PENERAPAN METODE TF-IDF UNTUK CHATBOT PADA SISTEM INFORMASI PELAYANAN PERCETAKAN ONLINE

Dimas Wahyu Wibowo[1], Moch. Zawaruddin Abdullah[2], Joshua Davian Kristanto[3]

[1,2,3] Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang
[1] dimas.w@polinema.ac.id, [2] zawaruddin@polinema.ac.id, [3] joshuadavian@gmail.com

***Abstrak*— Pesatnya perkembangan teknologi dalam bidang perekonomian di Indonesia berdampak besar terhadap kegiatan bisnis yang dilakukan oleh para pelaku usaha. Banyak pelaku usaha yang memanfaatkan perkembangan teknologi untuk menunjang usaha mereka. Sebagai hasilnya, banyak dari usaha-usaha yang telah mengalami digitalisasi, sehingga proses jual-beli produk bisa dilakukan secara *online*. *Customer Service* merupakan profesi yang berfungsi untuk memberikan informasi kepada *user* atau *customer* mengenai produk usaha atau pelayanan yang ditawarkan oleh sebuah tempat usaha. *Customer Service* diharapkan dapat memberikan informasi yang jelas dan akurat sesuai dengan pertanyaan yang diajukan oleh *customer*. Pada umumnya, *customer service* tidak bisa merespon pertanyaan dari *customer* secara tepat waktu. Karena adanya jam kerja tertentu yang ditetapkan tiap tempat usaha sehingga *customer service* tidak bisa merespon pertanyaan dari user setiap waktu.**

Berdasarkan dari permasalahan tersebut, penelitian ini bertujuan untuk membuat sebuah sistem *chatbot* yang dapat merespon setiap pertanyaan dari *user* tanpa terikat batas waktu tertentu. Hasilnya adalah *chatbot* dapat memberikan informasi yang akurat mengenai apa yang ditanyakan oleh *customer*. Pada penelitian ini digunakan metode TF-IDF dan *Cosine Similarity*. Algoritma TF-IDF berfungsi untuk memberikan bobot/*weight* pada setiap dokumen jawaban yang tersedia dalam *database*. Dilanjutkan dengan algoritma *Cosine Similarity* untuk mengukur kesamaan antara *input* dari *customer* dengan data jawaban yang ada di sistem. Sistem kemudian memberikan respon berdasarkan tingkat kesamaan tertinggi sehingga *customer* bisa memperoleh informasi yang mereka inginkan.

Kata kunci—Customer Service, Chatbot, TF-IDF, Cosine Similarity)

I. PENDAHULUAN

Seiring berjalannya dari waktu ke waktu, teknologi mengalami berbagai macam perkembangan yang pesat dalam kehidupan umat manusia. Perkembangan ini tentunya mempengaruhi banyak bidang, salah satunya dalam bidang perekonomian. Hal ini membuat para pelaku binis memanfaatkan teknologi yang ada untuk meningkatkan pemasaran produk mereka. Dari yang semula aktivitas jual-beli dilakukan secara manual, mulai digantikan dengan teknologi. Hal ini memudahkan antara calon pembeli dan penjual untuk melakukan transaksi melalui akses *website* [1].

Sebagai contoh dalam aktivitas jual beli dikenal dengan pekerjaan *Customer Service* (CS). *Customer Service* adalah setiap kegiatan yang ditujukan untuk memberikan kepuasan melalui pelayanan yang diberikan seseorang kepada klien dalam menyelesaikan masalahnya. *Customer Service* sendiri dimanfaatkan oleh pelaku usaha untuk memberikan informasi kepada pengguna mengenai produk usaha yang dijualnya. Seringkali, *Customer Service* mengalami banyak kendala dalam merespon permintaan informasi dari pengguna. Salah satu diantaranya adalah CS tidak selalu ON time saat pengguna menanyakan suatu informasi. CS juga tidak dapat merespon pengguna selama 24 jam dikarenakan terbatas oleh jam kerja dari tempat kerja yang bersangkutan.

