

## SISTEM REKOMENDASI PADA E-COMMERCE MENGGUNAKAN K-NEAREST NEIGHBOR

Chandra Saha Dewa Prasetya

Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada  
Email: chandrasahadp@gmail.com

(Naskah masuk: 19 Juli 2017, diterima untuk diterbitkan: 28 September 2017)

### Abstrak

Semakin banyaknya informasi produk yang ada di internet menghadirkan tantangan baik pembeli maupun pebisnis online dalam lingkungan e-commerce. Pembeli sering mengalami kesulitan saat mencari produk di internet karena banyaknya produk yang dijual di internet. Selain itu, pebisnis online sering mengalami kesulitan karena memiliki data mengenai produk, pembeli, dan transaksi yang sangat banyak, sehingga menyebabkan pebisnis online mengalami kesulitan untuk mempromosikan produk yang tepat pada target pembeli tertentu. Sistem rekomendasi dikembangkan untuk mengatasi permasalahan tersebut dengan berbagai metode seperti *Collaborative Filtering*, *Content based*, dan *Hybrid*. Metode *Collaborative Filtering* menggunakan data rating pembeli, *Content Based* menggunakan konten produk seperti judul atau deskripsi, dan *Hybrid* menggunakan keduanya sebagai dasar rekomendasi. Dengan menggunakan basis data graf, maka model sistem rekomendasi dapat dirancang dengan berbagai metode pendekatan sekaligus. Pada penelitian ini, algoritma *k-Nearest Neighbor* digunakan untuk menentukan top-n rekomendasi produk untuk setiap pembeli. Hasil dari penelitian ini metode *Content Based* mengungguli metode lain karena data yang digunakan *sparse*, yaitu kondisi dimana jumlah rating yang diberikan pembeli relatif sedikit terhadap banyaknya produk yang tersedia pada e-commerce.

**Kata kunci:** sistem rekomendasi, *k-nearest neighbor*, *collaborative filtering*, *content based*.

### Abstract

*The growing number of product information available on the internet brings challenges to both customer and online businesses in the e-commerce environment. Customer often have difficulty when looking for products on the internet because of the number of products sold on the internet. In addition, online businessman often experience difficulties because they has much data about products, customers and transactions, thus causing online businessman have difficulty to promote the right product to a particular customer target. A recommendation system was developed to address those problem with various methods such as Collaborative Filtering, ContentBased, and Hybrid. Collaborative filtering method uses customer's rating data, content based using product content such as title or description, and hybrid using both as the basis of the recommendation. In this research, the k-nearest neighbor algorithm is used to determine the top-n product recommendations for each buyer. The result of this research method Content Based outperforms other methods because the sparse data, that is the condition where the number of rating given by the customers is relatively little compared the number of products available in e-commerce.*

**Keywords:** recommendation system, *k-nearest neighbor*, *collaborative filtering*, *content based*.

## 1. PENDAHULUAN

Semakin banyaknya informasi produk yang ada di internet menghadirkan tantangan baik pembeli maupun pebisnis online dalam lingkungan e-commerce. Pembeli sering mengalami kesulitan saat mencari produk di internet karena banyaknya produk yang dijual di internet. Pebisnis online sering mengalami kesulitan karena memiliki data mengenai produk, pembeli, dan transaksi yang sangat banyak, sehingga menyebabkan pebisnis online mengalami kesulitan untuk mempromosikan produk yang tepat pada target pembeli tertentu. Menurut Knijnenburg et al. (2012) sistem rekomendasi secara otomatis dapat menganalisis penggunaan data calon pembeli untuk

menyaring konten halaman web, mengkategorisasi pesan newsgroup, dan merekomendasikan informasi. Sistem rekomendasi menganalisis data mengenai produk atau interaksi pengguna dan produk untuk menemukan hubungan antara produk dan pengguna. Hasil yang diterima akan ditampilkan sebagai rekomendasi.

