

Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta

Application of the Cosine Similarity Algorithm and Weighting of the TF-IDF System for New Student Admissions on Private Campuses

Apriani^{1*}, Hizbu Zakiyudin², Khairan Marzuki³

^{1,2,3} Universitas Bumigora

apriani@universitasbumigora.ac.id^{1*}, zacky66blackcloud@gmail.com²,

Khairan.marzuki@universitasbumigora.ac.id³

Abstrak – Era globalisasi ditandai dengan perkembangan teknologi dan informasi, hal ini berimbas pada kebutuhan manusia akan informasi. PMB (Penerimaan Mahasiswa Baru) merupakan aktivitas rutin perguruan tinggi pada setiap pembukaan ajaran baru. Pelaksanaan PMB tidak luput dari berbagai pertanyaan yang sudah pernah ditanyakan sebelumnya. Dengan memanfaatkan teknologi informasi maka lahirlah FAQ (Frequently Asked Question) yang berisikan jawaban dari pertanyaan yang sering dilontarkan oleh orang yang membutuhkan informasi. Tujuan penelitian ini adalah mengurangi beban pihak PMB dalam menjawab pertanyaan secara berulang-ulang, maka dibangunlah sistem penjawab FAQ dengan menerapkan pembobotan TF-IDF (Term Frequency- Inverse Document Frequency) dan algoritma cosine similarity. Metode dalam penelitian ini menggunakan cosine Similarity Algorithm dan pembobotan TF-IDF, Penelitian ini menggunakan 7 buah sampel data dari keseluruhan data FAQ yang didapat dari wawancara dengan pihak penerimaan mahasiswa baru Universitas Bumigora. Data sampel yang digunakan akan melalui proses preprocessing, pembobotan TF-IDF, dan metode cosine similarity untuk menentukan tingkat kesamaan tertinggi yang nantinya akan keluar sebagai hasil akhir. Dengan menggunakan pembobotan TF-IDF dan metode cosine similarity pada 7 data sampel bisa mendapatkan tingkat akurasi hingga mencapai 64.28%.

Kata Kunci : Cosine Similarity, FAQs, PMB, TF-IDF

Abstract – The era of globalization is marked by the development of technology and information, this has an impact on the human need for information. PMB (New Student Admission) is a university routine activity at every opening of a new teaching. The implementation of PMB does not escape various questions that have been developed previously. By utilizing information technology, a FAQ (Frequently Asked Question) was born which contains answers to questions that are often asked by people who need information. The purpose of this research is to reduce the burden on the PMB in answering questions repeatedly, so a FAQ answering system was built by applying TF-IDF (Term Frequency-Inverse Document Frequency) weighting and cosine similarity algorithms. The method in this study uses the Cosine Similarity Algorithm and TF-IDF weighting. This study uses 7 data samples from the total FAQ data obtained from interviews with new student admissions at Bumigora University. The sample data will be used through the preprocessing process, TF-IDF weighting, and the cosine similarity method to determine the best level of comfort that will be achieved later as the final result. By using TF-IDF weighting and the cosine similarity method on 7 data samples, an accuracy rate of up to 64.28% can be obtained.

Keywords: FAQs, TF-IDF, Cosine Similarity, PMB, Universitas Bumigora.

Jurnal Bumigora Information Technology (BITe)

Vol.3, No.1, Juni 2021, pp. 19-27

ISSN: 2685-4066

DOI: 10.30812/bite.v3i1.1110

1. Pendahuluan

Era globalisasi ditandai dengan perkembangan teknologi dan informasi yang semakin maju, hal tersebut berimbas pada kebutuhan manusia akan informasi yang merupakan hal penting dalam berbagai kegiatan manusia [1]. Perkembangan informasi juga selalu diikuti dengan meningkatnya kebutuhan manusia akan informasi oleh karena itu, berbagai cara bisa digunakan untuk memenuhi kebutuhan informasinya [2][3]. Kebutuhan akan informasi inilah yang membuat berbagai lembaga untuk menyediakan informasi yang beragam ditambah lagi dengan perkembangan teknologi yang semakin maju juga meningkatkan pola perilaku masyarakat dalam menggunakan teknologi untuk mencari, menggunakan, dan bertukar informasi [4], begitu juga dengan pelaksanaan Penerimaan Mahasiswa Baru (PMB) oleh perguruan tinggi.

