

SKRIPSI

**SISTEM REKOMENDASI UNTUK DATA IMPLISIT DENGAN
DISTRIBUSI NORMAL MULTIVARIAT**



Jordan Alexander

NPM: 6181901020

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2024**

UNDERGRADUATE THESIS

**RECOMMENDATION SYSTEM FOR IMPLICIT DATA WITH
MULTIVARIATE NORMAL DISTRIBUTION**



Jordan Alexander

NPM: 6181901020

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2024**

LEMBAR PENGESAHAN

SISTEM REKOMENDASI UNTUK DATA IMPLISIT DENGAN DISTRIBUSI NORMAL MULTIVARIAT

Jordan Alexander

NPM: 6181901020

Bandung, 25 Juni 2024

Menyetujui,

Pembimbing

Digitally signed
by Luciana
Abednego

Luciana Abednego, M.T.

Ketua Tim Penguji

Digitally signed
by Mariskha Tri
Adithia

Mariskha Tri Adithia, P.D.Eng

Anggota Tim Penguji

Digitally signed
by Elisati Hulu

Elisati Hulu, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed
by Lionov

Lionov, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

SISTEM REKOMENDASI UNTUK DATA IMPLISIT DENGAN DISTRIBUSI NORMAL MULTIVARIAT

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 25 Juni 2024



Jordan Alexander
NPM: 6181901020

ABSTRAK

Sistem rekomendasi adalah suatu sistem yang menyediakan rekomendasi kepada pengguna untuk mendapatkan produk yang diinginkannya. Sistem rekomendasi berguna untuk meningkatkan pengalaman pengguna dalam mencari produk atau sesuatu yang menarik atau yang sesuai dengan minat dan kesukaan masing-masing pengguna. Sistem rekomendasi ini bertujuan agar dapat meningkatkan penjualan produk dan meningkatnya pengalaman pengguna dalam mencari produk yang sesuai dengan minatnya.

Pada penelitian ini akan digunakan metode *collaborative filtering* yang memberi rekomendasi film yang berasal dari interaksi pengguna lainnya dengan film-film yang mirip berdasarkan interaksinya dengan film. Data yang bisa didapatkan dari interaksi pengguna dengan produk atau aplikasi dapat diklasifikasikan menjadi 2, yaitu eksplisit dan implisit. Data implisit merupakan data yang tidak diinput secara langsung atau sadar oleh pengguna seperti melihat produk, klik *item* produk, histori pembelian, dan lain-lain. Pada penelitian ini data implisit yang digunakan berasal dari MovieLens. Data implisit ini merupakan data interaksi pengguna dengan film. Interaksi ini berasal dari *rating* yang diberikan pengguna pada film-film. Apabila pengguna memberikan *rating* dengan suatu film, maka nilai interaksinya 1. Jika tidak, maka akan diberi nilai interaksi 0. Maka nantinya data yang dipakai akan berbentuk sebuah matriks interaksi yang dimana kolomnya merupakan *movieId* dan barisnya merupakan *userId*.

Interaksi antara pengguna-pengguna dengan *item* memiliki nilai interaksi yang berdistribusi normal multivariat. Nantinya dengan menerapkan probabilitas sebaran distribusi normal multivariat pada interaksi pengguna dan *item*, pada penelitian ini akan memprediksi interaksi yang bernilai 0 atau pengguna yang tidak memberikan *rating* film, sehingga film tersebut yang tidak diberi *rating* akan diprediksi nilainya dengan distribusi normal multivariat.

