

Perbandingan Algoritma *Content-Based Filtering* dan *Collaborative Filtering* dalam Rekomendasi Kegiatan Ekstrakurikuler Siswa

Diyo Sukma Pradana^{1*}, Prajoko², George Pri Hartawan³

Program Studi Teknik Informatika, Universitas Muhammadiyah Sukabumi

Jl. R. Syamsudin, S.H. No. 50 Cikole, Sukabumi, Indonesia

*e-mail *Corresponding Author*: diyosukmapradana@gmail.com

Abstract

Extracurricular activities play an important role in developing students' creativity. However, the problem that is often experienced by students in determining the choice of extracurricular activities is choosing the right type of activity and in line with the interests and talents of students. This study aims to test and compare the performance of the Naïve Bayes-based Content-based Filtering and Collaborative Filtering models in recommending appropriate extracurricular activities for students. Testing of each model is done by dividing the training data and test data in a ratio of 80% and 20%. The training process uses the RecommenderNET Library. The accuracy of the Content-based Filtering model was tested using Naïve Bayes of the Multinomial type, while the Collaborative Filtering model used the Gaussian type of Nave Bayes. The test results of the Naïve Bayes model for Content-based Filtering show an accuracy rate of 74%, while Collaborative Filtering obtains 56%.

Keywords: Recommendation System; Naïve Bayes; Library RecommenderNET

Abstrak

Kegiatan ekstrakurikuler memegang peran penting dalam mengembangkan kreativitas siswa. Namun demikian, permasalahan yang sering dialami oleh siswa dalam menentukan pilihan kegiatan ekstrakurikuler adalah memilih jenis kegiatan yang tepat dan sejalan dengan minat dan bakat siswa. Penelitian ini bertujuan untuk menguji dan membandingkan kinerja model *Content-based Filtering* dan *Collaborative Filtering* berbasis *Naïve Bayes* dalam merekomendasikan kegiatan Ekstrakurikuler yang tepat bagi siswa. Pengujian masing-masing model dilakukan dengan membagi data latih dan data uji dalam perbandingan 80% dan 20%. Proses pelatihan menggunakan *Library RecommenderNET*. Akurasi model *Content-based Filtering* diuji menggunakan *Naïve Bayes* jenis *Multinomial*, sedangkan model *Collaborative Filtering* menggunakan *Naïve Bayes* jenis *Gaussian*. Hasil uji model *Naïve Bayes* untuk *Content-based Filtering* menunjukkan tingkat akurasi 74%, sedangkan *Collaborative Filtering* memperoleh 56%.

Kata kunci: Sistem Rekomendasi; Naïve Bayes; Library RecommenderNET

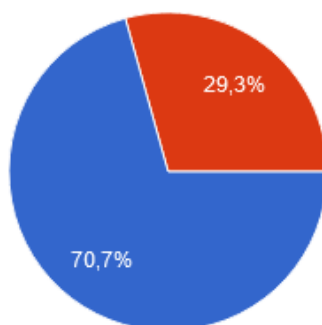
1. Pendahuluan

Pendidikan memiliki fungsi penting untuk meningkatkan kualitas sumber daya manusia [1], sekolah sebagai lembaga pendidikan formal bertugas untuk membentuk karakter penerus bangsa yang bermartabat dan bersahaja, sekolah juga berfungsi untuk mengembangkan potensi-potensi siswa dalam mewujudkan tujuan pendidikan nasional. Dalam sekolah ada berbagai jenis mata pelajaran yang bervariasi mulai dari sains, sosial sampai lingkungan. Di sekolah SMA sendiri ada 3 jenis disiplin ilmu yakni IPA, IPS, dan Bahasa. Selain belajar di kelas sekolah juga menyediakan kegiatan di luar kelas, yaitu kegiatan ekstrakurikuler. Kegiatan Ekstrakurikuler merupakan salah satu cara untuk menggali potensi diri dan berekspresi sesuai dengan minat dan bakat masing-masing [2].

Peraturan tentang kegiatan ekstrakurikuler disebutkan pada Permendikbud nomor 62 tahun 2014, bahwa “*Setiap peserta didik memiliki hak mengikuti kegiatan ekstrakurikuler sesuai dengan minat dan bakat mereka*”. Atas dasar tersebut, pemilihan ekstrakurikuler yang terjadi di Sekolah Menengah Atas (SMA) diserahkan sepenuhnya kepada Siswa. Namun sering terjadi

dimana terdapat siswa yang masih bimbang dalam pemilihan jenis kegiatan ekstrakurikuler yang sesuai dengan minatnya, serta campur tangan orang tuanya yang sering dianggap terlalu dominan membatasi kebebasan siswa [3].

Hasil survey pada penelitian awal yang kami lakukan melalui kuesioner, menunjukkan bahwa dari 358 Responden 70,7% menjawab Ya dan 29,3% menjawab Tidak untuk pertanyaan “Apakah anda kesulitan dalam memilih kegiatan Ekstrakurikuler saat SMA?”



