

# Sistem Rekomendasi Artikel Ilmiah Berbasis Web Menggunakan Content-based Learning dan Collaborative Filtering

Betharia Sri Fitrianti, Muhammad Fachurrozi\*, Novi Yusliani  
Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya  
Jl. Palembang-Prabumulih, KM 33, Indralaya, Ogan Ilir, Sumatera Selatan, Indonesia  
Email: obetsobets@gmail.com

**Abstract**— Penelitian ini mengimplementasikan metode content-based learning dan collaborative filtering pada sistem rekomendasi artikel ilmiah berbasis web untuk merekomendasikan artikel ilmiah berbahasa Inggris. Sistem memiliki empat komponen, yaitu analisa konten, profile learner, komponen penyaringan, dan pengambilan dokumen. Content-based learning diimplementasikan pada komponen analisa konten yang bertugas mengidentifikasi masukan berupa artikel ilmiah berbahasa Inggris. Profile learner dilakukan untuk menghitung kemiripan antar pengguna setelah sistem mendapatkan feedback berupa rating dari pengguna. Collaborative filtering diimplementasikan pada komponen penyaringan yang bertugas untuk merekomendasikan artikel kepada pengguna setelah sistem mendapatkan hasil kemiripan antar pengguna. Pengambilan dokumen dilakukan pada proses pengambilan artikel yang dicari oleh pengguna. Uji coba dilakukan pada 100 artikel ilmiah, 6 kelas kategori, serta melibatkan 35 pengguna. Hasil penelitian ini membuktikan bahwa implementasi metode content-based learning dan collaborative filtering pada sistem rekomendasi artikel ilmiah berbasis web mampu memberikan tingkat relevansi dan efektifitas sebesar 0.801 berdasarkan Mean Average Precision dan 0,851 berdasarkan Mean Absolute Error.

**Keywords**— sistem rekomendasi, content-based learning, collaborative filtering, hybrid recommendation

## I. PENDAHULUAN

Artikel ilmiah merupakan salah satu bentuk karya ilmiah, dan khusus untuk diterbitkan pada jurnal ilmiah. Terdapat berbagai macam karya ilmiah beserta bahasannya yang beragam. Karena keberagaman inilah seorang penulis biasanya mengalami kesulitan dalam mencari artikel yang sesuai dengan kebutuhannya. Oleh karena itu, dibutuhkan sistem rekomendasi yang dapat membantu penulis dalam mencari dan merekomendasikan artikel yang sesuai dengan bahasan yang dicari [1]

Sistem rekomendasi (*recommender system*) adalah sebuah sistem yang melakukan pencarian dan mendapatkan daftar rekomendasi berupa informasi, item, atau layanan yang bersifat personal [2] Pada pertengahan 1990 banyak riset tentang sistem rekomendasi untuk mengatasi masalah membanjirnya informasi yang ada pada internet [3]. Proses pencarian item yang akan direkomendasikan dapat dilakukan berdasarkan kemiripan, baik berupa kemiripan suatu item dengan item lainnya, berdasarkan konten, atau kemiripan

selera pengguna satu dan yang lain berdasarkan peringkat yang diberikan pada item tersebut.

Metode *Back Propagation* (BP) pada sistem rekomendasi, BP terbukti memiliki Mean Absolute Error (MAE) lebih rendah dibandingkan dengan Collaborative Filtering, tetapi membutuhkan waktu yang lebih lama untuk mengoptimalkan kinerja BP [4]. Lalu, beberapa peneliti seperti [5] dan [6] mengusulkan pendekatan Singular Value Decomposition (SVD) guna membantu meningkatkan akurasi sistem rekomendasi. Penggunaan metode tersebut menguntungkan di bidang akurasi, namun terlalu banyak menggunakan sumber daya sistem dan kehilangan informasi pengguna selama pengurangan dimensi. Metode matrix factorization yang diterapkan dapat mengintegrasikan konten dan informasi pengguna lainnya ke dalam pemodelan *Collaborative Filtering*, tetapi memiliki masalah yang sama dengan SVD.

