

## Sistem Rekomendasi Program Studi untuk Siswa SMA Sederajat Menggunakan Metode *Hybrid Recommendation* dengan *Content Based Filtering* dan *Collaborative Filtering*

Muthi Ishlah Rizky<sup>1</sup>, Ibnu Asror<sup>2</sup>, Yusza Reditya Murti<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>muthiish@students.telkomuniversity.ac.id, <sup>2</sup>iasros@telkomuniversity.ac.id, <sup>3</sup>yuszaa@telkomuniversity.ac.id

### Abstrak

Pendidikan tinggi merupakan sebuah wadah untuk menghasilkan SDM yang unggul dan berfikir, tidak heran banyak perguruan tinggi diminta untuk menghasilkan mahasiswa yang unggul dan mampu bersaing didunia kerja. Akan tetapi, faktanya adalah sebanyak 45% mahasiswa merasa salah dalam memilih program studi yang tepat, dan sebanyak 92% siswa SMA sederajat merasa kebingungan untuk menentukan masa depannya. Oleh karena itu perlu adanya sistem untuk merekomendasikan program studi, sehingga mampu mengurangi fenomena tersebut. Metode yang digunakan untuk memberikan rekomendasi program studi adalah *hybrid recommendation*. Metode *hybrid recommendation* ini dipilih karena mampu mempelajari profil *user*, dengan mengembangkan metode *naive bayes* untuk mengklasifikasi program studi dengan mengolah data nilai, sedangkan untuk mempertimbangkan minat siswa menggunakan *item-based collaborative filtering* dengan menghitung kemiripan pilihan prodi yang diberikan oleh siswa. Dalam menggabungkan kedua metode ini akan menghasilkan program yang sesuai dengan nilai dan minat siswa untuk memilih program studi. Hasil penelitian ini menunjukkan bahwa pengolahan data nilai dengan *naive bayes* menghasilkan akurasi sebesar 88.7% , sedangkan untuk pengolahan kemiripan pilihan siswa berupa rating dengan *item-based collaborative filtering* memiliki MAE sebesar 0.2%.

Kata kunci : sistem rekomendasi, *naive bayes* , *item- based collaborative filtering*, *hybrid recommendation*

### Abstract

*Higher education is a place to produce superior human resources and think, no wonder many universities are asked to produce students who are superior and able to compete in the world of work. However, the fact is that as many as 45% of students feel wrong in choosing the right study program, and as many as 92% of high school students or equivalent feel confused to determine their future. Therefore it is necessary to have a system to recommend study programs, to reduce the phenomenon. The method used to provide study program recommendations is a hybrid recommendation. This hybrid recommendation method was chosen because it can study user profiles, by developing the naive bayes method to classify study programs by processing value data, while to consider student interests using item-based collaborative filtering by calculating the similarity of study program choices provided by students. Combining these two methods will produce a program that is by the values and interests of students to choose study programs. The results of this study indicate that data processing with naive bayes produces an accuracy of 88.7%, while processing the similarity of student choices in the form of rating with item-based collaborative filtering has an MAE of 0.2%.*

*Keyword : recommendation system, naive bayes , item- based collaborative filtering, hybrid recommendation*

## 1. Pendahuluan Latar Belakang

Perguruan Tinggi adalah sebuah wadah pendidikan untuk menghasilkan SDM yang unggul dan berfikir sehingga dituntut untuk menghasilkan mahasiswa yang unggul dan mampu bersaing didunia kerja. Maka dari itu, salah satu perusahaan yakni youthmanual melakukan penelitian selama dua tahun untuk mendalami lebih dari 400.000 profil dan data siswa dan mahasiswa di seluruh Indonesia. Dari hasil penelitian tersebut, ditemukan fakta yang cukup menarik yakni 92% siswa SMA/SMK sederajat merasa bingung dengan masa depannya dan 45% mahasiswa merasa salah mengambil program studi [1]. Hasil dari penelitian tersebut diketahui bahwa mahasiswa sering melakukan pindah program studi, pindah universitas atau bahkan *drop out* karena ketidakmampuan untuk menyelesaikan pendidikannya di Perguruan Tinggi.

Menurut Data yang dimuat pada tempo.co, CEO aplikasi Aku Pintar menyatakan bahwa 87% mahasiswa Indonesia salah mengambil program studi, Selain itu faktor utama dari kesalahan siswa dan mahasiswa kesulitan dalam mengambil program studi adalah karena mereka tidak mengenali program studi yang akan dipilih [2]. Menurut hasil dari laporan buku tahunan Kemenrisetdikti 2017 dapat dikatakan bahwa sebanagan 6.024.324 mahasiswa yang mengalami salah program studi memiliki dampak seperti konflik batin, despresi hingga pindah program studi dan 195.176 mengalami *drop-out* [4].

Dengan adanya fakta tersebut, dibutuhkan suatu sistem untuk merekomendasikan program studi guna mempertimbangkan program studi yang akan diambil. Metode untuk membangun sistem rekomendasi ini adalah *hybrid recommendation system*, penelitian ini didukung dengan penelitian sebelumnya untuk dilakukan pertimbangan terkait metode yang akan digunakan. Dalam penelitian Arif Kurniawan berjudul Sistem Rekomendasi Produk Sepatu Dengan Menggunakan Metode *Collaborative Filtering* memiliki hasil akurasi sebesar 95,86% dengan menggunakan *item-based collaborative filtering* [5]. Penelitian dari Eri Eli Lavindi berjudul Aplikasi *Hybrid* dan *Naive bayes* untuk Sistem Rekomendasi Pembelian Laptop memiliki akurasi sebanyak 80 % dengan melakukan pengolahan data laptop yang diklasifikasi dalam *collaborative filtering* dan *content based filtering* [6]. Penelitian selanjutnya dilakukan oleh Riri Intan berjudul Sistem Rekomendasi Bacaan Tugas Akhir Jurusan Teknik Informatika Universitas Sriwijaya menggunakan metode *Collaborative Filtering* dan *Naive bayes* memiliki akurasi 81% dan MAE sebanyak 2.01 [7].

Dengan adanya penelitian tersebut, maka sistem rekomendasi yang akan dibangun menggunakan metode *content based filtering* dan *collaborative filtering*. *Content based filtering* menggunakan *naive bayes* untuk mengidentifikasi program studi dengan mengolah nilai, sedangkan *collaborative filtering* menggunakan *item-based collaborative filtering* untuk mengolah program studi pilihan siswa. Penelitian ini akan melakukan analisis *accuracy* untuk *naive bayes* dan MAE untuk *item-based collaborative filtering*.

