



Chatbot Informasi Penerimaan Mahasiswa Baru Menggunakan Metode FastText dan LSTM

Fahmi Yusron F.¹, Agus Komarudin², Melina³

^{1,2,3}Teknik Informatika, Fakultas Sains & Informatika, Universitas Jenderal Achmad Yani, Cimahi, Indonesia

¹fahmi.yusron@student.unjani.ac.id, ²agus.komarudin@lecture.unjani.ac.id, ³melina@lecture.unjani.ac.id

Abstract

New Student Admission (PMB) is an important stage in the continuity of education in an educational institution. The Faculty of Science and Informatics (FSI) at Jenderal Achmad Yani University (UNJANI) provides information services about PMB to prospective students and parents/guardians of prospective students but is still inefficient, so it is necessary to improve PMB information services by using Chatbots as a solution that is able to serve questions effectively and consistent. This study aims to develop a PMB information Chatbot system for FSI using the FastText and Long Short-Term Memory (LSTM) methods. Several methods have been used in Chatbot development research, such as Term Frequency–Inverse Document Frequency (TF-IDF), Bag of Words (BoW), and Convolutional Neural Networks (CNN). However, these studies still have certain limitations, such as the inability to grasp the meaning of words and difficulties in handling certain inputs. In this study, the text classification model uses the FastText method as the stage for representing words in vector form, then combined with several pre-processing methods (Tokenization & Casefolding) and LSTM for the classification stage. Then put it into the Chatbot component according to the architecture that was made. In testing, the Black Box Testing method is used to ensure the functionality of the Chatbot system. The test results show that the Chatbot system is able to understand the topic of questions asked by users properly. The interaction between users and Chatbots also runs smoothly, resulting in appropriate and informative responses. The results of this study are expected to be an effective and consistent solution for providing information about PMB to prospective students and parents/guardians of prospective students at FSI.

Keywords: Chatbot, Information, fastText, LSTM, Natural Language Processing.

Abstrak

Penerimaan Mahasiswa Baru (PMB) merupakan tahap penting dalam keberlangsungan pendidikan di sebuah institusi pendidikan. Fakultas Sains dan Informatika (FSI) di Universitas Jenderal Achmad Yani (UNJANI) memberikan pelayanan informasi seputar PMB kepada calon mahasiswa dan orang tua/wali calon mahasiswa tetapi masih kurang efisien dalam pelaksanaannya sehingga diperlukan peningkatan pelayanan informasi PMB dengan menggunakan Chatbot sebagai solusi yang mampu melayani pertanyaan-pertanyaan secara efektif dan konsisten. Penelitian ini bertujuan untuk mengembangkan sebuah sistem Chatbot informasi PMB FSI UNJANI dengan menggunakan metode FastText dan *Long Short-Term Memory* (LSTM). Beberapa metode yang telah digunakan dalam penelitian pengembangan Chatbot, seperti *Term Frequency–Inverse Document Frequency* (TF-IDF), *Bag of Words* (BoW), dan *Convolutional Neural Networks* (CNN). Namun, penelitian-penelitian tersebut masih memiliki keterbatasan tertentu, seperti ketidakmampuan menangkap makna kata dan kesulitan dalam menangani *input* tertentu. Pada penelitian ini, model klasifikasi teks menggunakan metode fastText sebagai tahap merepresentasikan kata ke bentuk vektor, kemudian digabungkan dengan beberapa metode *pre-processing* (*Tokenization & Casefolding*) dan LSTM untuk tahap pengklasifikasian. Selanjutnya dimasukkan ke dalam komponen Chatbot sesuai dengan arsitektur yang dibuat. Dalam pengujian, metode *Black Box Testing* digunakan untuk memastikan fungsionalitas sistem Chatbot. Hasil pengujian menunjukkan bahwa sistem Chatbot mampu memahami topik pertanyaan yang diajukan oleh pengguna dengan baik. Interaksi antara pengguna dan Chatbot juga berjalan dengan lancar, menghasilkan respons yang tepat dan informatif. Hasil penelitian ini diharapkan dapat menjadi solusi efektif dan konsisten untuk memberikan informasi seputar PMB kepada calon mahasiswa dan orang tua/wali calon mahasiswa di FSI UNJANI.