Hasil dari prediksi yang dilakukan akan dievaluasi dengan menggunakan metrik evaluasi $\text{precision}@k$ dan $\text{nDCG}@k$. $\text{Precision}@k$ merupakan metrik evaluasi yang digunakan untuk mengukur seberapa tepat sebuah sistem rekomendasi dalam memprediksi item yang akan diinteraksi oleh pengguna. Sedangkan $\text{nDCG}@k$ mirip seperti $\text{precision}@k$, tetapi $\text{nDCG}@k$ melihat urutan hasil prediksi pada *list* rekomendasi yang dihasilkan dari prediksi. Pada pemodelan yang dibuat pada penelitian ini, pemodelan dibuat dengan pembagian data *train* dan data *test* menggunakan *train test split* dan *KFold*. Pada percobaan, digunakan nilai rasio penghilangan interaksi untuk data *test* sebesar 0,3, 0,5, dan 0,7. Rasio penghilangan interaksi mengubah beberapa interaksi yang nilainya 1 menjadi 0 pada setiap pengguna yang dijadikan data *test* sebanyak rasio tersebut. Nilai $\text{precision}@k$ dan $\text{nDCG}@k$ dengan menggunakan ketiga rasio tersebut didapat berkisar nilai 0,3 hingga 0,5.

Dari hasil evaluasi yang didapat, hasil prediksi kurang baik karena nilai $\text{precision}@k$ dan $\text{nDCG}@k$ berkisar dibawah 0,5, namun dengan hasil yang kurang baik tersebut masih bisa digunakan untuk rekomendasi kepada pengguna karena pengguna akan terbuka untuk mengeksplorasi berbagai pilihan film, meskipun rekomendasinya tidak selalu sangat presisi.

Kata-kata kunci: Sistem rekomendasi, data implisit, interaksi, distribusi normal multivariat

ABSTRACT

A recommendation system is a system that provides recommendations to users to find the products they want. A recommendation system is useful for enhancing the user experience in finding products or something interesting or suitable for each user's interests and preferences. The purpose of this recommendation system is to increase product sales and improve the user experience in finding products that match their interests.

In this research, a collaborative filtering method will be used to recommend movies based on the interactions of other users with similar movies. Data obtained from user interactions with products or applications can be classified into two types: explicit and implicit. Implicit data refers to data that is not directly or consciously inputted by the user, such as viewing products, clicking on product items, purchase history, etc. In this study, the implicit data used comes from MovieLens. This implicit data represents user interactions with movies. These interactions are derived from the ratings users give to movies. If a user rates a movie, the interaction value is 1. If not, the interaction value is 0. Therefore, the data used will be in the form of an interaction matrix, with columns representing movieId and rows representing userId.

The interactions between users and items have interaction values that are multivariate normally distributed. By applying the probability of multivariate normal distribution to user and item interactions, this study will predict interactions with a value of 0, or users who do not rate movies. Thus, the unrated movies will be predicted using the multivariate normal distribution.

The prediction results will be evaluated using the precision@k and nDCG@k evaluation metrics. Precision@k is an evaluation metric used to measure how accurately a recommendation system predicts items that will be interacted with by users. Meanwhile, nDCG@k is similar to precision@k, but nDCG@k considers the order of the predicted results in the recommendation list generated from the prediction. In this study's modeling, the data is split into training and test sets using train-test split and KFold. The experiments use interaction removal ratios of 0.3, 0.5, and 0.7 for the test data. The interaction removal ratio converts some interactions with a value of 1 to 0 for each user in the test data according to the specified ratio. The precision@k and nDCG@k values using these three ratios range from 0.3 to 0.5.

From the evaluation results, the prediction results are not very good since the precision@k and nDCG@k values are below 0.5. However, despite the less optimal results, the system can still be used to recommend movies to users, as users are often open to exploring various movie options, even if the recommendations are not always highly precise

Keywords: Recommendation system, implicit data, interaction, multivariate normal distribution

Tugas akhir dipersembahkan kepada diri sendiri dan orang tua

KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus atas berkat dan penyertaan-Nya, penulis dapat menyelesaikan tugas akhir yang berjudul "Sistem rekomendasi, data implisit, interaksi, distribusi normal multivariat". Penulis menyadari bahwa dalam proses penyusunan skripsi ini, banyak bantuan dari teman-teman dan kerabat. Oleh karena itu, penulis ingin memberi ucapan terima kasih sebesar-besarnya kepada:

1. Keluarga penulis yaitu mama dan kakak penulis yang selalu memberikan dukungan selama menjalani perkuliahan, pembuatan tugas akhir, hingga sidang akhir, serta yang tak henti-hentinya mendoakan kelancaran studi penulis.
2. Ibu Natalia, selaku dosen pembimbing, yang telah memberikan banyak masukan, bimbingan, dan nasihat selama proses penyusunan tugas akhir ini. Ibu Mariskha Tri Adithia dan Bapak Elisati Hulu, selaku dosen penguji, yang telah memberikan saran dan kritik dalam penulisan tugas akhir ini.
3. Seluruh dosen Informatika Universitas Katolik Parahyangan yang telah memberikan ilmu kepada penulis selama perkuliahan.
4. Teman-teman dan rekan-rekan seperjuangan di Informatika Unpar, yang selalu memberikan dukungan moral dan juga berbagi pengetahuan dan pengalaman kepada penulis selama perkuliahan dan pembuatan tugas akhir.
5. Semua pihak yang telah memberikan dukungannya dan yang tidak disebutkan, baik secara langsung maupun tidak langsung.

Semoga segala bantuan dan dukungan yang telah diberikan selama ini senantiasa dibalas oleh Tuhan. Penulis menyadari bahwa tugas akhir ini masih memiliki banyak kekurangan dan jauh dari kata sempurna karena keterbatasan ilmu yang dimiliki. Oleh karena itu, penulis sangat mengharapkan saran dan kritik yang membangun untuk kepentingan pengembangan tugas akhir ini. Semoga tugas akhir ini dapat bermanfaat bagi yang membutuhkan.

Bandung, Juni 2024

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
DAFTAR TABEL	xxi
DAFTAR KODE PROGRAM	xxiii
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Batasan Masalah	3
1.5 Metodologi	3
1.6 Sistematika Pembahasan	4
DAFTAR NOTASI	1
2 LANDASAN TEORI	5
2.1 Sistem Rekomendasi[1]	5
2.2 Distribusi Normal[2]	6
2.2.1 Distribusi Normal Multivariat[3]	8
2.2.2 Distribusi Normal Multivariat pada Sistem Rekomendasi[4]	9
2.3 Metode Pengujian Model[5]	10
2.4 Evaluasi Model[4]	11
2.4.1 <i>Precision@k</i>	11
2.4.2 <i>The normalized Discounted Cumulative Gain at k (nDCG@k)</i>	12
3 ANALISIS PENYELESAIAN MASALAH	13
3.1 Penyelesaian Masalah	13
3.2 Contoh Perhitungan Penerapan Distribusi Normal Multivariat pada Sistem Rekomendasi	14
3.2.1 Menghitung Matriks Rata-rata ($\bar{\mu}_j$)	14
3.2.2 Menghitung Matriks Kovarian	15
3.2.3 Menghitung Prediksi	16
3.2.4 Menghitung Evaluasi Model	17
4 PEMODELAN SISTEM REKOMENDASI	23
4.1 Penyiapan Data	23
4.2 Pembuatan Model	24
4.3 Hasil Pemodelan dan Evaluasi Model	27
4.3.1 Evaluasi dengan <i>Train Test Split</i>	27

4.3.2	Evaluasi dengan Kfold	32
5	PEMBANGUNAN PERANGKAT LUNAK DAN PENGUJIAN	37
5.1	Deskripsi Perangkat Lunak	37
5.1.1	Tampilan Antar Muka Perangkat Lunak	37
5.2	Pengujian Perangkat Lunak	39
6	KESIMPULAN DAN SARAN	45
6.1	Kesimpulan	45
6.2	Saran	45
	DAFTAR PUSTAKA	47
	A KODE PROGRAM PEMODELAN	49
	B KODE WEBSITE	53