### Topik dan Batasannya

Sistem rekomendasi sendiri merupakan alat dan teknik perangkat lunak yang menyediakan saran untuk *item* untuk pengguna [10]. Dalam penelitian ini, sistem rekomendasi digunakan untuk memberikan rekomendasi program studi yang akan dipertimbangkan siswa berdasarkan data yang diberikan saat mendaftarkan diri ke Universitas Telkom. Pertimbangan menggunakan metode ini adalah atribut yang diberikan oleh Admisi Seleksi Mahasiswa Baru (SMB) berupa nilai dan pilihan program studi.

Universitas Telkom sebagai *study case*, adapun data yang akan diolah adalah siswa yang telah diseleksi oleh SMB melalui Jalur Prestasi Akademik (JPA) tahun ajaran 2018-2019. Dari 36 program studi yang dimiliki oleh Universitas Telkom diambil 20 program studi, karena 16 program studi lainnya merupakan kelas internasional dan kelas ekstensi dari diploma ke sarjana. Penelitian ini berfokus pada program studi dan tidak terkait jenjang, data yang digunakan adalah nilai Fisika, Matematika, Bahasa Indonesia dan Bahasa Inggris dan data pilihan program studi 1-5 yang dipilih oleh siswa. Pengolahan data nilai menggunakan *naive bayes* dengan mempertimbangkan variasi data nilai dan penyeleksian program studi, sedangkan untuk *item-based collaborative filtering* dalam pengolah rating dengan mempertimbangkan minat siswa pada program studi tersebut. Rumusan masalah yang akan dilakukan adalah analisis bagaimana sistem rekomendasi dibuat dengan menggunakan *hybrid recommendation* dan bagaimana mengukur tingkat performansi sistem rekomendasi dengan menggunakan *hybrid recommendation*.

### Tujuan

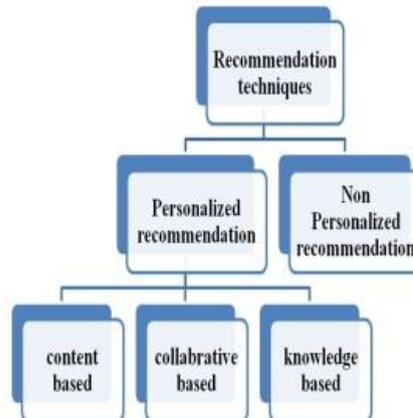
Penelitian ini bertujuan untuk membangun sistem rekomendasi pemilihan program studi untuk siswa/i SMA sederajat menggunakan rekomendasi hibrid dengan mengembangkan *content-based filtering* dan *collaborative filtering*. Pendekatan yang digunakan dalam *content-based filtering* adalah *naive bayes* dan *collaborative filtering* adalah *item-based collaborative filtering*. Pertama, data yang diperoleh berupa nilai fisika, matematika, Bahasa Indonesia dan Bahasa Inggris akan diolah dengan *naive bayes* sehingga mendapatkan program studi dengan probabilitas tertinggi. Kemudian metode *item-based collaborative filtering* dibangun dan diterapkan pada sistem rekomendasi dengan menghitung rating program studi alternatif yang akan diberikan kepada siswa sesuai dengan rating yang diberikan sebelumnya. Selain itu, penelitian ini juga bertujuan untuk mengetahui akurasi yang dihasilkan oleh sistem rekomendasi yang telah dibangun dengan rekomendasi hibrid dengan menghitung akurasi *naive bayes* dan *mean absolute error* dari *item-based collaborative filtering*.

### Organisasi Tulisan

Penulisan jurnal penelitian ini terdiri dari lima bab. Pada bab 1, berisi tentang latar belakang penelitian ini dibuat, topik dan batasan yang terkandung di dalamnya, serta tujuan dilakukannya penelitian. Pada bab 2, mengandung teori yang menunjang dilakukannya penelitian ini. Bab 3 merupakan implementasi sistem yang dibangun pada penelitian ini. Selanjutnya, bab 4 berisi analisis hasil dari sistem yang dibangun. Terakhir, pada bab 5 mengandung kesimpulan dari semua kegiatan yang dilakukan pada penelitian.

2. Studi Terkait

1. Teknik Rekomendasi adalah teknik untuk mengolah informasi dari sekumpulan pengguna untuk memprediksi *item* dan memberikan rekomendasi *item* yang terbaik kepada pengguna [11]. Tujuan dari sistem rekomendasi adalah untuk menghasilkan rekomendasi yang bermakna ke pengguna untuk barang atau produk yang mungkin menarik bagi mereka [12].



Gambar 1 Teknik Rekomendasi Sistem [11].

2. *Collaborative filtering* memberikan informasi berdasarkan kelompok pengguna yang memiliki kesamaan, perbedaan minat pada beberapa anggota kelompok menjadikan sumber informasi baru yang mungkin bermanfaat bagi anggota kelompok lainnya [13]. *Collaborative filtering* melakukan pengolahan data dengan rating dari user yang digunakan dalam sistem rekomendasi, dibedakan menjadi dua cara yaitu:

- 1) Secara Eksplisit, yaitu proses pengumpulan data dimana user memberikan data secara sadar/ sengaja.
- 2) Secara Implisit, yaitu proses pengumpulan data dimana user tidak menyadari bahwa ia telah memberikan masukan terhadap sistem

3. Pendekatan *user based collaborative filtering* menggunakan *nearest neighbor algorithm* yaitu teknik statistika untuk menemukan sekumpulan pengguna yang dikenal sebagai tetangga, tetangga ini harus mempunyai sejarah sama. Setelah sekumpulan tetangga terbentuk sistem menggunakan algoritma yang berbeda untuk menggabungkan kesukaan *neighbours* untuk menghasilkan predikis atau rekomendasi *N-teratas* untuk *active user* [14].

4. Pendekatan *item based collaborative filtering* memberikan rekomendasi berdasarkan kemiripan antar *item*. Metode ini merupakan metode rekomendasi yang didasari atas adanya kesamaan antara pemberian rating terhadap suatu *item* dengan *item* yang pernah dirating user lain, *item* yang telah di rating oleh user akan menjadi patokan untuk mencari sejumlah *item* lainnya yang berkorelasi dengan *item* yang telah dirating user [10] [15]. Implementasi *item based collaborative filtering* dilakukan dalam dua tahap sebagai berikut :

- o Melakukan pemrosesan data rating dari suatu *item* untuk mendapatkan data pengguna yang telah merating *item*.
- o Menghitung *similarity* antara *item* satu dengan lainnya menggunakan metode *cosine similarity* berdasarkan rating yang diberikan pengguna.