Gambar 1. Hasil Kuesioner Kesulitan Memilih Ekstrakurikuler

Hasil survey tersebut membuktikan bahwa sebagian besar siswa memang masih kesulitan dalam memilih kegiatan ekstrakurikuler yang sesuai dengan minat dan bakat masing-masing siswa. Atas dasar tersebut, dipandang perlu untuk mengembangkan model sistem rekomendasi sebagai alat untuk memberikan pertimbangan dalam memilih kegiatan ekstrakurikuler siswa khususnya siswa Sekolah Menengah Atas (SMA).

Dalam dunia Teknologi Informasi, terdapat 2 model sistem rekomendasi yang biasa dikembangkan, yaitu *Content-based Filtering* dan *Collaborative Filtering*. Metode *Content-based Filtering*, memiliki fungsi merekomendasikan sesuatu/item yang mirip dengan sesuatu/item yang disukai pengguna di masa lalu [4]. Misal, jika kita menyukai buku berjudul; “Laskar Pelangi Karya Andrea Hirata”, maka sistem yang dibuat akan merekomendasikan buku dengan penulis yang sama Andrea Hirata, sistem ini juga akan merekomendasikan buku dengan genre yang sama. Sedangkan contoh dari metode *Collaborative Filtering* ialah jika A menyukai buku Andrea Hirata, B menyukai buku Tere Liye dan Andrea Hirata, sedangkan C menyukai buku Tere Liye, maka sistem akan merekomendasikan A buku Tere Liye dan C buku Andrea Hirata, sistem ini bekerja bergantung pada pendapat komunitas pengguna [4].

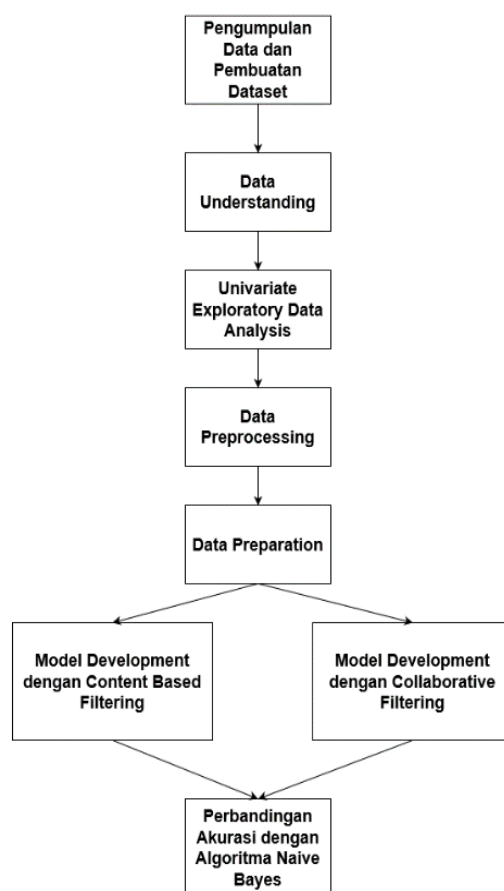
Perbandingan kedua metode menurut jurnal *Comparing Content Based and Collaborative Filtering Recommender Systems* berkesimpulan bahwa jika melirik dari hasil kemiripan deskripsi, produk sampai atribut lain maka *Content Based* lebih unggul [5]. Hal tersebut memberi peluang peneliti lebih lanjut membandingkan kedua sistem rekomendasi namun dalam kasus yang berbeda yaitu Sistem Rekomendasi Ekstrakurikuler Siswa SMA.

2. Tinjauan Pustaka

Sudah ada beberapa penelitian yang berkaitan dengan topik sistem rekomendasi kegiatan ekstrakurikuler siswa, seperti Sistem Rekomendasi Ekstrakurikuler dengan *Content-based Filtering* dalam jurnal Penerapan Metode Content-Based Filtering Pada Sistem Rekomendasi Kegiatan Ekstrakurikuler (Studi Kasus di Sekolah ABC) oleh Firmahsyah dan Gantini[6] dengan menggunakan algoritma Naïve Bayes sebagai evaluasi memperoleh akurasi rata-rata 0,64, dan juga Sistem rekomendasi buku dengan *Collaborative Filtering*, seperti penelitian oleh Ergina dalam jurnalnya yang berjudul *Sistem Rekomendasi Pemilihan Buku Menggunakan Algoritma Collaborative Filtering pada Perpustakaan Universitas Muhammadiyah Sukabumi* [7], dengan menggunakan MAE memperoleh akurasi optimal 93%.

Untuk menguji/evaluasi model, peneliti menelusuri berbagai jenis algoritma evaluasi dari beberapa penelitian sebelumnya menyebutkan bahwa algoritma Naive bayes itu lebih unggul daripada *Decision Tree* dan *Neural Network*, algoritma *Naïve Bayes* memperoleh hasil *training* sebesar 95,20% dibanding kedua algoritma yang lain [8], dibandingkan juga dengan KNN, Naive Bayes masih unggul [9], maka dari itu peneliti menggunakan algoritma *Naïve bayes* sebagai pengujian akurasi model.

