

**IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBORS*
PADA SISTEM REKOMENDASI BUKU MENGGUNAKAN
TEKNIK *COLLABORATIVE FILTERING***

SKRIPSI

OLEH:

**KORI ISABELLA HUTABARAT
188160022**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2022**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 13/12/22

Access From (repository.uma.ac.id)13/12/22

**IMPLEMENTASI ALGORITMA *K-NEAREST NEIGHBORS*
PADA SISTEM REKOMENDASI BUKU MENGGUNAKAN
TEKNIK *COLLABORATIVE FILTERING***

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana di Fakultas Teknik
Universitas Medan Area



Oleh:

**KORI ISABELLA HUTABARAT
188160022**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2022**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

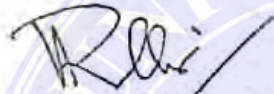
1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 13/12/22

Access From (repository.uma.ac.id)13/12/22

Judul Skripsi : Implementasi Algoritma *K-Nearest Neighbors* pada Sistem
Rekomendasi Buku Menggunakan Teknik *Collaborative
Filtering*
Nama : Kori Isabella Hutabarat
NPM : 188160022
Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing



Rizki Muliono, S.Kom., M. Kom.
Pembimbing I



Zulfikar Sembiring, S.Kom., M. Kom.
Pembimbing II



Dr. Rahman Syah, S.Kom., M. Kom.
Dekan Fakultas Teknik



Rizki Muliono S.Kom., M. Kom.
Ka. Prodi

Tanggal Lulus : 31 Agustus 2022

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana di Universitas Medan Area merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 31 Agustus 2022



Kori Isabella Hutabarat

188160022

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TEISIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Kori Isabella Hutabarat
NPM : 188160022
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis karya : Skripsi

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Implementasi Algoritma *K-Nearest Neighbors* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering*

berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada tanggal : 31 Agustus 2022

Yang menyatakan



(Kori Isabella Hutabarat)

ABSTRACT

A recommendation system is a system or technique that recommends a particular product, service, or entity that has been widely used. In the industrial sector, recommendation systems are an important tool to improve user experience and promote sales or services. The appliance recommends specific items instead a number of items that may match the user's preferences as "top-N" recommendations. In this study investigated the item-based data, the collaborative filtering technique was employed to provide recommendations based on ratings. Moreover, this study also supported the K-Nearest Neighbors as one of the algorithms to present the recommendations in a system. Mean Absolute Error (MAE) is used as the parameter to measure error, while the data used in this research is book data as a case study. The results of this test produce the smallest MAE value at $K = 26$, as much as 0.42. The research also pointed out that the value of K will affect the execution time to provide recommendations. This model is successful in providing book recommendations to users. This proves that the implementation of the K-Nearest Neighbors Algorithm on a Book Recommendation System Using Collaborative Filtering Techniques has succeeded in providing book recommendations to users with low errors. Henceforth, this recommendation system can be used in other case studies, either primary data or secondary data which have a rating attribute in the dataset.

Keywords: Recommendation System, K-Nearest Neighbors, Collaborative Filtering, Mean Absolute Error (MAE), Books.

ABSTRAK

Sistem rekomendasi merupakan suatu sistem atau teknik yang merekomendasikan produk, layanan, atau entitas tertentu. Sistem rekomendasi telah banyak digunakan oleh hampir setiap kalangan, misalnya pada industri sistem rekomendasi adalah alat penting untuk meningkatkan pengalaman pengguna dan mempromosikan penjualan atau layanan. Sistem rekomendasi tidak merekomendasikan item secara spesifik, namun merekomendasikan sejumlah item yang mungkin cocok dengan preferensi pengguna berupa “top-N” *recommendation*. Teknik *Collaborative Filtering* adalah teknik yang memberikan rekomendasi berdasarkan rating, dan pada penelitian ini menggunakan item-based. Algoritma *K-Nearest Neighbors* merupakan salah satu algoritma yang dapat digunakan untuk memberikan rekomendasi pada sistem rekomendasi. Penelitian ini menggunakan *Mean Absolute Error* (MAE) sebagai parameter untuk mengukur error. Data yang digunakan pada penelitian ini adalah data buku sebagai studi kasus. Hasil pengujian ini menghasilkan nilai MAE terkecil pada $K = 26$ yaitu sebesar 0,42. Hasil pengujian ini juga menunjukkan bahwa nilai K akan mempengaruhi waktu eksekusi untuk memberikan rekomendasi. Model ini berhasil memberikan rekomendasi buku pada pengguna. Hal ini membuktikan bahwa Implementasi Algoritma *K-Nearest Neighbors* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering* berhasil memberikan rekomendasi buku kepada pengguna dengan *error* yang rendah. Untuk selanjutnya, sistem rekomendasi ini dapat digunakan pada studi kasus lainnya, baik menggunakan data primer maupun menggunakan data sekunder yang memiliki atribut rating pada datasetnya.

Kata kunci: Sistem Rekomendasi, *K-Nearest Neighbors*, *Collaborative Filtering*, *Mean Absolute Error* (MAE), Buku

RIWAYAT HIDUP

Penulis merupakan putri atau anak ke-3 (ketiga) dari ayah Marincon Hutabarat dan ibu Rita Wati Situmeang yang dilahirkan di Pandalingan, pada tanggal 17 Oktober 2000. Pada tahun 2018, Penulis lulus dari SMK Sw. HKBP Sibolga Jurusan Teknik Komputer dan Jaringan, lalu kemudian terdaftar sebagai mahasiswa Fakultas Teknik Prodi Teknik Informatika Universitas Medan Area.

Selama mengikuti perkuliahan, Penulis aktif mengikuti beberapa organisasi internal maupun eksternal kampus, juga berbagai kompetisi mulai dari tingkat lokal hingga tingkat internasional. Selanjutnya Penulis melaksanakan Praktek Kerja Lapangan (PKL) di SMK Sw. HKBP Sibolga.

Selain itu, Penulis mendapatkan kesempatan untuk mengikuti dan melaksanakan program magang di beberapa perusahaan, salah satunya perusahaan di Korea Selatan, serta selanjutnya di Gojek yang merupakan salah satu *Decacorn* di Indonesia.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “**Implementasi Algoritma *K-Nearest Neighbors* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering***” ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Program Strata-1 pada Fakultas Teknik Program Studi Teknik Informatika di Universitas Medan Area.

Dalam proses menyelesaikan skripsi ini, penulis menyadari bahwa tulisan ini masih jauh dari kata sempurna dan juga terdapat banyak kekurangan. Penulis juga mengharapkan kritik dan saran yang sifatnya membangun dari para pembaca. Kemudian penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
2. Bapak Dr. Rahmad Syah, S.Kom., M.Kom. selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Kepala Program Studi Teknik Informatika yang juga selaku dosen pembimbing I yang telah membimbing, mengarahkan, dan memberikan motivasi kepada penulis dalam proses menyelesaikan skripsi ini.
4. Bapak Zulfikar Sembiring, S.Kom., M.Kom., selaku pembimbing II yang telah meluangkan waktu untuk membimbing penulis dan memberikan ilmu yang bermanfaat sehingga skripsi ini bisa diselesaikan.

5. Bapak Andre Hasudungan Lubis, S.Ti., M.Sc. selaku Ketua Panitia Sidang Tugas Akhir penulis yang telah memberikan ilmu yang bermanfaat dan motivasi kepada penulis.
6. Ibu Susilawati, S.Kom., M.Kom. selaku Sekretaris Panitia Sidang Tugas Akhir penulis yang telah memberikan ilmu yang bermanfaat.
7. Seluruh Dosen Teknik Informatika Universitas Medan Area yang selama ini telah membekali penulis dengan ilmu yang sangat bermanfaat.
8. Seluruh Pegawai Universitas Medan Area yang telah membantu dalam proses administrasi.
9. Teristimewa kepada kedua orang tua penulis, Bapak Marincon Hutabarat dan Ibu Rita Wati Situmeang yang dengan penuh kasih sayang telah mendidik penulis serta dengan doa restunya penulis dapat menyelesaikan pendidikan hingga perguruan tinggi.
10. Kedua kakak penulis, Kak Ika dan Kak Tanti yang selalu mendukung dan mendoakan serta sangat banyak membantu penulis hingga berhasil menyelesaikan pendidikan saat ini, dan juga untuk adik-adik penulis yang selalu mendoakan yang terbaik untuk penulis.
11. Reymunda, Teh Kia, dan Teh Goo yang selalu memberikan semangat, dukungan, serta sangat banyak membantu penulis dalam berbagai kendala yang penulis hadapi.
12. Teman-teman Teknik Informatika Reguler 2018, terima kasih atas persahabatan dan persaudaraannya selama ini. Semoga Allah memudahkan untuk menyelesaikan study S-1 ini.

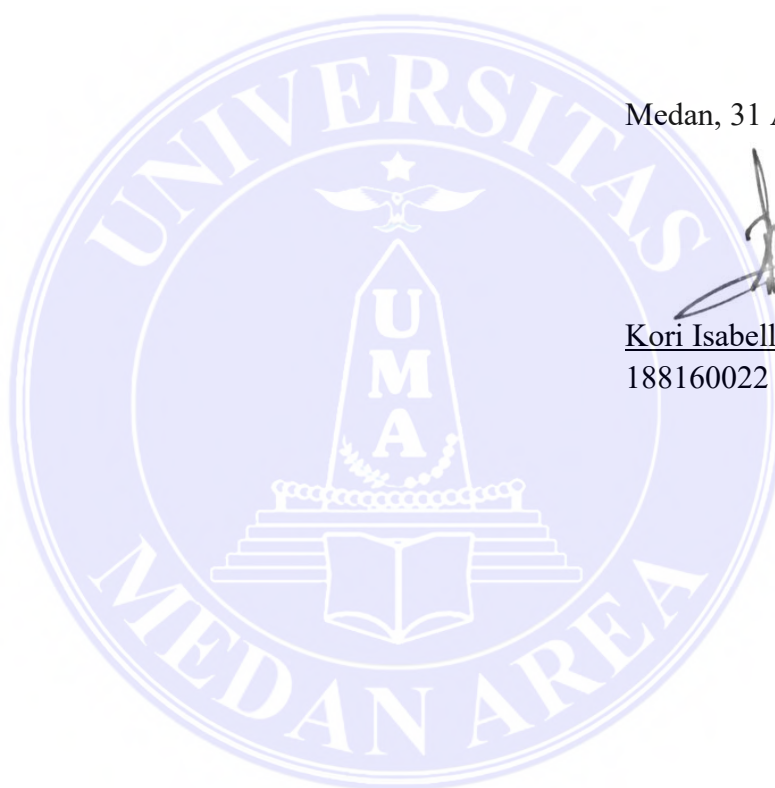
13. Kepada semua orang terdekat penulis, yang tidak terkait dalam penulisan skripsi ini tetapi memberikan dukungan dan motivasi untuk dapat menyelesaikan skripsi ini dengan baik.

Penulis menyadari dalam penyusunan skripsi ini, masih terdapat banyak kekurangan dan jauh dari kesempurnaan, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Semoga skripsi ini dapat berguna dan bermanfaat bagi kita semua.

Medan, 31 Agustus 2022



Kori Isabella Hutabarat
188160022



DAFTAR ISI

	Halaman
ABSTRACT.....	v
ABSTRAK.....	vi
RIWAYAT HIDUP.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	6
1.6 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI.....	9
2.1 Sistem Rekomendasi.....	9
2.1.1 <i>Content Based Filtering</i>	10
2.1.2 <i>Collaborative Filtering</i>	12
2.2 <i>K-Nearest Neighbors</i> (KNN).....	14
2.3 <i>Mean Absolute Error</i> (MAE).....	17
2.4 Buku.....	17
2.5 <i>Flowchart</i>	18
2.6 Penelitian Terdahulu.....	18
BAB III METODOLOGI PENELITIAN.....	23
3.1 Pengumpulan Data.....	23
3.1.1 Jenis dan Sumber Data.....	23
3.1.2 Metode Pengumpulan Data.....	25
3.2 Tahapan Penelitian.....	25
3.2.1 Kerangka Kerja.....	26
3.2.2 Kriteria Pengujian.....	30
3.2.3 Format Data Pengujian.....	30
3.2.4 Alat yang Digunakan.....	30
3.2.5 Pengujian Penelitian.....	31
3.2.6 Analisis Hasil Pengujian.....	31
3.3 Perhitungan Teknik <i>Item Based Collaborative Filtering</i> (IBCF)....	31
BAB IV HASIL DAN PEMBAHASAN.....	36
4.1 Hasil.....	36
4.1.1 Hasil Implementasi.....	41
4.2 Pembahasan.....	43
4.2.1 Mengekstrak Dataset.....	44
4.2.2 Eksplorasi Data.....	45

4.2.3 Prapemrosesan Data	47
4.2.4 Visualisasi Data	49
BAB V KESIMPULAN DAN SARAN	52
4.1 Kesimpulan	52
4.2 Saran	53
DAFTAR PUSTAKA	54
LAMPIRAN – LAMPIRAN	57

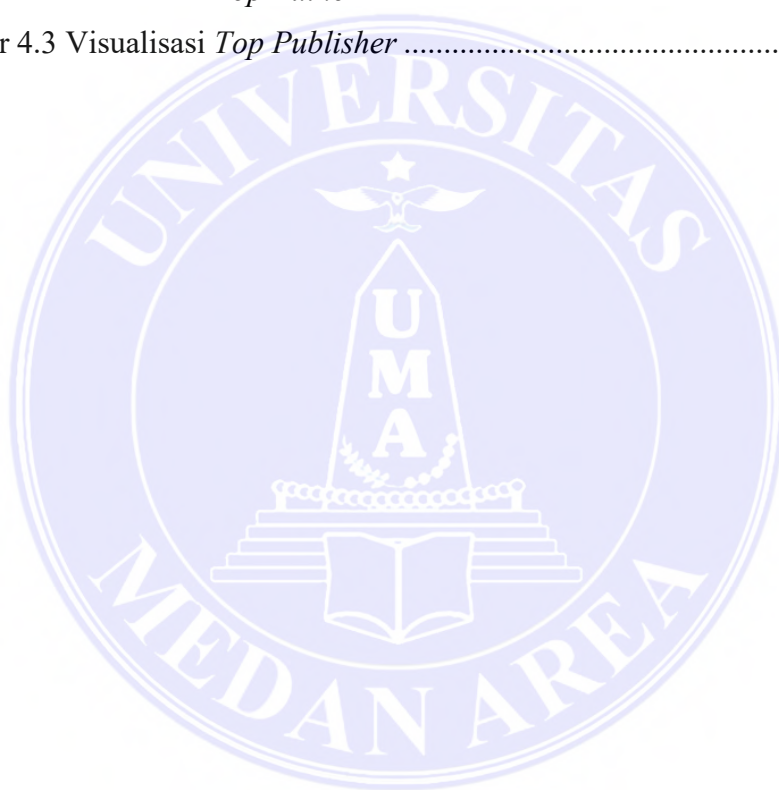


DAFTAR TABEL

	Halaman
Tabel 2.1 Simbol <i>Flowchart</i>	18
Tabel 2.2 Penelitian Terdahulu	22
Tabel 3.1 Data <i>Books.csv</i>	23
Tabel 3.2 Data <i>Ratings.csv</i>	24
Tabel 3.3 Data <i>Users.csv</i>	25
Tabel 3.4 Sampel Data – <i>Item Rating</i>	31
Tabel 3.5 Data <i>Rating</i> Perhitungan Teknik IBCF.....	32
Tabel 3.6 Data interseksi antara buku CM dan buku CC.....	33
Tabel 3.7 Matriks Kemiripan Buku	33
Tabel 3.8 Contoh Data Uji	34
Tabel 4.1 Hasil Implementasi	42
Tabel 4.2 Hasil MAE	43
Tabel 4.3 Informasi Dataset.....	43