*Collaborative Filtering* (CF) adalah metode yang memprediksi item untuk pengguna tertentu berdasarkan item yang sebelumnya dinilai oleh pengguna lain [3]. Namun terdapat kekurangan dalam sistem tersebut, diantaranya adalah item yang baru ditambahkan tidak akan pernah direkomendasikan karena tidak memiliki rekaman peringkat. Berbeda dengan metode Content-Based Learning yang bebas dari masalah sparse matrix karena melakukan proses pada tiap item yang memiliki atribut maupun tidak. Akan tetapi, sistem tersebut hanya merekomendasikan item yang serupa berdasarkan tingkat kemiripan.

Pada penelitian ini, penggabungan metode Collaborative Filtering dan Content-Based Learning akan diterapkan dalam sistem rekomendasi artikel ilmiah. Kombinasi user-activity yang terdapat pada Collaborative Filtering diharapkan dapat menutupi kelemahan Content-Based Learning. Penelitian ini bertujuan untuk mengetahui tingkat akurasi dari penggabungan metode Collaborative Filtering dan Content-Based Learning dalam sistem rekomendasi artikel ilmiah

## II. TINJAUAN PUSTAKA

Su dan Ye [6] melakukan penelitian mengenai item-based collaborative filtering menggunakan Rough Set Prediction untuk memprediksi rating yang kosong dalam matriks user-item. Penelitian tersebut dibandingkan dengan sistem rekomendasi dengan item-based collaborative filtering tanpa algoritma Rough Set Prediction. Dengan menggunakan Receiver Operating Characteristic (ROC), metode yang diusulkan terbukti memiliki tingkat akurasi lebih tinggi

dibandingkan dengan item-based collaborative tanpa Rough Set Prediction. Penelitian sistem rekomendasi oleh Tan dan Ye [7] mengenai user-based collaborative filtering menggunakan item classification. Dalam penelitian tersebut item classification digunakan untuk mengklasifikasi item berdasarkan atribut item. Berdasarkan Mean Absolute Error (MAE), metode tersebut terbukti lebih baik dibandingkan dengan metode collaborative filtering tanpa item classification.

Gong dan Ye [4] melakukan penelitian sistem rekomendasi item menggunakan metode Collaborative Filtering dan algoritma Back Propagation Neural Network. Penelitian tersebut membandingkan metode Collaborative Filtering tradisional dan metode Collaborative Filtering yang menggunakan algoritma Back Propagation (BP). CF dengan menggunakan algoritma BP terbukti memiliki MAE lebih rendah dibandingkan dengan collaborative filtering tanpa algoritma BP. Akan tetapi membutuhkan waktu yang lebih lama untuk mengoptimalkan kinerja BP. Uddin, Shrestha dan Jo [8] melakukan penelitian untuk meningkatkan metode Content-Based Learning dengan menggunakan Diverse Collaborative Prediction (DCP) dalam sistem rekomendasi film. Dalam penelitian tersebut DCP digunakan untuk menemukan item yang tidak mirip diantara item mirip yang direkomendasikan oleh collaborative filtering. Berdasarkan MAE, metode tersebut terbukti lebih baik dibandingkan dengan content-based learning tanpa DPC dan Naïve Hybrid Approach, serta dapat menekan nilai MAE hingga 3,77% dan 0,48%.

Yun dan Youn[8] melakukan penelitian mengenai sistem rekomendasi menggunakan metode collaborative filtering dan user information. User information digunakan untuk mengisi matriks user-item yang kosong. Penelitian ini terdiri dari beberapa tahap yaitu, pengisian non-rated data berdasarkan user information, perhitungan user similarity, dan prediksi. Hasil penelitian ini berupa nilai MAE sebesar 0,87, sedangkan metode collaborative filtering tanpa user information sebesar 1,04.

