

PENERAPAN METODE TF-IDF DAN N-GRAM PADA PENGEMBANGAN APLIKASI CHATBOT BERBASIS LINE UNTUK LAYANAN PUBLIK KESEHATAN DI KOTA MALANG

Dhebys Suryani Hormansyah¹, Indinabilah Aulia²

Teknik Informatika, Teknologi Informasi, Politeknik Negeri Malang,
¹dhebys.polinema@gmail.com, ²indinabilahaulia@gmail.com

Abstrak

Teknologi kecerdasan buatan saat ini dapat diolah dengan berbagai macam bentuk, seperti *ChatBot* dengan berbagai metode, salah satunya menggunakan TF-IDF dan N-GRAM. TF-IDF merupakan sebuah metode dengan menghitung bobot masing-masing kata dalam suatu pertanyaan yang nantinya akan dicocokkan dengan dataset, sedangkan N-GRAM merupakan metode dimana sebagai ekstrasi kalimat masukan dari user yang nantinya akan dimasukkan ke dalam dataset. Hasil penelitian dapat diperoleh bahwa *Question-Answering* dan pemberian informasi baru dari user dalam bentuk *ChatBot* menggunakan TF-IDF dan N-Gram proses pengurangan data yang relevan dengan dataset.

Kata kunci : *Chatbot, N-Gram, Tf-Idf, Chatbot, Bot Line*

1. Pendahuluan

Informasi yang dibutuhkan masyarakat makin hari akan terus meningkat. Efisiensi waktu serta penyampaian sangat mendukung akan berkembangnya sebuah platform informasi. Penyampaian informasi yang didukung oleh perkembangan teknologi terbaru akan memudahkan masyarakat pengguna dalam mendapatkan informasi yang dibutuhkan. Semakin dimudahkannya sebuah media informasi maka akan semakin berkembangnya Kota tersebut. Layanan Publik merupakan suatu media yang disediakan oleh pemerintah untuk dapat memberikan informasi terkini kepada masyarakat. Dengan tersedianya layanan publik masyarakat akan dipermudah dalam mencari informasi yang diinginkan dan diharapkan masyarakat dapat memanfaatkan layanan tersebut maksimal mungkin.

Terdapat website untuk menyediakan informasi layanan publik di Kota Malang yaitu www.malangkota.go.id. Pengunjung website akan memperoleh informasi mengenai layanan kesehatan pada website tersebut, serta dapat melihat informasi yang disediakan. Namun dalam penyampaian informasi dirasa kurang interaktif, dimana pengunjung diharuskan dengan jeli memilah satu-persatu data informasi yang tersedia hingga mendapatkan yang dibutuhkan. Penyampaian informasi pada website tersebut hanya berbentuk table dan tidak terdapat *customer service*. *Customer*

service akan sangat berguna pada sebuah website layanan publik dengan didukung mengikuti trend teknologi yang ada.

Terdapat sebuah sistem aplikasi yang dapat digunakan sebagai pengganti *customer service* berupa sistem aplikasi *chatbot*, dimana pengaplikasiannya akan ditaruh di website tersebut. Berdasarkan sistem yang sudah ada akan dikembangkan lagi dan juga meningkatkan hasil respon dari sistem yang sebelumnya tidak bisa menampilkan hasil sesuai detail yang diinginkan masyarakat pengguna. Pengembangan sistem informasi mengikuti kebutuhan masyarakat pengguna yaitu ChatBot Line.

Oleh karena itu untuk mengatasi masalah tersebut dapat dibuatkan sebuah system yang dapat digunakan sebagai pengganti *customer service* berupa system aplikasi *chatbot*. Sering kita ketahui bahwa *customer service* diharuskan untuk *standby* 24 jam non-stop. *Chatbot* sendiri merupakan sebuah program computer yang dirancang untuk mensimulasikan sebuah percakapan atau komunikasi yang interaktif kepada pengguna (manusia) melalui bentuk teks, suara, dan visual. Dengan menggunakan metode TF-IDF sebagai respon tanya-jawab dan metode N-gram sebagai metode yang memproses masukkan informasi dari user/masyarakat pengguna, dimana akan di terapkan dalam pengerjaan aplikasi *chatbot* Line pada layanan public kesehatan di Kota Malang.

2. Landasan Teori

a. ChatBot

Chatbot adalah salah satu sistem cerdas yang dihasilkan dari pemrosesan Bahasa Alami atau Natural Language Processing (NLP) yang merupakan salah satu cabang dari Kecerdasan Buatan atau Artificial Intelligence (AI). NLP mempelajari komunikasi antara manusia dengan computer melalui Bahasa alami Fanani (2012). Chatbot memungkinkan manusia dapat berkomunikasi dengan mesin menggunakan perantara Bahasa alami. Bentuk komunikasi yang terjadi adalah melalui percakapan menggunakan media tulisan.