Salah satu cara dalam mengurangi beban CS dalam menangani berbagai permintaan informasi dari pengguna adalah dengan menggunakan *chatbot*. *Chatbot* adalah program komputer yang dapat melakukan percakapan melalui media tulisan [2]. Dengan menggunakan *chatbot*, pengguna bisa memperoleh informasi tanpa harus dibatasi oleh jarak dan waktu. Pengguna bisa menanyakan suatu informasi dan *chatbot* akan merespon dengan memberikan jawaban sesuai kata kunci yang ditanyakan. Beberapa penelitian yang menggunakan *chatbot* sebagai media pendukung CS dalam mengolah informasi adalah penelitian Dhebys [3], penelitian tersebut menggunakan *chatbot* untuk layanan publik kesehatan di Kota Malang dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

TF-IDF merupakan metode algoritma yang menentukan frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan [4]. Metode ini untuk menghitung nilai TF dan IDF pada setiap kata di setiap dokumen. Semakin besar jumlah kemunculan suatu term (TF tinggi) dalam dokumen, semakin besar pula bobotnya atau akan memberikan nilai kesesuaian yang semakin besar [5]. Metode ini kemudian dikombinasikan dengan metode *Cosine Similarity* yang berfungsi untuk membandingkan kemiripan antar dokumen, dalam hal ini yang dibandingkan adalah *query* dengan dokumen latih.

Gerai Percetakan Radja Printing merupakan usaha percetakan yang berdiri tahun 2016. Berlokasi di kota Probolinggo, gerai percetakan ini bergerak di bidang percetakan dan penjualan *banner*. Permasalahan yang dihadapi oleh usaha ini yaitu tidak adanya sistem yang

memungkinkan pembeli untuk melakukan pemesanan secara *online*. Permasalahan lain yang dihadapi oleh usaha ini yaitu adanya keterbatasan jam kerja *customer service* dalam merespon pertanyaan yang diajukan oleh calon pembeli. Dikarenakan usaha ini memiliki jam kerja yang dibatasi oleh aturan perusahaan, sedangkan *customer* seringkali membutuhkan informasi yang cepat dan tepat. Berdasarkan pada studi kasus yang ada, penelitian ini menerapkan metode TF-IDF dalam Sistem Informasi Penjualan Online Gerai Radja Printing, yang diharapkan memudahkan serta menyelesaikan permasalahan yang telah dipaparkan diatas. Aplikasi *chatbot* ini diharapkan dapat menjawab pertanyaan *customer* dengan penjelasan yang mudah dipahami.

II. LANDASAN TEORI

A. Chatbot

Chatbot adalah program komputer yang dapat melakukan percakapan melalui media tulisan [6]. Percakapan dapat terjadi dengan manusia atau chatbot yang lain. Secara harfiah *chatbot* berasal dari dua kata, yaitu *chat* dan *bot*. Dalam dunia komputer, chat dapat diartikan sebagai kegiatan komunikasi yang menggunakan sarana tulisan. Sedangkan *bot* merupakan program yang memiliki sejumlah data yang bila diberikan input akan menghasilkan output sebagai jawaban [7].

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

Metode TF -IDF *adalah* cara pemberian bobot hubungan suatu kata (term) terhadap dokumen.

1) TF Murni (RAW TF)

Merupakan nilai TF diberikan berdasarkan jumlah kemunculan suatu term di dokumen.

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

Ket:

tf : *Term frequency*

tf_{ij} : *Banyaknya muncul term i pada dokumen j*

2) Inverse Document Frequency (IDF)

Merupakan sebuah perhitungan dari bagaimana term di distribusikan secara luas pada koleksi dokumen yang bersangkutan.IDF menunjukkan hubungan ketersediaan sebuah term dalam seluruh dokumen. Semakin sedikit jumlah dokumen yang mengandung term mengandung term yang dimaksud., maka nilai IDF semakin besar.

$$IDFj = log(D / dfj) \tag{2}$$

Ket:

IDF : *Inverse Document Frequency*

D : *Jumlah dokumen yang terambil dalam sistem*

dfj : *Banyak dokumen dalam koleksi memuat term*

3) Term Weightening TF-IDF

Adalah penggabungan dari formula perhitungan *raw* TF dengan formula IDF dengan cara mengalikan nilai TF dengan nilai IDF.

