

## **Analisa Sistem Rekomendasi Pada Film dengan Metode Content Based dan Collaborative Filtering**

**Muhammad Alfin Fikri**

Program Studi Teknik Informatika, Universitas Yos Soedarso, [fikrialfin@gmail.com](mailto:fikrialfin@gmail.com)

### **ABSTRACT**

*In this modern world, entertainment is a must for each person. With entertainment, you can avoid stress, one of which is by watching movies. Many people prefer to watch movies through applications or sites, such as Netflix or Youtube. they do not know that what they watch has applied a recommendation system. by utilizing the movie recommendation system, users do not need to search for movies again but have been presented with films that are similar to those watched before. The recommendation system itself works through each user's data and then matched with other users, this is also called the Content Based Recommender System, while there is also through the rating of the film watched with the rating, users will be presented with a film with a good rating according to the genre they like (Collaborative Filtering). However, there are many combinations of recommendation systems commonly called Hybrid Recommender Systems that can be better than Collaborative Filtering or Content Based with better MAE results and increasing dataset sparsity by 1-2%, the results can also vary when tested with a larger dataset.*

**Keywords:** *Recommender System; Collaborative filtering; Content Based; Hybrid System Recommender.*

### **ABSTRAK**

Dalam dunia yang modern ini, hiburan adalah suatu keharusan bagi masing-masing tiap orang. Dengan adanya hiburan, bisa menghindari dari stress salah satunya dengan menonton film. Banyak orang lebih senang menonton film melalui aplikasi ataupun situs, seperti Netflix maupun Youtube. mereka tidak tau bahwa yang mereka tonton sudah mengaplikasikan sistem rekomendasi. dengan memanfaatkan sistem rekomendasi film, pengguna tidak perlu untuk melakukan pencarian lagi film lagi melainkan sudah disodorkan film-film yang mirip dengan yang ditonton sebelumnya. Sistem rekomendasi sendiri bekerja melalui data setiap pengguna lalu dicocokkan dengan pengguna lain ini disebut juga dengan Recommender System Content Based sedangkan ada juga melalui rating film yang ditonton dengan adanya rating, pengguna akan disodorkan film yang rating bagus sesuai dengan genre yang disukai (Collaborative Filtering). Namun banyak sekali kombinasi dari sistem rekomendasi yang biasa disebut dengan Hybrid Recommender System yang bisa lebih baik dari pada Collaborative Filtering maupun Content Based dengan hasil MAE yang lebih baik dan meningkatkan sparsity dataset 1-2%, hasilnya juga dapat bervariasi ketika diuji dengan dataset yang lebih besar.

