

Prosiding
ANNUAL RESEARCH SEMINAR 2016
6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

Sistem Rekomendasi Bacaan Tugas Akhir Jurusan Teknik Informatika Universitas Sriwijaya menggunakan Metode *Collaborative Filtering* dan *Naive Bayes*

Riri Intan Aprilia¹

Teknik Informatika Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
Email : ririintanap@gmail.com

Muhammad Fachrurrozi²

Teknik Informatika Fakultas Ilmu Komputer
Universitas Sriwijaya
Palembang, Indonesia
Email : mfachrz@unsri.ac.id

Abstract— Sistem Rekomendasi merupakan aplikasi pemberi rekomendasi berupa informasi yang dibutuhkan oleh pengguna berdasarkan *feedback* dari pengguna lain. Fakultas Ilmu Komputer Universitas Sriwijaya khususnya jurusan Teknik Informatika memerlukan sistem rekomendasi bacaan tugas akhir mengingat sulitnya mengakses data kumpulan tugas akhir yang ada sesuai dengan kategori bidang yang diminati oleh mahasiswa. Untuk mengatasi masalah ini menggunakan metode *collaborative filtering* dan *naive bayes*. Metode *collaborative filtering* digunakan untuk memberi rekomendasi suatu *item* berupa tugas akhir berdasarkan *rate* oleh pengguna lainnya, sedangkan *naive bayes* digunakan untuk mengklasifikasikan setiap data tugas akhir ke dalam bidang yang sesuai. Berdasarkan hasil percobaan, penelitian ini memiliki akurasi hasil didapatkan 81% dari 100 data uji dan 48 data latih untuk metode *naive bayes* dan nilai MAE sebesar 2,01 untuk metode *collaborative filtering*.

Keywords: *Collaborative Filtering*, *Naive Bayes*, *Sistem Rekomendasi*

I. PENDAHULUAN

Sistem rekomendasi difungsikan sebagai aplikasi untuk mengusulkan informasi dan menyediakan fasilitas yang diinginkan pengguna dalam membuat suatu keputusan. Sistem ini diasumsikan seperti penggambaran kebutuhan dan keinginan pengguna melalui pendekatan metode rekomendasi dengan mencari dan merekomendasikan suatu *item* dengan menggunakan *rating* berdasarkan kemiripan dari karakteristik informasi pengguna. Salah satu sistem rekomendasi yang diperlukan khususnya untuk mahasiswa di Fakultas Ilmu

Komputer Universitas Sriwijaya adalah sistem pemberi rekomendasi bacaan tugas akhir mengingat sulitnya mengakses data mengenai kumpulan tugas akhir mahasiswa khususnya di jurusan Teknik Informatika yang mengharuskan mahasiswa datang ke perpustakaan untuk membaca tugas akhir yang ada.

Collaborative Filtering adalah teknik yang secara luas digunakan untuk menghasilkan rekomendasi pribadi yang merekomendasikan suatu *item* berdasarkan *rate* yang diberikan oleh pengguna lainnya [1].

Naive Bayes adalah metode probabilistik untuk klasifikasi. Metode ini digunakan untuk menentukan probabilitas dari contoh kelas yang diberikan nilai dari variable [3].

Tujuan dari sistem pemberi rekomendasi tugas akhir untuk mahasiswa Teknik Informatika ini adalah menghasilkan daftar rekomendasi tugas akhir dan dapat bermanfaat dalam mempermudah mahasiswa dalam mencari bacaan tentang tugas akhir mahasiswa Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya berdasarkan kategori dari tugas akhir tersebut.

Dengan kedua metode tersebut diharapkan masalah yang ada dapat diatasi dengan melalui beberapa tahap yang akan dilakukan agar mendapatkan keluaran berupa rekomendasi tugas akhir yang sesuai dan akurat untuk setiap *user*.

II. METODE PENELITIAN

2.1. Sistem Rekomendasi

Menurut [4] Sistem Rekomendasi adalah program perantara atau perwakilan yang secara cerdas menyusun daftar dari informasi yang diperlukan dan mencocokkan berdasarkan keinginan dari pengguna. Sistem rekomendasi bertujuan untuk

Prosiding
ANNUAL RESEARCH SEMINAR 2016
 6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

http://ars.ilkom.unsri.ac.id

menyarankan *item* kepada pengguna. Sistem rekomendasi langsung menyarankan pengguna ke *item-item* yang dapat memenuhi kebutuhan dan keinginan mereka dengan cara mempersempit informasi di *database* besar [5].