PMB merupakan aktivitas rutin perguruan tinggi pada setiap pembukaan ajaran baru. Pelaksanaan PMB tidak luput dari berbagai pertanyaan-pertanyaan yang sudah pernah ditanyakan sebelumnya oleh pihak calon mahasiswa baru. Dengan memanfaatkan teknologi agar masyarakat lebih dimudahkan untuk mendapatkan informasi, maka lahirlah Frequently Asked Question (FAQ) yang membantu mendapatkan informasi yang dibutuhkan. FAQ berisikan jawaban dari pertanyaan-pertanyaan yang sering dilontarkan oleh orang-orang yang membutuhkan informasi. Banyaknya informasi yang bersifat mirip dapat menjadi faktor pengguna mengalami kesulitan pada saat mencari kata yang diinginkan, metode Cosine Similarity dapat menjadi solusi untuk masalah ini.

Metode cosine similarity ini menghitung tingkat kesamaan (similarity) antara dua buah objek yang dinyatakan dalam dua buah vektor dengan menggunakan kata kunci dari sebuah dokumen sebagai ukuran [5]. Metode Cosine 2 Similarity digunakan untuk menghitung tingkat kesamaan antar dokumen tetapi tidak menyertakan frekuensi kata (term). Dengan adanya pembobotan TF-IDF (Term Frequency – Inverse Document Frequency) yang dapat memberikan nilai bobot pada setiap kata dalam dokumen akan membantu cosine similarity untuk memproses kata secara maksimal [6]. Berdasarkan pemaparan di atas, penulis tertarik untuk melakukan penelitian dalam menganalisa Sistem Penjawab FAQ Menggunakan Metode Cosine Similarity dengan pembobotan TF-IDF.

2. Metode Penelitian

Data yang digunakan pada penelitian ini merupakan data digital yang berisikan tentang FAQ seputar penerimaan mahasiswa baru Universitas Bumigora Mataram. Metode pengumpulan data dilakukan dengan 2 cara, yaitu dengan studi literatur yang dilakukan pada brosur, website serta data dari PMB, kemudian wawancara yang dilakukan dengan staf bagian Penerimaan Mahasiswa Baru (PMB) terkait dengan pertanyaan-pertanyaan yang sering diterima dari calon mahasiswa.

2.1. Analisis Data

Data yang digunakan pada penelitian ini berupa data yang berisikan tentang penerimaan mahasiswa baru di Universitas Bumigora berdasarkan pertanyaan-pertanyaan yang sering diterima oleh PMB. Data penelitian yang didapat lebih kepada yang belum mengenal Universitas Bumigora dan terlebih lagi diperuntukkan kepada calon mahasiswa baru. Wawancara dilakukan pada 8 Januari 2021 diruangan PMB dengan staf bagian PMB terkait dengan jalur pendaftaran, uang kuliah, persyaratan saat pandemi, beasiswa, dan masih banyak lagi.

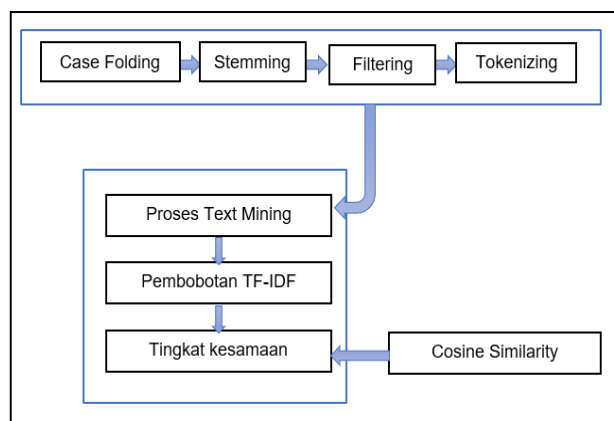
Berdasarkan hasil wawancara, 7 data sampel diambil untuk digunakan pada penelitian. Berikut uraian dari 7 data sampel :

Tabel 1 List FAQs.