DAFTAR GAMBAR

1.1	Rekomendasi lagu pada aplikasi Spotify	1
1.2	Ilustrasi perbandingan rekomendasi dengan metode <i>collaborative filtering</i> (kiri) dan <i>content-based filtering</i> (kanan)	2
2.1	Kurva lonceng bentuk sebaran normal random variabel	7
2.2	Peran μ dan σ pada grafik distribusi normal	8
2.3	Bentuk <i>skewness</i> positif dan negatif	8
2.4	Visualisasi distribusi normal bivariat	9
2.5	Bentuk 3D dari distribusi normal bivariat	10
2.6	Ilustrasi <i>cross validation</i> dengan $k = 5$	11
4.1	Contoh <i>dataframe</i> implisit kosong	24
4.2	Contoh <i>dataframe</i> implisit yang sudah diisi	24
4.3	Ilustrasi Data <i>Train</i>	27
4.4	Ilustrasi Data <i>Test</i>	27
4.5	Nilai $\text{precision}@k$ untuk setiap s	31
4.6	Nilai $\text{nDCG}@k$ untuk setiap s	31
4.7	Nilai $\text{precision}@k$ $\text{KFold} = 5$	33
4.8	Nilai $\text{nDCG}@k$ $\text{KFold} = 5$	33
4.9	Nilai perbandingan $\text{precision}@k$ KFold dan <i>train test split</i>	34
4.10	Nilai perbandingan $\text{nDCG}@k$ KFold dan <i>train test split</i>	34
5.1	Tampilan halaman awal	37
5.2	Tampilan <i>dropdown</i> <i>userId</i> ketika dibuka	38
5.3	Tampilan <i>dropdown</i> banyaknya rekomendasi ketika dibuka	38
5.4	Tampilan <i>loading</i> saat kedua <i>dropdown</i> dipilih	39
5.5	Tampilan halaman <i>website</i> dengan memilih <i>userId</i> 1 dan banyak rekomendasi 3	39
5.6	Tampilan <i>output</i> dari kode Python untuk <i>userId</i> 1 dan banyak rekomendasi 3	40
5.7	Tampilan halaman <i>website</i> dengan memilih <i>userId</i> 1 dan banyak rekomendasi 10	40
5.8	Tampilan <i>output</i> dari kode Python untuk <i>userId</i> 1 dan banyak rekomendasi 10	41
5.9	Tampilan halaman <i>website</i> dengan memilih <i>userId</i> 1 dan banyak rekomendasi 20	41
5.10	Tampilan <i>output</i> dari kode Python untuk <i>userId</i> 1 dan banyak rekomendasi 20	42
5.11	Tampilan halaman <i>website</i> dengan memilih <i>userId</i> 3055 dan banyak rekomendasi 10	42
5.12	Tampilan <i>output</i> dari kode Python untuk <i>userId</i> 3055 dan banyak rekomendasi 10	43

DAFTAR TABEL

3.1	Data implisit interaksi pengguna dengan <i>item</i>	14
3.2	Contoh data implisit	17
3.3	Contoh data implisit setelah dihilangkan interaksi sebagian	18
3.4	Contoh hasil prediksi	18
3.5	<i>Recommendation list</i> setiap pengguna	19
4.1	Matriks kovarian data contoh dengan perhitungan manual	29
4.2	Hasil prediksi menggunakan perhitungan manual	29
4.3	Matriks kovarian data contoh dengan perhitungan Python	29
4.4	Hasil prediksi menggunakan kode Python	29
4.5	Tabel nilai precision@k dan nDCG@k untuk 1000 pengguna dengan interaksi terbanyak	30
4.6	Tabel nilai precision@k dan nDCG@k untuk setiap s (rasio interaksi yang dihilangkan)	31
4.7	Tabel nilai precision@k dan nDCG@k untuk setiap <i>fold</i>	33
4.8	Tabel nilai rata-rata precision@k dan nDCG@k untuk Kfold = 5	34

