

# Penerapan Algoritma *Content Based Filtering* dan *Frequent Pattern Growth* pada Sistem Rekomendasi Program Mahasiswa Wirausaha di Universitas Negeri Surabaya

Atikah Adawiyyah<sup>1</sup>, Salamun Rohman Nudin<sup>2</sup>

<sup>1,2</sup> Teknik Informatika, Jurusan Teknik Informatika, Fakultas Teknik

Universitas Negeri Surabaya

<sup>1</sup>[atikah.17051204079@mhs.unesa.ac.id](mailto:atikah.17051204079@mhs.unesa.ac.id)

<sup>2</sup>[salamunrohman@unesa.ac.id](mailto:salamunrohman@unesa.ac.id)

**Abstrak**— Sistem rekomendasi mulai sering digunakan seiring dengan perkembangan aktivitas dan media berbasis internet. Saat ini belum ada sistem untuk mendukung mahasiswa yang ingin mengikuti kegiatan Program Mahasiswa Wirausaha (PMW) dengan memberikan informasi gambaran judul PMW yang pernah diterima dengan merekomendasikan judul dan perkiraan dana yang sesuai dengan ide dan minat mahasiswa. Sehingga perlu dilakukan analisis terhadap kesesuaian ide dan minat mahasiswa dengan merekomendasikan judul dan perkiraan dana berdasarkan mahasiswa yang sudah lolos PMW sebelumnya. Proses penentuan dan analisis program mahasiswa wirausaha di Universitas Negeri Surabaya dapat dikerjakan dengan menyusun sistem asosiasi. Di dalam data mining, untuk menyusun sistem asosiasi dapat menggunakan algoritma *Content Based Filtering* dan algoritma *Frequent Pattern Growth (FP-Growth)*. Pencarian judul dan perkiraan dana menjadi lebih mudah dengan adanya sistem yang dibangun. Tujuan dari penelitian ini adalah untuk mengembangkan sistem rekomendasi program mahasiswa wirausaha di Universitas Negeri Surabaya untuk mahasiswa yang mengikuti kegiatan PMW. Hasil dari penelitian ini adalah rekomendasi *Content Based Filtering* dengan mengukur nilai *Cosine Similarity* terhadap vektorisasi dokumen lebih memudahkan mahasiswa untuk melakukan pencarian terhadap judul dan perkiraan dana yang di butuhkan. Dari hasil 10 pengujian terhadap data baru kemiripan tertinggi yaitu 47.83% dengan rata-rata 44.2%. Dan kemiripan terendah yaitu 23.43% dengan rata-rata 19.29%.

**Kata Kunci**— Sistem Rekomendasi, *Content Based Filtering*, *FP-Growth*, Wirausaha, PMW

## I. PENDAHULUAN

Sistem rekomendasi telah tersedia dan sudah banyak digunakan saat ini. Pelayanan yang menggunakan sistem rekomendasi diantaranya *Amazon*, *Netflix*, dan *YouTube*. Beberapa sistem rekomendasi mengadopsi algoritma yang berbeda untuk merekomendasikan item kepada pengguna. Contohnya, *Collaborative Filtering* mengumpulkan dan menganalisis data dari semua pengguna untuk memprediksi apa yang pengguna sukai berdasarkan kemiripan dengan pengguna lain, sedangkan *Content Based Filtering* merekomendasikan item berdasarkan kemiripan deskripsi item dengan minat profil pengguna [1]. *Content Based Filtering* mencoba menggunakan atribut item, dengan beberapa dari kesamaan diantara dua buah item, untuk menghasilkan item serupa yang berhubungan dengan item yang diberikan [2]. *FP-Growth* merupakan algoritma lain yang banyak diketahui untuk menemukan item yang sering muncul. Algoritma ini

didasarkan pada data struktur *FP-Tree*, yang mengembalikan informasi frekuensi dari item secara konsisten. Tidak seperti algoritma *Apriori*, algoritma *FP-growth* hanya membutuhkan dua pemindaian per database dan tidak menghasilkan kandidat item set, maka proses algoritma *FP-Growth* ditinjau lebih efisien dibandingkan proses algoritma *Apriori* [3]. *Text Preprocessing* merupakan langkah penting dalam mengonversi teks menjadi teks terstruktur. Oleh karena itu, diputuskan untuk membandingkan kata query dengan kata indeks (istilah atau frasa penting) yang muncul dalam dokumen itu sendiri. Pada umumnya terdapat empat tahapan utama yaitu *Case Folding*, *Tokenizing*, *Filtering* dan *Stemming* [4]. Kedua algoritma tersebut sangat membutuhkan *text processing* atau pengolah kata untuk mempermudah pemrosesan data, seperti membersihkan beberapa kata yang terdapat imbuhan, terdapat simbol dan menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data.