Kata kunci: Chatbot, Informasi, fastText, LSTM, Natural Language Processing.



Lisensi

Lisensi Internasional Creative Commons Attribution-ShareAlike 4.0.

1. Pendahuluan

Fakultas Sains dan Informatika (FSI) di Universitas Jenderal Achmad Yani (UNJANI) awalnya didirikan sebagai Sekolah Tinggi Matematika dan Ilmu Pengetahuan Alam (STMIPA) pada tahun 1988. Pada 10 November 2017, namanya diubah menjadi Fakultas Sains dan Informatika (FSI) karena perkembangan ilmu pengetahuan dan teknologi. Saat ini, FSI memiliki 4 program studi, yaitu Kimia (akreditasi B), Informatika (akreditasi B), Sistem Informasi, dan Magister Kimia.

Penerimaan Mahasiswa Baru (PMB) di FSI UNJANI sangat penting, dan salah satu peran utama adalah memberikan informasi terkait PMB kepada calon mahasiswa. Bagian Tata Usaha (TU) sering kali harus menjawab pertanyaan yang sama berkali-kali, yang tidak efisien. Oleh karena itu, diperlukan peningkatan pelayanan informasi PMB dengan menggunakan Chatbot untuk memberikan jawaban yang konsisten dan terbaru kepada calon mahasiswa.

Chatbot adalah program komputer yang menanggapi pesan pengguna melalui teks atau suara dengan menggunakan *Natural Language Processing* (NLP) sebagai bagian dari Kecerdasan Buatan. Chatbot pertama kali diperkenalkan oleh Joseph Weizenbaum pada tahun 1966 dan mulai populer melalui tes Loebner untuk mengukur efektivitasnya. NLP adalah cabang kecerdasan buatan yang memungkinkan sistem untuk memahami teks atau perkataan seperti manusia. Aplikasi NLP mencakup perangkat pintar yang dapat mendengar dan merespons perintah manusia, mesin pencarian yang memberikan informasi, dan Chatbot yang merespons pengguna dengan cepat dan baik [1].

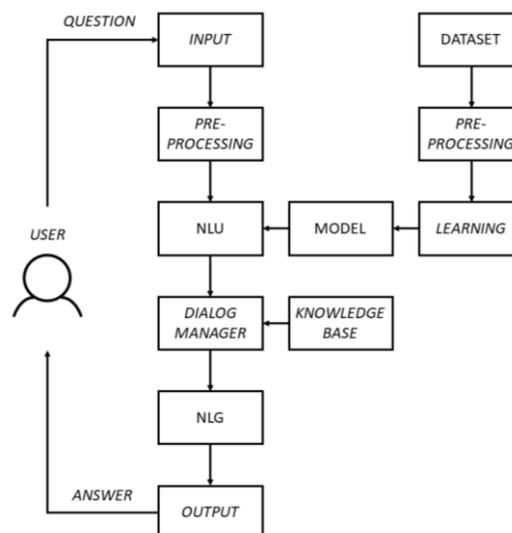
Pada penelitian terdahulu mengkaji penggunaan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dalam mengembangkan Chatbot untuk sistem pelayanan percetakan *online* dengan hasil penelitian menunjukkan sistem Chatbot ini dapat merespon pertanyaan pelanggan secara cepat tanpa batasan waktu, dengan tingkat *Recall* dan *Precision* sebesar 94% [2]. Pada penelitian sebelumnya mengkaji penggunaan metode *Bag of Words* (BoW) dan *Feed-forward Neural Networks* (FNN) untuk mengembangkan Chatbot yang digunakan di Rumah Sakit dengan hasil penelitian menunjukkan bahwa Chatbot memiliki akurasi 100% pada iterasi ke-1000 dan *loss* 0% pada iterasi ke-697 [3]. Pada penelitian sebelumnya mengkaji klasifikasi teks menggunakan metode TF-IDF dan *Word2Vec* untuk ekstraksi fitur, serta *Convolution Neural Networks* (CNN). Hasil penelitiannya mencapai akurasi optimal sebesar 82% [4].

