	2	2	3	1	2	3
66	1	2	2	2	3	2	1	2	1	3
68	1	2	2	2	4	2	2	2	1	3
70	1	2	2	2	4	2	1	2	2	3
71	1	1	2	3	1	2	7	2	2	3
72	1	2	2	2	1	2	1	2	1	3
74	1	2	2	4	5	2	2	2	2	3
75	1	2	2	2	5	2	6	2	2	3
...	...	...	...	...	...	...	...	...	...	...
198	2	1	2	2	2	2	1	2	1	3

Cluster 3										
NO	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
2	1	1	1	4	4	1	3	1	2	2
5	1	2	2	4	4	1	7	2	2	2
7	1	2	2	2	4	1	2	1	2	2
8	1	1	2	4	4	1	2	2	1	2
10	1	2	2	2	4	1	2	2	1	2
11	1	2	1	2	3	1	3	1	2	2
12	1	2	2	2	4	2	5	1	1	1
13	1	2	2	2	2	2	5	1	8	2
15	1	2	2	2	4	1	6	2	1	2
16	1	2	2	2	4	1	2	1	2	2
19	1	2	2	4	3	1	3	2	2	2
22	1	2	2	2	4	1	1	2	2	2
25	1	2	2	2	4	1	3	2	2	2
29	1	2	2	4	1	1	10	1	2	1
39	1	2	2	2	2	1	3	1	1	1
40	1	2	2	2	5	1	15	2	1	1
45	1	2	2	2	1	1	15	2	1	2
61	1	2	2	3	4	1	3	1	2	1
62	1	2	2	2	1	1	2	2	1	1
65	1	2	2	4	2	1	3	2	1	2
76	1	2	2	2	1	1	1	1	2	2
84	1	2	2	1	4	4	1	2	2	2
90	1	2	2	2	4	1	10	2	2	2
93	1	2	2	1	2	5	7	1	2	2
94	1	2	2	2	4	1	11	1	1	1
99	1	2	2	2	4	1	10	1	1	1
103	1	2	2	2	4	1	11	1	1	1
107	1	2	2	2	1	1	2	2	1	2
113	2	1	2	2	5	1	14	1	1	2
...	...	...	...	...	...	...	...	...	...	...
199	1	2	2	4	4	1	9	1	1	1

Gambar 0.8. Hasil Klasterisasi dengan Metode K-Means

Lampiran 12. *Listing* Program Simulasi Sistem Rekomendasi Pekerjaan

Tampilan Awal

```
function keluar_Callback(hObject, eventdata, handles)
respon=keluar('Title','Konfirmasi Keluar');
switch(respon)
case 'Tidak'
%tidak ada aksi
case 'Ya'
close
end
% --------------------------------------------------
function layout_ta_1_Callback(hObject, eventdata, handles)
close
layout_ta_1 % Grafik hasil klasterisasi
% --------------------------------------------------
function manual_Callback(hObject, eventdata, handles)
close
manual % Simulasi sistem rekomendasi pekerjaan
% --------------------------------------------------
function axes1_CreateFcn(hObject, eventdata, handles)
% Read the image
 % Read the image and the alpha channel
[img, ~, alpha] = imread('logo_uin.png');
% Set the desired background color
backgroundColor = [0.95, 0.97, 0.95];
% Create an axes with the specified background color
axes(hObject);
set(hObject, 'Color', backgroundColor);
% Create a background with the specified color
coloredBackground = cat(3, backgroundColor(1) *
 ones(size(img, 1), size(img, 2)), ...
backgroundColor(2) * ones(size(img, 1), size(img, 2)), ...
backgroundColor(3) * ones(size(img, 1), size(img, 2)));
```

```
% Display the colored background
imshow(coloredBackground, 'Parent', hObject);
hold on;
% Display the image with the specified background color
h = imshow(img, 'Parent', hObject);
set(h, 'AlphaData', alpha);
hold off;
```