Program Mahasiswa Wirausaha (PMW) adalah program kegiatan yang di adakan perguruan tinggi Universitas Negeri Surabaya yang dilaksanakan oleh Satuan Unesa Career Center (SUCC) untuk mengembangkan keterampilan mahasiswa yang ingin berwirausaha dan ingin memiliki profit dalam berwirausaha. Tujuan diadakannya PMW yaitu terciptanya pengusaha muda yang memiliki pendidikan tinggi yang siap dalam persaingan global [5]. Program mahasiswa wirausaha sangat membantu mahasiswa yang ingin memulai usaha tanpa mengeluarkan modal sepeser pun yaitu mendapat modal PMW yang bersifat hibah dan didampingi oleh tim PMW Unesa. Banyak mahasiswa yang kesulitan untuk mencari ide atau judul usaha yang akan diajukan, ide-ide di internet pun belum menjamin ide tersebut bisa lolos seleksi program mahasiswa wirausaha, hal ini disebabkan beberapa faktor utama yaitu ide atau judul yang diangkat terlalu umum, inovasi yang dikembangkan tidak mengikuti kebutuhan zaman, perencanaan usaha belum matang dan dana yang diajukan terlalu tinggi. Oleh karena itu dibutuhkan sistem untuk mengembangkan sistem rekomendasi program mahasiswa wirausaha di Universitas Negeri Surabaya untuk mahasiswa yang mengikuti kegiatan PMW.

Penelitian sebelumnya sangat memudahkan peneliti untuk menerapkan pada sistem rekomendasi program mahasiswa wirausaha. Banyak algoritma yang telah diterapkan oleh peneliti sebelumnya terkait sistem rekomendasi memakai algoritma *Content Based Filtering*, algoritma *Apriori* dan algoritma *FP-Growth*. Septiarnata (Murti et al., 2019)

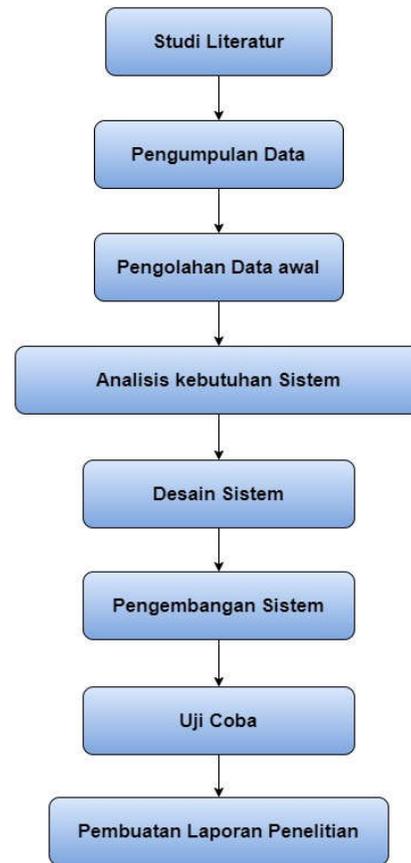
mengusulkan algoritma *FP-Growth* dan metode *Content Based Filtering* untuk mengidentifikasi pola peminjaman buku di Perpustakaan Fakultas Teknologi Informasi Universitas Stikubank Semarang. Aturan asosiasi dapat dimanfaatkan untuk memilih buku yang akan ditampilkan di katalog perpustakaan berdasarkan data riwayat peminjaman buku pengguna [6]. Pada penelitian (Tommy et al., 2019) mengusulkan implementasi aplikasi yang dibangun dilengkapi sistem rekomendasi, yang menggunakan algoritma Apriori (berdasarkan skema pembelian produk oleh pelanggan pada umumnya) dan metode *Content Based Filtering* (menurut hobi yang dipilih pelanggan waktu mendaftar akun dan karakteristik produk pada keranjang pelanggan) [7]. Pada penelitian lain (Pranata & Utomo, 2020) mengusulkan metode data mining, *Association Rules* dan algoritma *FP-Growth* untuk diterapkan pada analisis data mining untuk meningkatkan penjualan, khususnya pelayanan terhadap kebutuhan masyarakat. Kemudian untuk mengetahui suatu strategi peningkatan kualitas penjualan dan menganalisis proses yang berdampak kuat pada penjualan sparepart dengan menggunakan algoritma *FP-Growth* yaitu penyelesaian berupa pencarian *association rules mining*. [8].

Pada penelitian ini mengembangkan beberapa metode dari peneliti sebelumnya. Pengembangan yang akan dilakukan adalah menerapkan 2 metode yaitu algoritma *Content Based Filtering* dan algoritma *FP-Growth*, dengan 2 model perhitungan yang berbeda namun sama-sama mencari hasil berdasarkan kata yang sering muncul. Parameter analisis pada sistem rekomendasi program mahasiswa wirausaha yaitu berdasarkan kriteria judul, kriteria abstrak dan kriteria dana.

Proses implementasi pada penelitian ini yaitu memberikan informasi gambaran judul pmw yang pernah diterima dan di implementasikan pada sebuah sistem rekomendasi judul PMW serta memberikan standart perkiraan dana yang diajukan agar sesuai konsep yg diangkat. Adanya sistem ini diharapkan dapat membantu mahasiswa yang mengikuti kegiatan Program Mahasiswa Wirausaha dan dapat mengembangkan ide-ide dengan memberikan informasi dari hasil rekomendasi judul mahasiswa yang lolos PMW ditahun sebelumnya.

## II. METODE PENELITIAN

Tahapan yang digunakan dalam alur penelitian untuk membangun sistem rekomendasi program mahasiswa wirausaha di Universitas Negeri Surabaya dilakukan dalam beberapa tahapan pada Gbr. 1.



