

MODEL NATURAL LANGUAGE PROCESSING TF-IDF SEBAGAI CHATBOT TOUR GUIDE MUSEUM KONFERENSI ASIA AFRIKA (KAA)

Aurelia Zafira Putri
Teknik Elektro
Telkom University
Bandung, Jawa Barat
aureliazafiraputri.azp@gmail.com

Priskila Destriani B
Teknik Industri
Institut Teknologi Kalimantan
Medan, Sumatera Utara
priskilabanjarnahor@gmail.com

Syifa Nurul Asyfia
Teknik Elektro
Telkom University
Bandung, Jawa Barat
syifnrl@365.telkomuniversity.ac.id

Rizal Ashari
Teknik Elektro
Politeknik Negeri Ujung Pandang
Makassar, Sulawesi Selatan
rizalashari147@gmail.com

Muh Daffa Abbas
Teknik Telekomunikasi
Politeknik Negeri Ujung Pandang
Makassar, Sulawesi Selatan
daffaabbas07@gmail.com

Abstrak—Museum Konferensi Asia Afrika (KAA) telah dibuka untuk umum, namun kebijakan pembatasan sosial dari pihak museum tetap diberlakukan dan membuat wisatawan kesulitan dalam pencarian data wisatawan khususnya para pelajar dan mahasiswa/i. Strategi dalam membantu pencarian data tersebut dengan memanfaatkan kemajuan pada bidang teknologi. *Natural Language Processing* (NLP) merupakan cabang teknologi *Artificial Intelligence* (AI) yang diciptakan untuk membantu mesin/komputer dalam memahami bahasa manusia (Chopra, A., Prashar, A., & Sain, C., 2013). Maka dari itu dibuat suatu sistem *Chatbot* yang menggunakan konsep *tour guide* pada museum Konferensi Asia Afrika (KAA). Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode *cosine similarity* merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah *vector* dengan menggunakan *keywords* (kata kunci) dari sebuah dokumen sebagai ukuran. Metode penelitian yang digunakan adalah metode *Research and Development* (R&D). Data latih yang digunakan dalam penelitian ini berjumlah 2 (dua) file *dataset* dalam bentuk file csv dengan 17 pertanyaan yang berbeda. Hasil penelitian menunjukkan bahwa penggunaan metode TF-IDF dapat menghasilkan indeks dengan nilai tertinggi yang menghasilkan kata yang sama dengan kata input).

Kata kunci— *Chatbot*, *cosine similarity*, Museum KAA, *Machine Learning*, NLP, TF-IDF.

I. PENDAHULUAN

Dinamika kehidupan sosial mulai berubah semenjak terjadinya pandemi Covid-19, terkait kebijakan pembatasan sosial diterapkan pemerintah untuk menekan angka persebaran penyakit Covid-19 yang berdampak pada penutupan berbagai tempat yang memungkinkan terjadinya kerumunan. Hingga saat ini terdapat penurunan wisatawan yang berkunjung ke Kota Bandung dengan kurun waktu 3 (tiga) tahun. Dapat dilihat bahwa jumlah wisatawan di tahun 2018 sebanyak 7.585.345, yang berasal dari domestik maupun mancanegara menjadi 3.259.300 di tahun 2020. Data dibawah ini memperlihatkan bahwa penurunan yang terjadi hampir 50% dari kunjungan wisatawan di tahun 2018 (Badan Pusat Statistik Kota Bandung, 2020).