3. Metodologi



Gambar 2. Tahapan Metode Penelitian

Penelitian ini dilakukan selama 3 Bulan, pengumpulan data dilakukan selama 1 Bulan untuk mengumpulkan data-data alumni/siswa SMA yang sudah mengikuti ekstrakurikuler di seluruh Indonesia, data tersebut dipakai untuk keperluan pembuatan dataset. Data responden yang tercatat sudah mencapai 358 Orang, 175 Laki-laki, 183 Perempuan. Hal yang akan diteliti juga diuji ialah perbandingan akurasi kedua Sistem Rekomendasi dengan metode *Content-based Filtering* dan *Collaborative Filtering* menggunakan Algoritma Naive Bayes sebagai algoritma pengujian/evaluasi.

Hal pertama yang dilakukan ialah merapikan data dan analisis data ditahap *Data Understanding* dan *Univariate Exploratory Data Analysis*, memilih atribut atau kolom dataset yang dipakai untuk proses pembuatan sistem rekomendasi. Selanjutnya Data disiapkan dan dicek ulang apakah data ada nilai *null* atau kosong di tahap *Data Preprocessing* dan *Data Preparation*. Tahap terakhir ialah proses *Modeling* atau *Model Development* dan Perbandingan Akurasi menggunakan Naive Bayes. Data dibagi/*split* untuk *Content-based Filtering* data *trainnya* 42% dan 58% data *testnya* sedangkan untuk *Collaborative Filtering* data *train* 60% dan *test* 40%.

Entitas atau variabel utama yang diteliti adalah bidang Ekstrakurikuler, minat dan bakat juga variabel pendukung apa saja yang membuat seseorang memilih Bidang Ekstrakurikuler tersebut, adapun bidang ekstrakurikulernya ialah:

1. Bidang Akademik dan Kebahasaan, seperti *English club*, *Japanese club*, *debate club*, dan sebagainya
2. Bidang Bela Negara, seperti Pramuka, Paskibra dan PMR
3. Bidang IT/Teknologi, seperti komputer, robotic dan sebagainya
4. Bidang Kesenian, seperti paduan suara, tari, lukis, theater dan sebagainya
5. Bidang Lingkungan Hidup, eskul yang berhubungan tentang alam seperti pendakian dan sebagainya
6. Bidang Olah raga, seperti basket, sepak bola, futsal dan sebagainya

4. Hasil dan Pembahasan

4.1. Pengumpulan Data dan Pembuatan Dataset

Tahapan awal penelitian yaitu mengumpulkan data berupa responden kuesioner, data ini dibuat pada *Google Form* yang selanjutnya data diolah menjadi sebuah dataset yang bisa digunakan untuk penelitian.

4.2. Data Understanding dan Univariate Exploratory Data Analysis

Memahami data lebih lanjut, apa saja atribut didalamnya dan berfungsi sebagai apa. Berikut atribut dataset yang sudah diolah, bisa dilihat pada Tabel 1.

Tabel 1. Atribut Dataset

No.	Atribut	Keterangan
1.	Nama	Berfungsi sebagai id, peneliti mengubah nama menjadi Siswa1 dst untuk privasi data
2.	Asal Sekolah	Asal sekolah SMA Responden, di seluruh Indonesia
3.	Sulit Memilih	Pertanyaan apakah responden ketika memilih ekstrakurikuler kesulitan
4.	Gender	Jenis kelamin
5.	Agama	5 Agama di Indonesia
6.	Jurusan	Jurusan ketika sekolah SMA, ada IPA, IPS, dan Bahasa
7.	Bidang Ekstrakurikuler	Bidang Ekstrakurikuler yang dipilih ketika sekolah SMA, ada 5 Bidang yaitu : Bidang Akademik dan Bahasa, Bidang Olahraga, Bidang IT, Bidang Bela Negara, Bidang Lingkungan Hidup dan Bidang Kesenian
8.	Pekerjaan Ibu	Pekerjaan Ibu Responden
9.	Pekerjaan Bapak	Pekerjaan Bapak Responden
10.	Sungguh Berkegiatan	Pernyataan Ya atau Tidak, ketika mengikuti ekstrakurikuler itu sungguh-sungguh atau tidak
11.	Minat Bakat Tersalurkan	Pernyataan Ya atau Tidak, ketika mengikuti ekstrakurikuler minat dan bakat tersalurkan
12.	Alasan Memilih	Alasan memilih Ekstrakurikuler
13.	Teman Berpengaruh	Pertanyaan Ya atau Tidak, apakah teman berpengaruh terhadap pemilihan ekstrakurikuler
14.	Lama Pertemuan	Berapa lama pertemuan ekstrakurikuler
15.	Banyak Pertemuan	Berapa banyak pertemuan dalam seminggu kegiatan ekstrakurikuler tersebut
16.	Tes Minat Bakat	Pemilihan minat dan bakat responden sesuai dengan karakter pribadi, terdapat 48 pertanyaan seputar minat bakat ini.