Grafik Hasil Klasterisasi

```
% --------------------------------------------------
function pushbutton1_Callback(hObject, eventdata, handles)
% Reset
axes(handles.axes1);
plot(0);
set(handles.edit1,'string','');
% --------------------------------------------------
function pushbutton2_Callback(hObject, eventdata, handles)
% Process
% Membaca data dari file Excel
filename = 'data_kmeans.xlsx';
sheet = 'Sheet2';
xlRange = 'A1:K200';
[num, txt, raw] = xlsread(filename, sheet, xlRange);
...
% Selengkapnya sama dengan listing program klasterisasi
 data dengan metode k-means (lampiran 10)
% Jalankan K-means pada data hasil PCA dengan jumlah
 cluster optimal
[idx_optimal, C_optimal] = kmeans(score(:, 1:2),
 optimal_clusters, 'Replicates', 5, 'Display', 'final');
% Plot hasil clustering pada data PCA ke dalam axes1
axes(handles.axes1); % Mengarahkan plot ke axes1
cla; % Bersihkan axes sebelum plot baru
hold on;
% Warna untuk setiap cluster
custom_colors = [0 0 1; 0 1 0; 1 0 0];
```

```
 % Hijau, Merah, Biru
% Pastikan custom_colors sesuai dengan jumlah cluster optimal
if optimal_clusters <= 3
colors = custom_colors(1:optimal_clusters, :);
else
colors = lines(optimal_clusters);
end
% Plot hasil clustering pada data PCA ke dalam axes1
axes(handles.axes1); % Mengarahkan plot ke axes1
cla; % Bersihkan axes sebelum plot baru
hold on;
for i = 1:optimal_clusters
% Plot setiap cluster dengan warna yang sudah ditentukan
scatter(score(idx_optimal == i, 1), score(idx_optimal
 == i, 2), 36, colors(i, :), 'filled');
end
% Plot centroids
plot(C_optimal(:, 1), C_optimal(:, 2), 'kx', 'MarkerSize',
 15, 'LineWidth', 3);
% Tambahkan nomor data ke plot clustering dengan
 warna sesuai cluster
no = raw(2:end, 1); % Ekstrak kolom 'NO' dari data asli
for i = 1:optimal_clusters
cluster_points = find(idx_optimal == i);
text(score(cluster_points, 1), score(cluster_points, 2),
 num2str(cell2mat(no(cluster_points))), ...
'VerticalAlignment', 'bottom', 'HorizontalAlignment',
 'right', 'Color', colors(i, :));
end
legend(arrayfun(@(x) ['Cluster ' num2str(x)],
 1:optimal_clusters, 'UniformOutput', false), 'Centroids');
title(['Hasil Clustering dengan ', num2str(optimal_clusters),
 ' Cluster (PCA Data)']);
xlabel('Principal Component 1');
ylabel('Principal Component 2');
hold off;
```

```
keterangan = {...
'Data yang ada dalam cluster 1 merupakan data untuk',...
'',...
};
% Menampilkan keterangan di edit3
set(handles.edit1,'Max',1000);
set(handles.edit1, 'String', keterangan);
guidata(hObject, handles);
% --------------------------------------------------------
function pushbutton3_Callback(hObject, eventdata, handles)
% Back
close
cover % Tampilan awal
```

