

Optimisasi Chatbot Virtual Assistant di Stasiun Kereta Api KAI Cikarang Menggunakan *TF-IDF* dan *Cosine similarity* untuk Peningkatan Layanan Penumpang

Optimization of Chatbot Virtual Assistant at KAI Cikarang Train Station Using TF-IDF and Cosine similarity for Passenger Service Improvement

Alvian S. Pratama¹, Choky Andriano², Fadli Ramadan³, Viyan Qomarudin Noor⁴

¹Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

²Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

³Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

⁴Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa

¹vian0204@mhs.pelitabangsa.ac.id, ²choky29@mhs.pelitabangsa.ac.id, ³fadli.R10@mhs.pelitabangsa.ac.id,

⁴viyanqn21@mhs.pelitabangsa.ac.id

Abstract

The increasing number of train users at KAI Cikarang Station requires fast and accurate information services. The main problem is the limited resources to provide information manually. To overcome this, this research develops a chatbot based on TF-IDF (Term Frequency-Inverse Document Frequency) and cosine similarity methods to provide automated information services related to departure schedules, ticket prices, and ticket purchase procedures.

The dataset is taken from an official source, <https://commuterline.id> which provides information related to train services. The data is processed through text preprocessing, weighting with TF-IDF, and query matching using cosine similarity. The chatbot is implemented through Telegram by utilizing the BotFather feature, as well as Python integration to support system logic.

The test results show that the chatbot is able to provide relevant answers with an 80% accuracy rate, calculated from twenty test questions with 16 appropriate answers and 4 inappropriate answer. This accuracy rate shows the effectiveness of TF-IDF and cosine similarity methods in matching user questions with relevant answers. This chatbot not only improves the efficiency of information services at the station, but also provides a better user experience. This research offers an innovative solution to improve the quality of public transportation services in Indonesia.

Keywords: chatbot, TF-IDF, cosine similarity, virtual assistant, train service, text preprocessing

Abstrak

Peningkatan jumlah pengguna kereta api di Stasiun KAI Cikarang membutuhkan layanan informasi yang cepat dan akurat. Masalah utama adalah keterbatasan sumber daya untuk memberikan informasi secara manual. Untuk mengatasi hal tersebut, penelitian ini mengembangkan chatbot berbasis metode *TF-IDF* (Term Frequency-Inverse Document Frequency) dan *cosine similarity* untuk menyediakan layanan informasi otomatis terkait jadwal keberangkatan, harga tiket, dan prosedur pembelian tiket.

Dataset diambil dari sumber resmi, yaitu <https://commuterline.id> yang menyediakan informasi terkait layanan kereta api. Data tersebut diproses melalui tahapan *text preprocessing*, pembobotan dengan *TF-IDF*, dan pencocokan pertanyaan menggunakan *cosine similarity*. Chatbot diimplementasikan melalui Telegram dengan memanfaatkan fitur BotFather, serta integrasi *Python* untuk mendukung logika sistem.

Hasil pengujian menunjukkan chatbot mampu memberikan jawaban yang relevan dengan tingkat akurasi 80%, dihitung dari dua puluh pertanyaan uji dengan 16 jawaban sesuai dan 4 jawaban tidak sesuai. Tingkat akurasi ini menunjukkan efektivitas metode *TF-IDF* dan *cosine similarity* dalam mencocokkan pertanyaan pengguna dengan jawaban yang relevan. Chatbot ini tidak hanya meningkatkan efisiensi layanan informasi di stasiun, tetapi juga memberikan pengalaman pengguna yang lebih baik. Penelitian ini menawarkan solusi inovatif untuk meningkatkan kualitas layanan transportasi publik di Indonesia.

