

SISTEM INFORMASI WEDDING ORGANIZER FITUR CHATBOT DENGAN METODE *COSINE SIMILARITY*

Hurin In¹, Hardian Oktavianto², Daryanto³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
Hurin222@gmail.com¹, hardian@unmuhjember.ac.id², daryanto@unmuhjember.ac.id³

ABSTRAK

Penyelenggara pernikahan ini adalah salah satu layanan khusus yang membantu pengantin dan keluarga dalam perencanaan dan pengawasan pelaksanaan seluruh rangkaian acara pernikahan sesuai dengan jadwal yang telah ditentukan. Namun sebagian besar perusahaan pernikahan masih melakukan operasi pemasaran dan admin secara manual, seperti pemasaran melalui brosur, dan media sosial. Maka dapat disimpulkan bahwa layanan admin yang menggunakan operasi secara manual tidak cukup untuk memenuhi kualitas layanan. Maka kita membutuhkan solusi yang tepat yaitu memanfaatkan salah satu perkembangan teknologi terbaru seperti kecerdasan buatan seperti manusia yaitu Robot Chat (chatbot). Dengan fasilitas chatbot, admin tidak perlu membalas pertanyaan pelanggan setiap saat karena user merasa seakan-akan ada 2 pribadi manusia yang saling berkomunikasi. Nyatanya tidak, manusia berkomunikasi dengan *Bot*. *Bot* tersebut sudah di rancang untuk merespon segala jenis pertanyaan yang di inputkan oleh manusia (*user*). Oleh karena itu perlu metode pencarian string yang perlu dilakukan oleh Bot. Cosine Similarity merupakan salah satu metode pencarian string yang banyak saat ini. Metode *Cosine Similarity* tersebut dilakukan dengan mempertimbangkan frekuensi suatu kata dalam suatu dokumen (*term frequency*), dan penyebaran suatu kata pada sekumpulan dokumen (*Inverse Document Frequency*) serta kemiripan antar dokumen dengan metode *Cosine Similarity* dinyatakan dalam nilai bobot dari TF*IDF. Peneliti tertarik menerapkan Metode *Cosine Similarity* pada system informasi wedding organizer fitur *chatbot*. Hasil pengujian tersebut dari 30 data diuji 53 kali pertanyaan dengan model pertanyaan sederhana dengan 28 uji coba didapat hasil sesuai sebanyak 61% dan tidak sesuai 39% sedangkan model kompleks dengan 25 uji coba didapat hasil sesuai sebanyak 88% dan tidak sesuai 12%.

Kata Kunci: Wedding, Organizer, Chatbot, *Cosine Similarity*

1. Latar Belakang

Wedding organizer sendiri merupakan salah satu jasa khusus yang membantu calon pengantin dan keluarga dalam perencanaan dan supervisi pelaksanaan seluruh rangkaian acara pesta pernikahan sesuai dengan jadwal yang telah ditetapkan dari mulai akad nikah, upacara adat, pelaminan dan lain-lain. Namun kebanyakan perusahaan wedding masih melakukan pemasaran dan kegiatan operasional admin secara manual, seperti pemasaran masih melalui brosur, dan media sosial.

Chatbot adalah sebuah program komputer yang dirancang untuk mensimulasikan sebuah percakapan atau komunikasi yang interaktif kepada user melalui bentuk teks. Percakapan yang terjadi antara percakapan dengan manusia merupakan bentuk respon dari program yang telah dideklarasikan pada database program pada komputer. Respon yang dihasilkan merupakan hasil pemindaian dari kata kunci pada inputan user dan menghasilkan respon balasan yang di anggap paling cocok, atau pola kata-kata yang di anggap yang paling mendekati didalam database. *Chatbot* mampu meniru bahasa percakapan pada manusia. Jadi jika manusia sedang bercakap-cakap dengan program ini, maka user merasa seakan-akan ada 2 pribadi manusia yang saling berkomunikasi. Nyatanya tidak, manusia berkomunikasi dengan *Bot*. *Bot* tersebut sudah di rancang untuk merespon segala jenis pertanyaan yang di inputkan oleh manusia (*user*).

Oleh karena itu perlu pencarian kata yang perlu dilakukan oleh Bot dapat memanfaatkan metode *cosine similarity* yang membuat pengolahan data-d

data yang di inputkan menghasilkan output yang *valid*. Metode ini melakukan pencarian terhadap string yang mendekati dengan string lain yang terkumpul dalam sebuah kamus.