Penelitian ini membuat Chatbot dengan pendekatan klasifikasi teks menggunakan metode *fastText* sebagai tahap merepresentasikan kata ke bentuk vektor, lalu digabungkan dengan beberapa metode *pre-processing* (*Tokenization & Casefolding*) dan LSTM untuk tahap

pengklasifikasian. Model yang dibuat dimasukkan ke dalam komponen Chatbot sesuai dengan arsitektur yang dibuat sebagai solusi terhadap permasalahan pelayanan proses PMB di Fakultas Sains dan Informatika UNJANI

2. Metode Penelitian

Penelitian ini membangun sistem berupa Chatbot Informasi Penerimaan Mahasiswa Baru (PMB) di Fakultas Informatika dengan menggunakan metode LSTM dan *fastText*. Sistem Chatbot ini dapat mengklasifikasikan berbagai jenis pertanyaan terkait PMB, seperti biaya kuliah dan tanggal pendaftaran, mekanisme pendaftaran, dll. Penelitian ini melibatkan pengumpulan dan *pre-processing* data percakapan atau pertanyaan yang sering diajukan oleh calon mahasiswa baru, yang kemudian dilabeli dan dibagi menjadi kategori yang relevan dengan PMB. Metode LSTM digunakan untuk memahami pola dan hubungan antara pertanyaan dan kategori yang sesuai, dengan keunggulan dalam memodelkan dependensi jangka panjang pada teks. Selain itu, metode *fastText* juga digunakan sebagai representasi kata ke bentuk vektor-vektor yang sehingga dapat meningkatkan akurasi klasifikasi. Penelitian ini juga mencakup komponen-komponen penting dalam arsitektur Chatbot, seperti *Natural Language Understanding* (NLU), *Dialog Manager*, *Natural Language Generation* (NLG), dan *Knowledge Base*. Alur metode penelitian dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian Chatbot

2.1 Pengumpulan Data

Data yang digunakan adalah data hasil kuesioner yang disebar kepada mahasiswa-mahasiswi Teknik Informatika UNJANI. Jumlah data yang terkumpul adalah 1.128 *records*. Atribut dari dataset ini yaitu berupa teks pertanyaan mengenai suatu topik seputar PMB. Adapun topik pertanyaan yang diberikan terdapat 11 macam yaitu biaya, silabus, akreditasi, pencapaian, jalur pendaftaran, tanggal pendaftaran,

kelas, beasiswa, prospek kerja, macam prodi dan fasilitas pada suatu jurusan. Topik dan contoh kalimat dapat dilihat pada Tabel 1.

Tabel 1. Topik pertanyaan & contoh kalimat

Pertanyaan	Topik
Apa saja prodi yang ada di fakultas ini?	Macam Prodi
Berapa uang yang perlu disiapkan untuk masuk informatika?	Biaya
Permisi, izin tanya kalau jurusan sistem informasi, apa saja topik pelajarannya?	Silabus
Apakah jurusan kimia terakreditasi a?	Akreditasi
Pendaftaran untuk jurusan informatika lewat jalur rapot bisa ga?	Jalur
Bagaimana alur pendaftaran penerimaan mahasiswa baru?	Pendaftaran
halo kak mau tanya apakah jurusan kimia sudah mempunyai prestasi seperti mengikuti lomba atau yang lainnya?	Pencapaian
Buat yang bekerja, bisa ga kuliahnya di hari libur?	Program Kelas
Apakah fakultas sains dan informatika mempunyai program beasiswa bagi yang tidak mampu?	Beasiswa
Mau nanya kalau di jurusan kimia, nanti belajar apa aja?	Silabus
Apa saja fasilitas di jurusan informatika?	Fasilitas

Dataset akan dibagi menjadi dua bagian, yakni data *training* dan data *testing*. Pembagian data ini akan dilakukan dengan proporsi sebanyak 80% untuk data *training* dan 20% untuk data *testing*.

