

BAB II

TINJAUAN PUSTAKA

1.1 Literature Review

Terdapat penelitian yang telah dilakukan dalam penerapan system rekomendasi film menggunakan metode *collaborative filtering* dan content based filtering. Sehingga, dalam penelitian ini diperlukannya tinjauan pustaka sebagai alat dalam penerapan metode penelitian ini untuk mengidentifikasi kesenjangan, mengetahui penelitian dalam bidang yang sama, serta melanjutkan penelitian sebelumnya. Perbedaan yang terdapat pada penelitian ini dengan penelitian yang terdahulu adalah data menggunakan dataset yang diunduh pada kaggle dan metode yang digunakan adalah *collaborative filtering* dan content based filtering untuk mengetahui tingkat akurasi sebuah hasil rekomendasi yang diinginkan.

Tabel 2.1. Daftar Literature Review

No	Peneliti	Judul	Pembahasan
1.	Eggy Ryana Agustina, Munir, Eddy Prasetyo Nugroho, 2020.	Sistem Rekomendasi Film Menggunakan Metode <i>Collaborative filtering</i> dan K-Nearest Neighbors	Memberikan rekomendasi terhadap user aktif dengan cara menghitung rating prediksi dengan kondisi beberapa user lainnya sudah pernah memberi rating pada film yang akan diprediksi untuk user aktif. Memaksimalkan hasil rekomendasi dengan pengelompokan film berdasar beberapa atribut menggunakan metode k-Nearest Neighbors (kNN).
2.	Syandra Sari, Defry Tri Hendra, 2017.	Aplikasi Rekomendasi Film Menggunakan	Euclidean distance digunakan dalam pendekatan ini untuk menghitung kemiripan dan nilai rekomendasi berdasarkan rating yang di berikan

		Pendekatan <i>Collaborative filtering</i> dan Euclidean Distance sebagai Ukuran Kemiripan Rating	oleh penggemar film lain. Hasil evaluasi aplikasi menunjukkan rekomendasi dengan pendekatan ini dapat memberikan rekomendasi sesuai dengan keinginan seorang penggemar film dengan nilai Kendall Rank Correlation Coefficient sebesar 0,66 .
3.	Kiki Ratna Sari, Wildan Suharso, Yufiz Azhar, 2020.	Pembuatan Sistem Rekomendasi Film dengan Menggunakan Metode Item Based <i>Collaborative filtering</i> pada Apache Mahout	Metode Item Based <i>Collaborative filtering</i> dikembangkan pada framework Apache Mahout dan proses pengujiannya menggunakan algoritma k-fold cross validation Hasil dari penelitian ini menunjukkan bahwa metode Item based <i>Collaborative filtering</i> memberikan hasil rekomendasi yang sangat mendekati dengan preferensi nilai yang diberikan oleh penggunanya, hal ini ditunjukkan pada hasil pengujian sistem mendapatkan akurasi sebesar 97%.
4.	Arwin Halim, Hernawati Gohzali, Dita Maria Panjaitan, Ilham Maulana, 2017	Sistem Rekomendasi Film menggunakan Bisecting K-Means dan <i>Collaborative filtering</i>	Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi film menggunakan kombinasi dari bisecting K-Means dan <i>Collaborative filtering</i> . Data film yang digunakan pada penelitian ini berasal dari MovieLens yang terdiri dari 100.000 rating dari 668 user untuk 10329 judul film dalam 18 genre film
5.	Ajay	A Movie	Membahas tentang sistem

	Kaushik, Shubham Gupta, Manan Bhatia, 2018	Recommendation System using Neural Networks	rekomendasi dengan menggunakan Neural Networks. Studi tersebut melakukan tes terhadap dataset dari rating MovieLens yang berjumlah 100.000 evaluasi dengan skala 1 sampai 5, terdiri dari 1682 film dan 943 pengguna. Hasil yang didapatkan dari tes tersebut dengan menggunakan metode Convolutional Neural Network mendapatkan akurasi sebesar 92,7%.
--	---	---	---

2.1.1 Literature Review 1

Secara umum proses pemberian rekomendasi terdiri dari tiga langkah, yaitu: penemuan similar user, pembuatan ketetanggaan (neighborhood), dan penghitungan prediksi berdasarkan tetangga yang dipilih. Maka aplikasi yang bisa memberikan rekomendasi film kepada pengguna sangat diperlukan untuk mendapatkan rekomendasi film yang sesuai dengan keinginan. Content based filtering dan Collaborative filtering adalah pendekatan yang paling umum untuk membangun sebuah sistem rekomendasi.

Collaborative Filtering (CF) adalah suatu metode dalam membuat prediksi otomatis untuk memperkirakan ketertarikan atau selera seorang user terhadap suatu item dengan cara mengumpulkan informasi dari user-user yang lain yang direpresentasikan dalam bentuk nilai rating. Selain rating, CF merupakan proses penyaringan atau pengevaluasian item menggunakan opini orang lain.