2. Kajian Pustaka

a. Preprocessing

Pre-processing dalam proses klarifikasi dokumen digunakan untuk membangun sebuah *index* dari koleksi dokumen *index* dari koleksi dokumen. *Index* adalah himpunan *term* yang menunjukkan isi atau topic yang dikandung oleh dokumen. Pembuatan *invertedindex* harus melibatkan konsep *linguistic processing* yang bertujuan meng-ekstrak term-term penting dari dokumen yang dipresentasikan sebagai *bag-ofword*. Tahap *Pre-processing* yang digunakan antara lain :

- *Tokenizing*
- *Filtering*
- *Stemming*

1. Tokenizing

Tokenizing adalah proses pemecahan dokumen atau paragraph atau kalimat menjadi daftar kata (*token*) yang berdiri sendiri. Fungsi *token* tersebut akan melakukan pengecekan terhadap jarak antar kata untuk membuat daftar kata pada semua kata yang terdapat pada dokumen.

2. *Stoplist/Wordlist/Filtering*

Stoplist/Wordlist/Filtering adalah daftar kata yang tidak relevan dalam mengidentifikasi isi suatu dokumen seperti kata sambung, kata depan, kata ganti, dan

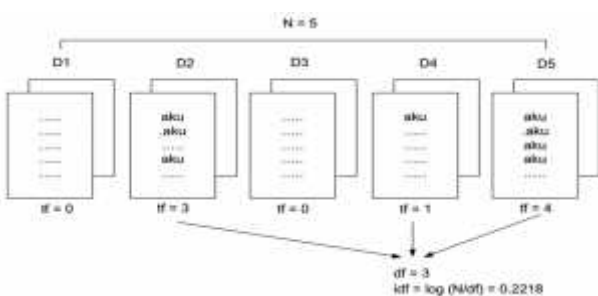
tanda baca. *Wordlist* adalah daftar kata yang relevan sebagai kata kunci sehingga akan digunakan dalam *Stemming*. *Stoplist/Wordlist* adalah proses pembuangan atau penyaringan atau *Stoplist* sehingga dihasilkan *Wordlist*.

3. *Stemming*

Kata-kata yang muncul didalam dokumen sering mempunyai banyak varian morfologik. Oleh karena itu setelah menjalani proses *tokenizing* dan *stopword removal*, kata-kata yang tersisa menjalani proses *stemming*. *Stemming* bertujuan untuk mengubah atau mengembalikan kata menjadi bentuk kata dasarnya (*root word*) dengan menghilangkan imbuhan. Pada penelitian ini menggunakan algoritma Nazief & Adriani yang disesuaikan atau dikembangkan dalam Bahasa Indonesia.

b. **TF-IDF**

Term Frequency (*tf*) merupakan frekuensi kemunculan suatu kata (*term*) dalam dokumen. Oleh sebab itu, *tf* memiliki nilai yang bervariasi dari satu dokumen ke dokumen lain tergantung dari tingkat kepentingan sebuah *term* dalam sebuah dokumen. Semakin sering suatu *term* muncul dalam suatu dokumen, *term* tersebut akan memiliki nilai *tf* yang lebih besar dari pada *term-term* lain yang jarang muncul. Penggunaan faktor *tf* belum mencukupi dalam menentukan pembobotan. Untuk itu, diperlukan faktor *Inverse Document Frequency* (*idf*) yang merupakan sebuah statistik “global” yang mengkarakteristikan sebuah *term* dalam keseluruhan koleksi dokumen. *idf* merupakan sebuah perhitungan dari bagaimana *term* didistribusikan secara luas pada koleksi dokumen yang bersangkutan. Semakin sedikit dokumen yang mengandung *term* yang dimaksud, maka nilai *idf* semakin besar. Jika setiap dokumen dalam koleksi mengandung *term* yang bersangkutan, maka nilai *idf* dari *term* tersebut adalah nol (0). Hal ini menunjukkan bahwa setiap *term* yang muncul pada dokumen dalam koleksi tidak berguna untuk membedakan dokumen berdasarkan topik tertentu. Ilustrasi algoritma *tf-idf* ditunjukkan pada gambar berikut :



Keterangan :