1. Algoritma *cosine-based similarity* Pada kasus ini dua *item* dianggap sebagai 2 vektor. Kesamaan antara 2 *item* ini diukur dengan menghitung kosinus dari sudut antara 2 vektor *item*. *Item* dibandingkan misalnya u dan v, dianggap sebagai sebuah vektor baris dengan anggotanya adalah nilai rating yang diberikan terhadap kedua *item* tersebut. Persamaan dari *cosine-based similarity* dapat dilihat pada persamaan (1).

$$similarity(u, v) = \frac{\sum_{i \in I} r_{ui} r_{vi}}{\sqrt{\sum_{i \in I} r_{ui}^2} \sqrt{\sum_{i \in I} r_{vi}^2}} \quad (1)$$

dimana:

$$similarity(u, v) = \frac{\sum_{i \in I} r_{ui} r_{vi}}{\sqrt{\sum_{i \in I} r_{ui}^2} \sqrt{\sum_{i \in I} r_{vi}^2}}$$

$$u \in U = \{r_{u1}, r_{u2}, \dots, r_{un}\}$$

$$v \in U = \{r_{v1}, r_{v2}, \dots, r_{vn}\}$$

$$I = \{1, 2, \dots, n\}$$

$$r_{ij} = \frac{r_{ij} + r_{ij}}{2}$$

2. Algoritma *weighted sum* digunakan untuk menghitung prediksi rating, menghitung rating dilakukan dengan cara membandingkan rating yang pernah diberikan pengguna pada suatu *item* dengan kemiripan *item* satu dengan *item* lainnya [15]. Persamaan algoritma *weighted sum* dapat dilihat dari **persamaan (2)**.

$$r_{ij} = \frac{r_{ij} \times r_{ij}}{r_{ij}} \quad (2)$$

Keterangan :

$$r_{ij} = \text{Rating pengguna } i \text{ pada item } j$$

$$r_{ij} = \text{Rating pengguna } i \text{ pada item } j$$

$$r_{ij} = \text{Rating pengguna } i \text{ pada item } j$$

5. *Content-based Filtering* adalah informasi konten untuk merekomendasikan *item* serupa menggunakan data yang independent dan tidak memiliki history. *Content-based Filtering* ini berfokus pada *probabilistic* seperti *naive bayes* dan klasifikasi probabilistik lainnya [16]. Teknik – teknik yang digunakan dalam content-based seperti TF-IDF, Bayesian Classifiers, Cluster analysis, decision trees dan artificial neural networks [12].

1. *Naive Bayes* merupakan penganalisis probabilistik sederhana yang menghitung satu set probabilitas dengan menjumlahkan kombinasi frekuensi dan nilai dari dataset yang diberikan. Algoritma ini menggunakan teorema *Bayes* dan mengasumsikan semua atribut independen atau non-dependen diberikan oleh nilai dalam variabel kelas [17]. Beberapa tipe yang dapat digunakan dengan *naive bayes* adalah :

- *Multinomial*, metode ini sebagian besar digunakan untuk masalah klasifikasi dokumen, yaitu apakah dokumen termasuk dalam kategori. Fitur / prediktor yang digunakan oleh pengklasifikasi adalah frekuensi dari kata-kata yang ada dalam dokumen.
- *Bernoulli*, metode ini mirip dengan multinomial naive bayes tetapi prediktornya adalah variabel boolean. Parameter yang kita gunakan untuk memprediksi variabel kelas hanya mengambil nilai ya atau tidak, misalnya jika ada kata dalam teks atau tidak
- *Gaussian*, metode ini ketika prediktor mengambil nilai kontinu dan tidak diskrit, dapat dianggap bahwa nilai-nilai ini diambil sampelnya dari distribusi gaussian. Tahapan dalam melakukan algoritma naive bayes adalah sebagai berikut [17]:

- a. Menghitung Prior tiap Kelas, pada **Persamaan 3:**

$$P(C_j) = \frac{\text{Jumlah Kemunculan Data Kelas } C_j}{\text{Total Data Keseluruhan}} \quad (3)$$

Keterangan :

P(Cj) = Nilai Prior untuk Kelas C

- b. Hitung Likelihood Tiap Kelas Data Kontinu, pada **Persamaan 4:**

$$P(X_i = x_i | C_j = c_j) = \frac{1}{\sigma_{ij} \sqrt{2\pi}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (4)$$

Keterangan:

P = Peluang

Xi = Atribut ke i

xi = Nilai atribut ke i

Cj = Kelas yang dicari

cj = Subkelas yang dicari

$\pi = \frac{22}{7}$  atau 3.14

e = nilai euler 2.71

$\mu$  = nilai MEAN

$\sigma$  = Nilai Standar Deviasi

- c. Menghitung MEAN, pada **Persamaan 5:**

$$\mu = \frac{\sum_{i=1}^n X_i}{n}$$

(5)

d. Menghitung Standar Deviasi, pada **Persamaan 6:**

$$\sigma = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1}} \quad (6)$$

Keterangan :

$\sum_{i=1}^n X_i$  = Sigma Penjumlahan data x dengan indeks 1 sampai n

n = jumlah data

e. Perhitungan Posterior, pada **Persamaan 7:**  
 $P(C|X_1, X_2, X_3, \dots, X_n) = P(C)\prod_{i=1}^n P(X_i|C) \quad (7)$

Keterangan:

$P(C|X_1, X_2, X_3, \dots, X_n)$  = nilai posterior data dengan variable

$X_1$  hingga  $X_n$  untuk kelas C

$P(C)$  = nilai prior untuk kelas C

$\prod_{i=1}^n P(X_i|C)$  = sigma perkalian untuk Likelihood kondisi X dengan indeks 1 hingga n untuk kelas C

f. Kelas dengan nilai Posterior tertinggi aka menjadi hasil klasifikasi dari data uji pada **Persamaan 8:**  
 $P(C|X_1, X_2, X_3, \dots, X_n) = \arg \max P(C)\prod_{i=1}^n P(X_i|C) \quad (8)$

Keterangan :

arg max = argument dengan nilai tertinggi untuk hasil posterior tiap kelas

$P(C)\prod_{i=1}^n P(X_i|C)$  = proses perhitungan posterior untuk kelas C

6. *Hybrid recommender* adalah penggabungan dua atau lebih metode secara bersamaan untuk menciptakan hasil yang diinginkan [18]. Dengan menggabungkan beberapa metode pendekatan ini dirasa lebih efektif untuk mendapatkan hasil yang sesuai [11]. *Hybrid recommendation System* mempunyai beberapa Teknik untuk melakukan penggabungan metode dalam memberikan rekomendasi [15]:

1. *Weighted* : bobot dari dua atau lebih metode rekomendasi digabungkan secara numerik
2. *Switching* : sistem memilih salah satu atau lebih metode rekomendasi dan menerapkan salah satu metode rekomendasi yang dipilih.
3. *Mixed* : teknik rekomendasi dari berbagai metode ditampilkan menampilkannya secara bersamaan.
4. *Feature Combination* : output dari salah satu teknik rekomendasi digunakan sebagai input yang lain.
5. *Feature Augmentatio* : fitur-fitur dari sumber data rekomendasi yang berbeda digabung bersama-sama ke dalam algoritma rekomendasi tunggal.
6. *Cascade* : satu rekomen der mengolah rekomendasi yang diberikan oleh yang lainnya.
7. *Meta-level* : model dipelajari oleh satu rekomen der yang digunakan sebagai inputan yang lainnya.