Algoritma ini dibagi menjadi dua kelas yaitu *user-based* dan *item-based* :

a) *User-Based Collaborative Filtering*

User-based nearest neighbour algorithm menggunakan teknik statistika untuk menemukan sekumpulan pengguna, dikenal sebagai tetangga (*neighbour*), yang memiliki sejarah setuju dengan pengguna yang menjadi sasaran. Setelah sekumpulan tetangga terbentuk, sistem menggunakan algoritma yang berbeda untuk menggabungkan kesukaan *neighbours* untuk menghasilkan prediksi atau rekomendasi *N-teratas* untuk *active user* [8].

b) *Item-to-Item Collaborative Filtering*

Item-based collaborative filtering merupakan metode rekomendasi yang didasari pada kesamaan antara pemberian *rating* terhadap suatu produk yang diberi *rating*. Metode ini dihitung dengan menggunakan langkah - langkah dari *similarities* dan kemudian langkah-langkah *similarities* ini digunakan untuk memprediksi peringkat. Hal ini dapat dilakukan melalui menemukan *k nearest neighbours* dan kemudian merekomendasikan *item*. [5]

2.2. Perhitungan *Similarities*

Pada penelitian ini penulis menggunakan persamaan *adjusted cosine similarity* untuk menghitung kemiripan karena menurut dari penelitian yang dilakukan algoritma yang dapat menghasilkan nilai *error* (MAE) terkecil adalah algoritma *adjusted cosine similarity* [8].

$$Sim(i, j) = \frac{\sum u \in U (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum u \in U (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum u \in U (R_{u,j} - \bar{R}_u)^2}} \quad (2.1)$$

- $Sim(i, j)$: Nilai kemiripan antara *item* i dan *item* j
- $u \in U$: Himpunan *user* u yang *merate* *item* i dan j
- $R_{u,i}$: *Rating user* u pada *item* i
- $R_{u,j}$: *Rating user* u pada *item* j
- \bar{R}_u : Nilai rata-rata *rating user* u

2.3. Perhitungan Prediksi

Pada perhitungan prediksi digunakan metode *weighted sum* pada sebuah *item* i untuk pengguna u dengan menghitung jumlah dari *rating* yang diberikan oleh pengguna pada *item* yang mirip dengan i. Setiap *rating* dipertimbangkan oleh yang berhubungan dengan *similarity* dari $s_{i,j}$ antara *item* i dan j. $P_{u,i}$ menunjukkan perhitungan dari sebuah prediksi [8] :

$$p_{u,i} = \frac{\sum_{i \in l} (S(i,j) * r(u,j))}{\sum_{i \in l} |S(i,j)|} \quad (2.2)$$

- $p_{u,i}$: Prediksi untuk *user* u pada *item* i
- $i \in l$: Himpunan *item* yang mirip dengan *item* i
- $S(i, j)$: Hasil *rating similarity user* i pada *item* j
- $r_{u,j}$: Nilai kemiripan antar *item* yang sama

2.4. Prapengolahan

Tahap ini merupakan tahap pertama yang dilakukan setelah mendapatkan hasil rekomendasi dari *Collaborative Filtering* untuk kemudian diolah. Tahap praproses terbagi menjadi 2 bagian, praproses untuk pengolahan data pengklasifikasian dokumen dan praproses untuk pengolahan data pencarian dokumen.

Penelitian ini melakukan klasifikasi teks yang mengandung emosi menggunakan algoritma Complement Naïve Bayes Classifier. Sebelum melakukan proses pelatihan dan klasifikasi dan yang akan memakan waktu

Gambar II.1 Dokumen Masukan

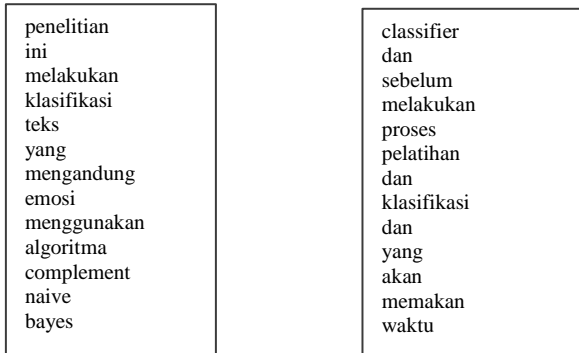
Praproses untuk pengolahan data pengklasifikasian dokumen.