Penelitian ini memberikan rekomendasi terhadap user aktif dengan cara menghitung rating prediksi dengan kondisi beberapa user lainnya sudah pernah memberi rating pada film yang akan diprediksi untuk user aktif. User aktif yang memiliki banyak film yang dihitung prediksi ratingnya, maka dengan perhitungan menggunakan metode weight sum dapat membuat urutan hasil prediksi dari rating terbesar ke rating terkecil untuk urutan perekomendasi film (Agustian, 2020).

Pada metode ini akan diketahui nilai kesamaan antar item dengan tingkat persebaran rating kecil dan nilai kesamaan antar item cenderung lebih jarang berubah dibandingkan dengan nilai kesamaan antar pengguna. Item-based collaborative filtering melakukan kesamaan dengan bentuk suatu model kesamaan secara offline, maka secara otomatis akan menghemat waktu dan penggunaan memori untuk perhitungan saat pengguna mengakses halaman situs.

Adapun kekurangan pada pendekatan ini yaitu tidak dapat merekomendasikan item baru kepada user apabila item tersebut belum pernah diberi rating oleh user manapun. Untuk proses perhitungan pencarian similarity menggunakan metode weighted sum agar dapat memberikan rating prediksi pada item film yang belum pernah dirating oleh user aktif. Untuk memaksimalkan hasil rekomendasi dilakukan pengelompokan film berdasarkan beberapa atribut dengan menggunakan metode k-Nearest Neighbors (kNN). Metode kNN melakukan klasifikasi terhadap item berdasarkan data latih dari item lain yang memiliki jarak terdekat. Dengan menggunakan dua metode tersebut hasil akhir yang didapat akan sesuai dengan harapan yaitu baik dari segi kualitatif maupun kuantitatif.

2.1.2 Literature Review 2

Film telah menjadi salah satu hiburan favorit utama masyarakat. Film terbagi atas berbagai macam genre. Jumlah film pertahun yang terhitung mencapai ribuan mengundang kesulitan dalam memilih film yang sesuai dengan kriteria penggemar film. Salah satu solusi dari permasalahan ini adalah sistem rekomendasi yang memanfaatkan opini atau rating orang lain terhadap suatu film. *Collaborative Filtering* adalah salah satu pendekatan untuk sistem rekomendasi yang mengevaluasi item menggunakan opini orang lain. *Euclidean distance* digunakan dalam pendekatan ini untuk menghitung kemiripan dan nilai rekomendasi berdasarkan rating yang di berikan oleh penggemar film lain. Hasil evaluasi aplikasi menunjukkan rekomendasi dengan pendekatan ini dapat memberikan rekomendasi sesuai dengan keinginan seorang penggemar film dengan nilai *Kendall Rank Correlation Coefficient* sebesar 0,66 .

Sistem rekomendasi pada awal perkembangannya menjalankan algoritma

yang menggunakan rekomendasi-rekomendasi yang diberikan oleh sekelompok pelanggan untuk memperoleh rekomendasi bagi seorang pelanggan tertentu. Pendekatan ini disebut *collaborative filtering*. Prinsip dari *collaborative filtering* adalah jika seorang pelanggan A bersepakat pada sekelompok pelanggan lain tentang barang-barang maka dari kelompok pelanggan tersebut dapat diperoleh informasi: barang apa yang akan relevan dan menarik bagi pelanggan A, inilah yang akan menjadi rekomendasi.

Ketika situs *e-commerce* mulai berkembang, penawaran berbagai barang semakin banyak tersedia di Internet. Pilihan yang sangat beragam ditambah ketersediaan informasi yang sangat banyak tidak selalu mengantarkan pembeli pada pilihan barang yang paling sesuai dengan kebutuhannya, namun kadang kadang justru membawa pembeli pada pilihan barang yang kurang tepat. Kemunculan sistem rekomendasi menjadi sarana yang bermanfaat untuk mengatasi situasi tersebut. Sistem rekomendasi memberikan masukan pada pembeli barang- barang apa yang mungkin sesuai dengan kebutuhan mereka. Sistem rekomendasi akan menyimpan semua data dan respon pembeli sehingga dapat bermanfaat untuk perhitungan rekomendasi yang akan datang. Kini sistem rekomendasi memainkan peranan yang penting dalam situs terkenal seperti: Amazon.com, YouTube, Yahoo, Tripadvisor

Euclidean distance digunakan dalam pendekatan ini untuk menghitung kemiripan dan nilai rekomendasi berdasarkan rating yang di berikan oleh penggemar film lain. Pelanggan harus melakukan pendaftaran terlebih dahulu sedangkan bila pelanggan belum pernah memberikan rating maka rekomendasi buku diberikan berdasarkan pilihan random saja. Hasil evaluasi aplikasi menunjukkan rekomendasi dengan pendekatan ini dapat memberikan rekomendasi sesuai dengan keinginan seorang penggemar film dengan nilai Kendall Rank Correlation Coefficient sebesar 0,66 (Wahyudi, 2017)

2.1.3 Literature Review 3

Pengujian dilakukan dengan tujuan untuk mengetahui nilai keakuratan pada hasil rekomendasi yang diberikan oleh sistem setelah diterapkannya metode Item based *Collaborative filtering* pada Apache Mahout. Pemilihan dataset yang

digunakan pada pengujian dengan nilai K sebesar 5 menghasilkan 5 grup dataset dengan masing-masing grup berjumlah 20167 data. 4 grup data tersebut nantinya digunakan sebagai data training dan 2 grup data digunakan sebagai data uji. Metode pengukuran nilai akurasi dan presisi yang digunakan yaitu Root Mean Square Evaluation yang telah tersedia pada Apache Mahout.