Konsep sistem rekomendasi telah digunakan oleh berbagai bisnis online seperti amazon.com dan ebay.com sebagai alat bisnis. Sistem rekomendasi dilaporkan telah meningkatkan penjualan produk dan membangun loyalitas pembeli (Mobasher, 2007). Dalam sistem rekomendasi terdapat beberapa metode yang sering digunakan yaitu *Collaborative Filtering*, *Content Based*, dan *Hybrid*. *Collaborative Filtering* menggunakan riwayat pemilihan atau riwayat penilaian sebagai dasar untuk menentukan

rekomendasi. *Content Based* menggunakan kesamaan produk untuk ditawarkan kepada pembeli. Sedangkan metode *Hybrid* menggabungkan metode dua atau lebih metode untuk menghasilkan rekomendasi yang lebih baik.

Pengembangan sistem rekomendasi sering mengalami 2 tantangan. Yang pertama bagaimana menggambarkan berbagai macam informasi mengenai produk dan pengguna dan bagaimana. Selain itu bagaimana membangun sebuah model yang fleksibel untuk digunakan dengan berbagai metode pendekatan yang berbeda. Pada penelitian ini akan digunakan algoritma k-nearest neighbor untuk menemukan top-n rekomendasi produk untuk setiap pembeli. Algoritma *K-Nearest Neighbor* (KNN) memiliki beberapa kelebihan yaitu kemudahan, efektivitas, intuitif, dan performa klasifikasi yang kompetitif dalam banyak domain (Imandoust dan Bolandraftar, 2013).

## 2. DASAR TEORI

Sistem rekomendasi membantu pengguna untuk mengidentifikasi

### 2.1 Sistem Rekomendasi

Sistem rekomendasi membantu pengguna untuk mengidentifikasi produk yang sesuai dengan kebutuhan, kesenangan, dan keinginan user. Sistem rekomendasi akan membimbing user untuk menemukan produk yang relevan dan berguna dari banyaknya produk yang tersedia.

Menurut sejak pertama Tang et al. (2013) kali ditemukan pada tahun 1990, ada banyak penelitian mengenai sistem rekomendasi dilakukan. Sistem rekomendasi mulai diaplikasikan ke berbagai bidang dengan metode yang berbeda seperti *Content Based*, *Collaborative Filtering*, dan *Hybrid*. Sistem rekomendasi yang menggunakan metode *Content Based* menggunakan kesamaan produk untuk ditawarkan kepada pengguna. Namun, metode *Content Based* memiliki kelemahan yaitu ketika fitur konten yang tersedia terbatas, maka akurasi rekomendasi yang dihasilkan cukup rendah (Yuan et al., 2014).

*Collaborative Filtering* adalah metode yang paling sering digunakan untuk membangun sistem rekomendasi. Metode ini bergantung pada riwayat pemilihan atau riwayat penilaian (Su dan Khoshgoftaar, 2009). Metode *Hybrid* menggabungkan metode *Content Based* dan *Collaborative Filtering* untuk menghasilkan rekomendasi yang lebih baik (Tang et al., 2013).

### 2.2 Vector Space Model

*Vector Space Model* adalah model aljabar untuk merepresentasikan teks dokumen sebagai vektor dan setiap dimensi sesuai dengan sebuah istilah yang terpisah dalam sebuah ruang vektor. Jika sebuah istilah ada dalam suatu dokumen, maka nilai dari

vektor tersebut tidak nol. Misalkan kita mempunyai  $n$  istilah berbeda pada *lexicon*. Kemudian *lexicon*,  $\ell$ , merepresentasikan sebuah kumpulan dari istilah dan dapat didefinisikan pada Persamaan (1).

$$\ell = \{t_1, t_2, t_3, \dots, t_n\} \quad (1)$$

Kemudian, sebuah vektor  $\vec{d}_i$  didefinisikan pada persamaan (2).

$$\vec{d}_i = \{w_{1i}, w_{2i}, w_{3i}, \dots, w_{ni}\} \quad (2)$$

dimana  $w_{ki}$  merepresentasikan bobot istilah ke- $k$  pada dokumen  $i$ . *Vector Space Model* merupakan dasar dalam operasi information retrieval seperti klasifikasi atau klustering dokumen (Danisman dan Alpkocak, 2008).