2.2 Praproses

Pre-processing dalam konteks Chatbot Informasi Penerimaan Mahasiswa Baru (PMB) bertujuan untuk membersihkan dan mempersiapkan data teks sebelum dilakukan pelatihan model klasifikasi menggunakan metode LSTM dan fastText. Dalam kasus ini, dua tahap *pre-processing* yang dilakukan adalah *Casefolding* dan *Filtering*.

Jika terdapat suatu kalimat "Berapa biaya pendidikan di Informatika??", maka kalimat tersebut akan dilakukan proses pembersihan untuk setiap karakternya dijadikan huruf kecil. Pada Tabel 2 terdapat beberapa kata yang mempunyai karakter huruf besar. Lalu karakter huruf besar tersebut akan diubah menjadi huruf kecil sesuai kode ASCII [5].

Tabel 2. Proses *Casefolding*

Kalimat	Casefolding
"Berapa biaya pendidikan di Teknik Informatika?"	"berapa biaya pendidikan di teknik informatika?"

Berikut merupakan algoritma pada proses *casefolding*.

Algoritma Casefolding
<pre> Input: sentence Output: lower_case_sentence lower_case_sentence <- empty_string for each character in sentence: if character is an uppercase letter: lower_case_character <- convert character to lowercase else: lower_case_character <- character append lower_case_character to lower_case_sentence return lower_case_sentence </pre>

Kemudian setiap karakter akan disaring sehingga hanya terdapat karakter yang mengandung huruf kecil saja. Pada Tabel 3 terdapat contoh dimana karakter selain huruf kecil akan dihilangkan dari kalimat tersebut [6].

Tabel 3. Proses *Filtering*

Kalimat	Filtering
"Berapa biaya pendidikan di Teknik Informatika?"	"berapa biaya pendidikan di teknik informatika"

Berikut merupakan algoritma pada proses *filtering*.

Algoritma Filtering
<pre> Input: sentence Output: clean_sentence chars <- "abcdefghijklmnopqrstuvwxy" clean_sentence <- empty_string for each character in sentence: if character is an alphabet character (exists in chars): append character to clean_sentence return clean_sentence </pre>

Namun khusus pada saat pelatihan akan dilakukan eksperimen model tanpa menggunakan proses *casefolding* pada tahap praproses.

Selanjutnya teks akan diubah menjadi representasi vektor yang dapat dipahami oleh model. Metode fastText digunakan untuk membangun *vector embedding*. fastText menggunakan pendekatan yang memperhitungkan informasi subkata dalam kata. Pada penelitian ini menggunakan metode fastText *pre-trained* untuk mendapatkan vektor *embedding* yang telah dilatih sebelumnya. *Vector embedding* fastText *pre-trained* ini dimasukkan ke dalam variabel *embedding matrix*. *Embedding matrix* merupakan matrik yang berisi vektor-vektor *embedding* untuk setiap kata dalam dataset. Setiap baris dalam *embedding matrix* akan merepresentasikan vektor *embedding* dari sebuah kata. Dengan memasukkan vektor *embedding* fastText *pre-trained* ke dalam *embedding matrix*, sistem dapat menggunakan representasi ini sebagai fitur dalam model. Setelah itu perlu ada penyesuaian dengan dataset, *embedding matrix* akan menyesuaikan dengan dataset yang telah ada sebelumnya sehingga perlu membuat *embedding index*. *Embedding index* adalah kamus yang memetakan setiap kata dalam dataset kita ke indeks yang sesuai dalam *embedding matrix*. Kamus ini akan memungkinkan sistem untuk mengakses vektor *embedding* yang sesuai dengan kata yang ada dalam dataset saat melakukan pemrosesan atau pelatihan model [7].