D1 – D5 : dokumen
 tf : banyaknya *term* yang dicari pada setiap dokumen

N : total dokumen
 Df : banyaknya dokumen yang mengandung *term* yang dicari

Dalam penelitian ini, algoritma pembobotan *Term Frequency* (*tf*) – *Inverse Document Frequency* (*idf*) ditetapkan pada tahap *similarity document*. Nilai *tf-idf* diperoleh dengan menggunakan persamaan :

$$W_{i,j} = tf_{i,j} \times idf_j = tf_{i,j} \times \log\left(\frac{N}{df_j}\right) \dots\dots\dots(1)$$

Keterangan :

$w_{i,j}$: bobot *term* ke-*j* terhadap dokumen ke-*i*
tf_{i,j} : jumlah kemunculan *term* *j* ke dalam dokumen *i*
 N : jumlah dokumen secara keseluruhan
df_j : Jumlah dokumen yang mengandung *term* *j*

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_j}\right) + 1 \dots\dots\dots(2)$$

Berdasarkan persamaan (1), berapapun besarnya nilai apabila $N = df$ maka akan didapat hasil nol (0) untuk perhitungan *idf*. Untuk itu, dapat ditambahkan nilai 1 pada sisi *idf*, sehingga perhitungan bobotnya menjadi :

Perhitungan bobot dari *term* tertentu dalam sebuah dokumen dengan menggunakan $tf \times idf$ menunjukkan bahwa deskripsi terbaik dari dokumen adalah *term* yang banyak muncul dalam dokumen tersebut dan sangat sedikit muncul pada dokumen lain (Sulistyo, 2015).

c. **Perhitungan Cosine Similarity**

✿ Kemiripan antar dokumen dihitung menggunakan suatu fungsi ukuran kemiripan yaitu *similarity measure*. Ukuran ini memungkinkan perbandingan dokumen sesuai dengan kemiripan relevansinya terhadap *query*. Salah satu ukuran kemiripan teks yang paling populer adalah *cosine similarity*. Ukuran ini menghitung nilai *cosinus* sudut antara dua vektor. Jika terdapat dua vektor dokumen *d_j* dan *query q*, serta *t term* diekstrak dari dari koleksi dokumen, maka nilai *cosinus* antara *d_j* dan *q* didefinisikan sebagai berikut (Muharromah, 2018).

$$\text{Cosine} \rightarrow \text{sim}(d_j, q) = \frac{\overline{d_j \cdot q}}{|d_j| |q|} = \frac{\sum_{i=1}^t (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^t w_{ij}^2 \cdot \sum_{i=1}^t w_{iq}^2}}$$

Keterangan :

d : dokumen
q : *query*
w : bobot *term* dalam sebuah dokumen

d. Threshold

Untuk memperoleh hasil pencarian dokumen yang maksimal diperlukan sebuah nilai ambang batas (*Threshold*) agar sistem dapat memilih mana dokumen yang mirip dan mana yang tidak mirip. Dokumen dengan nilai = *threshold* dapat dinyatakan mirip, sedangkan dokumen dengan nilai < *threshold* dinyatakan tidak mirip. Untuk mendapatkan nilai batas diperlukan suatu data training untuk melakukan uji coba (Muharromah, 2018).

e. Recall dan Precision

Sistem temu kembali informasi mengembalikan sekumpulan dokumen sebagai jawaban dari *query* pengguna. Terdapat dua kategori dokumen yang dihasilkan oleh sistem temu kembali informasi terkait pemrosesan *query*, yaitu *relevant documents* (dokumen yang relevan dengan *query*) dan *retrieved documents* (dokumen yang diterima pengguna). Ukuran umum yang digunakan untuk mengukur kualitas dari data *retrieval* adalah kombinasi *precision* dan *recall*.

Precision mengevaluasi kemampuan sistem temu kembali informasi untuk menemukan kembali data *top-ranked* yang paling relevan, dan didefinisikan sebagai persentase data yang dikembalikan yang benar-benar relevan terhadap *query* pengguna. *Precision* merupakan proporsi dari suatu set yang diperoleh yang relevan. (Agus, 2018) (Melita, 2018).