**Kata kunci :** *Sistem Rekomendasi; Collaborative filtering; Content Based; Hybrid System Recommender.*

### **PENDAHULUAN**

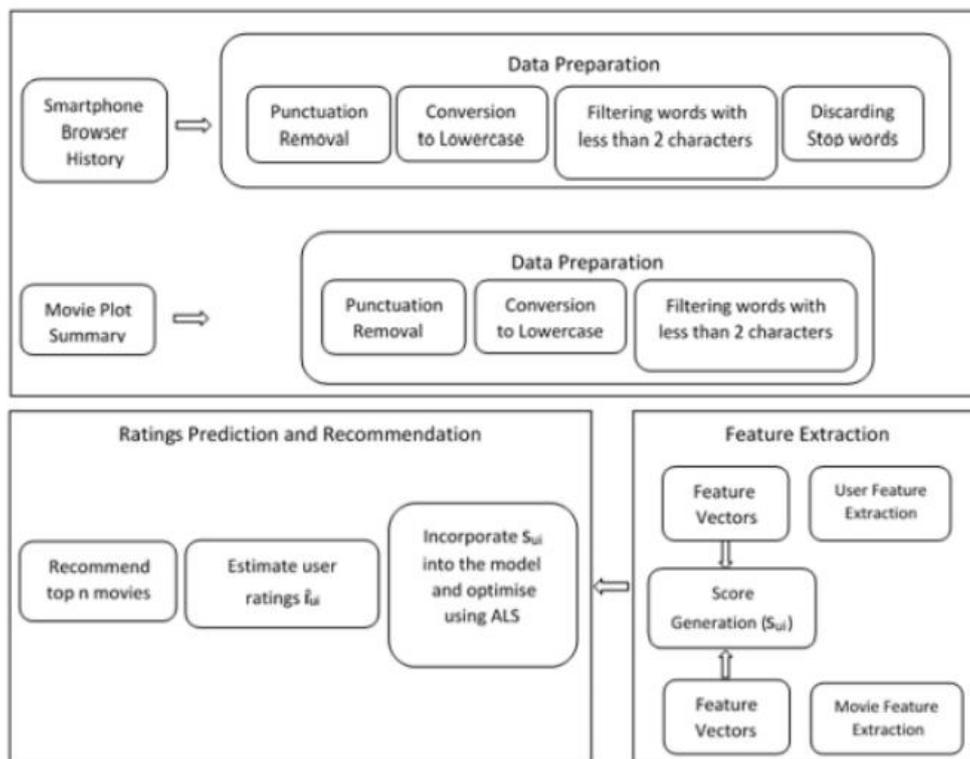
Perkembangan teknologi semakin hari semakin canggih, salah satunya Sistem Rekomendasi. Sistem ini bekerja dengan cara mengumpulkan data beberapa orang lalu memprediksi data yang cocok atau disukai untuk orang tersebut. Secara luas sistem rekomendasi sudah dipakai beberapa situs seperti Youtube, Netflix, Facebook bahkan Amazon. Dengan memanfaatkan sistem rekomendasi mereka bisa menghemat waktu pengguna untuk memilih produk yang cocok digunakan oleh pengguna tanpa harus mencari terlebih dahulu, dengan contoh saat kita memilih film action seperti "Captain Amerika" maka setelah menonton film tersebut akan ada rekomendasi film yang berkaitan dengan Captain Amerika seperti "Avegars:Infinty War" ataupun "Captain Marvel". Pada dasarnya sistem rekomendasi dibagi menjadi 2 yaitu Collaborative Filtering (CF) dan Content Based (CB). Namun banyak sekali penggabungan antara sistem rekomendasi CF maupun sistem rekomendasi CB untuk menghasilkan data yang cocok dan efisien bagi pengguna di situs web maupun film. Sistem

rekomendasi sendiri, sekarang menjadi suatu pilihan untuk memprediksi keinginan pengguna dengan datadata yang sudah tersedia maupun dari pengalaman bisa membantu proses sistem rekomendasi. Terdapat bermacam-macam tantangan dalam membuat sistem rekomendasi seperti struktur aktifitas pengguna yang tidak wajar atau berbeda-beda, pergabungan genre yang hampir sama maupun rating dari film yang tidak stabil membuat sistem rekomendasi menjadi lebih susah. Namun rating dari pengguna itulah sebagai acuan sistem rekomendasi bekerja lebih baik.

**RELATED WORK**

Dalam penelitian terdahulu berfokus pada sistem rekomendasi, beberapa peneliti menggunakan metode berbeda-beda untuk melakukan identifikasi seperti: Shreya Agrawal [1] Dalam penelitiannya meningkatkan pendekatan user kedalam film menggunakan rekomendasi sistem, berfokus pada 3 hal yaitu akurasi, kualitas, dan skalabilitas dari dataset movielens film. Dengan menggabungkan 2 metode berbeda yaitu content based filtering dan collaborative filtering dapat hasil yang signifikan dan juga sedikit ada peningkatan dengan mempertimbangkan genre film sebagai acuan dalam proses rekomendasi. Shreya [1] juga mengevaluasi tingkat dari kemajuan dengan precision recall dan hasilnya dari cukup bagus dari 3 konten tersebut dengan nilai memori dari CF adalah avarege dan nilai memori CB tergolong low.