Jawaban	Kata kunci
Siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora	Seleksi, SNMPTN.
Universitas Bumigora selalu memberikan informasi lowongan pekerjaan kepada mahasiswa	Lowongan pekerjaan.
Persyaratan daftar atau masuk Universitas Bumigora tidak berubah pasca pandemi Covid.	Syarat, covid.
Universitas Bumigora menyediakan bantuan keuangan seperti	Bantuan, Beasiswa.
<ol style="list-style-type: none"> 1. Beasiswa Bidik misi (KIP) 2. Peningkatan Prestasi Akademik (PPA) 3. Yayasan Pendidikan Eksekutif Komputer 	
Universitas Bumigora juga memiliki sistem UKT (Uang Kuliah Tunggal)	UKT, sistem, uang, kuliah, tunggal.
Jawaban	Kata kunci
Universitas Bumigora (UBG) menerima mahasiswa pindahan atau lanjutan dari internal maupun eksternal, jika kampus luar maka syaratnya akreditasi minimal B untuk kampus sebelumnya.	UBG, akreditasi, internal, eksternal.
Universitas Bumigora memiliki E-learning sendiri.	E-learning.

Data FAQ pada tabel 1. merupakan data sampel yang nantinya akan melalui beberapa proses untuk mendapatkan hasil dengan tingkat kesamaan (similarity) tertinggi sebagai output.

Berikut ini merupakan alur proses yang akan dilalui oleh data sampel pada tabel 1:



Gambar 1 Alur proses.

2.2. Preprocessing

Preprocessing merupakan proses text mining yang dilakukan untuk memperoleh informasi yang lebih berkualitas lebih tinggi dari text [7]. *Preprocessing* merupakan tahapan awal untuk menyiapkan teks menjadi data yang akan diolah lebih lanjut. Teks tidak bisa langsung di proses oleh algoritma pencarian, oleh karena itu dibutuhkan proses preprocessing text untuk mengubah teks menjadi data numerik [8][9][1].

Pada gambar 1 terdapat 4 tahapan pada *preprocessing* yaitu *case folding*, *stemming*, *filtering*, dan *tokenizing*.

A. Case Folding

Case Folding merupakan bentuk text *preprocessing* yang paling sederhana dan efektif. Case folding bertujuan untuk mengubah huruf-huruf dalam sebuah teks atau dokumen menjadi huruf kecil.

B. Stemming

Stemming adalah proses menghilangkan infleksi kata ke bentuk dasarnya. Contoh kata “mengatakan”, “katakan”, “dikatakan” akan ditransformasi menjadi kata “kata”.

C. Filtering

Filtering merupakan tahap untuk menyaring kata-kata kunci hasil dari proses tokenizing, proses ini menggunakan algoritma *Stopword*. *Stopword* adalah kumpulan kata-kata yang bersifat umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna penting sehingga tidak cocok untuk dijadikan sebagai kata kunci. Contoh *stopword* dalam bahasa Indonesia yaitu “di”, “yang”, “dan” dan masih banyak lagi. Tujuan menggunakan *stopword* yaitu menghapus atau mengeliminasi kata-kata yang tidak memiliki arti atau kurang penting sehingga tidak cocok untuk menjadi kata kunci.