Rumus *confusion matrix*

		Nilai Sebenarnya (pakar)	
		Relevant	Not Relevant
Nilai Prediksi (sistem)	Retrieved	True Positive (TP)	False Positive (FP)
	Not Retrieved	False Negative (FN)	True Negative (TN)

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Keterangan:

TP : Jumlah data terambil oleh sistem dan di anggap relevan oleh pakar

FP : Jumlah data terambil oleh sistem dan di anggap tidak relevan oleh pakar

FN : Jumlah data tidak terambil oleh sistem dan di anggap relevan oleh pakar

TN : Jumlah data tidak terambil oleh sistem dan di anggap tidak relevan oleh pakar

3. METODE PENELITIAN

a. Objek dan Lokasi Penelitian

Objek penelitian dalam penyusunan skripsi ini adalah sistem informasi yang berhubungan dalam hal penjualan jasa pada CV.Luluk yang berada di Kraksaan Probolinggo, beralamat di jalan Ahmad Yani Gang Perkutut No.56. CV.Luluk merupakan sebuah perusahaan yang bergerak dibidang wedding organizer yang melayani klien sesuai permintaan.

b. Metodologi Penelitian

Penelitian ini bertujuan merancang dan membangun aplikasi layanan berbasis web dengan penerapan fitur *Chatbot* menggunakan metode *cosine similarity* dimana setiap langkah penelitian dilakukan secara berurutan, mulai dari tahapan analisis, preprocessing, pembobotan, perhitungan cosine, treshold dan pengujian.

c. Preprocessing Query dan Dokumen

Preprocessing atau normalisasi dokumen dan query akan dilakukan beberapa tahapan yaitu dengan menggunakan tokenizing, filtering dan stemming.

d. Pembobotan *TF-IDF*, *cosine similarity* dan *Threshold*.

Data yang telah dipreprocessing akan dikukan beberapa tahapan untuk mendapatkan hasil dan keputusan yaitu:

- Pembobotan *TF-IDF*
- Perhitungan *cosine similarity*
- Threshold

e. Pengujian

Pengujian disini menggunakan metode dari *confusion matrix*. Dengan mengukur tingkat presisi dan recall. Dari hasil ini akan dipertimbangkan atau ditarik kesimpulan.

4. HASIL DAN PEMBAHASAN

a. Deskripsi data

Tahap ini adalah tahap awal dalam pengolahan data untuk menemukan hasil dalam penelitian. Data yang diambil berjumlah 30 data yang berupa pertanyaan beserta jawabannya.

Data awal

no	pertanyaan	Jawaban
1	hallo apa kabar ?	selamat datang, kak ada yang bisa kami bantu
2	halo	selamat datang, ada yang bisa kami bantu?
3	harganya berapa ?	harga yang mana ya kak?
4	Bagaimana metode pembayaran yang digunakan ?	sistem pembayaran DP 70%, kemudian kurangnya di lunasi pas acara
5	Bagaimana ketentuan jika acara pernikahan lewat dari kesepakatan waktu?	akan ada tambahan biaya atau denda
6	apakah anda mempunyai vendor rekanan?	kami memiliki bnyak rekanan mulai vendor rekanan sampai fotografi
7	berapa harga paket komplit atau lengkap?	250.000.000
8	berapa harga paket sederhana?	145.000.000
...		
30	Apakah bisa membatalkan salah satu item yang sudah dipesan?	Tinggal di cancel aja di keranjang yang sudah di sediakan

b. Preprocessing

Pada tahap ini data awal akan di normalisasi atau dipreprocessing dengan tahapan Tokenizing, Filtering dan Stemming.

Preprocessing query

Query	Tokenizing	Filtering	Stemming
Bagaimana cara melakukan pembayaran pesanan	bagaimana cara melakukan pembayaran pesanan	cara melakukan pembayaran pesanan	cara laku bayar pesan

Preprocessing data pertanyaan

No	Pertanyaan	Tokenizing	Filtering	Stemming
1	P1			
	hallo apa kabar ?	hallo apa kabar	hallo kabar	hallo kabar
2	P2			
	hallo	hallo	hallo	hallo
3	P3			
	harganya	harganya	harganya	harga

	berapa ?	berapa		
4	P4			
	Bagaimana metode pembayaran yang digunakan ?	bagaimana metode pembayaran yang digunakan	metode pembayaran digunakan	metode bayar guna
5	P5			
	Bagaimana ketentuan jika acara pernikahan lewat dari kesepakatan waktu?	Bagaimana ketentuan jika acara pernikahan lewat dari kesepakatan waktu	ketentuan acara pernikahan lewat kesepakatan waktu	Tentu acara nikah lewat sepakat waktu
6	P6			
	apakah anda mempunyai vendor rekanan?	Apakah anda mempunyai vendor rekanan	mempunyai vendor rekanan	punya vendor rekanan
7	P7			
	berapa harga paket komplit atau lengkap?	Berapa harga paket komplit atau lengkap	harga paket komplit lengkap	harga paket komplit lengkap
8	P8			
	berapa harga paket sederhana?	Berapa harga paket sederhana	harga paket sederhana	harga paket sederhana
...				
30	P30			
	Apakah ada denda jika membatalkan semua pesanan ?	Apakah ada denda jika membatalkan semua pesanan	denda membatalkan pesanan	Denda batal pesan