Ashish Pal [2] Penelitiannya berfokus pada peningkatan



Gambar 1. Tahapan movie recomender framework

Metode collaborative filtering dengan kasus spartial. Dalam penelitian tersebut membandingkan pendekatan populer yaitu Pure CF (Collaborative Filttering), SVD (Singular Value Decomposition) dan Hybird CF (Collaborative Filttering) dari membandingkan tersebut dapatlah hasil bahwa Hybird CF lebih bagus dengan initial sparsity 98,388% didapat improve sparsity 97,13%. Dalam uji evaluasi Arshih [2] menggunakan MAE (Mean Absolute Error). Dari 3 percobaan yang dilakukan algoritma hybrid CF berfungsi lebih baik daripada pure CF alasannya pure CF terlalu bergantung pada data yang tersedia melalui pengaktifan matrix.

User	URL	Topic Modelling (Bag of Words)
1	<a href="https://www.kaggle.com/c/word2vec-nlp-tutorial/data?unlabeledTrainData.tsv.zip">https://www.kaggle.com/c/word2vec-nlp-tutorial/data?unlabeledTrainData.tsv.zip</a>	[0,0,0,1,....,0,0]
	<a href="https://www.youtube.com/results?search_query=movie+bloopers">https://www.youtube.com/results?search_query=movie+bloopers</a>	[1,2,1,0,....,1,0]
	<a href="https://www.google.co.in/search?q=UnicodeDecodeError%3A+%27utf">https://www.google.co.in/search?q=UnicodeDecodeError%3A+%27utf</a>	[1,0,2,1,....,0,1]
2	<a href="http://www.hifivision.com/what-should-i-buy/39753-what-cable-buy-connect-laptop-amp.html">http://www.hifivision.com/what-should-i-buy/39753-what-cable-buy-connect-laptop-amp.html</a>	[0,0,0,0,....,0,0]
	<a href="https://www.google.co.in/webhp?sourceid=chrome-instant&amp;ion=1&amp;espv=2&amp;ie=UTF-8#q=vector+space+model+python+example">https://www.google.co.in/webhp?sourceid=chrome-instant&amp;ion=1&amp;espv=2&amp;ie=UTF-8#q=vector+space+model+python+example</a>	[0,0,1,0,....,0,0]
	.....	
3	<a href="https://www.youtube.com/results?search_query=mo+vlogs">https://www.youtube.com/results?search_query=mo+vlogs</a>	[1,1,2,1,....,0,1]
	<a href="https://github.com/search?p=2&amp;q=URL+clustering&amp;type=Repositories&amp;utf8=%E2%9C%93">https://github.com/search?p=2&amp;q=URL+clustering&amp;type=Repositories&amp;utf8=%E2%9C%93</a>	[3,0,0,0,....,0,0]
	.....	
n	<a href="https://github.com/pashikc/url-clusterer-with-k-means/tree/master/out1450039293_MyIterationForComparison">https://github.com/pashikc/url-clusterer-with-k-means/tree/master/out1450039293_MyIterationForComparison</a>	[0,0,0,0,....,1,1]
	<a href="http://www.ttdconline.com/hotelnamedisplay.do">http://www.ttdconline.com/hotelnamedisplay.do</a>	[1,1,1,1,....,1,0]
	<a href="https://www.google.co.in/webhp?sourceid=chrome-instant&amp;ion=1&amp;espv=2&amp;ie=UTF-8">https://www.google.co.in/webhp?sourceid=chrome-instant&amp;ion=1&amp;espv=2&amp;ie=UTF-8</a>	[0,1,0,1,....,0,0]
.....		