Berikut butir instrumen seputar tes minat bakat pada kuesioner, dimana pertanyaan diambil dari jurnal yang berkaitan dengan minat/bakat seseorang [10]:

Tabel 2. Pertanyaan Minat dan Bakat pada Dataset

No	Pertanyaan Minat dan Bakat
1	Gemar melakukan kegiatan yang berkaitan dengan menulis
2	Suka menulis baik dalam bentuk cerpen, puisi dan sejenisnya
3	Suka bercerita atau berinteraksi secara verbal
4	Tertarik dalam dunia tulis-menulis, khususnya menulis cerita fiksi
5	Suka berargumen
6	Suka ilmu linguistic
7	Mempunyai peliharaan, baik hewan maupun tanaman
8	Cinta dan suka mempelajari hal berbau alam
9	Suka berkunjung ke kebun binatang
10	Menyukai kegiatan pecinta alam
11	Suka dengan kegiatan yang tenang, contohnya memancing
12	Melindungi kelestarian alam atau lingkungan
13	Suka memikirkan cita-cita yang ingin dicapai

No	Pertanyaan Minat dan Bakat
14	Cenderung individualis atau melakukan kegiatan secara sendiri
15	Cenderung menyukai bekerja hanya dengan kemampuan sendiri
16	Suka menulis berdasarkan pengalaman dan perasaan
17	Mempertimbangkan sisi plus dan minus untuk keputusan yang akan diambil dalam menyelesaikan sebuah masalah
18	Paham akan sisi positif dan negatif yang ada dalam diri sendiri
19	Suka dengan permainan yang perlu pemikiran kritis
20	Menyukai sains
21	Suka dengan dunia eksakta
22	Menyukai komputer dan seluk beluknya
23	Lebih mudah untuk bekerja dan berpikir mengenai hal-hal yang ada hubungannya dengan angka
24	Menyukai matematika
25	Suka berinteraksi sesame
26	Mudah beradaptasi dengan orang baru
27	Lebih suka bekerja secara berkelompok
28	Peduli dan suka membantu orang lain
29	Mempunyai gagasan atau ide untuk membuat orang lain nyaman dalam bekerja sama
30	Seringkali menjadi tempat curhat untuk teman
31	Suka dan bisa memainkan alat music
32	Suka dengan dunia tarik suara dan music
33	Mudah meresapi suatu nada atau melodi dari lagu
34	Mampu dengan mudah mendefinisikan lagu dan instrumen dari berbagai musik dan lagu
35	Suka bekerja dengan mendengarkan suatu lagu atau irama
36	Suka merangkai kata dan memasukkannya ke dalam sebuah lagu
37	Suka dengan lego atau puzzle
38	Suka dengan dunia fotografi
39	Suka melukis
40	Meniru cara kerja dari orang yang lain jika menghadapi suatu permasalahan
41	Mempunyai kemampuan visualisasi dalam bentuk gambar yang cenderung baik
42	Mempunyai kemampuan yang bagus di dalam mapping/mempetakan suatu wilayah
43	Suka mengolah tubuh/olahraga
44	Jika sedang berpikir, cenderung menggerakkan anggota tubuh tertentu
45	Merasa jenuh jika harus berdiam diri secara terus menerus
46	Suka dengan dunia seni peran dan/atau menari
47	Suka kegiatan yang dilakukan di luar ruangan/outdoor
48	Lebih menyukai kegiatan praktik daripada hanya teori

4.3. Data Preprocessing dan Data Preparation

	Nama	Bidang Ekstrakurikuler	Gender	Jurusan	Teman Berpengaruh	Lama Pertemuan	Banyak Pertemuan	Minat Bakat
0	Siswa01	Kesenian	Laki-laki	IPS	Ya	Lebih dari 1 Jam	2 Kali Seminggu	Cenderung menyukai untuk bekerja secara berkel...
1	Siswa02	Olahraga	Laki-laki	IPS	Ya	Lebih dari 1 Jam	2 Kali Seminggu	Cenderung menyukai untuk bekerja secara berkel...
2	Siswa03	Akademik & Kebahasaan	Perempuan	IPA	Ya	1 Jam	1 Kali Seminggu	Menggemari kegiatan yang melibatkan tulisan, M...
3	Siswa04	Olahraga	Perempuan	IPA	Tidak	Lebih dari 1 Jam	2 Kali Seminggu	Mempertimbangkan baik sisi plus ataupun minus ...
4	Siswa05	Kesenian	Perempuan	IPA	Tidak	Lebih dari 1 Jam	2 Kali Seminggu	Mempertimbangkan baik sisi plus ataupun minus ...