### 2.3 Cosine Similarity

Setelah dokumen direpresentasikan sebagai vektor, maka derajat kesamaan antara dua dokumen dapat dihitung sebagai korelasi antara 2 vektor yang sesuai. Hal ini dapat diukur sebagai dua sudut antar vektor yang disebut *Cosine Similarity*. *Cosine similarity* adalah salah satu metode yang paling populer yang sering diterapkan pada teks dokumen untuk keperluan temu balik informasi dan klustering. Untuk mencari cosine similarity antara dua dokumen  $\vec{t}_a$  dan  $\vec{t}_b$  ditunjukkan pada Persamaan (3).

$$SIM(\vec{t}_a, \vec{t}_b) = \frac{\vec{t}_a \cdot \vec{t}_b}{|\vec{t}_a| \times |\vec{t}_b|} \quad (3)$$

dimana  $\vec{t}_a$  dan  $\vec{t}_b$  adalah vektor multi dimensional dari kumpulan istilah  $T = \{t_1, t_2, t_3, \dots, t_n\}$ . Setiap dimensi pada vektor merepresentasikan istilah dengan bobotnya pada dokumen yang bernilai non negatif. Hasil dari *Cosine Similarity* bernilai non negatif dan berada diantara nilai 0 dan 1 (Huang, 2008).

### 2.4 KNN

Algoritma KNN adalah salah satu algoritma yang sering digunakan untuk melakukan klasifikasi. Algoritma termasuk dalam algoritma lazy learning yang mudah untuk diimplementasikan (Alkhatib et al., 2013).

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 (Imandoust dan Bolandraftar, 2013). Data latih diubah menjadi vektor dan sebuah jarak dihitung menggunakan beberapa metode, seperti *euclidean distance* atau *cosine similarity*.

Langkah-langkah algoritma KNN:

1. Menentukan parameter  $k$  (jumlah tetangga terdekat)

2. Hitung jarak data latih dengan semua data uji
3. Urutkan jarak tersebut berdasarkan nilai yang terkecil sejumlah k.
4. Tentukan kelompok data uji berdasarkan label mayoritas pada k.

## 2.4 Evaluasi

Metode yang paling dasar dan sering digunakan untuk mengukur performa sistem adalah precision dan recall (Christopher et al, 2008). Gagasan ini dapat dijelaskan dengan confusion matrix pada Tabel 1.

Tabel 1. Confussion matrix

	Prediksi	
	Relevan	Tidak Relevan
Relevan	True Positive	False Positive
Tidak Relevan	False Negative	True Negative

Jika data positif dan diprediksi positif akan dihitung sebagai True Positive (TP), tetapi jika data itu diprediksi negatif maka akan dihitung sebagai False Negative (FN). Jika data negatif dan diprediksi negatif akan dihitung sebagai true negative, tetapi jika data tersebut diprediksi positif maka akan dihitung sebagai False Positive (FP).

Dari Tabel 1 dapat dihitung nilai *precision*, *recall*, dan *F-measure*. *Precision* adalah bagian dari dokumen yang diambil secara benar. Persamaan (4) adalah persamaan untuk menghitung nilai *precision*.

$$precision = \frac{TP}{TP+FP} \quad (4)$$

*Recall* adalah bagian dari dokumen yang relevan yang diambil. Persamaan (5) adalah persamaan untuk menghitung nilai *recall*.

$$recall = \frac{TP}{TP+FN} \quad (5)$$

*F-measure* adalah nilai yang mewakili kinerja sistem yang merupakan rata-rata dari nilai *precision* dan *recall* (Christopher et al, 2008). Persamaan (6) adalah persamaan untuk menghitung nilai *F-measure*.

$$F - measure = \frac{2PR}{P+R} \quad (6)$$

## 3. DATASET

Dataset yang digunakan diambil dari laman Stanford Network Analysis Platform (SNAP) yang berisi metadata produk produk pada e-commerce amazon pada musim panas 2005. Untuk setiap produk terdapat beberapa informasi berikut yaitu judul, salesrank, kategori, group, dan review produk seperti meliputi waktu, rating, dan pembeli. Cuplikan data ditunjukkan pada Gambar 1.