Penelitian yang dilakukan oleh Sumarlin, aplikasi sistem rekomendasi film sesuai dengan preferensi genre yang ditentukan oleh user berhasil dibuat dengan menggunakan metode *Simple Additive Weighting* pada platform Android dengan nilai *System Usefulness* sebesar 85,4% sehingga aplikasi ini sangat membantu pengguna dalam memilih film. Sari dengan penelitiannya berikut juga menjelaskan bahwa dengan menggunakan *Collaborative Filtering* bersama dengan *Euclidean Distance* sebagai pengukur kemiripan rating user dapat memberikan rekomendasi bagi pelanggan toko film online dengan hasil evaluasi menggunakan *Kendall Rank Correlation Coefficient* menghasilkan nilai 0,66 atau 66%, hal ini menunjukkan hasil rekomendasi yang diberikan sesuai dengan keinginan pelanggan. Begitu juga dengan penelitian yang dilakukan oleh Marthasari, sistem rekomendasi dibangun menggunakan aturan-aturan yang dihasilkan oleh Algoritma Apriori. Untuk dapat menampilkan barang rekomendasi digunakan nilai support 20, sedangkan nilai confidence digunakan untuk menentukan N-teratas barang untuk direkomendasikan.

Sistem rekomendasi saat ini masih dapat ditingkatkan lagi kualitasnya, yaitu dengan meningkatkan lagi nilai relevansi hasil rekomendasi yang diberikan dari penelitian sebelumnya. Berdasarkan uraian yang disampaikan diatas, maka pada penelitian kali ini akan dibangun sebuah sistem rekomendasi film dengan menggunakan metode *Item Based Collaborative Filtering* dengan *Apache Mahout* sebagai frameworknya dan proses pengujian nantinya menggunakan algoritma *k-fold cross validation*. Dengan diterapkannya metode tersebut pada penelitian ini dikarenakan metode tersebut dirasa sangat cocok dengan permasalahan yang ada, sehingga penulis ingin membuktikan seberapa cocok tingkat relevansi hasil rekomendasi yang dihasilkan oleh metode *Item Based Collaborative Filtering*.

Terlihat dari hasil pengujian bahwa nilai akurasinya sudah sangat memuaskan yaitu sebesar 97%. Hal ini menunjukkan bahwa sistem sudah sesuai dengan preferensi dari setiap penggunanya. Nilai evaluation score pada Apache Mahout dihitung dengan membandingkan seberapa relevan hasil rekomendasi yang dihasilkan dari data training dengan data uji yang digunakan dalam model pengujian. Sedangkan nilai precision dan recall yang masih rendah menunjukkan bahwa dataset yang digunakan masih perlu ditingkatkan lagi (Sari, Suharso and Azhar, 2020).

2.1.4 Literature Review 4

Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi film menggunakan kombinasi dari bisecting K-Means dan *Collaborative filtering*. Data film yang digunakan pada penelitian ini berasal dari MovieLens yang terdiri dari 100.000 rating dari 668 user untuk 10329 judul film dalam 18 genre film. Proses training terdiri dari proses kluster dengan algoritma bisecting K-Means dan perhitungan nilai similarity dengan *collaborative filtering* (item-based dan user-based). Proses testing dilakukan untuk menghitung nilai error sistem dengan menghitung nilai Mean Absolute Error (MAE). Hasil penelitian menunjukkan rekomendasi dengan bisecting K-Means dan user-based *collaborative filtering* mendapatkan nilai MAE yang lebih rendah dibandingkan dengan bisecting K-Means dan item-based *collaborative filtering*.

Sistem rekomendasi adalah suatu mekanisme yang dapat memberikan suatu informasi atau rekomendasi sesuai dengan kesukaan user berdasarkan informasi yang diperoleh dari user (Sarwar dkk, 2001). Oleh karena itu, diperlukan model rekomendasi yang tepat agar rekomendasi yang diberikan oleh sistem sesuai dengan kesukaan user, serta mempermudah user mengambil keputusan dalam menentukan item (film) yang akan dipilih (McGinty dan Smyth, 2006). Salah satu metode rekomendasi yang digunakan dalam sistem rekomendasi adalah Collaborative filtering. Collaborative filtering menghubungkan setiap user dengan kesukaan yang sama terhadap suatu item (film) berdasarkan rating yang diberikan user. Untuk meningkatkan keakurasian hubungan antara user dengan kesukaan yang sama terhadap suatu item (film) digunakan algoritma clustering

(Gupta, 2009).

