

Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF

by Nuzul Hikmah

Submission date: 02-Apr-2023 07:16AM (UTC+0500)

Submission ID: 2053145811

File name: niversitas_Panca_Marga_Probolinggo_menggunakan_Metode_TF-IDF.pdf (744.07K)

Word count: 4733

Character count: 24941



Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF

Nuzul Hikmah ¹, Dyah Ariyanti ² dan Ferry Agus Pratama ³

¹ Universitas Panca Marga¹; n.hikmah1807@upm.ac.id

² Universitas Panca Marga²; dyahariyanti@upm.ac.id

³ Universitas Panca Marga³; ferryaguspratama@gmail.com

* Correspondence : n.hikmah1807@upm.ac.id

Abstract: Panca Marga University has an information service called Siakad (campus academic information system), but this service is only limited to information about academics and academics, where student information needs include many such as tuition fees, classroom location, lecturer ID. And other information that is not contained in the Siakad. Problems like this can be overcome by building customer service, but customer service has a shortage of service time, where operators cannot serve customers for 24 hours. To overcome this, the researcher provides a solution by building a system called a chatbot as a virtual assistant, so that the system will be able to serve customers for 24 hours without any human intervention. In the chatbot development process, researchers used the waterfall method as an application development method. For text processing, the researcher uses TF-IDF, VSM and Cosine Similarity using the Python programming language. In ensuring the validity of the data, the researchers used academic manuals and frequently asked questions by students using a questionnaire as a medium for collecting data. The end result of this research is a chatbot application that can be used as a virtual assistant as customer service in serving and providing information about the academics of the Panca Marga University of Probolinggo. Based on the results of the accuracy and UAT testing, the accuracy rate obtained by the chatbot reached 85.7% and the UAT testing in the first test reached 84.1% with a total of 30 respondents, in the second test it reached 82.1% with a total of 92 respondents.

Keywords: Chatbot, Tf-Idf, Vector Space Model, Cosine Similarity.

Abstrak: Universitas panca marga memiliki layanan informasi yang disebut dengan siakad (sistem informasi akademik kampus), namun layanan ini hanya sebatas informasi seputar akademik dan akademik, Dimana kebutuhan informasi mahasiswa meliputi banyak seperti biaya pembayaran spp, lokasi ruang kelas, nidn dosen. Dan lain lain informasi yang tidak terdapat pada siakad. Permasalahan seperti ini dapat diatasi dengan dibangunnya *customer service*, namun *customer service* memiliki kekurangan pada waktu pelayanan, dimana operator tidak dapat melayani *customer service* selama 24 jam. Untuk menanggulangi hal ini, peneliti memberikan solusi dengan membangun sebuah sistem yang dinamakan *chatbot* sebagai *asisten virtual*, sehingga system akan mampu melayani customer selama 24 jam tanpa ada campur tangan manusia. Dalam proses pembangunan chatbot, peneliti menggunakan metode waterfall sebagai metode pengembangan aplikasi. untuk pengolahan teks, peneliti menggunakan TF-IDF, VSM dan Cosine Similarity dengan menggunakan Bahasa pemrograman python. Dalam menjamin kevalidan data, peneliti menggunakan buku pedoman akademik dan pertanyaan yang seri diajukan mahasiswa dengan menggunakan kuesioner sebagai media pengumpulan datanya. Hasil akhir dari penelitian ini adalah aplikasi chatbot yang dapat dijadikan sebagai virtual assistant sebagai customer service dalam melayani dan menyediakan informasi seputar akademik Universitas Panca Marga

Sitasi: Hikmah, N.; Ariyanti, N.; Pratama, N. (2022). Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan metode TF-IDF. JTIM: Jurnal Teknologi Informasi Dan Multimedia, 4(2), 133-148. <https://doi.org/10.35746/jtim.v4i2.225>



Copyright: © 2022 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (<https://creativecommons.org/licenses/by-sa/4.0/>).

Probolinggo. Berdasarkan hasil pengujian akurasi dan UAT, tingkat akurasi yang didapatkan chatbot mencapai 85,7% dan pengujian UAT pada pengujian pertama mencapai 84,1% dengan total responden 30 responden, pada pengujian kedua mencapai 82,1% dengan total responden 92 responden.

Kata kunci: Chatbot, Tf-Idf, Vector Space Model, Cosine Similarity.

1. Pendahuluan

ChatBot adalah layanan obrolan berbasis AI (*Artificial Intelligence*), atau robot virtual yang dapat mensimulasikan percakapan manusia. Teknologi ini disebut juga dengan asisten digital yang dapat memahami dan memproses request pengguna serta memberikan respon yang relevan dengan cepat. Itulah sebabnya hal ini disebut dengan "ChatBot", karena mampu memahami dan memproses request pengguna secara otomatis.

Cara kerja dari ChatBot sangat sederhana, ChatBot akan memindai kata kunci dalam input yang dikirimkan yakni request pengguna, lalu ChatBot akan mencari kata kunci yang paling cocok atau dengan pola kata yang paling memiliki kemiripan dengan inputan dan akan mengirim respon secara otomatis berdasarkan dengan kata kunci tersebut.

Selama penulis menempuh pendidikan di Universitas Panca Marga Probolinggo, Universitas Panca Marga Probolinggo belum mengadaptasi Chatbot ini dalam sebuah pelayanannya. dimana Kampus Universitas Panca Marga Probolinggo belum memiliki layanan komunikasi yang dapat diakses secara interaktif selama 24 jam. *Customer service* yang ada di PTIK (Pusat Teknologi Informasi dan Komunikasi) Universitas Panca Marga Probolinggo belum memiliki *Customer service* yang melayani mahasiswa hingga 24 jam. Yang dimana kebutuhan mahasiswa untuk mendapat informasi tersebut tidak terikat oleh jam kerja, sedangkan PTIK (Pusat Teknologi Informasi dan Komunikasi) dapat melayani kebutuhan informasi dan komunikasi tersebut hanya dalam jam kerja.