DAFTAR KODE PROGRAM

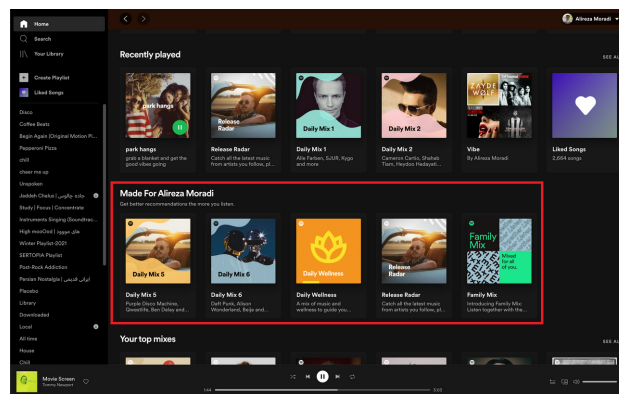
4.1	Kode untuk membuat dataframe kosong untuk data implisit	23
4.2	Kode untuk mengisi dataframe implisit	24
4.3	Kode untuk matriks rata-rata	24
4.4	Kode untuk membagi data menjadi data <i>train</i> dan data <i>test</i>	25
4.5	Kode untuk membuat matriks kovarian	25
4.6	Kode untuk melakukan prediksi atau mengisi nilai interaksi yang kosong	26
4.7	Kode untuk mengecek singularitas matriks	27
4.8	Kode untuk evaluasi model dengan precision@k	28
4.9	Kode untuk evaluasi model dengan precision@k	28
4.10	Kode untuk evaluasi model dengan nDCG@k	28
4.11	Kode untuk membuat data <i>train</i> dan data <i>test</i> untuk Kfold	32
5.1	Kode untuk menampilkan judul film pada userId 1 dan sebanyak 3 film	39
5.2	Kode untuk menampilkan judul film pada userId 1 dan sebanyak 10 film	40
5.3	Kode untuk menampilkan judul film pada userId 1 dan sebanyak 20 film	41
5.4	Kode untuk menampilkan judul film pada userId 3055 dan sebanyak 10 film	43
A.1	Kode untuk <i>import library</i> yang dibutuhkan	49
A.2	Kode untuk load semua data mentahdari Google Drive	49
A.3	Kode untuk membuat data implisit interaksi	49
A.4	Kode untuk membuat train test split	49
A.5	Kode untuk KFold	50
A.6	Kode untuk membuat matriks kovarian	50
A.7	Kode untuk cek matriks singular	50
A.8	Kode untuk memprediksi	50
A.9	Kode untuk membuat recommendation lists	51
A.10	Kode untuk menghitung precision@k	51
A.11	Kode untuk menghitung nDCG@k	51
A.12	Kode untuk plot precision KFold	52
A.13	Kode untuk nDCG@k Kfold	52
A.14	Kode untuk plot perbandingan precision@k Kfold dan train test split	52
A.15	Kode untuk perbandingan nDCG@K	52
A.16	Kode untuk perbandingan precision@k train test split untuk setiap s	52
A.17	Kode untuk perbandingan nDCG@k train test split untuk setiap s	52
B.1	Kode untuk index.php	53
B.2	Kode untuk script.js	54
B.3	Kode untuk getData.php	54
B.4	Kode untuk kode styles.css	55