D. Tokenizing

Tokenizing merupakan proses pemisahan teks atau kalimat menjadi potongan kata yang disebut token yang kemudian di analisa. *Tokenizing* juga dibagi menjadi 2 jenis, yaitu *tokenizing kata* dan *tokenizing kalimat*. *Tokenizing kata* memisah sebuah kalimat dengan hasil potongan-potongan kata dari kalimat yang dipotong yang kemudian dijadikan token, sedangkan *tokenizing kalimat* digunakan memisah kalimat-kalimat yang ada dalam sebuah paragraf.

2.3. Pembobotan Term Fequncy –Invers Document Frequency (TF_ID)

Tahap selanjutnya yaitu pembobotan TF-IDF yang digunakan untuk mengevaluasi seberapa pentingnya sebuah kata di dalam sebuah dokumen. Term Frequency (TF) yaitu semakin tinggi frekuensi kemunculan term pada sebuah dokumen maka akan semakin tinggi juga nilai bobot untuk term itu sendiri. Sementara itu, proses Inverse Document Frequency (IDF) merupakan kebalikan dari proses TF. Pada IDF, semakin tinggi frekuensi kemunculan term maka nilai bobot term itu sendiri akan semakin kecil. Berikut persamaan TF-IDF yang digunakan pada penelitian ini :

keterangan :

$w_{i,j}$ = bobot dokumen ke-i terhadap kata ke-j

$t_{f,i,j}$ = banyaknya kata i yang dicari pada sebuah dokumen j

N = total dokumen

$d_{f,i}$ = banyaknya dokumen yang mengandung kata ke-i

Jika frekuensi kemunculan term sama dengan jumlah dokumen, maka hasil perhitungan $IDF_j = 0$. Untuk menghindari hasil $w_{ij} = 0$, maka pada hasil perhitungan IDF akan ditambahkan dengan nilai 1 (satu). Library Sckit-Learn (Sklearn) merupakan salah satu library pada python yang digunakan untuk membantu proses pembobotan TF-IDF. Penambahan nilai 1 (satu) pada hasil perhitungan IDF dapat diantisipasi pada library sklearn.

2.4. Cosine Similarity

Metode *cosine similarity* merupakan metode yang digunakan untuk menghitung tingkat kesamaan (*similarity*) antar dua buah objek yang dinyatakan dalam dua buah vektor dengan menggunakan kata kunci (*keyword*) dari sebuah dokumen sebagai ukuran [5].

Tingkat *similarity* pada fungsi *cosine similarity* berkisar pada interval 0 (nol) sampai dengan 1 (satu). Nilai 0 (nol) melambangkan kedua objek sama sekali berbeda, sedangkan nilai 1 (satu) melambangkan kedua buah objek persis atau sama. Untuk mendapatkan hasil dari

fungsi tersebut dapat dilakukan normalisasi [10]. Berikut persamaan yang digunakan untuk menghitung tingkat kesamaan pada cosine similarity :

Keterangan :

$$x \cdot y = x \text{ dan } y \text{ dihitung dengan } \sum_{k=1}^n x_k y_k$$

$$\|x\| = \text{panjang vektor } x, \text{ dihitung dengan } \sum_{k=1}^n x_k^2$$

$$\|y\| = \text{panjang vektor } y, \text{ dihitung dengan } \sum_{k=1}^n y_k^2$$

3. Hasil dan Pembahasan

3.1. Persiapan Software

Software yang perlu disiapkan yaitu python sebagai bahasa pemrograman yang digunakan dan beberapa library pendukung seperti NLTK, Sklearn, dan Sastrawi.

3.2. Implementasi pada Python

Tahapan implementasi sistem pada python meliputi proses *preprocessing*, pembobotan TF-IDF, dan metode cosine similarity. Ketiga proses tersebut merupakan tahapan yang memang perlu dilakukan untuk mendapatkan hasil yang lebih maksimal.

Tabel 2 Hasil *case folding*.

Sebelum	Sesudah
Siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora	siswa yang gagal seleksi snmptn bisa mendaftar di universitas bumigora

Table 3 Hasil *stemming*.