Gbr. 1 Alur Penelitian

Penelitian ini terdapat beberapa tahapan yang harus dilakukan supaya penelitian dapat berjalan sesuai dengan alur penelitian dalam Gbr. 1. penjelasan masing-masing tahapan akan dijabarkan sebagai berikut :

### A. Studi Literatur

Tahap awal pada penelitian ini adalah melakukan studi literatur, tahap ini bertujuan untuk memperoleh berbagai informasi. Studi Literatur dilakukan untuk mencari teori mengenai *Recommender System*, *Text Preprocessing*, *Content Based Filtering*, *Association Rules*, dan *Frequent Patern Growth*. Pencarian teori dilakukan pada *text book* dan jurnal ilmiah.

### B. Tahap Pengumpulan Data

Pada tahap ini dilakukan pengumpulan dataset yang akan digunakan dalam penelitian ini diperoleh dari Divisi Kewirausahaan dan Pengembangan Karier pada Satuan Unesa Career Center (SUCC) di Universitas Negeri Surabaya. Data yang diperoleh pada tahap ini adalah data mahasiswa yang mengikuti kegiatan program mahasiswa wirausaha dan lolos sampai tahap didanai pada tahun 2018-2020 yaitu sebanyak 182 data, dengan rincian 42 data tahun 2018, 62 data tahun 2019 dan 72 data tahun 2020.

C. Pengolahan Data Awal

Pada tahap ini dilakukan pengolahan data awal. Data awal yang diperoleh dari tahap sebelumnya khususnya dari tahap pengumpulan data, terdiri dari beberapa kolom dan dari beberapa kolom data tersebut harus dipilih untuk digunakan dalam penelitian ini. Ada beberapa kriteria yang akan digunakan pada penelitian ini yaitu kriteria judul program mahasiswa wirausaha, kriteria abstrak atau deskripsi usaha, dan kriteria dana yang disetujui. Pembersihan data akan dilakukan terhadap data2 yang ganda, nul, dan data tipe data yang tidak sesuai.

D. Analisis Kebutuhan Sistem

Pada tahap ini dilakukan analisis kebutuhan sistem yang terdiri dari identifikasi masalah dan analisis kebutuhan fungsional.

1) Identifikasi Masalah

Berdasarkan latar belakang yang dibahas di atas, masalah dapat dirumuskan, bagaimana cara menerapkan sistem rekomendasi program mahasiswa wirausaha dengan algoritma Content Based Filtering dan algoritma Frequent Pattern Growth

2) Analisis Kebutuhan Fungsional

Sistem rekomendasi program mahasiswa wirausaha ini dibuat untuk memenuhi kebutuhan pengguna sebagai berikut :

- a. Pengguna mendapatkan rekomendasi judul referensi program berdasarkan masukan berupa judul dan abstrak.
- b. Pengguna mendapatkan perkiraan dana berdasarkan masukan berupa judul, abstrak dan perkiraan dana.

3) Analisis Kebutuhan Pengguna

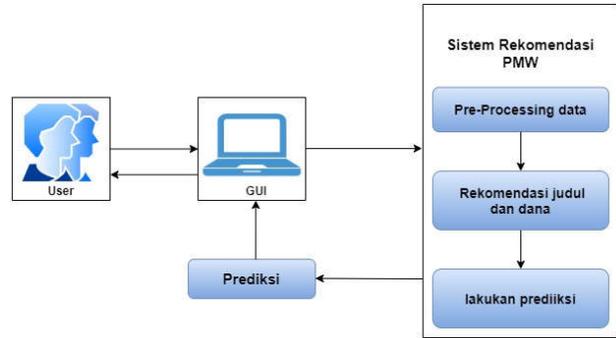
Dalam sistem rekomendasi ini, kebutuhan pengguna di atas diakomodir beberapa fungsi dalam program yaitu pada Tabel I

TABEL I  
 FUNGSI KEBUTUHAN PENGGUNA

kode	Form
F1	Form input judul program
F2	Form input abstrak
F3	Form input dana
F4	Menampilkan rekomendasi judul
F5	Menampilkan perkiraan dana

E. Desain Sistem

Sistem yang diterapkan dalam sistem rekomendasi ini yaitu straightforward, dimana user langsung bisa mendapatkan rekomendasi judul referensi dan perkiraan dana dengan memasukkan judul, abstrak dan dana pada halaman web yang telah disediakan. Gbr. 2 merupakan rancangan desain sistem rekomendasi.

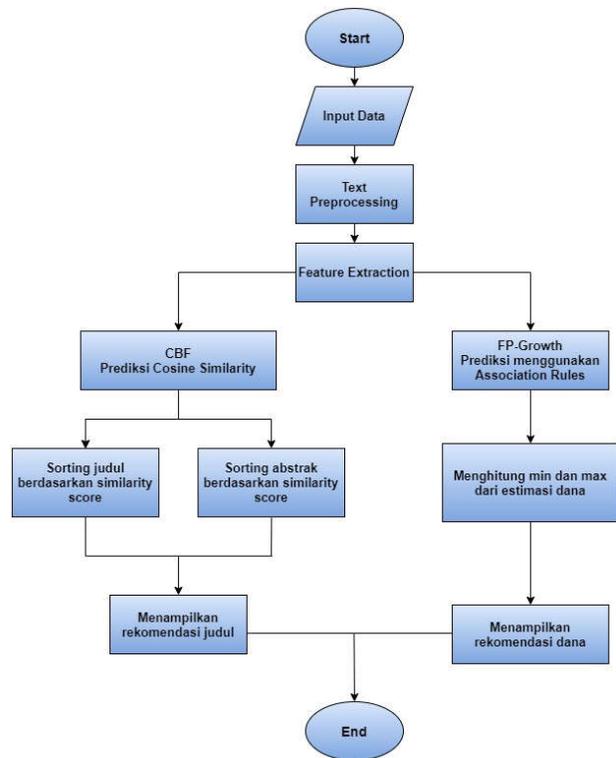


Gbr. 2 Desain Sistem

Gbr. 2 menjelaskan bahwa pengguna diarahkan ke GUI dan sistem akan memproses input data yang dimasukkan oleh pengguna dan menampilkan output rekomendasi untuk ditampilkan kepada pengguna.