Id: 1	
ASIN: 0827229534	
title: Patterns of Preaching: A Sermon Sampler	
group: Book	
salesrank: 396585	
similar: 5 0804215715 156101074X 0687023955 0687074231 082721619X	
categories: 2	
[Books[283155]]Subjects[1000]Religion & Spirituality[22]Christianity[12290]	
[Books[283155]]Subjects[1000]Religion & Spirituality[22]Christianity[12290]	
reviews: total: 2 downloaded: 2 avg rating: 5	
2000-7-28 customer: A2JW67OY8U6HHK rating: 5 votes: 10 helpful: 9	
2003-12-14 customer: A2VE83MZF98ITY rating: 5 votes: 6 helpful: 5	

Gambar 1. Cuplikan data yang digunakan

## 4. PERANCANGAN SISTEM

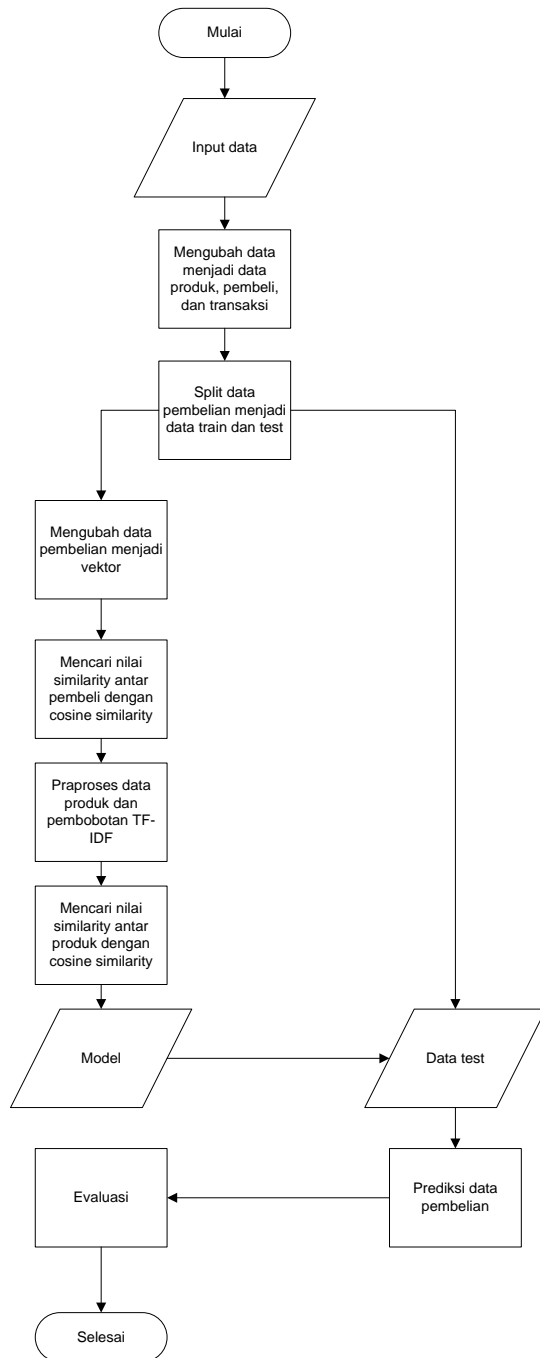
Pada penelitian ini, preprocessing dilakukan terhadap dataset yang berformat txt ke menjadi 3 data yaitu data produk, data pembeli, dan data transaksi. Data yang telah diolah kemudian disimpan dalam format *comma separated value* (csv).

Kemudian proses *5-fold cross validation* dilakukan pada data transaksi yang terdiri dari id pembeli, id produk, dan rating produk untuk membagi dataset menjadi data train dan data test. Data train adalah data yang digunakan untuk membentuk model. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Data test adalah data yang digunakan untuk mengukur sejauh mana model dapat melakukan prediksi dengan benar.

Selanjutnya data transaksi pembeli diubah menjadi vektor dengan dimensi union produk yang akan dibandingkan. Vektor tersebut kemudian dibandingkan dengan vektor lain menggunakan *cosine similarity*. Pada data produk, terdapat beberapa fitur yang dapat digunakan untuk mencari similarity yaitu judul, grup, dan kategori. Untuk masing-masing fitur tersebut dilakukan pembobotan TF-IDF.