Permasalahan Seperti ini dapat diatasi dengan mengadaptasi Teknologi Chatbot. Dimana Chatbot dimanfaatkan sebagai *Virtual Assistant* untuk menggantikan *Customer Service* sebagai sebuah layanan informasi dan komunikasi.

Penggunaan Chatbot tidak terlepas dari AI (*Artificial Intelligence*), AI (*Artificial Intelligence*) berguna untuk mengembangkan model percakapan Chatbot agar menjadi lebih natural. Penelitian yang dilakukan oleh suparno dalam jurnalnya yang berjudul "Chatbot sebagai implementasi pemanfaatan teknologi artificial intelligence dengan Channel Telegram". Yaitu penelitian tentang akses informasi sekolah yang diperuntukan kepada orang tua siswa untuk mendapatkan informasi sekolah menggunakan *Microsoft Azure* dan Channel Telegram.

Implementasi Chatbot pada pendaftaran mahasiswa baru menggunakan *Recurrent neural network*[1]. penelitian tentang implementasi chatbot untuk pendaftaran mahasiswa baru menggunakan metode RNN (Recurrent Neural Network). Aplikasi Chatbot (Milki Bot) yang terintegrasi dengan web CMS untuk Customer Service pada UKM MINSU[2]. Yaitu penelitian tentang implementasi Chatbot pada pelayanan pesanan pada UKM MINSU.

Peneliti menggunakan TF-IDF VSM (*Term Frequency dan Invers Document Frequency Vector Space Model*) sebagai penghitung bobot dokumen dan Cosine Similarity untuk melihat kemiripan query dengan dokumen.

TF-IDF memiliki hasil yang lebih baik dalam proses klasifikasi teks pendek dibandingkan dengan *document frequency difference*, *chi-square*, *bag-of-words* dan *n-grams*. TF-IDF dapat digunakan sebagai pembobotan kata menggunakan VSM (*Vector Space Model*) sehingga setiap dokumen memiliki nilai similarity antar dokumen, yang dapat digunakan sebagai masukan untuk proses klasifikasi.

Tujuan dari Penelitian ini adalah harapannya dapat mengimplementasikan chatbot sebagai virtual assistant sebagai *Customer Service* atau pelayanan komunikasi dan informasi di Universitas Panca Marga Probolinggo. Sehingga manfaat yang akan didapatkan

yaitu 1) Membantu peran Operator sebagai *Customer Service*. 2). Mendapat pelayanan secara real time yang dapat dilakukan melalui smartphone masing masing menggunakan aplikasi Whastapp. 3) Memberikan kemudahan bagi masyarakat kampus dalam mendapatkan Informasi. 4) Mendapatkan informasi Universitas Panca Marga.

2. Bahan dan Metode

2.1. Chatbot

Chatbot adalah layanan obrolan berbasis AI (kecerdasan buatan), atau robot virtual yang dapat mensimulasikan percakapan manusia. *Chatbot* secara harfiah berasal dari dua kata *chat* dan *Bot*.

Chatbot terdiri dari tiga kombinasi, dimana ketiga kombinasi ini membentuk sebuah ChatBot [3], diantaranya :

1. *User Interface* (Antarmuka pengguna)

Antarmuka pengguna dirancang untuk membuat pengguna merasa nyaman dan lebih mudah menggunakan chatBot.

2. *Artificial Intelegence* (Kecerdasan buatan)

AI atau kecerdasan buatan mengacu pada simulasi kecerdasan manusia membuat chatbot lebih pintar dalam menangkap setiap interaksi pengguna.

3. Integrasi

Integrasi dengan sistem lain dapat meningkatkan fitur yang kaya di chatbot. Pada perkembangan dunia bisnis atau teknologi saat ini, Chatbot menjadi salah satu solusi alternatif yang digunakan untuk memudahkan dalam melakukan intreraksi dengan pelanggan.

2.2. Text Preprocessing

Text Preprocessing merupakan teknik awal dalam data mining untuk mengubah data mentah atau biasa dikenal dengan raw data yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya[4]

Pada tahap *preprocessing* dilakukan penyiapan dokumen mentah menjadi dokumen atau representatif dokumen yang siap diproses untuk Langkah selanjutnya. Pada tahap ini proses untuk mengolah inputan dari pengguna dilakukan antara lain :

4. *Case Folding*

Case Folding adalah mengubah semua huruf menjadi huruf kecil. *Case Folding* hanya mengatasi huruf 'a' sampai dengan 'z', karakter Selain huruf akan dihilangkan atau tidak dianp[5].

Berikut contoh dari *case folding* :

Tabel 1. Hasil *case folding*.

Sebelum Case Folding	Sesudah Case Folding
Berapa biaya Pendaftaran di UNIVERSITAS PANCA MARGA ?	berapa biaya pendaftaran di universitas panca marga ?

5. *Tokenizing*

Tokenizing adalah Proses memisahkan kalimat menjadi susunan kata. Umumnya setiap kata teridentifikasi atau terpisah oleh karakter spasi, sehingga proses *tokenizing* mengadakan karakter spasi dalam melakukan pemisahan kata[4].

Berikut contoh dari *tokenizing* :

Tabel 2. Hasil Tokenizing.