Sebelum	Sesudah
Siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora	siswa yang gagal seleksi snmptn bisa daftar di universitas bumigora

Tabel 4 Hasil tokenizing

Sebelum	Sesudah
Siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora	siswa gagal seleksi daftar universitas bumigora

Tabel 5 Hasil tokenizing

Sebelum	Sesudah
Siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora	'siswa', 'gagal,' 'seleksi', 'snmptn', 'n', 'daftar', 'uni versitas', 'bumi gora

Berikut implementasi dari *preprocessing* :

- A. Pada tabel 2, teks sebelum melalui case folding “Siswa”, “SNMPTN”, “Universitas”, dan “Bumigora” masih memiliki huruf kapital, tetapi setelah melalui proses case folding huruf kapital akan berubah menjadi huruf kecil seperti “siswa”, “snmptn”, “universitas”, dan “bumigora” merupakan hasil yang didapat.
- B. Bisa dilihat pada tabel 3.2, pada kata “mendaftar” menjadi kata “daftar” setelah melalui proses stemming
- C. Kata yang di eliminasi dari kalimat pada tabel 3.3 adalah “yang”, “bisa”, dan “di”. Pada kalimat sebelum di filter masih mengandung kata “yang”, “bisa”, dan “di”, sementara kalimat setelah di filter sudah tidak mengandung kata tersebut.
- D. Bisa dilihat pada tabel 4, hasil kalimat setelah melalui proses tokenizing menjadi terpisah satu sama lain.

Tabel 2 sampai Tabel 5 menunjukkan hasil dari setiap tahapan preprocessing dengan salah satu kalimat sampel yang ada pada Tabel 6. Setelah ke-4 proses preprocessing dilakukan, proses akan berlanjut ke proses pembobotan TF-IDF.

Pada proses pembobotan TF-IDF, akan digunakan 2 buah sampel dari Tabel 1.

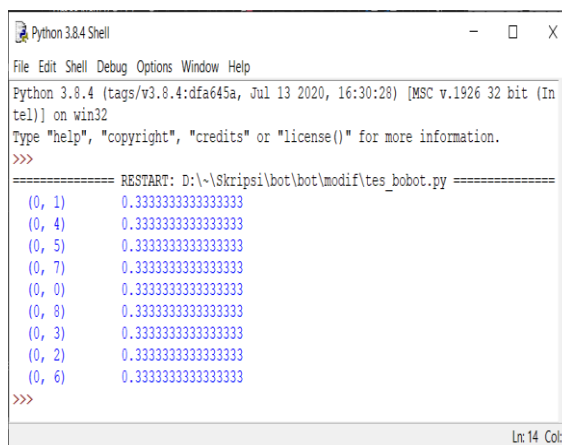
Tabel 6 Dua sampel

Sampel 1	Sampel 2
Siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora	Persyaratan daftar atau masuk Universitas Bumigora tidak berubah pasca pandemi Covid

Berikut hasil dari pembobotan yang dilakukan sampel 1 dan 2.