Jenis Wisatawan	Jumlah Wisatawan Menurut Jenis (Jawa)		
	2018	2019	2020
Wisatawan Mancanegara	227 560	252 842	30 210
Wisatawan Domestik	7 357 785	8 175 221	3 229 090
Jumlah	7 585 345	8 428 063	3 259 300
Jumlah Wisatawan Menurut Jenis di Kota Bandung			
Dinas Kebudayaan dan Pariwisata Kota Bandung			

Kondisi tersebut menyebabkan keterpurukan di industri pariwisata. Salah satunya pada Museum yang berlokasi di Kota Bandung, yaitu Museum Konferensi Asia Afrika (KAA). Meskipun di tahun 2022 ini Museum Konferensi Asia Afrika (KAA) telah dibuka untuk umum, namun tetap ada kebijakan pembatasan sosial dari pihak Museum yang tertera pada web resmi Museum Konferensi Asia Afrika (mkaa). Kebijakan tersebut diantaranya, Museum hanya buka di 4 (empat) hari saja yaitu Selasa, Kamis, Sabtu dan Minggu. Wisatawan yang akan berkunjung harus mengisi *Google Form* yang disediakan oleh pihak Museum dan harus mengakses aplikasi Peduli Lindungi. Wisatawan yang akan berkunjung dan memiliki tujuan untuk penelitian, harus membuat proposal terlebih dahulu. Kemudian hanya diperkenankan untuk berkeliling di Museum dengan batasan waktu maksimal 15 menit saja.

Dengan banyaknya aturan yang ketat dari pemerintah maupun pihak Museum, membuat Museum yang seharusnya dapat melakukan pelayanan informasi dan edukasi yang merupakan fungsi museum menjadi tidak dapat melakukan pelayanan dengan optimal. Selain mengurangi fungsi dari Museum, tentu saja merugikan banyak pihak tidak terkecuali para wisatawan khususnya para pelajar dan mahasiswa/i yang memiliki keterbatasan dalam pencarian data. Strategi yang digunakan untuk membantu pencarian data bagi wisatawan Museum Konferensi Asia Afrika (KAA) adalah dengan memanfaatkan kemajuan pada bidang teknologi. *Natural Language Processing* (NLP) merupakan cabang teknologi *Artificial Intelligence* (AI) yang diciptakan untuk membantu mesin/komputer dalam memahami bahasa manusia (Chopra, A., Prashar, A., & Sain, C., 2013). Salah satu contoh penggunaan teknologi NLP terdapat dalam aplikasi *Chatbot* yang dapat membantu permasalahan tersebut. *Chatbot* adalah sebuah program yang mendukung interaksi pengguna melalui percakapan di *Natural Language* dan biasanya dapat diakses melalui web atau jejaring sosial (S. Perez-Soler, S. Juarez-Puerta, E. Guerra and J. de Lara, 2021).

Maka dari itu, kami memilih untuk membuat sebuah sistem *chatbot tour guide* Museum dengan menggunakan

algoritma *Machine Learning* yaitu TF-IDF (*Term Frequency - Inverse Document Frequency*) yang menggabungkan dua proses untuk mendapatkan penghitungan nilai, yaitu durasi kemunculan kata dalam dokumen dan frekuensi kata (Riyani, Zidny & Burhanuddin, 2019). Harapan kami dengan dirancangnya sistem *chatbot* untuk *tour guide* museum dapat mengedukasi kepada para wisatawan khususnya pelajar dan mahasiswa/i agar mendapatkan data yang dicari pada Museum Konferensi Asia Afrika (KAA) sesuai fungsi dari Museum itu sendiri sebagai tempat pencarian informasi.

II. KAJIAN PUSTAKA

A. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah disiplin ilmu komputer yang bertujuan untuk memahami konsep dan maksud dari bahasa manusia. Sementara manusia cukup mahir memahami sintaks *linguistic* dan tata bahasa serta hubungan spasial tersirat, komputer memiliki kesulitan besar pengolahan *query* bahasa alami (Allen, 1995). Contoh sistem ini dan penggunaan API adalah *Natural Language Toolkit* (NLTK) dan *CoreNLP* dari Stanford University.

Natural Language Processing (NLP) merupakan salah satu bagian dari *Artificial Intelligence* (AI). NLP mempelajari komunikasi antara manusia dengan komputer melalui bahasa alami (Chopra, A., Prashar, A., & Sain, C., 2013).