Sebelum Tokenizing	Sesudah Tokenizing
berapa biaya pendaftaran di universitas panca marga ?	berapa biaya pendaftaran di universitas panca marga

28
6. *Stopword Removal*
Stopword Removal adalah suatu proses menghapus kata yang tidak perlu atau kata yang termasuk kedalam stopword [6], stopword adalah kata-kata yang biasanya dan sering muncul dan akan diabaikan dalam pemrosesan . Berikut beberapa contoh Stopword:

Tabel 3. Hasil Stopword removal.

Daftar Stopword
'ada', 'adalah', 'agak', 'agar', 'akan', 'amat', 'anda', 'antara', 'anu', 'apakah', 'apalagi', 'atau', 'bagaimanapun', 'bagi', 'bahwa', 'begitu', 'belum', 'bisa', 'boleh', 'dahulu', 'dalam', 'dan', 'dapat', 'dari', 'daripada', 'demi', 'demikian', 'dengan', 'di', 'dia', 'dimana', 'dll',

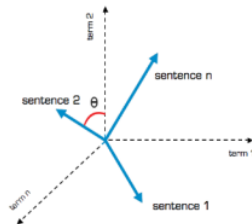
36
7. *Stemming*
Stemming adalah suatu proses dimana setiap kata akan dikembalikan menjadi kata dasarnya[4]. Berikut contoh dari *Stemming* :

Tabel 4. Hasil Stemming.

Sebelum Tokenizing	Sesudah Tokenizing
berapa	biaya
biaya	pendaftaran
pendaftaran	universitas
di	panca
universitas	marga
panca	
marga	

17
2.3. *TF-IDF dan VSM*

TF-IDF (Term Frequency – Inverse Document Frequency) adalah metode pembobotan kata dengan menentukan frekuensi kemunculan kata dalam sebuah dokumen[7]. TF-IDF melakukan pembobotan kata (term) pada setiap dokumen dengan memperhatikan frekuensi atau seringnya kata tersebut muncul dalam sebuah dokumen. TF-IDF melakukan pembobotan kata (term) pada setiap dokumen dengan memperhatikan frekuensi atau seringnya kata tersebut muncul dalam sebuah dokumen.



Gambar 1. vector space model.

Gambar 1. menunjukkan pemodelan dokumen teks di ruang dimensi dimana sentence adalah kalimat dokumen sedangkan term adalah term atau kata.

Sejumlah N kata yang berbeda sebagai kamus kata (*vocabulary*) atau indeks kata *terms*[index]. Kata-kata ini akan membentuk ruang vektor yang memiliki dimensi sebesar N . Setiap kata i dalam dokumen atau *query* diberikan bobot sebesar w_i [8].

2.3. Cosine Similarity

Cosine Similarity adalah metode yang digunakan untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek. Secara umum perhitungan metode ini didasarkan pada *vector space similarity measure*. Metode *cosine similarity* ini menghitung *similarity* antara dua buah objek (misalkan $D1$ dan $D2$) yang dinyatakan dalam dua buah vektor dengan menggunakan kata kunci dari sebuah dokumen sebagai ukuran[9].

2.4. Akurasi

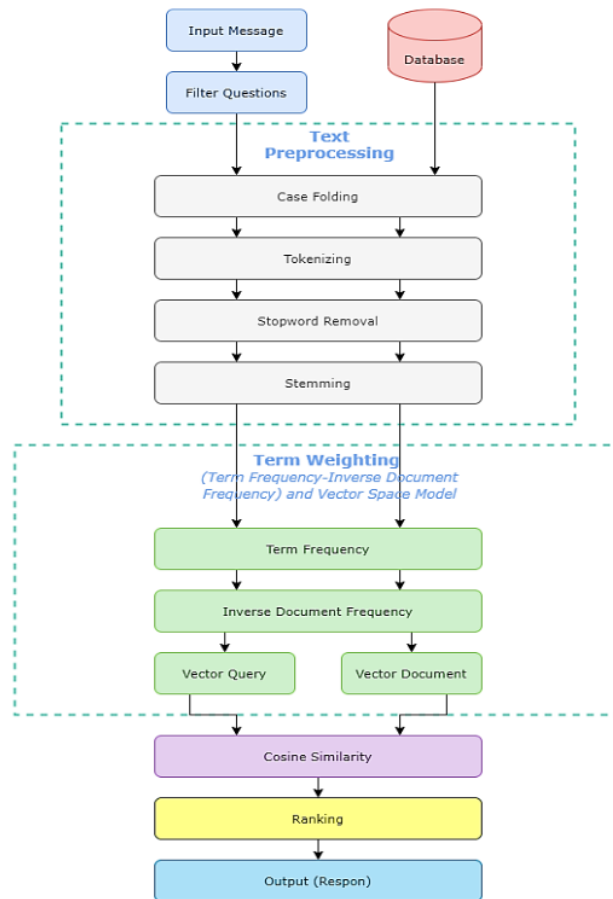
Akurasi merupakan kedekatan nilai yang diperoleh dari hasil pengukuran dengan nilai yang sebenarnya. Semakin dekat nilai hasil dari pengukuran, maka sistem dapat dikatakan semakin akurat. Nilai akurasi adalah nilai yang menggambarkan seberapa akurat hasil klasifikasi dari system[10].

2.3. User Acceptance Test

User Acceptance test adalah pengujian yang dilakukan untuk menyesuaikan perangkat lunak yang dibangun sudah sesuai dengan kebutuhan yang ada, dimana proses tersebut akan menentukan perangkat lunak yang telah dibangun dapat diterima atau tidak [3].

Sasaran dari UAT adalah untuk memberikan penjelasan tentang faktor-faktor penentu penerimaan computer yang bersifat umum, yang mampu menjelaskan perilaku pengguna diberbagai teknologi komputasi.