Misalnya, jika terdapat kalimat yang sudah dilakukan *pre-processing* "berapa biaya pendidikan di informatika", maka setiap kata dari kalimat tersebut akan mempunyai vektor bobot yang ada pada *variable embedding matrix*. Contoh dimensi dapat dilihat pada Tabel 4 Bobot per kata.

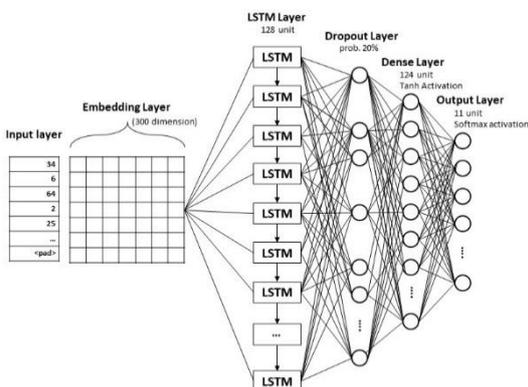
Tabel 4. Bobot per kata

Dimensi	Kata				
	Berapa	Biaya	Daftar	Di	Kimia
1	0.24	0.02	-0.21	0.02	-0.08
2	-0.70	0.24	0.02	-0.08	-0.70
3	0.06	-0.08	0.02	-0.47	-0.21
...
300	0.02	0.06	-0.24	-0.17	0.02

2.3 Learning

Model klasifikasi yang dibangun dengan menggunakan metode LSTM akan mampu memahami dan mengklasifikasikan beberapa pertanyaan berdasarkan kategori yang telah ditentukan sebelumnya, seperti biaya pendaftaran, jadwal pendaftaran, biaya kuliah, atau informasi pencapaian program studi [8].

Model klasifikasi LSTM akan mengambil *vector embedding* dari teks pertanyaan sebagai *input*. Kemudian, dengan bantuan layer LSTM dan *layer* tambahan seperti *layer Dropout* atau *layer Dense*, model akan memproses teks dan menghasilkan probabilitas klasifikasi untuk setiap kategori yang telah ditentukan. Kategori dengan probabilitas tertinggi akan dipilih sebagai kategori yang sesuai dengan pertanyaan tersebut. Arsitektur LSTM dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur LSTM

Proses klasifikasi ini memungkinkan Chatbot untuk memberikan respons yang tepat dan relevan terhadap pertanyaan yang diajukan oleh calon mahasiswa atau wali. Dengan memanfaatkan model LSTM dan fastText, Chatbot dapat memahami maksud pengguna dan mengarahkan mereka ke informasi yang sesuai dengan pertanyaan mereka. Proses klasifikasi merupakan inti dari kemampuan Chatbot dalam memberikan informasi yang akurat dan bermanfaat terkait PMB di Fakultas Sains dan Informatika.

2.4 Natural Language Understanding (NLU)

Natural Language Understanding (NLU) adalah cabang dari kecerdasan buatan yang berkaitan dengan kemampuan komputer untuk memahami dan menganalisis bahasa manusia secara alami. NLU melibatkan pemrosesan teks dan ucapan dengan tujuan untuk mengenali makna, struktur sintaksis, dan konteks

dari pesan yang disampaikan dalam bahasa manusia [9].

Metode fastText dan *Long Short-Term Memory* (LSTM) yang digunakan dalam penelitian ini digunakan sebagai alat atau teknik untuk mengimplementasikan komponen NLU ini. fastText digunakan untuk mengatasi masalah representasi kata dan memperhitungkan struktur morfologi dan informasi subkata. LSTM, di sisi lain, akan digunakan untuk memahami pola dan hubungan antara pertanyaan dan kategori yang sesuai, termasuk dalam konteks PMB

2.5 Dialog Manager

Dialog Manager adalah elemen utama yang bertanggung jawab untuk mengatur interaksi antara pengguna dan Chatbot. Fungsinya adalah mengelola percakapan, baik dalam hal memahami masukan pengguna maupun memberikan respons yang sesuai. *Dialog Manager* menggunakan pengetahuan yang ada dalam *Knowledge Base* untuk menghasilkan tanggapan yang relevan. Di samping itu, *Dialog Manager* juga bertugas untuk menciptakan representasi semantik atau pemahaman terhadap pesan yang diterima, dan menggunakan tindakan komunikatif untuk memberikan respons yang sesuai dengan kebutuhan percakapan [10].