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Sistem rekomendasi adalah suatu sistem yang menyediakan rekomendasi kepada pengguna untuk mendapatkan produk yang diinginkannya. Sistem rekomendasi berguna untuk meningkatkan pengalaman pengguna dalam mencari produk atau sesuatu yang menarik atau yang sesuai dengan minat dan kesukaan masing-masing pengguna. Sistem rekomendasi banyak dipakai dalam aplikasi film, buku, musik, produk, dan lain-lain. Sistem rekomendasi ini bertujuan agar dapat meningkatkan penjualan dan juga dalam meningkatkan *user engagement* atau tingkat perhatian pengguna terhadap produk atau *item* dengan berinteraksi atau menghabiskan waktu dengan produk atau *item*. Dengan meningkatnya pengalaman pengguna dalam mencari produk yang sesuai dengan minatnya, hal tersebut dapat membuat peningkatan penjualan dari sebuah perusahaan. Gambar 1.1 merupakan contoh dari ilustrasi sistem rekomendasi yang diberikan aplikasi untuk memberikan rekomendasi lagu yang sesuai dengan pengguna. Pada bagian tengah tampilan aplikasi spotify, terdapat bagian yang diberi kotak merah pada Gambar 1.1 yang bertuliskan "Made for Alireza Moradi" yang terdapat beberapa *playlist* lagu yang direkomendasikan untuk pengguna yaitu Alireza Moradi. Rekomendasi yang diberikan Spotify ini diharapkan sesuai dengan selera lagu yang disukai Alireza Moradi.

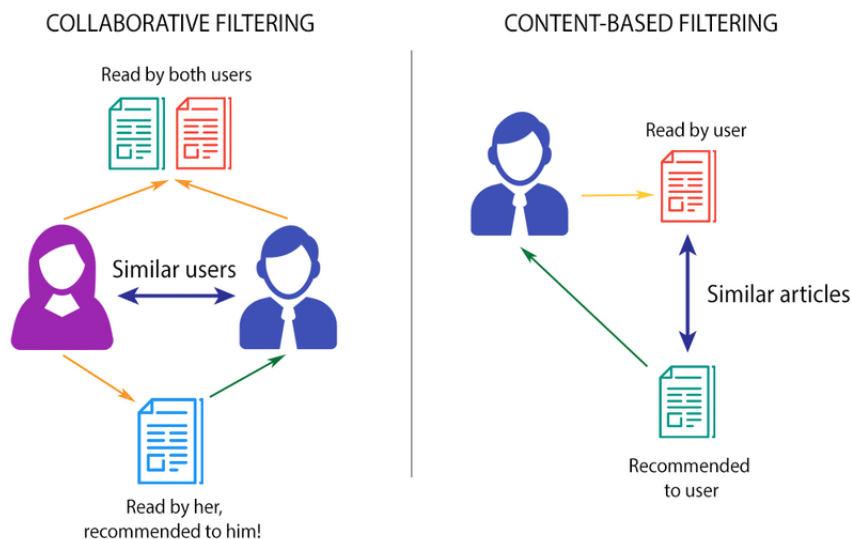


Gambar 1.1: Rekomendasi lagu pada aplikasi Spotify¹

Terdapat beberapa metode populer yang digunakan dalam sistem rekomendasi, dua diantaranya adalah *collaborative filtering* dan *content-based filtering* yang divisualisasikan pada Gambar 1.2. Pada *collaborative filtering*, rekomendasi akan berasal dari interaksi pengguna lainnya yang mirip berdasarkan interaksinya dengan produk atau *item* di aplikasi. Misalkan Budi dan Ani menyukai lagu Bruno Mars sehingga mereka berdua menyukai lagu yang mirip. Lalu Budi mendengarkan lagu Michael Jackson dan menyukainya, maka Ani akan diberikan rekomendasi lagu Michael Jackson juga karena Budi menyukainya. Lalu metode yang kedua adalah *content-based filtering* yang merupakan metode pada sistem rekomendasi yang memanfaatkan informasi pengguna sebelumnya

¹<https://community.spotify.com/t5/Desktop-Windows/Show-all-Made-for-you-Playlists/td-p/5198663?lightbox-message-images-5254579=132711i2ED11098D21D23FC>

untuk nantinya merekomendasikannya kepada pengguna. Misalkan Budi menyukai lagu bergenre pop, maka nanti sistem rekomendasi akan merekomendasikan lagu-lagu yang bergenre pop. Pada penelitian ini akan digunakan metode *collaborative filtering* karena menggunakan pengenalan kesamaan interaksi pengguna dengan *item*. Data yang bisa didapatkan dari interaksi pengguna dengan produk atau aplikasi dapat diklasifikasikan menjadi 2, yaitu eksplisit dan implisit.