3. Hasil



34
Gambar 2. Arsitektur Sistem Chatbot

Gambar 2. merupakan arsitektur sistem chatbot, pada tahap ini dilakukan proses text preprocessing, pembobotan kata, menghitung kemiripan dokumen dan juga ranking berdasarkan nilai bobot tertinggi. Pada saat pengguna mengirimkan pesan (*query*) sistem akan memproses data dengan melalui tahap text preprocessing, kemudian data akan dihitung dengan metode *Tf-Idf* dan *VSM* serta *similarity* (kemiripan) menggunakan *cosine similarity*.

Pada studi kasus ini, penulis akan menggunakan data berikut sebagai sampel data latih dan data uji. Bagian ini dapat dibagi dengan *subheading*. Ini harus memberikan deskripsi singkat dan tepat tentang hasil eksperimen, interpretasinya, serta kesimpulan eksperimen yang dapat ditarik.

Tabel 5. Data Latih.

Data Latih		
No	Dokumen	Tipe Dokumen
1	alamat Univerisitas Panca Marga (UPM) ada di Jl Yos Sudarso, No 107 Pabean, Dringu – Probolinggo	dimana
2	ruang pendaftaran mahasiswa baru terletak digedung utara lantai I	dimana
3	Dekan Fakultas Teknik adalah Ir. Haryono, M.H.	siapa
4	Wakil Dekan I (Wadek I) Fakultas Teknik adalah Ahmad Izzuddin, S.T., M.Kom.	siapa
5	Wakil Dekan II (Wadek II) Fakultas Teknik adalah Ira Aprilia, S.Pd., M.Pd.	siapa
6	Wakil Dekan III (Wadek III) Fakultas Teknik adalah Tri Prihatingsih, S.T., M.T.	siapa
7	Ketua Program Studi (Kaprodi) Teknik Elektro adalah Nuzul Hikmah, S.T., M.Kom.	siapa
8	Ketua Program Studi (Kaprodi) Teknik Industri adalah Yustina Suhandini Tjahjaningsih, S.T., M.T.	siapa
9	Ketua Program Studi (Kaprodi) Teknik Mesin adalah Djoko Wahyudi, S.T., M.T.	siapa
10	Registrasi Administrasi adalah layanan terhadap mahasiswa untuk memperoleh status terdaftar pada suatu jenis program studi / jurusan tertentu dilingkungan Universitas Panca Marga	Apa
11	Registrasi Akademik adalah layanan terhadap mahasiswa untuk memperoleh hak dan ijin mengikuti perkuliahan pada suatu program studi / jurusan tertentu dilingkungan Universitas Panca Marga.	apa
12	Registrasi Ulang atau Heregisi Adalah Registrasi bagi mahasiswa lama	apa
13	Registrasi Harus dilakukan untuk memberikan status terdaftar kepada mahasiswa, sehingga berhak menggunakan semua fasilitas yang ada di Universitas Panca Marga	kenapa
14	Penyelenggaraan ujian dimaksudkan untuk 1) Menilai apakah mahasiswa telah memahami atau menguasai bahan bahasan yang disajikan dalam kuliah. 2) Mengelompokkan mahasiswa ke dalam beberapa golongan berdasarkan kemampuan. 3) Menilai apakah bahan kuliah disajikan sesuai dalam kurikulum. 4) Mengetahui apakah cara penyajian dosen cukup baik.	kenapa
Data Uji		
No	Query	
1	Siapa ketua program studi Teknik elektro ?	

Sebelum dilakukan perhitungan pembobotan kata dan menghitung tingkat similarity antara document dengan *query*, *document* akan melewati tahap *filter questions* terlebih dahulu berdasarkan kata tanya pada *query*, hal ini berguna untuk mempersempit perhitungan dan meningkatkan tingkat atau nilai similarity yang akan dihasilkan. Karena *document* yang akan diproses hanya *document* dengan *document type* berdasarkan *query* yang ada.

Pada studi kasus ini kata tanya pada *query* adalah kata “siapa”, maka *document* yang akan diproses merupakan *document* dengan *document type* “siapa”.

Berikut ini adalah proses perhitungan menggunakan metode *Tf-Idf*, *VSM* dan *Cosine Similarity* berdasarkan data yang sudah melalui tahap preprocessing sesuai arsitektur sistem *chatbot* pada Gambar 2.

3.1. Term Frequency

$$tf = tf_{ij} \tag{1}$$

Keterangan :

tf = Term frequency

tf_{ij} = Banyaknya muncul term (kata) i pada dokumen j

Tabel 6. Hasil Term Frequency.

Term	D1	D2	D3	D4	D5	D6	D7
dekan	1	1	1	1	0	0	0
fakultas	1	1	1	1	0	0	0
teknik	1	1	1	0	0	0	0
ir	1	0	0	0	0	0	0
haryono	1	0	0	0	0	0	0
wakil	0	1	1	1	0	0	0
wadek	0	1	1	1	0	0	0
ahmad	0	1	0	0	0	0	0
izzuddin	0	1	0	0	0	0	0
kom	0	1	1	0	1	1	0
ii	0	0	2	0	0	0	0
ira	0	0	1	0	0	0	0
aprilia	0	0	1	0	0	0	0
pd	0	0	2	0	0	0	0
iii	0	0	0	2	0	0	0
tri	0	0	0	1	0	0	0
prihatingsih	0	0	0	1	0	0	0
ketua	0	0	0	0	1	1	1
program	0	0	0	0	1	1	1
studi	0	0	0	0	1	1	1
kaprodi	0	0	0	0	1	1	1
elektro	0	0	0	0	1	0	0
nuzul	0	0	0	0	1	0	0
hikmah	0	0	0	0	1	0	0
industri	0	0	0	0	0	1	0
yustina	0	0	0	0	0	1	0
suhandini	0	0	0	0	0	1	0
tjahjaningsih	0	0	0	0	0	1	0
mesin	0	0	0	0	0	0	1
djoko	0	0	0	0	0	0	1
wahyudi	0	0	0	0	0	0	1

3.2. Inverse Document Frequency

$$idf_i = \log\left(\frac{N}{D_{fi}}\right) \tag{2}$$

Keterangan :

idf_i = inverse document frequency

N = Jumlah dokumen

D_{fi} = Jumlah kemunculan term i terhadap Dokumen

Tabel 7. Hasil Inverse Document Frequency.