Selanjutnya dilakukan vektorisasi dengan dimensi union dari kata yang dibandingkan. *Cosine similarity* kemudian dilakukan untuk mengetahui similarity antara dua vektor yang dibandingkan.. Selanjutnya algoritma KNN dengan metode *Collaborative Filtering*, *Content Based*, atau *Hybrid* dijalankan untuk mencari top-n rekomendasi dengan kepada suatu pembeli. Metode *Collaborative Filtering* menggunakan similarity antar pembeli sebagai dasar rekomendasi.

Metode *Content Based* menggunakan similarity antar produk sebagai dasar rekomendasi. Sedangkan metode *Hybrid* menggabungkan hasil dari *Collaborative Filtering* dan *Content Based* sebagai dasar rekomendasi. Data test hasil dari proses *5-fold cross validation* digunakan sebagai masukan untuk melakukan prediksi. Hasil dari prediksi kemudian dihitung nilai *presisi*, *recall*, dan *F-measure* untuk mengetahui performa metode yang digunakan. Perancangan sistem pada penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Perancangan Sistem

## 5. PENGUJIAN DAN HASIL

Pengujian dilakukan untuk mengetahui performa metode *Collaborative Filtering*, *Content Based*, dan *Hybrid*. Pengujian dilakukan dengan menghitung nilai *precision*, *recall*, dan *F-measure* dari masing-masing metode. Proses pengujian dilakukan dengan metode 5-fold Cross Validation yaitu membagi data menjadi 5 bagian, kemudian satu bagian pertama dijadikan data latih dan data lainnya menjadi data uji.

Kemudian bagian kedua dijadikan data train dan bagian lainnya dijadikan data test, begitu seterusnya sampai bagian kelima. Sehingga persentase data yang

digunakan untuk penelitian ini adalah 80% data train dan 20% data test. Untuk setiap pembeli akan diberikan rekomendasi produk sebanyak 10, 30, 50, 80, dan 100 dari total 9467 produk yang digunakan pada penelitian ini.

### 5.1. Contoh Pengujian

Untuk setiap pembeli pada data pengujian akan diberikan rekomendasi produk. Selanjutnya akan dicari nilai *precision* dan *recall* untuk setiap pembeli. Pengujian dilakukan menggunakan parameter *k* yaitu 10, 30, 50, 80, 100. Contoh hasil pengujian ditunjukkan pada Tabel 2.

Tabel 2. Contoh pengujian

Kode pembeli	Rekomendasi	Produk dibeli
A1JTG	18059, 7448,	
	13886, 34371,	7448, 1231
	34772, 24475,	34772, 1213
	17361, 30919,	12021, 2341
	21314, 12432	10231, 2134
	12301, 12344	

Dari Tabel 2 nilai *precision* dan *recall* kemudian dihitung sebagai yaitu  $precision = 2/(2 + 10) = 0,16$  dan  $recall = 2/(2 + 8) = 0,20$ .

Proses pengujian untuk setiap percobaan dilakukan terhadap 1323 pembeli. Nilai rata-rata *precision* dan *recall* adalah nilai *precision* dan *recall* pada percobaan tersebut. Nilai *F-measure* untuk percobaan tersebut kemudian dicari menggunakan nilai rata-rata *precision* dan *recall*.

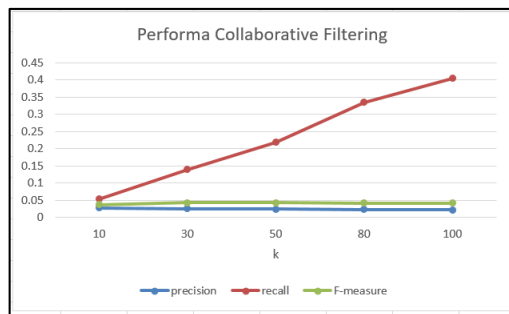
### 5.2. Pengujian Metode *Collaborative Filtering*

Pada sistem rekomendasi menggunakan metode ini, didapatkan hasil seperti Tabel 3.

