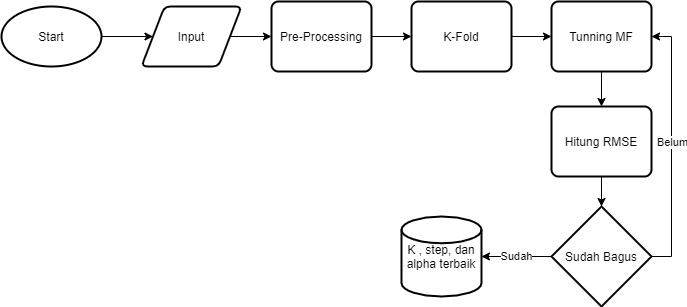
Laporan Tugas Besar *Recommender System*

1. Dataset

Dataset yang digunakan pada tugas besar ini berjumlah tiga data yaitu rating.csv yang memiliki jumlah data sebanyak 100836 yang akan dipakai pada ketiga metode *recommender system*, movies.csv yang memiliki data sebanyak 9742 yang akan dipakai pada proses clustering movie dengan genre, dan tags.csv yang memiliki data sebanyak 3683 yang akan dipakai pada proses clustering movie dengan tag.

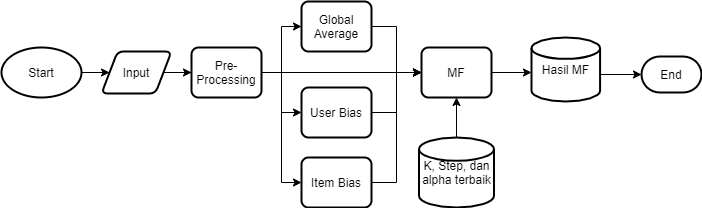
1. Matrix Factorisation

Metode ini akan dilakukan dua tahap yaitu proses latih Matrix Factorisation (MF) dan proses recommendation menggunakan metode MF ini, berikut adalaha skema latih MF.



Gambar Proses Latih MF

Input merupakan data rating.csv, seterusnya data input akan dilakukan Pre-Processing dimana proses ini akan membaca rating.csv dangan library pandas python yaitu pd.read\_csv(), setelah dilakukan pembacaan rating.csv proses pre-processing selanjutnya yaitu membuat matriks rating berdasarkan data yang sudah dibaca, setelah matriks rating telah terbuat maka proses selanjutnya yaitu melakukan k-fold dimna k=5 sehingga akan dilakukan k-kali iterasi dengan partisi ke-k sebagai test dan partisi lainnya akan diguanakn sebagai train, data yang dipakai sebagai test akan dilakukan perubahan nilai yaitu semua nilai rating pada data test akan dirubah menjadi nol (0). Pada proses k-fold akan dilakukan tunning MF dimana parameter yang digunakan pada proses ini yaitu k=[1,2,3,4,5], step=[3,10], dan alpha [0.0002, 0.0000002], dari hasil tunning MF didapatkan nilai rmse terbaik yaitu 1,3 dengan k = 1, step = 3, dan alpha = 0.0002. selanjutnya k,step, dan alpha akan diguanakan pada proses rekomndasi menggunakan MF, berikut adalah proses rekomendasi menggunakan MF.

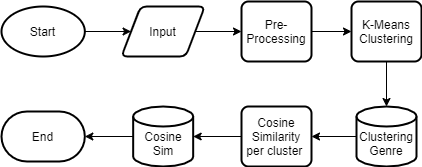


Gambar Skema Rekomendasi MF

Input merupakan data rating.csv, seterusnya data input akan dilakukan Pre-Processing dimana proses ini akan membaca rating.csv dangan library pandas python yaitu pd.read\_csv(), setelah dilakukan pembacaan rating.csv proses pre-processing selanjutnya yaitu membuat matriks rating berdasarkan data yang sudah dibaca. Proses selanjutnya yaitu menghitung global average, user bias, dan item bias, dimana data inputanya adalah matriks rating hasil dari pre-prosessing, setelah global average, user bias, dan item bias sudah didapatkan hasilnya, maka proses lanjutnya yaitu rekomendasi menggunakan metode MF, pada proses ini inputan yang di butuhkan yaitu matriks rating hasil dari preproses dan hasil dari perhitungan global average, user bias, dan item bias, serta menggunakan parameter k, step, dan alpha terbaik yang sudah didapatkan dari proses latih MF. Setelah proses MF sudah dijalankan maka akan dipilih top 1000 prediksi rating movie untuk setiap user.

