

SISTEM REKOMENDASI METODE CONTENT-BASED FILTERING

PADA ITEM PRODUK DI CASEME

SKRIPSI



PUJI SETIAWAN

17.11.1466

PROGRAM SARJANA

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2021

**SISTEM REKOMENDASI METODE CONTENT-BASED FILTERING
PADA ITEM PRODUK DI CASEME**

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

PUJI SETIAWAN

17.11.1466

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2021

PERSETUJUAN

SKRIPSI

SISTEM REKOMENDASI METODE CONTENT-BASED FILTERING

PADA ITEM PRODUK DI CASEME

yang dipersiapkan dan disusun oleh

Puji Setiawan

17.11.1466

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 04 Agustus 2021

Dosen Pembimbing,

Arif Dwi Laksito, M. Kom

NIK. 190302150

PENGESAHAN
SKRIPSI
SISTEM REKOMENDASI METODE CONTENT-BASED FILTERING
PADA ITEM PRODUK DI CASEME

yang dipersiapkan dan disusun oleh

Puji Setiawan

17.11.1466

telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Agustus 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Lukman, M. Kom.

NIK. 190302151

Bety Wulan Sari, M.Kom.

NIK. 190302254

Arif Dwi Laksito, M. Kom

NIK. 190302150

Skripsi ini telah diterima sebagai salah satu persyaratan

untuk memperoleh gelar Sarjana Komputer

Tanggal 26 Agustus 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S. Kom., M. Kom.

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan di bawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan tidak dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk mendapatkan gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu terkait dalam naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 9 September 2021



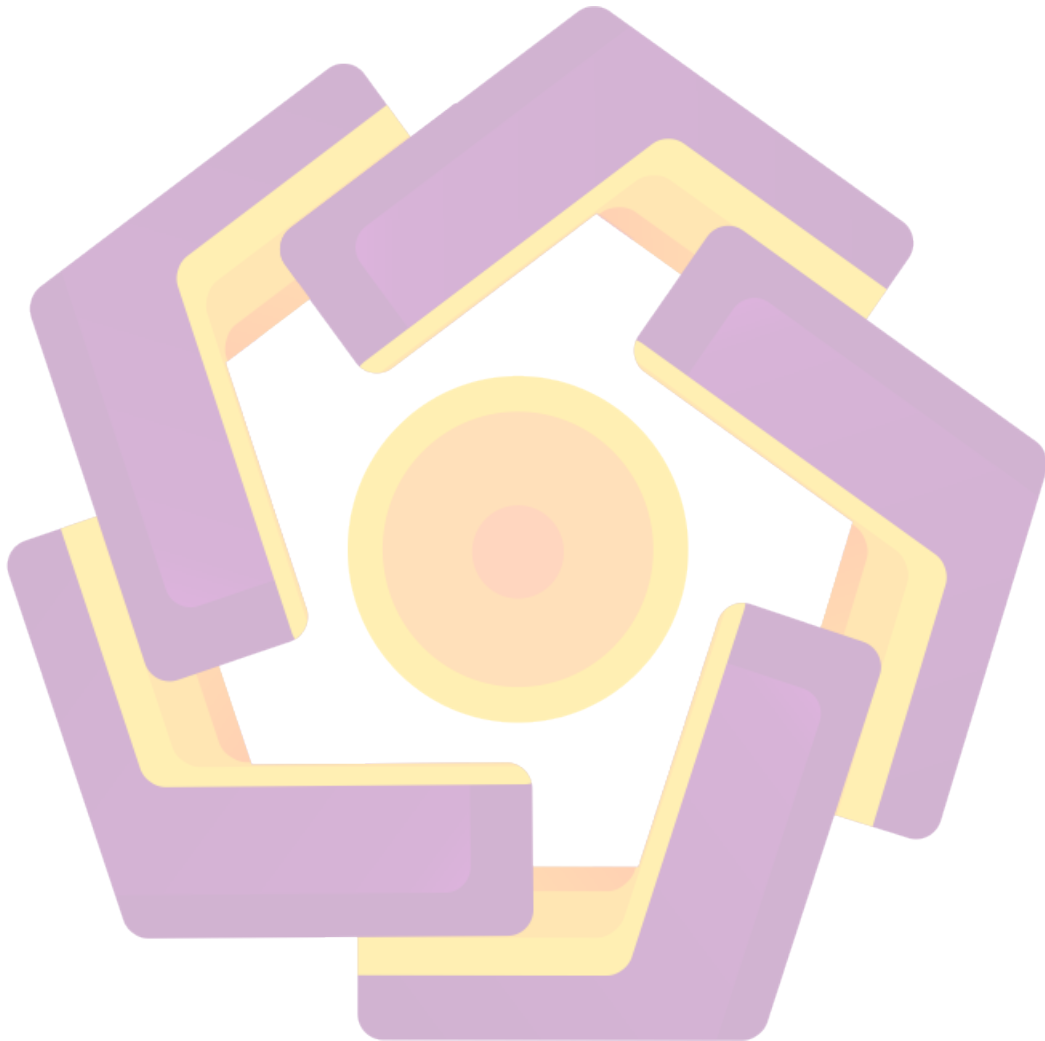
(Puji Setiawan)