Kata kunci: chatbot, *TF-IDF*, *cosine similarity*, asisten virtual, layanan kereta api, *text preprocessing*

Pendahuluan

Stasiun kereta api adalah salah satu fasilitas transportasi yang sering kali menjadi tempat bertemunya berbagai kebutuhan informasi bagi para penumpang[1]. Berdasarkan data PT Kereta Api Indonesia (KAI), kereta api merupakan transportasi nasional yang paling banyak digunakan masyarakat Indonesia dengan jumlah penumpang kereta api mencapai lebih dari 34 juta penumpang pada periode Januari 2024. Semakin meningkatnya jumlah pengguna transportasi kereta api di Indonesia, termasuk di Stasiun KAI Cikarang, kebutuhan akan layanan informasi yang cepat, akurat, dan dapat diakses kapan saja semakin penting. Solusi inovatif yang dapat diterapkan untuk memenuhi kebutuhan ini adalah dengan menggunakan chat bot sebagai virtual assistant[2].

Chat bot merupakan program komputer yang dibuat untuk menstimulasikan sebuah komunikasi atau percakapan yang interaktif melalui text, suara, maupun visual kepada pengguna (manusia)[3]. Teknologi ini telah berhasil diterapkan di berbagai bidang tetapi, penggunaannya di bidang transportasi publik, khususnya stasiun kereta api di Indonesia, masih belum banyak dikembangkan. Teknologi ini dapat memberikan bantuan untuk layanan informasi secara otomatis, seperti harga tiket, prosedur pembelian tiket, jadwal keberangkatan kereta, hingga informasi keadaan darurat. Penelitian ini menggunakan metode *TF-IDF* (*Term Frequency-Inverse Document Frequency*) dan *cosine similarity* agar chat bot dapat memahami dan menjawab pertanyaan pengguna dengan tepat dan memiliki tingkat akurasi yang tinggi.

TF-IDF merupakan metode yang umumnya digunakan untuk menentukan hubungan kata (term) terhadap kalimat atau dokumen dengan memberikan bobot atau nilai pada masing-masing kata[4]. Nilai TF (term frequency) menunjukkan seberapa sering kata muncul dalam dokumen, sementara IDF (inverse document frequency) menunjukkan seberapa unik kata tersebut di antara dokumen lain[5]. Sementara itu, *cosine similarity* adalah ukuran yang umum digunakan dalam bidang pengambilan informasi dan area studi serupa. Ukuran tersebut menggambarkan dokument teks sebagai vektor yang memiliki istilahnya masing-masing[6]. Model ini digunakan agar kesamaan antara dua dokumen dapat ditentukan dengan menghitung kosinus antara vektor istilah masing-masing dokumen. Nilai *cosine similarity* berkisar diantara 0 hingga 1, ketika nilai yang mendekati 1 menunjukkan tingkat kesamaan yang tinggi[7]. Dalam konteks chat bot, *cosine similarity* digunakan untuk mencocokkan pertanyaan pengguna dengan dokumen atau jawaban yang paling relevan dalam basis pengetahuan[8].

Studi ini bertujuan untuk mengimplementasikan chat bot berbasis metode *TF-IDF* dan *cosine similarity* di Stasiun KAI Cikarang. Tujuan lain dari penelitian ini juga adalah untuk mengukur efektivitas chat bot dalam meningkatkan pengalaman pengguna serta efisiensi layanan informasi di stasiun. Dengan implementasi chat bot ini, diharapkan dapat memberikan kontribusi nyata dalam meningkatkan efisiensi layanan informasi stasiun kereta api di Indonesia serta memberikan pengalaman pengguna yang lebih baik.

Metode Penelitian

Jenis penelitian yang digunakan dalam penelitian ini adalah penelitian kuantitatif. Metode penelitian kuantitatif merupakan salah satu jenis penelitian yang spesifikasinya sistematis, terencana, dan terstruktur dengan jelas sejak awal hingga terciptanya desain penelitian[9]. Jenis data yang digunakan dalam penelitian ini adalah data primer yang diambil dari sumber resmi, yaitu <https://commuterline.id/> yang menyediakan informasi terkait jadwal keberangkatan, harga tiket, dan prosedur pembelian tiket. Data tersebut digunakan

untuk membangun pustaka data chatbot yang berfungsi sebagai asisten virtual bagi pengguna layanan di Stasiun KAI Cikarang.

