

INTEGRASI *NATURAL LANGUAGE PROCESSING* DALAM *CHATBOT MARKETING* (STUDI KASUS TOKO CAHAYA FAJAR)

MUHAMMAD RIZAL H¹, HASRIANA², TAMRAH³, MURSALIM⁴,
BUTSIARAH⁵, NURAIIDA LATIF⁶, ELLY WARNI⁷, MUHAMMAD
ICHFAN ASKAR⁸

^{1,2,3,4,5,6,8}Program Studi Teknik Informatika, Universitas Teknologi Akba
Makassar

⁷Program Studi Teknik Informatika, Universitas Hasanuddin
Jln. Perintis Kemerdekaan KM. 9 No. 75 Makassar, Indonesia
e-mail: ¹rizal@unitama.ac.id, ²hasriana17@mhs.akba.ac.id

³tamrah@unitama.ac.id, ⁴mursalim@unitama.ac.id, ⁵butsirah@unitama.ac.id,
⁶nuraida@unitama.ac.id, ⁷elly@unhas.ac.id, tamsir@unitama.ac.id,
michfan@unitama.ac.id

ABSTRAK

Penelitian ini bertujuan untuk merancang aplikasi penjualan perlengkapan pernikahan dengan fitur chatbot dengan penerapan TF-IDF dan cosine similarity untuk meningkatkan akurasi sistem pendeteksi tanya jawab pada chatbot. Data ini diperoleh melalui observasi dan wawancara. Metode pengujian sistem menggunakan blackbox dan metode *Frequency-Inverse Document Frequency* (F-IDF). Hasil penelitian menunjukkan bahwa aplikasi untuk alat penjualan perlengkapan pernikahan dengan fitur chatbot menggunakan metode TF-IDF dan *cosine similarity* menampilkan hasil nilai dengan akurasi yang baik. Pengujian sistem menggunakan blackbox serta pengujian penerapan chatbot menggunakan pengujian validasi pada menggunakan *recall* dan *precision* menghasilkan nilai 90% dari hasil metode TF-IDF dan *cosine similarity* untuk akurasi tanya jawab chatbot.

Kata kunci: Aplikasi penjualan, *Chatbot*, TF-IDF, *Cosine similitarity*, Toko Cahaya Fajar terang.

I. PENDAHULUAN

Toko Cahaya Fajar Terang merupakan salah satu toko yang menyediakan perlengkapan pernikahan seperti baju pengantin, aksesoris penari dan aksesoris pengantin dan baju bodo serta jas tutup. Namun pada toko tersebut memiliki kendala data keuangan, data stok barang dan penjualan. Tidak adanya sistem yang memungkinkan untuk admin menyimpan data penjualan dan data stok barang oleh karena itu admin akan kesulitan menentukan apakah data stok barang masih ada. Admin mengalami keterbatasan dalam merespon pertanyaan yang diajukan oleh pelanggan, dikarenakan toko Cahaya Fajar memiliki jam kerja yang dibatasi oleh aturan pengelola New Makassar Mall, customer seringkali membutuhkan

informasi yang cepat dan tepat, oleh karena itu memerlukan bantuan sebuah aplikasi layanan seperti Fitur chatbot otomatis agar mempermudah Customer Servis (CS) dalam menjawab pertanyaan dari pelanggan mengenai hal tentang registrasi pembayaran, stok barang yang tersedia. CS mengalami banyak kendala dalam merespon pertanyaan yang berulang dari pelanggan.

Tidak diragukan lagi bahwa kecerdasan buatan (AI) telah menjadi teknologi baru yang menarik banyak perhatian dari para peneliti dan industri dalam dekade terakhir. Salah satu aplikasi AI yang diharapkan akan berkembang pesat dalam waktu dekat adalah agen percakapan, yang juga dikenal sebagai chatbot atau chatterbot. Agen percakapan merupakan sistem percakapan mesin yang dapat berinteraksi dengan pengguna manusia melalui bahasa percakapan alami. Banyak perusahaan dan organisasi saat ini menggunakan chatbot untuk berkomunikasi dengan pelanggan mereka, terutama dalam bidang e-commerce. Chatbot dapat membantu menjawab pertanyaan pelanggan secara cepat tanpa melibatkan manusia, sehingga meningkatkan kualitas layanan pelanggan dan mengalokasikan sumber daya manusia (Ngai et al., 2021).