Clustering adalah mengelompokkan item data kedalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial (Garcia-Molina dkk, 2002). Pada penelitian Gupta (2009) telah mencoba menggabungkan Collaborative filtering dengan algoritma kmeans yang menghasilkan sistem rekomendasi yang efisien untuk pemrosesan data dalam jumlah besar dan akurasi yang tinggi. Bisecting K-Means merupakan algoritma yang lebih baik dibandingkan algoritma K-Means karena memproduksi cluster yang seragam dan tidak memproduksi cluster kosong, tingkat keakurasian yang baik dan lebih efisien ketika jumlah cluster meningkat (Patil dkk, 2015). Penelitian ini menggabungkan Collaborative filtering dengan Bisecting KMeans untuk menghasilkan sistem rekomendasi yang baik.

Penelitian ini telah menghasilkan sebuah sistem rekomendasi film berbasis website dengan menggunakan kombinasi algoritma bisecting K-Means dan *Collaborative filtering*. Sistem rekomendasi berbasis website yang telah dikembangkan menggunakan informasi dari dataset MovieLens. Tingkat error pada sistem rekomendasi telah dihitung dengan menggunakan nilai MAE. Nilai rata-rata MAE kombinasi dari bisecting K-Means dan user-based CF adalah 1.63, lebih rendah dibandingkan dengan nilai rata-rata MAE kombinasi dari bisecting K-Means dan item-based (Halim *et al.*, 2017).

2.1.5 Literature Review 5

Pada penelitian ini membahas tentang sistem rekomendasi dengan menggunakan Neural Networks. Studi tersebut melakukan tes terhadap dataset dari rating MovieLens yang berjumlah 100.000 evaluasi dengan skala 1 sampai 5, terdiri dari 1682 film dan 943 pengguna. Hasil yang didapatkan dari tes tersebut dengan menggunakan metode Convolutional Neural Network mendapatkan hasil akurasi sebesar 92,7% (Rizqi Az Zayyad and Kurniawardhani, 2021).

Era digital ditandai dengan *information overload* membutuhkan cara penemuan kembali informasi yang efektif. Sistem rekomendasi muncul sebagai solusi memberikan informasi atau item yang bersifat personal dengan

mempelajari interaksi seorang pengguna dan pengguna lain dan items yang telah terekam sebelumnya tanpa memasukkan query seperti pada *search engine*. *Collaborative filtering* sebagai metode dalam memberikan rekomendasi personal.

Pada paper ini peneliti menyodorkan sebuah model mesin rekomendasi untuk user baru dengan metode *Collaborative Filtering* dengan algoritma alternating least square-weight regularization (ALS-WR) yang kemudian di filter kembali berdasarkan kemiripan genrenya yang menggunakan algoritma *cosine similarity* dengan tujuan memberikan error terkecil dengan presisi yang tinggi.

Untuk dataset peneliti menggunakan dataset dari movielens.org. Root Mean Squared Error (RMSE) pada saat training mendapatkan hasil prediksi pada dataset 100K adalah 0.96 (validasi) sementara 0.94 (test), pada dataset 1M nilai RMSE 0.86 (validasi) dan 0.96 (test), pada dataset 10M nilai RMSE 0.81 (validasi) sementara RMSE pada data test diperoleh 0.81 (test). Terlihat bahwa algoritma ALS-WR dapat mengatasi overfitting karena hasil dari validasi dan test pada saat training adalah sama. Terlihat juga semakin besar data, RMSE semakin kecil, dengan demikian ALS-WR dapat digunakan untuk data yang terus tumbuh dan bertambah.

Hasil dari *cosine similarity* untuk mendekatkan hasil *collaborative filtering* dengan genrenya juga didapatkan nilai 1 untuk kemiripan 100% dan nilai itu akan berkurang berdasarkan tingkat kemiripan suatu item film yang dipilih user. Untuk uji penerimaan user didapatkan hanya 28% dari hasil rekomendasi pertama yang dapat diterima user, nilai ini meningkat menjadi 62% tingkat penerimaan user terhadap rekomendasi kedua. Hasil akhir ternyata 75% responden lebih menyukai rekomendasi kedua yaitu hasil dari filtering dua tahap dibandingkan hanya collaborative filtering saja.

2.2 Sistem Rekomendasi

Sistem rekomendasi adalah sistem yang membantu pengguna dalam mengatasi informasi yang meluap dengan memberikan rekomendasi spesifik bagi pengguna dan diharapkan rekomendasi tersebut bisa memenuhi keinginan dan kebutuhan pengguna. Sistem rekomendasi banyak digunakan untuk membuat prediksi seperti buku, musik, film, tempat wisata dan lain-lain. Sebagai contoh

sistem rekomendasi digunakan untuk pemilihan film yang nantinya dapat membantu pengguna memilih film yang akan ditonton. Diperlukan model rekomendasi yang tepat agar rekomendasi yang diberikan sesuai dengan keinginan pengguna, serta mempermudah pengguna mengambil keputusan yang tepat dalam menentukan film yang akan dipilih. Terdapat tiga jenis sistem rekomendasi berdasarkan metode yang digunakannya yakni, *collaborative filtering*, *content-based filtering*, dan *hybrid*. *Collaborative filtering* dibagi menjadi dua bagian yaitu *item-based* dan *user-based*.