Metode analisis data yang digunakan adalah dengan memproses data survei menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk memberikan bobot pada kata-kata kunci. TF-IDF dihitung dengan rumus berikut:

$$TF\text{-}IDF(t, d) = TF(t, d) \times IDF(t)$$

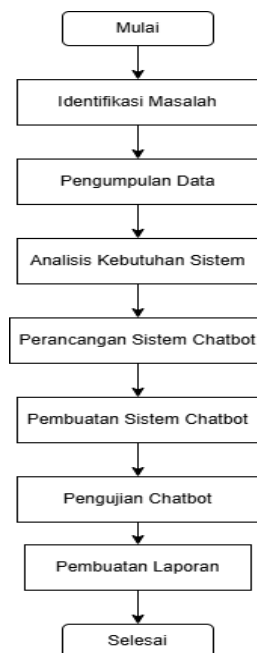
Gambar 1 Rumus TF-IDF

Sementara itu *Cosine similarity* dihitung dengan rumus berikut:

$$\text{Similarity}(A, B) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\| \|\vec{B}\|}$$

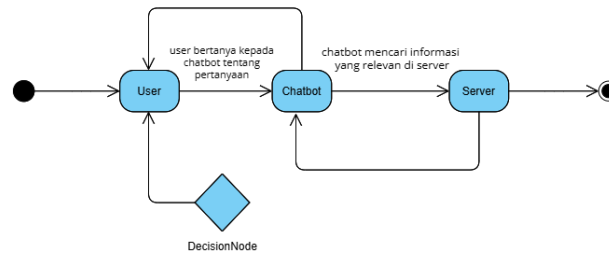
Gambar 2 Rumus Cosine similarity

cosine similarity digunakan untuk mencocokkan pertanyaan pengguna dengan jawaban yang paling relevan dalam basis pengetahuan. Sebelum diimplementasikan, data hasil survei diproses melalui tahapan pra-pemrosesan teks, untuk memastikan kualitas data yang digunakan.



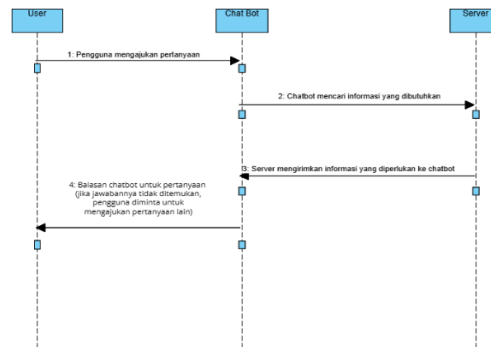
Gambar 3 Alur Penelitian Chatbot

Metode pengolahan data yang digunakan pada penelitian adalah menggunakan NLP yang meliputi *Text preprocessing*. *Preprocessing* merupakan proses yang digunakan algoritma *machine learning* untuk pengolahan text [10]. Kegunaan *text preprocessing* yaitu mengolah suatu kalimat atau pertanyaan dari user menjadi bahasa yang lebih mudah dipahami oleh chatbot. kemudian akan dilakukan pembobotan menggunakan metode TF-IDF dan metode *Cosine similarity* untuk mencari kesamaan antar dokumen sehingga chatbot dapat memberikan jawaban yang tepat apabila terdapat lebih dari satu jawaban yang memiliki kesamaan susunan kata[11].



Gambar 4 Activity Diagram Chatbot

Activity Diagram adalah model analisis yang digunakan untuk menggambarkan sebuah proses aktivitas[12]. Proses dimulai saat pengguna mengajukan pertanyaan ke chatbot. Chatbot mencari informasi di server jika ditemukan, jawaban diberikan, jika tidak, pengguna diberi tahu bahwa jawaban tidak tersedia. Server berperan penting dalam menyediakan data, menunjukkan alur kerja chatbot sebagai asisten virtual.