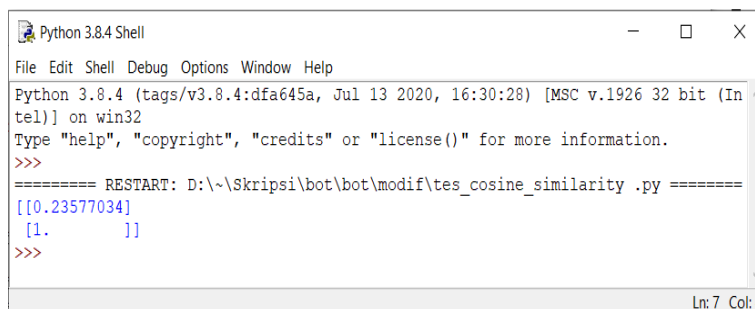


Gambar 2 Hasil pembobotan sampel 1



Gambar 3 Hasil pembobotan sampel 2

Pada gambar 1, angka di sebelah kanan menunjukkan nilai bobot untuk setiap kata yang ada pada kalimat sampel 1. Pada gambar 2, angka bagian kanan menunjukkan nilai bobot untuk setiap kata yang ada pada kalimat sampel 2. Perhitungan tingkat kesamaan antar dokumen dilakukan pada proses cosine similarity. Hasil dari pembobotan pada proses pembobotan TF-IDF yang akan digunakan untuk membandingkan tingkat kesamaan antar sampel 1 dengan sampel 2. Berikut hasil dari proses cosine similarity:



Gambar 4 Hasil similarity sampel 1 dan 2.

Gambar 4 merupakan hasil perhitungan *cosine similarity* berdasarkan bobot dari proses TF-IDF untuk sampel 1 dan sampel 2. Hasil akhir yang didapat yaitu 0.23577034, hasil tersebut karena angka tersebut sangat jauh dari 1 yang menjadi titik tingkat kemiripan tertinggi, maka sampel 1 dan sampel 2 bisa dikatakan tidak mirip.

3.3. Output Sistem

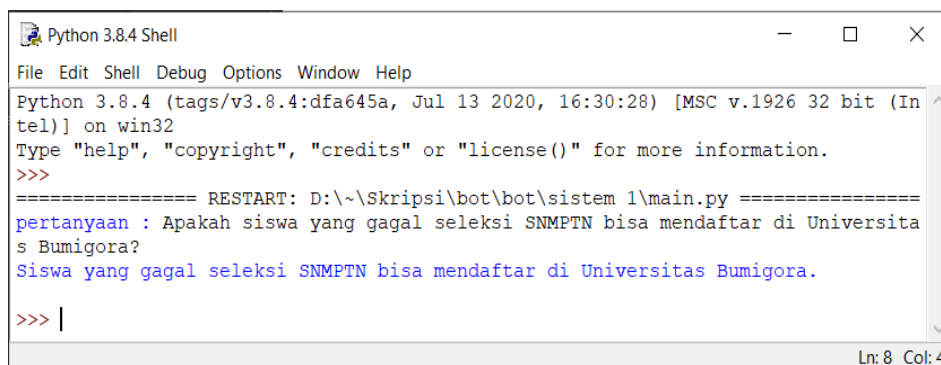
Percobaan dilakukan pada 7 data sampel dengan *keyword* berbeda. Perhatikan tabel berikut.

Tabel 7. Pertanyaan kuisioner

Query	Teks 1	Teks 2	Teks 3	Teks 4	Teks 5	Teks 6	Teks 7
Apakah siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora?	0.68	0.12	0.17	0.07	0.12	0.07	0.07
Universitas Bumigora selalu memberikan informasi lowongan pekerjaan kepada mahasiswa?	0.11	0.81	0.11	0.07	0.12	0.12	0.07
Apakah persyaratan daftar masuk Universitas Bumigora pasca pandemi berubah?	0.18	0.08	0.63	0.08	0.13	0.13	0.08
Apakah Universitas Bumigora menyediakan bantuan keuangan seperti beasiswa bidik misi PPA atau YPEK?	0.09	0.1	0.1	0.4	0.17	0.06	0.06
Apa Universitas Bumigora memiliki sistem UKT atau Uang Kuliah Tunggal?	0.11	0.07	0.11	0.12	0.81	0.07	0.16
Apakah Universitas Bumigora menerima mahasiswa pindahan kampus internal maupun eksternal?	0.11	0.07	0.11	0.07	0.12	0.59	0.07
Apakah Universitas Bumigora memiliki E-learning sendiri?	0.15	0.1	0.15	0.1	0.27	0.1	0.58

Tabel 7 menunjukkan seberapa mirip antara *query* dengan sampel teks kalimat jawaban FAQ. Contohnya pada *query* pertanyaan ke-2, nilai tingkat kemiripan tertinggi ada pada teks2 dengan nilai 0.81, sedangkan pada teks1 mendapat nilai kemiripan sebesar 0.11, nilai teks3 sebesar 0.11, pada teks4 sebesar 0.07, teks5 mendapat 0.12, nilai teks6 sebesar 0.12, dan pada teks7 mendapat nilai kemiripan sebesar 0.07.