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Konsep IBCF.....	14
Gambar 3.1 Tahapan Penelitian.....	26
Gambar 3.2 Kerangka Kerja Sistem Rekomendasi Buku Menggunakan Algoritma <i>K-Nearest Neighbors</i> dengan Teknik <i>Collaborative Filtering</i>	26
Gambar 3.3 Tahapan Implementasi <i>K-Nearest Neighbors</i>	27
Gambar 3.4 Tahapan Evaluasi Menggunakan MAE	28
Gambar 4.1 Visualisasi <i>Rating</i>	49
Gambar 4.2 Visualisasi <i>Top Author</i>	50
Gambar 4.3 Visualisasi <i>Top Publisher</i>	51



BAB I PENDAHULUAN

1.1 Latar Belakang

Sistem rekomendasi merupakan suatu sistem atau teknik yang merekomendasikan produk, layanan, atau entitas tertentu (Banik, 2018). Sistem rekomendasi telah banyak digunakan oleh hampir setiap kalangan. Sistem rekomendasi tidak merekomendasikan *item* secara spesifik, namun merekomendasikan sejumlah *item* yang mungkin cocok dengan preferensi pengguna. Oleh karena itu, keluaran pada sistem rekomendasi berupa berupa “*top-N*” *recommendation*. Artinya, mesin akan memberikan sejumlah rekomendasi dengan peringkat teratas sesuai preferensi pengguna. Dalam industri, sistem rekomendasi adalah alat penting untuk meningkatkan pengalaman pengguna dan mempromosikan penjualan atau layanan untuk banyak situs web online dan aplikasi seluler. Misalnya, 80% film yang ditonton di *Netflix* berasal dari rekomendasi, dan 60% klik video berasal dari rekomendasi beranda di *YouTube* (Zhang *et al.*, 2020). Ada beberapa alasan mengapa sistem rekomendasi banyak digunakan, yaitu sebagai berikut:

1. Dapat meningkatkan penjualan *item*.
2. *Item* yang terjual menjadi lebih beragam.
3. Dapat meningkatkan kepuasan pengguna.
4. Dapat meningkatkan kesetiaan pengguna.
5. Dapat meningkatkan pemahaman tentang apa yang diinginkan pengguna.

Ada beberapa metode yang umumnya digunakan dalam membuat sistem rekomendasi, diantaranya yaitu *content based filtering*, *collaborative filtering*, dan

hybrid recommendation. *Hybrid recommendation* merupakan gabungan dari dua atau lebih sistem rekomendasi. *Collaborative filtering* adalah teknik yang paling banyak digunakan dalam sistem rekomendasi, yang mana cara kerjanya adalah menjumlahkan *rating* atau pilihan dari suatu produk, menemukan profil pengguna dengan melihat *history rating* yang diberikan pengguna, lalu menghasilkan rekomendasi yang baru berdasarkan perbandingan antarpola pengguna. Nilai *rating* bisa dalam biner seperti suka atau tidak suka, dan bisa juga dalam bentuk *voting* (Aryani, 2019).

Content based filtering didasarkan pada profil preferensi pengguna dan deskripsi *item*. Pada *content based filtering*, untuk mendeskripsikan *item*, biasanya menggunakan kata kunci selain profil pengguna untuk menunjukkan preferensi suka atau tidak suka dari pengguna. *Content based filtering* mempelajari profil minat pengguna baru berdasarkan data dari objek yang telah dinilai pengguna. Dengan kata lain, *content based filtering* merekomendasikan *item* yang mirip dengan *item* yang disukai di masa lalu (Geetha *et al.*, 2018). Semakin banyak informasi yang diberikan pengguna, semakin baik akurasi sistem rekomendasi. *Content based filtering* bekerja dengan cara menganalisis deskripsi *item* untuk mengidentifikasi *item* yang menarik bagi pengguna. Teknik pemberian rekomendasi yang digunakan dalam pendekatan ini mengandalkan deskripsi *item* dan menghasilkan rekomendasi dari *item* yang serupa dengan yang disukai pengguna di masa lalu, tanpa secara langsung bergantung pada preferensi pengguna yang tersimpan pada profil pengguna. Pendekatan ini tidak memerlukan basis pengguna yang besar dan mengumpulkan data tentang pengguna. *Content based filtering* memungkinkan rekomendasi hanya berdasarkan deskripsi *item* itu

sendiri dan bukan berdasarkan ketertarikan pengguna (Mondi dkk., 2019). Ini merupakan salah satu kelemahan metode *content based filtering* yang mana hasil rekomendasi untuk masing-masing pengguna menjadi kurang baik.

Penelitian yang dilakukan oleh Sukmana dkk. (2019) pada Implementasi *Content-Based Filtering* Pada Aplikasi Radar Zakat Dalam Merekomendasikan Preferensi Mustahik menyatakan bahwa hasil rekomendasi yang diberikan tidak menunjukkan urutan prioritas. Dokumen pertama yang direkomendasikan bukanlah rekomendasi terbaik dari dokumen-dokumen hasil rekomendasi. Sistem rekomendasi hanya menghasilkan rekomendasi berdasarkan kemiripan antar dokumen yang ada dan tidak melakukan perhitungan tingkat prioritas. Hal ini menunjukkan bahwa hasil rekomendasi yang diberikan dengan menggunakan metode *content based filtering* masih kurang efektif.

Selain itu, penelitian yang dilakukan oleh Fitrianti dkk. (2020) tentang pembuatan sistem rekomendasi film dengan pendekatan *content based filtering* memaparkan bahwa berdasarkan hasil pengujian, sistem yang dibangun hanya dapat memberikan rekomendasi film dengan nilai presisi 45% dan *recall* 60%.

Dalam penelitian yang dilakukan oleh Aryani dkk. (2019) tentang sistem rekomendasi dengan menggunakan metode *collaborative filtering* dijelaskan bahwa sistem informasi rekomendasi cinderamata khas Bengkulu berbasis *e-marketplace* berhasil dibangun dengan menggunakan metode *collaborative filtering*, sesuai dengan teori yang ada memperoleh hasil pengujian 100% berhasil pada pengujian dengan menggunakan metode *black box*, yang mana dinilai dari 34 pengujian fungsional yang dapat dijalankan di *e-marketplace*.

Penelitian yang dilakukan oleh Syah (2020) pada Performa Algoritma *User K-Nearest Neighbors* pada Sistem Rekomendasi di Tokopedia memaparkan bahwa *collaborative filtering* memiliki kemampuan lebih baik dalam memberikan sistem rekomendasi daripada *content based filtering*. Hal ini dapat diketahui dari nilai MAE yang dihasilkan yaitu 0.488. Semakin kecil nilai MAE, maka model tersebut semakin bagus yang berarti akan semakin baik dalam memberikan rekomendasi.

K-Nearest Neighbors pada penelitian yang dilakukan oleh Gusti dkk. (2019) pada Rekomendasi Sistem Pemilihan Mobil Menggunakan *K-Nearest Neighbors Collaborative Filtering*, memaparkan bahwa hasil akurasi yang didapatkan sebesar 95,15% pada nilai $K = 10$. Tingkat akurasi ini sudah sangat baik dalam memberikan rekomendasi kepada pengguna.

Penelitian mengenai *K-Nearest Neighbors* pada *The User Personalization with KNN for Recommender System* yang dilakukan oleh Dharma & Samosir (2019) menunjukkan bahwa rekomendasi kepada pengguna dapat dilakukan dengan menggunakan algoritma KNN yang dapat dilihat dari hasil evaluasi algoritma telah dilakukan dengan rumus RMSE yaitu 1,237.

Algoritma KNN hingga saat ini menjadi salah satu algoritma yang populer untuk sistem rekomendasi karena dapat memberikan rekomendasi dengan tingkat akurasi yang baik. Penelitian yang dilakukan oleh Riswanto *et al.*, (2019) pada *Mobile Recommendation System for Culinary Tourism Destination using K-Nearest Neighbors* menyampaikan bahwa algoritma *K-Nearest Neighbors* berhasil memberikan rekomendasi untuk membantu pengguna dalam memilih destinasi

kuliner di Yogyakarta yang mana sistem dapat memberikan rekomendasi yang tepat kepada pengguna sesuai dengan parameter yang diinginkan pengguna.

Buku merupakan salah satu sumber informasi yang dapat memberikan wawasan kepada pembaca untuk mengetahui banyak hal. Banyak buku dengan berbagai jenis yang terbit setiap hari di seluruh penjuru dunia. Banyaknya buku yang terbit memberikan dampak positif diantaranya yaitu bisa membaca buku yang beragam dan menemukan banyak pilihan. Namun, hal ini juga dapat menimbulkan kebingungan dalam menemukan buku yang sesuai untuk dibaca. Untuk memudahkan pembaca menemukan buku yang diinginkan, maka diperlukan suatu sistem yang dapat merekomendasikan buku kepada pembaca.

Berdasarkan pemaparan tersebut, maka pada penelitian ini akan dilakukan Implementasi Algoritma *K-Nearest Neighbors* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering*. Dalam penelitian ini, akan dibuat suatu sistem rekomendasi menggunakan data buku sebagai studi kasus. Untuk selanjutnya, sistem rekomendasi ini dapat digunakan pada studi kasus lainnya, baik menggunakan data primer maupun menggunakan data sekunder yang memiliki atribut *rating* pada datasetnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah pada penelitian ini adalah bagaimana mengimplementasikan algoritma *K-Nearest Neighbors* pada sistem rekomendasi buku menggunakan teknik *collaborative filtering*.

Untuk menyelesaikan permasalahan yang disebutkan di atas, maka dalam penelitian ini dibagi dalam beberapa *research question*, antara lain:

1. Bagaimana mengimplementasikan algoritma *K-Nearest Neighbors* pada sistem rekomendasi buku menggunakan teknik *collaborative filtering*.
2. Bagaimana nilai akurasi menggunakan *Mean Absolute Error* pada algoritma *K-Nearest Neighbors* dalam memberikan rekomendasi buku.

1.3 Tujuan Penelitian

Adapun tujuan dilakukannya penelitian ini adalah sebagai berikut:

1. Mengimplementasikan algoritma *K-Nearest Neighbors* pada sistem rekomendasi buku menggunakan teknik *collaborative filtering*.
2. Mengetahui nilai akurasi menggunakan *Mean Absolute Error* pada algoritma *K-Nearest Neighbors* pada sistem rekomendasi buku.

1.4 Manfaat Penelitian

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberi pengetahuan kepada peneliti lain mengenai implementasi algoritma *K-Nearest Neighbors* pada sistem rekomendasi buku menggunakan teknik *collaborative filtering*.
2. Dapat memberikan rekomendasi kepada pengguna mengenai buku yang dapat dibaca selanjutnya.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan pada penelitian ini merupakan data sekunder yang diambil dari sumber terbuka, yaitu *Kaggle*.
2. Teknik yang digunakan adalah *item-based collaborative filtering*.

3. Nilai K yang digunakan pada penelitian ini adalah $K=13$, $K=15$, dan $K=26$ yang didapatkan dengan menentukan nilai K optimal dari rentang K tertentu.
4. Perhitungan nilai *similarity* menggunakan *Pearson Correlation*.

1.6 Sistematika Penulisan

Sistematika penulisan yang diajukan dalam penelitian ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab ini menguraikan tentang latar belakang, rumusan masalah yang didetailkan dengan *research question*, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini akan membahas seluruh dasar-dasar teori yang digunakan dalam penelitian, mulai dari sistem rekomendasi, *K-Nearest Neighbors*, definisi buku dan semua yang digunakan dalam tahapan analisis, perancangan, serta implementasi dalam penelitian.

BAB III : METODOLOGI PENELITIAN

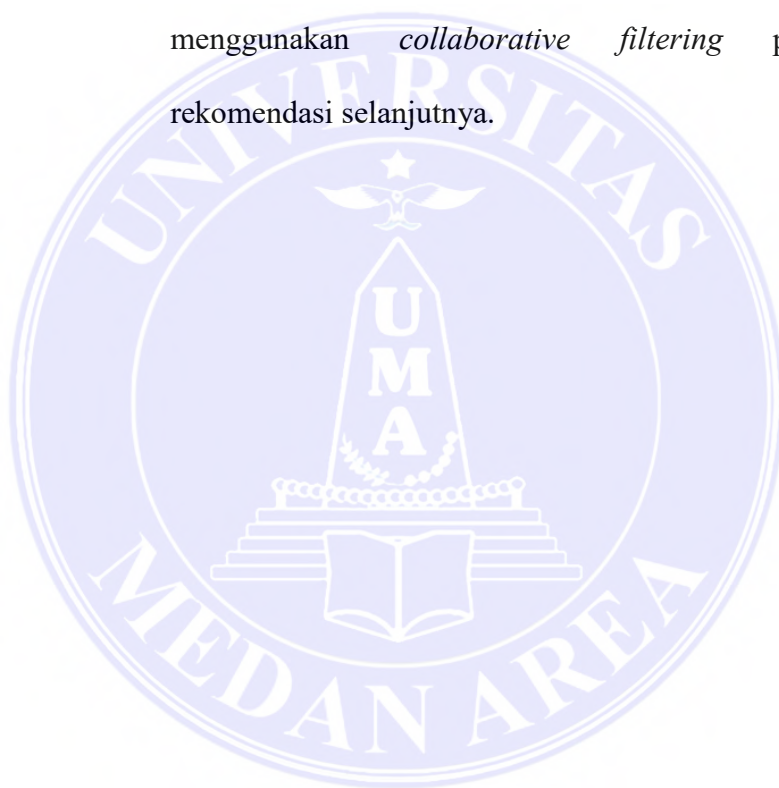
Bab ini akan membahas mengenai tahap-tahap yang akan diterapkan pada penelitian. Setiap rencana dari tahapan penelitian dideskripsikan secara rinci.