Gambar 5 Sequence Diagram Chatbot

Sequence Diagram merupakan sebuah diagram yang digunakan untuk menjelaskan dan menampilkan interaksi antar objek-objek dalam sebuah sistem secara terperinci [13]. *Sequence Diagram* diatas menjelaskan terkait alur komunikasi antara pengguna, chatbot, dan server. Metode *TF-IDF* sebagai metode pembobotan untuk setiap jawaban chatbot. Ini adalah lisensi *Creative Commons* untuk template tanya jawab chatbot yang digunakan untuk mencari jawaban untuk kata kunci dalam teks[14]. Hasil *preprocessing* akan disimpan pada data *library* dalam bentuk file *JSON*. Template pertanyaan dan jawaban yang sudah ada pada data *library* kemudian dikelompokkan atau dibuat subheading agar pencarian jawaban lebih cepat. Caranya adalah dengan mengelompokkan kata tanya dari pertanyaan seperti apa, siapa, bagaimana, di mana, dan berapa banyak. Prinsip metode *TF-IDF* ini adalah menghitung bobot kata atau frase [15].

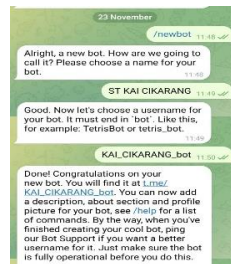
Hasil dan Pembahasan

Dalam penelitian ini, pengembangan chatbot memanfaatkan fitur bawaan Telegram, yaitu BotFather. Pembuatan bot dapat dilakukan dengan mengakses fitur BotFather melalui aplikasi Telegram dengan mencari kata kunci "BotFather". BotFather digunakan untuk mengatur bot, seperti memberikan nama, token API, dan mengelola fungsi dasar bot sebagai asisten virtual.



Gambar 6 Hasil Pencarian BotFather di Telegram

Gambar 6 menunjukkan pencarian BotFather yang digunakan untuk membuat bot melalui Telegram. Untuk membuat bot, klik tombol *Start* dan ketik perintah `"/newbot"`, yang harus dimulai dengan garis miring.



Gambar 7 Penamaan dan Nama Pengguna Bot

Seperti yang ditunjukkan pada Gambar 7, setelah chatbot dibuat, akan ada pemberitahuan yang mengandung nama pengguna dan token bot. Token bot harus disimpan dengan aman.



Gambar 8 Tampilan Perintah `"/mybot"`

Dengan menggunakan perintah `"mybots"` dan memilih pada bot, penulis hanya mengubah informasi chatbot, seperti nama, deskripsi, dan foto profil bot.



Gambar 9 Tampilan Info Bot

Setelah mengedit chatbot dengan BotFather, langkah berikutnya adalah menggunakan tools *Python* untuk coding. Proses coding ini bertujuan untuk mengintegrasikan chatbot Telegram. Berikut ini skripnya:

```
# Fungsi untuk memproses input dan mencari jawaban
def get_response(user_input):
    questions = [item['question'] for item in faq_data]
    answers = [item['answer'] for item in faq_data]

    # Menggunakan TF-IDF Vectorizer
    vectorizer = TfidfVectorizer()
    vectors = vectorizer.fit_transform(questions)

    # Hitung kesamaan cosine
    user_vector = vectorizer.transform([user_input])
    similarities = cosine_similarity(user_vector, vectors)

    # Temukan indeks dengan skor tertinggi
    best_match = np.argmax(similarities)
    best_score = similarities[0, best_match]

    # Jika skor kesamaan rendah, berikan jawaban default
    if best_score < 0.2:
        return "Maaf, saya tidak memahami pertanyaan Anda. Bisa dijelaskan lebih spesifik?"
    return answers[best_match]

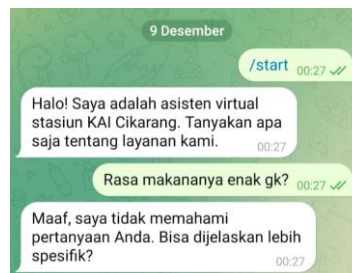
# Fungsi start
@asyncio.coroutine
def start(update: Update, context: CallbackContext) -> None:
    await update.message.reply_text("Halo! Saya adalah asisten virtual stasiun KAI Cikarang. Tanyakan apa saja tentang layanan kami.")

# Fungsi untuk menangani pesan
@asyncio.coroutine
def handle_message(update: Update, context: CallbackContext) -> None:
    user_input = update.message.text
    response = get_response(user_input)
    await update.message.reply_text(response)

# Fungsi utama
def main():
    # Token dari BotFather
    TOKEN = "3093784366:AAcWb3Ckq83E3hdtLW4d8VYH5Vg1kqgk"
```