Chatbot adalah aplikasi untuk berkomunikasi melalui pesan teks atau suara, yang digunakan dalam berbagai industri termasuk penjualan. Aplikasi chatbot dapat membantu perusahaan meningkatkan efisiensi penjualan dengan membantu pelanggan mencari produk yang sesuai dengan kebutuhan mereka, memberikan rekomendasi produk, dan membantu pelanggan dalam proses pembelian. Selain itu, chatbot dapat membantu perusahaan dalam menangani pertanyaan pelanggan dengan cepat dan efisien (Amalia & Wibowo, 2019; Mashud & Wisda, 2019). Aplikasi penjualan dengan chatbot dapat mengurangi beban CS dalam menjawab pertanyaan berulang dari pelanggan. Metode yang diperlukan untuk meningkatkan akurasi tanya jawab pada chatbot adalah dengan mengukur kesamaan kata. Chatbot membantu menangani masalah percakapan dengan pelayanan dan membalas pesan dari pelanggan secara otomatis.

Metode TF-IDF dan Cosine Similarity digunakan untuk meningkatkan akurasi chatbot dalam memberikan jawaban yang tepat. Metode klasifikasi juga dapat meningkatkan akurasi dengan presisi terbaik 95% dan recall terbaik 100%. Namun,

akurasi dipengaruhi oleh dataset yang digunakan. Penambahan proses koreksi kata tidak baku secara manual dapat membantu NLP dalam menghasilkan term yang dibutuhkan. Kekurangan pada penelitian ini adalah chatbot hanya dapat membalas pertanyaan secara otomatis dan cepat, tetapi kurang memastikan informasi yang disediakan oleh pelaku usaha tepat dan akurat (Saputro et al., 2021; Shobirin et al., 2020).

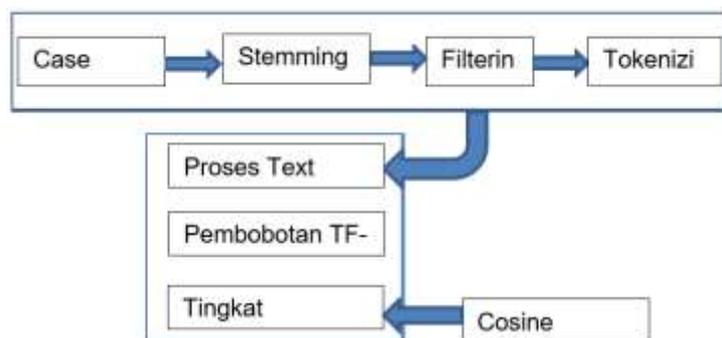
II. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini terdiri dari beberapa teknik, yaitu: (1) studi lapangan (observasi) sebagai teknik pengumpulan data, di mana informasi berupa dokumen maupun lisan yang dibutuhkan dalam pembuatan aplikasi penjualan perlengkapan pernikahan ini diambil untuk mendapatkan informasi yang diperlukan dalam penelitian; (2) wawancara sebagai teknik pengumpulan data melalui percakapan tatap muka antara narasumber dan pewawancara dengan tujuan untuk mendapatkan informasi yang tepat dari narasumber yang terpercaya; (3) dokumentasi sebagai teknik pengumpulan data dengan melihat catatan-catatan proses penjualan dan mengambil dokumentasi berupa gambar-gambar jualan; (4) analisis terhadap kelemahan sistem dalam proses penjualan, pencatatan laporan penjualan, dan data stok barang yang masih dilakukan secara manual. Pada proses pelayanan pelanggan mengenai komunikasi melalui sosial media dengan pelanggan masih terbatas waktu karena jam kerja toko Cahaya Fajar terbatas oleh jam kerja New Makassar Mall. Akan tetapi, pelanggan memerlukan informasi setiap saat tentang penyediaan stok barang yang dijual; (5) analisis kebutuhan perangkat keras dan lunak dalam pembuatan aplikasi; (6) pembobotan TF-IDF digunakan untuk menghitung frekuensi kemunculan kata, sedangkan Cosine Similarity digunakan untuk mengukur kesamaan antara dua vektor pada akurasi chatbot; (7) proses preprocessing teks dilakukan dengan melakukan seleksi data pada setiap dokumen yang sudah dibuat, meliputi case folding, tokenizing, filtering, dan stemming (Herwijayanti et al., 2018; Wahyuni et al., 2017).

