TUGAS AKHIR – KS184822

ANALISIS REKOMENDASI MENGGUNAKAN *SINGLE LINKAGE CLUSTERING* DAN *K-MODES CLUSTERING* DALAM PENDEKATAN *HYBRID FILTERING*

ANADIA RAHMAT SYIHAB HIDAYATULLAH
NRP 062115 4000 0001

Dosen Pembimbing
Dr. Dra. Kartika Fithriasari, M.Si.

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019

TUGAS AKHIR – KS184822

ANALISIS REKOMENDASI MENGGUNAKAN *SINGLE LINKAGE CLUSTERING* DAN *K-MODES CLUSTERING* DALAM PENDEKATAN *HYBRID FILTERING*

ANADIA RAHMAT SYIHAB HIDAYATULLAH
NRP 062115 4000 0001

Dosen Pembimbing
Dr. Dra. Kartika Fithriasari, M.Si.

PROGRAM STUDI SARJANA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019

FINAL PROJECT – KS184822

RECOMMENDATION ANALYSIS USING SINGLE LINKAGE CLUSTERING AND K-MODES CLUSTERING WITHIN HYBRID FILTERING

ANADIA RAHMAT SYIHAB HIDAYATULLAH
SN 062115 4000 0001

Supervisor
Dr. Dra. Kartika Fithriasari, M.Si.

UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS, COMPUTING, AND DATA SCIENCE
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2019

LEMBAR PENGESAHAN

ANALISIS REKOMENDASI MENGGUNAKAN *SINGLE LINKAGE CLUSTERING* DAN *K-MODES CLUSTERING* DALAM PENDEKATAN *HYBRID FILTERING*

TUGAS AKHIR

Diajukan untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Statistika
pada
Program Studi Sarjana Departemen Statistika
Fakultas Matematika, Komputasi, dan Sains Data
Institut Teknologi Sepuluh Nopember

Oleh:
Anadia Rahmat Syihab Hidayatullah
NRP. 062115 4000 0001

Disetujui oleh Pembimbing:
Dr. Dra. Kartika Fithriasari, M.Si. ( )
NIP. 19691212 199303 2 002

Mengetahui,
Kepala Departemen Statistika

Dr. Suhartono
NIP. 19710929 199512 1 001

SURABAYA, JULI 2019

ANALISIS REKOMENDASI MENGGUNAKAN *SINGLE LINKAGE CLUSTERING* DAN *K-MODES CLUSTERING* DALAM PENDEKATAN *HYBRID FILTERING*

Nama Mahasiswa	**: Anadia Rahmat Syihab Hidayatullah**
NRP	**: 062115 4000 0001**
Departemen	**: Statistika**
Dosen Pembimbing	**: Dr. Dra. Kartika Fithriasari, M.Si.**

Abstrak

Perkembangan teknologi telah membawa perubahan dalam bidang kehidupan manusia. Hingga pada saat era digital kini semua media informasi dapat diakses melalui internet. Salah satu media yang memanfaatkan teknologi internet dalam memperoleh keuntungan adalah dunia film dengan cara menyajikan film secara daring. Sejak awal perkembangannya hingga saat ini telah tercatat 3.361.741 judul film yang telah dikeluarkan industri perfilman. Sekian banyak film membawa dampak bagi user dan pemilik situs film daring terutama dalam hal efisiensi tampilan yang film yang relevan bagi user-user film. Oleh karena itu munculah pendekatan yang dapat memberikan bantuan bagi user dalam menentukan keputusan yakni menggunakan kombinasi sistem rekomendasi yakni teknik sistem rekomendasi Hybrid Filtering yaitu menggabungkan teknik Demographic Filtering (Single Linkage Clustering dan K-Modes Clustering) dan Collaborative Filtering. Hasil dari penelitian diperoleh jumlah kelompok user berdasarkan kemiripan karakteristik yang sama sejumlah 34 jenis klaster. Rekomendasi yang didapatkan sejumlah 10 film yang relevan melalui proses preferensi user lain dan prediksi rating.

Kata Kunci: Sistem Rekomendasi, Film, Single Linkage Clustering, K-Modes Clustering, Hybrid Filtering

(Halaman ini sengaja dikosongkan)

RECOMMENDATION ANALYSIS USING SINGLE LINKAGE CLUSTERING AND K-MODES CLUSTERING WITHIN HYBRID FILTERING

Name	**: Anadia Rahmat Syihab Hidayatullah**
Student Number	**: 062115 4000 0001**
Department	**: Statistics**
Supervisor	**: Dr. Dra. Kartika Fithriasari, M.Si.**

Abstract

Technological developments have brought changes in the field of human living. Until right now in this digital era, all media information can be accessed via the internet. One of the media that uses internet technology to gain profits is the world of movie by presenting movies online. Since the beginning of its development to date there have been 3,361,741 movie titles released by the movie industry. Those many movies have an impact on the users and owners of online movie sites, especially in terms of the efficiency of the appearance of the movies that are relevant to the users of the movie. Therefore, an approach that can provide assistance to users in determining decisions is using a combination of recommendation systems, it called a Hybrid Filtering recommendation system technique that combines Demographic Filtering techniques (Single Linkage Clustering and K-Modes Clustering) and Collaborative Filtering. The results of the study obtained the number of user groups based on similar characteristics of 34 types of clusters. The recommendations obtained are 10 relevant movies through other user preference processes and rating predictions.

Keywords: Recommendation System, Movies, Single Linkage Clustering, K-Modes Clustering, Hybrid Filtering

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadirat Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan penelitian Tugas Akhir yang berjudul "**Analisis Rekomendasi Menggunakan *Single Linkage Clustering* dan *K-Modes Clustering* Dalam Pendekatan *Hybrid Filtering***" dengan baik. Selama penyusunan laporan, penulis telah menerima bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT yang merahmati penulis dengan kekuatan, kesabaran dan kemudahan dalam menempuh Tugas Akhir.
2. Nabi Muhammad SAW, yang telah membawa dalam jalan indahnya Islam yang penuh ilmu.
3. Ibu Bapak yang sangat penulis cintai yang selalu mendoakan dan mendukung penulis dalam penyelesaian Tugas Akhir, dan tak lupa Adik tersayang Dheo Abid A.R. yang selalu menjadi motivasi penulis dalam belajar dan menjadi kakak yang terbaik.
4. Ibu Dr. Dra. Kartika Fithriasari, M.Si. selaku dosen pembimbing yang selalu sabar memberikan arahan dan masukan kepada penulis dalam proses penyelesaian Tugas Akhir.
5. Ibu Irhamah, M.Si., Ph.D. dan Ibu Pratnya Paramitha Oktaviana, S.Si., M.Si. selaku dosen penguji yang telah banyak membantu dan memberikan masukan untuk kesempurnaan Tugas Akhir ini.
6. Bapak Dr. Suhartono, M.Sc selaku Kepala Departemen Statistika FMKSD ITS dan Ibu Dr. Santi Wulan P., S.Si., M.Si. selaku Ketua Program Studi Sarjana Departemen Statistika FMKSD ITS yang telah memberikan nasehat dan keyakinan kepada penulis dalam menyelesaikan Tugas Akhir.

7. Bapak Imam Safawi, S.Si., M.Si. selaku dosen wali yang selalu sabar memberi dukungan dan motivasi.
8. Seluruh dosen dan karyawan Departemen Statistika ITS yang telah memberikan ilmu, pengalaman, dan semangatnya kepada penulis agar selalu positif dimana pun dan kapanpun.
9. Senior-senior Statistika ITS angkatan 2012 khususnya M. Rizky Fauzi yang dengan sabar memotivasi dan memberikan dukungan besar dalam pengerjaan Tugas Akhir penulis, 2013, dan 2014 khususnya Fazlur Rahman yang telah banyak memberikan masukan dan motivasi dalam proses penyelesaian Tugas Akhir.
10. Teman-teman Statistika ITS angkatan 2015 Vivacious yang telah memberikan bantuan dan dukungan kepada penulis selama penyelesaian laporan Tugas Akhir.
11. Lianna Dwi Rahmawati yang penyabar dan pengertian dalam memotivasi penulis untuk menyelesaikan Tugas Akhir.
12. Rizal Aditya, Yolan Setyo Utomo, Habib Jazuli, Barep Adji Widhi Pangestu, Alfattan Kurniawan, semua teman, relasi, dan berbagai pihak yang telah membantu penulis dalam penyelesaian laporan ini.

Penulis berharap hasil Tugas Akhir ini dapat memberikan sumbangan yang bermanfaat bagi semua pihak. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun dari semua pihak.

Surabaya, 24 Mei 2018

Penulis

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Halaman

(*Halaman ini sengaja dikosongkan)*

DAFTAR LAMPIRAN

Halaman

(*Halaman ini sengaja dikosongkan)*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi telah membawa perubahan dalam bidang kehidupan manusia. Hal ini tidak bisa dihindari, karena kemajuan teknologi akan berjalan sesuai dengan kemajuan ilmu pengetahuan. Setiap inovasi teknologi yang diciptakan untuk memberikan manfaat positif bagi kehidupan manusia seperti kemudahan melakukan pekerjaan dan cara yang baru dalam melakukan aktivitas seperti cara komunikasi (Ngafifi, 2014). Cara komunikasi manusia saat ini mengalami banyak perubahan diakibatkan perkembangan teknologi dari yang awalnya hanya sebuah lembar pengumuman hingga menjadi berita *daring*, dari berupa fotografi berubah menjadi video bergerak di televisi. Hingga akhirnya pada era digital kini semua media informasi dapat diakses melalui internet. Internet menyediakan informasi berupa data digital yang mudah diakses dari berbagai teknologi informasi dan komunikasi secara cerdas, cepat, praktis dan terintegrasi. Salah satu contoh adalah film yang awal perkembangannya diakses melalui televisi, kini banyak situs daring yang memberikan layanan tontonan daring menggunakan jaringan internet sehingga dapat diakses kapanpun dan dimanapun (Dwicahya, 2018).

Sejak kemunculannya, film sudah menjadi salah satu media hiburan yang populer di kalangan masyarakat. Sejak tahun 1874 hingga 2015, tercatat 3.361.741 judul film yang telah dikeluarkan industri perfilman yang dapat diakses secara daring. Jumlah yang banyak ini memanjakan penikmat film dalam menonton, akan tetapi terkadang dapat menimbulkan kesulitan bagi penikmat film untuk mendapatkan film yang sesuai dengan keinginan jika tanpa adanya bantuan (Halim dkk, 2017). Disamping itu jika diperoleh rekomendasi yang sesuai bagi setiap *user* yang mengunjungi situs film daring maka akan memudahkan pemilik situs untuk menampilkan rekomendasi film yang sesuai saja dengan ketertarikan *user* sehingga akan menghemat tempat di situs miliknya. Oleh karena itu, perlu adanya suatu sistem yang dapat mempermudah pencarian *item* (film) atau yang biasa disebut

dengan *search engine*. Cara kerjanya adalah mencari *item* yang dicari dalam tumpukan informasi lain. Namun seiring waktu muncul mesin rekomendasi sebagai pengembangan dari sistem pencarian *item*. Sistem rekomendasi adalah sistem yang bertanggungjawab atas mesin rekomendasi. Sistem ini mampu mengidentifikasi serta memberikan konten yang berpotensi besar untuk dipilih dan digunakan oleh pengguna berdasarkan proses penyaringan, pemilahan *item* dan informasi yang mengambil preferensi dari perilaku maupun riwayat pengguna (*user*) (Asanov, 2015).

Terdapat enam jenis teknik sistem rekomendasi namun yang populer pengembangannya adalah metode *Hybrid Filtering* karena menggunakan dua atau lebih teknik untuk mengatasi kekurangan pada masing-masing teknik sistem rekomendasi. Salah satu teknik *Hybrid Filtering* adalah gabungan *Demographic Filtering* dan *Collaborative Filtering*. Teknik ini memanfaatkan informasi profil *user* seperti jenis kelamin, pekerjaan usia dan sebagainya serta opini atau penilaian *user* lain berupa *rating* atau *feedback* lain yang ada untuk memprediksi *item* yang mungkin disukai/diminati oleh seorang *user* (Adomavicius, Gediminas, & Tuzhilin, 2005). Pendekatan *Collaborative Filtering* yang digunakan dalam penelitian ini adalah *User-Based Collaborative* dengan mengasumsikan bahwa untuk mencari sesuatu yang akan digemari oleh *user* A, maka harus mencari *user* lain dengan kecenderungan yang sama. Pengguna lain yang memiliki ketertarikan akan sesuatu yang sama dengan *user* A disebut *neighbors* (Arvid, Setyohadi, Budiyanto, & Ernawati, 2016).

Dalam rangka meningkatkan keakurasian hubungan antara *user* dengan kesukaan yang sama terhadap suatu item (film) digunakan algoritma *clustering* pendekatan *Demographic Filtering*. Metode klaster lain yang umum digunakan *Bayesian Clustering, K-Means* dan *Hierarchical Clustering* (Dapiah, Iriawan, & Fithriasari, 2018). Metode *clustering* yang digunakan dalam penelitian ini adalah metode *Single Linkage Clustering* dan *K-Modes*. Menurut Sharma dan Gaud (2015), metode *K-Modes* merupakan metode pengembangan dari *K-Means* yang mampu mengelompokkan data kategorikal dan menghasilkan klaster yang

lebih stabil dengan waktu komputasi yang lebih singkat jika dibandingkan dengan *K-means*. Sedangkan kelebihan dari metode *Single Linkage Clustering* adalah mampu mengatasi data pengamatan yang cukup besar dan pernyataan tersebut merujuk pada penelitian Stuetzle dan Nugent (2010). Oleh karena itu, dalam penelitian ini akan digunakan metode *Single Linkage Clustering* dan *K-Modes* yang diharapkan mampu meningkatkan kebaikan model klaster yang akan diperoleh, karena *K-Modes Clustering* digunakan untuk mengatasi permasalahan klaster data kategorik seperti data dalam penelitian ini.

Penelitian menggunakan teknik *Collaborative Filtering* dengan metode klaster partisional *K-Means Clustering* diperoleh hasil yang menyatakan bahwa algoritma tersebut merupakan yang terbaik pada kasus rekomendasi data film (Mahdavi & Dakhel, 2011). Penelitian yang serupa dengan tema perbandingan metode *user-based* dan *item-based collaborative filtering* pada data film diperoleh hasil prediksi *rating* dengan kedua metode tersebut tidak berbeda signifikan (Dwicahya, 2018). Serta penelitian pendukung lainnya dengan metode *clustering* pada data film menggunakan *Bisecting K-Means* menghasilkan bahwa metode *user-based* lebih rendah nilai tingkat kesalahannya dibandingkan *item-based* dalam prediksi *rating* (Halim, dkk., 2017). Sedangkan penggunaan teknik *Demographic Filtering* pada data film telah dilakukan Azizah (2016) pada Tugas Akhir Mahasiswa ITS Jurusan Sistem Informasi dengan menggunakan *K-Means Clustering* pada profil *user* film.