BAB IV : HASIL DAN PEMBAHASAN

Bab ini membahas tentang langkah-langkah implementasi dan hasil program yang terdiri dari tampilan program, alur program dan penjelasan program.

BAB V : KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan dari semua uraian-uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penerapan dan penelitian menggunakan *collaborative filtering* pada sistem rekomendasi selanjutnya.



BAB II LANDASAN TEORI

2.1 Sistem Rekomendasi

Sistem rekomendasi adalah sistem yang memprediksi peringkat dan preferensi yang mungkin diberikan pengguna pada suatu produk. Ini sering disajikan secara berurutan sebagai "*top-N*" recommendation (Kane, 2021). Sistem rekomendasi dirancang dengan tujuan untuk membantu pengguna dengan cara memberikan rekomendasi kepada pengguna ketika pengguna dihadapkan dengan jumlah informasi yang besar yang mana diharapkan rekomendasi yang diberikan dapat membantu pengguna dalam proses menentukan apa yang sesuai kepada pengguna atau apa yang pengguna inginkan (Wahyudi, 2017). Sistem rekomendasi menyediakan informasi kepada pengguna mengenai daftar *item* yang sesuai dengan keinginan pengguna. Pencarian *item* yang direkomendasikan bisa berdasarkan kemiripan, baik berdasarkan konten maupun berdasarkan korelasi antar pengguna berdasarkan *rating* yang diberikan pada *item*.

Ada beberapa alasan mengapa sistem rekomendasi banyak digunakan oleh berbagai kalangan, antara lain dapat memudahkan pengguna untuk menemukan *item* yang sesuai dengan keinginan pengguna, dapat mengingatkan pengguna tentang sesi yang mereka tutup sebelumnya pada suatu aplikasi, dan membantu pengguna menemukan produk yang lebih beragam (Ayyadevara, 2018). Sistem rekomendasi adalah metode yang kuat yang memungkinkan pengguna untuk menyaring sejumlah besar *item* dan informasi. Ada banyak sekali informasi yang tersedia di internet dan ini akan menimbulkan kebingungan kepada pengguna untuk memilih apa yang sesuai dengan yang mereka butuhkan.

Maka dari itu, sistem rekomendasi sangat membantu pengguna untuk memberikan pilihan *item* yang paling sesuai dengan kebutuhan pengguna (Sattar *et al.*, 2017).

Item-based dan *user-based* dikenal dalam sistem rekomendasi saat menentukan rekomendasi. Keduanya sering digunakan dalam penelitian yang disesuaikan dengan kasus yang ingin diteliti. Dengan menemukan *item* tambahan yang memiliki nilai kemiripan tertinggi dengan *item* yang dipilih, teknik *item-based* menemukan *item* lain yang mungkin disukai pengguna tersebut. Di sisi lain, *user-based* memberikan rekomendasi dengan membandingkan semua *item* pada semua pengguna aktif dengan pengguna tertentu dan menganalisis kesamaan (Nasuha dkk., 2021).

Secara umum, sistem rekomendasi memiliki tiga metode yang dapat digunakan, yaitu *content based filtering*, *collaborative filtering*, dan *hybrid recommendation* (Shuai Zhang, 2018). Setiap metode disesuaikan dengan permasalahan agar menghasilkan informasi yang tepat. Pendekatan yang digunakan dalam sistem rekomendasi ditentukan oleh masalah yang dihadapi.

2.1.1 *Content Based Filtering*

Content based filtering bekerja dengan mencari korelasi suatu *item* dan memberikan rekomendasi berdasarkan kemiripan atribut *item* dan preferensi profil pengguna. Dengan kata lain, *content based filtering* merekomendasikan item yang serupa dengan item yang disukai atau dicari pengguna di masa lalu (Theobald, 2018). Penggunaan *content based filtering* dalam memberikan rekomendasi memiliki kelebihan dan kelemahan. Beberapa kelebihan dari pendekatan ini, yaitu (Nasuha dkk., 2021):

1. *User Independence*

Pendekatan *content based filtering* memanfaatkan *rating* yang diberikan oleh pengguna aktif untuk membangun profilnya sendiri. Pendekatan ini tidak membutuhkan *rating* dari pengguna lain untuk mendapatkan rekomendasi suatu *item*. *Content based filtering* merekomendasikan *item* yang mirip dengan deskripsi *item* yang disukai pengguna di masa lalu.

2. *Transparency*

Pada *content based filtering* tidak akan ada *transparency* seperti yang terjadi pada *collaborative filtering* karena *transparency* akan terjadi apabila terdapat kesamaan selera antara pengguna yang satu dengan pengguna yang lain. Sementara itu, *content based filtering* tidak membutuhkan *rating* dari pengguna lain karena pengguna akan mendapatkan rekomendasi *item* berdasarkan profil pengguna itu sendiri.

Berikut adalah beberapa kelemahan dalam penggunaan pendekatan *content based filtering* dalam memberikan rekomendasi *item* (Mohanty *et al.*, 2020):

1. *Limited content analysis*

Pendekatan *content based filtering* memiliki keterbatasan dalam menganalisis deskripsi *item*. Pendekatan ini tidak dapat merekomendasikan suatu *item* apabila deskripsi *item* tersebut tidak lengkap.

2. *Over-specialization*

Pendekatan ini tidak dapat merekomendasikan *item* yang memiliki konten yang berbeda dengan *item* yang dipilih oleh pengguna.

3. *New user problem*

Pendekatan ini tidak dapat merekomendasikan suatu *item* kepada pengguna baru dikarenakan pengguna tersebut belum memiliki profil. Pengguna harus memberikan *rating* pada sejumlah *item* tertentu agar sistem rekomendasi dapat merekomendasikan *item* yang sesuai dengan keinginan pengguna.

2.1.2 *Collaborative Filtering*

Collaborative Filtering populer dalam bisnis maupun dalam penelitian karena kesederhanaannya dan tingkat kinerjanya yang tinggi. Sistem rekomendasi yang menerapkan pendekatan *collaborative filtering* memberikan rekomendasi *item* yang memiliki korelasi berdasarkan *rating* yang diberikan oleh tiap pengguna pada *item* (Hwangbo *et al.*, 2018). Hal ini didasarkan pada gagasan bahwa orang-orang yang setuju dengan evaluasi mereka tentang barang-barang tertentu di masa lalu cenderung setuju lagi di masa depan. Dalam merekomendasikan *item*, profil dari target pengguna akan disesuaikan dengan data untuk melihat profil pengguna lain yang memiliki selera yang mirip. *Item* yang disukai tersebut kemudian direkomendasikan kepada target yang kemungkinan menyukainya juga. Pendekatan *collaborative filtering* memiliki banyak kelebihan, salah satunya adalah penerapan metode ini relatif sederhana sehingga banyak digunakan oleh banyak kalangan.

Pada penelitian yang dilakukan oleh Wardana & Ananta Timur (2018) tentang *Collaborative Filtering Recommender System* pada Virtual 3D Kelas Cendekia, dipaparkan bahwa berdasarkan hasil pengujian yang dilakukan dengan melibatkan 10 pengguna menggunakan data dengan *sparsity* 80% dan standar

deviasi 1,14, sistem rekomendasi berbasis *collaborative filtering* yang diimplementasi dapat memprediksi *rating* dengan RMSE 1,06 atau dengan kata lain tingkat akurasinya sebesar 78,78%.

Collaborative filtering, dalam memberikan rekomendasi, memiliki banyak kelebihan (Geetha *et al.*, 2018). Berikut adalah beberapa kelebihan dari pendekatan ini:

1. *No domain knowledge necessary*

Pendekatan ini tidak memerlukan pengetahuan domain karena *embeddings* dipelajari secara otomatis.

2. *Serendipity*

Pendekatan ini dapat membantu pengguna menemukan minat baru. Secara terpisah, sistem *machine learning* mungkin tidak mengetahui bahwa pengguna tertarik pada *item* tertentu, tetapi pendekatan ini dapat merekomendasikannya karena pengguna serupa tertarik pada *item* tersebut.

3. *Great starting point*

Sampai batas tertentu, sistem hanya membutuhkan matriks umpan balik untuk melatih model faktorisasi matriks. Secara khusus, sistem tidak memerlukan fitur kontekstual. Dalam praktiknya, ini dapat digunakan sebagai salah satu dari beberapa kandidat generator.

Adapun kelemahan dari pendekatan ini adalah *cold-start problem* yaitu jika terdapat *item* baru, *item* baru tersebut tidak dapat direkomendasikan karena belum mendapatkan *rating*. Akibatnya, *item* tersebut tidak akan direkomendasikan kepada pengguna (Melville & Sindhvani, 2017).

2.1.2.1 Item-Based Collaborative Filtering

Item-based collaborative filtering adalah salah satu yang paling populer dari pendekatan *collaborative filtering* (Sattar *et al.*, 2017). *Item-based collaborative filtering* bekerja dengan cara menghitung kesamaan antara masing-masing *item* (Selvi & Priya., 2020) Misalnya, Kia menyukai buku “Pulang” dan buku “Pergi”, tetapi belum membaca buku “Tentang Kamu”. Selanjutnya, Kia memastikan bahwa Ica dan Goo memberikan *rating* yang sama pada buku “Tentang Kamu” dengan dua buku sebelumnya. Oleh karena itu, Kia menyimpulkan bahwa ia juga menyukai buku “Tentang Kamu”, seperti terlihat pada Gambar 2.1.



Gambar 2.1 Konsep IBCF

2.2 K-Nearest Neighbors (KNN)

Algoritma KNN merupakan metode pengklasifikasian objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Muhathir dkk., 2020). Terdapat beberapa jenis klasifikasi yang sering digunakan pada algoritma KNN, yaitu (Khairina dkk., 2022):

1. *Fine* KNN: Pengklasifikasi tetangga terdekat yang membuat perbedaan antara kelas dengan jumlah tetangga diatur ke 1.
2. *Medium* KNN: Kelas tetangga terdekat yang membuat perbedaan lebih sedikit dari pada *Fine* KNN dengan jumlah tetangga diatur ke 10.
3. *Coarse* KNN: Pengklasifikasi tetangga terdekat yang membuat perbedaan kasar antarkelas, dengan jumlah tetangga diatur ke 100.
4. *Cosine* KNN: Pengklasifikasi tetangga terdekat yang menggunakan jarak cosinus metrik.
5. *Cubic* KNN: Kelas tetangga terdekat yang menggunakan jarak kubik metrik.
6. *Weighted* KNN: Sebuah kelas tetangga terdekat yang menggunakan pembobotan jarak.

Dalam penggunaan algoritma KNN, data dibagi menjadi dua bagian, yaitu data latih dan data uji. Data latih digunakan algoritma untuk melakukan dasar prediksi, sedangkan data uji terdiri dari nilai yang diprediksi oleh algoritma (Prasetya, 2017). Alur utama dalam memprediksi *rating* adalah sebagai berikut (Nugroho dkk., 2020):

1. Melihat *user matrix item* pada dataset
2. Menghitung nilai *similarity* antar-*item* (*item-based*) atau *similarity* antar-*user* (*user-based*).
3. Menentukan banyaknya *neighbor* yang memiliki nilai *similarity* terbesar dengan *item A* atau *user A* dengan melihat nilai kemiripannya.
4. Memprediksi *rating* dari *user A* terhadap *item* tertentu dengan perhitungan yang melibatkan *neighbor* yang telah ditentukan.

5. Memberikan rekomendasi *item* kepada *user*

Algoritma KNN mengklasifikasikan objek berdasarkan data pembelajaran yang merupakan jarak terdekat atau memiliki karakteristik serupa dari suatu objek ke objek lain. Kedekatan karakteristik pengguna dapat ditemukan dengan menggunakan *pearson correlation*, *cosine similarity*, dan *adjusted cosine similarity* (Dharma & Samosir, 2019). Dalam penelitian ini, dipergunakan pendekatan *pearson correlation* yang dapat dilihat pada persamaan 2.1 berikut.

$$PC(i, j) = \frac{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_i)(r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_i)^2 \cdot \sum_{u \in U_{ij}} (r_{uj} - \bar{r}_j)^2}} \quad (2.1)$$

Keterangan:

$PC(i, j)$ = *similarity* antara *item* i dan *item* j

r_{ui}, r_{uj} = *rating* yang diberikan *user* u ke *item* i dan j

\bar{r}_i, \bar{r}_j = rata-rata *rating* dari *item* i dan j

Setelah proses ini, tahap perhitungan prediksi akan dilakukan untuk memprediksi *rating* yang akan diberikan pengguna untuk *item* tertentu. Setelah mengidentifikasi *neighbor* dengan nilai kemiripan tertinggi, akan dilakukan proses prediksi hingga sampai pada pemberian rekomendasi. Persamaan *Weighted Sum* merupakan salah satu rumus yang digunakan untuk menentukan nilai prediksi (Setiawan dkk., 2019).

Berikut adalah persamaan 2.2 untuk menghitung prediksi pada teknik IBCF.

$$P(a, j) = \frac{\sum_{i \in I} (R_{a,i} \cdot S_{i,j})}{\sum_{i \in I} |S_{i,j}|} \quad (2.2)$$

Keterangan:

(a,j) = prediksi *rating* pada *item* j untuk *user* a

$R_{a,i}$ = *rating* yang diberikan *user* a ke *item* i

$S_{i,j}$ = *similarity* antara *item* i dan j

2.3 Mean Absolute Error (MAE)

Analisis hasil pengujian dilakukan dengan melakukan analisa terhadap nilai MAE. Analisis hasil pengujian dipengaruhi oleh kemiripan antar nilai prediksi sistem dan nilai sebenarnya. Semakin mirip antara nilai prediksi sistem dan nilai *rating* sebenarnya, maka akurasi pada sistem semakin baik. Pada MAE, semakin besar nilai MAE yang dihasilkan menandakan hasil prediksi sistem semakin buruk. Semakin kecil nilai MAE dan mendekati nol, maka prediksi sistem semakin akurat. Persamaan MAE dijelaskan pada persamaan 2.3 (Wang & Lu, 2018).

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad 2.3$$

Keterangan:

p_i = *rating* yang diprediksi

q_i = *rating* yang sebenarnya

N = jumlah data

2.4 Buku






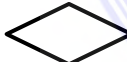

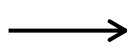
Buku adalah kumpulan informasi tentang topik yang berisikan hasil analisis. Buku dapat berupa buku pelajar resmi ataupun sekadar buku bacaan hiburan seperti majalah dan sejenisnya. Buku didefinisikan sebagai karya tulis dan cetak yang halaman-halamannya dijilid pada satu sisi atau hasil dari sebuah karya yang ditujukan untuk penerbitan (Nasuha dkk., 2021). Ada beberapa bagian

penting pada buku, seperti judul buku, nomor ISBN, pengarang, penerbit, dan kategori buku (Darwati, 2017).