Term	bobot
dekan	0.24303804868629
fakultas	0.24303804868629
teknik	0.0
ir	0.8450980400142
haryono	0.8450980400142
wakil	0.36797678529459
wadek	0.36797678529459
ahmad	0.84509804001425
izzuddin	0.84509804001425
kom	0.54406804435027
ii	0.84509804001425
ira	0.84509804001425
aprilia	0.84509804001425
pd	0.84509804001425
iii	0.84509804001425
tri	0.84509804001425
prihatingsih	0.84509804001425
ketua	0.36797678529459
program	0.36797678529459
studi	0.36797678529459
kaprodi	0.36797678529459
elektro	0.84509804001425
nuzul	0.84509804001425
hikmah	0.84509804001425
industri	0.84509804001425
yustina	0.84509804001425
suhandini	0.84509804001425
tjahjaningsih	0.84509804001425
mesin	0.84509804001425
djoko	0.84509804001425
wahyudi	0.84509804001425

3.2. Term Weighting

$$w_{ij} = tf_{ij} \times \log\left(\frac{N}{D_{fi}}\right) \tag{3}$$

Atau

$$w_{ij} = tf \times idf \tag{4}$$

Keterangan:

W_{ij} = Bobot term terhadap dokumen

tf = term frequency

idf = Inverse document frequency

Tabel 8. Hasil Term Weighting.

Term	D1	D2	D3	D4	D5	D6	D7
dekan	0.243	0.243	0.243	0.243	0.0	0.0	0.0
fakultas	0.243	0.243	0.243	0.243	0.0	0.0	0.0
teknik	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ir	0.845	0.0	0.0	0.0	0.0	0.0	0.0
haryono	0.845	0.0	0.0	0.0	0.0	0.0	0.0
wakil	0.0	0.367	0.367	0.367	0.0	0.0	0.0
wadek	0.0	0.367	0.367	0.367	0.0	0.0	0.0
ahmad	0.0	0.845	0.0	0.0	0.0	0.0	0.0
izzuddin	0.0	0.845	0.0	0.0	0.0	0.0	0.0
kom	0.0	0.544	0.0	0.0	0.544	0.0	0.0
ii	0.0	0.0	1.690	0.0	0.0	0.0	0.0
ira	0.0	0.0	0.845	0.0	0.0	0.0	0.0
aprilia	0.0	0.0	0.845	0.0	0.0	0.0	0.0
pd	0.0	0.0	1.690	0.0	0.0	0.0	0.0
iii	0.0	0.0	0.0	1.690	0.0	0.0	0.0
tri	0.0	0.0	0.0	0.845	0.0	0.0	0.0
prihatingsih	0.0	0.0	0.0	0.845	0.0	0.0	0.0
ketua	0.0	0.0	0.0	0.0	0.367	0.367	0.367
program	0.0	0.0	0.0	0.0	0.367	0.367	0.367
studi	0.0	0.0	0.0	0.0	0.367	0.367	0.367
kaprodi	0.0	0.0	0.0	0.0	0.845	0.367	0.367
elektro	0.0	0.0	0.0	0.0	0.845	0.0	0.0
nuzul	0.0	0.0	0.0	0.0	0.845	0.0	0.0
hikmah	0.0	0.0	0.0	0.0	0.845	0.0	0.0
industri	0.0	0.0	0.0	0.0	0.0	0.845	0.0
yustina	0.0	0.0	0.0	0.0	0.0	0.845	0.0
suhandini	0.0	0.0	0.0	0.0	0.0	0.845	0.0
tjahjaningsih	0.0	0.0	0.0	0.0	0.0	0.845	0.0
mesin	0.0	0.0	0.0	0.0	0.0	0.0	0.845
djoko	0.0	0.0	0.0	0.0	0.0	0.0	0.845
wahyudi	0.0	0.0	0.0	0.0	0.0	0.0	0.845

3.2. Cosine Similarity

$$\cos = \frac{A \cdot B}{|A||B|} \tag{5}$$

keterangan :

A = Vektor A, yang akan dibandingkan kemiripannya

B = Vektor B, yang akan dibandingkan kemiripannya

|A| = Panjang vektor A

|B| = Panjang vektor B

11
Tabel 8. Hasil Cosine Similarity.

Dokumen	Hasil Similarity
D1	0.0
D2	0.0
D3	0.0
D4	0.0
D5	0.613148046811475
D6	0.20817889174368312
D7	0.23424261815489641

3.2. Pengujian Tingkat akurasi

Tabel 9. Hasil Cosine Similarity.

Tipe	Benar	Salah	Jumlah
Apa	4	1	5
Dimana	19	1	20
Kapan	3	2	5
Kenapa	4	1	5
Siapa	18	2	20
Bagaimana	4	1	5
Berapa	8	2	10
Total	60	10	70

Sehingga dari hasil pengujian diatas dapat dihitung menggunakan rumus akurasi sebagai berikut :

$$\begin{aligned}
 \text{akurasi} &= \frac{\text{Jumlah jawaban benar}}{\text{Jumlah total pertanyaan}} 100\% \\
 \text{akurasi} &= \frac{60}{70} 100\% \\
 \text{akurasi} &= 85,7\%
 \end{aligned}
 \tag{6}$$

3.2. Pengujian UAT

1. pengujian pertama

Tabel 10. Hasil Pengujian UAT Pertama .