Gambar 2. Extraksi fitur Bag of Word

Dixon Prem [3] Dalam penelitiannya mengusulkan penggabungan sistem rekomendasi collaborative filtering dan content based untuk menghasilkan preferensi pengguna laten yang nantinya memberikan hasil personal bagi pengguna. Peneliti memakai browsing history sebagai acuan membuat sistem rekomendasi. Dixon [3] melakukan banyak tahapan dalam pembuatan sistem rekomendasi. Data preparation melakukan stopwords, lower casing maupun stemming agar data menjadi lebih ringkas dan bisa di proses, lalu feature extraction data yang sudah siap dilakukan ekstraksi fitur menggunakan Bag of Words. Sebuah vector yang mewakili setiap URL dibuat permodelan per topic. Bag of Word juga dikenal sebagai permodelan per topic yang digunakan dalam proses bahasa alami guna untuk mewakili sebuah dokumen, dalam hal ini dokumen berupa data dari ekstraksi fitur yang telah diolah menjadi model per topic. Seperti pada gambar 2.

MSE is computed as:

$$\frac{1}{|K|} \sum_{(u,i) \in K} |r_{ui} - (\mu + b_u + b_i + q_i^T(p_u + s_{ui}))| \quad (1)$$

RMSE is computed as:

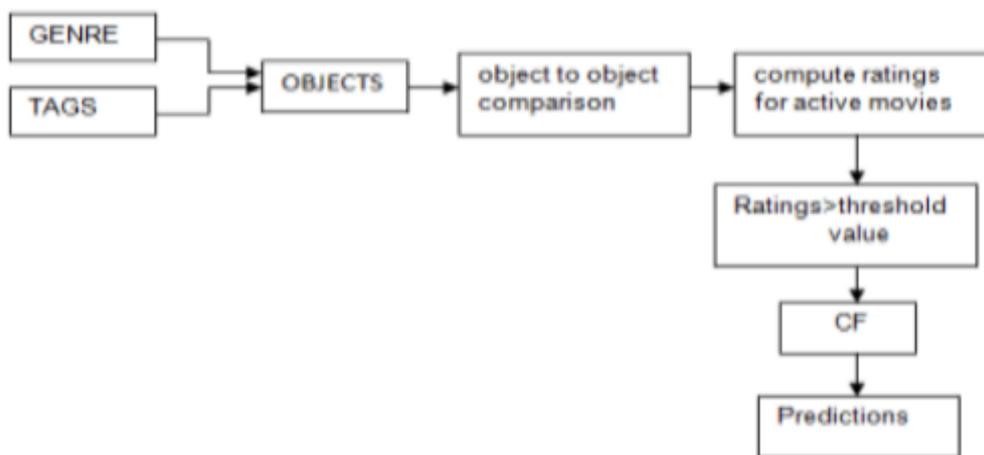
$$\sqrt{\frac{1}{|K|} \sum_{(u,i) \in K} (r_{ui} - (\mu + b_u + b_i + q_i^T(p_u + s_{ui})))^2} \quad (2)$$

Setelah melakukan tahapan ekstraksi fitur maka selanjutnya mulai melakukan tahapan rating prediction yang biasa dikenal sebagai collaborative filtering. Dixon [3] juga menilai kualitas sistem rekomendasi dengan beberapa metode salah satunya dengan Mean Squared Error (MSE) atau Root Mean Squared Error (RMSE), hasil yang di peroleh MSE yang lebih kecil akan terlihat lebih baik. Presisi, yang didefinisikan sebagai rasio jumlah film diterima ke jumlah film yang ditampilkan sebagai rekomendasi digunakan untuk mengevaluasi film top-n rekomendasi. Abinash Pujahari [4] penelitiannya tentang grup sistem rekomendasi yang mengkombinasikan antara pengguna dengan barang-barang dengan metode collaborative filtering. Membuat sistem rekomendasi yang

menyarankan grup mana saja yang cocok untuk pengguna dengan metode CF. Acuan dalam metode ini menggunakan similarity atau disebut juga kemiripan antara beberapa user lalu di cocokkan dengan pengguna, setelah di hitung skor kemiripan dari beberapa data yang cocok, pengguna akan direkomendasikan grup dan barang-barang yang cocok untuk dibeli. Dalam penelitiannya Pujahari [4] mengevaluasi tahap akhir dengan Precision Recall didapat pengelompokan berdasarkan grup lebih efektif. Dalam kumpulan data yang telah dibuat yaitu sistem rekomendasi grup film bisa dikatakan ada rekomendasi berkisar hingga 3-5 anggota tiap grup.