Percakapan dengan chatbot dapat berupa obrolan biasa atau obrolan pada tema - tema tertentu yang melibatkan disiplin ilmu yang lain. Percakapan yang terjadi antara komputer dengan manusia merupakan bentuk respon dari program yang telah dideklarasikan pada database program pada computer Fanani (2012). Kemampuan komputer dalam menyimpan banyaknya data tanpa melupakan satu pun informasi yang disimpannya digabungkan dengan kepraktisan bertanya pada sumber informasi langsung dibandingkan dengan mencari informasi sendiri serta kemampuan learning yang dimilikinya menyebabkan chatbot adalah customer service yang handal.

Contoh percakapan dengan chatbot:

User: siapa nama kamu ?

Bot : nama saya Bot

Chatbot akan menjawab sesuai yang tersedia pada dataset, dimana dataset dibangun oleh penulis dengan bantuan perhitungan metode sehingga sistem dapat menganalisa kemiripan antara pertanyaan harus di jawab dengan respon yang seperti apa.

b. Line

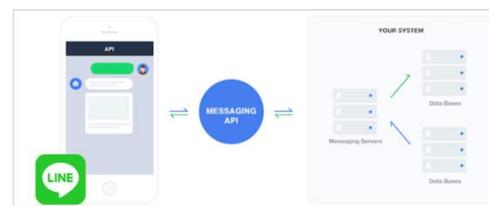
Line adalah suatu aplikasi yang digunakan untuk kegiatan berkirim pesan (messenger / chatting) secara gratis di perangkat smartphone. Namun, aplikasi Line sebenarnya juga bisa disebut sebagai aplikasi jejaring sosial karena terdapatnya fitur timeline sebagai wadah untuk berbagi status, pesan suara, video, foto, kontak dan informasi lokasi. Dengan aplikasi Line kita juga bisa melakukan voice call maupun video call secara real time dan gratis. Line disediakan di semua perangkat smartphone dan di semua sistem operasi mobile: Android, iPhone / iOS, Nokia / Windows Phone, Blackberry dan juga PC (komputer yang bersistemkan Mac OS ataupun

Windows) Amin (2012).

Line pada system chatbot kali ini akan berperan sebagai media perantara atau *user interface* yang akan berhubungan langsung dengan user/masyarakat pengguna. Dimana chatbot yang dibuat akan berupa bot chat dalam sebuah *grup-room* maupun *room-chatting* sendiri.

c. Messaging Api Line

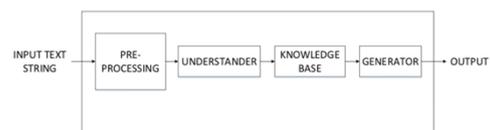
Messaging Api Line memungkinkan data yang dikirimkan melalui antar server aplikasi system bot dengan platform Line. Saat system bot mengirimkan pesan bot system, sebuah webhook akan dipicu dan platform Line akan mengirimkan permintaan ke URL webhook system bot. Server system bot kemudian akan mengirimkan permintaan ke platform Line untuk menanggapi pengguna. Permintaan dikirim melalui HTTPS dalam format JSON. Berikut alur system Messaging Api Line pada sebuah system bot : Amin (2012).



Gambar 2.1 Alur Sistem Messaging Api Line

d. Natural Language Processing

Natural Language Processing (NLP) merupakan salah satu cabang ilmu AI yang berfokus pada pengolahan bahasa natural. Bahasa natural adalah bahasa yang secara umum digunakan oleh manusia dalam berkomunikasi satu sama lain. Bahasa yang diterima oleh komputer butuh untuk diproses dan dipahami terlebih dahulu supaya maksud dari user bisa dipahami dengan baik oleh computer.



Gambar 2.2 Natural Language Processing

Ada berbagai terapan aplikasi dari NLP. Diantaranya adalah Chatbot (aplikasi yang membuat user bisa seolah-olah melakukan komunikasi dengan computer), Stemming atau Lemmatization (pemotongan kata dalam bahasa tertentu menjadi bentuk dasar pengenalan fungsi setiap kata dalam

kalimat), Summarization (ringkasan dari bacaan), Translation Tools (menterjemahkan bahasa) dan aplikasi-aplikasi lain yang memungkinkan komputer mampu memahami instruksi bahasa yang diinputkan oleh user Fanani (2012).

e. TF-IDF

Metode Term Frequency-Inverse Document Frequency (TF-IDF) adalah cara pemberian bobot hubungan suatu kata (term) terhadap dokumen.

1. Perhitungan Term Frequency (tf) menggunakan persamaan (2.1)

$$tf = tf_{ij} \quad (2.1)$$

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 .