Sistem rekomendasi merupakan sebuah (web) alat personalisasi yang menyediakan pengguna sebuah informasi daftar item-item yang sesuai dengan keinginan masing-masing pengguna. Sistem rekomendasi menyimpulkan preferensi pengguna dengan menganalisa ketersediaan data pengguna, informasi tentang pengguna dan lingkungannya. Oleh karena itu rekomendasi akan menawarkan kemungkinan dari penyaringan informasi personal sehingga hanya informasi yang sesuai dengan kebutuhan dan preferensi pengguna yang akan ditampilkan di sistem dengan menggunakan sebuah teknik atau model rekomendasi. (Wijaya and Alfian, 2018)

2.3 Collaborative filtering

Collaborative filtering adalah suatu konsep dimana opini dari pengguna lain yang ada digunakan untuk memprediksi item yang mungkin disukai/diminati oleh seorang pengguna (Ricci , 2011). *Collaborative filtering* memberikan rekomendasi berdasarkan kumpulan dari pendapat, minat dan ketertarikan beberapa user yang biasanya diberikan dalam bentuk rating yang diberikan user kepada suatu item. Untuk memperoleh data rating dari user yang digunakan dalam sistem rekomendasi, dibedakan menjadi dua cara (Wibowo,2010) yaitu:

1. Secara Eksplisit, yaitu proses pengumpulan data dimana user memberikan data secara sadar/ sengaja.
2. Secara Implisit, yaitu proses pengumpulan data dimana user tidak menyadari bahwa ia telah memberikan masukan terhadap sistem.

Pendekatan *collaborative filtering* pada dasarnya dibagi menjadi dua kategori yaitu *user-based collaborative filtering* disebut juga *memory-based*, dan

item based *collaborative filtering* yang disebut juga model-based, Pada pendekatan user based *collaborative filtering* sistem memberikan rekomendasi kepada user item-item yang disukai atau dirating oleh user – user lain yang memiliki banyak kemiripan dengannya. Kelebihan dari pendekatan user based *collaborative filtering* adalah dapat menghasilkan rekomendasi yang berkualitas baik. Sedangkan kekurangannya adalah kompleksitas perhitungan akan semakin bertambah seiring dengan bertambahnya pengguna sistem, semakin banyak pengguna (user) yang menggunakan sistem maka proses perekomendasi akan semakin lama. (Rizqi Az Zayyad and Kurniawardhani, 2021)

Pendekatan item based *collaborative filtering* memberikan rekomendasi berdasarkan kemiripan antar item. Metode ini merupakan metode rekomendasi yang didasari atas adanya kesamaan antara pemberian rating terhadap suatu item dengan item yang pernah dirating user lain. Item yang telah di rating oleh user akan menjadi patokan untuk mencari sejumlah item lainnya yang berkorelasi dengan item yang telah dirating user. Motivasi kunci dibalik metode ini adalah user akan cenderung menyukai item yang sejenis atau mempunyai korelasi dengan item yang telah disukainya. (Rizqi Az Zayyad and Kurniawardhani, 2021)

Secara umum proses pemberian rekomendasi pada *collaborative filtering* terdiri atas 2 langkah, yaitu: Penemuan similar item dan Penghitungan prediksi. Terdapat beberapa algoritma untuk menemukan similar item, yaitu:

1. Algoritma *Cosine-based Similarity*

Pada kasus ini dua item dianggap sebagai 2 vektor. Kesamaan antara 2 item ini diukur dengan menghitung kosinus dari sudut antara 2 vektor item. Item dibandingkan misalnya u dan v , dianggap sebagai sebuah vektor baris dengan anggotanya adalah nilai rating yang diberikan terhadap kedua item tersebut. Dua vektor dikatakan sama jika membentuk sudut atau nilai kosinusnya 1. Dengan kata lain dua item dikatakan mirip jika nilai kosinus dari perhitungan mendekati 1.

2. Algoritma *Correlation-based Similarity*

Pada algoritma ini kemiripan antara dua item i dan j diukur dengan menghitung korelasi Pearson-r correlation. Agar perhitungan korelasi yang diperoleh akurat, terlebih dahulu dilakukan pemisahan terhadap co-rated items (item-item yang kedua item i dan j nya di-rating oleh user).

3. Adjusted-cossine similarity

Persamaan adjusted cosine similarity digunakan untuk menghitung nilai kemiripan antar item. Perhitungan kemiripan ini merupakan modifikasi dari perhitungan kemiripan berbasis vektor dimana dengan melihat fakta bahwa setiap user memiliki skema rating yang berbeda-beda. Terkadang user memberi rating yang tinggi terhadap item a disisi lain user memberi rating yang sangat rendah pada item b . Maka dari itu untuk setiap rating dikurangi dengan rata-rata rating yang diberikan user.

2.4 *Content Based Filtering*

Sistem rekomendasi berbasis konten (Content-based Recommendation System) menggunakan ketersediaan konten (sering juga disebut dengan fitur, atribut atau karakteristik) sebuah item sebagai basis dalam pemberian rekomendasi (Prasetyo *et al.*, 2019). Sebagai contoh, sebuah film mempunyai konten seperti genre, author, tahun rilis, dan lain-lain, atau sebuah file dokumen memiliki konten berupa tulisan yang ada di dalamnya.