#### A. Karya Tulis Ilmiah

Karya tulis ilmiah atau Karya ilmiah adalah tulisan atau laporan tertulis yang memaparkan hasil penelitian atau pengkajian suatu masalah oleh seseorang atau sebuah tim dengan memenuhi kaidah dan etika keilmuan yang dikukuhkan dan ditaati oleh masyarakat keilmuan. Data, simpulan dan informasi lain yang terkandung dalam karya ilmiah tersebut dijadikan acuan (referensi) bagi ilmuwan lain dalam melaksanakan penelitian atau pengkajian selanjutnya. Karya ilmiah berfungsi sebagai sarana untuk mengembangkan ilmu pengetahuan dan teknologi berupa penjelasan, prediksi dan pengawasan. Ada beberapa jenis karya tulis ilmiah, yaitu artikel ilmiah, makalah, kertas kerja, paper, skripsi, tesis, disertasi, dan artikel ilmiah populer. Berikut ini merupakan karakteristik karya ilmiah:

- Mengacu pada teori sebagai landasan berpikir (kerangka pemikiran) dalam pembahasan masalah;
- Lugas, tidak emosional, bermakna tunggal, tidak menimbulkan interpretasi lain;
- Logis, disusun berdasarkan urutan yang konsisten;
- Efektif, ringkas dan padat;

- Efisien, hanya mempergunakan kata atau kalimat yang penting dan mudah dipahami;
- Objektif berdasarkan fakta, setiap informasi dalam kerangka ilmiah selalu ada adanya, sebenarnya, dan konkret;
- Sistematis, baik penulisan dan pembahasan sesuai dengan prosedur dan sistem yang berlaku.

#### B. Sistem Rekomendasi

Sistem rekomendasi adalah suatu program yang melakukan prediksi terhadap suatu item, seperti film, musik, buku, berita dan lain sebagainya yang menarik bagi pengguna, berdasarkan keadaan dan keinginan pengguna. Sistem ini berjalan dengan menggunakan data dari pengguna secara langsung maupun tidak

Dalam sistem rekomendasi terdapat beberapa metode dasar yang digunakan, diantaranya adalah:

- Content-based recommendation: merupakan sistem rekomendasi yang menerapkan metode Content-based Learning dimana pengguna akan direkomendasikan untuk memilih item yang mirip dengan item yang telah dipilih pengguna tersebut dimasa lalu.
- Collaborative recommendation: merupakan sistem rekomendasi yang menggunakan metode collaborative filtering dimana pengguna akan direkomendasikan untuk memilih item yang dipilih atau disukai oleh para pengguna lain (dengan selera yang sama) dimasa lalu.
- Hybrid approach: sistem rekomendasi yang menggunakan metode hybrid-based dimana metode tersebut adalah gabungan antara metode content-based dan collaborative.

#### C. Content-based Learning

Pembelajaran berbasis konten untuk sistem rekomendasi berawal dari studi temu kembali informasi (Information Retrieval), dan information filtering. Menurut Borgers dan Bosch[] content-based recommendation adalah hasil dari penelitian penyaringan informasi dalam sistem berbasis konten. Content-based memerlukan teknik yang tepat untuk mewakili item dan beberapa strategi untuk membandingkan user profile dan representasi item. Content-based terbukti dapat secara efektif menyaring informasi yang relevan dengan topik, namun terdapat beberapa kekurangan dari metode tersebut dalam sistem yaitu:

- Penyaringan berbasis konten sulit untuk menghasilkan rekomendasi tak terduga (serendipitous recommendation), karena semua informasi yang direkomendasikan dan dipilih berdasarkan konten.
- Hanya mengandalkan kinerja sistem dan mengabaikan masukan dari pengguna lain.

Pencarian kemiripan antara pengguna dan item adalah langkah penting dalam algoritma sistem rekomendasi. Menurut Fan, Pan dan Jiang [10] Cosine similarity merupakan algoritma yang memisalkan garis pada matriks user-item sebagai vektor. Kesamaan antara dua item yang didefinisikan sebagai sudut kosinus ditunjukkan pada Persamaan 1

$$\text{sim}(x, y) = \cos(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

Dimana  $\text{sim}(x, y)$  adalah kesamaan antara  $x$  dan  $y$ . dan  $x, y$  merupakan dua vektor dan  $n$  mewakili vektor dimensi.