Simulasi Sistem Rekomendasi Pekerjaan

```
% --------------------------------------------------------
function popupmenu1_Callback(hObject, eventdata, handles)
pilih_usia=get(hObject,'value');
switch pilih_usia
case 1
case 2
    pilih_usia=1;
    handles.mode.usia=pilih_usia;
    guidata(hObject,handles);
case 3
    pilih_usia=2;
    handles.mode.usia=pilih_usia;
    guidata(hObject,handles);
case 4
    pilih_usia=3;
    handles.mode.usia=pilih_usia;
    guidata(hObject,handles);
end
handles.mode.var1 = pilih_usia;
guidata(hObject, handles);
end
```

```
% --------------------------------------------------------
function popupmenu2_Callback(hObject, eventdata, handles)
jenis_kelamin=get(hObject,'value');
switch jenis_kelamin
case 1
case 2
    jenis_kelamin=1;
    handles.mode.gender=jenis_kelamin;
    guidata(hObject,handles);
case 3
    jenis_kelamin=2;
    handles.mode.gender=jenis_kelamin;
    guidata(hObject,handles);
end
handles.mode.var2 = jenis_kelamin;
guidata(hObject, handles);
end
...
% Selengkapnya tuliskan script yang sama untuk
 popupmenu3_Callback hingga popupmenu6_Callback

% --------------------------------------------------------
function pushbutton1_Callback(hObject, eventdata, handles)
% Reset
set(handles.popupmenu1,'value',1);
set(handles.popupmenu2,'value',1);
set(handles.popupmenu3,'value',1);
set(handles.popupmenu4,'value',1);
set(handles.popupmenu5,'value',1);
set(handles.popupmenu6,'value',1);
set(handles.edit1,'string','');
set(handles.edit2,'string','');
set(handles.edit3,'string','');
end
% --------------------------------------------------------
function pushbutton2_Callback(hObject, eventdata, handles)
```

```
% Process
% Ambil pilihan pengguna dari 'handles'
userChoices = [
    handles.mode.var1, ...
    handles.mode.var2, ...
    handles.mode.var3, ...
    handles.mode.var4, ...
    handles.mode.var5, ...
    handles.mode.var6
];
% Tentukan cluster dan informasi pekerjaan
[cluster, jobInfo] = determineCluster(userChoices);
% Tampilkan hasil cluster dan informasi pekerjaan
set(handles.edit1, 'String', cluster);
if ~isempty(jobInfo)
    set(handles.edit2,'Max',10);
    set(handles.edit2, 'String', jobInfo);
end
end
% ---define the determineCluster function---
function [cluster, jobInfo] = determineCluster(userChoices)
% Define clusters
cluster1 = {...
    [2, 3], ... % Variabel usia
    [2],    ... % Variabel jenis kelamin
    [2, 3], ... % Variabel ekonomi
    [1, 3, 4],... % Variabel pendidikan
    [2, 3, 5],... % Variabel hobi
    [2]};   % Variabel penghargaan
cluster2 = {...
    [3],    ... % Variabel usia
    [1],     ... % Variabel jenis kelamin
    [1],    ... % Variabel ekonomi
    [1, 2],... % Variabel pendidikan
    [2, 3],... % Variabel hobi
    [2]};   % Variabel penghargaan
```

```
cluster3 = {...
    [1],    ... % Variabel usia
    [1],    ... % Variabel jenis kelamin
    [2],     ... % Variabel ekonomi
    [5],    ... % Variabel pendidikan
    [1, 4],... % Variabel hobi
    [1]};   % Variabel penghargaan
% Initialize cluster and jobInfo as empty
cluster = 'Tidak termasuk ke dalam Cluster manapun';
jobInfo = 'Tidak ditemukan informasi pekerjaan terkait Cluster';
% Check for cluster 1
if all(cellfun(@(choices, selected) isempty(choices) ||
 ismember(selected, choices), cluster1, num2cell(userChoices)))
    cluster = 'Termasuk ke dalam cluster 1';
    jobInfo = sprintf(['- Pekerjaan yang cocok: PNS...\n' ...
                    ']);
    return;
end
...
% Selengkapnya tuliskan script yang sama untuk
 cluster 2 dan cluster 3

end
% --------------------------------------------------------
function pushbutton3_Callback(hObject, eventdata, handles)
% Back
close
cover % Tampilan awal
% --------------------------------------------------------
function pushbutton4_Callback(hObject, eventdata, handles) % Data
ket1 = {...
     'Kondisi ekonomi keluarga yang dimaksud dalam...',...
     '',...
     'Jenis hobi dibagi menjadi 5 kategori, yaitu:',...
     '',...
     'Capaian penghargaan yang dimaksud di sini ...',...
```

```
     '',...
     };
% Menampilkan keterangan di edit3
set(handles.edit3,'Max',110);
set(handles.edit3, 'String', ket1);
guidata(hObject, handles);
end
% --------------------------------------------------
function pushbutton5_Callback(hObject, eventdata, handles)
ket2 = {...
     'Jenis pekerjaan yang ada pada data sebelumnya...',...
     '',...
     };
% Menampilkan keterangan di edit3
set(handles.edit3,'Max',110);
set(handles.edit3, 'String', ket2);
guidata(hObject, handles);
end
```

Lampiran 13. Kombinasi Data Inputan untuk Setiap *Cluster* dalam Simulasi Sistem Rekomendasi Pekerjaan