7. Performansi Akurasi klasifikasi mengukur kualitas kedekatan dengan kebenaran atau nilai sebenarnya yang dicapai oleh suatu sistem [19]. Akurasi adalah metrik yang paling banyak digunakan dan terkenal dilapangan adalah kecerdasan buatan dan secara umum ditulis pada **persamaan 9** [19]. Di bawah ini, persamaan untuk akurasi:

$$Accuracy = \frac{\sum_{i=1}^n \text{Correct Predictions}}{\sum_{i=1}^n \text{Total Predictions}} \quad (9)$$

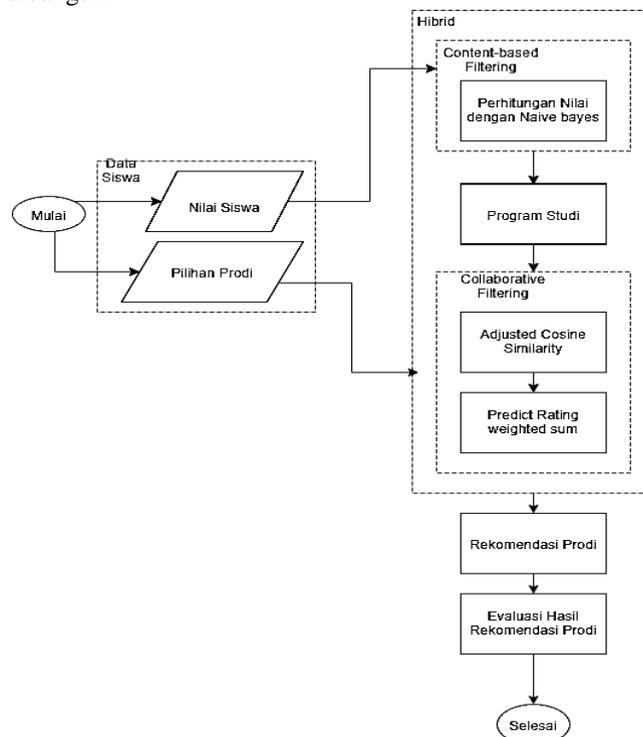
8. *Metrics Absolute Error* atau MAE adalah pengukuran kualitas dari pengevaluasian untuk membandingkan dengan hasil rekomendasi yang diberikan oleh sistem dan pengguna sebelumnya [12]. Evaluasi pada hasil sistem rekomendasi dilakukan dengan melihat rata-rata nilai eror yang dihasilkan pada setiap user. Persamaan *MAE* digunakan untuk menghitung nilai rata-rata selisih antara nilai prediksi dengan nilai yang sebenarnya [7], perhitungan MAE dapat di lihat pada **persamaan 10:**

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (10)$$

$$\begin{aligned} MAE &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ &= \frac{1}{n} \sum_{i=1}^n |h(x_i) - y_i| \\ &= \frac{1}{n} \sum_{i=1}^n |h(x_i) - y_i| \\ &= \frac{1}{n} \sum_{i=1}^n |h(x_i) - y_i| \end{aligned}$$

### 3. Sistem yang Dibangun

Sistem yang akan dibangun adalah sebuah sistem rekomendasi yang mampu memberikan rekomendasi berdasarkan nilai dan pilihan yang diminati. Dalam sistem ini terdapat 4 bagian besar yaitu mengolah data siswa, melakukan *hybrid recommendation*, menghasilkan rekomendasi dan evaluasi. Berikut adalah gambaran umum dari sistem yang dibangun:



**Gambar 2** Alur Perancangan Sistem Rekomendasi

Berikut adalah penjelasan terkait **Gambar 2**.

#### 3.1 Data Siswa

Dataset yang digunakan berupa data peserta yang lolos Seleksi Mahasiswa Baru (SMB), yang diselenggarakan oleh Universitas Telkom melalui Jalur Prestasi Akademik (JPA) untuk tahun ajaran 2018-2019. Dataset terdiri dari 13 kolom (terdiri dari kolom nomor, UserID, Nilai Bahasa Indonesia, Nilai Bahasa Inggris, Nilai Matematika, Nilai Fisika, Program Studi Pilihan 1, Program Studi Pilihan 2, Program Studi Pilihan 3, Program Studi Pilihan 4, Program Studi Pilihan 5, Keterangan Lulus). Dengan data 7 Fakultas terdiri atas 3 fakultas teknik dan 4 fakultas non teknik, 20 program studi dapat dilihat pada **tabel 2 dilampiran** dan program studi ekstensi dapat dilihat pada **lampiran tabel 1**. Data yang digunakan dalam penelitian ini sebanyak 1090 data, data terdiri dari nilai dan pilihan prodi yang siswa inginkan, dalam pemilihan program studi siswa harus mengetahui aturan yang berlaku di Telkom University yang berada pada lampiran **tabel 2**. Adapun aturan-aturan ini berlaku atas dasar peraturan Admisi Seleksi Mahasiswa Baru (SMB) Telkom University [20] tersebut diantaranya:

1. Setiap program studi dikelompokkan berdasarkan kategori kejuruan pada jenjang SMA, yakni IPA dan IPS, sehingga kategori program studi menjadi Teknik-Non Teknik.
2. Siswa/i dari kejuruan IPA dapat memilih program studi kategori Teknik dan Non-teknik, sedangkan siswa/i jurusan IPS hanya dapat memilih program studi kategori Non Teknik saja. Aturan ini diperoleh dari aturan yang dibuat oleh BAA Universitas Telkom [20]. Dengan demikian penulis menerapkan aturan yang sama yang dibuat oleh pihak Universitas Telkom.
3. Siswa/i yang memilih salah satu program studi kategori Program Studi Teknik wajib menginputkan 4 buah nilai rata-rata mata pelajaran, yakni Bahasa Indonesia, Bahasa Inggris, Matematika, dan Fisika. Sedangkan, program studi kategori Program Studi Non-teknik wajib menginputkan tiga buah nilai mata pelajaran yakni Bahasa Indonesia, Bahasa Inggris, dan Matematika. Mata pelajaran Fisika adalah opsional., Selanjutnya data nilai akan diolah pada algoritma *naive bayes*. Data yang diberikan oleh SMB Universitas Telkom dapat dilihat sebagai berikut :



**3.2 Perhitungan Nilai dengan Naive bayes**

Naive bayes akan dijalankan terlebih dahulu dengan melibatkan inputan nilai tiap mata pelajaran. Dimana data yang digunakan berdistribusi normal dengan perhitungan gaussian. Content-based filtering untuk membuat hybrid recommendation menggunakan naive bayes, naive bayes sering digunakan dalam content-based filtering yang mampu mengidentifikasi data sesuai dengan probabilitasnya [13].