1) Flowchart

Sistem rekomendasi program mahasiswa wirausaha memiliki flowchart untuk proses prediksi untuk mendapatkan sistem rekomendasi. Gbr. 3 merupakan Alur atau Flowchart dari proses sistem rekomendasi.



Gbr. 3 Flowchart Sistem Rekomendasi

Dari flowchart yang ditunjukkan pada Gbr.3 dapat diuraikan sebagai berikut yaitu hal pertama yang dilakukan adalah input data yaitu memasukkan data yang diperlukan dalam sistem dalam hal ini yaitu dataset PMW dalam bentuk .csv file.

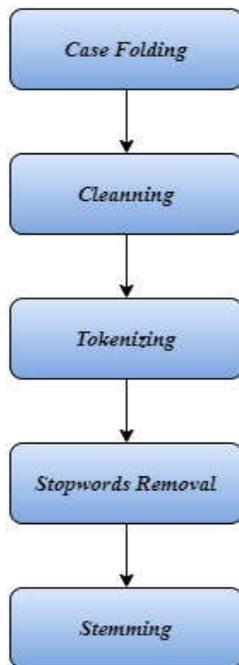
Setelah data berhasil dimasukkan selanjutnya dilakukan *text preprocessing* untuk menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data. Setelah dilakukan *text preprocessing*, dilakukan proses *feature extraction* yaitu mengubah text menjadi vektor. Selanjutnya dilakukan prediksi dengan 2 cara yaitu dengan menggunakan algoritma *CBF* yaitu prediksi *Cosine Similarity* dan dengan menggunakan *FP-Growth* yaitu prediksi menggunakan *Association Rules*. Pada algoritma *CBF* proses yang dilakukan yaitu *sorting* judul dan *sorting* abstrak berdasarkan *similarity score* sehingga dapat menghasilkan rekomendasi judul dan abstrak namun yang ditampilkan hanya rekomendasi judul. Dan pada algoritma *FP-Growth* dilakukan penghitungan nilai minimal dan maksimal dari perkiraan dana sehingga menghasilkan perkiraan dana yang sesuai dengan rekomendasi judul.

#### F. Tahap Pengembangan Sistem

Pada tahap ini dilakukan pengembangan sistem, dimana penulis merancang sistem yang akan dibentuk sebagaimana akan sesuai dengan tujuan penelitian artikel ilmiah ini. Pada tahap pengembangan sistem akan dilakukan *text preprocessing* dan implementasi algoritma.

##### 1) Text Preprocessing

Pada tahap ini yaitu *Text Preprocessing* yang digunakan untuk mengolah data dengan cara menghilangkan permasalahan-permasalahan yang dapat mengganggu hasil daripada proses data. Ada beberapa tahapan *Text Preprocessing* pada Gbr. 4.



Gbr. 4 Tahap *preprocessing*

- a. *Case Folding*  
 Pada tahapan *case folding* dilakukan perubahan dalam kumpulan data menjadi huruf kecil semua.
- b. *Cleanning*  
 Pada tahapan *cleaning* dilakukan proses pembersihan data menggunakan cara menghilangkan data yang tidak digunakan. Seperti menghapus simbol dan angka.
- c. *Tokenizing*  
 Pada tahapan *tokenizing* dilakukan proses pemisahan kalimat menjadi potongan-potongan kata, angka ataupun symbol. Seperti “produk makanan.” menjadi [‘produk’, ‘makanan’, ‘.’].
- d. *Stopwords Removal(Filtering)*  
 Pada tahapan *stopwords removal* dilakukan proses menemukan kata-kata penting dan menghapus istilah atau kata-kata yang kurang penting. Seperti “itu”, “yang”, “dan”. Intinya yaitu menghapus istilah yang mempunyai informasi rendah supaya bisa fokus dalam istilah yang penting saja.
- e. *Stemming*  
 Pada tahapan *stemming* dilakukan proses untuk mengubah istilah atau kata yang berimbuhan menjadi kata ke bentuk dasarnya, sesuai menggunakan KBBI. Seperti kata “makanan” menjadi kata “makan”.

##### 2) Tahap Implementasi Content Based Filtering (CBF)

Metode yang pertama yaitu menggunakan algoritma *Content Based Filtering* dengan teknik *Cosine Similarity*. *Cosine Similarity* adalah metode berbasis dokumen yang dibuat dengan menggunakan kata kunci untuk menghitung kemiripan antara 2 objek yang dinyatakan pada 2 vector [9].