NIM 17.11.1466

MOTTO

“Setiap orang punya gilirannya sendiri-sendiri, bersabarlah dan tunggulah”

-Gol D. Roger



PERSEMBAHAN

Alhamdulillah dengan semua kemudahan yang Allah SWT berikan kepada saya sehingga dapat menyelesaikan skripsi ini dengan baik dan tidak banyak pusingnya. Dengan ini saya mempersembahkan skripsi ini kepada semua pihak yang ikut terlibat secara langsung maupun tidak langsung, yaitu:

1. Allah SWT yang telah memberikan rahmat, nikmat, serta hidayahnya sehingga saya bisa menyelesaikan skripsi ini dengan baik.
2. Kedua orang tua saya yang selalu memberikan doa, dukungan serta semangat sehingga saya dapat menyelesaikan skripsi ini.
3. Kakak dan adik saya yang selalu mensupport saya dan memberi motivasi terhadap saya.
4. Bapak Arif Dwi Laksito, M.Kom selaku dosen pembimbing yang telah memberikan saran, bantuan, masukan, dan bimbingan untuk menyelesaikan skripsi ini.
5. Teman-teman saya Ibtisamah, Dinda Trisna Aristya, Melati Puspa C, M. Yanuar , yang selalu mendengarkan keluh kesah dan memberikan dukungan.
6. Teman-teman 17-IF08 yang telah mau diajak berdiskusi dan memberikan semangat untuk berjuang menyelesaikan skripsi ini.
7. Teman-teman Angkringan Suyud yang selalu support dan memberikan hiburan ketika penat.
8. Kepada Irfansyah selaku pemilik Caseme yang telah memberikan perijinan untuk melakukan penelitian.
9. Dan seluruh pihak yang sudah membantu saya yang tidak dapat saya sebutkan satu persatu, saya ucapkan terimakasih.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb

Puji syukur kita ucapkan kehadiran Allah SWT, yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini. Shalawat dan salam tidak lupa kita ucapkan kepada junjungan kita nabi Muhammad SAW, serta keluarga dan para sahabatnya yang memberi syafaatnya kepada kita semua sehingga kita dalam perlindungan-Nya.

Skripsi ini disusun sebagai salah satu syarat kelulusan mahasiswa UNIVERSITAS AMIKOM YOGYAKARTA. Selain itu juga merupakan bukti mahasiswa menyelesaikan kuliah jenjang Strata-1 dan untuk memperoleh gelar sarjana komputer.

Keberhasilan ini tidak lepas dari bantuan berbaai pihak. Dalam kesempatannya, penulis berterima kasih :

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Bapak Arif Dwi Laksito, M.Kom. selaku dosen pembimbing yang selalu memberikan kemudahan, nasehat baik, dan waktunya selama penulisan skripsi ini.
5. Bapak Lukman, M. Kom. dan ibu Bety Wulan Sari, M.Kom. selaku dosen penguji, terimakasih atas saran yang diberikan selama pengujian untuk memperbaiki penelitian ini menjadi lebih baik lagi.
6. Bapak dan Ibu dosen UNIVERSITAS AMIKOM YOGYAKARTA yang telah memberikan ilmu saat perkuliahan.