**Tabel 7 Dataset yang Digunakan**

userID	Bhs. Indonesia	Bhs. Inggris	Matematika	Fisika	Lulus Program studi
11001779	77,5	81,75	77,25	76,75	Teknologi Informasi
11001780	93	81,5	86	80,25	Teknologi Informasi
11001781	87,75	87,75	87,75	86,5	Teknologi Informasi
11001782	79,75	73,75	75,75	78	Teknologi Informasi
11001783	87	84	85,2	86,4	Teknologi Informasi
11000861	88,75	86	90,75	85,75	MBTI
11000862	93	92,5	89	85,5	MBTI
11000863	86	79,25	79,25	77,75	MBTI
11000865	85,75	79,5	81,25	79,5	MBTI
11000866	84,75	86,75	82,5	79,5	MBTI

**1. Menghitung Prior**

Setelah dataset dipisahkan dari rating, langkah berikutnya ialah menghitung peluang prior setiap kelas sesuai dengan persamaan (3). Perhitungan naive bayes maka hasil yang didapatkan seperti berikut :

$$P(\text{Kelas}) = \frac{5}{10} = 0.5$$

Setelah mendapatkan prior pada kelas Teknologi Informasi, selanjutnya mencari nilai likelihood pada persamaan 4 dengan menghitung rata-rata dan standar deviasinya. Untuk menghitung mean menggunakan persamaan 5 dan untuk standar deviasi menggunakan persamaan 6, berikut adalah contoh perhitungan mean dan standart deviasi pada seluruh nilai, pada perhitungan dibawah ini menggunakan nilai Bahasa Indonesia pada baris pertama.

**2. Menghitung MEAN**

Perhitungan ini dilakukan untuk mendapatkan rata-rata dari setiap inputan nilai, dengan menggunakan rumus gaussian. Data yang digunakan merupakan distrinusi normal dan mempunyai banyak variasi pecahan dengan desimal.

$$\mu(\text{Bahasa Indonesia}) = \frac{77.5 + 93 + 87.75 + 79.75 + 87}{5} = 8.5$$

**3. Menghitung Standar Deviasi**

Perhitungan ini dilakukan untuk mendapatkan standar deviasi untuk setiap inputan nilai.

$$\sigma(\text{Bahasa Indonesia}) = \sqrt{\frac{(77.5 - 8.5)^2 + (93 - 8.5)^2 + (87.75 - 8.5)^2 + (79.75 - 8.5)^2 + (87 - 8.5)^2}{5}} = 6.312$$

**4. Menghitung likelihood**

Perhitungan ini dilakukan untuk mendapatkan peluang nilai dalam kelas tersebut.

$$P(77.5 | \text{Bahasa Indonesia}) = \frac{1}{\sqrt{2 \times 3.14 \times 6.312^2}} \times \frac{(77.5 - 8.5)^2}{2 \times 6.312^2} = 0.0593$$

Selanjutnya perhitungan *posterior* dengan hasil dari perkalian *likelihood* dan *prior*, untuk menghitung *posterior* dapat dilihat pada **persamaan 7**, berikut adalah contoh perhitungan *posterior* pada siswa a pada jurusan teknologi informasi:

**Tabel 8 Tabel Likelihood**

Bahasa Indonesia	Bahasa Inggris	Matematika	Fisika
0,059342234	0,070162437	0,037125557	0,057647811

$$P(A) = 0.0593 \times 0.0701 \times 0.0371 \times 0.0576 \times 0.5 = 4.455 \times 10^{-6}$$

Setelah dilakukan perhitungan *posterior* hasil yang didapatkan adalah  $4.455 \times 10^{-6}$ , hasil dari *posterior* Siswa A pada prodi teknologi informasi akan dibandingkan dengan *posterior* MBTI. Hasil perbandingan antar *posterior* yang terbesar akan menjadi prodi lulusan dari perhitungan *naive bayes*. Setelah mendapatkan program studi dari perhitungan *naive bayes*, maka rating akan diolah dengan algoritma *item-based collaborative filtering* sesuai dengan kelompok program studi hasil *naive bayes*.

**3.3 Perhitungan Item-based collaborative filtering**

Data berupa rating yang dipilih oleh siswa untuk keterangan pilihan siswa 1-5 dengan nilai rating bisa dilihat pada **tabel 5**. *Collaborative filtering* sendiri merupakan pengidentifikasian kelompok yang sama dan merekomendasikan *item* yang sama. Sehingga menggunakan perhitungan *item-based collaborative filtering* ini untuk menghitung kesamaan rating yang diberikan siswa dengan prodi yang sama, prodi yang dihasilkan oleh *naive bayes*. Berikut merupakan contoh dari rating siswa A yang akan di bandingkan dengan siswa B dilihat pada **tabel 9**. Pada data ini Siswa A dan B lulus di program studi teknologi informasi.

**Tabel 9 Rating Siswa pada Prodi**

Kode Prodi Siswa	kode prodi							
	0	1	2	3	4	5	6	7
Siswa A	4		5					
Siswa B	3	1	5			2		4

Tahap ini adalah mencari nilai kemiripan antar rating prodi yang dibandingkan dengan *adjusted cosine similarity*. Perbedaan *adjusted cosine similarity* dengan *cosine similarity* adalah rata-rata rating, untuk *cosine similarity* tidak memakai rata-rata rating. Berikut perhitungan mencari rata-rata rating :

1. Menghitung rata-rata rating

$$\bar{r}_i = \frac{5 + 4}{2} = 4.5$$

$$\bar{r}_A = \frac{4 + 5}{2} = 4.5$$

$$\bar{r}_B = \frac{3 + 1 + 5 + 2 + 4}{5} = 3$$

**Tabel 10 Rata-rata Rating Siswa**

Kode Prodi Siswa	kode prodi								rata-rata rating
	0	1	2	3	4	5	6	7	
Siswa A	4	0	5	0	0	0	0	0	4,5
Siswa B	3	1	5	0	0	2	0	4	3

Setelah mendapatkan rata-rata dari rating setiap siswa dilanjutkan dengan menghitung *adjusted cosine similarity*, dengan **persamaan 1**.