Rekomendasi judul usaha menurut hasil dari mahasiswa yang sudah lolos mengikuti program mahasiswa wirausaha di tahun sebelumnya merupakan konsep dari algoritma *content based filtering*. Faktor yang akan merekomendasikan sebuah judul usaha akan diberikan bobot berdasarkan suatu parameter. Dan *output* rekomendasinya yaitu faktor yang mempunyai nilai *cosine similarity* tertinggi. Langkah-langkah perhitungan algoritma *Content Based Filtering* dengan teknik *Cosine Similarity* sebagai berikut :

- a. *Penyiapan Dataset*
- b. *Preprocessing terhadap semua dataset*
- c. *Proses perhitungan kemiripan menggunakan Cosine Similarity*

Berikut ini adalah rumus *Cosine Similarity* :

$$\text{cossim}(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (1)$$

Keterangan :

*A* = Vektor

*B* = Vektor

*A<sub>i</sub>* = bobot term *i* dalam blok *A<sub>i</sub>*

*B<sub>i</sub>* = bobot term *i* dalam blok *B<sub>i</sub>*

*i* = jumlah term dalam kalimat

- d. *Urutkan hasil perhitungan Similarity dari yang tertinggi (descending)*

### 3) Tahap Implementasi FP-Growth

Metode yang kedua yaitu menggunakan algoritma *Frequent Pattern Growth (FP-Growth)*. Algoritma yang bisa digunakan untuk memilih kumpulan data yang paling sering muncul (frequent item set) dari kumpulan data. Langkah-langkah perhitungan algoritma *FP-Growth* sebagai berikut :

- a. *Penyiapan Dataset*
- b. *Preprocessing terhadap semua dataset*
- c. *Pencarian Frequent Item set*
- d. *Urutkan berdasarkan priority*
- e. *Dataset diurutkan berdasarkan priority*
- f. *Cari confidence dari item baru*
- g. *Association Rules*

### G. Tahap Uji Coba

Pada tahap ini dilakukan uji coba pengujian sistem menggunakan algoritma *Content Based Filtering* dan algoritma *Frequent Pattern Growth* yang diterapkan pada sistem rekomendasi. Kemudian akan di analisis untuk kelebihan dan keterbatasan algoritma tersebut pada sistem rekomendasi program mahasiswa wirausaha.

### H. Tahap Pembuatan Laporan Penelitian

Tahap akhir dari penelitian ini dilakukan penyusunan laporan penelitian dalam bentuk artikel ilmiah yang berisi dasar teori, tahapan proses penelitian, serta hasil penelitian dengan dilakukan pengujian sistem menggunakan algoritma *content based filtering* dan *FP-Growth*.

## III. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan Sistem Rekomendasi Program Mahasiswa Wirausaha (PMW) di Universitas Negeri Surabaya dengan menggunakan algoritma *Content Based Filtering* dan *Frequent Pattern Growth*. Berikut adalah hasil dan pembahasan mengenai sistem rekomendasi PMW :

### A. Python

Sistem Rekomendasi PMW dibuat menggunakan bahasa pemrograman *python* dengan *user interface* berbasis *website*, dan memanfaatkan beberapa modul sebagai berikut :

- 1) *pandas* untuk mengolah data.
- 2) *numpy* untuk melakukan operasi matriks dan matematis lainnya.
- 3) *nlTK* untuk melakukan *text preprocessing*.
- 4) *sastrawi* untuk melakukan *text preprocessing* dalam bahasa Indonesia.
- 5) *sklearn* untuk melakukan *feature extraction*, perhitungan *cosine similarity*.
- 6) *mlxtend* untuk menghitung *frequent item set* dan *association rules*.
- 7) *jupyter notebook* untuk data analisis dan pembuatan model.
- 8) *dash* untuk membuat aplikasi web berbasis *flask microframework*.

- 9) *dash-bootstrap-components* sebagai modul yang menyediakan tema tampilan web berdasarkan *bootstrap*.

### B. Text Preprocessing

Untuk memperoleh rekomendasi, pengguna memasukkan judul, abstrak dan dana. Akan dilakukan *text preprocessing* dengan tahap sebagai berikut.:

#### 1) *Cleaning dan Case Folding*

Pada tahap ini data judul dan abstrak dibersihkan dari karakter-karakter yang tidak dibutuhkan dalam analisis, memanfaatkan fungsi fungsi seperti *replace*, *rstrip* dan *dropna*. Serta pada tahapan *case folding* menggunakan fungsi *lower* untuk mengubah seluruh huruf kapital menjadi huruf kecil.

#### 2) *Tokenisasi*

Tahapan *tokenisasi* memisahkan data menjadi kumpulan kata dalam format list, memanfaatkan fungsi *word\_tokenize* yang terdapat dalam modul *nlTK*. Agar fungsi *word\_tokenize* dapat digunakan, maka perlu diunduh terlebih dahulu *library punctuation* dengan menjalankan fungsi *nlTK.download('punct')*. Kemudian selanjutnya menjalankan perintah iterasi untuk melakukan tokenisasi terhadap setiap baris data.

#### 3) *Stopwords Removal*

Setelah data dilakukan tokenisasi, data perlu dibersihkan dari kata-kata yang kurang bermanfaat dalam analisis, atau dilakukan *stop words removal*. *Stop words removal* dilakukan dengan memanfaatkan *stopwords dictionary* bahasa Indonesia yang tersedia dalam modul *nlTK*, dengan sebelumnya mengunduh *dictionary* tersebut dengan perintah *nlTK.download('stopwords')*, kemudian melakukan iterasi terhadap setiap data untuk memfilter apakah ada atau tidaknya *stopwords* dalam data.