Tabel 3. Pengujian Metode *Collaborative Filtering*

k	<i>precision</i>	<i>recall</i>	<i>F-measure</i>
10	0.028	0.053	0.036
30	0.025	0.138	0.0424
50	0.023	0.218	0.0425
80	0.022	0.334	0.0420
100	0.021	0.404	0.041

Dari Tabel 3, metode *Collaborative Filtering* menghasilkan *precision* terbaik pada percobaan dengan parameter *k*=10 yaitu sebesar 0.028. *Recall* terbaik dihasilkan pada percobaan menggunakan *k*=100 yaitu sebesar 0.404. *F-measure* terbaik dihasilkan pada percobaan dengan parameter *k*=50 yaitu sebesar 0.0425. Nilai *precision* terendah dihasilkan pada percobaan dengan parameter *k*=100 yaitu sebesar 0.021. Nilai *recall* dan *F-measure* terendah dihasilkan pada percobaan dengan parameter *k*=10 yaitu sebesar 0.053 dan 0.036. Performa metode *Collaborative Filtering* ditunjukkan pada Gambar 3.



Gambar 3. Performa Collaborative Filtering

Gambar 3 menunjukkan semakin rendah nilai  $k$  maka nilai *precision* dan *F-measure* akan semakin tinggi. Namun nilai *recall* akan semakin rendah. Hal ini disebabkan karena semakin tinggi nilai  $k$ , maka nilai false positive yaitu barang yang direkomendasikan namun tidak dibeli akan semakin tinggi.

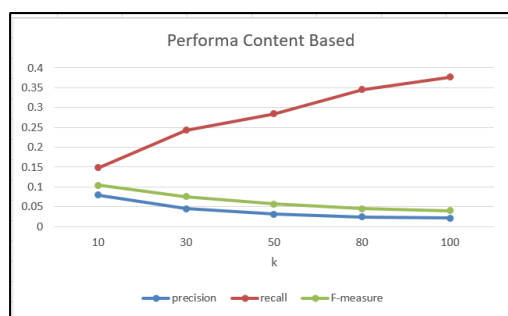
### 5.3. Pengujian Metode Content Based

Pada sistem rekomendasi menggunakan metode ini, didapatkan hasil seperti Tabel 4.

Tabel 4. Pengujian Metode Content Based

k	precision	recall	F-measure
10	0.080	0.148	0.103
30	0.044	0.242	0.075
50	0.031	0.284	0.057
80	0.024	0.345	0.045
100	0.021	0.377	0.040

Dari Tabel 4, metode *Content Based* menghasilkan *precision*, dan *F-measure* terbaik pada percobaan menggunakan parameter  $k=10$  yaitu sebesar 0.080 dan 0.103. *Recall* tertinggi dihasilkan pada percobaan menggunakan  $k=100$  yaitu sebesar 0.377. Sedangkan nilai *precision*, dan *F-measure* terendah dihasilkan pada percobaan menggunakan nilai  $k=100$  yaitu sebesar 0.0212 dan 0.040. *Recall* terendah dihasilkan pada percobaan menggunakan  $k=10$  yaitu sebesar 0.148. Performa metode *Content Based* ditunjukkan pada Gambar 4.



Gambar 4. Performa Content Based

Dari Gambar 4 dapat disimpulkan semakin rendah nilai  $k$  maka nilai *precision* dan *F-measure* akan semakin tinggi. Namun nilai *recall* akan semakin rendah. Hal ini disebabkan karena semakin

tinggi nilai  $k$ , maka nilai false positive yaitu barang yang direkomendasikan namun tidak dibeli akan semakin tinggi.

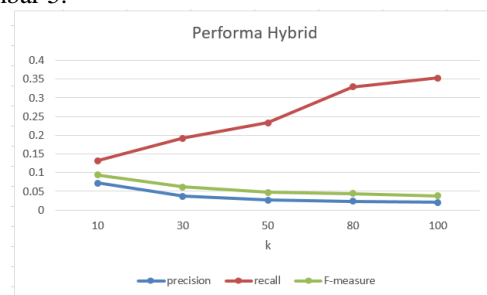
### 5.4. Pengujian Metode Hybrid

Pada sistem rekomendasi menggunakan metode ini, didapatkan hasil seperti Tabel 5.