Dialog Manager akan membantu Chatbot untuk memahami konteks percakapan, menjaga konsistensi dalam respon, dan menjalankan percakapan dengan pengguna dengan cara yang efektif. Sehingga memberikan peningkatan pengalaman pengguna dengan Chatbot, khususnya ketika calon mahasiswa baru mencari informasi tentang PMB.

2.6 Natural Language Generation (NLG)

Natural Language Generation (NLG) berfokus pada kemampuan Chatbot untuk menghasilkan teks yang terlihat seperti dibuat oleh manusia secara alami. NLG digunakan untuk mengubah data atau informasi yang tidak terstruktur menjadi teks yang dapat dimengerti oleh manusia [11].

NLG digunakan oleh *Dialog Manager* untuk menghasilkan teks atau kalimat yang kemudian akan disampaikan oleh Chatbot kepada pengguna. NLG berperan dalam mengubah instruksi dari *Dialog Manager* menjadi bahasa manusia yang dapat dimengerti oleh pengguna.

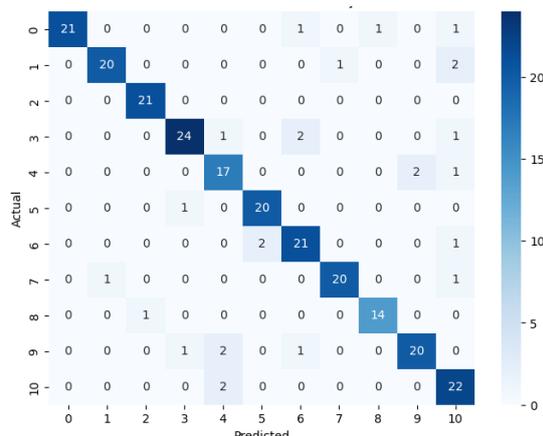
3. Hasil dan Pembahasan

3.1 Evaluasi Model

Penelitian ini membuat dua model yaitu model yang dibuat melalui proses *pre-processing* dengan algoritma *Casefolding* dan model yang dibuat tanpa melalui proses *pre-processing* dengan algoritma *Casefolding*.

Pengujian model pada penelitian ini dievaluasi menggunakan metode *Confusion Matrix*. Nilai-nilai

dalam tabel *confusion matrix* untuk model dengan *Casefolding* dapat dilihat pada Gambar 3.



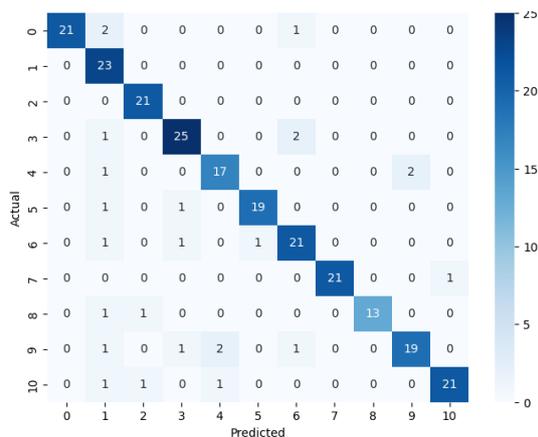
Gambar 3. *Confusion Matrix* Model dengan *Casefolding*

Dan beberapa metrik evaluasi model dapat dilihat pada Gambar 4.