B. Chatbot

Chatbot merupakan salah satu bentuk aplikasi *Natural Language Processing* (NLP) (Nila & Afrianto, 2015). *Chatbot* adalah sebuah aplikasi yang memungkinkan percakapan antara mesin dengan manusia menggunakan bahasa alami manusia. *Chatbot* sering digambarkan sebagai salah satu cara untuk mensimulasikan interaksi antara mesin dengan manusia. Untuk memahami bahasa alami manusia, *chatbot* menggunakan pendekatan NLP untuk mempelajari dan memproses setiap kata yang diucapkan oleh manusia berupa pesan teks (Khanna, Pandey, Kalia, & Bhale, 2015).

Chatbot terdiri dari komponen bot program dan komponen brain file. Bot program merupakan program utama pada *chatbot* yang akan mengakses input dari pengguna, melakukan parsing dan kemudian membawanya ke brain file untuk kemudian diberikan respon. Bot program terdiri dari komponen scanner dan parser. Brain file merupakan otak dari *chatbot* itu sendiri yang menentukan bagaimana cara *chatbot* berpikir dan akan memberikan respon. Brain file biasanya berupa file plain text. Brain file berfungsi sebagaimana tabel informasi pada kompilator bahasa pemrograman tingkat tinggi. Di dalam brain file inilah disimpan semua kosakata, kepribadian, dan pengetahuan (knowledge) dari *chatbot*. Semakin banyak pengetahuan yang dimiliki chat bot maka akan semakin besar ukuran file dari brain file tersebut (Rudiyanto, 2005).

C. Term Frequency - Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode yang digunakan pada proyek akhir ini dan berfungsi untuk mencari relevansi antar beberapa dokumen. Metode ini yang cukup mudah dipelajari dan diterapkan untuk permasalahan keakuratan sebuah dokumen. TF-IDF adalah perhitungan atau pembobotan kata

melalui teknik tokenisasi, stopwords, dan stemming. Frekuensi munculnya kata dalam dokumen yang diberikan menunjukkan pentingnya kata itu di dalam sebuah dokumen (Andayani & Ryansyah, 2017).

Metode ini melakukan penggabungan 2 (dua) konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut (Nurjannah, Hamdani & Astuti, 2013).

- a) Perhitungan *Term Frequency* (tf) menggunakan persamaan:

$$tf = tf_{ij} \quad (1)$$

Dengan tf adalah *term frequency*, dan tf_{ij} adalah banyaknya kemunculan *term* t_i dalam dokumen d_j , *Term frequency* (tf) dihitung dengan menghitung banyaknya kemunculan *term* t_i dalam dokumen d_j .

- b) Perhitungan *Inverse Document Frequency* (idf), menggunakan persamaan:

$$idf_i = \log \frac{N}{df_i} \quad (2)$$

Dengan idf_i adalah *inverse document frequency*, N adalah jumlah dokumen yang terambil oleh sistem, dan idf_i adalah banyaknya dokumen dalam koleksi dimana *term* t_i muncul di dalamnya.

- c) Perhitungan *term frequency Inverse Document Frequency* (tfidf), menggunakan persamaan:

$$W_{ij} = tf_i \cdot \log \frac{N}{df_i} \quad (3)$$

Dengan W_{ij} adalah bobot dokumen, N adalah Jumlah dokumen yang terambil oleh sistem, tf_{ij} adalah banyaknya kemunculan *term* t_i pada dokumen d_j , dan df_i adalah banyaknya dokumen dalam koleksi dimana *term* t_i muncul di dalamnya. Bobot dokumen (W_{ij}) dihitung untuk didapatkannya suatu bobot hasil perkalian atau kombinasi antara *term frequency* (tf_{ij}) dan *Inverse Document Frequency* (idf_i) (Amin, 2012).