2. Menghitung *adjusted cosine similarity*

Setelah mendapatkan rata-rata maka dihitung similarity dengan **persamaan 1**. Dengan membandingkan kode prodi 0 yang bernilai 4 dan prodi 2 bernilai 5 pada siswa A, pada siswa B kode prodi 0 bernilai 3 dan kode prodi 2 bernilai 5.

**Tabel 11 Perbandingan Rating Prodi 0 dan 1**

user	$r_{i,0}$	$r_{i,2}$	$\bar{r}_i$
Siswa A	4	5	4.5
Siswa B	3	5	3

$$W(0,2) = \frac{(4 - 4.5)(5 - 4.5) + (3 - 3)(5 - 3)}{\sqrt{((-0.5)^2 + (0)^2)} \sqrt{(0.5)^2 + (2)^2}} = \frac{-0.25}{\sqrt{0.25} \sqrt{4.25}} = -0.287$$

$$\sqrt{(4 - 4.5)^2 + (3 - 3)^2} \sqrt{(5 - 4.5)^2 + (5 - 3)^2} = \sqrt{0.25} \sqrt{4.25} = 1.03$$

Setelah melakukan perhitungan dengan menggunakan persamaan *adjusted cosine similarity* untuk mencari nilai kemiripan antara prodi 0 dan 2 maka didapat hasil kemiripan dengan nilai -0.287. Setelah nilai kemiripan didapat maka tahap selanjutnya perhitungan pencarian similarity setiap prodi yang dirating siswa A dan B. Sehingga semua kemiripan prodi selesai dapat dilihat pada **tabel 12**.

**Tabel 12 Kesamaan rating prodi yang dibandingkan**

Prodi yang dibandingkan	Prodi yang dibandingkan	Nilai Kemiripan
0	1	0.70711
0	3	-0.82
0	4	0
0	5	0
0	6	1
0	7	-0.5477
1	2	-1
1	3	-1
1	4	0
1	5	1
1	6	0
1	7	-0.948
2	3	1
2	4	0
2	5	-1
2	6	-1
2	7	0.3229
3	4	0
3	5	-1
3	6	1
3	7	-0.0521
4	5	0
4	6	0
4	7	0
5	6	1
5	7	-1
6	7	1

Setelah mendapatkan semua hasil kemiripan rating, dimana semua rating dalam program studi dihitung. Selanjutnya dibuat matriks kemiripan antar rating prodi seperti pada **tabel 13** berikut:

**Table 13 Matrix Item Similarity**

Kode Prodi	0	1	2	3	4	5	6	7
0	1	0,707106781	-0,28735	-0,82		0,82	1	-0,547659175
1	0,707106781	1	-1	-1		0	0	-0,948683298
2	-0,287347886	-1	1	1		1	-1	0,322900628
3	-0,82	-1	1	1		-1	-1	-0,052103046
4	0	0	0	0	1	0	0	0
5	0	1	-1	-1		1	1	-1
6	1	0	-1	-1		1	1	1
7	-0,547659175	-0,948683298	0,322901	0,0521		-1	1	1

Pada tabel matriks kemiripan item pada *adjusted cosine similarity* mempunyai range dari 1,0 dan -1.

- Menghitung prediksi rating dengan *weighted sum* menggunakan **persamaan 2** pada siswa A untuk memprediksi rating yang kosong pada **tabel 9** berikut, pada kasus ini hanya prodi 0 dan 2 yang memiliki nilai, sehingga untuk menghitung prediksi rating yang kosong sebagai berikut:

$$P_{(0,1)} = \frac{(4 \times 0.7) + (5 \times -1) + (0 \times -1) + (0 \times 0) + (0 \times 0) + (0 \times 0.9)}{(|0.7| + |-1| + |-1| + |0| + |0| + |0.9|)} = -0.6$$

Prediksi rating pada prodi 1 adalah -0,6 artinya rating yang diberikan kurang dari 0 sehingga tidak dapat direkomendasikan, dimana Perhitungan terus berlanjut hingga semua rating prodi yang kosong pada siswa A mempunyai nilai seperti **tabel 14**.

**Table 14 Hasil Prediksi**

user	kode prodi							
	0	1	2	3	4	5	6	7
Siswa A	4	-0.6	5	5	0	0	4	5

Hasil rekomendasi dari *collaborative filtering* adalah rating yang tertinggi, sehingga rating awal yang lebih rendah akan disisihkan dan digantikan oleh rating yang lebih besar.

**3.4 Rekomendasi Program Studi Alternatif**

Hasil perhitungan *weighted sum* pada **tabel 13**. Dengan melakukan perhitungan *naive bayes* dan *item-based collaborative filtering* dengan rekomendasi hibrid sehingga bisa disimpulkan bahwa alternatif yang akan direkomendasikan kepada siswa/i ialah prodi **2, 3 dan 7** dimana representasi dari kode prodi tersebut adalah **Teknik Informatika, Teknik Industri dan Sistem Informasi**.

**3.5 Perhitungan Evaluasi**

**3.5.1 Akurasi Naive bayes**

Evaluasi akurasi hasil perhitungan *naive bayes* dilihat dari berapa banyak hasil dari data *naive bayes* yang relevan dengan hasil didataset. Berikut adalah contoh perhitungan akurasi menggunakan **persamaan 9**, hasil 10 dataset dapat dilihat pada **tabel 19**.

**Tabel 11 Contoh Akurasi Naive Bayes**

Hasil <i>naive bayes</i>	Kelas dataset	True / False
0	0	T
8	0	F
0	0	T
0	0	T
8	0	F
8	8	T
8	8	T
0	8	F
8	8	T
8	8	T

$$\frac{7}{10} \times 100\% = 70\%$$

Maka hasil akurasi dari perhitungan *naive bayes* adalah 70% dengan menggunakan contoh dataset diatas. Selanjutnya menghitung akurasi MAE pada *item-based collaborative filtering*.