7. Semua pihak yang membantu baik dukungan moral maupun materil, pikiran, dan tenaga dalam penyelesaian skripsi ini.

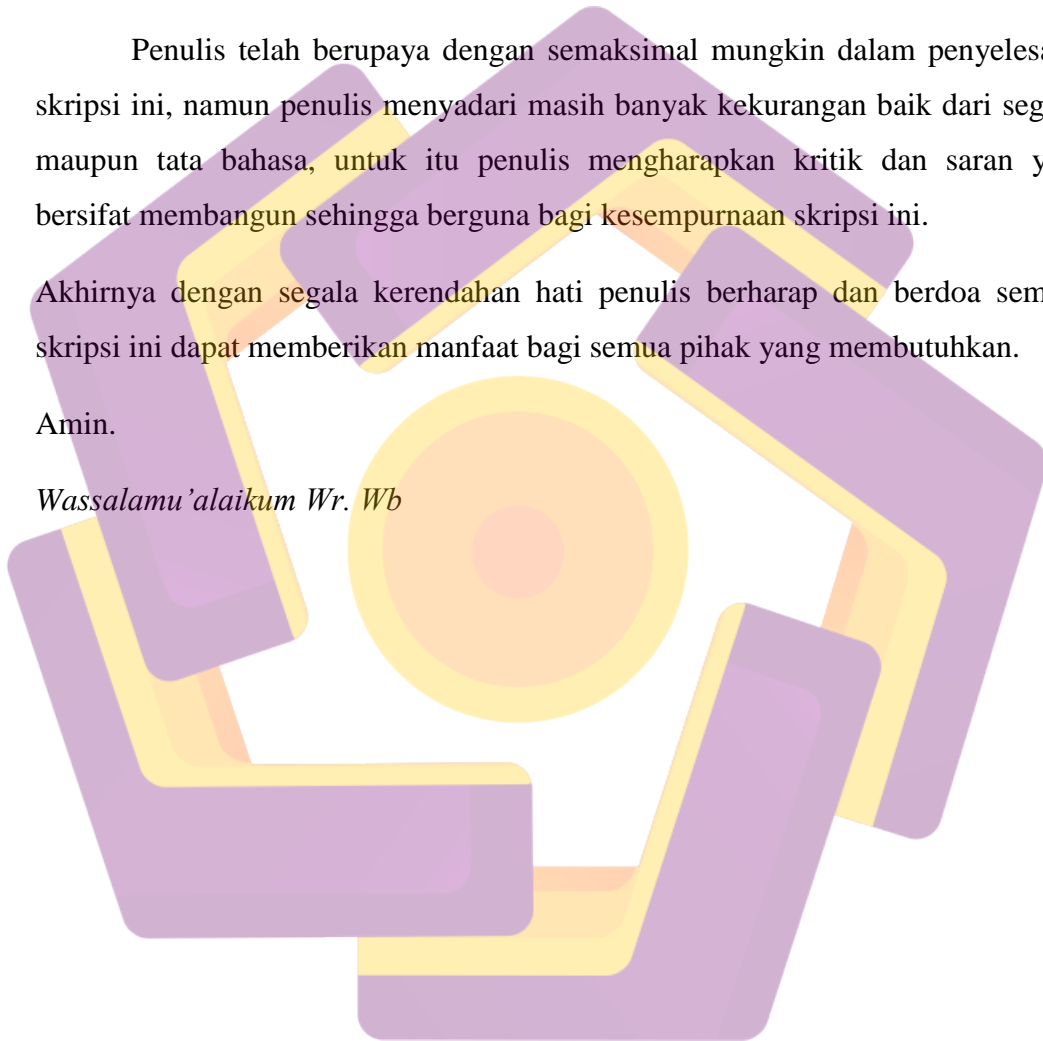
Ucapan terima kasih juga penulis sampaikan kepada pihak yang tidak bisa disebutkan satu per-satu semoga Allah SWT membalas kebaikan dan melimpahkan rahmat dan hidayah-Nya kepada kita semua.