c. Hasil pembobotan TF-IDF

no	dokumen	hasil
1	P1	0
2	P2	0
3	P3	0
4	P4	1,265487
5	P5	0
6	P6	0
7	P7	0
8	P8	0
9	P9	0
10	P10	0
11	P11	0
12	P12	0
13	P13	0
14	P14	0
15	P15	0
16	P16	1,265487
17	P17	0
18	P18	0
19	P19	0
20	P20	0
21	P21	0
22	P22	0
23	P23	0
24	P24	0
25	P25	0
26	P26	0
27	P27	0
28	P28	1
29	P29	1
30	P30	1

Pembobotan

Berikut langkah retrieval menggunakan data pada tabel perhitungan TF-IDF:

- Mengkalikan bobot antara bobot query dengan bobot term pada setiap dokumen.

$$Q * P = (1,602059991 * 0) + (1,602059991 * 0) \dots = 1,265487161$$

$$Q * P16 = (1,602059991 * 0) + (1,602059991 * 0) \dots = 1,265487161$$

$$Q * P28 = (1,602059991 * 0) + (1 * 1) \dots = 1$$

$$Q * P29 = (1,602059991 * 0) + (1 * 1) \dots = 1$$

$$Q * P30 = (1,602059991 * 0) + (1 * 1) \dots = 1$$

- Menghitung panjang query dengan mengakarkan jumlah kuadrat bobot query.

$$|Q| = \sqrt{1,602059991^2 + 1,602059991^2 + \dots} = 2,720051395$$

- Menghitung panjang dokumen dengan mengakarkan jumlah dari bobot dokumen yang dikuadratkan pada setiap dokumen.

$$|P4| =$$

$$\sqrt{(1,602059 + 1,602059^2 + \dots) * (0^2 + 0^2 + \dots)} = 5,86601346$$

$$|P16| =$$

$$\sqrt{(1,602059^2 + 1,602059^2 + \dots) * (0^2 + 0^2 + \dots)} = 4,678303536$$

$$|P28| =$$

$$\sqrt{(1,602059^2 + 1,602059^2 + \dots) * (0^2 + 1^2 + \dots)} = 6,70416624$$

$$|P29| =$$

$$\sqrt{(1,602059^2 + 1,602059^2 + \dots) * (0^2 + 1^2 + \dots)} = 6,465971969$$

$$|P30| =$$

$$\sqrt{(1,602059^2 + 1,602059^2 + \dots) * (0^2 + 1^2 + \dots)} = 6,237923739$$

d. Cosine Similarity

Menghitung cosine similarity dengan membagi bobot Q.Dn dengan hasil perkalian antara panjang query (|Q|) dan panjang dokumen (|Dn|).

$$(P4) \text{ Cos}(Q, P4) = \frac{1,2}{5,8}$$

$$= 0,21573205$$

$$(P16) \text{ Cos}(Q, P16) = \frac{1,2}{4,6}$$

$$= 0,27050129$$

$$(P28) \text{ Cos}(Q, P28) = \frac{1}{6,7}$$

$$= 0,149160979$$

$$(P29) \text{ Cos}(Q, P29) = \frac{1}{6,4}$$

$$= 0,15465579$$

$$(P30) \text{ Cos}(Q, P30) = \frac{1}{6,2}$$

$$= 0,16030975$$

Ambang batas (threshold) adalah 0,031678662 yang di dapat dari jumlah total cosine di masing-masing dokumen dibagi banyaknya dokumen yaitu 30 Dokumen (P1, P2, P3 sampai P30).

e. Pengujian

Model pengujian pada system ada dua yang pertama pengujian pertanyaan sederhana dan pengujian pertanyaan kompleks