Hasil akhir penelitian yang diperoleh adalah suatu list rekomendasi film yang relevan bagi setiap *user* dengan teknik rekomendasi *Top N Recommendation* yang mengacu pada ketetanggaan (*neighborhood-based*) yang telah terbentuk antar *user* (*user-based*). *Top N Recommendation* memilih sejumlah rekomendasi yang paling revelan dengan kualitas terbaik. Teknik ini dapat dilakukan dengan menggunakan *Ranked-based Technique*. *RBT* yang mana memberikan sejumlah N rekomendasi judul film yang relevan dengan *user* terkait dengan menggunakan *Ranked-Based*. Metode ini bekerja dengan cara pemeringkatan *rating movie* terbaik yang diberikan oleh *user* secara menurun dan memberikan sejumlah n peringkat *rating* tertinggi yang diberikan

pada sasaran *user*. Hasil penelitian ini diharapkan mampu mengaplikasikan metode statistik dalam hal pengelompokkan data sehingga mampu memberikan rekomendasi bagi *user* dalam memilih film yang akan ditonton (Hapsari, Wibowo, & Baizal, 2015).

1.2 Rumusan Masalah

Perkembangan sistem rekomendasi di dunia film daring menjadi suatu penelitian yang masih memerlukan banyak penyempurnaan. Suatu ide yang menggunakan data-data *user* dan *user* lain di masa lalu serta penilaian-penilaian terhadap suatu film untuk memprediksi suatu film lain yang cocok dengan selera *user* akan membantu memberikan kemudahan bagi *user* dalam mengambil keputusan yang baik. Kompleksitas selera *user* dapat didekati dengan suatu metode pengelompokkan profil demografi dengan harapan sekelompok *user* dengan latar belakang yang sama akan memperbesar peluang kesamaan selera film daring yang ditonton. Sehingga penting adanya pengelompokkan profil demografis dalam pertimbangan pengelompokkan *user* film daring, metode yang dapat diterapkan dalam pengelompokkan ini adalah *Single Linkage Clustering* dan *K-Modes Clustering*. Tak kalah pentingnya tanggapan *user* terhadap film daring yang pernah ditonton akan mempengaruhi kualitas film salah satunya adalah *rating*. Pada umumnya film dengan *rating* yang tinggi akan sangat direkomendasikan untuk ditonton bagi *user* lain. Metode yang dapat diterapkan pada data tersebut salah satunya adalah *Hybrid Filtering*. Berdasarkan uraian tersebut maka metode yang digunakan dalam penelitian ini adalah *Single Linkage Clustering* dan *K-Modes Clustering* dengan pendekatan *Hybrid Filtering*.

1.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Diperoleh karakteristik *users* dalam data *movielens* berdasarkan usia, jenis kelamin, pekerjaan dan *rating* yang diberikan.

2. Didapatkan hasil implementasi metode *Single Linkage Clustering* dengan *K-Modes* dalam *Hybrid Filtering* pada data *movielens.*
3. Dihasilkan rekomendasi bagi *users* berdasarkan sistem rekomendasi.

1.4 Manfaat

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Bagi Pihak Terkait

Manfaat yang diperoleh penonton film (*user*) dalam penelitian ini yaitu mendapatkan kemudahan dalam pengambilan keputusan dengan beberapa referensi film yang relevan kepada *user* dengan pandangan historis sendiri maupun kesamaan minat film *user* lain. Bagi pelapak film daring memperoleh manfaat berupa daftar rekomendasi yang relevan diberikan kepada *user* atau penonton film daring sehingga keterbatasan menampilkan secara visual daftar film-film kesukaan tiap *user* dapat teratasi.

2. Bagi Peneliti

Manfaat yang diperoleh peneliti dalam penelitian ini adalah mampu memberikan solusi pengambilan keputusan berdasarkan pendekatan statistik pada studi kasus sistem rekomendasi film kepada *users*.

1.5 Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini yaitu data yang digunakan adalah *movielens dataset* bersifat *opensource* yang diunduh secara daring terdiri dari 100.836 *rating* yang berskala rasio (0,5-5) yang telah diberikan oleh 610 *user* terhadap 9.742 film pada laman (http://movielens.umn.edu) pada 20 April 2019. Metode *clustering* yang digunakan adalah *Single Linkage Clustering* dan *K-Modes* dengan menggunakan pendekatan sistem rekomendasi *Hybrid Filtering*.

(Halaman ini sengaja dikosongkan)

BAB II
TINJAUAN PUSTAKA

2.1 Statistika Deskriptif

Metode statistik adalah prosedur-prosedur yang digunakan dalam pengumpulan, penyajian, analisis, dan penafsiran data. Metode-metode tersebut terbagi menjadi dua yaitu statistika deskriptif dan statistika inferensia. Statistika deskriptif adalah metode yang berkaitan dengan pengumpulan dan penyajian data sehingga memberikan informasi hanya mengenai data dan tidak menarik kesimpulan (Walpole, 1995).

2.2 *Demographic Filtering*

Jenis sistem ini menyarankan item tergantung pada profil demografis pengguna. Anggapannya adalah bahwa rekomendasi yang berbeda harus dibuat untuk catatan demografis yang berbeda. Banyak situs *web* menggabungkan solusi kustomisasi yang sederhana dan efektif tergantung pada demografis. Misalnya, pelanggan dialihkan ke situs *web* tertentu tergantung pada bahasa atau negara mereka. Atau rekomendasi dapat dipersonalisasi sesuai dengan usia *user*. Sementara metode ini telah cukup populer dalam aspek pemasaran, namun relatif sedikit penelitian sistem rekomendasi yang relevan dalam hal sistem rekomendasi demografis (Nilashi, dkk., 2017).

2.3 *Collaborative Filtering*

Cara kerja algoritma sistem rekomendasi *Collaborative Filtering* (*CF*) adalah memanfaatkan opini *user* lain yang ada untuk memprediksi item yang mungkin akan disukai atau diminati oleh seorang *user* (Ricci dkk, 2011). Tujuan utama algoritma ini adalah merekomendasikan produk atau memperkirakan kegunaan produk tertentu untuk *user* tertentu berdasarkan kesukaan *user* di masa lalu dan pandangan dari pengguna lain yang berpikiran sama. Ada dua jenis hasil yang diberikan metode ini, yaitu prediksi *rating* dan tugas rekomendasi (Nilashi, 2012). Pada sebuah skenario *CF*, terdapat sebuah daftar *user* **u** $\{u_1,u_2, \ldots, u_m\}$ dan daftar *movie* $m\{m_1,m_2, \ldots, m_n\}$, dan setiap *user* u_i memiliki daftar *movie* m_{ui} yang mana telah dinilai oleh *user* seperti berupa *rating*.

2.3.1 *User-Based Collaborative Filtering*

User-Based Collaborative Filtering adalah menemukan sekumpulan *user neighbour* memiliki historis kesukaan yang sama dengan *user* yang akan dijadikan sasaran rekomendasi. Setelah sekumpulan tetangga terbentuk, sistem menggabungkan kesukaan tetangga (*neighbour*) untuk menghasilkan rekomendasi kepada *user* (Sarwar dkk, 2001). Perhitungan kemiripan antar *user* dan *item* digunakan persamaan *Adjusted Cosine* sebagai berikut.

$$S_{(u,u')} = \frac{\sum_{i=1}^{I}(R_{u,i})(R_{u',i})}{\sqrt{\sum_{i=1}^{I}(R_{u,i})^2}\sqrt{\sum_{i=1}^{I}(R_{u',i})^2}}, i \in I \qquad (2.1)$$

Keterangan:

$S_{(u,u')}$ = Nilai kemiripan antara *user u dan user u'*
$i \in I$ = Himpunan *movie* yang mirip *movie i*
$R_{u,i}$ = *Rating user u pada movie i*
$R_{u',i}$ = *Rating user u' pada movie i*

2.3.2 Prediksi *Rating*

Cara menghitung nilai prediksi untuk *user* atau *item* baru digunakan persamaan *Weighted Sum* (Sarwar dkk, 2001).

$$P_{(u,i)} = \frac{\sum_{i=1}^{I}(R_{(u',i)} \cdot S_{(u,u')})}{\sum_{i=1}^{I} S_{(u,u')}}, i \in I \qquad (2.2)$$

Keterangan :

$P_{(u,i)}$ = Prediksi *rating movie i* pada *user u*
$S_{(u,u')}$ = Nilai kemiripan antara *user u* dan *user u'*
$i \in I$ = Himpunan *movie* yang mirip *movie i*
$R_{(u',i)}$ = Rata-rata *rating user u'* (selain *u*) pada *movie i*

2.4 *Single Linkage Clustering*

Metode *Single Linkage Clustering* merupakan teknik pengelompokan yang bekerja berdasarkan prinsip algoritma *Agglomerative Hierarchical Clustering*. Prinsip kerja dari pengelompokan *Agglomerative Hierarchical Clustering* dilakukan secara bertahap. Pada setiap iterasi dari pengelompokan *Hierarchical Clustering* hanya ada satu pemilihan penggabungan suatu item terhadap item lainnya (Handoyo, Rumani, & Nasution, 2014). Input untuk algoritma *Single Linkage* bisa berwujud jarak atau persamaan antara pasangan-pasangan dari objek. Klaster yang akan terbentuk berasal dari penggabungan jarak terpendek atau *similarities* (kemiripan) yang paling besar.

Algoritma dalam *clustering* dengan menggunakan metode *Single Linkage Clustering* adalah sebagai berikut (Handoyo, Rumani, & Nasution, 2014).

1. Menentukan jumlah klaster yang ingin dibentuk.
2. Menghitung matriks jarak untuk pasangan klaster yang terdekat. Untuk metode *Single Linkage Clustering*, penentuan jarak terdekat dapat dihitung dengan formula berikut.

$$d_{uv} = min\{d_{uv}\}, d_{uv} \in D_\epsilon \qquad (2.3)$$

dengan D_ϵ merupakan *Euclidian Distance* dengan formula berikut.

$$D_\epsilon = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \qquad (2.4)$$

Dimana,

D_ϵ = *Euclidean Distance*
i = Banyaknya objek
(x,y) = Koordinat objek
(s,t) = Koordinat *centroid*

3. Menggabungkan kedua klaster yang memiliki jarak terdekat kemudian menghapus baris dan kolom matriks jarak yang bersesuaian dengan kedua klaster. Lalu tambahkan baris dan kolom yang memberikan jarak-jarak antara kedua klaster yang telah digabung dengan klaster-klaster yang tersisa.
4. Mengulangi langkah 2 dan 3 hingga hanya tersisa satu klaster.

2.5 *K-Modes Clustering*

K-Modes merupakan pengembangan dari metode K-Means agar dapat di gunakan untuk klasterisasi data kategorikal. K-Modes menggunakan sebuah ukuran jarak *(dissimilarity)* berupa kecocokan suatu nilai atribut tiap dimensi terhadap titik pusat klaster, menggantikan mean dengan modus, dan menggunakan metode berbasis frekuensi untuk memutakhirkan modus dalam proses meminimalkan jarak (*dissimilarity*) dari seluruh data ke pusat klaster masing-masing serta dapat menghasilkan klaster yang lebih stabil dengan waktu komputasi yang lebih singkat jika dibandingkan dengan *K-means*.

Algoritma yang digunakan dalam melakukan analisis klaster dengan metode *K-Modes* adalah sebagai berikut (Nduru, Buulolo, & Pristiwanto, 2018).

1. Pilih k data sebagai inisialisasi *centroid* (modus), satu untuk setiap klaster.
2. Hitung jarak antara masing-masing objek dan mode klaster, tetapkan objek ke klaster yang pusatnya memiliki jarak terdekat ke objek ulangi langkah ini sampai semua objek ditetapkan ke kelompok.

$$d(x, y) = \sum_{j=1}^{r} \in (x_j, y_j) \qquad (2.5)$$

Dimana :

D*(x,y)* = jarak data x ke y

x_j = nilai fitur ke-j dari x

y_j = nilai fitur ke-j dari y

R adalah jumlah fitur dan berikut adalah nilai pencocokan seperti pada persamaan berikut :

$$\in (x_j, y_j) = \begin{cases} 0, & x_j = y_j \\ 1, & x_j \neq y_j \end{cases} \qquad (2.6)$$

3. Perbarui modus (sebagai *centroid*) dari setiap klaster dengan nilai kategoris yang sering muncul pada setiap klaster.
4. Ulangi langkah 2 dan 3 untuk memenuhi syarat, yaitu data klaster tersebut tidak bergerak atau posisi pusat *centroid* tidak berubah.

2.6 *Silhoutte Coefficient*

Silhouette coefficient digunakan untuk mengukur persamaan yang terjadi pada suatu objek dengan membandingkan antara klasternya tertentu dengan klaster lainnya. Hal yang perlu diketahui ketika menentukan nilai *silhouette* adalah hasil dari partisi atau *clustering result* dan pengelompokan semua kedekatan antar objek (Jannah, Fithriasari, & Usagawa, 2018). Langkah-langkah yang dilakukan dalam menghitung nilai *silhouette* adalah sebagai berikut (Han, Kamber, & Pei, 2012).

1. Menghitung rata-rata jarak dari suatu *user* misalkan *i* dengan semua *user* lain yang berada dalam satu klaster C_i.

$$a(i) = \frac{1}{|C_i|-1} \sum_{j \in C_{i,i \neq j}} d(i,j) \quad (2.6)$$

Dimana,

j = *user* lain dalam satu klaster C_i

d(i,j) = Jarak antara *user i* dengan *user j*

2. Menghitung rata-rata jarak dari *user i* tersebut dengan semua *user* di klaster lain kemudian diambil nilai terkecilnya.

$$d(i) = \frac{1}{|C_k|} \sum_{k \in C_k} d(i,k) \quad (2.7)$$

Dimana,

d(i) = Jarak rata-rata *user i* dengan semua *user* lain pada klaster C_k

$$b(i) = \min C_{k,k \neq i}\{d(i)\} \quad (2.8)$$

Dimana,

b(i) = Nilai terkecil dari jarak rata-rata *user i* dengan semua *user* pada klaster lain C_k

3. Diperoleh nilai *silhouette* sesuai dengan formula:

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \quad (2.9)$$

2.7 *Hybrid Filtering*

Sistem rekomendasi menggunakan Hybrid Filtering didapatkan dari kombinasi dua atau lebih teknik rekomendasi lainnya dengan tujuan mencoba untuk mengatasi kekurangan masing-masing teknik sistem rekomendasi. Secara lebih lanjut, kombinasi dari pengembangan pendekatan sistem rekomendasi *hybrid* bergantung pada karakteristik data yang digunakan.