**3.5.2 Mean Absolute Error**

Pada perhitungan MAE digunakan untuk melihat apakah rating yang diberikan oleh *item-based collaborative filtering* sudah sesuai dengan rating yang diberikan oleh siswa, selisih antara rating sebelumnya dengan rating yang diberikan oleh *item-based collaborative filtering* dikatakan baik jika rata-rata MAEnya mendekati 0, *range* 0 sampai  $\infty$ . Perhitungan MAE menggunakan **persamaan 10** dengan menggunakan data siswa A dan 20 kode program studi dimana kode 8-19 bernilai 0 atau tidak rating.

user	kode prodi							
	0	1	2	3	4	5	6	7
Siswa A	4	0	5	0	0	0	0	0
Siswa A	4	-0.6	5	5	0	0	4	5

$$\frac{|(4 - 4) + |0 - (-0.6)| + (5 - 5) + (0 - 5) + (0 - 0) + (0 - 0) + (0 - 4) + (0 - 5)|}{20} = 0.67$$

**4. Evaluasi**

Evaluasi yang dilakukan pada penelitian in adalah mengukur akurasi dan MAE dari sejumlah data latih dan data uji. Setelah dilakukan pengujian dengan melibatkan 763 baris data training dan 327 data uji dalam 20 program studi. Maka dibuat skenario sebagai berikut:

1. Melakukan analisis kesesuaian hasil program studi yang diberikan oleh *naive bayes* dengan lulusan dataset dengan akurasi.

Skenario ini bertujuan untuk mengetahui ketidakcocokan hasil rekomendasi yang diberikan oleh sistem untuk user berdasarkan inputan berupa nilai dan rating. Sebanyak 14,5% data uji yang ada menunjukkan ketidakcocokan antara jurusan yang sedang ditempuh dengan rekomendasi yang diberikan oleh *naive bayes*. Hal ini disebabkan karena probabilitas nilai siswa pada prodi tersebut sedikit, sehingga posterior yang dihasilkan juga kecil.

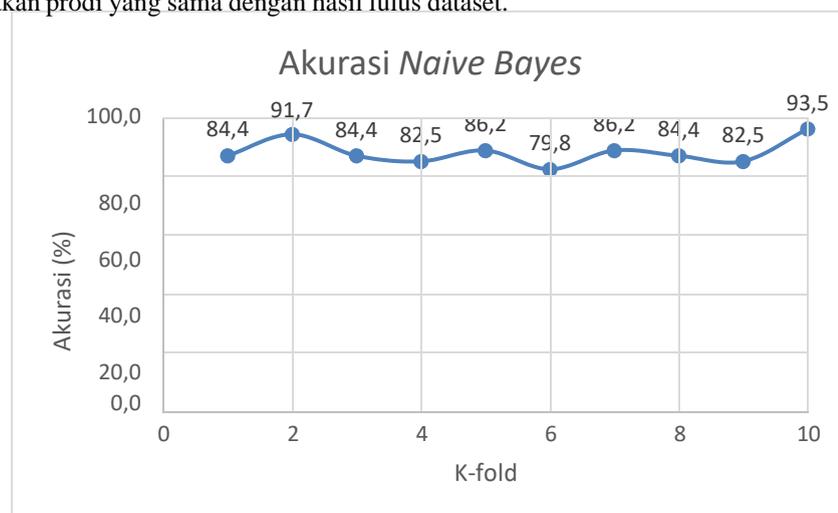
2. Melakukan analisis rata-rata error dari hasil *item-based collaborative filtering* dengan rating yang diberikan siswa dengan MAE.

Skenario ini bertujuan untuk mengetahui error dalam kemiripan rating dari hasil rekomendasi yang diberikan oleh sistem, data rating yang dilakukan dengan *item-based collaborative filtering* antara rating sebelumnya dengan rating baru.

#### 4.1 Analisis Hasil Pengujian

##### 4.1.1 Akurasi

Pada analisis hasil pengujian ini, terdapat data uji sebanyak 327 dari dataset nilai yang dimiliki dengan 20 program studi.. Hasil akurasi dari *naive bayes* yang program studi menggunakan *k-fold cross validation* dengan  $K=10$ . Dimana dapat dilihat pada **diagram 1** dibawah ini. Dari hasil pengujian tersebut memiliki rata-rata 85.5%, hasil ini disebabkan perhitungan *naive bayes* dengan hasil kelulusan siswa didataset relevan artinya hasil akhir posterior yang terbesar merupakan prodi yang sama dengan hasil lulus dataset.



**Diagram 1 Hasil Naive Bayes**

##### 4.1.2 MAE

Hasil perhitungan akurasi menggunakan *Mean Absolute Error* (MAE) terhadap metode *Collaborative Filtering* adalah 0.21945%. 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 cukup baik, range untuk MAE sendiri adalah 0 sampai  $\infty$  dimana semakin mendekati 0 maka MAE akan semakin baik. Berikut adalah hasil rata-rata MAE setiap siswa pada data testing bisa dilihat pada **diagram 2**.

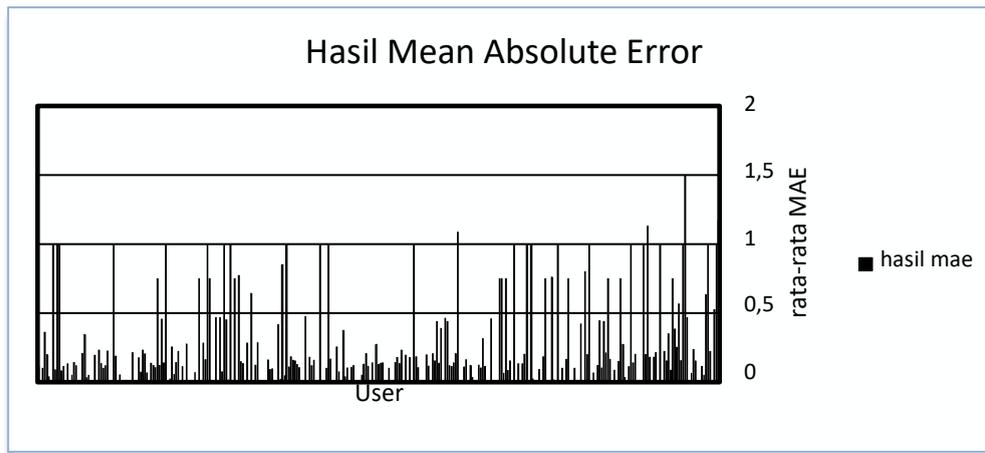


Diagram 2 Hasil MAE

Terdapat beberapa alasan terjadinya kesalahan dalam hasil rekomendasi program studi ini, diantaranya adalah sedikitnya jumlah dataset dan sistem menampilkan hasil yang tidak sesuai. Berikut adalah penjelasannya.

1. Sistem menampilkan hasil yang tidak sesuai

Pada penelitian ini, jumlah dataset yang digunakan adalah 327 tugas akhir dengan 20 program studi. Kelas pada program studi tersebut terbilang terlalu banyak untuk membangun sebuah sistem rekomendasi dengan menggunakan *naive bayes*. Dampaknya adalah 14,5% data tidak sesuai dengan hasil rekomendasi sistem.