Sebelum menghitung kesamaan antar artikel, artikel yang dimasukkan oleh pengguna terlebih dahulu harus melewati proses pra pengolahan dan pembobotan. Pada penelitian ini pembobotan akan dilakukan menggunakan rumus tf-idf atau term frequency inverse document frequency, dirumuskan dalam persamaan 2.2 [10]

Sebelum menghitung kesamaan antar artikel, artikel yang dimasukkan oleh pengguna terlebih dahulu harus melewati proses pra pengolahan dan pembobotan. Pada penelitian ini pembobotan akan dilakukan menggunakan rumus tf-idf atau term frequency inverse document frequency, dirumuskan dalam persamaan 2 [10].

$$\text{TFIDF} = \text{weight} = \text{tf} \cdot \log\left(\frac{n}{\text{df}}\right) \quad (2)$$

Pada persamaan 2,  $\text{tf}$  merupakan term frequency atau jumlah term yang terdapat pada satu artikel,  $n$  merupakan jumlah artikel secara keseluruhan, dan  $\text{df}$  atau document frequency merupakan jumlah suatu term yang ada pada semua artikel.

#### D. Collaborative Filtering

Tidak seperti metode Content-based, sistem rekomendasi dengan menggunakan metode collaborative filtering mencoba untuk memprediksi item yang mungkin berguna untuk pengguna tertentu berdasarkan item yang sebelumnya telah dinilai oleh pengguna lain.

- Li dan Kim (2002) menyatakan terdapat dua masalah utama yang ada pada metode collaborative filtering, yaitu:
- Cold start problem, dimana item baru tidak akan direkomendasikan karena belum memiliki rekaman nilai.
- Meskipun collaborative filtering dapat meningkatkan kualitas rekomendasi berdasarkan rating pengguna, namun sistem tidak merekomendasikan item yang serupa atau relevan.

*User-based collaborative filtering* adalah metode yang berasumsi bahwa pengguna dapat dikelompokkan berdasarkan selera mereka. Suatu item yang disukai oleh seseorang dalam suatu kelompok kemungkinan akan disukai oleh orang lain pada kelompok yang sama. Berdasarkan catatan rating dari semua pengguna, sistem rekomendasi dapat menghasilkan matriks user-item seperti pada Tabel 1

Dimana himpunan  $\{u_1, u_2, \dots, u_m\}$  menunjukkan himpunan pengguna, dan himpunan  $\{i_1, i_2, \dots, i_n\}$  menunjukkan himpunan item.  $R_{u,i}$  merepresentasi rating dari pengguna  $u$  terhadap item  $i$ .

Salah satu teknik utama dalam sistem rekomendasi adalah menemukan kemiripan antar pengguna atau barang. Dan persamaan II-3 adalah rumus perhitungan kesamaan dengan Pearson Correlation Coefficient menurut Fan, Pan, dan Jiang (2014).  $\text{sim}(x, y) =$

$$\frac{\sum_{c \in I_{xy}} (R_{x,c} - \bar{R}_x) \times (R_{y,c} - \bar{R}_y)}{\sqrt{\sum_{c \in I_x} (R_{x,c} - \bar{R}_x)^2} \times \sqrt{\sum_{c \in I_y} (R_{y,c} - \bar{R}_y)^2}}$$

Dimana  $R_{x,c}$  adalah nilai dari pengguna  $x$  terhadap item  $c$ ,  $I_{xy}$  menunjukkan himpunan item yang dinilai oleh pengguna  $x$  dan  $y$ ,  $(\bar{R}_x)$  dan  $(\bar{R}_y)$  adalah nilai rata-rata dari masing-masing pengguna  $x$  dan  $y$ .