Gambar 3. Dataset untuk *Content-based Filtering*

Pada tahapan *Data Preprocessing*, peneliti memproses data apa saja yang akan digunakan untuk proses modelling, untuk metode *Content-based Filtering* data yang digunakan ialah atribut: Nama, Bidang Ekstrakurikuler, Gender, Jurusan, Teman Berpengaruh, Lama Pertemuan, Banyak Pertemuan, dan Tes Minat Bakat (48 pertanyaan Minat dan Bakat), sedangkan untuk *Collaborative Filtering* Dataset yang digunakan hanya 48 Pertanyaan Minat dan

Bakat yang sudah diolah (bernilai 1 jika Ya, 0 jika Tidak), juga 6 bidang ekstrakurikuler. Pada *Data Preparation*, mengecek sekali lagi data jika ada yang bernilai *null* atau kosong.

	esku1	p1	p2	p3	p4	p5	p6	p7	p8	p9	...	p39	p40	p41	p42	p43	p44	p45	p46	p47	p48
0	Kesenian	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	1	1	0	0	1
1	Olahraga	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	1	1	0	0	1
2	Akademik & Kebahasaan	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	Olahraga	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	1	0	1	1	1
4	Kesenian	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	1	0	1	1	1

Gambar 4. Dataset *Collaborative Filtering*

4.4. Model Development Content-based Filtering

Dalam proses development metode ini, tahap pertama ialah mengubah representasi text menjadi vektor dengan *CountVectorizer*, contoh sederhana ada 3 buah dokumen/teks:

1. Siswa1 Menyukai Musik dan Membaca
2. Siswa2 Menyukai Membaca
3. Siswa3 Menyukai Komputer

Dari 3 dokumen/teks di atas tersusunlah kata kunci/*vocabulary*: Siswa1, Siswa2, Siswa3, Menyukai, Musik, Membaca, Komputer.

Tabel 3. Contoh Representasi *CountVectorizer*

Teks/Dokumen	Kata Kunci						
	Siswa1	Siswa2	Siswa3	Menyukai	Musik	Membaca	Komputer
Siswa1 Menyukai Musik dan Membaca	1	0	0	1	1	1	0
Siswa2 Menyukai Membaca	0	1	0	1	0	1	0
Siswa3 Menyukai Komputer	0	0	1	1	0	0	1

Maka representasi vektornya ialah:

1. Dokumen/teks 1 : [1,0,0,1,1,1,0]
2. Dokumen/teks 2 : [0,1,0,1,0,1,0]
3. Dokumen/teks 3 : [0,0,1,1,0,0,1]

Berikut implementasi pada program, contoh dokumen 1 yang sudah diekstraksi dan gabungkan teksnya dalam kolom baru **bag_of_words**:

"Kesenian Laki-laki IPS Ya Lebih dari 1 Jam 2 Kali Seminggu cenderung menyukai untuk bekerja secara berkelompok dan memiliki kemampuan dalam bermain alat musik mampu dengan mudah meresapi nada atau melodi dari sebuah lagu gemar sambil mendengarkan suatu irama kegiatan olah tubuh menggerakkan anggota tertentu ketika berpikir sering merasa jenuh diharuskan berdiam diri terus menerus lebih praktik daripada teori."

```
[ 'ada', 'baru',
  'akademik', 'bekerja',
  'akan', 'bela',
  'alam', 'bentuk',
  'alat', 'beradaptasi',
  'anggota', 'berargumen',
  'angka', 'berbagai',
  'atau', 'berbagi',
  'ataupun', 'bercerita',
  'bagi', 'berdian',
  'bahasa', 'berhubungan',
  'baik', 'berinteraksi',
  'berkelompok',
```

Gambar 5. Contoh *Vocabulary*/kata kunci metode *Content-based Filtering*

```
Count Matrix: [[0 0 0 ... 0 1 0]
               [0 0 0 ... 0 1 0]
               [0 1 1 ... 0 1 1]
               ...
               [0 0 0 ... 0 1 1]
               [1 0 0 ... 0 1 1]
               [0 0 0 ... 0 1 0]]
```

Gambar 6. Matriks Representasi Vektor dengan CountVectorizer

Untuk membandingkan kemiripan antara dokumen/teks 1 dengan yang lain, peneliti menggunakan *Cosine Similarity*, berikut rumus perhitungannya [11]:

$$Similarity(A, B) = \cos(\theta) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \dots\dots\dots(1)$$

- Keterangan : $\cos(\theta)$ = tingkat kesamaan dokumen
 A = term ke i untuk vektor dokumen ke A
 B = term ke i untuk vektor dokumen ke B
 n = jumlah data

Contoh sederhana kita pakai representasi vektor CountVectorizer:

- 1) Siswa1 Menyukai Musik dan Membaca: [1,0,0,1,1,1,0]
- 2) Siswa2 Menyukai Membaca: [0,1,0,1,0,1,0]

Jika Kita ingin melihat kemiripan antara A dan B, maka:

$$A \times B = (1 \times 0) + (0 \times 1) + (0 \times 0) + (1 \times 1) + (1 \times 0) + (1 \times 1) + (0 \times 0) = 2$$

$$\sqrt{A^2} = \sqrt{1^2 + 0^2 + 0^2 + 1^2 + 1^2 + 1^2 + 0^2} = \sqrt{4} = 2$$

$$\sqrt{B^2} = \sqrt{0^2 + 1^2 + 0^2 + 1^2 + 0^2 + 1^2 + 0^2} = \sqrt{3} = 1,7$$

Maka Hasil dari similarity (A,B) = 2/(2) x (1,7) = 2/3,4 = 0,59

Maka kemiripan antara dokumen A dan B adalah 59%.