Gambar 11 Skrip *TF-IDF* dan Kesamaan Kosinus

Gambar 12 merupakan contoh Penggunaan chatbot dengan memulai interaksi dengan mengetik *"/start"*, yang menunjukkan cara untuk mendapatkan informasi lebih lanjut.



Gambar 12 Chatbot Ketika tidak memahami pertanyaan

Berikut ini adalah tabel akurasi jawaban chatbot terhadap pengguna yang telah dirangkum dalam sebuah tabel:

Tabel 1 Hasil Akurasi Pengujian Chatbot

No.	Pertanyaan Pada Chatbot	Chatbot Responsif	Hasil
1	Apa kereta api yang bertujuan stasiun Surabaya Pasar Turi?	Untuk tujuan Surabaya tersedia kereta api Kertajaya, Pandalungan, dan Airlangga.	Sesuai
2	Kapan jadwal keberangkatan kereta api Jaka Tingkir?	Jam keberangkatan kereta api Jaka Tingkir yaitu pukul 12:40 dari stasiun Cikarang	Sesuai
3	Apa KRL ke Stasiun Manggarai?	Untuk tujuan Stasiun Manggarai bisa naik jurusan Angke	Sesuai
4	Berapa Harga tiket kereta tujuan Bandung?	Harga tiket kereta tujuan Bandung yaitu Rp. 45.000. Tersedia kereta api Cikuray.	Sesuai
5	Rasa makanannya enak gk?	Maaf, saya tidak memahami pertanyaan Anda. Bisa dijelaskan lebih spesifik?	Tidak Sesuai
6	Apa nama kereta tujuan Yogyakarta dari Cikarang?	Untuk kereta api tujuan Cikarang-Yogyakarta, anda bisa memilih beberapa jenis kereta api. Salah satu opsi yang tersedia yaitu kereta api bengawan.	Sesuai
7	Apakah tersedia layanan parkir di Stasiun KAI Cikarang?	Stasiun Cikarang menyediakan area parkir untuk penumpang yang ingin meninggalkan kendaraan saat bepergian dengan kereta.	Sesuai
8	Berapa harga tiket KRL dari Stasiun Cikarang ke Manggarai?	tarif krl menuju stasiun tersebut sebesar Rp.4000	Sesuai
9	Fasilitas kereta ekonomi apa aja?	Kereta ekonomi menyediakan tempat duduk yang nyaman, toilet, dan penjualan makanan dan minuman di dalam kereta.	Sesuai
10	tiket kereta cepat berapa?	Harga tiket kereta lokal dari Cikarang adalah Rp5.000 untuk semua rute.	Tidak Sesuai
11	Wifi ada gak?	Stasiun Cikarang menyediakan fasilitas Wi-Fi gratis di area tertentu. Anda bisa mengaksesnya dengan mengikuti petunjuk di stasiun.	Sesuai
12	fasilitas di stasiun apa aja?	stasiun kereta api kai cikarang menyediakan wifi, toilet, mushola, tempat parkir dan tempat makan.	Sesuai
13	Apakah ada kereta malam dari Cikarang ke Semarang?	Saat ini, kereta api malam dari Stasiun Cikarang tidak tersedia. Kereta umumnya beroperasi hingga malam hari, namun tidak ada keberangkatan larut malam.	Sesuai
14	Kereta apa dari cikarang ke surabaya?	Saat ini, tidak ada kereta api langsung dari Stasiun Cikarang menuju Surabaya. Anda bisa transit di Jakarta untuk melanjutkan perjalanan ke Surabaya.	Sesuai
15	dimana letak mushola?	Maaf, saya tidak memahami pertanyaan Anda. Bisa dijelaskan lebih spesifik?	Tidak Sesuai