No	Jenis Pengujian	SS	S	RR	TS	STS
1 Pengujian Kinerja						
1	Apakah anda setuju respon time (waktu respon) chatbot cukup cepat ?	14	16	-	-	-
2	Apakah anda setuju respon yang diberikan chatbot sesuai dengan pertanyaan yang anda ajukan ?	6	20	4	-	-

No Jenis Pengujian						
1 Pengujian Kinerja						
No	Pertanyaan	SS	S	RR	TS	STS
3	Apakah anda setuju <i>chatbot</i> dapat memperkenalkan diri dengan baik sehingga pengguna mengerti dan memahami cara penggunaan dari <i>chatbot</i> ?	10	15	4	1	-
2 Pengujian Kepuasan Pengguna						
1	Apakah anda setuju <i>chatbot</i> dapat membantu anda mendapatkan informasi seputar akademik ?	12	17	1	-	-
2	Apakah anda setuju <i>chatbot</i> ini mempermudah dan mempercepat anda mendapatkan informasi seputar akademik universitas panca marga probolinggo ?	10	19	-	1	-
3	Apakah anda setuju <i>chatbot</i> dapat menjadi prioritas utama anda sebagai virtual assistant untuk mendapatkan informasi seputar akademik universitas panca marga probolinggo ?	6	17	6	1	-
4	Apakah anda setuju <i>chatbot</i> dapat membantu anda mendapatkan informasi seputar akademik ?	12	17	1	-	-

berdasarkan data pada Tabel 10, dapat dianalisis hasil perhitungan sebagai berikut :

- Jumlah skor dari responden yang menjawab SS = $58 \times 5 = 290$
- Jumlah skor dari responden yang menjawab S = $104 \times 4 = 416$
- Jumlah skor dari responden yang menjawab RR = $15 \times 3 = 45$
- Jumlah skor dari responden yang menjawab TS = $3 \times 2 = 6$
- Jumlah skor dari responden yang menjawab STS = $0 \times 1 = 0$

Jumlah skor Total		= 757
--------------------------	--	--------------

Presentase menjawab SS	: $290 / 757 \times 100\%$	= 38.3 %
Presentase menjawab S	: $416 / 757 \times 100\%$	= 54.9 %
Presentase menjawab RR	: $45 / 757 \times 100\%$	= 5.9 %
Presentase menjawab TS	: $6 / 757 \times 100\%$	= 0.7 %
Presentase menjawab STS	: $0 / 757 \times 100\%$	= 0

Hasil jawaban dari responden sebanyak 30 orang, kemudian dapat dihitung nilai tertinggi dan terendah seperti berikut :

Nilai Tertinggi = $30 \times 6 \times 5 = 900$ (seandainya semua menjawab SS).

Nilai Terendah = $30 \times 6 \times 1 = 180$ (seandainya semua menjawab SS).

Berdasarkan perhitungan yang menyatakan nilai tertinggi adalah 900, dapat dicari presentase sebagaimana berikut :

$$\begin{aligned} \text{Presentase} &= \text{Jumlah Skor total} / \text{Nilai tertinggi} \times 100\% \\ &= 757 / 900 \times 100\% \\ &= 84,1\% \end{aligned}$$

2. pengujian kedua

Tabel 11. Hasil Pengujian UAT Kedua.

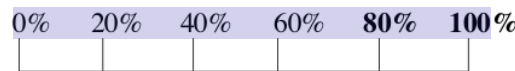
No Jenis Pengujian						
1 Pengujian Kinerja						
No	Pertanyaan	SS	S	RR	TS	STS
1	Apakah anda setuju respon time (waktu respon) chatbot cukup cepat ?	27	59	5	1	-
2	Apakah anda setuju respon yang diberikan chatbot sesuai dengan pertanyaan yang anda ajukan ?	15	57	16	4	-
3	Apakah anda setuju chatbot dapat memperkenalkan diri dengan baik sehingga pengguna mengerti dan memahami cara penggunaan dari chatbot ?	21	50	18	3	-
4	Apakah anda setuju chatbot dapat berjalan dengan baik ?	17	59	15	1	-
2 Pengujian Kepuasan Pengguna						
1	Apakah anda setuju chatbot dapat membantu anda mendapatkan informasi seputar Akademik Universitas Panca Marga?	7	68	17	-	-
2	Apakah anda setuju chatbot ini mempermudah dan mempercepat anda mendapatkan informasi seputar akademik universitas panca marga probolinggo ?	74	9	9	-	-
3	Apakah anda setuju chatbot dapat menjadi prioritas utama anda sebagai virtual assistant untuk mendapatkan informasi seputar akademik universitas panca marga probolinggo ?	5	46	26	15	-

No Jenis Pengujian						
1 Pengujian Kinerja						
No	Pertanyaan	SS	S	RR	TS	STS
4	Apakah anda setuju chatbot ini bermanfaat apabila digunakan sebagai asisten virtual ?	74	10	8	-	-

Pada Tabel 11 merupakan hasil *UAT* pengujian kedua, pada pengujian diambil 92 responden sebagai data penguji. Berdasarkan rumus sebelumnya yang digunakan untuk menghitung *UAT* pada Tabel 10, maka didapatkan hasil presentase sebesar 82,1%.

Nilai 84,1% dan nilai 82,1% merupakan nilai yang cukup tinggi sehingga berdasarkan pengujian user acceptance test tergolong sangat kuat.