Gambar 1.2: Ilustrasi perbandingan rekomendasi dengan metode *collaborative filtering* (kiri) dan *content-based filtering* (kanan)³

Pada metode *collaborative filtering*, terdapat beberapa model yang dapat digunakan, tetapi pada penelitian kali ini akan digunakan pemodelan dengan menggunakan algoritma distribusi normal multivariat. Distribusi normal multivariat ini adalah distribusi normal yang memiliki 2 atau lebih variabel. Variabel berdistribusi normal artinya variabel tersebut memiliki penyebaran dengan probabilitas fungsi normal. Variabel yang berdistribusi normal akan memiliki sebaran yang berbentuk seperti lonceng yang berpusat pada tengah atau berpuncak di tengah. Algoritma distribusi normal multivariat ini merupakan algoritma yang simpel dan memiliki waktu training yang cepat. Nantinya dengan algoritma distribusi normal multivariat akan mengisi ketidakadaan interaksi pengguna dengan item. Data akan dibagi menjadi data *training* dan data *testing* seperti pemodelan pada umumnya dan akan dihitung akurasi dengan metrik-metrik akurasi model, seperti $\text{precision}@k$ dan $\text{nDCG}@k$. Hasil pemodelan dari algoritma distribusi normal multivariat ini yang akan dipakai dalam memberikan rekomendasi dalam sebuah sistem rekomendasi[4].

Data yang didapatkan dari pengguna dibagi menjadi 2 jenis, yaitu data eksplisit dan data implisit. Data eksplisit merupakan data yang diinput langsung oleh pengguna seperti *rating* dan *like/dislike*. Data eksplisit biasanya mengandung *missing value* karena pengguna bisa saja tidak melakukan interaksi seperti tidak memberikan *rating* atau *like/dislike* sehingga datanya tidak berisi *value*. Sedangkan data implisit merupakan data yang tidak diinput secara langsung atau sadar oleh pengguna. Contohnya adalah melihat produk, klik *item* produk, histori pembelian, dan lain-lain. Kedua jenis data ini yang nantinya akan ditemukan ketika dalam pembuatan sistem rekomendasi.

Pada penelitian ini akan dibuat sebuah sistem rekomendasi film, dimana nantinya pengguna akan disajikan rekomendasi yang banyaknya sesuai keinginan pengguna. Rekomendasi yang diberikan menggunakan metode *collaborative filtering* dengan menggunakan algoritma distribusi normal multivariat. Pertama-tama akan dilakukan pemrosesan data supaya dapat dilakukan pemodelan. Lalu yang kedua akan dilakukan pemodelan untuk memprediksi item atau film mana yang akan

³https://www.researchgate.net/figure/Content-based-filtering-vs-Collaborative-filtering-Source_fig5_323726564

direkomendasikan kepada masing-masing pengguna. Lalu yang terakhir akan dilakukan evaluasi model yang telah dibuat dengan menggunakan metrik-metrik evaluasi. Pada penelitian ini digunakan 2 metrik evaluasi, yaitu $\text{precision}@k$ dan $\text{nDCG}@k$. $\text{Precision}@k$ merupakan metrik evaluasi untuk mengukur seberapa tepat prediksi rekomendasi yang diberikan pada *top k* rekomendasi atau sebanyak k rekomendasi yang diberikan. Lalu $\text{nDCG}@k$ merupakan metrik evaluasi yang melihat posisi peringkat prediksi yang tepat pada *top k* rekomendasi atau sebanyak k rekomendasi yang diberikan.