Classification Report - Data Uji:				
	precision	recall	f1-score	support
0	1.00	0.88	0.93	24
1	0.95	0.87	0.91	23
2	0.95	1.00	0.98	21
3	0.92	0.86	0.89	28
4	0.77	0.85	0.81	20
5	0.91	0.95	0.93	21
6	0.84	0.88	0.86	24
7	0.95	0.91	0.93	22
8	0.93	0.93	0.93	15
9	0.91	0.83	0.87	24
10	0.76	0.92	0.83	24
accuracy			0.89	246
macro avg	0.90	0.90	0.90	246
weighted avg	0.90	0.89	0.90	246

Gambar 4. Metrik Evaluasi Model dengan *Casefolding*

Adapun nilai-nilai dalam tabel *confusion matrix* untuk model dengan tanpa *Casefolding* dapat dilihat pada Gambar 5.



Gambar 5. *Confusion Matrix* Model tanpa *Casefolding*

Dan beberapa metrik evaluasi model tanpa *Casefolding* dapat dilihat pada Gambar 6.

Classification Report - Data Uji:				
	precision	recall	f1-score	support
0	1.00	0.88	0.93	24
1	0.72	1.00	0.84	23
2	0.91	1.00	0.95	21
3	0.89	0.89	0.89	28
4	0.85	0.85	0.85	20
5	0.95	0.90	0.93	21
6	0.84	0.88	0.86	24
7	1.00	0.95	0.98	22
8	1.00	0.87	0.93	15
9	0.90	0.79	0.84	24
10	0.95	0.88	0.91	24
accuracy			0.90	246
macro avg	0.91	0.90	0.90	246
weighted avg	0.91	0.90	0.90	246

Gambar 6. *Confusion Matrix* Model tanpa *Casefolding*

Dari evaluasi dua mode di atas, dapat disimpulkan bahwa kualitas model Chatbot yang menggunakan *casefolding* menghasilkan nilai akurasi sebesar 89%, dan model Chatbot yang tidak menggunakan *casefolding* tidak berbeda jauh yaitu sebesar 90%.

3.2 Eksperimen *Threshold Confidence*

Threshold confidence merupakan ambang batas skor kepercayaan (*confidence score*) yang harus terpenuhi agar suatu prediksi dianggap valid atau dapat diterima. Pada penelitian ini, saat melakukan klasifikasi pada suatu pertanyaan, model akan membandingkan skor kepercayaan dengan nilai probabilitas label klasifikasi yang paling tinggi. *Threshold confidence* digunakan untuk membatasi dan mengontrol ketepatan klasifikasi yang diterima sebagai *output* dari model.

Terdapat 30 pertanyaan dimana 15 pertanyaan sesuai dengan topik dan 15 pertanyaan diluar topik PMB dengan tujuan untuk mengukur skor kepercayaan yang paling bagus untuk diterapkan nantinya ke dalam model. Adapun ambang batas yang akan diuji sebanyak 7, berikut nilai ambang batas serta eksperimen yang dilakukan.

Tabel 5. Pengujian *Threshold Confidence*

Kalimat ke-	Ambang Batas (Sesuai [S] / Tidak Sesuai [TS])						
	0,4	0,5	0,6	0,65	0,7	0,8	0,9
1	TS	TS	S	S	S	S	S
2	S	TS	TS	TS	TS	TS	TS
3	TS	S	S	S	S	S	S
4	S	S	S	S	S	S	S
5	TS	TS	TS	S	S	S	S
6	S	S	S	S	S	TS	TS
7	S	S	S	S	S	S	S
8	S	S	S	S	S	S	S
9	S	S	S	S	S	S	S
10	S	S	S	S	S	S	S
11	S	S	S	S	S	TS	TS
12	S	S	S	S	S	S	S
13	S	S	S	S	TS	TS	TS
14	TS	TS	TS	TS	TS	S	S
15	S	S	S	S	S	S	S
16	TS	S	S	S	S	S	S
17	S	S	S	S	S	S	S
18	S	S	S	S	S	S	S
19	TS	S	S	S	S	S	S
20	S	S	S	S	S	S	S

21	S	S	S	S	S	S	TS
22	S	S	S	S	S	S	TS
23	S	S	S	S	S	S	S
24	S	S	S	S	S	S	S
25	S	S	S	S	S	S	TS
26	S	S	S	S	S	S	S
27	S	S	S	S	S	S	S
28	S	S	S	S	S	S	S
29	S	S	S	S	S	S	S
30	S	S	S	S	S	S	S
Total Sesuai	24	26	27	28	27	25	23

Dari hasil yang didapatkan, ambang batas yang optimal berada pada 0,65 dengan jumlah sesuai sebanyak 28.