Tabel 5. Pengujian Metode Hybrid

K	precision	recall	F-measure
10	0.072	0.1323	0.093
30	0.036	0.192	0.061
50	0.026	0.233	0.057
80	0.023	0.329	0.043
100	0.020	0.353	0.038

Dari Tabel 5, metode *Hybrid* menghasilkan *precision*, dan *F-measure* terbaik pada percobaan menggunakan parameter  $k=10$  yaitu sebesar 0.072 dan 0.093. *Recall* tertinggi dihasilkan pada percobaan menggunakan  $k=100$  yaitu sebesar 0.353. Sedangkan nilai *precision*, dan *F-measure* terendah dihasilkan pada percobaan menggunakan nilai  $k=100$  yaitu sebesar 0.020 dan 0.038. *Recall* terendah dihasilkan pada percobaan menggunakan  $k=10$  yaitu sebesar 0.132. Performa metode *Hybrid* ditunjukkan pada Gambar 5.



Gambar 5. Performa Hybrid

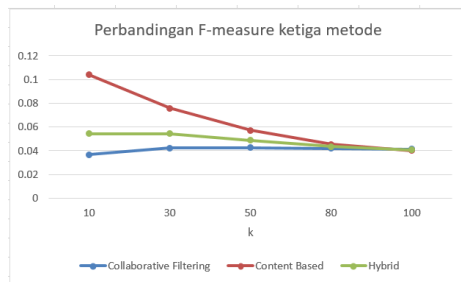
Dari Gambar 5 dapat disimpulkan semakin rendah nilai  $k$  maka nilai *precision* dan *F-measure* akan semakin tinggi. Namun nilai *recall* akan semakin rendah. Hal ini disebabkan karena semakin tinggi nilai  $k$ , maka nilai false positive yaitu barang yang direkomendasikan namun tidak dibeli akan semakin tinggi.

### 5.5. Perbandingan Performa Ketiga Metode

Dari ketiga metode yang diujikan, maka nilai rata-rata *F-measure* yang dihasilkan dapat dibandingkan untuk mengetahui performa metode mana yang menghasilkan hasil terbaik. Pada Gambar 6 merupakan hasil perbandingan sistem rekomendasi dari ketiga metode yang diaplikasikan.

Tabel 6.4 menunjukkan bahwa metode *Content Based* menghasilkan nilai *F-measure* tertinggi pada parameter  $k=10$  yaitu sebesar 0.103. Sedangkan sistem rekomendasi menggunakan metode *Collaborative Filtering* dengan parameter  $k=10$

menghasilkan nilai *F-measure* paling rendah yaitu sebesar 0.036.



Gambar 6. Perbandingan *F-measure* ketiga metode

Pada penelitian ini metode *Content Based* menghasilkan *F-measure* yang lebih baik dari metode *Collaborative Filtering* dan *Hybrid* karena data pembelian sangat *sparse*. Kondisi tersebut terjadi karena jumlah transaksi setiap pembeli relatif sedikit dibandingkan banyaknya produk yang tersedia. Pada penelitian ini, seorang pembeli rata-rata hanya memberikan rating terhadap 38 produk dari 9467 produk yang tersedia. Data yang *sparse* tersebut mengakibatkan turunnya performa metode *Collaborative Filtering*.

Sedangkan metode *Content Based* dapat menghasilkan performa yang lebih baik saat data pembelian yang dilakukan oleh seorang pembeli sedikit. Hal ini disebabkan karena metode ini menentukan rekomendasi berdasarkan data produk seperti judul, grup, dan kategori.

## 6. KESIMPULAN DAN SARAN

Berdasarkan dataset yang digunakan pada penelitian ini, didapatkan bahwa metode *Content Based* menghasilkan nilai rata-rata precision dan *F-measure* paling tinggi dibandingkan *Collaborative Filtering* dan *Hybrid* pada  $k=10$  yaitu sebesar 0.080 dan 0.148. Sedangkan *recall* tertinggi dihasilkan dengan metode *Collaborative Filtering* pada  $k=100$  yaitu 0.404.