D. Cosine Similarity

Cosine Similarity digunakan untuk melakukan perhitungan dari kesamaan suatu file/dokumen. Berikut merupakan rumus dari perhitungan *cosine similarity* :

$$\cos \alpha = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (4)$$

Dengan A dan B merupakan Vektor dari A dan B, yang akan dibandingkan kemiripannya. Kemudian $A \cdot B$ merupakan hasil *dot* antara vektor A dan vektor B, $|A|$ adalah panjang vektor A, $|B|$ = panjang vektor B dan $|A||B|$ merupakan hasil dari *cross* antara $|A|$ dan $|B|$ (Ye, 2014). Metode klasifikasi ini dilakukan dengan mencari setiap kesamaan antara kata input dalam program dengan kata input dalam *dataset* berturut-turut sesuai urutan kata input yang dimasukan pada program.

E. Museum Konferensi Asia-Afrika (KAA)

Museum Konferensi Asia Afrika (KAA) atau dikenal saat ini dengan Gedung Merdeka merupakan sebuah tempat sejarah yang berlokasi di Jl. Asia Afrika, Kota Bandung, Jawa Barat. Pembangunan museum ini diawali dengan ide dari Menteri Luar Negeri Republik Indonesia, Prof. Dr. Mochtar Kusumaatmadja, S.H., LL.M yang disampaikan pada forum rapat Panitia Peringatan 25 tahun Konferensi Asia Afrika (1980). Tujuan pembuatan museum untuk memperingati peristiwa Konferensi Asia – Afrika (KAA) pada tanggal 18 April 1955 – 24 April 1955. Kemudian museum ini diresmikan pada tanggal 24 April 1980 bertepatan dengan peringatan 25 tahun Konferensi Asia Afrika (KAA).

III. METODOLOGI

A. Observasi dan Studi Literatur

Langkah pertama yang dilakukan adalah melakukan observasi pada Museum Konferensi Asia – Afrika. Objek – objek yang terdapat pada museum dapat dikategorikan sebagai berikut:

- Sejarah Gedung Konferensi Asia Afrika
- Latar Belakang Konferensi Asia – Afrika
- Pertemuan – Pertemuan yang mendasari Konferensi Asia Afrika
- Tokoh – tokoh pada Konferensi Asia Afrika
- Dampak Konferensi Asia Afrika
- Piagam Dasasila Bandung
- Alat – alat yang dipakai ketika Konferensi Asia – Afrika berlangsung.

Objek yang ditampilkan sebagian besar adalah gambar yang dilengkapi dengan narasi panjang. Dari objek – objek yang telah diamati, akan dibangun sebuah *dataset* untuk model TF-IDF yang akan digunakan.

B. Dataset

Dataset yang dirancang terdiri dari pertanyaan – pertanyaan yang mewakili setiap kategori objek pada Museum Konferensi Asia – Afrika. Kumpulan pertanyaan yang telah dibuat kemudian disusun dalam tabel menjadi beberapa baris. Setiap kolom mewakili jawaban pertanyaan yang terdapat pada baris tersebut. *Dataset* yang telah dibentuk adalah sebagai berikut:

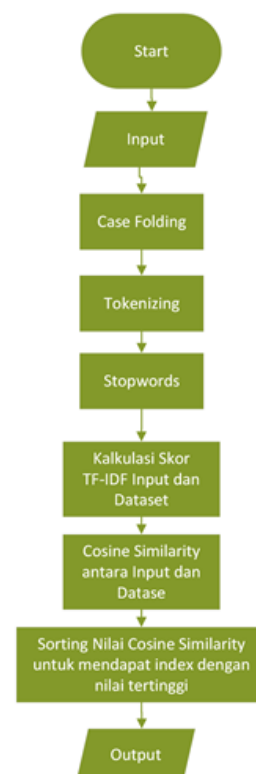
List Pertanyaan Museum KAA	List jawaban berdasarkan <i>dataset</i>
Apa tujuan berdirinya museum Asia - Afrika?	Memperingati peristiwa Konferensi Asia Afrika pada tanggal 18 April 1955 - 24 April 1955
Gedung Merdeka	Gedung Merdeka dibangun pada tahun 1895 dengan nama Gedung Concordia. Dulunya merupakan tempat hiburan dan perkumpulan kelompok sosial para saudagar Belanda. Pada tahun 1954, pemerintah Indonesia mengambil alih gedung ini dan menamakannya Gedung Merdeka.
Kejadian yang melatar belakangi Konferensi Asia - Afrika	Perang Dingin, Perang Dunia II, Sistem Apartheid, dan Kolonialisme
Perang Dunia II	Perang Dunia II (1 September 1939) merupakan pertarungan ideologi antara negara-negara Eropa, Amerika, dan Asia, yang melibatkan lebih dari 30 negara. Diawali dengan ambisi Jerman dalam menguasai Eropa dengan melakukan invasi terhadap Polandia, menyerang Denmark, Luxembourg, Belanda, Belgia, Norwegia, Perancis, dan Uni Soviet. Perang ini berakhir dengan Sekutu mengalahkan Jerman dan Jepang.
Perang Dingin	Perang dingin merupakan pertarungan ideologi antara Amerika Serikat dan Uni Soviet setelah Perang Dunia II berakhir. Dinamakan perang dingin karena dari kedua pihak tidak melancarkan aksi militer langsung terhadap satu sama lain, namun masing - masing pihak memiliki senjata nuklir yang dapat menyebabkan kehancuran. Hal ini memunculkan masalah baru yang mengakibatkan permusuhan bahkan pada tingkat perang terbuka. Seperti yang terjadi di Korea, Indo - Cina, Palestina, Afrika Selatan dan Afrika Utara. Perang Dingin ini juga melahirkan dua blok ideologis kuat; Blok Barat dan Blok Timur.
Apartheid	Adalah bentuk sistematis diskriminasi ras yang legal, yang mengelompokkan manusia berdasarkan ras dan warna. Orang - Orang berkulit hitam tidak memiliki hak dasar/fundamental