Hal ini dapat dilihat pada Gambar 3 di bawah ini :

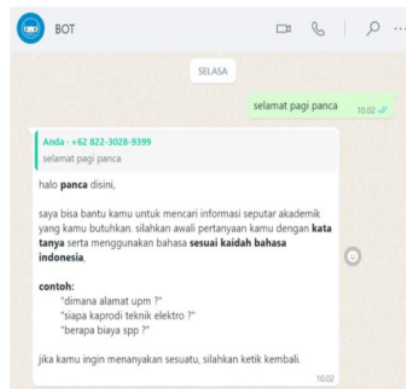


Gambar 3. Tingkat usability sistem pengujian *UAT*.

Keterangan :

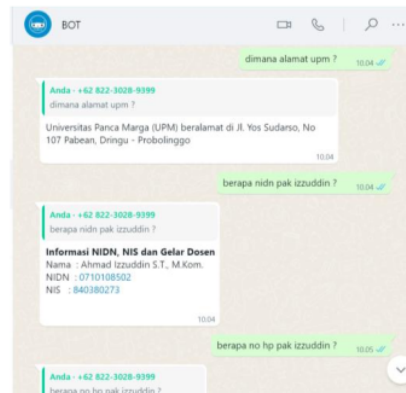
- 0 – 20 % = Sangat Lemah
- 20% - 40% = Lemah
- 41% - 60% = Cukup
- 61% - 80% = Kuat
- 80% - 100% = Sangat Kuat

Berikut merupakan beberapa dokumentasi percakapan antara pengguna dengan chatbot .



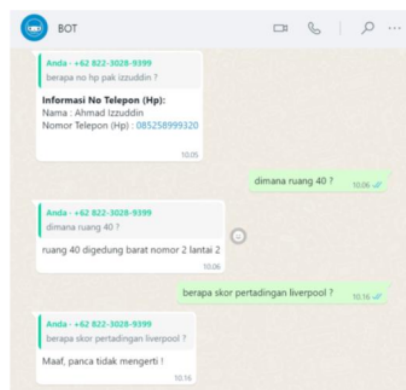
Gambar 4. Tingkat usability sistem pengujian *UAT*.

Gambar 4. menunjukkan percakapan antara pengguna dengan chatbot, dimana chatbot mengirimkan pesan default untuk memperkenalkan diri kepada pengguna, Hal ini bisa terjadi karena pengguna baru pertama kali menggunakan chatbot.



Gambar 5. percakapan pengguna dengan chatbot.

Gambar 5 Menunjukkan dimana pengguna mengajukan beberapa pertanyaan kepada chatbot. respon time dari percakapan tersebut tidak sampai 1 (satu) menit, hal ini menunjukkan bahwa chatbot sangat responsive dalam merespon pengguna.



Gambar 6. percakapan pengguna dengan chatbot.

Gambar 6. menunjukkan respon default bahwa chatbot tidak mengerti pertanyaan dari pengguna, Hal ini terjadi karena perhitungan antara query dan dokumen yang ada didalam database tidak menemukan kemiripan, maka secara otomatis chatbot akan mengirimkan respon default "Maaf, panca tidak mengerti!" untuk menunjukkan bahwa query atau pertanyaan dari pengguna tidak ditemukan atau tidak memiliki topik yang dianggap mirip dengan dokumen yang terdapat didalam database.

4. Kesimpulan

Bersadarkan hasil pengujian tingkat akurasi yang mencapai 85,7% dan pengujian UAT pada tahap pertama mencapai 84,1%, pada pengujian tahap kedua mencapai 82,1%, maka dapat disimpulkan bahwa chatbot membantu peranan operator sebagai *customer service* dengan memanfaatkan metode *tf-idf*, *vsm* dan *cosine similarity*.

Tingkat akurasi dapat ditingkat berdasarkan semakin banyak nya data yang terdapat dalam database, semakin banyak data maka pertanyaan akan semakin banyak yang terjawab oleh chatbot.

Selain itu term, pada suatu document sangat mempengaruhi tingkat keberhasilan yang dihasilkan oleh metode Tf-Idf dan Cosine Similarity, karena metode tersebut mengukur atau menghitung bobot kata pada sebuah document kemudian menghitung tingkat kemiripan document terhadap query berdasarkan nilai atau skor yang diperoleh dari pembobotan kata.