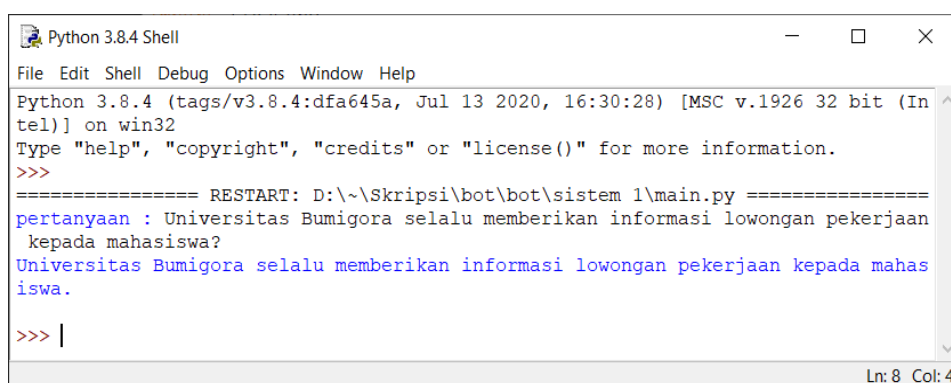
Berikut beberapa gambar yang memperlihatkan uji coba hasil pencarian kemiripan menggunakan pembobotan TF-IDF dan metode *cosine similarity* berdasarkan *query* pada tabel:



```
Python 3.8.4 Shell
File Edit Shell Debug Options Window Help
Python 3.8.4 (tags/v3.8.4:dfa645a, Jul 13 2020, 16:30:28) [MSC v.1926 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: D:\~\Skripsi\bot\bot\sistem 1\main.py =====
pertanyaan : Apakah siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora?
Siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora.
>>> |
```

Gamba 5. Hasil uji coba *similarity* sampel 1

Pertanyaan menggunakan *query* “Apakah siswa yang gagal seleksi SNMPTN bisa mendaftar di Universitas Bumigora?” memberikan jawaban yang sesuai dengan pertanyaan, yaitu informasi yang menyatakan bahwa siswa yang gagal dalam seleksi SNMPTN juga bisa mendaftar di Universitas Bumigora.



```
Python 3.8.4 Shell
File Edit Shell Debug Options Window Help
Python 3.8.4 (tags/v3.8.4:dfa645a, Jul 13 2020, 16:30:28) [MSC v.1926 32 bit (Intel)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: D:\~\Skripsi\bot\bot\sistem 1\main.py =====
pertanyaan : Universitas Bumigora selalu memberikan informasi lowongan pekerjaan kepada mahasiswa?
Universitas Bumigora selalu memberikan informasi lowongan pekerjaan kepada mahasiswa.
>>> |
```

Gambar 6. Hasil uji coba *similarity* sampel 2

Query dengan kata kunci “lowongan pekerjaan” juga memberikan jawaban yang sesuai dengan informasi yang dibutuhkan, yaitu informasi bahwa universitas selalu memberikan informasi lowongan pekerjaan kepada mahasiswa.