1. Content Based Filtering

Metode ini akan mengklastering movie dengan genre,pada metode ini akan dilakukan dua tahap yaitu pembentukan database klastering genre serta database cosine sim, dan proses rekomendasi, berikut adalah skema untuk mendapatkan database.



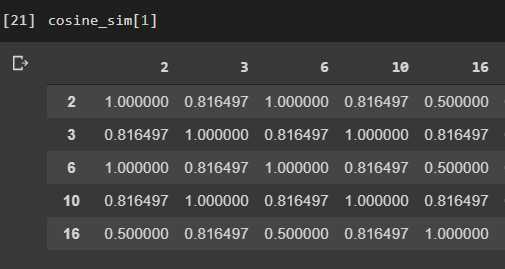
Gambar Skema Mendapatkan Database

Input merupakan data movies.csv, seterusnya akan dilakukan pre-processing dimana akan ada beberapa tahap pada proses ini yaitu pertama akan dilakukan pembacaan rating.csv dangan library pandas python yaitu pd.read\_csv(), tahap selanjutnya yaitu akan dilakukan split data pada genre dikarenakan data pada genre untuk setiap movie masih terpisah dengan karakter ‘|’ sehingga perlu displit data genre menggunakan data.split(‘|’) untuk mendapatkan semua kategori genre yang tersedia, setelah sudah didapatkan semua kategori genre maka proses selanjutnya dari preprocessing yaitu membuat matriks denga kolom di set kategori genre dan index diset sebanyak jumlah data movie pada movie.csv, setelah matriks sudah dibuat langkah selanjutnya yaitu melakukan pengisian matriks dimana movie yang mempunyai genre tersebut akan diset satu pada matriks dengan posisi index movie tersebut dengan kolum genre tersebut, berikut adalah hasil dari matriks yang sudah dibuat.



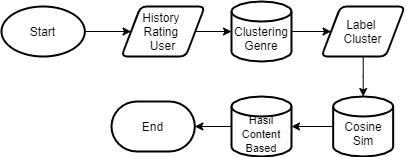
Gambar Hasil Matriks movie genre

Setelah matriks sudah selesai diisi maka proses preprocessing berhasil dilakukan, proses selanjutnya yaitu melakukan klaster terhadap matriks yang sudah dibuat tadi, pada metode ini klaster yang dipakai sebanyak 8 klaster dengan menggunakan library K-means sklearn. Setelah klaster sudah dilakukan maka langkah terakhir yaitu membuat matriks similarity untuk setiap klaster, jadi setiap klaster akan dihitung similarity setiap movie didalam klaster tersebut, berikut contoh hasil similarity dari klaster satu.



Gambar Hasil similarity klaster 1

Setelah database similarity dan klastering sudah didapat maka tahap pembentukan database sudah berhasil dilakukan,selanjutnya masuk ke tahap rekomendasi, berikut skema dari tahap rekomendasi.