Tabel 0.1. 42 Kombinasi data inputan pengguna

Cluster 1						
No	Usia	Jenis kelamin	Kondisi ekonomi keluarga	Pendidikan terakhir	Hobi	Capaian penghargaan
1.	25-29	Perempuan	Kelas menengah	SMP	*Activity*	Tidak pernah
2.	30-34	Perempuan	Kelas menengah	SMP	*Activity*	Tidak pernah
3.	25-29	Perempuan	Kelas atas	SMP	*Activity*	Tidak pernah
4.	30-34	Perempuan	Kelas atas	SMP	*Activity*	Tidak pernah
5.	25-29	Perempuan	Kelas menengah	Diploma	*Activity*	Tidak pernah
6.	30-34	Perempuan	Kelas menengah	Diploma	*Activity*	Tidak pernah
7.	25-29	Perempuan	Kelas atas	Diploma	*Activity*	Tidak pernah
8.	30-34	Perempuan	Kelas atas	Diploma	*Activity*	Tidak pernah
9.	25-29	Perempuan	Kelas menengah	Sarjana	*Activity*	Tidak pernah
10.	30-34	Perempuan	Kelas menengah	Sarjana	*Activity*	Tidak pernah
11.	25-29	Perempuan	Kelas atas	Sarjana	*Activity*	Tidak pernah
12.	30-34	Perempuan	Kelas atas	Sarjana	*Activity*	Tidak pernah
13.	25-29	Perempuan	Kelas menengah	SMP	*Play*	Tidak pernah
14.	30-34	Perempuan	Kelas menengah	SMP	*Play*	Tidak pernah
15.	25-29	Perempuan	Kelas atas	SMP	*Play*	Tidak pernah
16.	30-34	Perempuan	Kelas atas	SMP	*Play*	Tidak pernah
17.	25-29	Perempuan	Kelas menengah	Diploma	*Play*	Tidak pernah
18.	30-34	Perempuan	Kelas menengah	Diploma	*Play*	Tidak pernah
19.	25-29	Perempuan	Kelas atas	Diploma	*Play*	Tidak pernah
20.	30-34	Perempuan	Kelas atas	Diploma	*Play*	Tidak pernah
21.	25-29	Perempuan	Kelas menengah	Sarjana	*Play*	Tidak pernah
22.	30-34	Perempuan	Kelas menengah	Sarjana	*Play*	Tidak pernah
23.	25-29	Perempuan	Kelas atas	Sarjana	*Play*	Tidak pernah
24.	30-34	Perempuan	Kelas atas	Sarjana	*Play*	Tidak pernah
25.	25-29	Perempuan	Kelas menengah	SMP	*Lainnya*	Tidak pernah
26.	30-34	Perempuan	Kelas menengah	SMP	*Lainnya*	Tidak pernah
27.	25-29	Perempuan	Kelas atas	SMP	*Lainnya*	Tidak pernah
28.	30-34	Perempuan	Kelas atas	SMP	*Lainnya*	Tidak pernah
29.	25-29	Perempuan	Kelas menengah	Diploma	*Lainnya*	Tidak pernah
30.	30-34	Perempuan	Kelas menengah	Diploma	*Lainnya*	Tidak pernah
31.	25-29	Perempuan	Kelas atas	Diploma	*Lainnya*	Tidak pernah
32.	30-34	Perempuan	Kelas atas	Diploma	*Lainnya*	Tidak pernah
33.	25-29	Perempuan	Kelas menengah	Sarjana	*Lainnya*	Tidak pernah
34.	30-34	Perempuan	Kelas menengah	Sarjana	*Lainnya*	Tidak pernah
35.	25-29	Perempuan	Kelas atas	Sarjana	*Lainnya*	Tidak pernah
36.	30-34	Perempuan	Kelas atas	Sarjana	*Lainnya*	Tidak pernah
Hasil	Pekerjaan: PNS, nakes, karyawan, asisten, tenaga pendidik, *chef*, PHL Gaji tetap, Tidak berhubungan dengan hobi dan penghargaan					

Cluster 2						
No	Usia	Jenis kelamin	Kondisi ekonomi keluarga	Pendidikan terakhir	Hobi	Capaian penghargaan
1.	30-34	Laki-laki	Kelas bawah	SMP	*Activity*	Tidak pernah
2.	30-34	Laki-laki	Kelas bawah	SMA/SMK	*Activity*	Tidak pernah
3.	30-34	Laki-laki	Kelas bawah	SMP	*Play*	Tidak pernah
4.	30-34	Laki-laki	Kelas bawah	SMA/SMK	*Play*	Tidak pernah
Hasil	Pekerjaan: Wiraswasta dan *freelance* Gaji tidak tetap, Berhubungan dengan hobi, Tidak berhubungan dengan penghargaan					

Cluster 3						
No	Usia	Jenis kelamin	Kondisi ekonomi keluarga	Pendidikan terakhir	Hobi	Capaian penghargaan
1.	20-24	Laki-laki	Kelas menengah	Magister	*Making*	Pernah
2.	20-24	Laki-laki	Kelas menengah	Magister	*Arts*	Pernah
Hasil	Pekerjaan: Staff perusahaan, *part time*, pegawai magang, *entertainer*, TNI, lainnya Gaji tidak tetap, Berhubungan dengan hobi, Ada yang berhubungan dan ada yang tidak berhubungan dengan penghargaan					

Lampiran 14. Daftar Riwayat Hidup

DAFTAR RIWAYAT HIDUP

1. Identitas Diri

Nama Lengkap	: Marcela Andarista Maharani
Tempat, Tanggal Lahir	: Sukoharjo, 18 Maret 2002
Alamat	: Wonosari RT 04/RW 08, Kel. Wonosari, Kec. Ngaliyan, Kota Semarang
Nomor HP	: 0896-0373-5005
Email	: marcelareaverista1832@gmail.com

2. Riwayat Pendidikan
   (a) SD Negeri Wonosari 03 lulus tahun 2014
   (b) SMP Negeri 28 Semarang lulus tahun 2017
   (c) SMA Negeri 08 Semarang lulus tahun 2020
   (d) UIN Walisongo Semarang

Semarang, 16 September 2024

Marcela Andarista Maharani

NIM. 2008046035

Scanned by TapScanner