Penulis telah berupaya dengan semaksimal mungkin dalam penyelesaian skripsi ini, namun penulis menyadari masih banyak kekurangan baik dari segi isi maupun tata bahasa, untuk itu penulis mengharapkan kritik dan saran yang bersifat membangun sehingga berguna bagi kesempurnaan skripsi ini.

Akhirnya dengan segala kerendahan hati penulis berharap dan berdoa semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang membutuhkan.

Amin.

Wassalamu 'alaikum Wr. Wb



Yogyakarta, 29 April 2021

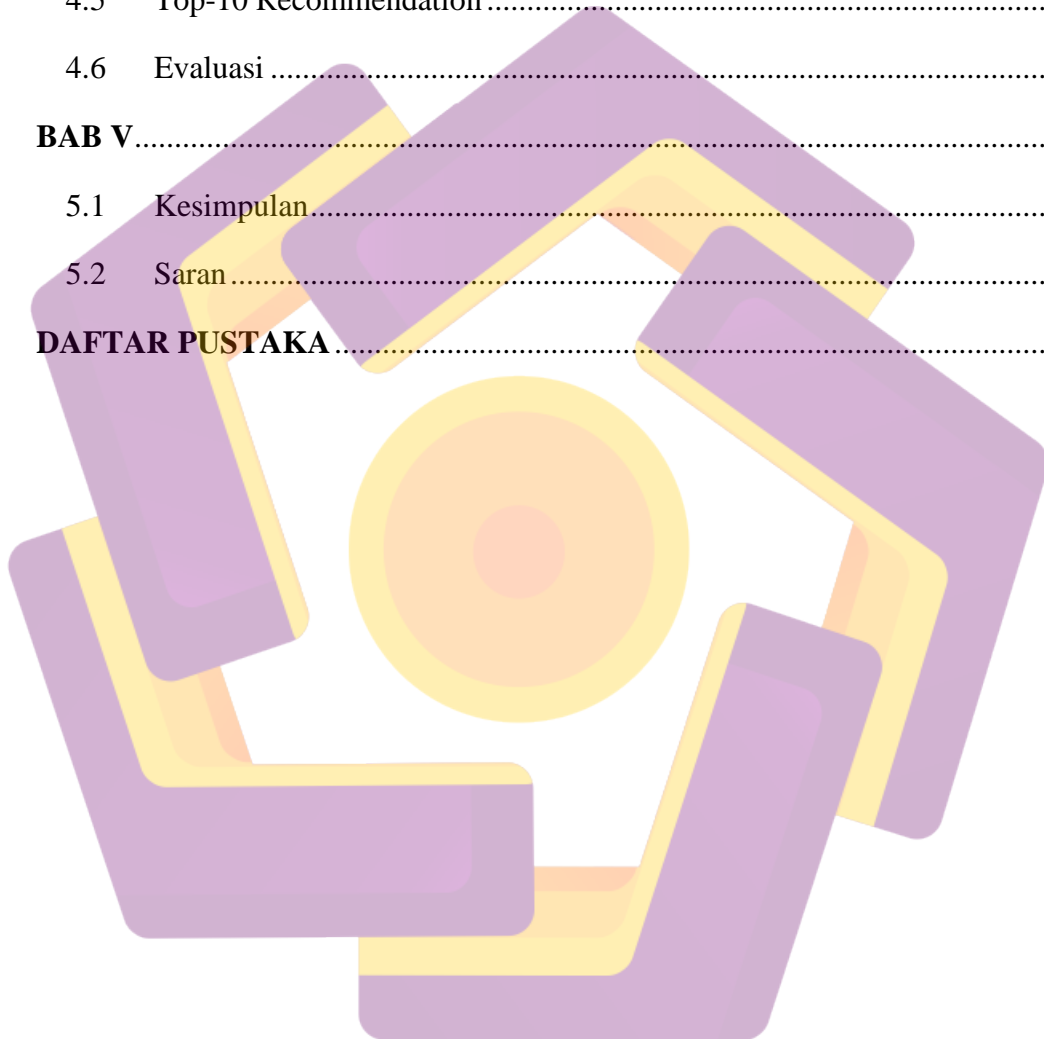
Puji Setiawan

DAFTAR ISI

SKRIPSI	i
PERSETUJUAN	iii
PERNYATAAN	iii
PENGESAHAN	iv
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
INTISARI	xv
ABSTRACT	xvi
BAB 1	2
1.1 Latar Belakang.....	2
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat Penelitian.....	4
1.5 Metode Penelitian.....	4
1.5.1 Metode Pengumpulan Data.....	4
1.5.2 Metode Analisis.....	4
1.5.3 Metode Perancangan.....	5
1.5.5 Metode Implementasi.....	5
1.5.6 Metode Testing.....	5

1.6	Sistematika Penulisan.....	5
BAB II	7
2.1	Tinjauan Pustaka	7
2.2	Landasan Teori.....	9
2.2.1	Sistem rekomendasi	9
2.2.2	Content-Based Filtering	10
2.2.3	Preprocessing	10
2.2.4	Pembobotan Term Frequency-Invers Document Frequency (TF-IDF)	12
2.2.5	Cosine Similarity.....	14
2.2.6	Confusion Matrix	14
BAB III	16
3.1	Alur Penelitian.....	16
3.1.1	Pengumpulan data	17
3.1.2	Preprocessing Data.....	18
3.1.3	TF-IDF Vectorizer	19
3.1.4	Cosine Similarity.....	21
3.1.5	Top-10 Recommendation.....	23
3.1.6	Evaluasi.....	24
3.2	Alat dan Bahan Penelitian	26
BAB IV	27
4.1	Pengumpulan data	27
4.2	Preprocessing.....	28
4.2.1	Case Folding	28