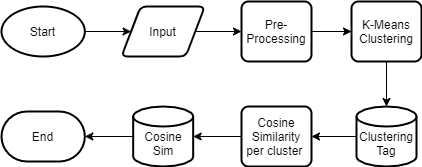


Gambar Tahap Rekomendasi

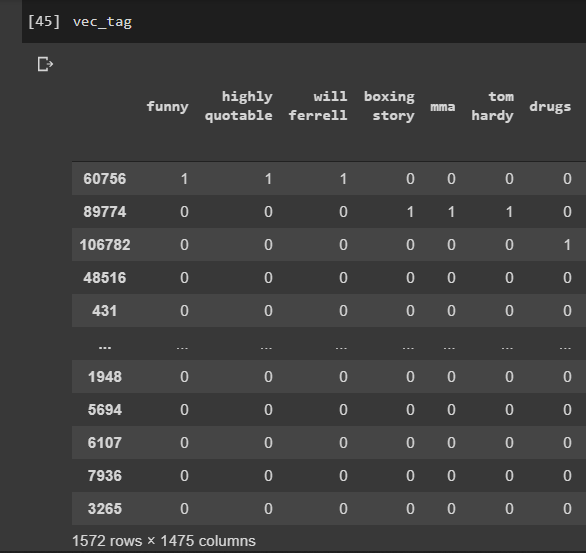
Tahap pertama pada tahap ini yaitu mendapatkan movieId dari history rating setiap user yang sudah merating lebih besar sama dengan 4.0 , setelah di dapatkan movieidnya maka selanjutnya akan mencari movieid tersebut masuk kedalam klaster ke-berapa, setelah mendapatkan label dari klasternya maka selanjutnya akan mengambil hasil similarity dari klaster tersebut dan akan mengambil top 5 movieid yang memiliki similarity tertinggi terhadap movieid yang dimasukan.

1. Clustering Movie Tags

Metode ini akan mengklastering movie dengan tag, pada metode ini akan dilakukan dua tahap yaitu pembentukan database klastering tag serta database cosine sim, dan proses rekomendasi, berikut adalah skema untuk mendapatkan database.

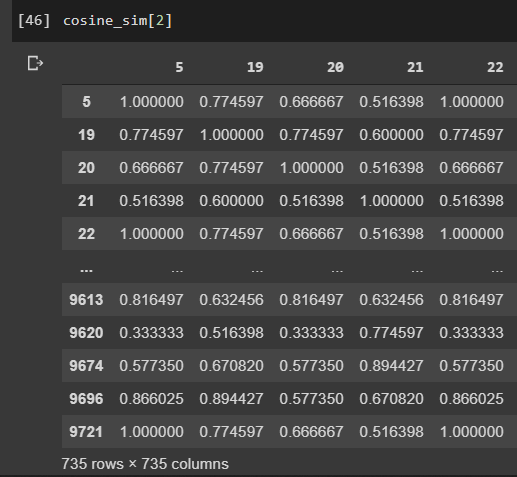


Input merupakan data tags.csv, seterusnya akan dilakukan pre-processing dimana akan ada beberapa tahap pada proses ini yaitu pertama akan dilakukan pembacaan tags.csv dangan library pandas python yaitu pd.read\_csv(), tahap selanjutnya yaitu akan dilakukan lower case pada data tag dikarenakan pada tag masih memiliki tag yang sama tetapi tidak sama casenya sehingga nntinya tag tersebut akan menjadi duplikat, setelah tag sudah di lowercase maka akan dilakukan proses np.unique untuk memastikan bahwa data yang didapat untuk setiap kategori tag itu tepat satu data, setelah sudah didapatkan semua kategori tag maka proses selanjutnya dari preprocessing yaitu membuat matriks denga kolom di set kategori tag dan index diset sebanyak jumlah dari np.unique terhadap movieid, setelah matriks sudah dibuat langkah selanjutnya yaitu melakukan pengisian matriks dimana movie yang mempunyai tag tersebut akan diset satu pada matriks dengan posisi index movie tersebut dengan kolum tag tersebut, berikut adalah hasil dari matriks yang sudah dibuat.



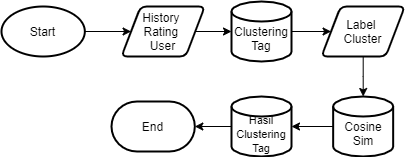
Gambar Hasil Matriks Movie Tags

Setelah matriks sudah selesai diisi maka proses preprocessing berhasil dilakukan, proses selanjutnya yaitu melakukan klaster terhadap matriks yang sudah dibuat tadi, pada metode ini klaster yang dipakai sebanyak 50 klaster dengan menggunakan library K-means sklearn. Setelah klaster sudah dilakukan maka langkah terakhir yaitu membuat matriks similarity untuk setiap klaster, jadi setiap klaster akan dihitung similarity setiap movie didalam klaster tersebut, berikut contoh hasil similarity dari klaster dua.