Metode content-based filtering biasa digunakan untuk merekomendasikan berita, artikel maupun situs web. Metode tersebut akan mengekstrak informasi yang terdapat pada item kemudian membandingkannya dengan informasi item yang pernah dilihat atau disukai oleh user.

Teknik – teknik yang digunakan dalam content-based seperti TF-IDF, Bayesian Classifiers, Cluster analysis, decision trees dan artificial neural networks. Sistem rekomendasi berbasis konten memiliki beberapa kelebihan, yaitu :

1. Sistem rekomendasi berbasis konten dapat menjelaskan bagaimana hasil rekomendasi didapatkan.

2. Sistem rekomendasi berbasis konten dapat merekomendasikan item-item yang bahkan belum pernah di-rate oleh siapapun.

Namun, sistem rekomendasi berbasis konten juga memiliki beberapa kelemahan, yaitu sistem rekomendasi berbasis konten memerlukan sebuah profil user yang berisikan ketertarikan dan minat pengguna. Bagi pengguna baru yang belum pernah melakukan aktivitas apapun dan tidak memiliki profil user yang cukup, sistem rekomendasi tidak dapat memberikan rekomendasi yang handal kepadanya (Cold Start Problem).

2.5 Hybrid Recommendation

Hybrid recommendation Masing-masing teknik sistem rekomendasi memiliki kelebihan dan kelemahannya tersendiri. Karenanya, sistem rekomendasi hybrid yang menggabungkan kekuatan dari model dan algoritma yang berbeda untuk mengatasi masalah-masalah yang telah disebutkan di atas menjadi target penelitian yang sedang marak dikembangkan (Jannach, 2011). Adapun tujuh teknik kombinasi untuk hybridization (burke, n.d):

1. Weighted : bobot dari dua atau lebih metode rekomendasi digabungkan secara numerik
2. Switching : sistem memilih salah satu atau lebih metode rekomendasi dan menerapkan salah satu metode rekomendasi yang dipilih.
3. Mixed : teknik rekomendasi dari berbagai metode ditampilkan menampilkannya secara bersamaan.
4. Feature Combination : output dari salah satu teknik rekomendasi digunakan sebagai input yang lain.
5. Feature Augmentatio : fitur-fitur dari sumber data rekomendasi yang berbeda digabung bersama-sama ke dalam algoritma rekomendasi tunggal.
6. Cascade : satu rekomender mengolah rekomendasi yang diberikan oleh yang lainnya.
7. Meta-level : model dipelajari oleh satu rekomender yang digunakan sebagai inputan yang lainnya

2.6 IMDb

Internet Movie Database (IMDb) adalah situs web yang menyediakan informasi mengenai film dari seluruh dunia, termasuk orang-orang yang terlibat di dalamnya mulai dari aktor/aktris, sutradara, penulis sampai penata rias dan musikus. Situs web ini sekarang dimiliki oleh Amazon.com. Di dalam IMDb juga ada komunitas yang dapat berkontribusi langsung untuk menuangkan review tentang film dan memberikan rating pada film tersebut. Pengguna yang telah terdaftar (registered users) bisa memberikan penilaiannya atas suatu film dengan menggeser kursornya di atas rentetan bintang-bintang dari 0 sampai 10, atau memilih dari dropdown menu untuk situs mobile IMDb.

Rating yang pengguna input akan memengaruhi nilai sebuah film. Nilai yang diinputkan seorang pengguna tidak langsung merubah nilai rating sebuah film namun jika dipopulasikan dengan registered user lainnya yang jumlahnya mencapai jutaan itu, tentu akan ada perubahan secara signifikan. menggunakan sistem weighted average (pembobotan). Banyak parameter yang diikutsertakan dalam formula IMDb untuk menghasilkan nilai akhir user rating sebuah film. Alasan di balik penggunaan sistem ini adalah untuk menghilangkan unsur ballot stuffing atau vote stuffing, yakni pemberian vote berkali-kali oleh satu user dengan mengubah-ubah vote yang sudah diberikan (karena pada dasarnya ketika satu user telah memberi vote, maka vote-nya akan tercatat di account IMDb-nya walaupun itu masih bisa diubah-ubah di kemudian hari).

2.7 Google Colaboratory

Google Colab atau Google Colaboratory, adalah sebuah *executable document* yang dapat digunakan untuk menyimpan, menulis, serta membagikan program yang telah ditulis melalui Google Drive. *Software* ini pada dasarnya serupa dengan Jupyter Notebook gratis berbentuk *cloud* yang dijalankan menggunakan *browser*, seperti Mozilla Firefox dan Google Chrome.

Ia memungkinkan penggunaanya untuk menjalankan kode Python tanpa perlu melakukan proses instalasi dan *setup* lainnya. Justru, semua keperluan *setting* dan *adjustment* akan diserahkan ke *cloud*. Maka dari itulah, aplikasi ini merupakan tempat yang baik bagi *programmer* yang ingin mengasah

pengetahuan mengenai Python.

Selain itu, Google Colaboratory juga terkenal karena dapat mendorong kebutuhan kolaborasi tim. Di mana *notebook* yang akan dibuat nantinya juga dapat diedit secara bersamaan oleh anggota tim lain, seperti halnya menyunting dokumen di Google Documents.