	dan kebebasan dalam hidup.		
Kolonialisme	Berawal dari ditemukannya jalur laut pada abad ke-15, dimulailah penjajahan bangsa - bangsa Eropa yang menyebar ke pelosok - pelosok pedalaman Afrika, Asia, dan benua lainnya pada abad ke-19. Ambisi bangsa Eropa untuk menguasai dunia, mengeruk sumber daya alam, dan memeras tenaga penduduk asli. Kolonialisme pada bangsa Asia masih berlangsung di beberapa negara, bahkan setelah Perang Dunia II berakhir.	masalah dunia pasca Perang Dunia II	(Mohammed Ali) untuk mengadakan pertemuan informal. Pertemuan tersebut didukung dengan baik oleh seluruh pimpinan negara dan pemerintahan negara-negara undangan. Saat itu, Presiden Indonesia Soekarno meminta Perdana Menteri Indonesia, Ali Sastroamidjojo, untuk membahas gagasan penyelenggaraan Konferensi Asia Afrika dalam Konferensi Kolombo. Pertemuan ini diadakan untuk mewujudkan gagasan membangun solidaritas Asia Afrika dan gerakan nasional melawan kolonialisme.
Peristiwa sebelum Konferensi Asia Afrika (KAA)	Pertemuan Tugu (9 Maret 1954 - 22 Maret 1954), Konferensi Kolombo (28 April 1954 - 2 Mei 1954), Konferensi Bogor (28 Desember 1954 - 29 Desember 1954), dan Pernyataan bersama Indonesia dengan PM India dan PM Birma	Konferensi Bogor	Pada tanggal 28 - 29 Desember 1954, atas undangan Perdana Menteri Indonesia, Perdana Menteri Konferensi Kolombo (Burma, Ceylon, India, Indonesia, dan Pakistan) mengadakan pertemuan di Bogor untuk membahas persiapan Konferensi Asia Afrika. Konferensi tersebut berhasil merumuskan rancangan agenda, tujuan, dan negara-negara yang diundang Konferensi Asia Afrika. Lima peserta Konferensi Bogor menjadi sponsor Konferensi Asia Afrika, dan Indonesia terpilih sebagai tuan rumah konferensi. Konferensi diputuskan untuk diadakan pada minggu terakhir bulan April 1955 dan Presiden Soekarno memilih Bandung sebagai tempat konferensi.
Pertemuan Tugu	Diketahui oleh Menteri Luar Negeri Sunario, pertemuan ini dihadiri oleh perwakilan Indonesia di Asia, Afrika, dan Pasifik untuk membahas rumusan-rumusan yang akan menjadi bahan bagi Perdana Menteri Ali Sastroamidjojo untuk dibawa ke Konferensi Kolombo, sebagai dasar usulan Indonesia untuk meluaskan gagasan kerjasama regional di tingkat Asia Afrika.		
Konferensi Kolombo	Merupakan cikal bakal terjadinya Konferensi Asia - Afrika. Pada tanggal 28 April - 2 Mei 1954, diadakan Konferensi Kolombo. Dibahas masalah-masalah yang menjadi Common Conference. Pada konferensi tersebut, Perdana Menteri Indonesia mengusulkan untuk mengadakan konferensi serupa lainnya, yang cakupannya lebih luas untuk Negara-negara Asia dan Afrika. Gagasan itu didukung oleh seluruh anggota konferensi meskipun beberapa masih ragu-ragu. Konferensi tersebut memberikan kesempatan kepada Indonesia untuk menjajaki kemungkinan penyelenggaraannya dan hal itu tertulis dalam paragraf terakhir dari Komunike Terakhir Konferensi Kolombo.	Delegasi Partisipan Konferensi Asia Afrika	U Nu (Perdana Menteri Birma), Prince Norodom Sihanouk (Mantan Raja Kamboja), Sir John Kotelawala (Perdana Menteri Ceylon), Chou En Lai (Perdana Menteri Republik Rakyat Tiongkok), Gamal Abdel Nasser (Perdana Menteri Mesir), Yilma Deressa (Ketua Delegasi Ethiopia di PBB), Kojo Botsio (Menteri Negara Pantai Emas), Pandit Jawaharlal Nehru (Perdana Menteri Indi), Sastroamidjojo (Perdana Menteri Indonesia), Ali Amini (Menteri Keuangan Iran), Mohammad Fadhil El Jamali (Anggota Parlemen Irak), Tatsunosuke Takasaki (Menteri Negara Jepang), Hazza Al Majali (Perdana Menteri Yordania), Katay Don Sasorith (Perdana Menteri Laos), Sami Bek Solh (Perdana Menteri Libanon), Momolu Dukuly (Pejabat Sekretaris Negara Liberia), Mahmoud Bey Muntasser (Duta Besar Libya di London), Sovag Jung Thapa (Menteri Luar
Pernyataan bersama menteri negara ASEAN terhadap	Pada awal tahun 1954, Perdana Menteri Ceylon, Sir John Kotelawala, mengundang Perdana Menteri Burma (U Nu), India (Jawaharlal Nehru), Indonesia (Ali Sastroamidjojo), dan Pakistan		