2.5 Flowchart

Flowchart atau diagram alur adalah representasi diagram dari langkah-langkah suatu algoritma. Dalam diagram alur, kotak dengan bentuk berbeda digunakan untuk menunjukkan jenis operasi yang berbeda. Kotak-kotak ini kemudian dihubungkan oleh garis dengan panah yang menunjukkan aliran atau arah yang harus dilalui untuk mengetahui langkah selanjutnya. Berikut ini adalah simbol standar yang digunakan dalam diagram alur program (Chaudhuri, 2020):

Tabel 2.1 Simbol *Flowchart*

Simbol	Nama	Keterangan
	Terminator	Simbol yang menyatakan permulaan atau akhir suatu program
	Input/Output Data	Simbol yang menyatakan proses input/output
	Proses	Simbol yang menyatakan proses yang dilakukan computer
	Sub Program (<i>Predefined Process</i>)	Simbol untuk sub program/proses menjalankan sub program
	Input/Output Dokumen	Simbol yang menyatakan input/output dalam bentuk dokumen
	<i>Decision</i>	Simbol yang menunjukkan kondisi tertentu yang akan menghasilkan atau mengambil keputusan
	<i>On Page Connector</i>	Simbol untuk menghubungkan bagian-bagian dari diagram alur yang dilanjutkan pada lembar kerja yang sama
	Garis Alir (<i>Flow Line</i>)	Simbol yang digunakan untuk menggabungkan antara simbol yang satu dengan simbol yang lain

2.6 Penelitian Terdahulu

Penelitian terdahulu merupakan penelitian yang telah dilakukan oleh peneliti sebelumnya yang memiliki keterkaitan dengan penelitian yang akan dilakukan oleh penulis. Penelitian terdahulu juga menjadi salah satu bahan pertimbangan sehingga dapat memberi referensi dalam menulis ataupun mengkaji

penelitian yang akan dilakukan. Berikut adalah penelitian terdahulu tentang sistem rekomendasi dengan menggunakan algoritma KNN yang menjadi acuan dan referensi penulis dalam melakukan penelitian ini:

1. Penelitian yang dilakukan Fitrianti dkk. (JOINS (*Journal of Information System*), Tahun 2022) membahas tentang Sistem Rekomendasi Film Berbasis *Website* dengan Metode *Prototype* Menggunakan Algoritma *K-Nearest Neighbors*. Data yang digunakan pada penelitian ini memiliki atribut *genres*, *rate*, *title*, *imdbid*, *plot*, *popularity*, *poster*, dan *length*. Namun, data yang diperlukan adalah *genres* dan *rate*, sehingga atribut lainnya dihapus. Total baris pada data ini adalah 1.554. Data ini kemudian dibagi menjadi data *training* dan data *testing* dengan proporsi 80:20. Perbedaan penelitian ini dengan penelitian yang akan penulis lakukan adalah pada penelitian ini menggunakan teknik *content based filtering* dan menggunakan presisi dan *recall* untuk mengukur akurasi, sedangkan pada penelitian yang akan penulis lakukan menggunakan teknik *collaborative filtering* dan menggunakan MAE untuk mengukur nilai *error*.
2. Penelitian yang dilakukan Aryani dkk. (Jurnal Rekursif, Tahun 2019) membahas tentang Perancangan Sistem Rekomendasi Pemilihan Cinderamata Khas Bengkulu Berbasis E-Marketplace. Penelitian ini menggunakan data 6 produk dengan rentang *rating* 1–5. Perbedaan penelitian ini dengan penelitian yang akan penulis lakukan adalah pada penelitian ini menggunakan *Adjusted Cosine Similarity* untuk menentukan kemiripan antarproduk, sedangkan pada penelitian yang akan penulis lakukan menggunakan *Pearson Correlation*.

3. Penelitian yang dilakukan Rama Dian Syah (Jurnal Informatika Universitas Pamulang, 2020) membahas tentang Performa Algoritma User *K-Nearest Neighbors* pada Sistem Rekomendasi di Tokopedia. Data yang digunakan pada penelitian ini adalah data sekunder yang diambil dari *Kaggle* yaitu Tokopedia Product Reviews yang berjumlah 40.608 ulasan pengguna Tokopedia. Data ini kemudian dibagi menjadi data latih dan data uji dengan proporsi 80:20. Perbedaan penelitian ini dengan penelitian yang akan penulis lakukan adalah pada penelitian ini menggunakan *user-based*, sedangkan pada penelitian yang akan penulis lakukan menggunakan *item-based*.
4. Penelitian yang dilakukan Gusti dkk. (TEKTRIKA - Jurnal Penelitian dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, dan Elektronika, Tahun 2019) membahas tentang Rekomendasi Sistem Pemilihan Mobil Menggunakan *K-Nearest Neighbors Collaborative Filtering*. Penelitian ini menggunakan 20 data uji dengan 1.088 data latih. Perbedaan penelitian ini dengan penelitian yang akan penulis lakukan adalah pada penelitian ini menggunakan *Euclidean Distances* untuk menghitung jarak *item*, sedangkan pada penelitian yang akan penulis lakukan menggunakan *Pearson Correlation* untuk menghitung kemiripan *item*.
5. Penelitian yang dilakukan oleh Arie Satia Dharma (SinkrOn: Penelitian dan Jurnal Teknik Informatika, Tahun 2019) membahas tentang *The User Personalization with KNN for Recommender System*. Data yang diamati pada penelitian ini berasal dari 39 pengguna yang berbeda yang diperoleh

saat melacak dari web log server dalam bentuk *IP address*, *access time*, *status code*, *HTTP request method*, *page size/number of bytes*, *referer*, dan *user agent*. Perbedaan penelitian ini dengan penelitian yang akan penulis lakukan adalah pada penelitian ini menggunakan RMSE untuk menghitung nilai *error*, sedangkan pada penelitian yang akan penulis lakukan menggunakan MAE.

6. Penelitian yang dilakukan oleh Riswanto *et al.*, (*Journal of Physics: Conference Series*, Tahun 2019) membahas tentang *Mobile Recommendation System for Culinary Tourism Destination using K-Nearest Neighbors*. Dalam penelitian ini, data yang digunakan untuk memberikan rekomendasi kepada pengguna terdiri dari 5 kriteria, yaitu rasa makanan, suasana lingkungan, harga, pelayanan, dan jarak. Perbedaan penelitian ini dengan penelitian yang akan penulis lakukan adalah pada penelitian ini evaluasi dilakukan dengan melakukan perhitungan kemiripan secara manual pada data sampel, sedangkan pada penelitian yang akan penulis lakukan evaluasi dilakukan dengan menghitung nilai MAE untuk mengukur nilai *error*.

Hasil penelitian dari penelitian terdahulu dapat dilihat pada tabel 2.2.

Tabel 2.2 Penelitian Terdahulu

No	Penulis & Tahun	Judul Penelitian	Hasil
1	Alifia Rahma Fitrianti, Asih Rohmani, Widjanarto (2020)	Sistem Rekomendasi Film Berbasis <i>Website</i> Dengan Metode <i>Prototype</i> Menggunakan <i>Algoritma K-Nearest Neighbors</i>	Sistem rekomendasi mampu memberikan rekomendasi film dengan nilai presisi 45% dan <i>recall</i> 60%.
2	Aryani, Boko Susilo, dan Yudi Setiawan (2019)	Perancangan Sistem Rekomendasi Pemilihan Cinderamata Khas Bengkulu Berbasis <i>E-Marketplace</i>	Sistem rekomendasi ini memperoleh hasil pengujian 100% berhasil pada pengujian dengan menggunakan metode <i>black box</i> , yang dinilai dari 34 pengujian fungsional yang dapat dijalankan di <i>e-marketplace</i> .
3	Rama Dian Syah (2020)	Performa Algoritma <i>User K-Nearest Neighbors</i> pada Sistem Rekomendasi di Tokopedia	Sistem rekomendasi berhasil memberikan rekomendasi dengan nilai MAE yang dihasilkan sebesar 0.488.
4	Ilham Gumantung Gusti, Muhammad Nasrun, Ratna Astuti Nugrahaeni (2019)	Rekomendasi Sistem Pemilihan Mobil Menggunakan <i>K-Nearest Neighbor Collaborative Filtering</i>	Hasil akurasi yang didapatkan sebesar 95,15% pada nilai K = 10 dan berhasil memberikan rekomendasi.
5	Arie Satia Dharma (2019)	<i>The User Personalization with KNN for Recommender System</i>	Rekomendasi kepada pengguna berhasil didapatkan dengan RMSE (<i>Pearson Correlation</i>) sebesar 1,346 dan RMSE (<i>Cosine Correlation</i>) sebesar 1,237.
6	Riswanto, E., Robi'in, B, Suparyanto (2019)	<i>Mobile Recommendation System for Culinary Tourism Destination using K-Nearest Neighbors</i>	Algoritma KNN berhasil memberikan rekomendasi untuk membantu pengguna dalam memilih destinasi kuliner di Yogyakarta sesuai dengan parameter yang diinginkan pengguna.

BAB III METODOLOGI PENELITIAN

3.1 Pengumpulan Data

Pada subbab ini akan membahas mengenai jenis data dan sumber data serta metode pengumpulan data.

3.1.1 Jenis dan Sumber Data

Jenis data yang dipakai pada penelitian ini merupakan data sekunder. Data ini diambil dari *website* penyedia data *open source*, yaitu *Kaggle*. *Kaggle* merupakan sebuah *website* yang menyediakan dataset secara *open source*. Data yang digunakan pada penelitian ini, yaitu *Book Recommendation Dataset*. Dataset ini berisikan tiga *file*, yaitu *Books.csv*, *Ratings.csv*, dan *Users.csv*. *Books.csv* memuat data sebanyak 271.360 baris dan 8 kolom, *Ratings.csv* memuat data sebanyak 1.149.780 baris dan 3 kolom, dan *Users.csv* memuat data sebanyak 278.858 baris dan 3 kolom. Nilai *rating* bersifat eksplisit, dinyatakan dalam skala 1-10 atau implisit, yang dinyatakan dengan 0.

Berikut ini adalah tabel data yang digunakan pada penelitian ini:

Tabel 3.1 Data *Books.csv*

No	ISBN	Book Title	Book Author	Year of Publication	Publisher
1	195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press
2	2005018	Clara Callan	Richard Bruce Wright	2001	Harper Flamingo Canada
3	60973129	Decision in Normandy	Carlo D'Este	1991	Harper Perennial
4	393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company
5	399135782	The Kitchen God's Wife	Amy Tan	1991	Putnam Pub Group
6	887841740	The Middle Stories	Sheila Heti	2004	House of Anansi Press
7	1552041778	Jane Doe	R. J. Kaiser	1999	Mira Books

8	1881320189	Goodbye to the Buttermilk Sky	Julia Oliver	1994	River City Pub
9	440234743	The Testament	John Grisham	1999	Dell
10	452264464	Beloved (Plume Contemporary Fiction)	Toni Morrison	1994	Plume
11	439095026	Tell Me This Isn't Happening	Robynn Clairday	1999	Scholastic
12	689821166	Flood : Mississippi 1927	Kathleen Duey	1998	Aladdin
13	971880107	Wild Animus	Rich Shapero	2004	Too Far
14	345402871	Airframe	Michael Crichton	1997	Ballantine Books
15	345417623	Timeline	MICHAEL CRICHTON	2000	Ballantine Books

Berdasarkan tabel 3.1, dapat dilihat bahwa Data *Books.csv* memiliki 5 atribut atau kolom, yaitu ISBN, Book Title, Book Author, Year of Publication, dan Publisher.

Tabel 3.2 Data *Ratings.csv*

No	User ID	ISBN	Book Rating
1	276725	034545104X	0
2	276726	155061224	5
3	276727	446520802	0
4	276729	052165615X	3
5	276729	521795028	6
6	276733	2080674722	0
7	276736	3257224281	8
8	276737	600570967	6
9	276744	038550120X	7
10	276745	342310538	10
11	276751	3596218098	8
12	276754	684867621	8
13	276755	451166892	5
14	276760	8440682697	10
15	276762	034544003X	0

Pada tabel Data *Ratings.csv* dapat dilihat bahwa tabel ini memiliki 3 atribut atau kolom yaitu User ID, ISBN, dan Book Rating.

Tabel 3.3 Data *Users.csv*

No	User ID	Location	Age
1	1	NYC, New York, USA	21
2	2	Stockton, California, USA	18
3	3	Moscow, Yukon Territory, Russia	20
4	4	Porto, V.N.Gaia, Portugal	17
5	5	Farnborough, Hants, United Kingdom	18
6	6	Santa Monica, California, USA	61
7	7	Washington, Dc, USA	22
8	8	Timmins, Ontario, Canada	24
9	9	Germantown, Tennessee, USA	15
10	10	Albacete, Wisconsin, Spain	26
11	11	Melbourne, Victoria, Australia	14
12	12	Fort Bragg, California, USA	16
13	13	Barcelona, Barcelona, Spain	26
14	14	Mediapolis, Iowa, USA	17
15	25	Oakland, California, USA	55

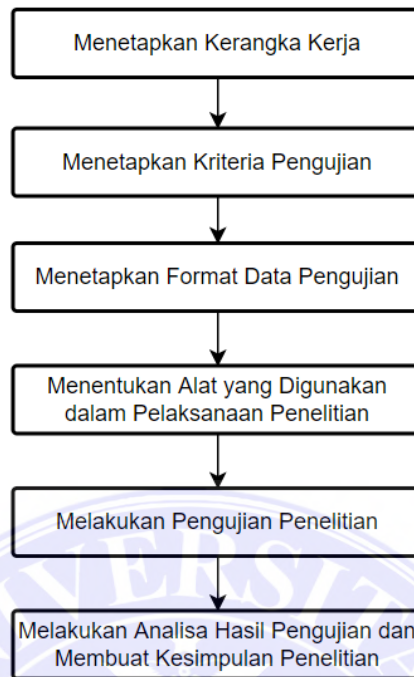
Pada tabel Data *Users.csv* di atas dapat dilihat bahwa tabel ini memiliki 3 atribut atau kolom yaitu User ID, Location, dan Age.

3.1.2 Metode Pengumpulan Data

Dataset diunduh dari *Kaggle* (<https://www.kaggle.com/arashnic/book-recommendation-dataset>) berdasarkan kriteria data yang diinginkan. Dalam hal ini yaitu ketiga *file* pada *Book Recommendation Dataset*.

3.2 Tahapan Penelitian

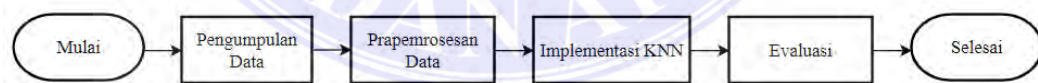
Untuk mengerjakan penelitian dengan judul Implementasi Algoritma *K-Nearest Neighbors* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering* ini, maka penelitian akan dibagi menjadi beberapa tahapan yang akan dijelaskan pada subbab selanjutnya.