Pada penelitian ini digunakan 1 dataset yang berasal dari MovieLens. MovieLens merupakan situs penelitian yang dikembangkan oleh GroupLens Research di University of Minnesota. MovieLens memberikan layanan sistem rekomendasi dengan menggunakan teknologi *collaborative filtering* dalam membuat rekomendasi film. Rekomendasi yang diberikan oleh pengguna berdasarkan dari *rating* film yang diberikan pengguna⁴.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dibentuk rumusan masalah sebagai berikut:

1. Bagaimana memanfaatkan distribusi normal multivariat untuk sistem rekomendasi?
2. Bagaimana cara mengimplemetasikan algoritma distribusi normal multivariat pada sistem rekomendasi?
3. Bagaimana hasil prediksi dari rekomendasi yang dihasilkan dari algoritma distribusi normal multivariat?

1.3 Tujuan

Berdasarkan rumusan masalah di atas, tujuan dari penelitian adalah sebagai berikut:

1. Mempelajari cara kerja algoritma distribusi normal multivariat.
2. Membangun perangkat lunak yang mengimplementasikan metode distribusi normal multivariat.
3. Mengevaluasi model yang telah dibuat dengan metrik evaluasi *precision at k* ($\text{precision}@k$) dan *The normalized Discounted Cumulative Gain at k* ($\text{nDCG}@k$)

1.4 Batasan Masalah

Batasan masalah pada skripsi ini adalah sebagai berikut:

- Model hanya berlaku untuk dataset Movie Lens

1.5 Metodologi

1. Melakukan studi literatur mengenai sistem rekomendasi.
2. Melakukan studi literatur mengenai metode distribusi normal multivariat.
3. Melakukan pencarian data yang cocok untuk penelitian.
4. Melakukan eksplorasi dan analisis terhadap data.
5. Melakukan eksplorasi dan studi literatur mengenai evaluasi model yang cocok.
6. Melakukan perancangan perangkat lunak yang akan dibangun.
7. Membangun perangkat lunak untuk menampilkan hasil sistem rekomendasi menggunakan metode distribusi normal multivariat.
8. Melakukan pengujian fungsional dan eksperimental perangkat lunak.
9. Melakukan analisis hasil pengujian dan mengambil kesimpulan.

⁴<https://movielens.org/info/about>

1.6 Sistematika Pembahasan

Penelitian ini tersusun ke dalam 6 bab secara sistematis sebagai berikut:

- Bab 1 Pendahuluan
Membahas latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
- Bab 2 Landasan Teori
Membahas hasil studi literatur mengenai dasar teori yang digunakan pada penelitian ini untuk menjawab rumusan dan tujuan masalah seperti penjelasan mengenai distribusi normal multivariat.
- Bab 3 Analisis Penyelesaian Masalah
Membahas analisis masalah yang ditemukan dalam proses mencapai hasil penelitian, membahas eksplorasi terkait *tools* yang digunakan pada penelitian ini, serta membahas penggunaan dan cara kerja *tools* yang digunakan seperti *library* yang digunakan, membahas eksplorasi terkait cara kerja algoritma distribusi normal multivariat, serta perhitungan manual menggunakan data berskala kecil.
- Bab 4 Pemodelan Sistem Rekomendasi
Membahas pengumpulan data dan transformasi data. Membahas proses analisis data yang dilakukan untuk menjawab tujuan penelitian, membahas hasil analisis data yang telah dilakukan pada penelitian ini, dan membahas interpretasi hasil analisis terkait informasi apa saja yang dapat ditemukan dari penelitian yang telah dilakukan.
- Bab 5 Pembangunan Perangkat Lunak dan Pengujian
Membahas perancangan dan pembuatan perangkat lunak dan pengujian perangkat lunak.
- Bab 6 Kesimpulan dan Saran
Membahas kesimpulan penelitian dan saran untuk penelitian berikutnya.