	Negeri Nepal), Mohammed Ali (Perdana Menteri Pakistan), Carlos P. Romulo (Utusan Khusus dan Pribadi Presiden Filipina untuk Amerika Serikat dengan Pangkat Duta Besar), Emir Faisal Ibn Abdul Aziz Al Saud (Ketua Dewan Menteri Arab Saudi), Sayed Ismail El Azhari (Perdana Menteri Sudan), Khaled El Azm (Menteri Luar Negeri Suriah), Prince Wan Waithayakon Krommun Naradhip Bongspraband (Menteri Luar Negeri Thailand), Fatin Rostu Zorlu (Wakil Perdana Menteri Turki), Pham Van Dong (Wakil Perdana Menteri Republik Demokrasi Vietnam [Vietnam Utara]), Nguyen Van Thoai (Menteri Perencanaan dan Pembangunan Negara Vietnam [Vietnam Selatan]), dan Emir Seif El Islam Al Hassan (Perdana Menteri Yaman).
Tokoh Konferensi Asia Afrika	Ketua: Ali Sastroamidjojo (Perdana Menteri Indonesia) Sekretaris Jendral: Roestan Abdoel Gani
Dampak Konferensi Asia Afrika	Lahirnya Gerakan Non - Blok, pertambahan jumlah negara yang merdeka di Asia - Afrika, adanya Konferensi Islam Asia - Afrika, dan lahirnya semangat solidaritas antara negara Asia - Afrika
Gerakan Non Blok	Gerakan yang menyikapi lahirnya dua blok ideologi kuat pasca Perang Dunia II; Blok Barat dan Blok Timur. Gerakan ini menunjukkan ketidakberpihakan terhadap dua blok tersebut.
Komite Konferensi Asia - Afrika	Hasil dari Konferensi Asia – Afrika membentuk Komite Kebudayaan, Komite Politik dan Komite Ekonomi untuk meningkatkan kerja sama antara negara Asia – Afrika. Ketua Komite Politik: Ali Sastroamidjojo; Ketua Komite Ekonomi: Ali Sastroamidjojo; Ketua Komite Kebudayaan: Mr. Muhammad Yamin.

C. Algoritma Chatbot

Selanjutnya dilakukan perancangan algoritma dari chatbot yang akan dibangun. Flowchart yang akan digunakan adalah sebagai berikut:



Gambar 1. Flowchart Algoritma Chatbot

Tahap pertama sebelum melakukan metode TF-IDF adalah melakukan *text preprocessing* terhadap input yang diberi oleh user serta *dataset* yang digunakan. *Text preprocessing* meliputi langkah – langkah berikut:

1. Case Folding

Tahap penelitian dengan perubahan huruf dari huruf kapital menjadi huruf kecil.

2. Tokenizing

Proses memecah dokumen menjadi kumpulan kata. Tokenization dapat dilakukan dengan menghilangkan tanda baca dan memisahkannya per spasi. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua *token* ke bentuk huruf kecil (*lower case*) (Kurniawan, Solihin, dan Hastarita, 2014). Data yang akan dilakukan *tokenizing* adalah kolom – kolom pertanyaan pada *dataset* dan juga input user.

3. Stopwords/Filtering

Proses penghilangan kata tidak penting pada deskripsi melalui pengecekan kata-kata hasil *parsing* deskripsi apakah termasuk di dalam daftar kata tidak penting (*stoplist*) atau tidak. Jika termasuk di dalam *stoplist* maka kata-kata tersebut akan di-*remove* dari deskripsi sehingga kata-kata yang tersisa di dalam deskripsi dianggap sebagai kata-kata penting atau *keywords* (Nurdiana, Jumadi, dan Nursantika, 2016).

Setelah melakukan *text preprocessing*, dilakukan pembobotan TF-IDF terhadap *dataset* dan input user. Proses dari TF-IDF akan menghasilkan frekuensi kemunculan setiap kata dalam kalimat tersebut. Kata

Setelah itu, untuk mengetahui kemiripan dari kata input user dan *dataset*, digunakan metode *cosine similarity*. Semakin besar nilai dari *cosine similarity* antara kata input dan *dataset*, semakin mirip kedua data tersebut. Dan bisa diketahui pertanyaan apa yang paling mirip dengan pertanyaan yang ada pada *dataset*.

Hasil dari algoritma yang telah disusun kemudian dites dengan menggunakan kata “perang dingin”. Berikut hasil dari tokenizing kolom pertanyaan pada *dataset*:

Gambar 2. Hasil uji coba tokenizing (perang dingin)

[illegible]

Kemudian, hasil dari vektorisasi TF-IDF direduksi menjadi *array* satu dimensi dan dilakukan proses *cosine similarity*. Hasil dari *cosine similarity* adalah berupa *array* yang berisi nilai yang mewakilkan kemiripan dari kata input dan kolom pertanyaan pada *dataset*. *Array* tersebut kemudian diurut dari nilai terkecil ke nilai terbesar. Terlihat indeks empat dan tiga memiliki nilai tertinggi diantara kolom pertanyaan lain.