$$w_{ij} = tf_{ij}\ x\ \log(\frac{D}{df_j}) \tag{3}$$

w_{ij} : *Bobot term terhadap dokumen*

tf_{ij} : *Jumlah kemunculan term dalam dokumen*

D : *Jumlah dokumen yang terambil dalam sistem*

dfj : *Banyak dokumen dalam koleksi memuat term*

C. Cosine Similarity

Cosine similarity berfungsi untuk membandingkan kemiripan antar dokumen, dalam hal ini yang dibandingkan adalah query dengan dokumen latih [8]. Dalam menghitung cosine similarity, pertama yang di lakukan yaitu melakukan perkalian skalar antara query dengan dokumen kemudian dijumlahkan, setelah itu melakukan perkalian antara panjang dokumen dengan panjang query yang telah dikuadratkan, setelah itu di hitung akar pangkat dua. Selanjutnya hasil perkalian skalar tersebut di bagi dengan hasil perkalian panjang dokumen dan query.

$$\cos Sim(\mathrm{d_j}, \mathrm{q_k}) = \frac{\sum_{i=1}^{n}(\mathrm{td_{ij}} \times \mathrm{tq_{ik}})}{\sqrt{\sum_{i=1}^{n} \mathrm{td_{ij}}^2 \times \sum_{i=1}^{n} \mathrm{tq_{ik}}^2}} \tag{4}$$

Ket:

cosSim(d_j, q_k) : *Cosine Similarity*

td_{ij} : *term ke-i dalam vektor untuk dokumen j*

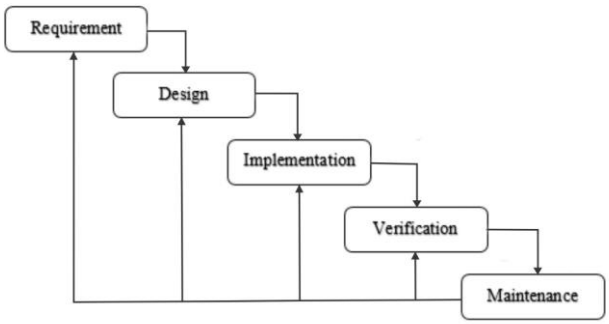
tq_{ik} : *term ke-I dalam vektor untuk query k*

n : *jumlah term yang unik dalam data set*

D. Recall and Precision

Recall adalah proporsi jumlah dokumen yang dapat ditemukan-kembali oleh sebuah proses pencarian di sistem IR. Rumusnya: Jumlah dokumen relevan yang ditemukan dibagi dengan jumlah semua dokumen relevan di dalam koleksi [9]. *Precision* adalah proporsi jumlah dokumen yang ditemukan dan dianggap relevan untuk kebutuhan si pencari informasi. Rumusnya: Jumlah dokumen relevan yang ditemukan dibagi dengan jumlah semua dokumen yang ditemukan. Sedangkan *Precision* dapat diartikan sebagai kepersisan atau kecocokan (antara permintaan informasi dengan jawaban terhadap permintaan itu) [10].

III. METODOLOGI PENELITIAN

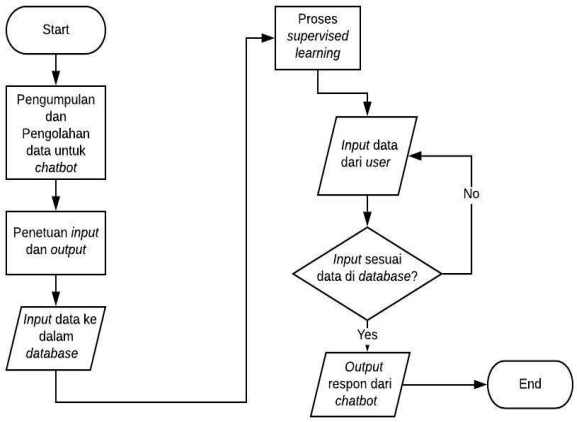


Gambar 1. Metode Waterfall

Dalam penelitian ini, pengembangan sistem aplikasi menggunakan metode pengembangan waterfall, dimana setiap tahapan di metode Waterfall menghasilkan satu atau lebih dokumen yang sudah disetujui. Tahap berikutnya tidak dapat dimulai sebelum tahapan sebelumnya selesai. Pada Gambar 1, metode *waterfall* dimulai dari analisa kebutuhan sistem yang akan dibuat. Dilanjutkan dengan proses desain rancangan aplikasi berdasarkan analisa kebutuhan. Selanjutnya hasil dari desain tersebut diubah menjadi kode program yang saling diintegrasikan. Tahap berikutnya yaitu melakukan *test* atau pengujian terhadap aplikasi yang telah dibuat. Terakhir adalah proses *maintenance* dimana dilakukan pemeliharaan dan perbaikan terhadap sistem apabila masih terjadi *bug*.