Setelah dimuat, teks akan melalui tahap *case folding* dimana teks diubah menjadi bentuk *lower case* dan penghilangan karakter yang tidak berkaitan dengan abjad.

penelitian ini melakukan klasifikasi teks yang mengandung emosi menggunakan algoritma complement naive bayes classifier dan sebelum melakukan proses pelatihan dan klasifikasi dan yang akan memakan waktu

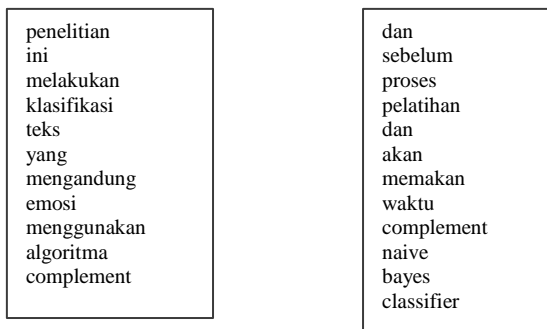
Gambar II.2 Casefolding

Kemudian teks akan melalui tahap tokenizing. Pada tahap ini teks diubah menjadi token-token kata. *Output* dari tahap ini adalah satuan kata.



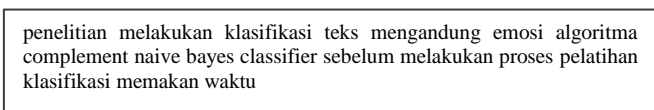
Gambar II.3 *Tokenizing*

Tahap terakhir dalam praproses ini adalah klasifikasi kata *unique*. Pada tahap ini, token-token yang memiliki frekuensi lebih dari 1 yang didapatkan dari proses *tokenizing* akan dihilangkan. Hasil keluaran dari tahap ini adalah sebagai berikut :



Gambar II.4 Klasifikasi Kata *Unique*

Lalu setelah melalui beberapa tahapan maka didapatkanlah teks hasil proses pra-pengolahan :



Gambar II.5 Teks Hasil Praproses dengan *Casefolding, Tokenizing*, dan Klasifikasi kata *unique*

2.5. Teorema Bayes

Naive Bayes adalah metode probabilistik untuk klasifikasi. Metode ini digunakan untuk menentukan probabilitas dari contoh kelas yang diberikan nilai dari variabel. *Naive Bayes* ini adalah algoritma paling sukses di bidang klasifikasi [7]. *Naive Bayes* memiliki kelebihan yaitu metode yang mudah diimplementasikan, dikarenakan metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang sedikit untuk menentukan estimasi parameter yang diperlukan dalam proses pengklarifikasian dan waktu *training* yang lebih cepat dibandingkan metode lainnya.

Menurut [2] terdapat beberapa proses yang dilakukan pada metode *Naive Bayes* yaitu :

1. Baca data latih
2. Hitung jumlah dan probabilitas, tetapi apabila data bersifat numerik maka :
 - a. Cari nilai mean dan standar deviasi masing-masing parameter yang merupakan data numerik,
 - b. Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
 - c. Mendapatkan nilai dalam tabel *mean*, standar deviasi dan probabilitas.

Algoritma ini menggunakan teorema *Bayes* oleh karena itu probabilitas dokumen d berada di kelas C_j yang dihitung berdasarkan [2] :

$$P(C_j|d) = \frac{P(C_j)P(d|C_j)}{P(d)} \tag{2.3}$$

$P(C_j|d)$: Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis C_j terjadi jika diberikan bukti (*evidence*) d terjadi.

$P(C_j)$: Probabilitas awal(priori) hipotesis C_j terjadi tanpa memandang bukti apapun.

$P(d|C_j)$: Probabilitas sebuah bukti d terjadi akan mempengaruhi hipotesis C_j .

$P(d)$: Probabilitas awal(priori) bukti d terjadi tanpa memandang hipotesis/bukti yang lain.

Prosiding
ANNUAL RESEARCH SEMINAR 2016
 6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

http://ars.ilkom.unsri.ac.id

III. HASIL PENGUJIAN

3.1. Perhitungan Akurasi Proses Klasifikasi

Masukan pada penelitian ini menggunakan 148 sampel data yang terdiri dari 48 data *training* dan 100 data *testing* serta melibatkan 30 pengguna. Hasil klasifikasi dideskripsikan pada tabel dibawah

Tabel III-1. Jumlah tugas akhir berdasarkan kategori

Kategori	A	C
Kriptografi	3	1
Sistem Pendukung Keputusan	1	0
Sistem Pakar	10	1
Pengolahan Citra	25	3
Bahasa Pemrosesan Alami	11	9
Data Mining	3	1
Jaringan Syaraf Tiruan	24	2
Game	3	2

Keterangan:

A = jumlah dokumen relevan yang ditemukan

C = jumlah dokumen tidak relevan yang ditemukan

Dengan menggunakan rumus :

$$\begin{aligned}
 \text{Akurasi} &= \frac{\text{Jumlah Kategorisasi relevan}}{\text{Jumlah Data Uji}} \times 100\% \\
 &= \frac{81}{100} \times 100\%
 \end{aligned}$$

hasil yang diperoleh pada sistem ini adalah 81 %.