Gambar 3.1 Tahapan Penelitian

3.2.1 Kerangka Kerja

Kerangka kerja merupakan rangkaian kerja yang akan dilakukan untuk menyelesaikan masalah penelitian. Gambar 3.2 adalah kerangka kerja yang akan digunakan pada penelitian ini:

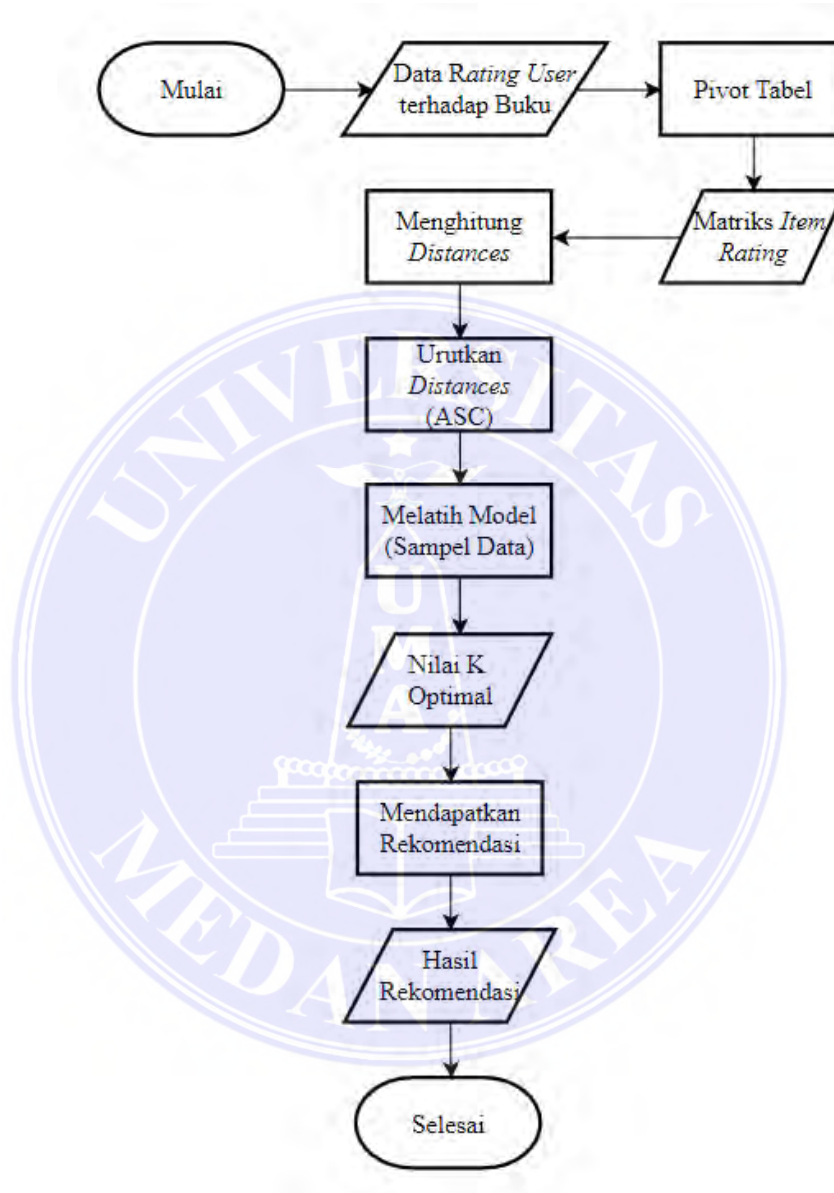


Gambar 3.2 Kerangka Kerja Sistem Rekomendasi Buku Menggunakan Algoritma *K-Nearest Neighbors* dengan Teknik *Collaborative Filtering*

Berdasarkan gambar 3.2, kerangka kerja pada sistem rekomendasi ini dimulai dengan tahapan pengumpulan data. Data yang digunakan diambil dari *website* sumber terbuka yaitu *Kaggle*. Setelah tahapan pengumpulan data, dilakukan prapemrosesan data yaitu mempersiapkan dataset menjadi data yang

siap digunakan untuk tahapan implementasi. Tahapan selanjutnya yaitu implementasi Algoritma KNN, kemudian dilakukan evaluasi.

Gambar 3.3 adalah proses implementasi *K-Nearest Neighbors*:

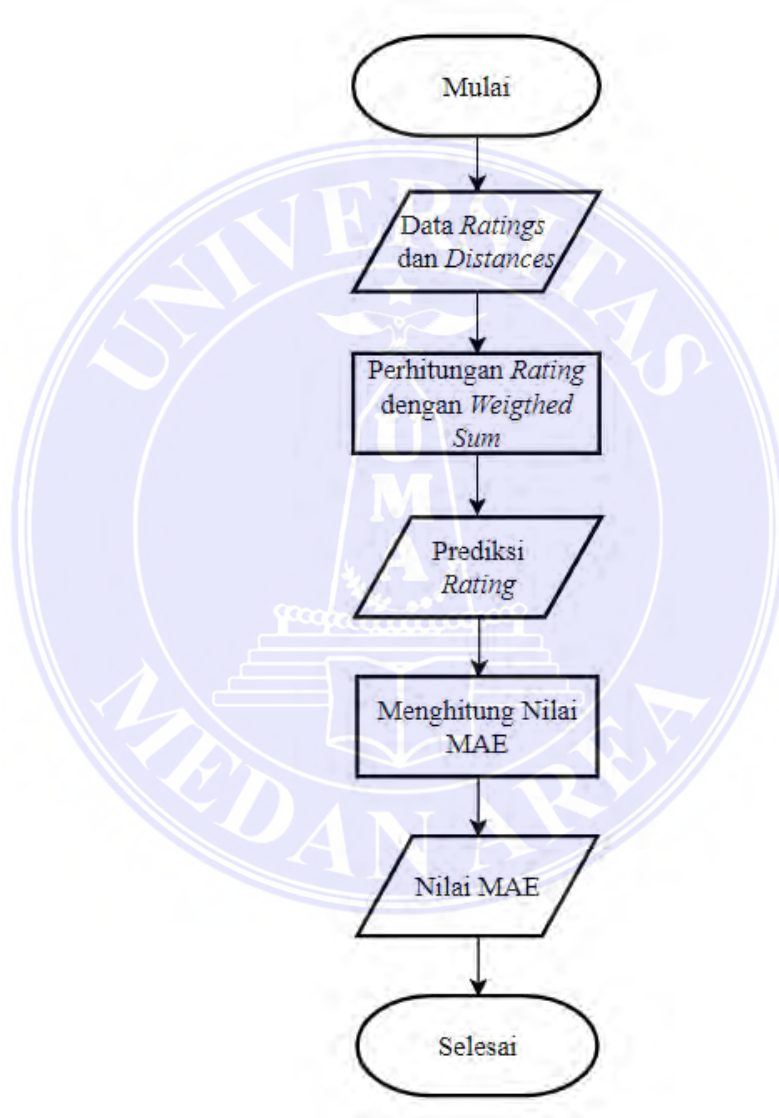


Gambar 3.3 Tahapan Implementasi *K-Nearest Neighbors*

Berdasarkan gambar 3.3, tahapan implementasi algoritma KNN dimulai dari membuat tabel pivot dengan menggunakan data *rating* pengguna terhadap buku. Luaran dari proses ini adalah matriks *item rating*. Selanjutnya menghitung *distances* antarbuku yang mana *distances* ini akan diurutkan dari yang terkecil

hingga yang terbesar. Setelah *distances* diurutkan, tahapan selanjutnya adalah melatih model dengan menggunakan sampel data untuk mendapatkan nilai K optimal. Setelah nilai K optimal didapatkan, nilai K ini kemudian digunakan untuk mendapatkan rekomendasi terhadap *item* yang dipilih.

Gambar 3.4 adalah proses evaluasi nilai *error* menggunakan MAE:



Gambar 3.4 Tahapan Evaluasi Menggunakan MAE

Berdasarkan gambar 3.4, tahapan evaluasi menggunakan MAE, dimulai dari menghitung *rating item* dengan *Weighted Sum* menggunakan data *ratings* dan *distances* yang telah didapatkan pada tahapan implementasi algoritma KNN yang

dijelaskan pada gambar 3.3. Luaran dari proses ini adalah prediksi *rating* terhadap *item*, yang selanjutnya dipakai untuk menghitung nilai MAE untuk mendapatkan hasil evaluasi dari model yang dibuat.

Gambaran besar dari kerangka kerja pada penelitian ini sebagaimana telah dijelaskan pada tiap kerangka kerja dimulai dari tahapan pengumpulan data yang mana pada penelitian ini menggunakan data buku yang didapat dari *website Kaggle*, yaitu *Book Recommendation Dataset*. Dataset ini berisikan tiga *file* yaitu *Books.csv*, *Ratings.csv*, dan *Users.csv*. *Books.csv* memuat data sebanyak 271.360 baris dan 8 kolom, *Ratings.csv* memuat data sebanyak 1.149.780 baris dan 3 kolom, dan *Users.csv* memuat data sebanyak 278.858 baris dan 3 kolom.

Setelah pengumpulan data, akan dilakukan tahapan prapemrosesan data yang meliputi:

1. *Data Understanding* adalah tahap memahami data yang dimiliki, yaitu berupa eksplorasi data untuk memahami variabel-variabel pada data serta korelasi antarvariabel.
2. *Data Preprocessing* adalah tahap persiapan data sebelum data digunakan untuk proses selanjutnya. Pada tahap ini, akan dilakukan penggabungan beberapa *file* sehingga menjadi satu kesatuan *file* yang utuh dan siap digunakan dalam tahap pemodelan.
3. *Data Preparation* adalah tahap mempersiapkan data dan melakukan beberapa teknik seperti mengatasi *missing value*, melakukan pembagian data menjadi data training dan data validasi.

4. Latih Model adalah tahap pembuatan model atau pembuatan sistem rekomendasi menggunakan algoritma KNN dengan teknik *collaborative filtering*.

Setelah tahap prapemrosesan data, akan dilakukan tahapan evaluasi model dengan menggunakan MAE.

3.2.2 Kriteria Pengujian

Kriteria pengujian yang dilakukan pada penelitian ini adalah pengujian pada nilai MAE dan waktu eksekusi. Perhitungan nilai MAE dipengaruhi oleh *rating* yang sebenarnya, *rating* yang diprediksi oleh sistem, dan jumlah data yang dihitung.

3.2.3 Format Data Pengujian

Format data yang digunakan dalam pengujian adalah data berformat *.csv*.

3.2.4 Alat yang Digunakan

Untuk melaksanakan penelitian dengan judul Implementasi Algoritma *K-Nearest Neighbors* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering* ini, penulis menggunakan *hardware* sebagai berikut:

1. Processor Intel(R) Core(TM) i3-1005G1 CPU @ 1.20GHz 1.19 GHz
2. RAM 8 GB
3. SSD 256 GB

Software yang digunakan adalah sebagai berikut:

1. Sistem Operasi Windows 11 64 bit
2. *Google Colab*
3. *Python*

3.2.5 Pengujian Penelitian

Pengujian nilai MAE dilakukan dengan menghitung selisih dari nilai *rating* sebenarnya dan nilai prediksi *rating* yang diberikan oleh sistem. Pengujian penelitian dilakukan untuk mengetahui nilai *error* atau kesalahan dari model yang telah dibuat. Semakin kecil nilai MAE, semakin baik model tersebut dalam memprediksi *rating* dan memberikan rekomendasi. Perhitungan nilai MAE dapat dilihat pada persamaan 2.1.

3.2.6 Analisis Hasil Pengujian

Analisis hasil pengujian dilakukan dengan melakukan analisa terhadap nilai MAE dan waktu eksekusi. Analisis hasil pengujian dipengaruhi oleh kemiripan antarnilai prediksi sistem dan nilai sebenarnya. Semakin mirip antara nilai prediksi sistem dan nilai *rating* sebenarnya, maka akurasi pada sistem semakin baik. Pada MAE, semakin besar nilai MAE yang dihasilkan menandakan hasil prediksi sistem semakin buruk. Semakin kecil nilai MAE dan mendekati nol, maka prediksi sistem semakin akurat.

3.3 Perhitungan Teknik *Item Based Collaborative Filtering* (IBCF)

Subbab ini akan menjelaskan tentang penerapan algoritma ke perhitungan dengan sampel data. Berikut adalah sampel data yang akan digunakan untuk perhitungan manual:

Tabel 3.4 Sampel Data – *Item Rating*

	CM	CC	DN	MU	MS	LA	AF	TT	JD	KW
Ica	10	10	8	8	10	10	10	8		
Nur		7	8			7	8		8	6
Goo	6		7	7	6	7	6	4		5
Tiwi		8		8	8	10	10	8	9	7
Kia	7	8	8	8	8	8	9	7	4	

Keterangan:

CM = Classical Mythology

CC = Clara Callan

DN = Decision in Normandy

MU = The Mummies of Urumchi

MS = The Middle Stories

LA = Lying Awake

AF = Airframe

TT = The Testament

JD = Jane Doe

KW = The Kitchen God's Wife

Berdasarkan tabel 3.4, dapat dilihat bahwa pada perhitungan manual ini akan digunakan 5 data user dengan 10 *items*, dalam hal ini buku.

Proses pertama adalah proses perhitungan untuk menemukan kesamaan antara dua buku yang telah dinilai pengguna. Bagian ini menjelaskan perhitungan antarbuku pada tabel 3.4, yaitu antara Classical Mythology dan Clara Callan.

Langkah pertama adalah mendapatkan semua data *rating* dari semua pengguna kedua buku tersebut, seperti terlihat pada tabel 3.5 di bawah ini.

Tabel 3.5 Data *Rating* Perhitungan Teknik IBCF

	CM	CC
Ica	10	10
Nur		7
Goo	6	
Tiwi		8
Kia	7	8

Berdasarkan tabel 3.5, dapat dilihat bahwa pengguna atas nama Ica dan Kia telah memberikan *rating* pada kedua buku.

Langkah kedua yang dilakukan adalah menghitung nilai *similarity* buku Classical Mythology dan buku Clara Callan dengan persamaan 2.1. Langkah perhitungan *similarity* adalah sebagai berikut:

1. Menghitung rata-rata *rating* yang dimiliki oleh buku:

Rata-rata *rating* buku Classical Mythology =

$$\frac{10 + 6 + 7}{3} = 7,67$$

Rata-rata *rating* buku Clara Callan =

$$\frac{10 + 7 + 8 + 8}{4} = 8,25$$

2. Mengambil semua interseksi *rating* antara buku Classical Mythology dan buku Clara Callan:

Tabel 3.6 Data interseksi antara buku CM dan buku CC

	CM	CC
Ica	10	10
Kia	7	8

Berdasarkan tabel 3.6, dapat dilihat bahwa pengguna atas nama Ica dan Kia adalah pengguna yang telah memberikan *rating* terhadap kedua buku sehingga data dari kedua pengguna inilah yang diambil sebagai data interseksi.