Gambar 7 menyajikan implementasi pada program, contoh hasil dari *Cosine Similarity*:

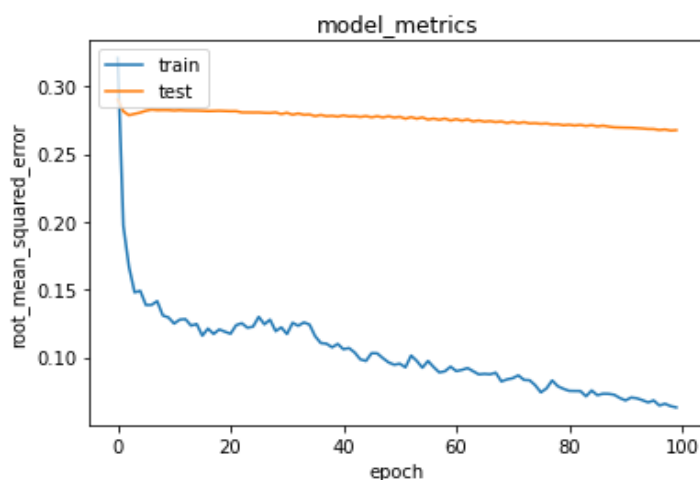
```
array([[1.          , 0.984375 , 0.28653494, ..., 0.525      , 0.39583333,
        0.57765089],
       [0.984375 , 1.          , 0.28653494, ..., 0.55      , 0.41666667,
        0.55901699],
       [0.28653494, 0.28653494, 1.          , ..., 0.21574396, 0.24720662,
        0.3216121 ],
       ...,
       [0.525      , 0.55      , 0.21574396, ..., 1.          , 0.66666667,
        0.17888544],
       [0.39583333, 0.41666667, 0.24720662, ..., 0.66666667, 1.          ,
        0.2236068 ],
       [0.57765089, 0.55901699, 0.3216121 , ..., 0.17888544, 0.2236068 ,
        1.          ]])
```

Gambar 7. Matriks *Cosine Similarity Content-based Filtering*

4.5. Model Development Collaborative Filtering

Metode ini menggunakan hasil dari jumlah ceklis yang dilakukan oleh responden tentang minat bakat mereka dan ekstrakurikuler yang mereka pilih, Misalkan Siswa01 yang masuk ke dalam ekstrakurikuler Olahraga menceklis 7 dari 48 pertanyaan yang sesuai dengan minat dan bakatnya.

Peneliti membagi data train dan test sebesar 80% dan 20%. Menggunakan *Library RecommenderNET* dari *Keras* sebagai proses pelatihan, berikut hasil grafik pelatihan model:



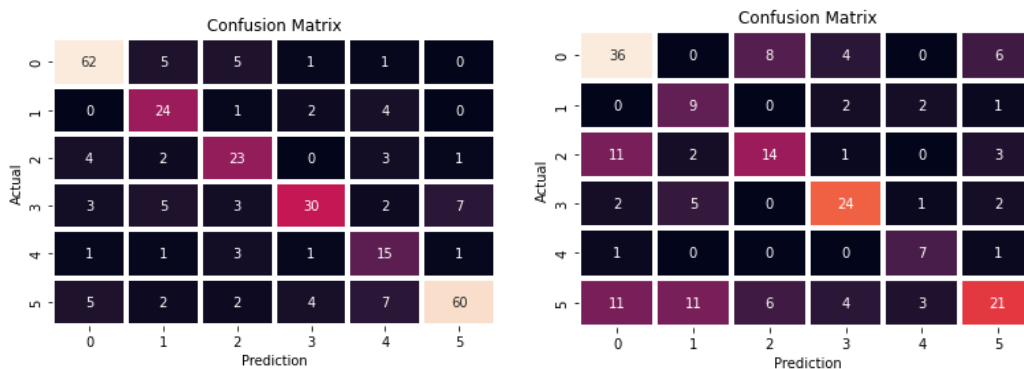
Gambar 8. Model Metrics Collaborative Filtering

Semakin nilai RMSE (*Root Mean Squared Error*) mendekati nol maka akurasi model semakin baik [12], bisa dilihat nilai grafik RMSE paling kecil $<0,10$ dan validasi data $0,27$, ini membuktikan model bekerja dengan baik.