3.2. Mean Absolute Error

Tingkat akurasi pada hasil sistem rekomendasi dilakukan dengan melihat nilai eror pada hasil. Dalam penelitian ini menggunakan persamaan MAE (*Mean Absolute Error*).

Persamaan ini termasuk jenis *statistical accuracy metrics* dimana MAE akan menghitung nilai rata-rata selisih antara nilai prediksi dengan nilai yang sebenarnya [8].

Persamaan MAE :

$$\text{MAE} = \frac{\sum_{i=1}^n |p_i - q_i|}{N} \quad (2.4)$$

Keterangan :

MAE : nilai rata-rata kesalahan hitungan

N : jumlah *item* yang dihitung

p_i : nilai prediksi *item* ke i

q_i : nilai *rating* sebenarnya *item* ke i

Hasil yang diperoleh ditunjukkan pada tabel III.2 yaitu 2,01.

Tabel III.2 Hasil Data Uji

Pengguna	Rata-rata rating	MAE
u1	0,525	0,8
u2	0,37	0,638
u3	0,115	0,75
u4	0,185	0,25
u5	0,245	1,97
u6	0,44	3
u7	0,45	1,13
u8	0,28	0
u9	0,205	0,109
u10	0,29	1,704
u11	0,46	0,554
u12	0,395	0,327
u13	0,325	1,277
u14	0,185	0,715
u15	0,44	2,666
u16	0	0
u17	0,33	0,6
u18	0,195	1,203
u19	0,35	2,5
u20	0,395	2,303
u21	0,42	1,853
u22	0,175	1
u23	0,28	1,3304
u24	0,52	1,761
u25	0,315	1,777
u26	0,275	0,269
u27	0,13	1,222
u28	0,295	1,333
u29	0,275	2,055
u30	0,3	1,144
Total		2,01

Berdasarkan tabel III.2, hasil perhitungan akurasi terhadap metode *Naive Bayes* adalah 81 %. Dari hasil yang didapat pada penelitian ini menunjukkan bahwa penggunaan metode *Naive Bayes* memiliki performa yang cukup baik, dimana sistem dapat mengklasifikasikan 69 dokumen dari 100 dokumen dengan tepat.

Prosiding
ANNUAL RESEARCH SEMINAR 2016
6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

Pada tabel III.1, hasil perhitungan akurasi menggunakan Mean Absolute Error (MAE) terhadap metode *Collaborative Filtering* adalah 2,01. Pada perhitungan MAE, semakin kecil hasil MAE yang diperoleh maka kesalahan pada sistem juga semakin sedikit. Jadi dapat disimpulkan bahwa penggunaan metode *Collaborative Filtering* dalam sistem rekomendasi bacaan tugas akhir dinilai masih memiliki beberapa kelemahan.

IV. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa hasil pengujian akurasi menggunakan *Mean Absolute Error* pada penelitian *Collaborative Filtering* adalah 2,01. Dan hasil pengujian akurasi pada penelitian *Naive Bayes* adalah 81%. dimana dari 100 data uji yang digunakan, terdapat 19 data yang tidak relevan ditemukan. Hal ini disebabkan karena sedikitnya jumlah keragaman data yang digunakan sehingga sistem menampilkan hasil yang tidak sesuai.

REFERENSI

- [1] Bundasak, S., & Chinnasarn, K. (2013). eMenu Recommender System using Collaborative. *10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 37-42.
- [2] Ghazanfar, M. A., & Prügel-Bennett, A. (2010). An Improved Switching Hybrid Recommender System Using Naive Bayes Classifier and Collaborative Filtering.
- [3] Robles, V., Larrañaga, P., Menasalvas, E., Pérez, M. S., & Herves, V. (2003). Improvement of naïve Bayes Collaborative Filtering using Interval Estimation. *IEEE*, 1-7.
- [4] Mittal, N., Nayak, R., Govil, M., & Jain, K. (2010). Recommender System Framework using Clustering and Collaborative Filtering. *IEEE*, 555-558.
- [5] Parvatikar, S., & Joshi, D. B. (2015). Online Book Recommendation System by using Collaborative filtering and Association Mining. *IEEE*.
- [6] Pratama, Y. A., Wijay, D., Paulus, & Halim, A. (2013). Digital Cakery dengan Algoritma Collaborative Filtering. 79-88.
- [7] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based Collaborative Filtering Recommendation Algorithms. *WWW10*, 285-295.
- [8] Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2010). *Recommender Systems Handbook*. Springer.