III. HASIL DAN PEMBAHASAN

3.1 Metode Perancangan Sistem

Beberapa tahapan akan dilakukan pada penelitian ini yaitu tahap perancangan sistem untuk mengklasifikasi dokumen gambar 3.1 menjelaskan tentang alur proses perancangan sistem:



Gambar III.1 Alur Proses Penerapan Metode

proses penerapan metode yang melibatkan beberapa tahap. Tahap pertama adalah Case Folding, yang mengubah semua huruf kapital menjadi huruf kecil. Tahap kedua adalah Stemming, yaitu menghilangkan imbuhan kata dasar dari arti yang serupa seperti mengubah "terbaru" menjadi "baru". Tahap ketiga adalah Filtering, yang mengambil kata-kata tidak penting seperti "di", "dan", "saya", dan sebagainya dari hasil konten. Tahap keempat adalah Tokenizing, yaitu memasukkan setiap kata dari suatu kalimat. Selanjutnya, terdapat tahap pembobotan TF-IDF, di mana dilakukan penghitungan frekuensi kemunculan kata, dan tahap Cosine Similarity, yang digunakan untuk mengukur kesamaan antara dua vektor pada akurasi chatbot.

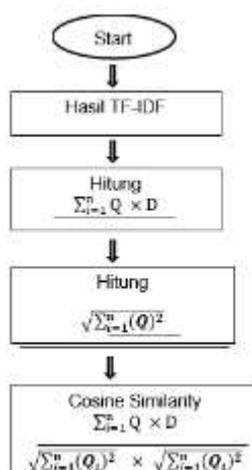
3.2 Pembobotan TF-IDF

Pada tahap ini dilakukan pembobotan kata/term pada dokumen testing dan dataset dokumen untuk menghitung nilai frekuensi kemunculan setiap kata dalam masing-masing dokumen yang ada dalam dataset. Gambar 3.3 menunjukkan proses pembobotan TF-IDF dengan flowchart perhitungan bobot berdasarkan nilai TF-IDF. Proses pembobotan TF-IDF dimulai dengan memasukkan dokumen uji atau Q dan data D sebagai isi data yang akan dibandingkan dengan dokumen Q. Kemudian, dilakukan penghitungan kemunculan kata di Q dan D untuk mendapatkan hasil

Term Frekuensi/kemunculan kata, selanjutnya hasil perhitungan kemunculan kata pada setiap dokumen DF dijumlahkan. Selanjutnya, dilakukan perhitungan IDF dimana $IDF = \text{LOG}(\text{jumlah dokumen/Dokumen frekuensi}) + 1$, dan terakhir dihitung nilai TF-IDF dengan rumus $TF-IDF = TF * IDF$ (term frekuensi * Index dokumen frekuensi) dan menampilkan hasil.

3.3 Hitung *Similaritas* menggunakan *Cosine similarity*

Pembobotan *TF-IDF* hasil tersebut digunakan untuk menghitung kemiripan dokumen *tes* dengan *dataset* dokumen menggunakan *cosine similarity*. Flowchart dapat dilihat pada Gambar 3.2 menjelaskan tentang tahapan menghitung *Similaritas* dengan *Cosine similarity*.



Gambar III.2 Tahapan Menghitung *Similaritas* Dengan *Cosine*

3.4 Pengujian Sistem

Untuk melakukan pengujian perangkat lunak menggunakan metode blackbox testing, perlu dilakukan analisis terhadap kebutuhan dan spesifikasi pengujian yang diperlukan. Penelitian ini membutuhkan beberapa hal untuk dapat melakukan blackbox testing, yaitu aplikasi yang akan diuji, sistem operasi yang digunakan, dan basis data yang digunakan dalam pengujian. Hasil dari analisis kebutuhan dan spesifikasi yang digunakan pada pengujian perangkat lunak dengan metode *blackbox testing*.