Tabel 1 Matriks user-item [10]

	1	...	i	...	n
1					
...					
u			$R_{u,i}$		
...					
m					

#### E. Hybrid-based recommendation

Diketahui proses sistem rekomendasi ditampilkan dalam tiga langkah utama, yaitu:

- *Content Analyzer*. Tugas utama komponen ini adalah merepresentasikan isi item yang berasal dari sumber-sumber informasi dalam bentuk yang sesuai untuk langkah-langkah pengolahan berikutnya. Data item dianalisis dengan teknik ekstraksi ciri untuk mengubah representasi item dari sumber informasi ke target representasi. Representasi ini adalah masukan untuk profile learner dan filtering component.
- *Profile Learner*. Modul ini bertugas untuk mengumpulkan perwakilan data preferensi pengguna dan mencoba untuk menggeneralisasi data tersebut dalam rangka membangun user profile. Generalisasi diwujudkan melalui teknik pembelajaran mesin yang mampu menyimpulkan model ketertarikan pengguna, mulai dari item yang disukai atau tidak disukai.
- *Filtering Component*. Modul ini memanfaatkan user profile untuk menyarankan item yang relevan dengan mencocokkan representasi profile terhadap barang yang akan direkomendasikan. Hasilnya adalah keputusan berupa relevansi yang dihitung dengan menggunakan matriks kemiripan (similarity matrix), langkah terakhir adalah menghasilkan daftar item yang berpotensi menarik. matriks kemiripan (*similarity matrix*), langkah terakhir adalah menghasilkan daftar *item* yang berpotensi menarik.

#### F. K-Nearest Neighbor

Algoritma k-nearest neighbor (k-NN atau KNN) adalah salah satu metode tertua dan paling sederhana untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Amatriain, Jaimes, & Pujol, 2011). Data learning dideskripsikan sebagai atribut numerik n-dimensi. Tiap data learning merepresentasikan sebuah titik, yang ditandai dengan  $c$ , dalam ruang n-dimensi. Jika sebuah data kueri yang labelnya tidak diketahui diinputkan, maka K-Nearest Neighbor akan mencari k buah data learning yang jaraknya

paling dekat dengan data kueri dalam ruang n-dimensi. Jarak antara data kueri dan data learning dapat dihitung dengan cara mengukur jarak antara titik yang merepresentasikan data kueri dengan semua titik yang merepresentasikan data learning menggunakan rumus Euclidean Distance pada persamaan  $d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$  (Sun dan Huang, 2010). Dimana vektor  $x = (x_1, x_2, x_3, \dots, x_n)$  dan  $y = (y_1, y_2, y_3, \dots, y_n)$ , n merupakan jumlah atribut.  $d(x,y)$  adalah dua data yang mirip.

Pada fase pembelajaran, algoritma ini hanya melakukan penyimpanan vektor-vektor fitur dan klasifikasi dari data pembelajaran. Pada fase klasifikasi, fitur-fitur yang mirip digunakan sebagai data tes (yang klasifikasinya tidak diketahui). Jarak dari vektor yang baru ini terhadap seluruh vektor data learning akan dihitung dan sejumlah k titik yang paling dekat diambil sebagai nearest neighbor. Titik yang baru akan dimasukkan kedalam kelas yang memiliki titik-titik terbanyak dari klasifikasi tersebut.

### III. HASIL DAN PEMBAHASAN

Masukan yang digunakan pada penelitian ini adalah artikel ilmiah berbahasa Inggris. Data yang digunakan meliputi judul, abstrak, dan kata kunci dari tiap artikel. Berdasarkan Gambar III-2, tahap pra-pengolahan meliputi casefolding, stopword removal, stemming, dan tokenizing. Tahap ini dilakukan terhadap judul, abstrak dan/atau kata kunci dari tiap artikel. Tabel III-1 menunjukkan contoh hasil pra-pengolahan terhadap 7 judul artikel.

Tabel 2 Contoh hasil pra-pengolahan

#### a. Dokumen

1. *Mutual Authentication mechanism for RFID System*
2. *The Web Security Password Authentication Based the Single Block Hash Function*
3. *Security of A Sessional Blind Signature Based On Quantum Cryptograph*
4. *Research On Intelligence Decision Support System for Power System*
5. *Overview of The New Types of Intelligent Decision Support System*
6. *Ontology Based Web Spatial Decision Support System*
7. *Using A Group of Decision Support System To Make Investment Prioritisation Decision*