#### 4) *Stemming*

Tahapan selanjutnya adalah melakukan proses *stemming*, yaitu mengembalikan kata-kata dalam data menjadi akar kata. Proses ini dilakukan dengan menggunakan modul *sastrawi*, *stemming* khusus bahasa Indonesia karena fungsi khusus bahasa Indonesia ini tidak tersedia dalam *nlTK*. Proses dijalankan dengan melakukan iterasi terhadap setiap data kemudian melakukan *stemming* terhadap setiap kata yang terdapat pada baris data.

### C. Pembuatan User Interface

Untuk pembuatan *user interface* sistem rekomendasi PMW, digunakan modul utama yaitu *dash* versi 1.21.0 yang merupakan perangkat lunak *opensource* dikembangkan oleh *plotly* dengan memanfaatkan *flask* serta teknologi *javascript* dari *D3.js* dalam memvisualisasikan data, serta *dash-bootstrap-components* yang merupakan modul tambahan dari *dash*, sehingga program dapat menggunakan mode tampilan dari *bootstrap* dengan lebih mudah.

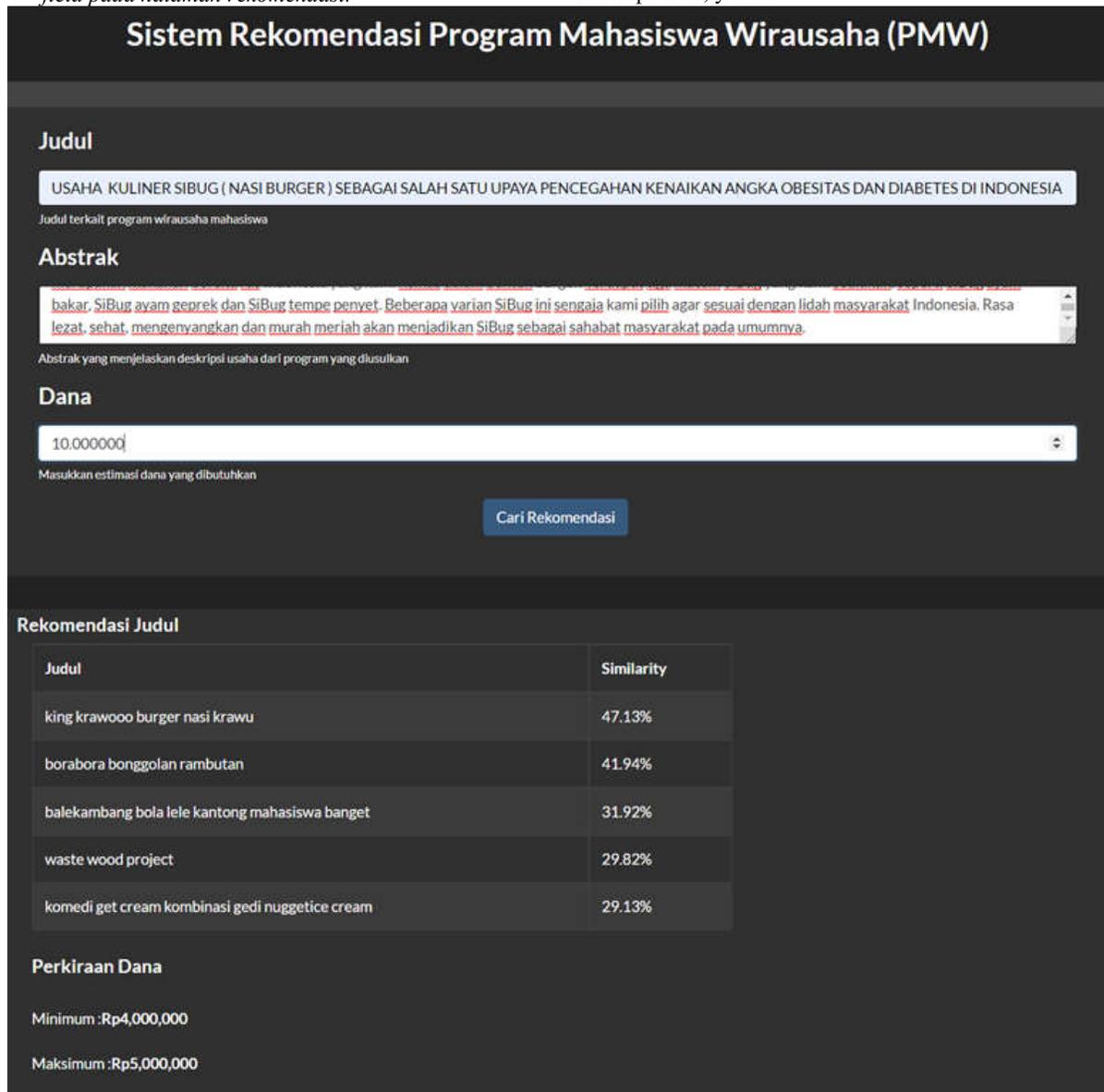
Beberapa komponen yang digunakan dalam pembuatan user interface adalah sebagai berikut :

- 1) Fungsi penyusun layout html dan css baik menggunakan fungsi dari dash-html-components maupun dash-bootstrap-components, yaitu Div, Container, Row dan Col.
- 2) Fungsi Card, FormGroup, dan Input dari dash-bootstrap-components untuk menyusun tampilan input field pada halaman rekomendasi.