3. Menghitung *similarity* dengan rumus *Pearson Correlation* :

PC (Classical Mythology, Clara Callan) =

$$\frac{(10 - 7,67)(10 - 8,25) + (7 - 7,67)(8 - 8,25)}{\sqrt{(10 - 7,67)^2 + (7 - 7,67)^2} \sqrt{(10 - 8,25)^2 + (8 - 8,25)^2}} = \frac{4,25}{4,29} = 0,99$$

Setelah dilakukan perhitungan kemiripan didapatkan hasil *similarity* buku Classical Mythology dan buku Clara Callan adalah 0,99. Perhitungan kemiripan buku yang satu dengan buku yang lain juga dilakukan dengan langkah-langkah perhitungan yang sama sehingga menghasilkan matriks kemiripan, seperti terlihat pada tabel 3.7 di bawah ini.

Tabel 3.7 Matriks Kemiripan Buku

	CM	CC	DN	MU	MS	LA	AF	TT	JD	KW
CM	1,00	0,99	0,68	0,68	0,96	0,99	0,84	0,82	1,00	1,00
CC	0,99	1,00	0,07	0,40	0,98	0,72	0,60	0,57	-0,21	-0,20
DN	0,68	0,07	1,00	1,00	0,85	0,45	0,85	0,97	-0,45	0,95
MU	0,68	0,40	1,00	1,00	0,82	0,75	0,96	0,97	-0,20	0,89
MS	0,96	0,98	0,85	0,82	1,00	0,79	0,86	0,86	-1,00	0,71

LA	0,99	0,72	0,45	0,75	0,79	1,00	0,87	0,88	0,37	0,83
AF	0,84	0,60	0,85	0,96	0,86	0,87	1,00	1,00	0,17	0,94
TT	0,82	0,57	0,97	0,97	0,86	0,88	1,00	1,00	0,38	0,94
JD	1,00	-0,21	-0,45	-0,20	-1,00	0,37	0,17	0,38	1,00	0,89
KW	1,00	-0,20	0,95	0,89	0,71	0,83	0,94	0,94	0,89	1,00

Tabel 3.7 menunjukkan nilai kemiripan antara satu buku dengan buku lainnya pada skala 1 hingga -1.

Perhitungan prediksi *rating* dilakukan pada tahap ketiga. Prediksi yang dibuat pada tahap ini, yang merupakan yang terakhir dari teknik ini, dihitung dengan menggunakan nilai kesamaan dari tahap sebelumnya.

Pada tabel 3.4 diasumsikan bahwa Tiwi belum pernah memberikan *rating* terhadap CM sehingga data tersebut akan dijadikan data uji seperti yang tersaji pada tabel 3.8:

Tabel 3.8 Contoh Data Uji

	CM	CC	DN	MU	MS	LA	AF	TT	JD	KW
Tiwi	?	8	?	8	8	10	10	8	9	7

Berdasarkan tabel 3.8, dapat dilihat bahwa pengguna atas nama Tiwi belum memberikan *rating* pada CM dan DN dan dianggap sebagai data yang masih kosong. Diketahui bahwa data yang kosong sudah diberikan penilaian, namun CM dan DN akan dijadikan sebagai data uji, dan Tiwi diasumsikan belum memberikan penilaian terhadap kedua buku tersebut. Pada kenyataannya, penilaian yang telah Tiwi berikan terhadap CM adalah 8 dan terhadap DN adalah 9.

Langkah selanjutnya adalah menentukan jumlah *neighbor* untuk melakukan perhitungan. Dalam contoh ini, akan digunakan 6 *neighbor* pada

perhitungan prediksi. Tahapan berikutnya adalah menentukan kondisi dalam menentukan jumlah *neighbor*, yaitu:

- a. *Neighbor* buku yang sudah dinilai Tiwi.
- b. Kemiripan antara CM dengan N-nya 0. Pada tabel 3.7 dapat dilihat bahwa ada 9 N yang nilainya lebih besar dari 0, yaitu Clara Callan, Decision in Normandy, The Mummies of Urumchi, The Middle Stories, Lying Awake, Airframe, The Testament, Jane Doe, dan The Kitchen God's Wife. Diketahui juga bahwa terdapat nilai kemiripan yang sama, yaitu CM dengan DN dan CM dengan The Mummies of Urumchi dengan kemiripan 0.68. Akan tetapi, karena Tiwi belum memberikan penilaian pada DN, maka kemiripan yang dipakai adalah kemiripan CM dengan MU.

Langkah berikutnya, yaitu melakukan perhitungan prediksi menggunakan *Weighted Sum* sesuai dengan persamaan 2.2.

Prediksi *rating* Tiwi terhadap buku Classical Mythology:

$$\frac{(7 \times 1,0) + (9 \times 1,0) + (10 \times 0,99) + (8 \times 0,99) + (8 \times 0,96) + (10 \times 0,84)}{|(1,0 + 1,0 + 0,99 + 0,99 + 0,96 + 0,84)|}$$

$$= 8,63$$

Hasil prediksi *rating* yang akan diberikan Tiwi terhadap buku Classical Mythology adalah 8,63 sedangkan *rating* aslinya adalah 8. Selisih inilah yang akan dihitung untuk mendapatkan seberapa besar nilai *Mean Absolute Error*.

$$MAE = \frac{|8,63 - 8|}{1} = 0,63$$

Berdasarkan perhitungan di atas, dapat dilihat bahwa nilai MAE dari prediksi *rating* Tiwi terhadap buku Classical Mythology adalah sebesar 0,63.

BAB V KESIMPULAN DAN SARAN

4.1 Kesimpulan

Setelah dilakukan pengujian pada *Book Recommendation Dataset* dengan algoritma *K-Nearest Neighbors*, maka dapat disimpulkan bahwa:

1. Metode *Item-Based Collaborative Filtering* berhasil diimplementasikan dalam memberikan rekomendasi buku.
2. Dengan menggunakan *rating* berskala 1 hingga 10, *Item-Based Collaborative Filtering* mampu memberikan prediksi *rating* yang cukup akurat dan mampu memberikan rekomendasi.
3. Dengan menggunakan nilai *K* yang berbeda, dapat disimpulkan bahwa semakin tinggi nilai *K* maka *Mean Absolute Error* yang dihasilkan juga semakin kecil dan semakin mendekati nol yang berarti nilai akurasi semakin tinggi.
4. Nilai *K* berpengaruh pada waktu eksekusi untuk mendapatkan hasil rekomendasi.
5. Hasil penelitian ini menunjukkan bahwa algoritma *K-Nearest Neighbors* dapat digunakan sebagai algoritma dalam memberikan rekomendasi *item* pada sistem rekomendasi, yang mana mampu memberikan *top-N recommendations* kepada pengguna.
6. Penelitian ini juga membuktikan bahwa untuk mendapatkan nilai kemiripan *item* pada implementasi Algoritma *K-Nearest Neighbors* dapat dilakukan dengan menggunakan rumus *Pearson Correlation*.

4.2 Saran

Berdasarkan penelitian yang telah dilakukan, penulis menemukan beberapa hal yang dapat dikembangkan untuk penelitian selanjutnya, yaitu:

1. Perhitungan *similarity* menggunakan *Pearson Correlation* memiliki kelemahan yaitu ketika hasil perhitungan pada rumus ini menghasilkan 0, maka tidak dapat dihitung nilai kemiripannya. Hal ini disebabkan karena *Pearson Correlation* menghitung *similarity* berdasarkan selisih *rating* asli dengan *rating* rata-rata. Penelitian selanjutnya dapat menggunakan metode perhitungan *similarity* lain seperti *cosine similarity* atau dengan perhitungan *distance*.
2. Dapat melakukan penelitian dengan menggunakan algoritma *K-Nearest Neighbors* yang ada pada *library scikit-learn* dan *tensorflow*.
3. Menggunakan cara atau perhitungan lain untuk mendapatkan nilai *K* yang optimal.
4. Dapat menggunakan data latih dengan jumlah yang lebih banyak untuk mendapatkan akurasi yang lebih baik atau nilai *error* yang lebih rendah.
5. Sistem rekomendasi yang telah dibuat belum dapat merekomendasikan *item* yang belum memiliki *rating*. Penelitian selanjutnya dapat menggunakan teknik yang lain seperti *content based filtering*, *hybrid recommendation*, atau teknik lainnya yang dapat merekomendasikan *item* yang belum memiliki *rating*.

DAFTAR PUSTAKA

- Aryani, B. S. (2019, Maret). Perancangan Sistem Rekomendasi Pemilihan Cinderamata Khas Bengkulu Berbasis E-Marketplace. *Jurnal Rekursif*, 7.
- Aryani, B. S. (2021). Sistem Rekomendasi Menggunakan Item-based Collaborative Filtering pada Konten Artikel Berita. *JURNALTIO*, 2.
- Ayyadevara, V. K. (2018). Recommender Systems. In *Pro Machine Learning Algorithms* (pp. 299–325). Apress. https://doi.org/10.1007/978-1-4842-3564-5_13
- Banik, R. (2018). *Hands-On Recommendation Systems with Python: Start Building Powerful and Personalized, Recommendation Engines with Python*. (S. Editing (ed.); 1st ed.). Packt Publishing.
- Chaudhuri, A. B. (2020). *Flowchart and Algorithm Basics*. Mercury Learning and Information.
- Darwati, I. (2017). Rancang Bangun Peminjaman dan Pengembalian Buku pada Perpustakaan Sekolah Dasar. 2(1), 139–144.
- Dharma, A. S., & Samosir, T. (2019). The User Personalization with KNN for Recommender System. 3(2), 45–48.
- Fitrianti, A. R., Rohmani, A., & Widjanarto. (2020). Sistem Rekomendasi Film Berbasis Website dengan Metode Prototype Menggunakan Algoritma K-Nearest Neighbors (KNN). *JOINS (Journal of Information System)*, 5(2), 278–287. <https://doi.org/10.33633/joins.v5i2.4168>
- Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018). A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System. *Journal of Physics: Conference Series*, 1000(1). <https://doi.org/10.1088/1742-6596/1000/1/012101>
- Gupta, S. C., & Goel, N. (2020). Performance Enhancement of Diabetes Prediction by Finding Optimum K for KNN Classifier with Feature Selection Method. *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020, Iccsit*, 980–986. <https://doi.org/10.1109/ICSSIT48917.2020.9214129>
- Gusti, I. G., Nasrun, M., & Astuti Nugrahaeni, R. (2019). Rekomendasi Sistem Pemilihan Mobil Menggunakan K-Nearest Neighbor (KNN) Collaborative Filtering. *TEKTRIKA - Jurnal Penelitian Dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, Dan Elektronika*, 4(1), 26. <https://doi.org/10.25124/tektrika.v4i1.1846>
- Hwangbo, H., Kim, Y. S., & Cha, K. J. (2018). Recommendation System Development for Fashion Retail E-commerce. *Electronic Commerce*

Research and Applications, 28, 94–101.
<https://doi.org/10.1016/j.elerap.2018.01.012>

Khairina, N., Tri Saputra Sibarani, T., Muliono, R., Sembiring, Z., & Muhathir. (2022). Identification of Pneumonia using The K-Nearest Neighbors Method using HOG Feature Extraction. *JITE (Journal of Informatics and Telecommunication Engineering)*, 5(January).
<http://ojs.uma.ac.id/index.php/jite>

Melville, P., & Sindhvani, V. (2017). Encyclopedia of Machine Learning and Data Mining. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning and Data Mining*. Springer US. <https://doi.org/10.1007/978-1-4899-7687-1>

Mohanty, S. N., Chatterjee, J. M., Jain, S., Elngar, A. A., & Gupta, P. (2020). Recommender System with Machine Learning and Artificial Intelligence. In S. N. Mohanty, J. M. Chatterjee, S. Jain, A. A. Elngar, & P. Gupta (Eds.), *Recommender System with Machine Learning and Artificial Intelligence*. Wiley. <https://doi.org/10.1002/9781119711582>

Mondi, R. H., Wijayanto, A., & Winarno. (2019). Recommendation System With Content-Based Filtering Method for Culinary Tourism in Mangan Application. *ITSMART: Jurnal Ilmiah Teknologi Dan Informasi*, 8(2), 65–72.

Muhathir, Tri Saputra Sibarani, T., & Al-Khowarizmi. (2020). Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction. *International of Computer Science and Information Technology (AIOCSIT) JournalISSN*, 1(1), 33–38.

Nasuha, D. E., Yusliani, N., & Marieska, M. D. (2021). *Sistem Rekomendasi Buku Menggunakan Item-Based Clustering Hybrid Method (Ic hm)*.
<https://repository.unsri.ac.id/40359/>

Nugroho, R., Polina, A., & Mahendra, Y. (2020). Tourism Site Recommender System Using Item-Based Collaborative Filtering Approach. *International Journal of Applied Sciences and Smart Technologies*, 2(2), 209–216.
<https://doi.org/10.24071/ijasst.v2i2.2987>

Prasetya, C. S. D. (2017). Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 4(3), 194. <https://doi.org/10.25126/jtiik.201743392>

Riswanto, E., Robi'in, B., & Suparyanto. (2019). Mobile Recommendation System for Culinary Tourism Destination using KNN (K-Nearest Neighbor). *Journal of Physics: Conference Series*, 1201(1), 012039.
<https://doi.org/10.1088/1742-6596/1201/1/012039>

Sattar, A., Ghazanfar, M. A., & Iqbal, M. (2017). Building Accurate and Practical Recommender System Algorithms Using Machine Learning Classifier and

- Collaborative Filtering. *Arabian Journal for Science and Engineering*, 42(8), 3229–3247. <https://doi.org/10.1007/s13369-016-2410-1>
- Selvi G., C., & Priya G.G., L. (2020). Rating Prediction Method for Item-based Collaborative Filtering Recommender Systems Using Formal Concept Analysis. *EAI Endorsed Transactions on Energy Web*, 8(33). <https://doi.org/10.4108/eai.2-12-2020.167291>
- Setiawan, Y., Nurwanto, A., & Erlansari, A. (2019). Implementasi Item Based Collaborative Filtering Dalam Pemberian Rekomendasi Agenda Wisata Berbasis Android. *Jurnal Pseudocode*, 6(1), 13–20. <https://doi.org/10.33369/pseudocode.6.1.13-20>
- Sukmana, H. T., Atinah, S., & Wardhani, L. K. (2019). Implementasi Content-Based Filtering Pada Aplikasi Radar Zakat Dalam Merekomendasikan Preferensi Mustahik. *Jurnal Teknik Informatika*, 12(2), 167–176. <https://doi.org/10.15408/jti.v12i2.13172>
- Syah, R. D. (2020). Performa Algoritma User K-Nearest Neighbors pada Sistem Rekomendasi di Tokopedia. *Jurnal Informatika Universitas Pamulang*, 5(3), 302–306. <https://doi.org/10.32493/informatika.v5i3.6312>
- Theobald, O. (2018). *Machine Learning for Beginners*. Scatterplot Press.
- Wahyudi, I. S. (2017). *Mesin Rekomendasi Film Menggunakan Metode Kemiripan Genre Berbasis Collaborative Filtering*.
- Wang, W., & Lu, Y. (2018). Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model. *IOP Conference Series: Materials Science and Engineering*, 324(1), 012049. <https://doi.org/10.1088/1757-899X/324/1/012049>
- Wardana, A. S., & Ananta Timur, M. I. (2018). Collaborative Filtering Recommender System pada Virtual 3D Kelas Cendekia. *IJEIS (Indonesian Journal of Electronics and Instrumentation Systems)*, 8(1), 73. <https://doi.org/10.22146/ijeis.28729>
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2020). Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Computing Surveys*, 52(1), 1–38. <https://doi.org/10.1145/3285029>