Selanjutnya perhitungan akurasi sistem dengan Pengujian menggunakan metode *Recall* and *Precision*. Pengujian dilakukan terhadap 10 dokumen dengan *query* yang berbeda yang ada pada tabel 3.3 hasil pengujian *recall and precision* yang telah dilakukan. Dari hasil pengujian *Black Box* maka dijelaskan sebagai berikut :

$$Recall = \frac{90}{100} \times 100\% = 90\%$$

$$Precision = \frac{90}{100} \times 100\% = 90\%$$

Berdasarkan hasil di atas, sistem chatbot memiliki tingkat akurasi sebesar 90% dalam mengoreksi dan menjawab pertanyaan dari pengguna. Pengujian akurasi dilakukan dengan membandingkan kesamaan kata antara pertanyaan yang diajukan oleh pengguna dengan koreksi (manual) yang dilakukan oleh chatbot. Kasus Sederhana yaitu, terdapat sebuah *query* (Q), dan 3 buah dokumen yaitu dokumen 1 (D1), dokumen 2 (D2), dan dokumen 3 (D3) .

Tabel III.1 Pertanyaan

Dokumen	Pertanyaan
D1	Ada baju pengantin koleksi terbaru ?
D2	Ada baju paseppi kita jual ?
D3	Ada baju bodo adat bugis ?
Q	Ada koleksi terbaru untuk baju pengantin ?

Penelitian akan dilakukan pada chatbot menggunakan metode text mining, dengan membandingkan dan memilih pertanyaan yang lebih spesifik dari pelanggan. Prosesnya meliputi *Preprocessing text*, *Tokenizing* untuk memecah kalimat menjadi kata, *Filtering* untuk menghilangkan tanda baca dan karakter, serta *Stemming* untuk mencari kata dasar.

Tabel III.2 Tabel *Text preprocessing*

Tokenizing	Stemming
ADA	ADA
BAJU	BAJU
PENGANTIN	PENGANTIN
KOLEKSI	KOLEKSI

TERBARU	BARU
PASEPPI	PASEPPI
KITA	KITA

Tabel 3.2 hasil Text preprocessing akan digunakan untuk menghitung Vector Space Model (TF-IDF) yang mencari jumlah term frekuensi dari kata dan index dokumen frekuensinya. Proses ini dilakukan pada tiga dokumen yang telah memiliki jawaban, ditambah satu dokumen query yang belum memiliki jawaban. Pembobotan term frekuensi dijelaskan pada Tabel 3.3

Tabel III.3 Proses Pembobotan *Term Frequency*

kata	TF				DF	IDF
	D1	D2	D3	Q		
ADA	1	1	1	1	4	0
BAJU	1	1	1	1	4	0
PENGANTIN	1	0	0	1	2	0.30103
KOLEKSI	1	0	0	1	2	0.30103
BARU	1	0	0	1	2	0.30103
PASEPPI	0	1	0	0	1	0.60206
KITA	0	1	0	0	1	0.60206

Dari hasil pembobotan Term Frekuensi dari tabel 3.3 proses selanjutnya yaitu melakukan pembobotan metode *TF-IDF* term frekuensi setelah itu masuk di *TF-IDF* jadilah vektor istilah pembobotan untuk menentukan nilai kalimat mana yang memiliki bobot nilai yang menghampiri pertanyaan dari Q membandingkan pada dokumen D1, D2, dan D3 pada proses perhitungan ini melakukan menggunakan metode *Cousin Similarity* proses. Tabel 3.4 menjelaskan proses perhitungan *TF-IDF*.