4.6. Perbandingan Akurasi dengan Naïve Bayes

Untuk metode *Content-based Filtering*, peneliti menggunakan *Naïve Bayes* jenis *Multinomial Naïve Bayes*, karena jenis ini cocok digunakan pada data yang bentuknya teks dan memprediksi setiap klasifikasi dari teks-teks [13] tersebut, yakni menghitung akurasi *Count Vectorizer* dalam kolom *bag_of_words* terhadap 6 bidang ekstrakurikuler, data dibagi menjadi 42% untuk train dan 58% untuk test,

Sedangkan untuk metode *Collaborative Filtering* peneliti menggunakan *Naïve Bayes* tipe *Gaussian*[14], untuk menghitung nilai dari setiap jawaban responden dan memprediksi hasil responden. Berikut visualisasi confusion matrix hasil prediksi vs aktual kedua metode:



Gambar 9. Confusion Matrix Content-Based Filtering (kiri) dan Collaborative Filtering (kanan)

Keterangan lebih lanjut :

- Actual = Aktual Nilai Data yang benar
- Prediction = Nilai Prediksi Data oleh Algoritma *Naïve Bayes*
- 0 = Bidang Akademik & Kebahasaan
- 1 = Bidang Bela Negara
- 2 = Bidang IT
- 3 = Bidang Kesenian
- 4 = Bidang Lingkungan Hidup
- 5 = Bidang Olahraga

Sebelum menghitung akurasi peneliti menghitung dahulu *precision*, *recall* dan *f1-score* dari *confusion matrix* di atas, dengan total prediksi yang benar diperlihatkan oleh TP (*True*

Positive) dan TN (True Negative). Sedangkan banyaknya prediksi yang salah diperlihatkan oleh FP (False Positive) dan FN (False Negative) [13], berikut contoh cara perhitungan:

Precision, menghitung tingkat kepastian sebagai perbandingan sampel kelas positif yang diklasifikasikan dengan benar/true terhadap keseluruhan sampel kelas positif.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(2)$$

$$Precision \text{ Akademik \& Kebahasaan (Pa)} = \frac{TPa}{TPa+FPa}$$

$$= \frac{62}{(62+0+4+3+1+5)} = 0,8267 = 0,83$$

Recall, mengukur rasio diantara sampel kelas positif yang diklasifikasikan dengan benar terhadap sampel kelas positif yang salah diklasifikasikan ke dalam kelas negatif.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(3)$$

$$Recall \text{ Akademik \& Kebahasaan (Ra)} = \frac{TPa}{TPa+FN}$$

$$= \frac{62}{(62+5+5+1+1+0)} = 0,8378 = 0,84$$

F1-Score, merupakan harmonic mean dari precision dan recall.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \dots\dots\dots(4)$$

$$F1\text{-score} = 2 \times ((Precision \times Recall) / (Precision + Recall))$$

$$F1\text{-score Akademik \& Kebahasaan} = 2 \times ((Pa \times Ra) / (Pa + Ra))$$

$$= 2 \times ((0,83 \times 0,84) / (0,83 + 0,84)) = 2 \times 0,42 = 0,83$$

Accuracy Multiclass, perhitungan akurasi multiclass berbeda dengan perhitungan akurasi binary, berikut rumus akurasi dari multiclass, yang akan digunakan untuk menghitung akurasi model [15]:

$$Accuracy \text{ Multiclass} = \frac{TP \text{ All Class}}{Total \text{ Support}} \dots\dots\dots(5)$$

$$Akurasi \text{ Content-based} = \frac{(62+24+23+30+15+60)}{290} = \frac{214}{290} = 0,74$$

Secara lebih detail, Gambar 10 menyajikan perbandingan precision, recall, f1-score dan akurasi:

	precision	recall	f1-score	support		precision	recall	f1-score	support
Akademik & Kebahasaan	0.83	0.84	0.83	74	0	0.59	0.67	0.63	54
Bela Negara	0.62	0.77	0.69	31	1	0.33	0.64	0.44	14
IT	0.62	0.70	0.66	33	2	0.50	0.45	0.47	31
Kesenian	0.79	0.60	0.68	50	3	0.69	0.71	0.70	34
Lingkungan Hidup	0.47	0.68	0.56	22	4	0.54	0.78	0.64	9
Olahraga	0.87	0.75	0.81	80	5	0.62	0.38	0.47	56
accuracy			0.74	290	accuracy			0.56	198
macro avg	0.70	0.72	0.70	290	macro avg	0.54	0.60	0.56	198
weighted avg	0.76	0.74	0.74	290	weighted avg	0.58	0.56	0.56	198

Gambar 10. Perbandingan Confusion Matrix Metode Content-based (kiri) dan Collaborative (kanan)

Pada Gambar 10 dapat dilihat bahwa akurasi model Content-based Filtering dapat mencapai 74%, sedangkan model Collaborative Filtering hanya mencapai 56%. Kedua hasil uji akurasi tersebut menunjukkan bahwa dalam kasus sistem rekomendasi pemilihan jenis kegiatan ekstrakurikuler, algoritma Content-based Filtering lebih rasional digunakan jika dibandingkan dengan algoritma Collaborative Filtering. Temuan ini tidak sejalan dengan temuan Rizky at al. [16] yang telah menguji algoritma Content Based Filtering dan Collaborative Filtering dalam kasus Rekomendasi Program Studi untuk Siswa Sekolah Menengah Atas Sederjat. Pada penelitian [16] tersebut, akurasi algoritma Content Based Filtering mencapai 88,7%, sedangkan collaborative filtering mencapai 98%.