Terdapat tujuh kategori sistem rekomendasi *hybrid* terdiri atas *weighted*, *switching*, *mixed*, *feature combination*, *feature augmentation*, *cascade* and *meta-level* telah diperkenalkan (Burke, 2007).

2.8 *Top N Recommendation*

Top N Recommendation adalah teknik yang dilakukan bertujuan memberikan sekumpulan N *item* yang relevan untuk disarakan pada *user* tertentu. Rekomendasi dapat dilakukan berdasarkan item yang paling banyak dipilih atau disukai, dapat pula berdasarkan peringkat kualitas suatu *item*. Sebagai contoh pada kualitas film, dilakukan pemeringkatan nilai *rating* tertinggi hingga terendah dan sejumlah N peringkat tertentu diberikan pada suatu *user* yang disebut *Ranking-Based Technique* (Zolakhaf, Babanezhad, & Pottinger, 2018).

BAB III
METODOLOGI PENELITIAN

3.1 Sumber Data

Penelitian ini akan menggunakan sumber data sekunder yang diperoleh dari situs *web* GroupLens Research Project yakni *movielens dataset* yang terdiri atas 100.836 *rating* yang berskala rasio (0,5-5) yang telah diberikan oleh 610 *user* terhadap 9.742 film *Box Office*. Terdapat catatan bahwa setiap *user* sedikitnya telah memberikan *rating* terhadap 20 film. Terdapat informasi demografi sederhana setiap *user* yang berisikan keterangan usia, jenis kelamin, pekerjaan dan kode pos. *Movielens* dataset diunduh melalui laman (http://*movielens*.umn.edu). Dataset berekstensi file zip dengan nama (ml-latest-small.zip).

3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini disajikan pada **Tabel 3.1**.

Tabel 3. 1 Variabel Penelitian

Var.	Definisi	Keterangan	Skala
X_1	*User id*	Berisikan tentang *user id* yang merupakan penonton film daring	Nominal
X_2	Usia	Berisi data usia setiap *user* (tahun)	Rasio
X_3	Jenis Kelamin	0: “F” (perempuan) 1 : “M” (laki-laki)	Nominal
X_4	Pekerjaan	Terdiri dari 21 jenis pekerjaan	Nominal

Tabel 3. 1 Variabel Penelitian (Lanjutan)

Var.	Definisi	Keterangan	Skala
X_5	*Movie Title*	Judul film yang telah tercatat dalam *website* milik *GroupLens Research*	Nominal
X_6	*Rating*	Penilaian *user* film kepada suatu film (0,5 - 5)	Rasio

3.3 Struktur Data

Struktur data *movielens* yang diambil berasal dari GroupLens Research Project akan digunakan pada penelitian disusun dan disajikan pada **Tabel 3.2**.

Tabel 3. 2 Struktur Data

No	X_1	X_2	X_3	X_4	X_5	X_6
1	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$	$X_{4,1}$	$X_{5,1}$	$X_{6,1}$
2	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	$X_{4,2}$	$X_{5,2}$	$X_{6,2}$
...	...	...	...	...	...	...
100.836	$X_{1,100.836}$	$X_{2,100.836}$	$X_{3,100.836}$	$X_{4,100.836}$	$X_{5,100.836}$	$X_{6,100.836}$

Data pada **Tabel 3.2** akan digunakan dalam proses analisis klaster menggunakan metode *Single Linkage Clustering* dan *K-Modes Clustering* dengan melibatkan variabel X_1 hingga X_4 yaitu *user id,* usia, jenis kelamin dan pekerjaan. Setelah didapatkan klaster yang terbentuk maka selanjutnya akan dilakukan proses analisis rekomendasi dengan melibatkan variabel hasil klaster yakni *user id* (X_1), *movie title* (X_5) dan *rating* (X_6). Untuk mempermudah dalam proses analisis rekomendasi maka variabel *user id* akan disimbolkan dengan U, variabel *movie title* disimbolkan dengan M dan variabel *rating* disimbolkan dengan R. Susunan data tersebut selanjutnya disebut dengan matriks *rating* yang disajikan pada **Tabel 3.3**.

Tabel 3. 3 Matriks *Rating*

Klaster	***User Id***	***Movie Title***				
		M_1	M_2	M_3	...	$M_{9.742}$
K_1	U_1	$R_{1,1}$	$R_{1,2}$	$R_{1,3}$	...	$R_{1,\ 9.742}$
	U_3	$R_{3,1}$	$R_{3,2}$	$R_{3,3}$	...	$R_{3,\ 9.742}$
	⋮	⋮	⋮	⋮		⋮
	U_p	$R_{p,1}$	$R_{p,2}$	$R_{p,3}$	...	$R_{p,\ 9.742}$
⋮	⋮	⋮	⋮	⋮		⋮
K_2	U_4	$R_{4,1}$	$R_{4,2}$	$R_{4,3}$	...	$R_{4,\ 9.742}$
	U_6	$R_{6,1}$	$R_{6,2}$	$R_{6,3}$	...	$R_{6,\ 9.742}$
	⋮	⋮	⋮	⋮		⋮
	U_q	$R_{q,1}$	$R_{q,2}$	$R_{q,3}$	...	$R_{q,\ 9.742}$
⋮	⋮	⋮	⋮	⋮		⋮
K_n	U_8	$R_{8,1}$	$R_{8,2}$	$R_{8,3}$	...	$R_{8,\ 9.742}$
	U_{11}	$R_{11,1}$	$R_{11,2}$	$R_{11,3}$	...	$R_{11,\ 9.742}$
	⋮	⋮	⋮	⋮		⋮
	U_r	$R_{r,1}$	$R_{r,2}$	$R_{r,3}$	...	$R_{r,\ 9.742}$

Keterangan:

K_i = Jenis klaster yang terbentuk, i = 1, 2 ..., n

p,q,r = *user* ke- pada setiap klaster terbentuk, misal p = 3 , maka p adalah *user* ke 3 pada klaster terbentuk tersebut.

Struktur data hasil pengunduhan dari *website* disertakan agar dapat memahami data secara jelas. *File* unduhan berupa u.*user* adalah sebuah tabel yang berisikan keterangan profil demografis dari *user* film daring. Disajikan dalam **Tabel 3.4**.

Tabel 3. 4 File u.*user*

User id	Usia	Jenis Kelamin	Pekerjaan
U_1	24	M	*Technician*
U_2	53	F	*Other*

Tabel 3. 4 File u.*user* (Lanjutan)

User id	Usia	Jenis Kelamin	Pekerjaan
U_3	23	M	*Writer*
...	...	...	...
U_{610}	22	M	*Student*

Sedangkan susunan tabel yang berisikan penilaian *user* film daring terhadap masing – masing daftar film (*movieid*) disajikan pada **Tabel 3.5** berikut.

Tabel 3. 5 File u.*data*

User id	*Movie Title*	*Rating*
U_1	1	4
U_2	1	5
U_3	1	4
...	...	...
$U_{100.836}$	9.742	3

3.4 Langkah Analisis

Langkah analisis yang akan dilakukan pada penelitian ini yaitu sebagai berikut:

1. Mengumpulkan data *movielens* yang terdiri dari tiga *file*, sebagai berikut:
   a. u.data, berisikan *userid, movieid, rating,* dan *timestamp*
   b. u.*item*, berisikan informasi tentang film terdiri dari *movieid, movie title, genre,* dan *release date*
   c. u.*user*, berisikan infomasi demografi sederhana *users* berupa *age*, *gender*, *occupation* dan *timestamp*.
2. Melakukan *preprocessing* data dengan tujuan menghilangkan atau mentransformasi atribut yang tidak diperlukan dalam proses pengolahan data dan analisis. Proses ini dilakukan dengan urutan sebagai berikut :
   a. Pada *file* u.*data* dihapuskan *timestamp*. *Timestamp* merupakan *unique second* yang tidak ada hubungannya dengan proses rekomendasi.

b. Pada *file* u.*user* dihapuskan *timestamp*. *Timestamps* merupakan *unique code* yang tidak ada hubungannya dengan proses rekomendasi.
c. Pada *file* u.*item* dihapuskan *release date* yang merupakan keterangan waktu rilis film yang tidak ada hubungannya dengan proses rekomendasi.
d. Atribut umur dibedakan berdasarkan rentang umur sebagai berikut.
   1 : " < 12 tahun"
   2 : "12-15 tahun"
   3 : "16-21 tahun"
   4 : "22-40 tahun"
   5 : "41-60 tahun"
   6 : " > 60 tahun"
e. Atribut jenis kelamin diubah ke dalam bentuk angka biner 0 dan 1 karena data kategori dan berskala nominal menjadi sebagai berikut.
   1) 0 = "F", simbol untuk jenis kelamin perempuan
   2) 1 = "M" simbol untuk jenis kelamin laki-laki
f. Atribut pekerjaan dibedakan atas beberapa jenis, diantaranya.
   1 : "Manajer Perusahaan"
   2 : "Tenaga Professional"
   3 : "Teknisi dan Asisten Tenaga Professional"
   4 : "Tenaga Tata Usaha"
   5 : "*Others*"

3. Mengeksplorasi karakteristik *users* berdasarkan usia, jenis kelamin, pekerjaan dan *rating* dengan menggunakan visualisasi diagram.
4. Melakukan *clustering* data,
   a. Analisis klaster dengan menggunakan metode *Single Linkage Clustering*, yaitu menentukan jumlah klaster *userid* optimum pada data *movielens* dengan memilih jarak minimal antar klaster yang terbentuk sehingga diperoleh pusat klaster. Langkah kerja metode ini dijelaskan sebagai berikut.
      1) Menentukan jumlah klaster yang ingin dibentuk.

2) Menghitung matriks jarak untuk pasangan klaster yang terdekat.
3) Menggabungkan kedua klaster yang memiliki jarak terdekat kemudian menghapus baris dan kolom matriks jarak yang bersesuaian dengan kedua klaster. Lalu tambahkan baris dan kolom yang memberikan jarak-jarak antara kedua klaster yang telah digabung dengan klaster-klaster yang tersisa.
4) Mengulangi langkah 2 dan 3 hingga hanya tersisa satu klaster.

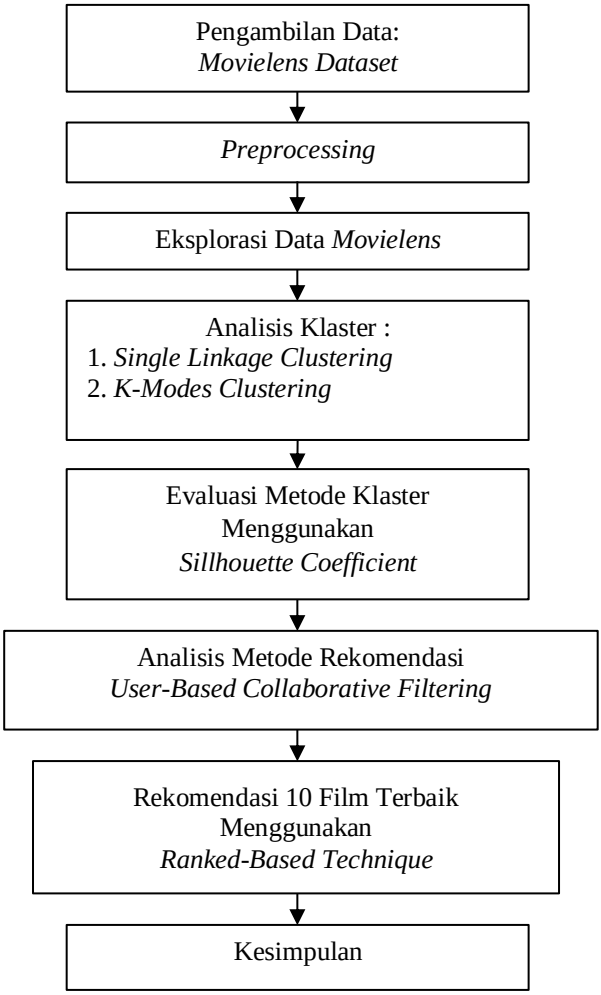
b. Analisis klaster metode *K-Modes* menggunakan pusat klaster secara random menjadi pusat klaster awal pada perhitungan metode *K-Modes*. Langkah analisis *K-Modes Clustering* dijelaskan sebagai berikut.
1) Menentukan jumlah klaster yang ingin dibentuk.
2) Menentukan nilai *centroid* secara acak
3) Alokasikan vektor ke cluster dengan modus terdekat.
4) Perbarui modus (sebagai *centroid*) dari setiap cluster dengan nilai kategori yang sering muncul pada setiap cluster.
5) Ulangi langkah 2 dan 3 untuk memenuhi syarat, yaitu data klaster tersebut tidak bergerak atau posisi pusat *centroid* tidak berubah.

5. Melakukan perbandingan hasil evaluasi kebaikan klaster *Single Linkage Clustering* dan *K-Modes* dengan menggunakan nilai *Sillhouete Coefficient* untuk mengetahui seberapa baik suatu obyek ditempatkan dalam suatu jumlah klaster oleh metode yang digunakan. Langkah evaluasi sebagai berikut.
1) Menghitung rata-rata jarak dari suatu item misalkan *i* dengan semua item lain yang berada dalam satu klaster.
2) Menghitung rata-rata jarak dari item *i* tersebut dengan semua item di klaster lain kemudian diambil nilai terkecilnya.
3) Mendapatkan nilai *Sillhoute* dengan perhitungan menggunakan rumus *Sillhoute Coefficient*.

6. Analisis metode rekomendasi *User-Based Collaborative Filtering* menggunakan persamaan *Adjusted Cosine* antar *user* dalam klaster. Langkah-langkah memprediksi dilakukan sebagai berikut.
    1) Pada setiap klaster *user* yang terbentuk, dihitung nilai korelasi antar *user* dengan tabel *matrix adjusted cosine*.
    2) Setelah diketahui nilai similarity antar *user*, maka dilakukan perhitungan prediksi nilai *rating* yang belum terisi oleh *user* akibat beberapa hal, misalnya belum pernah menonton film tersebut. Prediksi dilakukan dengan rumus *Weighted Sum* sehingga semua film akan memiliki semua penilaian dan dapat dilanjutkan pada proses rekomendasi menggunakan *Ranked-Based Technique*.
7. Merekomendasikan sekumpulan item relevan dalam *Top-N Recommendation* yang sesuai dengan *user preference* yakni kesamaan *preference* yang ada pada masing-masing klaster *users* terbentuk dengan pendekatan pemeringkatan atau *Ranked-Based Technique*. Banyaknya rekomendasi yang diberikan ditentukan terlebih dahulu.
8. Interpretasi dan menarik kesimpulan.

3.5 Diagram Alir Penelitian

Berikut merupakan diagram alir yang dilakukan pada penelitian.



Gambar 3. 1 Diagram Alir Penelitian

BAB IV
ANALISIS DAN PEMBAHASAN

Pada analisis dan pembahasan ini akan dibahas mengenai gambaran secara umum karakteristik data *user* film daring dan penilaian historis masing-masing *user* setelah menonton beberapa film.