4. Kesimpulan

Berdasarkan hasil dan pembahasan pada bab sebelumnya, dapat peneliti simpulkan bahwa dengan menggunakan kombinasi pembobotan TF-IDF dan metode *Cosine Similarity* bisa mendapatkan tingkat akurasi hingga 64,28. Angka tersebut didapat dengan menjumlahkan akurasi tertinggi pada masing-masing *query* kemudian dibagi dengan jumlah *query* lalu dijadikan persen. Tingkat akurasi ini sangat dipengaruhi oleh bobot kata kunci yang terkandung pada *query*. Jika kemunculan kata pada dokumen tinggi maka kata tersebut menjadi penting, sehingga bobot hubungan antara *query* dengan dokumen akan semakin tinggi apabila frekuensi kemunculan kata tersebut tinggi pada dokumen. Seperti yang bisa dilihat pada tabel 4.7, tingkat kesamaan untuk *query* ke-7 dengan teks2 hanya 0,1 (10%), angka ini didapat karena pada teks2 hanya memiliki 2 kata kunci yang juga dimiliki oleh *query*, yaitu “Universitas” dan “Bumigora”. Sedangkan pada teks7 bisa mencapai 0,58 (58%) karena teks7 mengandung banyak kata kunci yang ada pada *query*.

Peneliti sangat menyadari banyaknya kekurangan dalam tulisan maupun penelitian ini, saran peneliti ke depannya yaitu penggabungan metode yang lebih bervariasi untuk

meningkatkan tingkat akurasi *similarity*, dan juga memiliki *interface* sehingga mudah digunakan.

Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih kepada pihak – pihak yang telah banyak membantu dalam penyelesaian penelitian terutama pihak penerimaan mahasiswa baru Universitas Bumigora.

Referensi

- [1] M. Rifauddin, “Pengelolaan Arsip Elektronik Berbasis Teknologi,” *Khizanah Al-Hikmah Jurnal Ilmu Perpustakaan, Informasi, dan Kearsipan*, vol. 4, no. 2, pp. 168–178, 2016.
- [2] E. Mulyawati, “Model Perilaku Pencarian Informasi guna Memenuhi Kebutuhan Informasi (Studi Literatur),” *Publis*, vol. 1, no. 2, pp. 14–20, 2011.
- [3] K. Marzuki and A. Apriani, “Evaluasi Penerapan Teknologi Informasi E-Learning Pada Kampus Swasta Menggunakan Cobit 4.1,” *Jurnal Bumigora Information Technology (BITE)*, vol. 1, no. 2, pp. 161–166, 2019.
- [4] S. Y. Bayquni, N. Kurniasih, and R. K. Anwar, “Pertukaran Informasi Oleh Mahasiswa Jurusan Ilmu Jurnalistik Melalui Media Kompasiana,” *Jurnal Kajian Informasi dan Perpustakaan*, vol. 3, no. 1, p. 71, 2015, doi: 10.24198/jkip.v3i1.9490.
- [5] O. Nurdiana, J. Jumadi, and D. Nursantika, “Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur’an Dalam Bahasa Indonesia,” *Jurnal Online Informatika*, vol. 1, no. 1, p. 59, 2016, doi: 10.15575/join.v1i1.12.
- [6] R. Afandi, “Sistem Penjawab FAQ (Frequently Asked Question) Seputar Universitas Bumigora Menggunakan Metode Pembobotan TD - IDF dan Jaccard Similarity,” pp. 1–11, 2020.
- [7] A. Deolika, K. Kusriani, and E. T. Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” *Jurnal Teknologi Informasi*, vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.
- [8] V. Amrizal, “Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim),” *Jurnal Teknik Informatika*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
- [9] I. Oyong, K. Marzuki, T. A. Lorosae, and Kusriani, “Prediksi Popularitas Artikel Berdasarkan,” in *Seminar Nasional Teknologi Informasi dan Multimedia 2018 ISSN: 2302-3805*, 2018, pp. 43–48, [Online]. Available: <https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/2055>.
- [10] M. E. Sulistyono, R. Saptono, A. Asshidiq, J. Informatika, and U. S. Maret, “Penilaian Ujian Bertipe Essay Menggunakan Metode Text Similarity,” vol. 12, no. 02, pp. 146–158, 2015.