No.	Pertanyaan Pada Chatbot	Chatbot Responsif	Hasil
16	Kereta ke Bandung apa aja?	Untuk saat ini yang tersedia hanya kereta api Cikuray.	Sesuai
17	KRL terakhir jam berapa?	Jadwal keberangkatan KRL terakhir dari Stasiun Cikarang adalah pukul 22:00 malam.	Sesuai
18	Kereta Api Bengawan berangkat jam berapa?	Jam keberangkatan kereta api Bengawan yaitu pukul 06:46 dari stasiun Cikarang.	Sesuai
19	Ke Jakarta berapa Jam?	Perjalanan kereta dari Stasiun Cikarang ke Jakarta biasanya memakan waktu sekitar 1 hingga 1,5 jam, tergantung jenis kereta.	Sesuai
20	ada berapa loket?	Harga tiket KRL dari Stasiun Cikarang ke Jakarta mulai dari Rp 8.000 hingga Rp 15.000, tergantung pada tujuan stasiun.	Tidak Sesuai

Dari hasil Tabel 1, dapat diperoleh perhitungan akurasi sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah Jawaban Sesuai}}{\text{Jumlah Total Hasil}}$$

sebagai berikut Jumlah Jawaban yang sesuai = 16

Jumlah Total hasil = 20

$$\text{Akurasi} = \frac{16}{20} \times 100\% = 80\%$$

Hasil ini menunjukkan metode yang efektif untuk menghitung akurasi, dengan menggunakan rumus yang jelas dan sistematis. Akurasi yang diperoleh adalah 80%, menunjukkan bahwa pendekatan ini tidak hanya sederhana tetapi juga memberikan hasil yang informatif untuk evaluasi kinerja.

Kesimpulan

Penelitian ini berhasil mengembangkan chatbot sebagai asisten virtual di Stasiun Kereta Api KAI Cikarang, menggunakan metode *TF-IDF* dan *cosine similarity*. Chatbot yang dirancang mampu memberikan informasi terkait jadwal keberangkatan, harga tiket, dan prosedur pembelian tiket secara otomatis. Hasil evaluasi menunjukkan bahwa chatbot dapat memberikan jawaban yang akurat dengan tingkat akurasi 80%. Metode yang digunakan terbukti efektif dalam mencocokkan pertanyaan pengguna dengan jawaban yang relevan, meningkatkan efisiensi layanan informasi, dan memberikan pengalaman pengguna yang lebih baik.

Daftar Rujukan

- [1] R. Balqis, P. Putra, N. R. Oktadini, A. Meiriza, and P. E. Sevtiyuni, "Penerapan Metode SOSTAC dalam Perancangan Sistem Informasi Space Rent UMKM Stasiun," *Journal of Information System Research (JOSH)*, vol. 4, no. 3, Art. no. 3, Apr. 2023, doi: 10.47065/josh.v4i3.3247.
- [2] S. H. Bariah, W. Pratiwi, and K. A. N. Imania, "Pengembangan Virtual Assistant Chatbot Berbasis Whatsapp Pada Pusat Layanan Informasi Mahasiswa Institut Pendidikan Indonesia - Garut," *Petik: Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, vol. 8, no. 1, Art. no. 1, Mar. 2022.
- [3] D. W. Harahap and L. Fitria, "Aplikasi Chatbot Berbasis Web Menggunakan Metode Dialogflow," *JICOM*, vol. 1, no. 1, pp. 6–13, Apr. 2020, doi: 10.33059/j-icom.v1i1.2796.
- [4] N. Arifin, U. Enri, and N. Sulistiyowati, "Penerapan Algoritma Support Vector Machine (SVM) dengan *TF-IDF* N-Gram untuk Text Classification," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 6, no. 2, Art. no. 2, Dec. 2021, doi: 10.30998/string.v6i2.10133.
- [5] O. I. Gifari, M. Adha, I. R. Hendrawan, and F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan *TF-IDF* dan Support Vector Machine," *Journal of Information Technology*, vol. 2, no. 1, Art. no. 1, Mar. 2022, doi: 10.46229/jifotech.v2i1.330.