- 3) Fungsi Button dari dash-bootstrap-components untuk menampilkan tombol pencarian.

#### D. Hasil Pengujian Sistem

Sistem dibuat dengan bahasa pemrograman Python dan pada tampilan rekomendasi untuk client yang merupakan mahasiswa yang mengikuti kegiatan PMW. Gbr. 5 ini adalah tampilan dari halaman rekomendasi : rekomendasi terdapat beberapa fitur, yaitu :



Gbr. 5 Tampilan Halaman Rekomendasi

Pada tampilan rekomendasi di Gbr. 5 dapat ditunjukkan hasil prediksi untuk rekomendasi judul dan perkiraan dana yang paling mirip atau yang paling sesuai yang telah dimasukkan adalah “king krawooo burger nasi krawu” dengan

similarity (kemiripan) 47.13% dengan rata2 kemiripan dari top 5 rekomendasi judul yaitu 35.99% dan perkiraan dana minimum Rp4.000.000,- dan maksimum Rp5.000.000,-

- 1) Content Based Filtering (CBF)

Metode yang pertama yaitu menggunakan algoritma *Content Based Filtering* dengan teknik *Cosine Similarity*. Langkah-langkah perhitungan algoritma *Content Based Filtering* dengan teknik *Cosine Similarity* untuk rekomendasi judul PMW merupakan sebagai berikut :

- a. Dataset PMW 2018-2020 sebanyak 179 data
- e. Preprocessing terhadap semua dataset
- f. Proses perhitungan kemiripan judul dan abstrak menggunakan *Cosine Similarity*
- g. Urutkan hasil perhitungan *Similarity* judul dari yang tertinggi (*descending*)
- h. Menampilkan hasil rekomendasi judul 5 teratas serta *Similarity score* dan *mean (rata-rata)*

## 2) FP-Growth

Metode yang kedua yaitu menggunakan algoritma *Frequent Pattern Growth (FP-Growth)*. Langkah-langkah perhitungan algoritma *FP-Growth* untuk perkiraan dana PMW merupakan sebagai berikut :

- a. *Association Rules* memunculkan *Atencenden* dan *Consequens*
- b. Hasil dari *Frequent Item set* dan *Association Rules* di filter yang *Consequens nya* memunculkan item dana

- c. Input baru dari form input dana, diproses bersamaan dengan judul dan abstrak
- d. Preprocessing input baru dan outputnya term *frequency*
- e. Dari term *frequency* menghasilkan *similarity* yang paling dekat dengan *frequent item set* dan *antencenden*
- f. Pada *frequent item set* dan *antencenden* tercantum *Consequens* beberapa dana yang *similarity*
- g. Perhitungan dana diambil *Confidence* dari dua peringkat tertinggi
- h. Dua peringkat tersebut dijadikan minimum dan maximum dana
- i. Menampilkan hasil rekomendasi perkiraan dana dengan minimum dan maksimum dana tertinggi.

Berikut ini adalah hasil pengujian sistem pada halaman rekomendasi yaitu untuk melakukan pencarian rekomendasi judul dan dana. Tabel II adalah beberapa hasil pengujian terhadap beberapa judul yang berbeda :

TABEL II  
HASIL PENGUJIAN TERHADAP DATA BARU

No	Judul	Abstrak	rekomendasi judul tertinggi	similarity tertinggi	mean dari 5 judul teratas
1	tireng (tiwul goreng) milenial	tireng (tiwul goreng) milenial merupakan...	Kubale	47.83%	44.2%
2	usaha kuliner sibug ( nasi burger ) ...	sibug seperti namanya sibug merupakan nasi burger...	king krawooo burger nasi krawu	47.13%	35.99%
3	waysho ( wayang shoes )	sepatu sneakers merupakan sepatu yang banyak dipilih dan ...	the amazing shoes	45.17%	34.1%
4	"v-co: dry shampoo gel" sampo alami...	v-co: dry shampoo gel adalah jenis produk sampo berbentuk...	sublim mask masker bubuk susu dan daun belimbing wuluh ...	42.81%	40.06%
5	naturally drink	minuman herbal merupakan minuman...	kepo kedelai puding kombinasi kesegaran ...	37.53%	33.38%
6	xi rembo (rempah boba)	xi rembo merupakan minuman inovatif...	silky almond a pudding made with almond milk	35.92%	26.68%
7	inovasi totebag komunikatif sebagai media...	totebag adalah sebuah produk tas bahu...	tutonsi totebag two tone multifungsi	32.09%	27.54%
8	cinnamoo(vegan cinnamon-moringa ice cream)	cinnamoo merupakan olahan inovasi minuman...	kumat ice cream	30.36%	27.11%
9	arival bag (tas damar kurung anti virus dan bakteri )	era new normal mengharuskan kita untuk selalu menjaga kebersihan...	tamasya tas makrame Sylvia	24.86%	23.11%
10	v-citife (aplikasi	di era serba digital saat ini,	halo pahlawan web design and	23.43%	19.29%

	manajemen sertifikat untuk event berbasis web)	teknologi masuk mem...	digital marketing agency		
--	--	------------------------	--------------------------	--	--