4.1 *Preprocessing*

Data biografi *user* yang digunakan dalam penelitian terkandung beberapa variabel yakn*i age, occupation,* dan *gender* . Data *age* atau usia bertipe data numerik dan sangat beragam. Karakteristik data yang cukup beragam adalah *occupation* atau jenis pekerjaan yang terdiri dari 21 jenis pekerjaan. Kondisi data tersebut memberikan kesulitan dalam penggalian informasi agar bermanfaat dan proses analisis, sehingga diperlukan beberapa perlakuan agar data yang telah tersedia dapat diolah menjadi informasi yang bermanfaat. Hal-hal yang dapat dilakukan adalah terlebih dahulu data *user* film daring diubah ke dalam golongan *age* atau usia *user* berdasarkan periodisasi perkembangan psikologis manusia menurut Elizabeth B. Hurlock (1980), sehingga data *age* atau usia digolongkan ke dalam 6 kategori yakni :

Tabel 4. 1 Periodisasi Perkembangan Manusia

Golongan	Kategori	Keterangan Periode
1	< 12 tahun	Kanak-kanak
2	12-15 tahun	Pubertas
3	16-21 tahun	Remaja
4	22-40 tahun	Dewasa Awal
5	41-60 tahun	Dewasa Madya
6	> 60 tahun	Usia Lanjut

Tahapan kategorisasi dilakukan bertujuan untuk mempermudah dalam memahami data *age* atau usia dari *user* film daring sehingga dapat dilakukan proses eksplorasi dan analisis dengan metode yang sesuai dalam penelitian ini.

Tahapan selanjutnya adalah mengubah golongan pekerjaan dalam data *occupation* atau pekerjaan dari *user* yang awalnya

terdiri dari 21 jenis pekerjaan *user* dikategorikan menjadi 5 jenis golongan pekerjaan berdasarkan Buku Klasifikasi Baku Jenis Pekerjaan Indonesia (KBJI) (Statistik, 2002). Hasil pengubahan yang dilakukan dicantumkan ke dalam **Tabel 4.2** sebagai berikut,

Tabel 4. 2 Golongan Pekerjaan

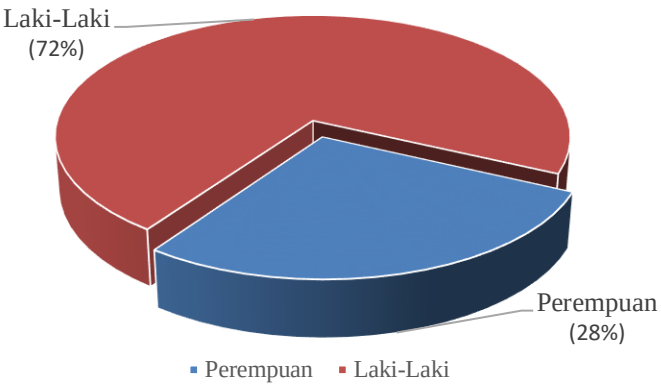
Golongan	*Occupation*	Panduan BPS
1	*Executive*	Manajer Perusahaan
2	*Doctor, Educator, Engineer, Healthcare, Lawyer, Programmer, Scientist, Writer*	Tenaga Profesional
3	*Artist, Entertainer,Marketing, Technician*	Teknisi dan Asisten Tenaga Profesional
4	*Administrator, Librarian*	Tenaga Tata Usaha
5	*Student, Salesman, Homemaker, Retired, Other, None,*	*Others*

Tahapan pengubahan jenis golongan pekerjaan akan lebih mempermudah proses eksplorasi data dan analisis menggunakan metode klaster dalam penelitian.

4.2 Karakteristik Data *User* Film Daring

Karakteristik data *user* film daring yang diperoleh secara daring di *web GroupLens Research* dapat digambarkan dengan menggunakan pendekatan analisis eksplorasi data. Analisis tersebut dilakukan berdasarkan informasi biografis dari *user*. Informasi tersebut berupa jenis kelamin, usia, dan pekerjaan *user*.

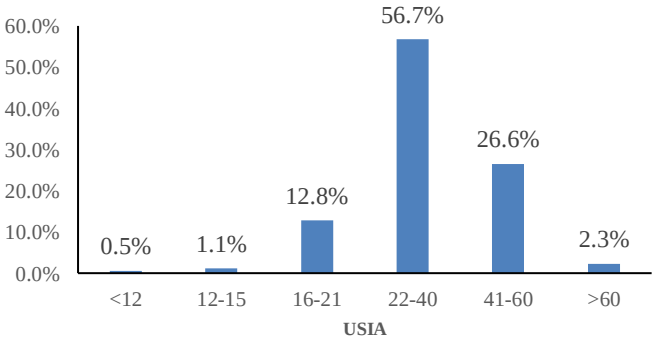
4.2.1 Karakteristik Jenis Kelamin



Gambar 4. 1 Jenis Kelamin User

Data yang tercatat menunjukkan bahwa laki-laki lebih dominan menyaksikan film daring dibandingkan perempuan. Terdapat 440 orang laki-laki menyaksikan film secara daring dan memberikan penilaian sedangkan perempuan hanya sejumlah 170 orang yang menonton film secara daring.

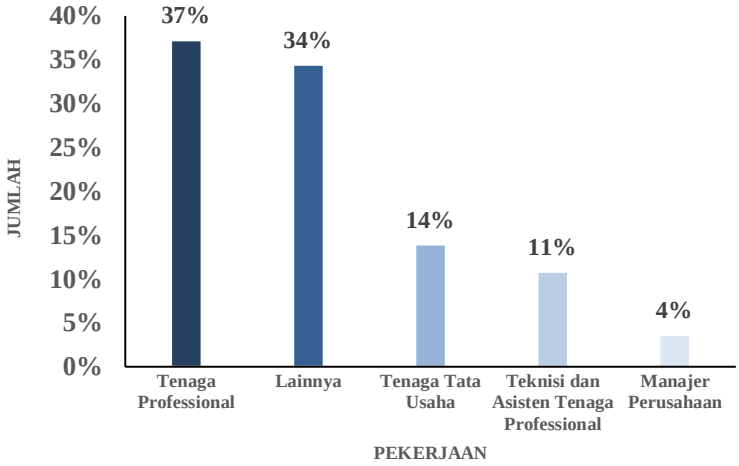
4.2.2 Karakteristik Usia



Gambar 4. 2 Usia *User*

Golongan *user* yang mengunjungi situs film daring didominasi penonton berusia 22 hingga 40 tahun yang tergolong dewasa awal dengan jumlah 346 orang. Adapun penonton yang usia kurang dari 12 tahun hanya terdapat 3 orang yang berarti penonton dari golongan kanak-kanak masih minim. Pada urutan kedua dan ketiga penonton terbanyak ditempati oleh usia berturut-turut 41 hingga 60 tahun (dewasa pertengahan) dan 16 hingga 21 tahun (remaja).

4.2.3 Karakteristik Pekerjaan



Gambar 4. 3 Pekerjaan User

Tinjauan dari segi jenis golongan pekerjaan *user* yang menonton film daring tenyata penonton terbanyak merupakan golongan tenaga profesional yang berjumlah 227 orang terdiri dari *doctor*, *educator*, *engineer*, *healthcare*, *lawyer*, *programmer*, *scientist*, dan *writer* sedangkan penonton yang paling sedikit berasal dari golongan pekerjaan manajer perusahaan yakni berasal dari *executive* hanya 22 penonton. Berada pada urutan kedua penonton terbanyak dengan selisih 17 orang dari urutan pertama, ditempati oleh jenis golongan pekerjaan lainnya yang terdiri dari *other*, *none*, *homemaker*, *retired*, *salesman* dan *student*.

Tabel 4. 3 Pekerjaan *User* Berdasarkan Jenis Kelamin

Pekerjaan / Jenis Kelamin	Manajer perusahaan	Tenaga profesional	Teknisi dan asisten	Tenaga tata usaha	*Others*	Total
Perempuan	2	43	15	44	66	170
Laki-laki	20	184	51	41	114	440
Total	22	227	66	85	210	610

Telah diketahui bahwa proporsi laki-laki lebih banyak daripada penonton perempuan. Akan tetapi jika ditinjau dari segi pekerjaan berdasarkan jenis kelamin, ternyata terdapat jenis pekerjaan yang lebih banyak perempuan daripada laki-laki yakni pada jenis golongan pekerjaan tenaga tata usaha dengan 44 orang atau selisih 3 orang dengan laki-laki dengan golongan pekerjaan serupa. Fakta lainnya adalah diantara 440 penonton laki-laki terdapat 20 orang yang ternyata menjabat sebagai manajer suatu perusahaan. Sedangkan hanya terdapat 2 perempuan yang menjabat sebagai manajer perusahaan diantara 170 *user* perempuan lainnya. Berdasarkan jumlahnya penonton film daring dari golongan perempuan adalah perempuan yang golongan pekerjaannya *others* yang meliputi pekerjaan *Student, Salesman, Homemaker, Retired, Other, dan None* sebanyak 66 orang. Sedangkan laki-laki yang merupakan *user* terbanyak menonton film daring sebanyak 227 orang bergolongan pekerjaan tenaga profesional yang didalamnya meliputi pekerjaan seperti *Doctor, Educator, Engineer, Healthcare, Lawyer, Programmer, Scientist* dan *Writer*.

Tabel 4. 4 Pekerjaan *User* Berdasarkan Usia

Pekerjaan / Usia	Manajer perusahaan	Tenaga profesional	Teknisi dan asisten	Tenaga tata usaha	*Others*	Total
Kanak-kanak	0	0	0	0	3	3
Pubertas	0	0	1	0	6	7
Remaja	0	6	6	1	65	78

Tabel 4. 4 Pekerjaan *User* Berdasarkan Usia (Lanjutan)

Pekerjaan / Usia	Manajer perusahaan	Tenaga profesional	Teknisi dan asisten	Tenaga tata usaha	*Others*	Total
Dewasa awal	12	135	44	52	103	346
Dewasa madya	9	83	15	30	25	162
Usia lanjut	1	3	0	2	8	14
Total	22	227	66	85	210	610

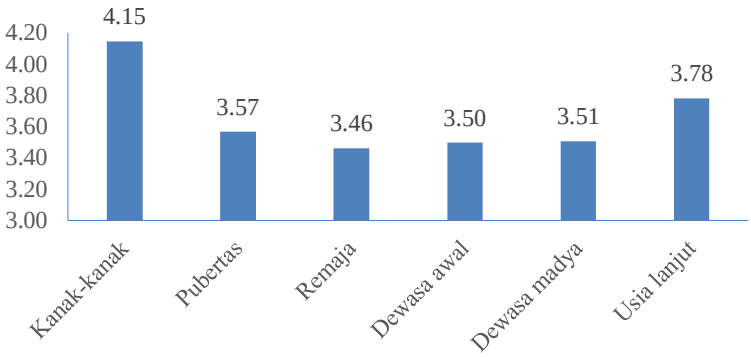
Tabel kontingensi golongan pekerjaan terhadap golongan usia *user* film daring memberikan informasi bahwa persebaran golongan pekerjaan pada *user* dengan golongan usia kanak-kanak dan pubertas merupakan yang terkecil. Hal ini dapat diketahui dari jumlah *user* golongan usia kanak-kanak hanya tediri dari 3 orang dengan pekerjaan *others* dan golongan usia pubertas ada 1 orang dengan pekerjaan teknisi dan asisten tenaga profesional serta 6 orang dengan golongan pekerjaan *others*. Sedangkan golongan usia remaja hingga usia lanjut memiliki keberagaman golongan pekerjaan *user* lebih besar. Paling dominan dengan jumlah *user* terbesar yakni 346 orang adalah golongan usia dewasa awal meliputi 135 *user* tenaga profesional, 103 *user* others, 52 *user* tenaga tata usaha, 44 *user* teknisi dan asisten tenaga profesional, dan 12 *user* manajer perusahaan.

4.2.4 Karakteristik *Rating User*

Penilaian historis dari suatu *user* film daring biasa disebut dengan *rating* yang menunjukkan tingkat kepuasan dalam menikmati film yang ditonton. Hal ini berpengaruh terhadap kualitas film. Subbab ini akan memberikan gambaran dari data *rating* yang diberikan oleh *user* film daring.

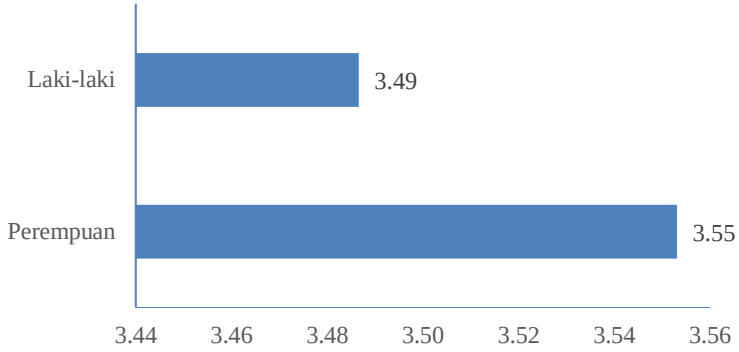
Rata-rata *rating* yang diberikan *user* menjadi salah satu informasi yang penting untuk mengetahui nilai atau kualitas film yang pernah ditontonnya. Sedangkan golongan usia *user* / penonton yang melihat suatu film memiliki ketertarikan tersendiri dalam pemilihan film yang disukai. Apabila dua komponen ini dipadukan maka dapat memberikan informasi terkait

kecenderungan pemberian *rating* oleh *user* / penonton dalam golongan usia tertentu terhadap film yang pernah ditonton. Informasi tersebut dapat diketahui melalui **Gambar 4.4** berikut.



Gambar 4. 4 Rata-rata *Rating* Berdasarkan Usia

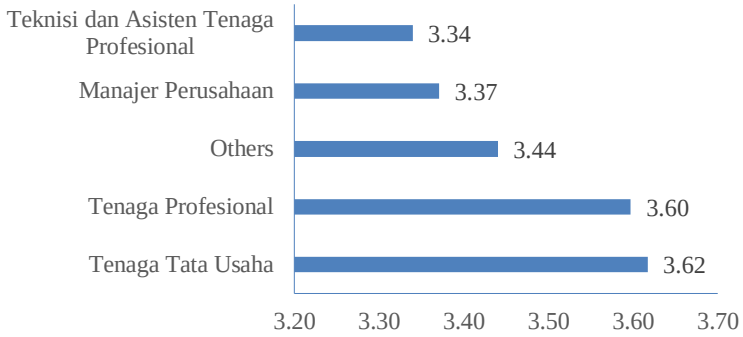
Nilai *rating user* terhadap film apabila ditinjau dari golongan usianya maka dapat diketahui informasi bahwa ternyata film-film daring disukai oleh berbagai golongan usia, dapat terlihat dari nilai *rating* yang diberikan oleh semua golongan usia lebih dari 3 bahkan mendekati 4. Hal tersebut mengindikasikan bahwa penonton merasa puas dalam menonton film sesuai dengan golongan usianya. Nilai *rating* tertinggi diberikan oleh golongan usia kanak-kanak yakni 4,15.