Tabel III.4 Proses Perhitungan TF-IDF

TF*IDF				
KATA	D1	D2	D3	Q
ADA	0	0	0	0
BAJU	0	0	0	0
PENGANTIN	0.301029996	0	0	0.30103
KOLEKSI	0.301029996	0	0	0.30103
BARU	0.301029996	0	0	0.30103
PASEPPI	0	0.60206	0	0
KITA	0	0.60206	0	0

Proses perhitungan akurasi manual terhadap *TF-IDF* yang telah dijabarkan

$$D1 : TFIDF_{t,d} = 1 \times 0$$

$$IDF_t = \log_{10} (4/1) = 0$$

Tabel III.5 Panjang Vektor

Panjang Vektor				
KATA	D1	D2	D3	D4
ADAT	0	0	0	0
BAJU	0	0	0	0
PENGANTIN	0.090619	0	0	0.090619
KOLEKSI	0.090619	0	0	0.090619
BARU	0.090619	0	0	0.090619
PASSEPI	0	0.362476	0	0
KITA	0	0.362476	0	0
	0.521399	1.042798	1.042798	0.796451

Tahap selanjutnya adalah melakukan perhitungan cosine similarity untuk membandingkan similaritas antara dokumen. Dalam perhitungan ini, dicari nilai similaritas tertinggi antara dokumen. Jika total similaritas adalah 0, maka dokumen tidak memiliki kemiripan, dan jika maksimal 1, maka dokumen memiliki kemiripan. Perhitungan cosine similarity dapat dilakukan secara manual.

$$sim(d_j, q) = \frac{\sum_{i=1}^N w_{i,j} w_{i,q}}{\sqrt{\sum_{i=1}^N w_{i,j}^2} \sqrt{\sum_{i=1}^N w_{i,q}^2}}$$

Tabel III.5 Proses Perhitungan dokumen dan query

Pembagi			
KATA	D1	D2	D3
ADA	0	0	0
BAJU	0	0	0
PENGANTIN	0.090619058	0	0
KOLEKSI	0.090619058	0	0
BARU	0.090619058	0	0
PASEPI	0	0	0
KITA	0	0	0
	0.271857175	0	0

Tahap ini adalah perhitungan frekuensi kata dalam koleksi dokumen yang cocok dengan query dan menentukan vektor berdasarkan term frekuensi. Tabel 3.6 menjelaskan proses pembagian nilai dokumen.

Tabel III.6 Proses Dokumen Dibagi

DIBAGI	
D1,Q	0.415269
D2,Q	0.830537
D3,Q	0.830537

Tabel III.7 Proses Hasil Perhitungan

CS,D1,Q	0.654654
CS,D2,Q	0
CS,D3,Q	0

Tabel 3.7 menjelaskan perhitungan cosine similarity setelah pembagian nilai dokumen. Hasilnya menunjukkan bahwa dokumen 1 memiliki ranking paling tinggi dan paling relevan dengan query.

IV. KESIMPULAN

Dalam penelitian ini, telah dibuat sebuah aplikasi penjualan perlengkapan pernikahan yang dilengkapi dengan fitur chatbot. Aplikasi ini menggunakan metode TF-IDF dan *cosine similarity* untuk menampilkan hasil dengan akurasi yang baik pada fitur chatbot. Selain itu, aplikasi ini juga telah diuji menggunakan metode blackbox testing dan fitur chatbot diuji menggunakan metode validasi dengan menghitung nilai *recall* dan *precision*. Hasil pengujian menunjukkan nilai akurasi tanya jawab chatbot mencapai 90% dengan menggunakan metode TF-IDF dan *cosine similarity*.

DAFTAR PUSTAKA

- Amalia, E. L., & Wibowo, D. W. (2019). Rancang Bangun Chatbot Untuk Meningkatkan Performa Bisnis. *Jurnal Ilmiah Teknologi Informasi Asia*, 13(2), 137–142.
- Herwijayanti, B., Ratnawati, D. E., & Muflikhah, L. (2018). Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity. *Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(1), 306–312. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/796>
- Mashud, M., & Wisda, W. (2019). Aplikasi Chatbot Berbasis Website sebagai Virtual Personal Assistant dalam Pemasaran Properti. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 9(2), 99. <https://doi.org/10.35585/inspir.v9i2.2497>
- Ngai, E. W. T., Lee, M. C. M., Luo, M., Chan, P. S. L., & Liang, T. (2021). Electronic Commerce Research and Applications An intelligent knowledge-based chatbot for customer service. *Electronic Commerce Research and Applications*, 50(April), 101098. <https://doi.org/10.1016/j.elerap.2021.101098>