4.2.2	Tokenisasi	28

4.2.3	Filtering	29
4.2.4	Stemming	32
4.3	TF-IDF.....	32
4.4	Cosine Similarity	33
4.5	Top-10 Recommendation	35
4.6	Evaluasi	36
BAB V	41
5.1	Kesimpulan.....	41
5.2	Saran.....	42
DAFTAR PUSTAKA	43



DAFTAR TABEL

Tabel 1 Studi Literatur	7
Tabel 2.1 Case Folding	9
Tabel 2.2 Tokenizing	10
Tabel 2.3 Filtering	10
Tabel 2.4 Stemming	10
Tabel 2.5 Confusion Matriks	14
Tabel 3.1 Data Sampel	16
Tabel 3.2 Pre-processing data sampel	18
Tabel 3.3 TF-IDF data sampel	18
Tabel 3.4 Panjang vektor data sampel	19
Tabel 3.5 Hasil presentase kemiripan data sampel	22
Tabel 3.6 Urutan kemiripan dari terbesar data sampel	23
Tabel 3.7 Evaluasi data sampel	24
Tabel 4.1 Confusion matrix metadataA	36
Tabel 4.2 Confusion matrix metadataB	37

DAFTAR GAMBAR

Gambar 4.1 Dataset	26
Gambar 4.2 MetadataA	26
Gambar 4.3 MetadataB	26
Gambar 4.4 Case folding metadataA	27
Gambar 4.5 Case folding metadataB	27
Gambar 4.6 Tokenisasi metadataA	28
Gambar 4.7 Tokenisasi metadataB	28
Gambar 4.8 Stopwords removal metadataA	28
Gambar 4.9 Stopwords removal metadataB	29
Gambar 4.10 Punctuation removal metadataA	29
Gambar 4.11 Punctuation removal metadataB	29
Gambar 4.12 Number removal metadataA	30
Gambar 4.13 Number removal metadataB	30
Gambar 4.14 Stemming metadataA	31
Gambar 4.15 Stemming metadataB	31
Gambar 4.16 TF-IDF metadataA	32
Gambar 4.17 TF-IDF metadataB	32
Gambar 4.18 Cosine similarity metadataA	33
Gambar 4.19 Cosine similarity metadataB	33
Gambar 4.20 Top-10 recommendation metadataA	34
Gambar 4.21 Top-10 recommendation metadataB	35
Gambar 4.22 Confusion matrix	36
Gambar 4.23 Diagram rata-rata akurasi tiap kategori	39