Gambar 4. 5 Rata-rata *Rating* Berdasarkan Jenis Kelamin

Pada **Gambar 4.5**, nilai apresiasi yang diberikan oleh *user* berdasarkan jenis kelamin mencapai nilai lebih dari 3 yang berarti

bahwa baik perempuan maupun laki-laki yang menikmati film di situs daring merasa puas. Perempuan memberikan apresiasi *rating* pada film yang telah ditonton lebih tinggi dari nilai *rating* yang diberikan oleh laki-laki yakni 3,55.



Gambar 4. 6 Rata-rata *Rating* Berdasarkan Pekerjaan

Setelah diketahui pada **Gambar 4.6** bahwa baik perempuan maupun laki-laki memberikan penilaian yang baik terhadap film yang telah ditonton dan juga berdasarkan golongan usia pun memberikan asil yang serupa, maka tidak jauh beda dengan hasil tersebut. Berdasarkan golongan pekerjaannya pun *user* memberikan nilai yang cukup baik yakni lebih dari 3 yang menunjukkan kepuasan dalam menonton film. Golongan pekerjaan yang memberikan rata-rata *rating* tertinggi adalah *user* golongan tenaga tata usaha sebesar 3,62.

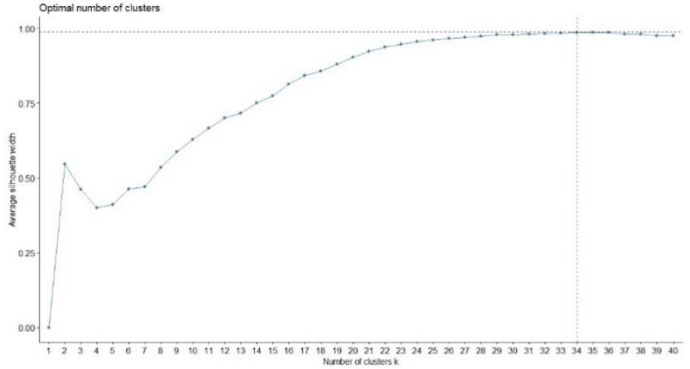
4.3 Sistem Rekomendasi

Sistem rekomendasi *hybrid* bekerja melalui beberapa langkah, yakni analisis klaster terbaik dalam menggelompokkan *user*, yang dicapai dengan menggunakan hasil perbandingan antara metode *Single Linkage Clustering* dengan *K-Modes Clustering*. Hasil klaster yang terbentuk akan diolah *menggunakan User-Based Collaborative Filtering* (*UBCF*). Langkah pertama dari *UBCF* adalah penyusunan matriks *rating* pada masing-masing klaster, selanjutnya dilakukan prediksi nilai *rating* yang belum terdapat nilainya dengan menggunakan *weighted sum* pada masing-masing klaster dan kemudian langkah terakhir adalah menggunakan

metode *ranked-based technique* untuk mendapatkan 10 rekomendasi film terbaik yang relevan bagi *user*. Langkah yang pertama adalah melakukan analisis klaster terhadap data demografik *user* dengan *Single Linkage Clustering*.

4.3.1 *Single Linkage Clustering*

Metode analisis klaster yang pertama digunakan dalam melakukan analisis klaster *Single Linkage Clustering* pada data biografi diri *user* film yang terdiri dari tiga variabel yaitu jenis kelamin, usia dan pekerjaan. Metode ini merupakan salah satu dari teknik klasterisasi secara *Hierarchical*. Langkah awal yang dilakukan adalah melakukan inisiasi jumlah klaster yang akan digunakan di dalam analisis. Pendekatan yang dilakukan dapat secara visual dengan pendekatan nilai *Silhouette Coefficient*. Penentuan jumlah klaster optimal yang dilakukan berdasarkan nilai *Silhouette* adalah melihat nilai tertinggi dari grafik yang terbentuk. Nilai tertinggi tersebut menggambarkan nilai *Silhouette* yang dimiliki oleh suatu jumlah klaster merupakan nilai terbaik dibandingkan dengan jumlah klaster lainnya. *Range* nilai *Silhouette* adalah berkisar antara -1 hingga 1 dan diperoleh pola yang konvergen. Sesuai dengan **Gambar 4.7**, jumlah klaster optimal yang terbentuk berdasarkan nilai *Silhouette* adalah sebanyak 34 klaster. Terlihat dari petunjuk garis putus-putus yang diberikan. Nilai *Silhouette* dengan jumlah klaster 34 adalah sebesar 0,985. Hasil tersebut dapat digunakan untuk membentuk klaster dengan menggunakan metode *Single Linkage Clustering*.



Gambar 4. 7 *Single Linkage Sillhouette Coefficient*

Setelah diperoleh inisiasi k klaster, maka dilakukan penyusunan matriks jarak berdasarkan data demografik *user* dengan menggunakan rumus perhitungan jarak *euclidean*. Pada **Tabel 4.5** dicantumkan cuplikan data demografik *user*.

Tabel 4. 5 Data u.*user*

userId	Jenis Kelamin (j)	Pekerjaan (p)	Usia (u)
1	1	3	4
2	0	5	5
3	1	2	4
4	1	3	4
	...		

Perhitungan jarak *euclidean* mampu memberikan nilai kedekatan antara 2 *user* yang berbeda. Sebagai contoh, perhitungan jarak *euclidean* antara *user* 1 terhadap *user* 2 dengan rumus perhitungan seperti pada **Persamaan 2.4**,

$$D_{1,2} = \sqrt{(1-0)^2 + (3-5)^2 + (4-5)^2}$$
$$= \sqrt{6} = \mathbf{2,44}$$

Setelah memahami cara kerja jarak *euclidean* maka dilakukan penyusunan matriks jarak *euclidean* antara semua *user* agar lebih mempermudah analisis selanjutnya, matriks tersebut dicantumkan pada **Tabel 4.6** berikut.

Tabel 4. 6 Cuplikan Matriks *Euclidean*

userId \ *userId*	1	2	3	4
1	0			
2	**2,4**	0		
3	1	3,3	0	
4	0	2,4	1	0
...		...		

Langkah selanjutnya adalah menggabungkan dua *user* berbeda yang memiliki nilai jarak *euclidean* terkecil. **Tabel 4.4** yang bertanda merah menunjukkan bahwa nilai *euclidean* terkecil adalah nol yang artinya nilai kedekatan antara *user* 1 dengan *user* 4 bernilai nol. Sehingga *user* 1 dengan *user* 4 menjadi *user* (1,4)

dan nilai jarak *user* lainnya terhadap *user* (1,4) memperhatikan aturan jarak minimum sehingga iterasi selanjutnya membentuk **Tabel 4.7** berikut.

Tabel 4. 7 Matriks *Euclidean* Iterasi 2

userId / *userId*	1	2	3	4
1,4	0			
2	2,4	0		
3	1	3,3	0	
...		...		

Penggabungan selanjutnya berdasarkan nilai jarak *euclidean* terkecil adalah *user* 3 ke dalam *user* (1,4) sehingga menjadi *user* (1,3,4) dengan nilai satu. Proses ini dilakukan berulang-ulang / iterasi hingga hanya tersisa satu klaster. Setelah tersisa satu klaster maka telah selesai langkah analisis klaster *Single Linkage Clustering*. Hasil akhir yang diperoleh adalah terbentuk 34 klaster, dengan keanggotan seperti pada **Tabel 4.8** berikut.

Tabel 4. 8 Klaster Optimum *Single Linkage*

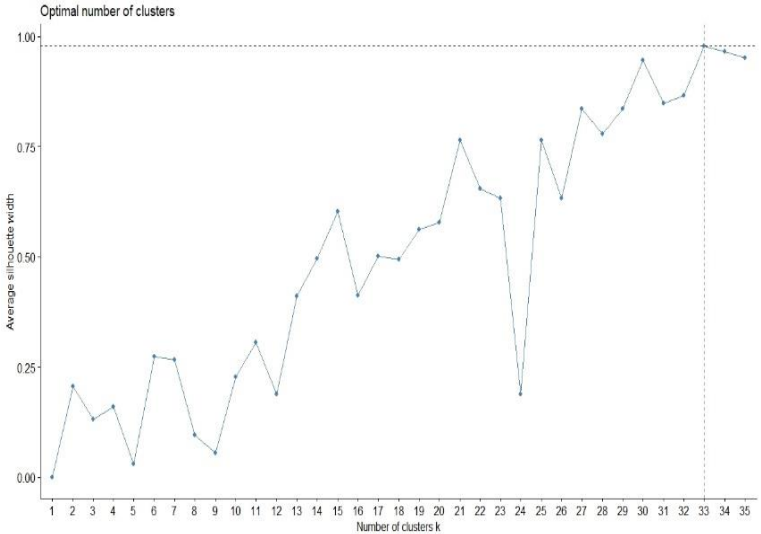
Klaster	**Jumlah Anggota**	**Anggota (*userId*)**
1	171	1, 3, 4, 8, 17, 19, 21, ..., 605, 606
2	10	2, 20, 120, 169, 273, 316, 418, 460, 544, 602
3	30	5, 11, 12, 18, 32, 38, 49, ..., 577, 599
⋮	⋮	⋮
34	1	585

Pada keanggotaan klaster satu terdapat 171 anggota, hal ini menunjukkan ternyata terdapat banyak *user* yang berkarakteristik demografik sejenis dengan karakteristik klaster satu sehingga banyak anggota yang terkumpul di dalamnya.

4.3.2 *K-Modes Clustering*

Analisis metode *K-Modes Clustering* adalah metode klaster yang kedua dalam penelitian ini. Analisis masih menggunakan data bertipe kategorik yang terdiri dari tiga variabel tiga variabel yaitu jenis kelamin, usia dan pekerjaan. *K-Modes* sangat disarankan

untuk digunakan untuk mengatasi permasalahan data kategorik dengan menggantikan ukuran *mean* menjadi modus. Langkah pertama dalam metode ini adalah menentukan k klaster inisiasi yang ingin dibentuk, dalam hal ini adalah 33 klaster.



Gambar 4. 8 *K-Modes Sillhouette Coefficient*

Nilai k klaster inisiasi didasarkan dari klaster optimum yang dapat diketahui dengan evaluasi klaster dengan pendekatan nilai *Silhouette* yang memiliki *range* antara -1 hingga 1. Semakin besar dan konvergen nilai *Silhouette* yang diperoleh, maka semakin tinggi pula peluang suatu klaster menjadi klaster optimal di dalam analisis. Sesuai dengan **Gambar 4.8**, nilai *Silhouette* tertinggi diperoleh jumlah klaster sebanyak 33. Terlihat dari garis putus-putus pada grafik yang menunjukkan keberadaan klaster optimal. Nilai *Silhouette* yang diperoleh adalah 0,978. Hasil tersebut adalah klaster optimum yang terbentuk dari metode *K-Modes Clustering.*

Kemudian menentukan *centroid* secara random. *Centroid* yang terpilih dalam metode *K-Modes Clustering* tercantum pada **Tabel 4.9** berikut**.**

Tabel 4. 9 *Centroid* Awal *K-Modes*

userId	Usia (u)	Jenis Kelamin (j)	Pekerjaan (p)
1	4	1	3
2	5	0	5
3	4	1	2
4	4	1	3
5	4	0	5
33	1	5	4

Centroid berfungsi sebagai sebuah pusat yang menjadi acuan terbentuknya kelompok dengan mengukur jarak setiap *user* terhadap *centroid* tersebut sehingga dapat ditentukan suatu *user* masuk dalam klaster mana. Perhitungan jarak menggunakan rumus seprti pada **Persamaan 2.5**. Contoh perhitungan jarak dilakukan dengan menggunakan k=3 dan *centroid* yang lebih sedikit mengacu pada tiga data awal **Tabel 4.9**, sehingga dapat memberikan kemudahan dalam memahami langkah analisis *K-Modes Clustering*. Perhitungan jarak *user* 1 terhadap *centroid* sebagai berikut, Dengan $\in$ adalah nilai pencocokan dengan aturan seperti pada **Persamaan 2.6**. Maka jarak *user* 1 terhadap *centroid* dihitung dengan cara sebagai berikut

$$d(x_1, c_1) = \in (1,1) + \in (3,3) + \in (4,4)$$
$$= 0 + 0 + 0$$
$$= \mathbf{0}$$

Berdasarkan hasil perhitungan jarak menunjukkan bahwa *user* 1 ke *centroid* 1 adalah 0. Perhitungan yang sama juga berlaku terhadap jarak setiap *user* terhadap *centroid*. Hasil perhitungan jarak dirangkum dalam **Tabel 4.10** sehingga memudahkan penentuan anggota klaster. Perhitungan nilai $\in$ apabila pada bernilai sama antara jarak obyek ke *centroid* maka akan bernilai nol, misalkan $\in (1,1) = 0$. Sedangkan apabila obyek ke *centroid* berbeda angka maka akan bernilai satu, contohnya $\in (2,3) = 1$. Perhitungan tersebut dilakukan pada semua obyek sehingga dapat diketahui kedekatan obyek terhadap *centroid* yang terbentuk pada klaster tersebut. Apabila *centroid* klaster diperbaiki maka dilakukan perhitungan jarak setiap obyek ke *centroid* yang baru.

Tabel 4. 10 Jarak ke *Centroid*

UserId	Jarak ke *Centroid*			Terdekat	Klaster
	1	2	3		
1	**0**	3	1	0	1
2	3	0	3	0	2
3	1	3	0	0	3
4	0	3	1	0	1
5	2	1	2	1	2
		...			

Kolom bertanda biru bergaris-garis menunjukkan nilai minimun jarak suatu *user* terhadap *centroid*. Jarak paling dekat *user* 1 ke *centroid* adalah jarak *user* 1 ke *centroid* 1 yang bernilai nol sehingga *user* 1 masuk ke dalam klaster 1. Hal yang sama berlaku pada setiap *user* sehingga dapat diketahui letak klaster dimana suatu *user* berada. Sehingga tersusun keanggotaan masing-masing klaster seperti pada **Tabel 4.11**.

Tabel 4. 11 Keanggotaan Klaster

userId	klaster 1	*userId*	klaster 2	*userId*	klaster 3
1	1 3 4	2	0 5 5	3	1 2 4
4	1 3 4	5	0 5 4		
	...		...		...