Gambar 3. Hasil uji coba pembobotan TF-IDF (2)
(perang dingin)

REFERENCES

- [1] Allen, J. (1995). *Natural Language Understanding* 2 nd Edition. Benjamin-Cummings Publishing Company.
- [2] Amin, F., 2012. Sistem Temu Kembali Informasi dengan Metode Vector Space Model. *Jurnal Sistem Informasi Bisnis*, 2(2), pp. 78-83.
- [3] Badan Pusat Statistik Kota Bandung. Jumlah Wisatawan Menurut Jenis (Jiwa), 2018-2020. Diakses pada 15 Juni, 2022, dari <https://bandungkota.bps.go.id/indicator/16/251/1/jumlah-wisatawan-menurut-jenis.html>
- [4] Chopra, A., Prashar, A., & Sain, C. (2013). Natural Language Processing. *INTERNATIONAL JOURNAL OF TECHNOLOGY ENHANCEMENTS AND EMERGING ENGINEERING RESEARCH*
- [5] Khanna, A., Pandey, B., Vashishta, K., Kalia, K., Pradeepkumar, B., & Das, T. (2015). A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence *International Journal of u- and e- Service, Science and Technology*, Vol. 8(7), pp. 277-284..
- [6] Kurniawan, A. Solihin, F., dan Hastarita, F. 2014. Perancangan dan Pembuatan Aplikasi Pencarian Informasi Beasiswa dengan Menggunakan Cosine Similarity. *Jurnal SimanteC Volume 4*, Nomor 2.
- [7] Museum Konferensi Asia Afrika (mkaa). Registrasi Sim Kuring. Diakses pada 15 Juni, 2022, dari <https://www.asiafricamuseum.org/halaman/registrasi-simkuring>
- [8] Nila, E., & Afrianto, I. (2015) RANCANG BANGUN APLIKASI CHATBOT INFORMASI OBJEK WISATA KOTA BANDUNG DENGAN PENDEKATAN NATURAL LANGUAGE PROCESSING. *Jurnal Ilmiah Komputer & Informatika (KOMPUTA)*.
- [9] Nurdiana, O., Jumadi., dan Nursantika, D. 2016. Perbandingan Metode Cosine Similarity dengan Metode Jaccard Similarity pada Aplikasi Pencarian Terjemahan Al-Qur'an dalam Bahasa Indonesia. *Jurnal Online Informatika Volume 1*, Nomor 1.
- [10] Nurjannah M., Hamdani & Astuti, I. F. 2013. Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Text Mining. *Jurnal Informatika Mulawarman*, Vol 8, No 3. Universitas Mulawarman, (Online), (<http://e-journals.unmul.ac.id/index.php/JIM/article/view/113>), diakses 15 Juni 2022
- [11] Riyani, A., Zidny, M., & Burhanuddin, A. (2019). Penerapan Cosine Similarity Dan Pembobotan Tf-Idf Untuk Mendeteksi Kemiripan Dokumen. 2(1), 23-27

- [12] Rudiyanto, N. (2005). PERANCANGAN DAN IMPLEMENTASI PERANGKAT LUNAK NATURAL LANGUAGE PROCESSING UNTUK PENGEMBANGAN CHAT BOT BERBAHASA INDONESIA
- [13] S. Andayani and A. Ryansyah, "Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen," *JuSiTik J. Sist. dan Teknol. Inf. Komun.*, vol. 1, no. 1, p. 53, 2017
- [14] S. Perez-Soler, S. Juarez-Puerta, E. Guerra and J. de Lara, "Choosing a Chatbot Development Tool," in *IEEE Software*, vol. 38, no. 4, pp. 94-103, July-Aug. 2021, doi: 10.1109/MS.2020.3030198.
- [15] Ye, J. 2014. Vector Similarity Measures of Simplified Neutrosophic Sets and Their Application in Multicriteria Decision Making. *Internasional Journal of Fuzzy Systems* Volume 16, Nomor 2.