#### b. Hasil proses Casefolding :

1. *mutual authentication mechanism for rfid system*
2. *the web security password authentication based the single block hash function*
3. *security of a sessional blind signature based on quantum cryptograph*
4. *research on intelligence decision support system for power system*
5. *overview of the new types of intelligent decision support system*
6. *ontology based web spatial decision support system*
7. *using a group of decision support system to make investment prioritisation decision*

#### c. Hasil proses Stop word Removal :

1. *mutual authentication mechanism rfid system*
2. *web security password authentication based single block hash function*
3. *security sessional blind signature based quantum cryptograph*
4. *research intelligence decision support system power system*
5. *overview types intelligent decision support system*
6. *ontology based web spatial decision support system*
7. *decision support system investment prioritisation decision*

#### d. Hasil proses Stemming :

1. *mutual authent mechanism rfid system*
2. *web secur password authenticate base single block hash function*
3. *secure session blind signature base quantum cryptograph*
4. *research intelligenc decision support system power system*
5. *overview type intellig decision support system*
6. *ontology base web spatial decision support system*
7. *decision support system invest priority decision*

Proses pertama yang dilakukan setelah mengetahui hasil Pra Pengolahan adalah pembobotan tf-idf. Term frequency inverse document frequency atau tf-idf digunakan untuk menghitung bobot dari tiap term yang ada dalam suatu artikel. Tabel 2 memperlihatkan daftar term berdasarkan contoh hasil pra-pengolahan (ditunjukkan pada tabel 1). Tabel 3 menunjukkan hasil perhitungan tf-idf, berdasarkan tabel 4 menggunakan persamaan tf-idf.

Komponen document retriever bertugas untuk mengumpulkan dokumen-dokumen yang mengandung kata kunci yang dicari oleh pengguna. Pada penelitian ini, pengumpulan dokumen dilakukan hanya dengan mengecek keberadaan kata kunci dalam dokumen. Apabila dokumen tersebut mengandung kata kunci yang dicari, maka dokumen tersebut dianggap sebagai dokumen yang dibutuhkan. Tabel 4 menunjukkan contoh hasil proses pencarian artikel.

Tabel 2 daftar term berdasarkan contoh hasil pra-pengolahan

term	tf							df	n/df	idf
	d1	d2	d3	d4	d5	d6	d7			
authent	1	1						2	3.5	0.5440
decision				1	1	1	2	5	1.4	0.1461
invest							1	1	7	0.8450
mechanism	1							1	7	0.8450
mutual	1							1	7	0.8450
priority							1	1	7	0.8450
rfid	1							1	7	0.8450
secur		1	1					2	3.5	0.5440
support				1	1	1	1	4	1.75	0.2430
system	1			1	1	1	1	5	1.4	0.1461
web		1					1	2	3.5	0.5440
...										

Tabel 3 hasil perhitungan tf-idf

term	weight						
	d1	d2	d3	d4	d5	d6	d7
authent	0.54 40	0.54 40	0	0	0	0	0
decision	0	0	0	0.14 61	0.14 61	0.14 61	0.29 22
invest	0	0	0	0	0	0	0.84 50
mechani sm	0.84 50	0	0	0	0	0	0
mutual	0.84 50	0	0	0	0	0	0
priority	0	0	0	0	0	0	0.84 50
rfid	0.84 50	0	0	0	0	0	0
secur	0	0.54 40	0.54 40	0	0	0	0
support	0	0	0	0.24 30	0.24 30	0.24 30	0.24 30
system	0.14 61	0	0	0.14 61	0.14 61	0.14 61	0.14 61
web	0	0.54 40	0	0	0	0.54 40	0
...							