Langkah selanjutnya, perbarui *centroid* diambil dari modus setiap klaster yang nilainya sering muncul pada setiap klaster sehingga terbentuk *centroid* baru sebagai berikut

Tabel 4. 12 *Centroid* Baru

userId	Jenis Kelamin (j)	Pekerjaan (p)	Usia (u)
1	1	3	4
2	0	5	5
3	1	2	4

Ulangi langkah perhitungan jarak *user* terhadap *centroid* dan pembaharuan *centroid* berdasarkan modus dari masing-masing klaster terbentuk hingga memenuhi syarat, yaitu data klaster tersebut tidak bergerak atau posisi pusat *centroid* tidak berubah.

Dengan bantuan *software R* didapatkan hasil klaster optimum metode *K-Modes* adalah 33 klaster dengan melalui serangkaian prosedur yang telah dijelaskan. **Tabel 4.13** merupakan hasil klaster menggunakan metode *K-Modes*.

Tabel 4. 13 Klaster Optimum *K-Modes*

Klaster	*User* Anggota Klaster	Jumlah
1	179	1
2	9, 30, 33, 37, 66, 73, 83, ..., 608, 610	76
3	206, 281, 609	3
4	266	1
5	2, 20, 98, 120, 169, 273, ..., 544, 602	11
...	...	...
33	6, 93, 113, 227, 298, 335, 474, 576	8

Langkah analisis klaster yang terakhir adalah melakukan perbandingan hasil klaster optimum yang terbentuk menggunakan nilai evaluasi *Silhouette Coefficient* antara *Single Linkage Clustering* dengan metode *K-Modes Clustering* dalam keterkaitannya pada data demografik *user* film daring. Nilai *Sillhoutte Coefficient* yang diperbandingkan dicantumkan pada **Tabel 4.14** berikut.

Tabel 4. 14 Evaluasi Metode Klaster

Metode	*Sillhouette Coefficient*	Klaster Terbentuk
Single Linkage Clustering	0,985	34
K-Modes Clustering	0,978	33

Evaluasi klaster dengan pendekatan nilai *Silhouette* yang memiliki *range* antara -1 hingga 1. Semakin besar dan konvergen nilai *Silhouette* yang diperoleh, maka semakin tinggi pula peluang suatu klaster menjadi klaster optimal di dalam analisis. Berlandaskan evaluasi klaster maka metode *Single Linkage Clustering* mampu memberikan hasil pembentukan klaster lebih baik daripada metode *K-Modes Clustering* dengan nilai 0,985 dan membentuk 34 klaster. Maka hasil yang akan digunakan untuk analisis sistem rekomendasi adalah hasil klaster dengan metode *Single Linkage Clustering*. Hasil klaster juga memiliki manfaat

yaitu mampu mengelompokkan *user-user* film daring dalam suatu kelompok karakteristik yang sama, sehingga diharapkan sistem rekomendasi dapat lebih mewakili ketertarikan kalangan jenis kelamin, usia dan pekerjaan yang sejenis terhadap suatu film.

4.3.3 Hasil Analisis Klaster

Analisis klaster *Single Linkage* menjadi metode terbaik dan menghasilkan 34 klaster. Hasil klaster dievaluasi dengan *Sillhouette Coefficient* menghasilkan bahwa klaster karakteristik *user* terbentuk tergolong sangat baik. Penilaian ini dapat diketahui dengan karakteristik yang mewakili masing-masing klaster. Apabila kemiripan antar *user* dalam satu klaster sangat mirip/dekat maka klaster makin baik. Sedangkan ditinjau dari kemiripan antar klaster yang berbeda semakin kecil maka semakin baik pengelompokkan klaster. Berdasarkan parameter klaster yang baik tersebut maka pada **Tabel 4.15** dicantumkan contoh karakteristik beberapa klaster hasil metode *Single Linkage Clustering* agar mampu mengetahui apakah klaster terbentuk sudah memenuhi kriteria klaster yang baik. Karakteristik masing-masing klaster diperoleh melalui nilai modus karena tipe data kategorik.

Tabel 4. 15 Karakteristik Anggota per Klaster

klaster	*user*	usia	jenis kelamin	pekerjaan
1	1	4	1	2
	3	4	1	2
	...	...	...	...
	606	4	1	2
2	2	5	0	5
	20	5	0	5
	...	...	...	...
	602	5	0	5
34	34	6	1	4

Modus masing-masing variabel pada klaster-klaster jika merujuk pada pengertian golongan pada **Tabel 4.1** dan **Tabel 4.2** maka karakteristik klaster 1 adalah seorang laki-laki golongan usia dewasa awal dengan pekerjaan Tenaga Profesional. Klaster 2 adalah seorang perempuan golongan usia dewasa madya dengan pekerjaan *Others*/lainnya. Sedangkan klaster 34 adalah seorang

laki-laki dengan usia tergolong usia lanjut berprofesi sebagai Tenaga Tata Usaha. Diperoleh kesimpulan bahwa karakteristik masing-masing *user* dalam satu klaster sangat mirip sehingga memenuhi parameter kebaikan klaster bahwa kemiripan antar anggota dalam satu klaster sangat mirip/kuat.

Tabel 4. 16 Karakteristik Anggota antar Klaster

Klaster	usia	jenis kelamin	pekerjaan	jumlah anggota
1	4	1	2	171
2	5	0	5	10
34	6	1	4	1

Dapat disimpulkan contoh hasil ketiga klaster di atas sangat berbeda satu sama lain sehingga memenuhi parameter klaster dikatakan baik karna kemiripan antarklaster sangat kecil. Contohnya meski klaster 34 beranggotakan hanya satu *user* tetapi memang karakteristiknya sangat berbeda dengan karakteristik klaster yang lain sehingga layak untuk membentuk klaster tersendiri. Hasil klaster terbaik metode Single Linkage Clustering ini akan dipergunakan dalam analisis rekomendasi *UBCF*.

4.3.4 *User-Based Collaborative Filtering*

Langkah sistem rekomendasi selanjutnya adalah mengolah data hasil klaster dengan menggunakan metode *User-Based Collaborative Filtering* (*UBCF*) yang memiliki konsep menemukan sekumpulan *user* yang memiliki historis kesukaan yang sama dengan *user* yang akan dijadikan sasaran rekomendasi. Setelah sekumpulan tetangga terbentuk, sistem menggabungkan film kesukaan tetangga (*neighbour*) untuk menghasilkan rekomendasi kepada *user* (Sarwar dkk, 2001). Variabel yang terlibat dalam analisis metode ini adalah *userId, movieId,* dan *rating*. Dalam melakukan analisis metode *UBCF*, pertama-tama dilakukan penyusunan matriks *rating*.

4.3.5 Penyusunan Matriks *Rating*

Langkah penyusunan matriks *rating* dalam sistem rekomendasi melibatkan tiga komponen yakni keterangan kolom horizontal berupa *movieId*, keterangan baris vertikal merupkan *userId* dan isi dari sel matriks *rating* merupakan nilai *rating* yang

diberikan oleh seorang *user* terhadap suatu film. Matriks *rating* yang terbentuk berasal dari hasil klaster metode *Single Linkage Clustering* yang berjumlah 34 klaster. Masing-masing klaster akan disusun matriks *rating* sehingga susunan matriks *rating* pun berjumlah 34. Penyusunan ini bertujuan untuk memudahkan pengolahan pada langkah selanjutnya yakni perhitungan *Adjusted Cosine* antara *user* dengan *user* yang lain. Sebelum menyusun matriks tersebut maka perlu diketahui jumlah keanggotaan setiap klaster hasil metode *Single Linkage Clustering* yang tercantum pada **Tabel 4.17**.

Tabel 4. 17 Anggota Hasil Klaster

Klaster (i)	**Jumlah Anggota (n)**	**Anggota (*userId*)**	**Jumlah Film Ditonton (A)**
1	171	1, 3, 4, 8, 17, 19, 21, ..., 605, 606	6839
2	10	2, 20, 120, 169, 273, 316, 418, 460, 544, 602	726
⋮	⋮	⋮	⋮
34	1	585	140

Setelah mengetahui anggota masing-masing klaster penyusunan matriks *rating* berdimensi (n_i x A_i) sehingga terbentuk 33 buah matriks *rating*. Matriks *rating* pada klaster 1 digunakan sebagai contoh memiliki dimensi (171 x 6839) seperti pada **Tabel 4.18** berikut.

Tabel 4. 18 Matriks *Rating*

movieId / *userId*	1	2	3	4	5	6	...	193609
1	**4**	0	4	0	0	4		0
3	0	0	0	0	0	0	...	0
4	0	0	0	0	0	0		0
8	0	4	0	0	0	0		0
...				...				
606	2,5	0	0	0	0	0	...	0

Nilai 4 yang terdapat pada sel matriks pertama artinya adalah *user* 1 memberikan nilai *rating* sebesar 4 terhadap *movie* 1 yang telah ditonton olehnya. Setelah disusun matriks *rating* maka dilakukan perhitungan menggunakan *adjusted cosine* untuk memperoleh nilai kemiripan antar *user* berdasarkan *rating* yang diberikan dan disusun dalam matriks *adjusted cosine* yang tercantum pada **Tabel 4.19**. Perhitungan jarak bertujuan menentukan kedekatan *user* dengan *user* yang lainnya menggunakan data *rating*.

4.3.6 Perhitungan *Adjusted Cosine*

Perhitungan nilai *adjusted cosine* dilakukan dengan menggunakan rumus seperti **Persamaan 2.1.** Sebagai contoh dihitung jarak *user* 1 dengan *user* 3,

$$S_{(1,3)} = \frac{(4)(0) + (0)(0) + \cdots + (0)(0)}{\sqrt{(4)^2 + (0)^2 + \cdots + (0)^2}\sqrt{(0)^2 + (0)^2 + \cdots + (0)^2}}$$

$$S_{(1,3)} = 0{,}0597$$

Semua jarak antar *user* dilakukan perhitungan dengan rumus yang sama dan kemudian hasil perhitungan tersebut disusun dalam matriks *Adjusted Cosine* seperti pada **Tabel 4.19** berikut. Sehingga dapat diketahui kemiripan karakteristik *rating* antar *user* atau penonton film daring.

Tabel 4. 19 Matriks *Adjusted Cosine*

userId / *userId*	1	3	4	8	...	606
1	1	0,059	0,194	0,136		0,164
3	**0,059**	1	0,002	0,004		0,012
4	0,194	0,002	1	0,062	...	0,200
8	0,136	0,004	0,062	1		0,099
...			...			
606	0,164	0,012	0,200	0,099	...	1

Perhitungan dengan cara yang sama, dilakukan pada setiap *user* dan semua hasil nilai koefisien *adjusted cosine* disusun dalam matriks *adjusted cosine* serta matriks tersebut dijadikan acuan

untuk melakukan prediksi pada *movie* yang belum memiliki *rating*. Diketahui bahwa *user* 3 belum pernah menonton *movie* 1 sehingga *rating user* 3 untuk *movie* 1 bernilai kosong. Dalam proses rekomendasi diperlukan untuk mengetahui nilai *rating* pada semua *movie* sehingga dilakukan prediksi *rating*. Hal yang diperlukan dalam prediksi adalah nilai koefisien kemiripan karakteristik *rating* antar *user* yakni *user* 3 terhadap semua *user* dan nilai *rating user* lain terhadap *movie* 1 tersebut yang akan dihitung dengan rumus *weighted sum* pada **Persamaan 2.2**.

4.3.7 Prediksi Nilai *Rating*

Tahap selanjutnya adalah melakukan perhitungan untuk memprediksi *rating* film yang masih belum terisi. Perhitungan dilakukan menggunakan *weighted sum* dengan melihat pada tabel matriks *rating* dan *adjusted cosine*, contoh akan dihitung nilai prediksi rating bagi user 3 pada film 1 sebagai berikut.

$$P_{(u,i)} = \frac{(0)(0{,}059) + (0)(0{,}194) + \cdots + (2{,}5)(0{,}164)}{(0{,}059)^2 + (0{,}194)^2 + \cdots + (0{,}164)^2}$$

$$P_{(u,i)} = 2{,}737$$

Dengan bantuan *software* R sehingga menghasilkan hasil prediksi seperti pada **Tabel 4.20**.

Tabel 4. 20 Prediksi *Rating*

movieId / *userId*	1	2	3	4	5	6	...	193609
1	4	**4,409**	4	**4,31**	**4,33**	4		**4,366**
3	**2,737**	**2,419**	**2,410**	**2,43**	**2,35**	**2,5**	...	**2,435**
4	**3,695**	**3,598**	**3,537**	**3,47**	**3,56**	**3,537**		**3,555**
8	**3,658**	4	**3,562**	**3,57**	**3,57**	**3,56**		**3,574**
...				...				
606	2,5	**3,723**	**3,624**			**3,658**	**...**	**3,657**

Keterangan : angka bercetak tebal merupakan hasil prediksi

Proses selanjutnya setelah didapatkan prediksi *rating* adalah menggunakan *Ranking-Based Technique* untuk mengurutkan *rating* film dari tertinggi hingga terendah pada masing-masing dan memberikan rekomendasi 10 film terbaik yang berpotensi besar untuk ditonton oleh *user* tertentu. *Ranking-Based Technique*

dilakukan pada semua *user* pada masing-masing klaster sehingga setiap *user* memiliki 10 rekomendasi film yang relevan untuk ditonton.

4.4 Hasil Rekomendasi Film

User 1 diketahui memiliki kedekatan dengan *user* 288 yang berada pada klaster yang sama yaitu klaster 1. Sehingga prediksi *rating* film *user* 1 dilakukan dengan menggunakan hasil jarak kedekatan kedua *user* tersebut dengan perhitungan *weighted sum*. Terakhir dengan *Ranking-Based Technique* diperoleh 10 rekomendasi film terbaik untuk *user* 1 yang dicantumkan pada **Tabel 4.21**.

Tabel 4. 21 Contoh Rekomendasi

No	*Title*	*Genres*
1	The Imitation Game (2014)	*Drama|Thriller|War*
2	WALLÂ·E (2008)	*Adventure|Animation|Children|Romance|Sci-Fi*
3	Graduate, The (1967)	*Comedy|Drama|Romance*
4	Shawshank Redemption, The (1994)	*Crime|Drama*
5	Interstellar (2014)	*Sci-Fi|IMAX*
6	Dark Knight, The (2008)	*Action|Crime|Drama|IMAX*
7	Rumble in the Bronx (Hont faan kui) (1995)	*Action|Adventure|Comedy|Crime*
8	Shakespeare in Love (1998)	*Comedy|Drama|Romance*
9	Star Trek: First Contact (1996)	*Action|Adventure|Sci-Fi|Thriller*
10	Wolf of Wall Street, The (2013)	*Comedy|Crime|Drama*

Hasil Rekomendasi *user* 1 pada klaster 1 menunjukkan bahwa rekomendasi film yang cocok disaksikan oleh *user* atau penonton dengan karakteristik demografik sebagai seorang laki-laki yang berusia 24 tahun dengan pekerjaan sebagai *technician* adalah film-film dengan judul seperti pada **Tabel 4.21** salah satu contohnya adalah Interstellar (2014) atau tipe-tipe *genre* film *comedy, action, drama, sci-fi* dan *adventure*.