Keuntungan terbesar dari Google Colaboratory adalah bahwa ia memiliki kumpulan *built-in-library machine learning* paling populer yang dapat dimuat dengan mudah dalam *notebook*.

2.8 YouTube

YouTube adalah perusahaan yang mengumpulkan koleksi user generated content, memuat ribuan film pendek dan episode televisi, dan ratusan film full-length. Melayani lebih dari dua miliar video per hari, telah menjadi pemimpin yang jelas dalam berbagi video online. YouTube terutama memperoleh pendapatan dengan menjual iklan pada halaman homepage dan pencarian hasil-hasilnya, serta dalam video-nya. Situs ini memungkinkan pengguna mengunggah, menonton, dan berbagi video.

Perusahaan ini berkantor pusat di San Bruno, California. Perusahaan ini didirikan pada tahun 2005 oleh Steve Chen (CTO mantan) dan Chad Hurley (mantan CEO). Youtube hari ini adalah anak perusahaan dari Internet pencari raksasa Google. Pada November 2006, YouTube, LLC dibeli oleh Google dengan nilai US\$1,65 miliar dan resmi beroperasi sebagai anak perusahaan Google.

Perusahaan youtube berkantor pusat di San Bruno, California, dan memakai teknologi Adobe Flash Video dan HTML5 untuk menampilkan berbagai macam konten video buatan pengguna, termasuk klip film, klip TV, dan video musik. Selain itu ada pula konten amatir seperti blog video, video orisinal pendek, dan video pendidikan. 29 YouTube berawal sebagai sebuah perusahaan teknologi rintisan yang didanai oleh investasi senilai \$11,5 juta dari Sequoia Capital antara November 2005 dan April 2006.

Kantor pertama YouTube terletak di atas sebuah pizzeria dan restoran Jepang di San Mateo, California. Nama domain www.youtube.com aktif pada 14

Februari 2005 dan situs ini dikembangkan pada bulan-bulan berikutnya. Video pertama di YouTube berjudul Me at the zoo. Video ini menampilkan pendiri pendamping Jawed Karim di San Diego Zoo. Video ini diunggah pada tanggal 23 April 2005 dan masih ada sampai sekarang di situs ini. YouTube menawarkan uji beta pada Mei 2005, enam bulan sebelum peluncuran resmi pada November 2005. Pertumbuhan situs ini meroket dan pada bulan Juli 2006, perusahaan ini mengumumkan bahwa lebih dari 65.000 video diunggah setiap harinya dan situs ini menerima 100 juta kunjungan video per hari. Kebanyakan konten di YouTube diunggah oleh individu, meskipun perusahaan-perusahaan media seperti CBS, BBC, Vevo, Hulu, dan organisasi lain sudah mengunggah material mereka ke situs ini sebagai bagian dari program kemitraan YouTube. Pengguna tak terdaftar dapat menonton video, sementara pengguna terdaftar dapat mengunggah video dalam jumlah tak terbatas.

Pada November 2011, jejaring sosial Google+ terintegrasi langsung dengan YouTube dan penjelajah web Chrome, sehingga videovideo YouTube bisa ditonton di Google+. Bulan Desember 2011, 30 YouTube meluncurkan antarmuka baru. Kanal video ditampilkan di kolom tengah halaman utama, sama seperti umpan berita situs-situs jejaring sosial. Pada saat yang sama, versi baru logo YouTube dipasang dengan bayangan merah yang lebih gelap. Inilah perubahan desain pertama mereka sejak Oktober 2006.

2.9 Netflix

Netflix adalah layanan streaming yang menawarkan berbagai acara televisi, film, film dokumenter, dan anime yang diakses dengan perangkat yang terhubung ke internet. Pengguna Netflix bisa menonton sepuasnya, kapan pun, dimana pun, dengan media apa pun dengan biaya langganan per bulan.

Kantor pusat Netflix berada di Los Gatos, California. Netflix didirikan pada tahun 1997 oleh Reed Hasting sand Marc Randolph di Scotts Valley, California. Model bisnis awal Netfilix adalah penjualan Digital Video Disc (DVD) dan rental melalui pengiriman. Satu tahun setelah berdiri, Netfilix fokus kepada penyewaan DVD dari pada penjualan DVD, sehingga bisnis usaha penjualan DVD ditinggalkan.

Pada tahun 2007, Netflix memperluas bisnisnya dengan mengenalkan media streaming namun tetap mempertahankan layanan penyewaan DVD dan Blu-ray. Perusahaan ini memperluas usahanya secara internasional, dengan layanan streaming tersedia di Canada pada tahun 2010 dan sejak saat itu layanan tersebut semakin berkembang. Sejak Januari 2016, layanan Netflix telah beroperasi lebih dari 190 negara, layanan ini tersedia secara bebas di Internet kecuali daratan China, Suriah, Republik Krimea, dan Indonesia (beberapa penyedia layanan internet memblokirnya karena masalah sensor).