Tabel 4. Contoh hasil pencarian artikel

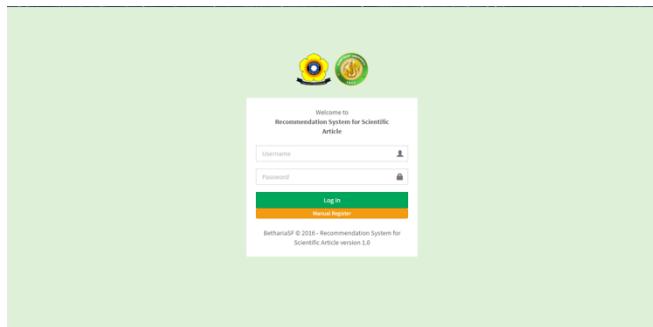
Search : Cryptograph
Result :
1. Title : Security of a sessional blind signature based on quantum cryptograph
Abstract : We analyze the security of a sessional blind signature protocol based on quantum cryptograph and show that there are two security leaks in this protocol. One is that the legal user Alice can change the signed message after she gets a valid blind signature from the signatory Bob, and the other is that an external opponent Eve also can forge a valid blind message by a special attack, which are not permitted for blind signature. ...
Keyword: Blind signature, Sessional blind signature, Non-forgeability, Quantum cryptography
2. Title : Mutual-Authentication Mechanism for RFID Systems
Abstract : ... Our advanced mutualauthentication protocol uses an AES algorithm as its cryptograph primitive. Since our AES algorithm has a relatively low cost, is fast, and only requires simple hardware, our proposed approach is feasible for use in RFID systems. In addition, the relatively low computational cost of our proposed algorithm compared to those currently used to implement similar levels of system security makes our proposed system especially suitable for RFID systems that have a large number of tags.
Keyword: RFID, AES, cryptograph.

Kelas-kelas yang telah dirancang pada tahap sebelumnya diimplementasikan dalam bahasa pemrograman PHP. Daftar kelas dapat dilihat pada tabel 5.

Tabel 5 Daftar kelas yang di rancang

No.	Nama Kelas	Nama Berkas	Keterangan
1.	Index	index.html	Kelas <i>boundary</i> yang berperan sebagai modul antarmuka utama.
2.	Upload_training	upload_training.php	Kelas ini berfungsi untuk menerima masukan file dari admin.
3.	Upload_testing	upload_testing.php	Kelas ini berfungsi untuk menerima masukan file dari admin.
4.	tfidf	upload_testing.php	Kelas ini berfungsi untuk menghitung bobot dari tiap term yang ada dalam artikel.
5.	cosine	upload_testing.php	Kelas ini berfungsi untuk menghitung kemiripan antar artikel
6.	knn	upload_testing.php	Kelas ini berfungsi untuk mengklasifikasikan artikel testing yang telah diupload
7.	preprocess	preprocess.php	Kelas ini bertujuan untuk melakukan pra-pengolahan terhadap artikel.
8.	Detail_artikel	detail_artikel.php	Kelas entitas yang berfungsi untuk menampung data artikel pada database
9.	Rating_artikel	rating_artikel.php	Kelas entitas yang menampung nilai rating serta menghitung kemiripan antar pengguna.
10.	Tampil_artikel	tampil_artikel.php	Kelas ini digunakan untuk menampilkan artikel keseluruhan
11.	Pilih_kategori	pilih_kategori.php	Kelas yang digunakan untuk menampung kata kunci yang dicari oleh pengguna
12.	Tampil_hasil	tampil_hasil.php	Kelas yang bertugas untuk menampilkan artikel yang dicari oleh pengguna
13.	Artikel_model	artikel_model.php	Merupakan kelas yang digunakan untuk menghubungkan sistem dan database
14.	Tampil_rekomendasi	tampil_rekomendasi.php	Kelas yang digunakan untuk menampilkan daftar artikel yang direkomendasikan

Gambar 1 merupakan implementasi antarmuka berdasarkan perancangan.



Gambar 1. Tampilan *login*

Jenis pengujian yang dilakukan pada penelitian ini adalah pengujian black box. Tabel 6, 7, 8 merupakan rencana pengujian pada perangkat lunak.