Dari percobaan yang dilakukan pada ketiga metode apabila nilai  $k$  yang semakin tinggi, maka nilai *recall* yang dihasilkan juga semakin tinggi. Hal ini disebabkan karena semakin banyak rekomendasi produk yang diberikan kepada pembeli, maka nilai true positive yaitu barang yang direkomendasikan dan dibeli akan meningkat. Selain itu nilai false positive tidak mempengaruhi *recall* sehingga nilainya tetap tinggi.

Pada penelitian ini, masih terdapat sejumlah keterbatasan dan kekurangan. Salah satu kelemahan menggunakan algoritma KNN adalah nilai parameter  $k$  perlu dicari terlebih dahulu untuk mendapatkan hasil yang paling optimal. Selain itu biaya komputasinya juga cukup tinggi karena perhitungan jarak dilakukan antara data uji dengan setiap data latih.

Penelitian selanjutnya dapat menggunakan data yang lebih besar dari penelitian ini. Selanjutnya

metode lain seperti *adjusted cosine similarity* dapat diterapkan untuk mencari bobot similarity antar produk maupun antar pembeli untuk meningkatkan nilai *precision*, *recall*, dan *F-measure*.

## 7. DAFTAR PUSTAKA

- ALKHATIB, K., NAJADAT, H., HMEIDI, I. & SHATNAWI, M.K.A. 2013. Stock price prediction using k-nearest neighbor (kNN) algorithm. *International Journal of Business, Humanities and Technology*, 3(3), 32-44.
- CHOI, K., YOO, D., KIM, G. & SUH, Y. 2012. A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis. *Electronic Commerce Research and Applications*, 11(4), 309-317.
- CHRISTOPHER, D.M., PRABHAKAR, R. & HINRICH, S.C.H.Ü.T.Z.E. 2008. Introduction to information retrieval. *An Introduction To Information Retrieval*, 151, 177.
- DANISMAN, T. & ALPKOÇAK, A. 2008, April. Feeler: Emotion classification of text using vector space model. In *AISB 2008 Convention Communication, Interaction and Social Intelligence* (Vol. 1, p. 53).
- DESYAPUTRI, D.M., ERWIN, A., GALINIUM, M. & NUGRAHADI, D. 2013, October. News recommendation in Indonesian language based on user click behavior. In *Information Technology and Electrical Engineering*, 164-169.
- HUANG, A. 2008, April. Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008)*, Christchurch, New Zealand. 49-56.
- IMANDOUST, S.B. & BOLANDRAFTAR, M. 2013. Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International Journal of Engineering Research and Applications*, 3(5), 605-610.
- KNIJNENBURG, B. P., WILLEMSSEN, M. C., GANTNER, Z., SONCU, H., & NEWELL, C., 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5), 441-504.
- LUND, S.S. & TANDBERG, Ø. 2015. Design of a Hybrid Recommender System: A Study of the Cold-Start User Problem (Master's thesis, NTNU).

- MA, K. 2016. Content-based Recommender System for Movie Website.
- MOBASHER, B. 2007. Data mining for web personalization. In *The adaptive web*, 90-135. Springer Berlin Heidelberg.
- SAHAL, R., SELIM, S. & ELKORANY, A. 2014. An Adaptive Framework for Enhancing Recommendation Using Hybrid Techniques. *International Journal of Computer Science & Information Technology*, 6(2), 51.
- SU, X. & KHOSHGOFTAAR, T.M. 2009. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 4.
- TANG, J., HU, X. & LIU, H. 2013. Social recommendation: a review. *Social Network Analysis and Mining*, 3(4), 1113-1133.
- VAINIONPÄÄ, I., & DAVIDSSON, S. 2014. Stock market prediction using the K Nearest Neighbours algorithm and a comparison with the moving average formula.
- YANG, X., GUO, Y. & LIU, Y. 2013. Bayesian-inference-based recommendation in online social networks. *IEEE Transactions on Parallel and Distributed Systems*, 24(4), 642-651.
- Yin, H., Sun, Y., Cui, B., Hu, Z. & Chen, L. 2013, August. Lcars: a location content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, 221-229. ACM.
- YUAN, Q., CONG, G. & LIN, C.Y. 2014, August. COM: a generative model for group recommendation. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 163-172. ACM.