Ilustrasi *Ranking-Based Technique* untuk 10 rekomendasi dilakukan dengan cara melakukan pemeringkatan hanya pada *rating* hasil prediksi saja dari tertinggi hingga terendah menggunakan **Tabel 4.20** kemudian urutan peringkat *rating* dirangkum pada **Tabel 4.22** untuk rekomendasi *movie* terbaik bagi *user* 1 dengan bantuan *syntax R* pada **Lampiran 6**.

Tabel 4. 22 Peringkat *Movie* Berdasarkan *Rating*

MovieId	*Title*	Hasil Prediksi
116797	The Imitation Game	4,514
60069	WALLÂ·E	4,511
1247	Graduate, The	4,498
318	Shawshank Redemption, The	4,484
109487	Interstellar	4,477
588559	Dark Knight, The	4,475
112	Rumble in the Bronx (Hont faan kui)	4,474
2396	Shakespeare in Love	4,471
1356	Star Trek: First Contact	4,466
106782	Wolf of Wall Street, The	4,462

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut.

1. Eksplorasi karakteristik data yang dilakukan menggunakan data u.*user* berisi informasi demografik *user* yang terdiri dari tiga variabel yaitu jenis kelamin. usia, dan pekerjaan. Sedangkan data u.*data* berisikan variabel *rating*. Hasil eksplorasi diperoleh kesimpulan sebagai berikut :
   a. laki-laki lebih dominan menyaksikan film daring dibanding perempuan karena jumlahnya 72 persen dari keseluruhan *user* / penonton.
   b. Golongan *user* yang mengunjungi situs film daring didominasi penonton berusia 22 hingga 40 tahun yang tergolong dewasa awal sebesar 56,7 persen.
   c. Jenis golongan pekerjaan *user* yang menonton film daring tenyata penonton terbanyak merupakan golongan tenaga profesional dengan persentase jumlah 37 persen. Terdiri dari pekerjaan *doctor*, *educator*, *engineer*, *healthcare*, *lawyer*, *programmer*, *scientist*, dan *writer*
   d. Kecenderungan semua golongan usia dari *user* memberikan nilai *rating* lebih dari 3,5 dapat diartikan bahwa nilai di atas 3,5 merupakan nilai apresiasi yang cukup bagus terhadap film daring yang ditonton oleh masing-masing golongan usia *user*.
2. Hasil analisis klaster metode *Single Linkage Clustering* tebentuk 34 klaster dengan nilai evaluasi kebaikan klaster *Sillhouette Coefficient* sebesar 0,985. Sedangkan analisis klaster metode *K-Modes Clustering* tebentuk 33 klaster dengan nilai evaluasi kebaikan klaster *Sillhouette Coefficient* sebesar 0,978. Hasil perbandingan metode klaster terbaik diperoleh hasil bahwa metode *Single Linkage*

Clustering lebih baik dalam memodelkan hasil klaster data *user* film.

3. Analisis sistem rekomendasi yang digunakan adalah *hybrid* karena menggunakan hasil klaster metode *Single Linkage Clustering* pada data demografik *user* sehingga termasuk *Demographic Filtering* dan dilanjutkan dengan metode *User-Based Collaborative Filtering* (*UBCF*). Langkah analisis yang dilakukan metode *UBCF* adalah penyusunan matrik *rating*, perhitungan jarak *Adjusted Cosine*, prediksi *rating*, dan perekomendasian menggunakan *Ranked-Based Technique* dengan jumlah 10 rekomendasi film terbaik untuk ditonton oleh masing-masing *user*.
4. Hasil Rekomendasi diberikan kepada *user* tertentu sejumlah 10 rekomendasi film, sebagai contoh adalah *user* 1 pada klaster 1 menunjukkan bahwa rekomendasi film yang cocok disaksikan oleh *user* atau penonton dengan karakteristik demografik sebagai seorang laki-laki yang berusia 24 tahun dengan pekerjaan sebagai *technician* adalah film-film dengan judul seperti Interstellar (2014) atau tipe-tipe *genre* film *comedy, action, drama, sci-fi* dan *adventure*.

5.2 Saran

Saran yang dapat diberikan oleh peneliti terkait analisis yang telah dilakukan adalah cara *pre-processing* memegang peranan penting dalam proses analisis lebih lanjut maupun interpretasi. Oleh karena itu perlu adanya landasan yang tepat yang dijadikan sebagai rujukan dalam setiap tindakan penggolongan kategori data. Bagi peneliti lain dapat menggunakan metode analisis klaster lain yang mampu memberikan hasil yang lebih baik dan basis metode rekomendasi lain.

DAFTAR PUSTAKA

Adomavicius, Gediminas, dan Tuzhilin. (2005). Toward the Next Generation of Recommender Systems : A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering,* vol 17(6), pp. 734-745.

Alfina , T., Santosa, B., dan Barakbah, A. R. (2012). *Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data* . Surabaya: Institut Teknologi Sepuluh Nopember.

Arvid, T., Setyohadi, Budiyanto, D., dan Ernawati. (2016). *User-Based Collaborative Filtering Dengan Memanfaatkan Pearson-Correlation Untuk Mencari Neighbors Terdekat Dalam Sistem Rekomendasi*. Dipetik Maret 05, 2018, dari http://e-journal.uajy.ac.id/8924

Asanov, D. (2015). Algorithms ang Methods in Recommender System. *International Journal of Computer Applications,* vol 118, pp. 1-5

Azizah, Z. (2016). *Sistem Rekomendasi Film Dengan Teknik Collaborative Filtering Menggunakan Bayesian Classifier*. Surabaya: Jurnal Sains dan Seni ITS, hal 10-22.

Berry, M. W., dan Kogan, J. (2010). *Text Mining Application and Theory*. United Kingdom: WILEY.

Burke, R. (2007). Hybrid Web Recommender System. *Lect. Notes Comput.Sc,* 6(4321), pp. 377-408.

Dapiah, Iriawan, N., dan Fithriasari, K. (2018). On The Modelling of The Average Value of High School Examination in West Java Using Bayesian Hierarchical Mixture Normal Approach. *International Conference on Information and Communication Technology, ICOIACT,* 1, pp. 689-694.

Dwicahya, I. (2018). *Perbandingan Sistem Rekomendasi Film Metode User-Based dan Item-Based Collaborative Filtering.* Yogyakarta: Media Teknika Jurnal Teknologi Universitas Sanata Dharma, vol 1, hal. 20-35

Feldman, R., dan Sanger, J. (2007). *The Text Mining Handbook : Advanced Approaches in Analyzing Unsrtructed Data*. New York: Cambridge University Press.

Goldberg, d. (1992). Using Collaborative Filtering to Weave An Information Tapestry. *Communication of the ACM*, 35(12), pp. 61-70.

Halim, A., Gohzali, H., dan Panjaitan , D. M. (2017). Sistem Rekomendasi Film Menggunakan Bisecting K-Means dan Collaborative Filtering. *CITISEE*, vol 1, hal. 38-41.

Han, J., Kamber, M., dan Pei, J. (2012). *Data Mining Concepts and Techniques*. USA: Morgan Kaufmann.

Handoyo, R., Rumani, M. R., dan Nasution, S. M. (2014). Perbandingan Metode Clustering menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen. *Jurnal SIFO Mikroskil (JSM)*,vol 15(2), hal. 73-82

Hapsari, Y. G., Wibowo, A. T., dan Baizal, Z. A. (2015). Analisis dan Implementasi Sistem Rekomendasi Menggunakan Most-Frequent Item dan Association Rule Technique. *e-Proceeding of Engineering*, vol 2(3), hal. 7757-7764.

Jannah, S. Z., Fithriasari, K., dan Usagawa, T. (2018). *Clustering and Visualizing Surabaya Citizen Aspirations by Using Text Mining*. Surabaya: Institut Teknologi Sepuluh Nopember.

Mahdavi, M., dan Dakhel, G. M. (2011). A New Collaborative Filtering Algorithm Using K-Means Clustering and Neighbors' Voting. *International Conference on Hybrid Intelligent System*, vol 2, pp. 179-184.

Nduru, E. K., Buulolo, E., dan Pristiwanto. (2018). Implementasi Algoritma K-Modes Untuk Menentukan Strategi Marketing STMIK Budi Dharma. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*. STMIK Budi Dharma, vol 2, hal. 12-19

Ngafifi, M. (2014). Kemajuan Teknologi dan Pola Hidup Manusia Dalam Prespektif Sosial Budaya. *Jurnal Pembangunan Pendidikan : Fondasi dan Aplikasi, 2*(1), 33-47.

Nilashi, M., Bagherifard, K., Ibrahim, O., Alizadeh, H., Nojeem, L. A., dan Roozegar, N. (2012). Collaborative Filtering Recommender System. *Research Journal of Applied Science, Engineering and Technology*, vol 5(16), pp. 4168-4182.

P, L., Gemmis, d. M., dan G, S. (2010). Content-Based Recommender Systems, *State of the Art and Trends*. vol 1 pp. 74-105

Ricci, L., dan Saphira, B. (2011). *Recommender System Handbook.* New York, USA: Springer Science and Bussiness Media.

Sarwar, B., Karypis, G., Konstan, J., dan Riedl, J. (2001). *Item-Based Collaborative Filtering Recommendation Algorithms.* GroupLens Research Group.

Sharma, N., dan Gaud, N. (2015). K-Modes Clustering Algorithm for Categorical Data. *International Journal of Computer Applications*, vol 127, pp. 1-6.

Srivastava, A., dan Sahami, M. (2009). *Text Mining Classification, Clustering, and.* USA: Taylor and Francis Group, LLC.

Statistik, B. P. (2002). *Klasifikasi Baku Jenis Pekerjaan Indonesia (KBJI)* . Jakarta: Badan Pusat Statistik.

Stuetzle, W. &. (2010). A Generalized Single Linkage Method for Estimating the Cluster Tree of a Density. . *The Journal of Computational and Graphical Statistics*, vol 19(2), pp. 397-418

Walpole, R. (1995). *Pengantar Metode Statistika.* Diterjemahkan oleh Ir. Bambang Sumantri. Edisi Ketiga. Jakarta: Gramedia Pustaka Utama.

Zhang, C. (2013). An Improved K-Means Clustering Algorithm. *Journal of Information & Computational Science*, vol 10, pp. 193-199.

Zolakhaf, Z., Babanezhad, R., dan Pottinger, R. (2018). A Generic Top-N Recommendation Framework For Trading-off Accuracy, Novelty, and Coverage.*IEEE*, vol 1, pp. 149-160.

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran 1 Struktur Data

Obs.	*userid*	usia	jenis kelamin	pekerjaan	*movieId*	*rating*
1	1	4	1	3	1	4
2	1	4	1	3	3	4
3	1	4	1	3	6	4
4	1	4	1	3	47	5
5	1	4	1	3	50	5
6	1	4	1	3	70	3
7	1	4	1	3	101	5
8	1	4	1	3	110	4
9	1	4	1	3	151	5
10	1	4	1	3	157	5
11	1	4	1	3	163	5
12	1	4	1	3	216	5
13	1	4	1	3	223	3
14	1	4	1	3	231	5
15	1	4	1	3	235	4
16	1	4	1	3	260	5
17	1	4	1	3	296	3
18	1	4	1	3	316	3
19	1	4	1	3	333	5
20	1	4	1	3	349	4
21	1	4	1	3	356	4
22	1	4	1	3	362	5
...	...	...	...	...	...	...
100.834	610	4	1	5	168250	5
100.835	610	4	1	5	168252	5
100.836	610	4	1	5	170875	3

Lampiran 2 Contoh Data *Movie*

movieId	*title*	*genres*
1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
2	Jumanji (1995)	Adventure|Children|Fantasy
3	Grumpier Old Men (1995)	Comedy|Romance
4	Waiting to Exhale (1995)	Comedy|Drama|Romance
5	Father of the Bride Part II (1995)	Comedy
6	Heat (1995)	Action|Crime|Thriller
7	Sabrina (1995)	Comedy|Romance
8	Tom and Huck (1995)	Adventure|Children
9	Sudden Death (1995)	Action
10	GoldenEye (1995)	Action|Adventure|Thriller
	...	
193585	Flint (2017)	Drama
193587	Bungo Stray Dogs: Dead Apple (2018)	Action|Animation
193609	Andrew Dice Clay: Dice Rules (1991)	Comedy

Lampiran 3 *Syntax R* untuk *Single Linkage*

```
--Single Linkage Clustering--
user.hc<-hclust(dist(user, method = "euclidean"),"single")
plot(user.hc)
summary(user)

--Evaluasi Klaster SLC--
fviz_nbclust(user,FUN=hcut,method="wss",k.max=20)
fviz_nbclust(user,FUN=hcut,method="silhouette",k.max=40)
+ geom_hline(yintercept = 0.988, linetype = 2)
user.cut<-cutree(user.hc,34)

--Isi dalam klaster--
table(user.cut)
rownames(user)[user.cut==15]

--Dendogram SLC--
plot(user.hc,hang=-10)
rect.hclust(user.hc,k=33,border=33:34)
fviz_cluster(list(data=user,cluster=user.cut))
```

Lampiran 4 *Syntax R* untuk *K-Modes*

```
--Kmodes Clustering--
result<-kmodes(user,34)
fviz_nbclust(user, kmodes, method = "wss", k.max=20)
fviz_nbclust(user, kmodes, method = "silhouette",
k.max=35)+ geom_hline(yintercept = 0.978, linetype = 2)
result_kmodes<-kmodes(user,33)

--Isi dalam klaster--
table(result_kmodes$cluster)
rownames(user)[result_kmodes$cluster==2]

--Plot hasil K-Modes--
clusplot (user, result_kmodes$cluster, color=TRUE,
shade=FALSE, labels=2, lines=0)
```

Lampiran 5 *Syntax R* untuk Sistem Rekomendasi

```
--Sistem Rekomendasi UBCF--
--Masukkan data rating--
rating <- read.csv("D:/Data1/KULIAH PER
SEMESTER/SEMESTER VIII/Data Spesial
TA/KUMPULAN DATA MOVIES/Data Pakai
TA/rating.csv",sep=",",header=TRUE)
rating$timestamp=NULL
ratings <-rating[,-c(4)]

--Masukkan data movies--
movies <- read.csv("D:/Data1/KULIAH PER
SEMESTER/SEMESTER VIII/Data Spesial
TA/KUMPULAN DATA MOVIES/Data Pakai
TA/movies.csv")

rec_sys = list()
for (i.Cluster in 1:length(unique(user.cut))){
   rec_sys[[i.Cluster]] = rating[which(rating$userId %in%
as.numeric(rownames(user)[user.cut==i.Cluster]) ),]}

--Susun Matriks Rating per Klaster yang Diinginkan—
(contoh user 1)
cast_data <-
acast(rec_sys[[1]],rec_sys[[1]]$userId~rec_sys[[1]]$movieId)
rating_mat <- as.matrix(cast_data)
rating_real_mat <- as(rating_mat,"realRatingMatrix")
rating_real_mat
recommender <-
Recommender(rating_real_mat[1:nrow(rating_real_mat)],met
hod="UBCF",
                    param=list(normalize="Z-
score",method="cosine"
                               ,nn=25,minRating=1))
```