Tabel 6 Rencana Pengujian *input* dokumen

No.	Identifikasi	Pengujian	Tingkat Pengujian
1.	U-4-101	Memasukkan rating	Pengujian unit

Tabel 7 Rencana Pengujian *use case* input dokumen

No.	Identifikasi	Pengujian	Tingkat Pengujian
1.	U-1-101	Memasukkan dokumen	Pengujian unit

Tabel 8 Rencana Pengujian *use case* klasifikasi dokumen

No.	Identifikasi	Pengujian	Tingkat Pengujian
1.	U-2-101	Mengklasifikasikan dokumen	Pengujian unit

Berdasarkan hasil pengujian *black box* pada tahap sebelumnya, dapat disimpulkan bahwa performa sistem rekomendasi artikel ilmiah telah sesuai dan bekerja dengan baik. Hal ini dapat dilihat dari kesimpulan skenario kasus uji yang menghasilkan hasil “diterima” atau sesuai dengan hasil yang diharapkan untuk seluruh kasus uji. Berikutnya, penjelasan akan dilanjutkan dengan menjabarkan hasil, analisis dan eksperimen metode terhadap data set. Gambar 2 menunjukkan hasil pengujian.



Gambar 2. Pengujian Sistem

Pengujian performa sistem rekomendasi dilakukan dengan mengukur nilai kesalahan rata-rata absolut (*Mean*

*Absolute Error*) terhadap perbandingan antara rating yang diprediksi dan rating nyata yang diberikan oleh pengguna. Pengujian dilakukan pada 100 dokumen dan melibatkan 35 pengguna aktif yang akan memberikan rating terhadap artikel.

Analisis akurasi menggunakan 100 sampel data yang terdiri dari 60 data *training* dan 40 data *testing*, serta melibatkan 35 pengguna. Pengukuran akurasi dilakukan dengan menghitung *Mean Absolute Error* (persamaan 2.6) antara rating yang diprediksi oleh rumus (persamaan 2.6) dan rating nyata yang diberikan oleh pengguna.

#### IV. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan beberapa hal sebagai berikut:

- Hasil pengujian menggunakan Recall pada metode Content-based Learning adalah 91,6%.
- Hasil pengujian menggunakan Mean Absolute Error pada penelitian Hybrid Recommendation (Content-based Learning dan Collaborative Filtering) adalah 0.85.

Untuk penelitian selanjutnya, saran yang ingin penulis sampaikan untuk pengembangan lebih lanjut dari penelitian ini adalah :

- Memperbanyak keragaman artikel, agar sistem dapat merekomendasikan artikel yang lebih beragam.
- Mengembangkan penelitian untuk membandingkan waktu proses dari algoritma *collaborative filtering*, algoritma *content-based learning*, dan/atau algoritma *hybrid* lainnya.
- Sistem dikembangkan untuk artikel ilmiah berbahasa Indonesia.

#### DAFTAR PUSTAKA

- [1] Winarno. (2013). *Rekayasa Text Mining Guna Membantu Referensi Pencarian Daftar Pustaka Menggunakan Metode Bayesian*. Semarang: Universitas Dian Nuswantoro.
- [2] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based Collaborative Filtering Recommendation Algorithms. *Proc. 10th Int'l WWW Conference*.
- [3] Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender System: A Survey of the State-of-the-Art and Possible Extension. *IEEE Transaction on Knowledge and Data Engineering*.
- [4] Gong, S. J., & Ye, H. W. (2009). An Item Based Collaborative Filtering Using BP Neural Networks Prediction. *IEEE Computer Society*, 146-148.
- [5] Patarek, A. (2007). Improving regularized singular value decomposition for collaborative filtering. *Proceedings of KDD cup and workshop*, 5-8.
- [6] Su, P., & Ye, H. W. (2009). An Item based Collaborative Filtering Recommendation Algorithm Using Rough Set Prediction. *JCAI*, 308-311.
- [7] Tan, H. S., & Ye, H. W. (2009). A Collaborative Filtering Recommendation Algorithm Based on Item Classification. *PACCS*, 694-697.

- [8] Uddin, M. N., Shrestha, J., & Jo, G. S. (2009). Enhance Content-based Filtering using Diverse Collaborative Prediction for Movie Recommendation. ACIIDS, 132-137.
- [9] Yun, S. Y., & Youn, S. D. (2010). Recommender System Based on User Information. IEEE.
- [10] Fan, J., Pan, W., & Jiang, L. (2014). An Improved Collaborative Filtering Algorithm Combining Content-based Algorithm and User Activity. IEEE BigComp 2014, 88-91.