LAMPIRAN – LAMPIRAN

Dataset				
Data <i>Books.csv</i>				
ISBN	Book Title	Book Author	Year of Publication	Publisher
195153448	Classical Mythology	Mark P. O. Morford	2002	Oxford University Press
2005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada
60973129	Decision in Normandy	Carlo D'Este	1991	HarperPerennial
374157065	Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It	Gina Bari Kolata	1999	Farrar Straus Giroux
393045218	The Mummies of Urumchi	E. J. W. Barber	1999	W. W. Norton & Company
399135782	The Kitchen God's Wife	Amy Tan	1991	Putnam Pub Group
425176428	What If?: The World's Foremost Military Historians Imagine What Might Have Been	Robert Cowley	2000	Berkley Publishing Group
671870432	PLEADING GUILTY	Scott Turow	1993	Audioworks
679425608	Under the Black Flag: The Romance and the Reality of Life Among the Pirates	David Cordingly	1996	Random House
074322678X	Where You'll Find Me: And Other Stories	Ann Beattie	2002	Scribner
771074670	Nights Below Station Street	David Adams Richards	1988	Emblem Editions
080652121X	Hitler's Secret Bankers: The Myth of Swiss Neutrality During the Holocaust	Adam Lebor	2000	Citadel Press
887841740	The Middle Stories	Sheila Heti	2004	House of Anansi Press
1552041778	Jane Doe	R. J. Kaiser	1999	Mira Books
1558746218	A Second Chicken Soup for the Woman's Soul (Chicken Soup for the Soul Series)	Jack Canfield	1998	Health Communications
1567407781	The Witchfinder	Loren D.	1998	Brilliance Audio -

	(Amos Walker Mystery Series)	Estleman		Trade
1575663937	More Cunning Than Man: A Social History of Rats and Man	Robert Hendrickson	1999	Kensington Publishing Corp.
1881320189	Goodbye to the Buttermilk Sky	Julia Oliver	1994	River City Pub
440234743	The Testament	John Grisham	1999	Dell
452264464	Beloved (Plume Contemporary Fiction)	Toni Morrison	1994	Plume
609804618	Our Dumb Century: The Onion Presents 100 Years of Headlines from America's Finest News Source	The Onion	1999	Three Rivers Press
1841721522	New Vegetarian: Bold and Beautiful Recipes for Every Occasion	Celia Brooks Brown	2001	Ryland Peters & Small Ltd
1879384493	If I'd Known Then What I Know Now: Why Not Learn from the Mistakes of Others? : You Can't Afford to Make Them All Yourself	J. R. Parrish	2003	Cypress House
61076031	Mary-Kate & Ashley Switching Goals (Mary-Kate and Ashley Starring in)	Mary-Kate & Ashley Olsen	2000	HarperEntertainment
439095026	Tell Me This Isn't Happening	Robynn Clairday	1999	Scholastic
689821166	Flood : Mississippi 1927	Kathleen Duey	1998	Aladdin
971880107	Wild Animus	Rich Shapero	2004	Too Far
345402871	Airframe	Michael Crichton	1997	Ballantine Books
345417623	Timeline	MICHAEL CRICHTON	2000	Ballantine Books
684823802	OUT OF THE SILENT PLANET	C.S. Lewis	1996	Scribner
375759778	Prague : A Novel	ARTHUR PHILLIPS	2003	Random House Trade Paperbacks
425163091	Chocolate Jesus	Stephan Jaramillo	1998	Berkley Publishing Group
3404921038	Wie Barney es sieht.	Mordecai Richler	2002	LÄf?Ä,Ä¼bbe
3442353866	Der Fluch der Kaiserin. Ein Richter- Di- Roman.	Eleanor Cooney	2001	Goldmann
3442410665	Sturmzeit. Roman.	Charlotte Link	1991	Goldmann
3442446937	Tage der Unschuld.	Richard North	2000	Goldmann

		Patterson		
375406328	Lying Awake	Mark Salzman	2000	Alfred A. Knopf
446310786	To Kill a Mockingbird	Harper Lee	1988	Little Brown & Company
449005615	Seabiscuit: An American Legend	LAURA HILLENBRAND	2002	Ballantine Books
60168013	Pigs in Heaven	Barbara Kingsolver	1993	Harpercollins
038078243X	Miss Zukas and the Raven's Dance	Jo Dereske	1996	Avon
055321215X	Pride and Prejudice	Jane Austen	1983	Bantam
067176537X	The Therapeutic Touch: How to Use Your Hands to Help or to Heal	Dolores Krieger	1979	Fireside
61099686	Downtown	Anne Rivers Siddons	1995	HarperTorch
553582909	Icebound	Dean R. Koontz	2000	Bantam Books
671888587	I'll Be Seeing You	Mary Higgins Clark	1994	Pocket
553582747	From the Corner of His Eye	Dean Koontz	2001	Bantam Books
425182908	Isle of Dogs	Patricia Cornwell	2002	Berkley Publishing Group
042518630X	Purity in Death	J.D. Robb	2002	Berkley Publishing Group
440223571	This Year It Will Be Different: And Other Stories	Maeve Binchy	1997	Dell

Data Ratings.csv

User ID	ISBN	Book Rating
276725	034545104X	0
276726	155061224	5
276727	446520802	0
276729	052165615X	3
276729	521795028	6
276733	2080674722	0
276736	3257224281	8
276737	600570967	6
276744	038550120X	7
276745	342310538	10
276746	425115801	0
276746	449006522	0
276746	553561618	0
276746	055356451X	0
276746	786013990	0
276746	786014512	0
276747	60517794	9

276747	451192001	0
276747	609801279	0
276747	671537458	9
276747	679776818	8
276747	943066433	7
276747	1570231028	0
276747	1885408226	7
276748	747558167	6
276748	3442437407	0
276751	033390804X	0
276751	3596218098	8
276754	684867621	8
276755	451166892	5
276760	8440682697	10
276762	034544003X	0
276762	380000059	0
276762	380711524	5
276762	451167317	0
276762	451454952	0
276762	843920262	0
276762	3404122879	0
276762	3404182928	0
276762	3404611306	0
276762	342662429	0
276762	3426690179	0
276762	3442424216	0
276762	3442425573	0
276762	3453092007	8
276762	3453157745	0
276762	3453176944	0
276762	3453185137	0
276762	3453185323	0
276762	3453213025	3

Data Users.csv

User ID	Location	Age
2	stockton, california, usa	18
4	porto, v.n.gaia, portugal	17
6	santa monica, california, usa	61
10	albacete, wisconsin, spain	26
11	melbourne, victoria, australia	49
13	barcelona, barcelona, spain	59
18	rio de janeiro, rio de janeiro, brazil	37
19	weston, ,	15
20	langhorne, pennsylvania, usa	28

21	ferrol / spain, alabama, spain	44
24	cologne, nrw, germany	39
25	oakland, california, usa	35
27	chicago, illinois, usa	36
28	freiburg, baden-wuerttemberg, germany	15
29	cuernavaca, alabama, mexico	20
30	anchorage, alaska, usa	47
31	shanghai, n/a, china	43
33	costa mesa, california, usa	31
35	grifton, wisconsin, usa	46
36	montreal, quebec, canada	14
37	san sebastian, n/a, spain	36
38	viterbo, lazio, italy	41
40	tonawanda, new york, usa	35
41	santee, california, usa	48
42	appleton, wisconsin, usa	55
44	black mountain, north carolina, usa	27
46	heidelberg, baden-wuerttemberg, germany	21
47	vicenza, veneto, italy	52
50	london, england, united kingdom	45

Source Code

```
# -*- coding: utf-8 -*-
"""book_recommendation

##### Nama : [Kori Isabella
Hutabarat] (https://www.linkedin.com/in/koriisabellaa/)
##### Email : corryissabella@gmail.com
##### Sumber Data : https://www.kaggle.com/datasets/arashnic/book-
recommendation-dataset

Import library yang akan digunakan, yaitu NumPy untuk memproses
operasi matematika yang akan digunakan pada dataframe dan Pandas
untuk membangun dataframe dari file, dalam hal ini file .csv.
Files digunakan untuk mengupload kaggle.json untuk menghubungkan
Google Colab dengan Kaggle.
"""

import numpy as np
import pandas as pd

from google.colab import files

"""# Kaggle Setup"""

#install kaggle
!pip install -q kaggle

#unggah file json yang diunduh dari akun kaggle
uploaded = files.upload()

"""Buat direktori Kaggle dan pindahkan file yang diunggah ke
folder baru. Kemudian berikan izin baca agar bisa diakses di
Google Colab."""

!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle

!chmod 600 /root/.kaggle/kaggle.json

"""## Mengunduh dan Menyiapkan Dataset



#####**Informasi Dataset :**
```

```

Jenis | Informasi
--- | ---
Sumber | [Kaggle Dataset : Book Recommendation
Dataset] (https://www.kaggle.com/datasets/arashnic/book-
recommendation-dataset)
Lisensi | CC0: Public Domain
Kategori | Literature
Rating Penggunaan | 10.0 (Bronze)
Jenis dan Ukuran Berkas | CSV (106.94 MiB)
"""

!kaggle datasets download -d arashnic/book-recommendation-dataset

!unzip -q /content/book-recommendation-dataset.zip -d .

"""# Univariate Exploratory Data Analysis

Pada tahap ini, eksplorasi data dilakukan untuk memahami variabel-
variabel dalam data dan korelasi antarvariabel. Eksplorasi
dilakukan terhadap tiap
file pada dataset yaitu Books.csv, Ratings.csv, dan Users.csv.
"""

# Memuat data pada sebuah dataframe menggunakan Pandas

Books = pd.read_csv('/content/Books.csv')
Ratings = pd.read_csv('/content/Ratings.csv')
Users = pd.read_csv('/content/Users.csv')

Books.head(10)

Books.info()

Books = Books.drop(columns=['Image-URL-S', 'Image-URL-M', 'Image-
URL-L'])

Books.head(10)

print('Bentuk data (baris, kolom):'+ str(Books.shape))
print('Bentuk data (baris, kolom):'+ str(Ratings.shape))
print('Bentuk data (baris, kolom):'+ str(Users.shape))

"""### Variabel Books"""
# Memuat informasi dataframe
Books.info()

```

```

Books.describe()

# Melihat jumlah data kosong pada setiap kolom
Books.isnull().sum()

#Drop data yang kosong pada setiap kolom
Books = Books.dropna(axis=0)
Books.isnull().sum()

print('Total jumlah ISBN:', len(Books['ISBN'].unique()))
print('Total jumlah Book-Title :', len(Books['Book-
Title'].unique()))
print('Total jumlah Publisher :',
len(Books['Publisher'].unique()))

"""### Variabel Ratings"""

Ratings.head(5)

Ratings.info()

Ratings.describe()

"""Berdasarkan output di atas, dapat disimpulkan bahwa rating
maksimum adalah 10 dan rating minimum adalah 0. Artinya, ada
rating implisit yang ditunjukkan dengan nilai 0 yang selanjutnya
akan kita anggap sebagai outlier karena skala rating yang
sebenarnya berkisar antara 1 hingga 10. """

Ratings.isnull().sum()

print('Total Jumlah User :', len(Ratings['User-ID'].unique()))
print('Total Jumlah Rating :', len(Ratings['Book-
Rating'].unique()))

"""### Variabel Users"""

Users.head(6)

Users.info()

Users.describe()

Users.isnull().sum()

# Drop data yang kosong pada setiap kolom
Users = Users.dropna(axis=0)

```

```

Users.info()

Users.isnull().sum()

print('Total Jumlah User :', len(Users['User-ID'].unique()))
print('Total Jumlah Lokasi :', len(Users['Location'].unique()))
print('Total Jumlah Umur:', len(Users['Age'].unique()))

"""# Data Preprocessing

Pada proses ini dilakukan penggabungan data pada variabel Books
dan variabel Ratings dengan merge pada kondisi left yaitu menjaga
semua baris dari dataframe pertama dan menambahkan kolom apa pun
yang cocok berdasarkan kolom ISBN di dataframe kedua.
"""

book_rating = pd.merge(Books, Ratings, on='ISBN', how='left')

book_rating.head(5)

book_rating.isnull().sum()

"""Pada proses ini dilakukan penggabungan data pada variabel
book_rating dan variabel Users dengan merge pada kondisi left
yaitu menjaga semua baris dari dataframe pertama dan menambahkan
kolom apa pun yang cocok berdasarkan kolom User-ID di dataframe
kedua."""

books = pd.merge(book_rating, Users, on='User-ID', how='left')
books

books.head(2)

books = books[['User-ID', 'Age', 'Location', 'ISBN', 'Book-
Title', 'Book-Author', 'Year-Of-
Publication', 'Publisher', 'Book-Rating']]
books.head()

books.shape

"""# Data Preparation

Pada tahap ini dilakukan persiapan data dan beberapa teknik
seperti mengatasi missing value dan menghapus NULL data.
"""

books.isnull().sum()

```



```

books = books.dropna()

books.isnull().sum()

books.shape

books.columns

"""Nama kolom di-rename sesuai keinginan, untuk lebih
merekpresentasikan kolom atau untuk memudahkan penulisan nama kolom
pada tahapan selanjutnya."""

books.rename(columns={'User-ID':'userID', 'Book-Title':'Title',
'Book-Author':'Author', 'Year-Of-
Publication':'PublicationYear', 'Book-Rating':'Rating'},
inplace=True)
books

"""Pada data preparation, akan diperiksa rating dengan nilai < 3
yang kemudian akan dihapus untuk kepentingan distribusi rating
agar lebih merata."""

# cek data yg memiliki rating <3
print('Jumlah rating < 3 :', books['Rating'].lt(3).sum())
books.shape

# menghilangkan data yang memiliki rating <3
books = books[books['Rating']>=3]
print('Jumlah total rating < 3 :', books['Rating'].lt(3).sum())
books.shape

"""Untuk mempercepat proses pemodelan, data difilter hanya dengan
tahun 2003 dan 2004"""

# filter data dengan tahun 2003 dan 2004
books['PublicationYear'] = books['PublicationYear'].astype('str')
books = books[books['PublicationYear'].str.contains("2003|2004")]

"""Cek data duplikat pada kolom ISBN lalu bersihkan seluruh data
duplikat"""

books_unique = books.drop_duplicates('ISBN')
books_unique

"""# Data Visualization