**METODOLOGY**

Sistem rekomendasi sangatlah flexbile banyak pergabungan dengan metode-metode lain dan berkembang menjadi Hybird Recomender System. Ashish Pal [2] mencoba mengembang Hybird Recommender System dengan pergabungan antara metode CF dan CB tentang film. Yang di ilustrasikan pada gambar 4.



Gambar 4. Alur Hybird Recomender System

- A. Data Review(content based)
 

Sistem rekomendasi yang disampaikan berdasarkan kesamaan antara profil pengguna atau item-item yang ada diprofil seperti genre [3] seperti content based. Namun dalam kasus ini hanya melakukan filtering terhadap genre dan tags lalu digabung menjadi suatu object.
- B. Data pre-processing
 

Pemrosesan data yang dilakukan untuk menghilangkan kata atau lower case pada judul, genre maupun tags. Proses ini bisa dilewati jikalau data yang digunakan sudah berbentuk cell ataupun bersih dari noise.
- C. Collaborartive Filtering
 

Sistem rekomendasi yang menggunakan data seperti rating atau pendekatan item-item yang bernilai [3]. Dengan contoh rating pengguna A dengan pengguna B berbeda maka dapatlah nilai dari tiap pengguna tersebut. Lalu sistem akan memproses dan membuatkan daftar rekomendasi yang cocok untuk pengguna dari kemiripan rating tersebut. Ilustraasi bisa dilihat pada gambar 5.

	Item 1	Item 2	Item 3	Item 4
User 1	3	4	0	1
User 2	4	3	4	1
User 3	3	4	0	4
User 4	4	3	4	1
User 5	3	4	3	4

Gambar 5. Contoh pengelompokan collaborative filltering

D. Prediction (Pearson Correlation)

Prediksi mengetahui kesamaan antara pengguna atau itemitem lainnya bisa dilakukan dengan syarat item tersebut terdapat nilai atau variable yang jelas seperti contoh rating film. metode di bawah ini yang digunakan untuk menghitung kesamaan di antara pengguna atau item

$$sim(u, v) = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u) (r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \quad (3)$$

Dalam Persamaan di atas, ukuran kemiripan dihitung berdasarkan informasi peringkat atau rating, dimana  $r_u$  mewakili peringkat yang diberikan pengguna  $u$  ke item  $i$ . Perhitungan correlation ini dilakukan berulang setiap kali item yang berbeda dengan membandingkan antara user 2 hingga 5 agar hasil semakin akurat bahwa rekomendasi sistem berjalan dengan baik

E. Evaluation

Evaluasi tetap harus dilakukan guna untuk mengetahui tingkat akurasi dari sistem. Dalam Penelitian ini hampir setiap peneliti menggunakan metode precision karna selain ringkas juga mudah di pakai.

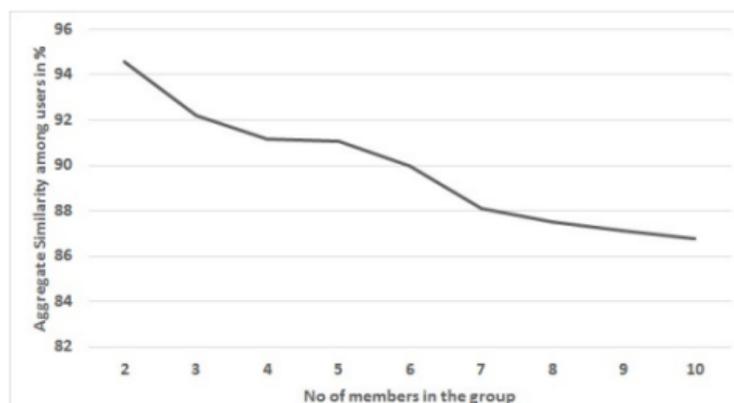
$$Precision = \frac{tp}{tp + fp} \quad (4)$$