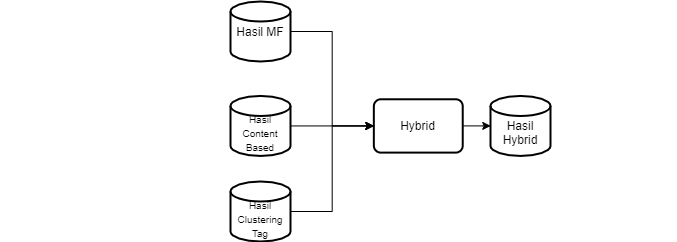
Gambar Hasil Similarity Klaster Dua

Setelah database similarity dan klastering sudah didapat maka tahap pembentukan database sudah berhasil dilakukan,selanjutnya masuk ke tahap rekomendasi, berikut skema dari tahap rekomendasi.



Tahap pertama pada tahap ini yaitu mendapatkan movieId dari history rating setiap user yang sudah merating lebih besar sama dengan 4.0, setelah di dapatkan movieidnya maka selanjutnya akan mencari movieid tersebut masuk kedalam klaster ke-berapa, setelah mendapatkan label dari klasternya maka selanjutnya akan mengambil hasil similarity dari klaster tersebut dan akan mengambil top 5 movieid yang memiliki similarity tertinggi terhadap movieid yang dimasukan.

1. Hybrid

Metode yang digunakan untuk melakukan hybrid yaitu *paralleized hybridization* dimana metode ini akan membuat recommender filtering berkerja secara berdampingan serta secara parallel sehingga output atau hasil dari masing-masing metode filtering akan dilakukan proses hybrid untuk mendapatkan hasil rekomendasi terakhir, berikut adalah skema hybrid pada tubes ini. 

Pada proses hybrid ini menggunakan asumsi bobot yaitu hasil mf dengan bobot = 0.4, hasil content-based dengan bobot = 0.3, serta hasil clustering tag dengan bobot = 0.3, dari bobot tersebut akan dilakukan setiap kombinasi untuk mendapatkan bobot terbesar, pada tahap ini meiliki beberapa tahap untuk melakukan hybrid, pertama akan dilakukan intersection atau irisan dari ketiga hasil metode filtering setelah itu hasilnya akan dimasukan kedalam variable bernama bobot\_kedua, selanjutnya akan dicek apakah isi dari variable bobot\_kedua sama dengan 20 atau tidak, jika sama dengan 20 maka hasil tersebut akan di recommendasikan terhadap user tersebut, jika kurang dari 20 maka akan dilakukan pengambilan data dari irisan MF dengan content-based, dan irisan dari MF dengan Clustering tag dan akan digabungkan kedua irisan tersebut, dari gabungan tersebut akan di hapus movieid yang sudah ada di didalam variable bobot\_kedua, setelah sudah dilakukan penghapusan data maka data tersebut akan di masukan ke dalam variable bobot\_ketiga.

Setelah bobot ketiga di dapatkan maka akan dilakukan penambahan data pada bobot\_kedua dengan menambahkan data dari bobot\_ketiga, jika hasil penambahan data tersebut sudah samadengan 20 maka hasil penambahan tersebut yang dijadikan rekomendasi pada user tersebut, tetapi jika hasilnya masih kurang dari 20 maka akan dilakukan proses irisan antara hasil content based dengan cluster tag. Hasil irisan tersebut akan dilakukan proses penghapusan movieid yang ternyata ada juga pada hasil penambahan data bobot\_kedua dengan bobot\_ketiga, setelah sudah dilakukan penghapusan maka hasilnya akan disimpan pada variable bobot\_keempat, setelah itu, data hasil dari penambahan bobot\_kedua dengan bobot\_ketiga akan di lakukan penambahan data dengan bobot\_keempat, jika hasil penambahan datatersebut samadengan 20 maka hasilnya penambahan tersebut yang akan dijadikan recomendasi untuk user tersebut, jika masih kurang dari 20 item maka akan dilakukan pengambilan data dari hasil MF untuk menjadikan ite rekomendasi menjadi 20 item rekomendasi, berikut contoh recommendasi dari user dengan userid 1.