2. Perhitungan Inverse Document Frequency (idf), menggunakan persamaan (2.2)

$$idf_i = \log N/df_i \quad (2.2)$$

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

3. Perhitungan term frequency Inverse Document Frequency (tfidf), menggunakan persamaan (2.3)

$$W_{ij} = tf_i \times \log (D/df_i) \quad (2.3)$$

Dengan W_{ij} adalah bobot dokumen, N adalah Jumlah dokumen yang diambil 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 (df_i) [8].

f. N-GRAM

Ekstraksi akan didasarkan pada algoritma pembagian N-gram. Disini N berartikan besaran

nilai dari kata untuk dianggap sebagai satu kesatuan untuk berhubungan metadata-nya. Sebagai contoh, untuk kalimat " the cow jumps over the moon ". Jika N = 2 (dikenal sebagai bigrams), maka n-grams akan menjadi:

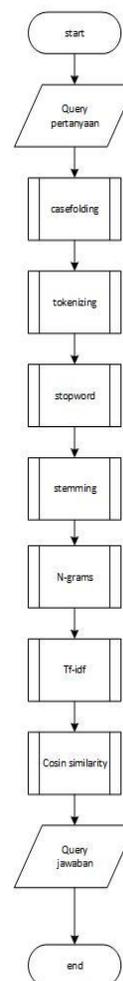
- the cow
- cow jumps
- jumps over
- over the
- the moon

Jika X = Jumlah kata dalam kalimat (K), jumlah ngrams untuk kalimat K akan menjadi:

$$N \text{ grams}_K = X - (N-1)$$

g. Proses Tanya-Jawab

Kumpulan data informasi tersebut disebut dengan *knowledge base*. *Knowledge Base*(dataset) merupakan sebuah tempat dimana data yang di masukkan sudah melalui proses *tokenized* dan dianalisis lebih jauh, disinilah NLP melakukan prosesnya.

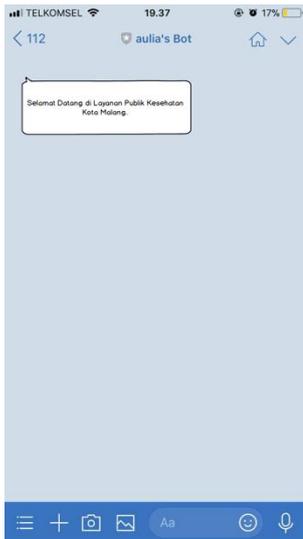


Gambar 2.3 Alur proses tanya jawab pengguna dengan sistem

3. Hasil dan Pembahasan

a. Tampilan pada Line

Merupakan proses sistem mengalokasikan kebutuhan-kebutuhan sistem baik perangkat keras maupun perangkat lunak dengan membentuk arsitektur sistem secara keseluruhan. Perancangan perangkat lunak melibatkan identifikasi dan penggambaran abstraksi sitem dasar perangkat lunak dan yang berhubungan.



Gambar 3.1 Tampilan awal chatbot

Pada tampilan awal tersebut akan muncul *greeting message* sebagai default system (tampilan pada chat room)



Gambar 3.2 Tampilan menu (1)

Pada tampilan didalam menu cek terdapat 4 fitur menu salah satunya RS-Puskemas, dengan mengklik menu tersebut akan muncul contoh text.

b. Konteks Pertanyaan

Konteks pertanyaan merupakan kumpulan pertanyaan yang nantinya akan membantu system dalam mengolah jawaban dan memberikan jawaban yang sesuai. Konteks pertanyaan sebagai berikut :

Tabel 3.1 Konteks Pertanyaan

No	Pertanyaan	Keterangan System Menjawab
1	Dimana alamat praktek dokter ..(nama dokter)..	System akan menampilkan alamat praktek dokter sesuai nama dokter yang di tanyakan
2	Alamat praktek dokter ..(nama dokter)..	System akan menampilkan alamat praktek dokter sesuai nama dokter yang di tanyakan
3	Dimana praktek dokter ..(nama dokter)..	System akan menampilkan alamat praktek dokter sesuai nama dokter yang di tanyakan
4	Jadwal praktek dokter ..(nama dokter)..	System akan menampilkan jadwal (jam, hari, buka dan tutup) dan alamat praktek dokter sesuai nama dokter yang di tanyakan
5	Jam berapa praktek dokter ..(nama dokter)..	System akan menampilkan jam berapa saja dokter tersebut melayani pasien beserta alamat praktek dokter

4. Kesimpulan dan Saran

Dari hasil penelitian dan perancangan yang telah dibuat, maka dapat diperoleh kesimpulan bahwa Question-Answering dalam bentuk ChatBot menggunakan N-Gram, TF-IDF dan Cosine Similarity dapat berkomunikasi dan menyampaikan informasi.

Daftar Pustaka:

Amin, Fatkhul, 2012. "Sistem Temu Kembali Informasi dengan Metode Vector Space Model". Ejournal undip. [1]
 Anonim. <https://developers.line.me/>, diakses pada 23 Januari 2018.] Betha Sidik. 2001. "Pemrograman Web PHP". Bandung : Penerbit Informatika.

- Fanani, A.Z. 2012. "Sistem *Costumer Service* Cerdas Menggunakan Metode *Fuzzy String Matching* Pada *E-Commerce*". Jurusan Teknik Elektro. Institut Sepuluh Nopember
- Shah, R., Lahoti, S., dan Prof Lavanya. K. 2017. "An Intelligent Chat-bot using Natural Language Processing". Department of Computer Engineering. VIT University.
- Utama, Y.P. 2017. "Aplikasi *Chatbot* Berbasis Web Pada Sistem Informasi Layanan Publik Kesehatan Di Malang Dengan Menggunakan Metode *Tf-Idf Dan Cosine Similarity*". Jurusan Teknologi Informasi. Politeknik Negeri Malang.