Lampiran 6 *Syntax R* untuk Sistem Rekomendasi (Lanjutan)

```
recommend_topN <-
predict(recommender,rating_real_mat[1:nrow(rating_real_mat
)],type="topNList",n=10)
recommend_top_list <-as(recommend_topN,"list")
recommend_movies <- function(rec_list,movies,user_id){
rec_movie_ids <- rec_list[[user_id]];
m1<- movies[which(movies$movieId==rec_movie_ids[1]),];
m2<- movies[which(movies$movieId==rec_movie_ids[2]),];
m3<- movies[which(movies$movieId==rec_movie_ids[3]),];
m4<- movies[which(movies$movieId==rec_movie_ids[4]),];
m5<- movies[which(movies$movieId==rec_movie_ids[5]),];
m6<- movies[which(movies$movieId==rec_movie_ids[6]),];
m7<- movies[which(movies$movieId==rec_movie_ids[7]),];
m8<- movies[which(movies$movieId==rec_movie_ids[8]),];
m9<- movies[which(movies$movieId==rec_movie_ids[9]),];
m10<-
movies[which(movies$movieId==rec_movie_ids[10]),];
movie_names=rbind(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10)
return(movie_names);
}

--10 Rekomendasi Film—
(contoh user 1 di klaster 1)
Sort(recommend, decreasing = TRUE)
recommend_movies(recommend_top_list,movies,1)
```

Lampiran 7 *Syntax R* untuk *GUI*

```
#-- Install Package ----
library(shiny)
library(cluster)
library(klaR)
library(dendextend)
library(factoextra)
library(ggplot2)
library(tadaatoolbox)
library(matlib)
library(car)
library(rgl)
library(recommenderlab)
library(reshape2)
library(SnowballC)
library(lsa)
#-- Load Saved Model
# setwd('D:/9. Others/Help/Syihab - TA/')
load("D:/Data1/KULIAH PER SEMESTER/SEMESTER
VIII/Data Spesial TA/SYNTAX R/TA/Model_GUI.RData")
# Define UI for dataset viewer app ----
ui <- fluidPage(
  # App title ----
  titlePanel("Movie Recommendation"),
  # Sidebar layout with a input and output definitions ----
  sidebarLayout(
    # Sidebar panel for inputs ----
    sidebarPanel(
      # Input: Selector for choosing gender ----
      selectInput(inputId = "Gender",
                  label = "Gender:",
                  choices = c("Male", "Female")),
      # Input: Numeric entry for age ----
      numericInput(inputId = "Age",
                   label = "Age:",
                   value = 17),
```

Lampiran 8 *Syntax R* untuk *GUI* (Lanjutan)

```
# Input: Selector for choosing job ----
    selectInput(inputId = "Job",
                label = "Occupation:",
                choices = c('None',
                            'Administrator',
                            'Artist',
                            'Doctor',
                            'Educator',
                            'Engineer',
                            'Entertainment',
                            'Executive',
                            'Healthcare',
                            'Homemaker',
                            'Lawyer',
                            'Librarian',
                            'Marketing',
                            'Programmer',
                            'Retired',
                            'Salesman',
                            'Scientist',
                            'Student',
                            'Technician',
                            'Writer',
                            'Other')),
      # Input: Selector for choosing movie title ----
    selectInput(inputId = "Title",
                label = "Choose movie:",
                choices = NULL),
    # Input: Selector for choosing rating ----
    selectInput(inputId = "Rating",
                label = "Your rating: (Higher Better)",
                choices =
c("0.5","1","1.5","2","2.5","3","3.5","4","4.5","5"))    ),
```

Lampiran 9 *Syntax R* untuk *GUI* (Lanjutan)

```
# Main panel for displaying outputs ----
    mainPanel(

      # Output: HTML table with requested number of
observations ----
      tableOutput("view")
    )))
# Define server logic to summarize and view selected dataset -
---
server <- function(input, output, session) {
  #-- Movie Title Selection ----
  updateSelectizeInput(session, 'Title', choices = movies$title,
                       server = TRUE)
 #-- Prediction ----
  predictions<-reactive({
    # Converting Data (Mengubah Input user menjadi input
sesuai program)
    xGender = if (input$Gender == "Male"){1}else{0}
    xAge    = if (input$Age < 12){1}else
      if(input$Age < 16){2}else
        if(input$Age < 22){3}else
          if(input$Age < 41){4}else
            if(input$Age < 61){5}else{6}
    xJob    = if (input$Job == "Executive"){1}else
      if(input$Job == "Doctor"){2}else
        if(input$Job == "Educator"){2}else
          if(input$Job == "Engineer"){2}else
            if(input$Job == "Healthcare"){2}else
              if(input$Job == "Lawyer"){2}else
                if(input$Job == "Progammer"){2}else
                  if(input$Job == "Scientist"){2}else
                    if(input$Job == "Writer"){2}else
                      if(input$Job == "Artist"){3}else
                        if(input$Job == "Entertainment"){3}else
                          if(input$Job == "Marketing"){3}else
                            if(input$Job == "Technician"){3}else
```

Lampiran 10 *Syntax R* untuk *GUI* (Lanjutan)

```
if(input$Job == "Administrator"){4}else
                        if(input$Job == "Librarian"){4}else
                          if(input$Job == "Other"){5}else
                            if(input$Job == "Homemaker"){5}else
                              if(input$Job == "None"){5}else
                                if(input$Job == "Retired"){5}else
                                  if(input$Job ==
"Salesman"){5}else{5}
  xMovie = movies$movieId[which(movies$title ==
input$Title)]
   xRating = input$Rating
   # Predicting New Data
   # Masukkan data user baru
   new_data = c(xGender,xAge,xJob,xMovie,xRating)
   new_data = data.frame(matrix(new_data, nrow = 1))
   colnames(new_data) =
c('Jenis.Kelamin','job','agecode','movieId','rating')
   # Mencari cluster user baru
   new_cluster =
unique(dataCluster$cluster[which(dataCluster$Jenis.Kelamin
== new_data$Jenis.Kelamin &
dataCluster$job == new_data$job &
dataCluster$agecode == new_data$agecode)])
if (anggota_cluster[new_cluster] > 1){
    # Mencari rekomendasi untuk semua user yang masuk
dalam cluster baru
recommend_topN <-
predict(model_recommender[[new_cluster]],
model_input[[new_cluster]][1:nrow(model_input[[new_cluste
r]])],
type="topNList",n=10)
# Mencari user baru lebih deket ke userId mana
xTemp = rec_sys[[new_cluster]]
xTemp = rbind(c(0,0,0),xTemp)
rownames(xTemp) = c(1:nrow(xTemp))
xTemp[1,2] = new_data$movieId
xTemp[1,3] = new_data$rating
```

Lampiran 11 *Syntax R* untuk *GUI* (Lanjutan)

```
cast_data <- acast(xTemp,xTemp$userId~xTemp$movieId)
      rating_mat <- as.matrix(cast_data)
      rating_mat[which(is.na(rating_mat)==TRUE)] = 0
      cosine_mat <- cosine(t(rating_mat))
      x_similar = which(cosine_mat[-1,1] == max(cosine_mat[-
1,1]))
      user_similar <- names(x_similar[1])
      all_recommend = as(recommend_topN,"list")
      recommend_topN =
all_recommend[[which(names(all_recommend) ==
user_similar)]]
    }else{
      # spesial case untuk cluster yang cuma satu anggota
      recommend_topN <-
model_recommender[[new_cluster]][c(1:10)]
    }

    #-- Mengeluarkan Rekomendasi Final ----
    # movies <- read.csv("D:/Data1/KULIAH PER
SEMESTER/SEMESTER VIII/Data Spesial
TA/KUMPULAN DATA MOVIES/Data Pakai
TA/movies.csv")
    final_recommend = movies[movies$movieId %in%
as.numeric(recommend_topN),-1]
    final_recommend
  })
    # Show the first "n" observations ----
  output$view = renderTable({   # the last 6 rows to show
    pred = predictions()
    head(pred, n = 10)
  })}
# Create Shiny app ----
shinyApp(ui = ui, server = server)
```

Lampiran 12 Contoh Rekomendasi Hasil Klaster 1

User	No	Title	Genres
1	1	The Imitation Game (2014)	Drama|Thriller|War
	2	WALLÂ·E (2008)	Adventure|Animation|Children|Romance|Sci-Fi
	3	Graduate, The (1967)	Comedy|Drama|Romance
	4	Shawshank Redemption, The (1994)	Crime|Drama
	5	Interstellar (2014)	Sci-Fi|IMAX
	6	Dark Knight, The (2008)	Action|Crime|Drama|IMAX
	7	Rumble in the Bronx (Hont faan kui) (1995)	Action|Adventure|Comedy|Crime
	8	Shakespeare in Love (1998)	Comedy|Drama|Romance
	9	Star Trek: First Contact (1996)	Action|Adventure|Sci-Fi|Thriller
	10	Wolf of Wall Street, The (2013)	Comedy|Crime|Drama
3		...	

Lampiran 13 Contoh Hasil Rekomendasi Klaster 2

User	No	Title	Genres
2	1	To Die For (1995)	Comedy|Drama|Thriller
	2	Matrix, The (1999)	Action|Sci-Fi|Thriller
	3	Schindler's List (1993)	Drama|War
	4	Lord of the Rings: The Fellowship of the Ring, The (2001)	Adventure|Fantasy
	5	Lord of the Rings: The Two Towers, The (2002)	Adventure|Fantasy
	6	Beautiful Mind, A (2001)	Drama|Romance
	7	Wizard of Oz, The (1939)	Adventure|Children|Fantasy|Musical
	8	West Side Story (1961)	Drama|Musical|Romance
	9	Christmas Story, A (1983)	Children|Comedy
	10	Chicago (2002)	Comedy|Crime|Drama|Musical
20		...	

Lampiran 14 Contoh Hasil Rekomendasi Klaster 3

User	No	Title	Genres
5	1	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Sci-Fi
	2	Silence of the Lambs, The (1991)	Crime|Horror|Thriller
	3	Back to the Future (1985)	Adventure|Comedy|Sci-Fi
	4	Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)	Action|Adventure
	5	LÃ©on: The Professional (a.k.a. The Professional) (LÃ©on) (1994)	Action|Crime|Drama|Thriller
	6	Godfather, The (1972)	Crime|Drama
	7	Lord of the Rings: The Fellowship of the Ring, The (2001)	Adventure|Fantasy
	8	Indiana Jones and the Last Crusade (1989)	Action|Adventure
	9	Saving Private Ryan (1998)	Action|Drama|War
	10	Heat (1995)	Action|Crime|Thriller
11		...	

Lampiran 15 Contoh Hasil Rekomendasi Klaster 4

User	No	Title	Genres
6	1	Saint, The (1997)	Action|Romance|Sci-Fi|Thriller
	2	Player, The (1992)	Comedy|Crime|Drama
	3	Go (1999)	Comedy|Crime
	4	World Is Not Enough, The (1999)	Action|Adventure|Thriller
	5	13th Warrior, The (1999)	Action|Adventure|Fantasy
	6	Grand Day Out with Wallace and Gromit, A (1989)	Adventure|Animation|Childr en|Comedy|Sci-Fi
	7	Mystery Men (1999)	Action|Comedy|Fantasy
	8	Carrie (1976)	Drama|Fantasy|Horror|Thrill er
	9	American Werewolf in London, An (1981)	Comedy|Horror|Thriller
	10	Matrix, The (1999)	Action|Sci-Fi|Thriller
93		...	

Lampiran 16 Contoh Hasil Rekomendasi Klaster 5

User	No	Title	Genres
7	1	The Imitation Game (2014)	Drama|Thriller|War
	2	WALLÂ·E (2008)	Adventure|Animation|Children|Romance|Sci-Fi
	3	Graduate, The (1967)	Comedy|Drama|Romance
	4	Shawshank Redemption, The (1994)	Crime|Drama
	5	Interstellar (2014)	Sci-Fi|IMAX
	6	Dark Knight, The (2008)	Action|Crime|Drama|IMAX
	7	Rumble in the Bronx (Hont faan kui) (1995)	Action|Adventure|Comedy|Crime
	8	Shakespeare in Love (1998)	Comedy|Drama|Romance
	9	Star Trek: First Contact (1996)	Action|Adventure|Sci-Fi|Thriller
	10	Wolf of Wall Street, The (2013)	Comedy|Crime|Drama
48		...	

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS,

Nama : Anadia Rahmat Syihab Hidayatullah
NRP : 062115 4000 0001

menyatakan bahwa data yang digunakan dalam Tugas Akhir ini merupakan data sekunder yang diambil dari ~~penelitian / buku / Tugas Akhir / Thesis / Publikasi /~~ lainnya yaitu :

Sumber : Web GroupLens Research Project
Keterangan : Data *movielens* diambil dari laman (http://movielens.umn.edu) pada 20 April 2019

Surat pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Surabaya, Mei 2019

Mengetahui,
Pembimbing Tugas Akhir

Dr. Kartika Fithriasari, M.Si.
NIP. 19691212 199303 2 002

Mahasiswa

Anadia Rahmat Syihab Hidayatullah
NRP. 062115 4000 0001

85

(Halaman ini sengaja dikosongkan)

BIODATA PENULIS

Penulis dengan nama lengkap Anadia Rahmat Syihab Hidayatullah yang akrab disapa Syihab ini, dilahirkan di Lamongan pada 13 Nopember 1996. Penulis menempuh pendidikan formal di SD Negeri Kutorejo 1 Tuban, dilanjutkan menempuh pendidikan di SMP Negeri 3 Tuban, dan di SMA Negeri 1 Tuban. Kemudian penulis diterima sebagai Mahasiswa Departemen Statistika, Fakultas Matematika, Komputasi, dan Sains Data, Institut Teknologi Sepuluh Nopember di Surabaya melalui jalur SNMPTN pada tahun 2015. Selama masa perkuliahan, penulis aktif dalam berbagai organisasi kampus seperti Dewan Perwakilan Angkatan 2016-2019. Departemen Badan Pelaksana Mentoring sebagai Staff pada tahun kepengurusan 2016/2017, sebagai Kepala Departemen Kaderisasi pada tahun kepengurusan 2017/2018. Selain itu, penulis juga aktif mengikuti berbagai kompetisi Statistika Nasional Olimpiade Statistika Nasional. Penulis sangat terbuka akan kritik dan saran terkait hasil Laporan Tugas Akhir ini dengan menghubungi penulis melalui surel atau email anadiarahmat11@gmail.com.