Perhitungannya juga cukup mudah, dimana true positive dibagi dengan true positive+ false positive, maka dapat lah nilai dari precision. Jika ditullis dengan rumus maka akan menjadi seperti berikut

$$Precision(U) = \frac{R \cap U}{R} \quad (5)$$

Dimana  $R$  menunjukkan item yang direkomendasikan untuk pengguna dan  $U$  menunjukkan barang yang digunakan oleh pengguna saat ini.  $R \cap U$  menandakan item umum antara  $R$  dan  $U$ . Nilai presisi diperoleh hanya untuk satu pengguna saja. Jika ingin mencari keluar presisi untuk sistem rekomendasi kelompok perlu di hitung kembali ketepatan rekomendasi untuk tiap masing-masing anggota kelompok dan kemudian menumukan ketepatan rata-rata kelompok total.

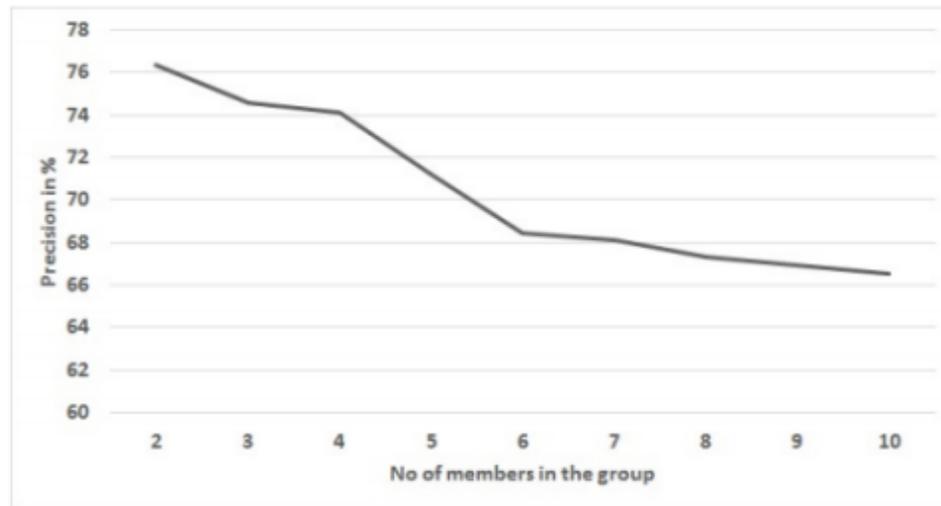
Dengan Precision lebih mudah mengetahui tingkat hasil dari perkembangan metode ini. Gambar 6 Dimana terlihat bahwa tingkat Similarity dari metode collaborative filtering sudah mendekati precision.