2. Sistem menampilkan hasil rekomendasi yang kurang mirip.

Pada penelitian ini, jumlah rating dari siswa tidak seimbang sehingga kemiripan rating tersebut terbilang kurang mirip karena masih memiliki error sebesar 0,2914% yang diberikan oleh sistem rekomendasi. Dampak dari kesalahan penginputan *rating* yang tidak seimbang, yaitu terdapat beberapa program studi yang dihasilkan oleh rekomendasi memiliki nilai error atau kurang mirip dengan program studi yang diminati user.

## 5. Kesimpulan dan Saran

### 5.1 Kesimpulan

Berdasarkan hasil analisis pengujian, dapat disimpulkan bahwa pendekatan *Hybrid Recommendation* dengan metode *Content-based Filtering* dan *Collaborative filtering* mampu diterapkan dalam membangun sistem rekomendasi pemilihan program studi untuk menyeleksi beragam program studi. Dimana pada uji sistem pada *naive bayes* ditemukan bahwa akurasi yang diberikan *naive bayes* adalah 85.5%, Sedangkan untuk MAE didapatkan sebanyak 0.2194%.

### 5.2 Saran

Penelitian selanjutnya diharapkan

1. Menggunakan evaluasi kepuasan pengguna untuk melihat apakah hasil rekomendasi dari sistem sudah sesuai dengan nilai dan minat siswa.
2. Menggunakan metode klasifikasi lain, karena kekurangan dari *naive bayes* belum optimal untuk menyeleksi lebih dari 10 kelas.
3. Menambah data profil user untuk memberikan rekomendasi dengan banyak pertimbangan sesuai dengan yang user inginkan.

## Daftar Pustaka

- [1] Nike Putri, "Youthmanual: Angka Siswa yang Salah Pilih Jurusan Masih Tinggi," skystarventures, 14 04 2018. [Online]. Available: <http://www.skystarventures.com/youthmanual-angka-siswa-yang-salah-pilih-jurusan-masih-tinggi/>. [Accessed 17 03 2018].
- [2] Mardiana Makmun, "CEO Aku Pintar: 87 Persen Mahasiswa Merasa Salah Jurusan," tempo.co, 22 08 2017. [Online]. Available: <https://bisnis.tempo.co/read/1144950/ceo-aku-pintar-87-persen-mahasiswa-merasa-salah-jurusan/full&view=ok>. [Accessed 17 03 2018].

- [3] Ayu Tri Arum Murti, "RMOL JATENG," Republik Merdeka, 18 12 2018. [Online]. Available: <http://www.rmoljateng.com/read/2018/12/15/15017/Masih-Adanya-Fenomena-Salah-Jurusan-Di-Kalangan-Mahasiswa->. [Accessed 24 11 2019].
- [4] Buku-Statistik-Pendidikan-Tinggi-2017, Kemenrisetdikti, 2017..
- [5] A. Kurniawan, "Sistem Rekomendasi Produk Sepatu dengan Menggunakan Metode Collaborative Filtering," *Seminar Nasional Teknologi Informasi dan Komunikasi 2016*, 2016.
- [6] E. E. Lavindi, Wijanarto and A. Rohmani, "Aplikasi Hybrid Filtering dan Naive Bayes untuk Sistem Rekomendasi Pembelian Laptop," *Jurnal of Information System*, vol. 4, pp. 54-64, 2019.
- [7] R. I. Aprilia and M. Fachrurrozi, "Sistem Rekomendasi Bacaan Tugas Akhir Jurusan Teknik Informatika Universitas Sriwijaya menggunakan Metode Collaborative Filtering dan Naive Bayes," *Prosiding Annual Research Seminar 2016*, vol. 2 no. 1, 2016.
- [8] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez,, "Recommender systems survey," *Knowl. Syst.*, vol. 49, pp. 109-132, 2013.
- [9] P. VALDIVIEZO-DIAZ, F. ORTEGA, E. COBOS and R. LARA-CABRERA, "A Collaborative Filtering Approach Based on Naive bayes Classifier," *Digital Object Identifier 10.1109/ACCESS.2019.2933048*, vol. 7, p. 12, 2019.
- [10] F. Ricci, L. Rokach and B. Shapira, *Recommender Systems Handbook*, springer science, 2010.
- [11] S. Akshita and A. Smita., " Recommender system: review. International Journal of Computer Applications," p. 38–42, 2013.
- [12] P. Melville and V. Sindhvani., " Recommender systems. In Encyclopedia of Machine Learning," 2010.
- [13] Schafer, J.B., Frankowski, D.,Herlocker,J. dan, "Collaborative Filtering Recommender System," *Springer-Verlag*, 2007.
- [14] Sarwar, B. M., Karypis, G., Konstan, J. A., and Riedl. J, "Analysis Of Recommendation Algorithms For E-Commerce," *In Proceedings of the ACM EC'00 Conference*, pp. 158-167, 2000.
- [15] A. E. Wijaya and D. Alfian, "SISTEM REKOMENDASI LAPTOP MENGGUNAKAN COLLABORATIVE FILTERING DAN CONTENT-BASED FILTERING," *Jurnal Computech & Bisnis, Vol 12, No 1, Juni 2018, 11-27*.
- [16] Firmahsyah and . T. Gantini, "Penerapan Metode Content-Based Filtering Pada Sistem Rekomendasi Kegiatan Ekstrakurikuler," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. Volume 2 Nomor 3, Desember 2016.
- [17] R. Fitriawanti, I. Cholissodin and . R. Karti, "KLASIFIKASI DAN REKOMENDASI JURUSAN KULIAH BAGI PELAJAR SMA MENGGUNAKAN ALGORITME NAIVE BAYES -WP," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, pp. 1-9, 2018.
- [18] G. Geetha, M. Safa, C. Fancy, and D. Saranya, "hybrid approach using collaborative filtering and content based filtering for recommender system," *Journal of Physics: Conference Series*, p. 1000:012101, 2018.
- [19] F. H. del Olmo and . E. Gaudioso, "Evaluation of recommender systems: A new approach," *Expert Systems with Applications 35 (2008) 790–804*, pp. 1-15, 2008.
- [20] BAA, Writer, "Aturan Akademik Universitas Telkom,". [Performance]. Telkom University., 2014.
- [21] A. Saleh, "Implementasi metode klasifikasi na'ive bayes dalam memprediksi besarnya penggunaan listrik rumah tangga," *Creative Information Technology Journal*, vol. 2, p. 207–217, 01 2015.