INTISARI

Banyaknya produk dalam sebuah *e-commerce* merupakan salah satu masalah bagi pengguna untuk memilih produk yang disukai. Untuk itu maka diperlukan sebuah sistem rekomendasi. Salah satu sistem rekomendasi yaitu dengan metode *content-based filtering*. Metode ini dapat mempermudah pengguna dalam melihat produk lain sesuai dengan kemiripan preferensi pengguna. Keuntungan sistem rekomendasi ini yaitu dapat mempengaruhi pengguna dalam memilih item produk tertentu. Peneliti melakukan penelitian tentang bagaimana membuat sistem rekomendasi dengan metode *content-based filtering* kemudian melakukan beberapa kombinasi metadata untuk mencari hasil yang optimal.

Tahap pertama pembuatan metode *content-based filtering* adalah pengumpulan data. Kemudian data itu dibuat kombinasi beberapa field dengan nama metadata A dan metadata B. Metadata A berisi gabungan dari dua *field* yaitu nama dan deskripsi. Metadata B berisi gabungan dari tiga *field* yaitu nama, kategori dan deskripsi. Sebelum data diolah maka dilakukan *preprocessing* terlebih dahulu. Dalam *preprocessing* ini terdapat beberapa proses yaitu case folding, tokenisasi, filtering dan stemming. Data hasil *preprocessing* tersebut diubah menjadi vektor dan diberi pembobotan kata. Setelah data menjadi vector kemudian dicari nilai kemiripan tertinggi menggunakan algoritma *cosine similarity*. Data tersebut kemudian diurutkan dan diambil urutan 10 tertinggi sebagai hasil rekomendasi untuk pengguna. Untuk melihat apakah sistem ini bekerja dengan baik dilakukan evaluasi menggunakan metode *confusion matrix*.

Dari hasil evaluasi didapatkan nilai rata-rata akurasi sebesar 90,4% untuk metadata A dan 67,6% metadata B. Akurasi lebih besar pada metadata A dikarenakan pada metadata B terdapat field deskripsi produk yang memiliki kesamaan antar kategori. Karena nilai metadata A lebih besar maka dari itu dapat disimpulkan bahwa kombinasi dari 2 field yaitu nama dan kategori hasilnya lebih baik untuk rekomendasi produk.

Kata kunci : sistem, rekomendasi, metadata

ABSTRACT

The number of products in an e-commerce is one of the problems for users to choose the preferred product. For this reason, a recommendation system is needed. One of the recommendation systems is the content-based filtering method. This method can make it easier for users to see other products according to the similarity of user preferences.

The advantage of this recommendation system is that it can influence users in choosing certain product items. Researchers conducted research on how to create a recommendation system with content-based filtering methods and then performed several combinations of metadata to find optimal results.

The first stage of making content-based filtering method is data collection. Then the data is made a combination of several fields with the name metadata A and metadata B. Metadata A contains a combination of two fields, namely name and description. Metadata B contains a combination of three fields namely name, category and description. Before the data is processed, preprocessing is carried out first. In this preprocessing there are several processes, namely case folding, tokenization, filtering and stemming. The preprocessed data is converted into vectors and given word weights. After the data becomes vector then the highest similarity value is searched using the cosine similarity algorithm. The data is then sorted and the highest 10 is taken as a result of recommendations for users. To see if this system works well, an evaluation is carried out using the confusion matrix method.

From the evaluation results, the average accuracy value is 90.4% for metadata A and 67.6% for metadata B. The accuracy is greater in metadata A because in metadata B there is a product description field that has similarities between categories. Because the metadata value of A is greater, it can be concluded that the combination of 2 fields, namely name and category, results in better product recommendations.

Keywords: *system, recommendation, metadata*