Referensi

- [1] T. A. Zuraiyah, D. K. Utami, and D. Herlambang, "IMPLEMENTASI CHATBOT PADA PENDAFTARAN MAHASISWA BARU MENGGUNAKAN RECURRENT NEURAL NETWORK," *J. Ilm. Teknol. Dan Rekayasa*, vol. 24, no. 2, pp. 91–101, 2019, doi: 10.35760/tr.2019.v24i2.2388.
- [2] A. Dwi R, "APLIKASI CHATBOT (MILKI BOT) YANG TERINTEGRASI DENGAN WEB CMS UNTUK CUSTOMER SERVICE PADA UKM MINSU," *J. Cendikia*, vol. XVI, 2018.
- [3] A. T. Nugraha, "JURUSAN TEKNIK INFORMATIKA FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG 2020," p. 81.
- [4] N. Hikmah, "Jurnal ATLAS NUSANTARA Teknik Informatika Volume 8, Nomor 1, April 2018," p. 14.
- [5] M. A. Rofiqi, Abd. C. Fauzan, A. P. Agustin, and A. A. Saputra, "Implementasi Term-Frequency Inverse Document Frequency (TF-IDF) Untuk Mencari Relevansi Dokumen Berdasarkan Query," *Ilk. J. Comput. Sci. Appl. Inform.*, vol. 1, no. 2, pp. 58–64, Dec. 2019, doi: 10.28926/ilkomnika.v1i2.18.
- [6] S. Mulyatun, H. Utama, and A. Mustopa, "PENDEKATAN NATURAL LANGUAGE PROCESSING PADA APLIKASI CHATBOT SEBAGAI ALAT BANTU CUSTOMER SERVICE," *J. Inf. Syst. Manag. JOISM*, vol. 3, no. 1, pp. 12–17, Jan. 2021, doi: 10.24076/JOISM.2021v3i1.404.
- [7] Pande Made Risky Cahya Dinatha and Nur Aini Rakhmawati, "Komparasi Term Weighting dan Word Embedding pada Klasifikasi Tweet Pemerintah Daerah," *J. Nas. Tek. Elektro Dan Teknol. Inf.*, vol. 9, no. 2, pp. 155–161, May 2020, doi: 10.22146/jnteti.v9i2.90.
- [8] I. R. Priandono, M. Hakimah, and N. F. Rozi, "Implementasi Vector Space Model Dengan Pembobotan Berbasis Kelas Pada Mesin Pencari Dokumen Skripsi," *J. Inform.*, p. 5, 2020.
- [9] D. Kurniadi, S. F. C. Haviana, and A. Novianto, "Implementasi Algoritma Cosine Similarity pada sistem arsip dokumen di Universitas Islam Sultan Agung," *J. Transform.*, vol. 17, no. 2, p. 124, Jan. 2020, doi: 10.26623/transformatika.v17i2.1613.
- [10] F. Azhmi, "APLIKASI CHATBOT SEKRETARIS JURUSAN BERBASIS ARTIFICIAL INTELLIGENCE MARKUP LANGUAGE (AIML)," in *Skripsi*, UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU PEKANBARU, 2020.

Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF

ORIGINALITY REPORT

20%

SIMILARITY INDEX

18%

INTERNET SOURCES

8%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1	jurnal.fkip-uwgm.ac.id Internet Source	2%
2	core.ac.uk Internet Source	2%
3	Rio Feriangga Kurniawan. "IMPLEMENTASI TEXT MINING MENGGUNAKAN METODE COSINE SIMILARITY UNTUK KLASIFIKASI KONTEN BERITA DI POSTINGAN GRUP FACEBOOK INFO LANTAS DAN KRIMINAL PASURUAN", JAMI: Jurnal Ahli Muda Indonesia, 2022 Publication	1%
4	students.if.itb.ac.id Internet Source	1%
5	Submitted to Universitas Kristen Duta Wacana Student Paper	1%
6	www.slideshare.net Internet Source	1%

peternakan.unitri.ac.id

7	Internet Source	1 %
8	repository.uksw.edu Internet Source	1 %
9	www.researchgate.net Internet Source	1 %
10	e-journal.unper.ac.id Internet Source	1 %
11	text-id.123dok.com Internet Source	1 %
12	repository.ummat.ac.id Internet Source	<1 %
13	repository.nusamandiri.ac.id Internet Source	<1 %
14	journal.ugm.ac.id Internet Source	<1 %
15	ecampus.pelitabangsa.ac.id Internet Source	<1 %
16	repository.uir.ac.id Internet Source	<1 %
17	conference.upnvj.ac.id Internet Source	<1 %
18	media.neliti.com Internet Source	<1 %

19	repository.unib.ac.id Internet Source	<1 %
20	Submitted to Universitas Jember Student Paper	<1 %
21	Submitted to Deptford Township High School Student Paper	<1 %
22	Submitted to Universitas Islam Indonesia Student Paper	<1 %
23	moam.info Internet Source	<1 %
24	Submitted to Universitas Islam Majapahit Student Paper	<1 %
25	dspace.uui.ac.id Internet Source	<1 %
26	warstek.com Internet Source	<1 %
27	ejournal.stkip-mmb.ac.id Internet Source	<1 %
28	1library.net Internet Source	<1 %
29	ejournal.gunadarma.ac.id Internet Source	<1 %
30	researchbank.rmit.edu.au Internet Source	<1 %

31	Submitted to Universitas Gunadarma Student Paper	<1 %
32	eprints.umm.ac.id Internet Source	<1 %
33	fisip.upm.ac.id Internet Source	<1 %
34	Ida Bagus Gede Sarasvananda, Ida Bagus Ary Indra Iswara. "Tanda Tangan Elektronik Menggunakan Algoritma Rivest Shamir Adleman (RSA) pada Sistem Informasi Surat Menyurat LPIK INSTIKI", Jurnal Sisfokom (Sistem Informasi dan Komputer), 2022 Publication	<1 %
35	id.123dok.com Internet Source	<1 %
36	journal.untar.ac.id Internet Source	<1 %
37	docobook.com Internet Source	<1 %
38	www.nor-ijournal.com Internet Source	<1 %
39	Fitra Adhitya Sandi, Laila Septiana. "Perancangan Sistem Informasi Point Of Sales Berbasis Web Pada Perusahaan Sablon", Computer Science (CO-SCIENCE), 2021 Publication	<1 %

40

Markides. Encyclopedia of Health and Aging

Publication

<1 %

41

Muhd. A.R. Imam Riauan, Tessa Shasrini.
"Dampak Komunikasi Terapeutik terhadap
Citra Pelayanan Kesehatan (Studi Kasus di
Rumah Sakit Umum Daerah Arifin Achmad
Pekanbaru)", Jurnal The Messenger, 2017

Publication

<1 %

42

repositori.uin-alauddin.ac.id

Internet Source

<1 %

Exclude quotes On

Exclude matches < 2 words

Exclude bibliography On