Gambar 6. Kemiripan di antara pengguna group

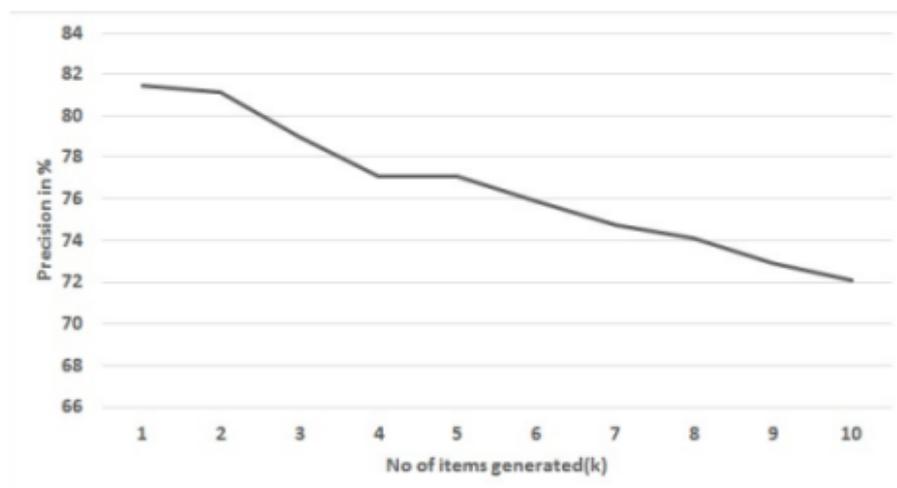
**F. Obtain Result**

Presisi tiap masing-masing kelompok berbeda karna tiap kelompok punya peran penting. Dengan menggunakan Algoritma 1, Algoritma 2 dan evaluasi kriteria telah menemukan nilai presisi berikut untuk sistem rekomendasi film kelompok kami untuk kelompok yang berbeda bisa dilihat di gambar no 7. Menunjukkan bahwa, ketika kita meningkatkan no. anggota dalam kelompok nilai presisi dari rekomendasi kelompok sistem menurun secara bertahap. Setelah beberapa kenaikan kurva. Namun itu menghasilkan hasil yang baik



Gambar 7. Presisi antara beberapa kelompok

Hal lain dalam prosedur rekomendasi yang diusulkan, telah menggunakan variabel  $k$ , namun tidak ada item yang dihasilkan oleh sistem. Misalkan sudah diperbaiki kelompok memiliki 5 anggota telah memberikan nilai-nilai yang berbeda, berarti tidak ada film yang dihasilkan akan berubah dan juga ketepatan kelompok akan berubah. Hasilnya adalah ditunjukkan pada Gambar 8. Ini menunjukkan bahwa ketika no. item untuk direkomendasikan meningkatkan ketepatan kelompok rekomendasi menurun.



Gambar 8. Presisi terhadap gambar no item

Saturation juga meningkatkan nilai  $k$ - sampai batas tertentu dalam kumpulan data yang digunakan digunakan  $k$  antara 2-4 dan memperoleh hasil yang baik

### KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka diperoleh kesimpulan bahwa penggabungan antara metode collaborative filtering dan content based lebih baik daripada hanya menggunakan satu metode. Masing-masing setiap preferensi pengguna menjadi pertimbangan tersendiri dan menghasilkan hasil yang efektif. Metode ini membutuhkan sumber daya (dataset) yang banyak. Dibutuhkan pencarian parameter yang tepat untuk menghasilkan akurasi yang tinggi dengan waktu komputasi yang cepat. Hasil pengujian pada model yang dibangun menunjukkan bahwa hybrid recommender system lebih unggul daripada menggunakan satu metode. Dalam penelitian Anshil Pal, rekomendasi konten hybrid menghasilkan MAE yang lebih baik dan meningkatkan sparsity dataset antara 1% - 2%, hasilnya juga dapat bervariasi ketika diuji dengan dataset yang lebih besar.

### DAFTAR PUSTAKA

1. S. Agrawal, "An Improved Approach for Movie Recommendation System," in 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2013.
2. A. Pal, P. Parhi, dan M. Aggarwal, "An improved content based collaborative filtering algorithm for movie recommendations," 2017 10th Int. Conf. Contemp. Comput. IC3 2017, vol. 2018–Janua, no. August, hal. 1–3, 2018.
3. D. P. D. R., "A Latent Factor Model based Movie Recommender using Smartphone Browsing History," in 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), 2017, hal. 1–6
4. A. Pujahari dan V. Padmanabhan, "Group Recommender Systems: Combining User-User and Item-Item Collaborative Filtering Techniques," Proc. - 2015 14th Int. Conf. Inf. Technol. ICIT 2015, hal. 148–152, 2016.
5. G. Sbrana and A. Silvestrini, "Temi di Discussione Random switching exponential smoothing and inventory forecasting," 2014.