```

```

Visualisasi data bertujuan untuk mendapatkan insights dari
dataset. Pada dataset ini, akan dilakukan visualisasi rating, top
author, dan top publisher
"""

import seaborn as sns

with sns.axes_style('white'):
    g = sns.catplot('Rating', data=books, aspect=2.0,
kind='count')
    g.set_ylabels('Total Rating')

import matplotlib.pyplot as plt

plt.subplots(figsize=(12,10))

ax=pd.Series(books_unique['Author']).value_counts()[:10].sort_valu
es(ascending=True).plot.barh(width=0.9,color=sns.color_palette('mu
ted',40))
for i, v in
enumerate(pd.Series(books_unique['Author']).value_counts()[:10].so
rt_values(ascending=True).values):
    ax.text(.8, i, v,fontsize=10,color='white',weight='bold')
plt.title('Top Author')
plt.show()

plt.subplots(figsize=(12,10))

ax=pd.Series(books_unique['Publisher']).value_counts()[:10].sort_v
alues(ascending=True).plot.barh(width=0.9,color=sns.color_palette(
'muted',40))
for i, v in
enumerate(pd.Series(books_unique['Publisher']).value_counts()[:10]
.sort_values(ascending=True).values):
    ax.text(.8, i, v,fontsize=10,color='white',weight='bold')
plt.title('Top Publishers')
plt.show()

"""# Remove unused column

Drop kolom yang tidak digunakan pada pemodelan. Variabel
books_unique akan digunakan pada tahapan implementasi KNN. Tujuan
menggunakan variabel books_unique adalah untuk memperkecil jumlah
data dan mempercepat proses pemodelan.
"""

books_unique=books_unique.drop('Age', axis=1)

```

```

books_unique=books_unique.drop('Location', axis=1)
books_unique

"""# KNN Implementation

## Creating Pivot

Pada tahapan ini, dibuat pivot tabel untuk mengkonversi tabel
menjadi matriks 2D dan mengisi missing value dengan nol karena
akan dilakukan
perhitungan distance antara vektor rating.

Pivot table dibuat dengan menggunakan variabel ratings dengan
atribut userID, ISBN, dan Rating. Adapun tujuan menggunakan
variabel baru ini adalah untuk memperkecil data agar proses
pembuatan pivot lebih ringan dan aman pada CPU.
"""

ratings = books[['userID', 'ISBN', 'Rating']]
ratings

rating_pivot = ratings.pivot(index='ISBN', columns='userID',
values='Rating').fillna(0)

"""## Model Development

Import libraries yang akan digunakan pada tahap pemodelan.
Operator digunakan untuk menentukan atribut/kolom/elemen (data
tuple) yang akan digunakan untuk sorting. Correlation digunakan
untuk mendapatkan jarak antaritem. MAE digunakan untuk evaluasi
model.
"""

import operator

from scipy.spatial.distance import correlation
from sklearn.metrics import mean_absolute_error as mae

"""Proses ini bertujuan untuk mendapatkan jarak (distance) antara
item yang dipilih dengan semua item pada dataset"""

def get_distances(target_book):
    distances = []

    isbn = target_book['ISBN'].values[0]

    for index, book in books_unique.iterrows():

```

```

        if book['ISBN'] != isbn:
            dist = correlation(rating_pivot.loc[isbn].values,
                              rating_pivot.loc[book['ISBN']].values)
            distances.append((book['ISBN'], book['Title'], dist))

        distances.sort(key=operator.itemgetter(2))
        return distances

"""### Finding the optimal K

Proses ini bertujuan untuk mendapatkan daftar rata-rata nilai MAE
dengan melatih sebanyak 50 data secara acak dengan nilai K pada
range yang ditentukan
"""

def get_error_rate(K_maks):
    result = []
    avg_error = 0

    for j in range(50):
        index = np.random.randint(len(books_unique))
        new_book = books_unique.iloc[index].to_frame().T
        distances = get_distances(new_book)

        total_rating = 0

        for i in range(K_maks):
            if j == 0:
                result.append([])

            neighbor = distances[i]
            current_book =
books_unique[books_unique['ISBN'].str.contains(neighbor[0])].iloc[
0]

            total_rating = total_rating + current_book[6]

            avg_rating = total_rating/(i+1)
            error = mae([new_book['Rating']], [avg_rating])

            result[i].append(error)

        for i in range(K_maks):
            result[i] = np.mean(result[i])

    return result

"""Proses ini bertujuan untuk mendapatkan nilai K optimal pada K =

```

```

11 - K = 173, dengan berdasarkan data latih sebanyak 50 data, dan
nilai K optimal yang didapatkan adalah K = 15"""

import matplotlib.pyplot as plt

rate = 173

error_rate = get_error_rate(rate)
error_rate_range = error_rate[10:]

K = error_rate.index(min(error_rate_range))+1

plt.figure(figsize=(10,6))
plt.plot(range(11,rate+1), error_rate_range, color='blue',
linestyle='dashed',
marker='o', markerfacecolor='red', markersize=10)
plt.title('Error Rate vs. K Value')
plt.xlabel('K')
plt.ylabel('Error Rate')
print("Minimum error:-", min(error_rate_range), "at K =", K)

"""

# Predictor

Proses ini bertujuan untuk mendapatkan Top-10 Recommendations dan
memprediksi rating, serta nilai MAE sesuai dengan nilai K optimal.
"""

def predict(query):
    new_book =
books_unique[books_unique['Title'].str.contains(query)].iloc[0].to
_frame().T
    print('Selected Book: ', new_book.Title.values[0])

    total_rating = 0
    distances = get_distances(new_book)

    print('\nTop 10 Recommended Books: \n')

    for i in range(K):
        neighbor = distances[i]
        current_book =
books_unique[books_unique['ISBN'].str.contains(neighbor[0])].iloc[
0]

```

```

        total_rating = total_rating + current_book[6]
        if i < 10:
            print(current_book[2]+" | Author:
"+str(current_book[3])+" | Rating: "+str(current_book[6]))

        print('\n')
        avg_rating = total_rating/K
        print('The predicted rating for %s is: %f'
%(new_book['Title']. values[0], avg_rating))
        print('The actual rating for %s is: %f' %(new_book['Title'].
values[0], new_book['Rating']))

        error = mae([new_book['Rating']], [avg_rating])

        # display
        print("Mean absolute error: " + str(error))

predict("Hard")

import matplotlib.pyplot as plt

rate = 179

error_rate = get_error_rate(rate)
error_rate_range = error_rate[10:]

K = error_rate.index(min(error_rate_range))+1

plt.figure(figsize=(10,6))
plt.plot(range(11,rate+1), error_rate_range, color='blue',
linestyle='dashed',
        marker='o', markerfacecolor='red', markersize=10)
plt.title('Error Rate vs. K Value')
plt.xlabel('K')
plt.ylabel('Error Rate')
print("Minimum error:-", min(error_rate_range), "at K =", K)

""""""

def predict(query):
    new_book =
books_unique[books_unique['Title'].str.contains(query)].iloc[0].to
_frame().T
    print('Selected Book: ', new_book.Title.values[0])

```

```

total_rating = 0
distances = get_distances(new_book)

print('\nTop 10 Recommended Books: \n')

for i in range(K):
    neighbor = distances[i]
    current_book =
books_unique[books_unique['ISBN'].str.contains(neighbor[0])].iloc[
0]
    total_rating = total_rating + current_book[6]
    if i < 10:
        print(current_book[2]+" | Author:
"+str(current_book[3])+" | Rating: "+str(current_book[6]))

    print('\n')
    avg_rating = total_rating/K
    print('The predicted rating for %s is: %f'
%(new_book['Title'].values[0], avg_rating))
    print('The actual rating for %s is: %f' %(new_book['Title'].
values[0], new_book['Rating']))

    error = mae([new_book['Rating']], [avg_rating])

    # display
    print("Mean absolute error: " + str(error))

predict("Hard")

rate = 167

error_rate = get_error_rate(rate)
error_rate_range = error_rate[10:]

K = error_rate.index(min(error_rate_range))+1

plt.figure(figsize=(10,6))
plt.plot(range(11,rate+1), error_rate_range, color='blue',
linestyle='dashed',
        marker='o', markerfacecolor='red', markersize=10)
plt.title('Error Rate vs. K Value')
plt.xlabel('K')
plt.ylabel('Error Rate')
print("Minimum error: -", min(error_rate_range), "at K =", K)

"""![image](https://user-
images.githubusercontent.com/87566521/180288333-89be2716-717e-
</pre>
</div>
<div data-bbox=)UNIVERSITAS MEDAN AREA
```

```

4693-b048-99756de7f276.png">"""

def predict(query):
    new_book =
books_unique[books_unique['Title'].str.contains(query)].iloc[0].to
_frame().T
    print('Selected Book: ', new_book.Title.values[0])

    total_rating = 0
    distances = get_distances(new_book)

    print('\nTop 10 Recommended Books: \n')

    for i in range(K):
        neighbor = distances[i]
        current_book =
books_unique[books_unique['ISBN'].str.contains(neighbor[0])].iloc[
0]
        total_rating = total_rating + current_book[6]
        if i < 10:
            print(current_book[2]+" | Author:
"+str(current_book[3])+" | Rating: "+str(current_book[6]))

        print('\n')
        avg_rating = total_rating/K
        print('The predicted rating for %s is: %f'
%(new_book['Title']. values[0], avg_rating))
        print('The actual rating for %s is: %f' %(new_book['Title'].
values[0], new_book['Rating']))

    error = mae([new_book['Rating']], [avg_rating])

    # display
    print("Mean absolute error: " + str(error))

predict("Hard")

"""# Penutup

Mean Absolute Error (MAE) mengukur besarnya kesalahan pada
prediksi rating terhadap data. Semakin rendah nilai MAE (Mean
Absolute Error) maka semakin baik dan akurat model yang dibuat.

Berikut rumusnya :

![image](https://user-
images.githubusercontent.com/87566521/139152819-30500f63-40a3-

```


40ed-86fd-a62e517adbb4.png)

Hasil implementasi dari model yang telah dibangun akan digunakan untuk memberikan rekomendasi buku serta memprediksi rating yang akan diberikan oleh pengguna. Berikut adalah hasil implementasi Algoritma K-Nearest Neighbors pada sistem rekomendasi buku dengan judul buku "Hard Eight : A Stephanie Plum Novel (A Stephanie Plum Novel)":

Nilai K dan Waktu Eksekusi

```

```

Hasil Rekomendasi

```

```

Hasil Evaluasi MAE

```

"""
```



Similarity Report ID: oid:29477:20753205

PAPER NAME

[SIDANG] 188160022_KORI ISABELLA H
UTABARAT.pdf

AUTHOR

KORI ISABELLA HUTABARAT

WORD COUNT

9238 Words

CHARACTER COUNT

54559 Characters

PAGE COUNT

54 Pages

FILE SIZE

1.2MB

SUBMISSION DATE

Aug 15, 2022 11:20 PM GMT+7

REPORT DATE

Aug 15, 2022 11:21 PM GMT+7

● **28% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- 23% Internet database
- 5% Publications database
- Crossref database
- Crossref Posted Content database
- 19% Submitted Works database

● **Excluded from Similarity Report**

- Small Matches (Less than 10 words)

Summary



UNIVERSITAS MEDAN AREA

FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PSSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225802, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 111/FT.6/01.10/V/2022
Lamp : -
Hal : **Perubahan Judul Tugas Akhir**

9 Mei 2022

Yth, Pembimbing Tugas Akhir
Rizki Muliono, S.Kom, M.Kom
Zulfikar Sembiring, S. Kom, M. Kom
di
Tempat

Dengan hormat, Sehubungan dengan adanya perubahan judul tugas akhir maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa tersebut :

N a m a : Kori Isabella Hutabarat
N P M : 188160022
Jurusan : Informatika

Maka dengan hormat kami mengharapkan kesediaan saudara :

1. **Rizki Muliono, S.Kom, M.Kom** (Sebagai Pembimbing I)
2. **Zulfikar Sembiring, S. Kom, M. Kom** (Sebagai Pembimbing II)

Adapun Tugas Akhir Skripsi berjudul :

"Implementasi Algoritma *K-Nearest Neighbor* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering*".

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.



Dr. **Rahmat Salsal, S. Kom, M. Kom**



UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 (061) 7366878, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 115 /FT.6/01.10/V/2022
Lamp : -
Hal : **Penelitian Dan Pengambilan Data Tugas Akhir**

13 Mei 2022

Yth. Kepala Sekolah SMK Swasta HKBP Sibolga
Jln. Mayjend Siswomiharjo No. 6
Di
Sibolga

Dengan hormat,
Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Kori Isabella Hutabarat	188160022	Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana lengkap pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :

Implementasi Algoritma *K-Nearest Neighbors* pada Sistem Rekomendasi Buku Menggunakan Teknik *Collaborative Filtering*

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.



Dr. Rahmad Syah, S. Kom, M. Kom

Tembusan :
1. Ka. BAMAI
2. Mahasiswa
3. File



DEPARTEMEN PENDIDIKAN HURIA KRISTEN BATAK PROTESTAN
SEKOLAH MENENGAH KEJURUAN (SMK)
SMK SWASTA HKBP SIBOLGA
Jl. Mayjend Siswomiharjo No. 6 Telp. (0631) 22853 Sibolga 22521



SURAT KETERANGAN

NOMOR : 642/I 05.17/SMK.H/2022

Berdasarkan surat Dekan Fakultas Teknik Universitas Medan Area Nomor : 115/FT.6/01.10/V/2022 tanggal 13 Mei 2022 tentang "Penelitian dan Pengambilan Data Tugas Akhir", maka dengan ini Kepala SMK Swasta HKBP Sibolga menerangkan bahwa :

N a m a : KORI ISABELLA HUTABARAT
NIM : 188160022
Jenjang/Jurusan : S1/Teknik Informatika

Benar telah melaksanakan dan menyelesaikan Penelitian dan Pengambilan Data Tugas Akhir dengan judul :

"Implementasi Algoritma K-Nearest Neighbors pada Sistem Rekomendasi Buku Menggunakan Teknik Collaborative Filtering" di SMK Swasta HKBP Sibolga.

Demikian Surat Keterangan ini dibuat dengan sebenarnya untuk dapat digunakan sebagaimana mestinya.



Dra. HOTNA BATUBARA