Sejak Juli 2018, Netflix memiliki lebih dari 130 juta total pelanggan secara internasional, termasuk 57.38 juta di Amerika Serikat sendiri. Upaya mereka adalah memproduksi konten baru, mengamankan hak untuk konten tambahan, dan perbedaan melalui 190 negara telah mengakibatkan perusahaan untuk mengajukan miliaran utang jangka panjang.. <http://repository.unimus.ac.id> Kantor pusat Netflix beralamat di 121 Albright Way, Los Gatos, California, Amerika Serikat. Mereka juga memiliki kantor di Belanda, Brasil, India, Jepang dan Korea Selatan.

2.10 Data Mining

Data Mining merupakan suatu teknik yang membahas mengenai penggalian atau pengumpulan data.informasi yang dikumpulkan. Hal tersebut dapat berupa pola-pola yang tersembunyi dalam data, selain itu hubungan dari setiap elemenelemen data, ataupun model untuk keperluan data (Adinugoro and Sari, 2018). Pada tahun 1990-an istilah Data mining mulai populer di komunitas pengguna basis data, akan tetapi sebenarnya perkembangan awal data mining berawal pada tahun 1763 ketika Thomas Bayes mempublikasikan Teorema Bayes.

Ada juga yang berpendapat bahwa, data mining merupakan suatu disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data. Data mining sering juga disebut Knowledge Discovery in Database (KDD), yaitu kegiatan pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Prasetyo, 2014). Sedangkan Menurut Daryl Pregibon

data mining merupakan campuran dari ilmu statistik, kecerdasan buatan, dan riset basis data.

Teknik-Teknik Data Mining sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahapan dalam pemrosesannya. Teknik-teknik tersebut bersifat interaktif, dimana pemakai terlibat langsung atau dengan perantara knowledge base. Teknik-teknik data mining adalah sebagai berikut:

1. Predictive modelling yang merupakan pengolahan data mining dengan melakukan prediksi/peramalan. Tujuan metode ini untuk membangun model prediksi suatu nilai yang mempunyai ciri-ciri tertentu.
2. Association (Asosiasi) merupakan teknik dalam data mining yang mempelajari hubungan antar data.
3. Clustering (Klastering) atau pengelompokan merupakan teknik untuk mengelompokkan data ke dalam suatu kelompok tertentu.
4. Classification (Klasifikasi) merupakan teknik mengklasifikasikan data. Perbedaannya dengan metode clustering terletak pada data, dimana pada clustering variabel dependen tidak ada, sedangkan pada classification diharuskan ada variabel dependen (Prasetyo, 2014).

Adapun proses Data Mining menurut secara sistematis ada tiga langkah utama dalam data mining:

1. Pemrosesan awal data (Eksplorasi)

Eksplorasi terdiri dari pembersihan data, normalisasi data, transformasi data, penanganan data yang kurang tepat, reduksi dimensi, pemilihan subset atribut, dan lain-lain.

2. Membangun model dan melakukan validasi terhadapnya

Membangun model dan melakukan validasi berarti melakukan analisis berbagai model kemudian menentukan model mana yang mempunyai kinerja prediksi yang paling baik. Metode atau model yang bisa digunakan antara lain klasifikasi, regresi, analisis cluster, deteksi anomali, analisis asosiasi, analisis pola sekuensial, dan sebagainya.

3. Penerapan

Setelah dilakukan analisis terhadap model yang akan digunakan, proses yang selanjutnya adalah menerapkan model yang baru untuk menghasilkan perkiraan/prediksi masalah yang diinvestigasi.

Kemudian klasifikasi yang merupakan suatu teknik dalam data mining yang digunakan untuk penggolongan atau pengelompokan berdasarkan target yang telah ditentukan. Klasifikasi merupakan salah satu metode machine learning. Metode ini tercantum ke dalam jenis supervised learning. Sebutan dari klasifikasi didapat dari tujuan utama metode ini buat memprediksikan sesuatu jenis dari input informasi.

Di kutip dari jurnal (Deni Arifanto, M.Kom, 2016) Arti kalsifikasi menurut Bertalya (2009: 5) ialah suatu teknik dengan melihat pada kelakuan serta atribut dari kelompok yang telah didefinisikan. teknik ini bisa memberikan klasifikasi pada informasi baru dengan memanipulasi data yang terdapat yang sudah diklasifikasi serta dengan memakai hasilnya buat membagikan sejumlah ketentuan. Aturan- ketentuan tersebut digunakan pada data data baru buat diklasifikasi. Metode ini memakai supervised induction yang menggunakan kumpulan pengujian dari informasi set yang terklasifikasi.

Tahapan dari klasifikasi dalam data mining adalah :

1. Pembangunan model, dalam tahapan ini dibuat sebuah model agar menuntaskan permasalahan klasifikasi class ataupun atribut dalam data, model ini dibentuk bersumber pada training set- sebuah contoh informasi dari kasus yang dialami, training set ini telah memiliki data yang lengkap baik attribut ataupun classnya.
2. Pelaksanaan model, pada tahapan ini model yang telah dibentuk tadinya digunakan agar memastikan atribut/ class dari suatu data baru yang atribut/ class- nya belum dikenal tadinya.
3. Penilaian, pada tahapan ini hasil dari pelaksanaan model pada tahapan tadinya dievaluasi memakai parameter terukur agar memastikan apakah model tersebut bisa diterima