IV. ANALISIS DAN PERANCANGAN

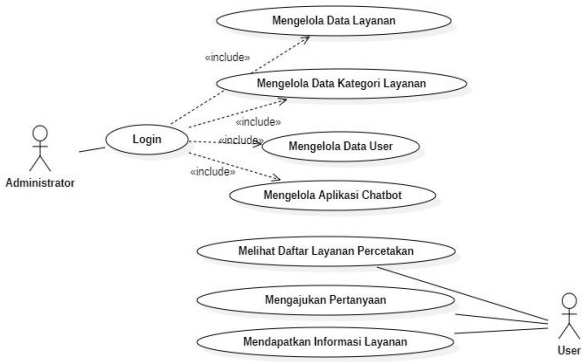
A. *Flowchart*



Gambar 2. *Flowchart*

Flowchart adalah suatu bagan dengan simbol-simbol tertentu yang menggambarkan urutan proses secara mendetail dan hubungan antara suatu proses (instruksi) dengan proses lainnya dalam suatu program. Gambar 2 menunjukkan alur proses yang dilakukan, dimulai dengan proses pengumpulan data dan pengolahan data, kemudian dilanjutkan dengan proses penentuan *input* dan *response* dari chatbot yang akan dibuat. Setelah itu dilanjutkan dengan proses *input* data ke dalam *database*, dan kemudian *chatbot* akan melakukan proses *supervised learning* untuk menambah pengetahuan pada sistem *chatbot* itu sendiri. *User* kemudian melakukan *input* pertanyaan pada sistem *chatbot*. Setelah itu, sistem akan memproses *input* pertanyaan dari *user*, apabila pertanyaan tersebut cocok dengan yang ada di database, maka sistem akan memberikan respon jawaban sesuai dengan *input* yang telah diberikan. Apabila tidak cocok, maka *user* akan diberikan pemberitahuan untuk melakukan *input* pertanyaan ulang.

B. *Use Case*



Gambar 3. *Use Case*

Use Case adalah interaksi yang saling berkesinambungan antara aktor dan juga sistem guna memperoleh kebutuhan fungsional dari sistem yang ada. Gambar 3 menunjukkan *use case* yang dimiliki dari masing-masing aktor. Pada sistem ini, terdapat 2 aktor yaitu *administrator* dan *user*. *Administrator* dapat melakukan pengelolaan data *chatbot*, *user*, dan layanan percetakan. Sedangkan *user* dapat mengajukan pertanyaan ke *chatbot* dan memperoleh informasi jawaban.

V. IMPLEMENTASI

Aplikasi ini dibangun menggunakan bahasa pemrograman PHP, *JavaScript*, *HTML*, *CSS* serta menggunakan database MySQL.

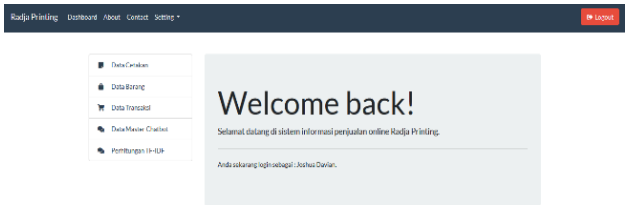
A. *Halaman Login Admin*

Berikut adalah halaman utama admin sistem aplikasi *chatbot*, halaman ini ditampilkan pada awal sistem ditampilkan. Terdapat halaman login untuk admin sebelum bisa mengakses sistem.

Gambar 4. *Login* Admin

B. *Halaman Dashboard Admin*

Halaman menu admin berisikan mengenai aksi-aksi yang digunakan untuk mengolah data dengan menggunakan TF-IDF.



Gambar 5. *Dashboard* Admin

C. *Halaman Chatbot User*

Halaman *chatbot* digunakan untuk media komunikasi antara *customer* dengan sistem dimana *customer* bisa memasukkan pertanyaan tentang informasi yang mereka inginkan.