5. Simpulan

Akurasi yang dihasilkan oleh Algoritma Naïve Bayes pada metode *Content-based Filtering* sebesar 74% ini unggul dibanding dengan *Collaborative Filtering* hanya 56%. Dengan tingkat akurasi tersebut, diperlukan upaya lanjutan agar tingkat akurasi kedua algoritma tersebut dapat lebih baik sebelum benar-benar digunakan sebagai sistem penunjang pengambilan keputusan berkaitan dengan peminatan. Pengujian lebih lanjut juga perlu dilakukan, dengan menambah porsi data uji, agar karakteristik kedua algoritma tersebut dapat dipahami secara lebih mendalam.

Daftar Referensi

- [1] S. M. Tinambunan, "Hubungan Kegiatan Ekstrakurikuler Terhadap Prestasi Belajar Siswa/i di SMA Negeri 1 Sunggal," Skripsi, Program Studi Sosiologi, Universitas Sumatera utara, Kota Medan, 2018.
- [2] A. Louis and D. Putri, "Aplikasi Sistem Pakar Untuk Menentukan Ekstrakurikuler Terhadap Minat Bakat Siswa (Studi Kasus Di Smpn 5 Kota Jambi)," *J. Acad.*, vol. 11, no. 2, pp. 34–37, 2019.
- [3] D. K. Sari and P. Simanjuntak, "Sistem Pakar Penentuan Minat Dan Bakat Ekstrakurikuler Siswa," *J. Comasie*, vol. 3, no. 3, pp. 21–30, 2020.
- [4] Dicoding Indonesia, "Machine Learning Terapan," *Dicoding Indonesia*, 2021. <https://www.dicoding.com/academies/319/tutorials/171116> (accessed Feb. 07, 2022).
- [5] P. Aggarwal, V. Tomar, and A. Kathuria, "Comparing Content Based and Collaborative Filtering in Recommender Systems," *Int. J. New Technol. Res.*, vol. 3, no. 4, pp. 65–67, 2017.
- [6] F. Firmahsyah and T. Gantini, "Penerapan Metode Content-Based Filtering Pada Sistem Rekomendasi Kegiatan Ekstrakurikuler (Studi Kasus di Sekolah ABC)," *J. Tek. Inform. dan Sist. Inf.*, vol. 2, no. 3, pp. 414–427, 2016.
- [7] E. K. Putri, "Sistem Rekomendasi Pemilihan Buku Menggunakan Algoritma Collaborative Filtering pada Perpustakaan Universitas Muhammadiyah Sukabumi," vol. 20, no. 2, pp. 338–343, 2021.
- [8] D. Xhemali, C. J. Hinde, and R. G. Stone, "Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *Int. J. Comput. Sci.*, vol. 4, no. 1, pp. 16–23, 2019.
- [9] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016.
- [10] M. H. Prami Swari, R. W. Arianti, and F. Muttaqin, "Case-Based Reasoning Pemberian Rekomendasi Profesi Berdasarkan Minat Dan Bakat Siswa Menggunakan Simple Matching Coefficient Similarity," *SINTECH (Science Inf. Technol. J.)*, vol. 3, no. 1, pp. 35–45, 2020.
- [11] G. Ferio, R. Intan, and S. Rostianingsih, "Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Metode User Based Collaborative Filtering Berbasis Algoritma Adjusted Cosine Similarity," *J. Infra*, vol. 7, no. 1, pp. 39–45, 2019.
- [12] S. Rahmawati, D. Nurjanah, and R. Rismala, "Analisis dan Implementasi pendekatan Hybrid untuk Sistem Rekomendasi Pekerjaan dengan Metode Knowledge Based dan Collaborative Filtering," *Indones. J. Comput.*, vol. 3, no. 2, pp. 11–20, 2018.
- [13] A. F. Hidayatullah, A. Aulia, F. Yusuf, K. P. Juwairi, R. Abida, and N. Nayoan, "Identifikasi konten kasar pada tweet bahasa Indonesia," *J. Linguist. Komputasional*, vol. 2, no. 1, pp. 1–5, 2019.
- [14] S. Syarli and A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, pp. 22–26, 2016.
- [15] E. Bagli, G. Visani, and M. Grandini, "Metrics For Multi-Class Classification: An Overview," arXiv preprint, Bologna, Italy, pp. 1-17, 2020.
- [16] M.I. Rizky, I. Asror, & Y.R. Murti. "Sistem Rekomendasi Program Studi untuk Siswa SMA Sederajat Menggunakan Metode Hybrid Recommendation dengan Content Based Filtering dan Collaborative Filtering". *eProceedings of Engineering*, vol. 7, no. 1, pp. 2776-2792, 2020.