Skenario Uji Pertanyaan Sederhana

NO	Pertanyaan	Jawaban	Keterangan
1.	Halo	Selamat datang, ada yang bisa kami bantu ?	Jawaban yang sesuai
2.	Hai	Selamat datang, ada yang bisa kami bantu ?	Jawaban yang sesuai
3.	Halo kk	Selamat datang, ada yang bisa kami bantu ?	Jawaban yang sesuai
4.	Hy ap kbr?	Selamat datang, ada yang bisa	Jawaban yang sesuai

		kami bantu ?	
5.	Bagaimana bayarnya?	Sistem pembayaran Dp 70%, kemudian kurangnya dilunasi pas hari H	Jawaban yang sesuai
..			
28	Make up akad	Silahkan cek di produk kak	Jawaban yang tidak sesuai

Scenario pertanyaan kompleks

No	Pertanyaan	Jawaban
1.	Apa anda menjalankan bisnis ini sendiri tanpa vendor rekanan?	Kami memiliki banyak vendor rekanan mulai dari catering sampai fotografi
2.	Untk boking sebaiknya brp bulan sebelum hari h?	Idealnya 6 bulan sebelum hari H. Namun untuk bulan-bulan yang padat atau tanggal cantik disarankan 1 tahun sebelum hari H.
3.	Pemesnaan brp bln sebelum hari h??	Idealnya 6 bulan sebelum hari H
4.	Pemesanan sebaiknya brp buln sbelum hari h??	Idealnya 6 bulan sebelum hari H
5.	Harga paket sederhana berapa ? bisa booking diluar kota?	Bisa kak, kita sering menggarap event-event diluar kota seperti Jombang, Lumajang, Bondowoso, Situbondo dll
...		
25	Membatalkan semua pesanan apa ada denda?	Ada, karena pesanan yang telah di pesan sudah direkap

lebih bervariasi sehingga dapat digunakan dengan banyak macam keluaran respon *chatbot*.

Daftar Pustaka

- Al Fatta, H. 2007. Analisi dan Perancangan Sistem Informasi. Penerbit: Andi.
- Jogiyanto, HM. 2005. Analisis dan Desain Sistem Informasi. Penerbit: Andi. Yogyakarta.
- Jogiyanto, HM. 2009. Sistem Informasi Manajemen. Penerbit: Andi. Yogyakarta.
- Kadir, A. 2003. Konsep dan Tuntunan Praktis Basis Data. Penerbit: Andi. Yogyakarta.
- Muharromah, Lailiyatul. 2018. *Penilaian Esai Otomatis Ujian Setengah Semester di SMK Asrama Pembina Masyarakat Jatimulyo Jember Menggunakan Metode Cosine Similarity*. Skripsi tidak diterbitkan Jember: Program Studi Teknik Informatika.
- Melita, Ria. 2018. Penerapan Metode *term frequency inverse document frequency (tf-idf)* dan *Cosine similarity* pada system temu kembali informasi untuk mengetahui *syarah hadits* berbasis web (studi kasus: *syarah umdatil ahkam*). Jurnal Teknik Informatika.
- Prahasta, E. 2005. Konsep – konsep Dasar Sistem Informasi Geografis. Penerbit Informatika: Bandung.
- Rudiyanto, N. 2005. Perancangan dan Implementasi Perangkat Lunak Natural Language Processing untuk pengembangan chatbot berbahasa Indonesia. Bandung: Universitas Komputer Indonesia.
- Sulistyo, M.E. 2015. Penilaian Ujian bertipe Essay Menggunakan Metode *Text Similarity*. Jurnal Telematika.
- Salim, Agus. 2018. Pencarian Link informasi pada Aplikasi *Ensiklopedia Buah dan Sayuran Lokal dengan Metode Cosine Similarity*. Skripsi tidak diterbitkan Jember: Program Studi Teknik Informatika.

f. Penutup

1. kesimpulan

Dari hasil uji coba yang telah dilakukan, memperoleh kesimpulan bahwa chatbot dengan metode cosine similarity merupakan penyelesaian proses pencarian jawaban berdasarkan kata kunci dari pertanyaan user. dari hasil bobot beberapa dokumen kita dapat memfilter untuk mengambil jawaban berdasarkan bobot yang paling tinggi. Dari 30 data diuji 53 kali pertanyaan dengan model pertanyaan sederhana dengan 28 uji coba didapat hasil sesuai sebanyak 61% dan tidak sesuai 39% sedangkan model komplek dengan 25 uji coba didapat hasil sesuai sebanyak 88% dan tidak sesuai 12%

2. SARAN

Disarankan penggunaan kalimat sebagai data gabungan kata kunci dan respon yang