Gambar 6. *Chatbot User*

VI. PENGUJIAN DAN PEMBAHASAN

Pengujian yang dilakukan pada sistem aplikasi *chatbot* yaitu menggunakan *Blackbox Testing* yang dimaksudkan agar hasil respon dari sistem sesuai dengan yang diinginkan. Pengujian juga dilakukan menggunakan metode *Recall and Precision*. Pengujian dilakukan terhadap 25 dokumen dengan *query* yang berbeda sebanyak 3 kali. Berikut adalah hasil pengujian *recall and precision* yang telah dilakukan.

A. *Percobaan 1*

$$Recall = \frac{94}{100} \times 100\% = 94\%$$

$$Precision = \frac{94}{100} \times 100\% = 94\%$$

B. *Percobaan 2*

$$Recall = \frac{60}{100} \times 100\% = 60\%$$

$$Precision = \frac{60}{100} \times 100\% = 60\%$$

C. *Percobaan 3*

$$Recall = \frac{48}{100} \times 100\% = 48\%$$

$$Precision = \frac{48}{100} \times 100\% = 48\%$$

VII. KESIMPULAN DAN SARAN

A. *Kesimpulan*

Berdasarkan pembahasan yang telah dilakukan pada bab I hingga bab VI, maka dapat disimpulkan bahwa:

- Sistem ini dapat menerima *input* data dari admin sebagai dataset awal.
- Sistem ini merubah dataset awal menjadi *array* dan melakukan *preprocessing* dengan menghilangkan simbol dan kata yang tidak perlu/penting.
- Sistem ini dapat menghitung bobot TF-IDF pada term dalam setiap dokumen yang telah diproses sebelumnya.
- Dilakukan proses pengukuran kemiripan menggunakan *Cosine Similarity* untuk mencari respon terbaik berdasarkan *input* dari *user*.

B. *Saran*

- Optimasi sistem lebih ditingkatkan ketika memproses pencarian jawaban.
- Menggunakan atau menambah metode lain agar *chatbot* dapat menerima respon berupa permintaan pemesanan produk.

DAFTAR PUSTAKA

[1] R. Natadian Astuti dan M. Fatchan, “Perancangan Aplikasi Teknologi Chatbot untuk Industri Komersial 4.0,” *e-Prosiding SNasTekS 2019*, pp. 339-348, 2019.

[2] D. Suryani dan E. Larasati, “Aplikasi Chatbot Objek Wisata Jawa Timur Berbasis AIML,” *SMARTICS Journal*, pp. 47-54, 2017.

[3] D. Suryani dan Y. Putera, “Aplikasi Chatbot Berbasis Web Pada Sistem Informasi Layanan Publik Kesehatan di Malang dengan Menggunakan Metode TF-IDF,” *Jurnal Informatika Polinema*, pp. 224-228, 2018.

[4] A. Riyani dan A. Burhanuddin, “Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen,” *Jurnal Linguistik Komputasional*, pp. 23-27, 2019.

[5] D. Suryani dan I. Aulia, “Penerapan Metode TF-IDF dan N-Gram pada Pengembangan Aplikasi Chatbot Berbasis LINE untuk Layanan Publik Kesehatan di Kota Malang,” *Jurnal Informatika Polinema*, pp. 7-11, 2018.

[6] M. Astiningrum dan M. Shoburu Rohmah, “Implementasi NLP dengan Konversi Kata pada Sistem Chatbot Konsultasi Laktasi.,” *Jurnal Informatika Polinema*, pp. 46-52, 2018.

[7] H. Yamaguchi dan M. Mozgoyov, “A Chatbot Based on AIML Rules Extracted From Twitter Dialogues,” *Communication Paper Poznan*, pp. 37-42, 2018.

[8] R. Melita dan T. Dirjam, “Penerapan Metode Term Frequency Inverse Document Frequency dan Cosine Similarity pada Sistem Temu Kembali Informasi untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus:

Syarah Umdatil Ahkam)," *Jurnal Teknik Informatika*, pp. 149-164, 2018.

[9] Y. Dwi Listio, A. Zulkarnain dan A. Tirtana, "Pembuatan Sistem Pencarian Pekerjaan Menggunakan TF-IDF," *Jurnal Ilmiah Teknologi Informasi Asia*, pp. 91-100, 2019.

[10] I. N. Satria Paliwahet dan I. K. Gede Darma Putra, "Pencarian Informasi Wisata Daerah Bali Menggunakan Teknologi Chatbot," *Lontar Komputer*, pp. 144-153, 2017.