- [6] R. A. Rasyid and D. H. U. Ningsih, "Penerapan Algoritma *TF-IDF* dan *Cosine similarity* untuk Query Pencarian Pada Dataset Destinasi Wisata," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 8, no. 1, Art. no. 1, Jan. 2024, doi: 10.35870/jtik.v8i1.1416.
- [7] S. A. Zulvian, K. Prihandani, and A. A. Ridha, "Perbandingan Metode MSD dan *Cosine similarity* pada Sistem Rekomendasi Item-Based Collaborative Filtering," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 4, no. 2, pp. 340–347, Dec. 2021, doi: 10.31539/intecom.v4i2.2781.
- [8] M. Habibi, "Implementation of *Cosine similarity* in an Automatic Classifier for Comments," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 3, no. 2, Art. no. 2, 2018, doi: 10.14421/jiska.2018.32-05.
- [9] S. Efendi, H. Sibyan, and N. Hasanah, "IMPLEMENTASI CHATBOT SEBAGAI VIRTUAL ASSISTANT PADA PELAYANAN PUBLIK DAN BANTUAN DI DINAS SOSIAL KABUPATEN WONOSOBO DENGAN METODE *TF-IDF*," *Clean Energy and Smart Technology*, vol. 3, no. 1, Art. no. 1, Oct. 2024, doi: 10.58641/cest.v3i1.103.
- [10] A. Muhidin, M. Danny, and E. Rilvani, "Algoritme Multinomial Naïve Bayes Pada Aplikasi Chatbot Layanan Informasi Berbasis Teks," *Progresif: Jurnal Ilmiah Komputer*, vol. 19, no. 1, Art. no. 1, Feb. 2023, doi: 10.35889/progresif.v19i1.1113.
- [11] Nilawati, Husaini, and J. Salat, "PENGUNAAN METODE *COSINE SIMILARITY* DAN *TF-IDF* UNTUK KLASIFIKASI JUDUL SEMINAR PROPOSAL PADA FAKULTAS TEKNIK UNIVERSITAS JABAL GHAFUR," *Sagita Academia Journal*, vol. 2, no. 1, Art. no. 1, Jan. 2024, doi: 10.61579/sagita.v2i1.60.
- [12] T. A. Zuraiyah, D. K. Utami, and D. Herlambang, "IMPLEMENTASI CHATBOT PADA PENDAFTARAN MAHASISWA BARU MENGGUNAKAN RECURRENT NEURAL NETWORK | Zuraiyah | Jurnal Ilmiah Teknologi dan Rekayasa." Accessed: Dec. 22, 2024. [Online]. Available: <https://ejournal.gunadarma.ac.id/index.php/tekno/article/view/2388/1895>
- [13] L. Permata Sari, "View of *Cosine similarity* -based Plagiarism Detection on Electronic Documents." Accessed: Dec. 22, 2024. [Online]. Available: <https://journal.lenterailmu.com/index.php/josapen/article/view/14/20>
- [14] R. Wati, S. Ernawati, and H. Rachmi, "Pembobotan *TF-IDF* Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH," *Jurnal Manajemen Informatika (JAMIKA)*, vol. 13, no. 1, pp. 84–93, Apr. 2023, doi: 10.34010/jamika.v13i1.9424.
- [15] S. Nida Cahyani and G. Wilujeng Saraswati, "IMPLEMENTATION OF SUPPORT VECTOR MACHINE METHOD IN CLASSIFYING SCHOOL LIBRARY BOOKS WITH COMBINATION OF *TF-IDF* AND WORD2VEC", Accessed: Dec. 22, 2024. [Online]. Available: <https://core.ac.uk/reader/596301450>