#### UCAPAN TERIMA KASIH

Ucapan terima kasih penulis ucapkan kepada Tuhan Yang Maha Esa atas segala karunia yang telah diberikan sehingga dapat menyelesaikan artikel ilmiah ini dengan sebaik-baiknya dari awal sampai akhir. Penyusunan artikel ilmiah ini tidak terlepas dari bimbingan, arahan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada orang tua dan keluarga yang selalu mendukung studi penulis. Penulis mengucapkan banyak terima kasih kepada seluruh civitas akademika jurusan Teknik Informatika Universitas Negeri Surabaya, khususnya teman-teman S1 Teknik Informatika angkatan 2017. Semoga segala bantuan yang sudah diberikan kepada penulis mendapat balasan yang setimpal dari Allah SWT. Semua penjelasan dalam laporan ini merupakan hasil dari rangkaian kajian studi dan analisis yang dilakukan oleh penulis dari awal hingga akhir. Penulis telah berusaha sebaik-baiknya dalam menyelesaikan artikel ilmiah ini.

#### REFERENSI

- [1] Ng, Y. K. (2020). CBRec: a book recommendation system for children using the matrix factorisation and content-based filtering approaches. *International Journal of Business Intelligence and Data Mining*, 16(2), 129. doi:10.1504/ijbidm.2020.104738
- [2] Dangeti, P. (2017). *Statistics for Machine Learning*. Birmingham:Packt Publishing.
- [3] Bagui, S., Devulapalli, K., & Coffey, J. (2020). A heuristic approach for load balancing the FP-growth algorithm on MapReduce. *Array*, 7, 100035. doi:10.1016/j.array.2020.100035
- [4] Hermawan, L., & Ismiati, M. B. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. *Jurnal Transformatika*, 17(2), 188-199.
- [5] Tim PMW 2021. (2021). *Buku Pedoman Program Mahasiswa Wirausaha (PMW)*. Surabaya: Satuan Unesa Career Center.
- [6] Murti, H., & Lestariningsih, E. (2019). Perancangan Sistem Rekomendasi Buku Pada Katalog Perpustakaan Menggunakan Pendekatan Content-Based Filtering Dan Algoritma Fp-Growth.
- [7] Tommy, L., Kirana, C., & Lindawati, V. (2019). Recommender System dengan Kombinasi Apriori dan Content-Based Filtering pada Aplikasi Pemesanan Produk. *Jurnal Teknoinfo*, 13(2), 84-95.
- [8] Pranata, B. S., & Utomo, D. P. (2020). Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service). *Bulletin of Information Technology (BIT)*, 1(2), 83-91.
- [9] Wahyuni, R. T., Prastiyanto, D., & Suprptono, E. (2017). Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi. *Jurnal Teknik Elektro*, 9(1), 18-23.

Dari Tabel II dapat dilihat bahwa dari 10 pengujian terhadap data baru (10 dataset pmw 2021) tingkat kemiripan atau *similarity score* terhadap pengujian data baru memiliki tingkat kemiripan yang rendah, dikarenakan judul dan abstraksi yang beragam. Hal tersebut terlihat dari hasil persentase tertinggi dan terendah pengujian data baru. Hasil *Similarity* tertinggi yaitu 47.83% dengan rata-rata (*mean*) 44.2% dan *similarity* terendah yaitu 23.43% dengan rata-rata (*mean*) 19.29%.

#### IV. KESIMPULAN

Penelitian ini menggunakan algoritma *Content Based Filtering* dan *Frequent Pattern Growth* untuk memberikan rekomendasi judul dan perkiraan dana yang diajukan pada sistem rekomendasi program mahasiswa wirausaha. Berdasarkan *output* pengujian sistem yang telah dilakukan maka bisa diambil kesimpulan bahwa :

1. Rekomendasi *content based filtering* dengan mengukur nilai *cosine similarity* terhadap *vectorized document* lebih memudahkan mahasiswa untuk melakukan pencarian terhadap judul. Dari hasil 10 pengujian terhadap data baru, dapat disimpulkan *similarity* tertinggi yaitu 47.83% dengan *mean* 44.2% dan *similarity* terendah yaitu 23.43% dengan *mean* 19.29%.
2. *Association rules* yang dibangun berdasarkan model *Frequent Pattern Growth* dan data yang telah dikumpulkan, dibutuhkan nilai *minimum support 0.03* dan *minimum confidence 0.1* sehingga dapat menghasilkan rekomendasi dana. Hal ini disebabkan *similarity* dari *frequent item set* sangat berbeda satu dengan yang lainnya.

#### V. SARAN

Berikut ini merupakan saran-saran yang perlu dipertimbangkan untuk pengembangan sistem lebih lanjut.

1. Sistem Rekomendasi PMW dibuat untuk memberikan informasi gambaran judul pmw yang pernah diterima dan harus dikembangkan untuk keoriginalitas karya yang telah dibuat.
2. Mengembangkan sistem pada penelitian ini dengan menambahkan beberapa fitur lainnya, seperti : rekomendasi proposal yang dapat diunduh.
3. Mengembangkan sistem pada penelitian ini yaitu menerapkan algoritma lain dengan data yang lebih besar.