3.3 Pengujian Perangkat Lunak

Pada penelitian ini pengujian perangkat lunak menggunakan *Black Box Testing* yaitu metode pengujian perangkat lunak yang berfokus pada pengujian fungsionalitas perangkat lunak tanpa memperhatikan implementasi internal atau struktur kode di dalamnya.

Dalam menguji kemampuan Chatbot dalam memahami pertanyaan, serangkaian pertanyaan yang telah disusun diberikan kepada Chatbot. Pertanyaan-pertanyaan tersebut mencakup berbagai topik terkait penerimaan mahasiswa baru di Fakultas Sains dan Informatika UNJANI. Dalam pengujian ini, Chatbot akan dievaluasi berdasarkan kemampuannya dalam mengidentifikasi inti pertanyaan. Pengujian tanya jawab dapat dilihat pada Tabel 6 Pengujian tanya jawab

Tabel 6. Pengujian tanya jawab

No.	Pertanyaan	Jawaban	Hasil (Sesuai [S] / Tidak Sesuai [TS])
1	“Berapa biaya pendaftaran untuk mahasiswa baru di Fakultas Sains dan Informatika?”	Respons <i>Chatbot</i> berkaitan dengan topik biaya	S
2	“Apa saja mata kuliah yang termasuk dalam silabus jurusan Informatika?”	Respons <i>Chatbot</i> berkaitan dengan topik silabus	S
3	“Apakah jurusan Informatika di perguruan tinggi ini telah terakreditasi? Jika ya, oleh lembaga akreditasi mana dan berapa lama akreditasinya?”	Respons <i>Chatbot</i> berkaitan dengan topik akreditasi	S
4	“Apa pencapaian terkini yang telah diraih oleh jurusan Informatika dalam bidang penelitian dan pengembangan teknologi informasi?”	Respons <i>Chatbot</i> berkaitan dengan topik pencapaian	S
5	“Kapan tanggal pembukaan pendaftaran untuk jurusan Informatika di tahun ini?”	Respons <i>Chatbot</i> berkaitan dengan topik tanggal pendaftaran	S

6	“Apa persyaratan dokumen yang harus disertakan dalam proses pendaftaran jurusan Informatika?”	Respons <i>Chatbot</i> berkaitan dengan topik jalur pendaftaran	S
7	“Apakah ada kelas dengan fokus khusus, seperti kelas internasional atau kelas dengan kurikulum tambahan untuk pengembangan keahlian tertentu?”	Respons <i>Chatbot</i> berkaitan dengan topik program kelas	S
8	“Apakah jurusan Informatika menawarkan beasiswa kepada mahasiswa yang berprestasi? Jika ya, apa persyaratan dan kriteria untuk mendapatkan beasiswa tersebut?”	Respons <i>Chatbot</i> berkaitan dengan topik beasiswa	S
9	“Bagaimana tingkat permintaan tenaga kerja di bidang Informatika, baik di sektor industri maupun sektor pemerintahan?”	Respons <i>Chatbot</i> berkaitan dengan topik prospek kerja	S
10	“Bagaimana kondisi dan kelengkapan perangkat keras dan perangkat lunak di laboratorium Informatika?”	Respons <i>Chatbot</i> berkaitan dengan topik fasilitas	S

Dari keseluruhan 10 pertanyaan yang diajukan kepada Chatbot, respons yang diberikan sesuai dengan harapan yang telah ditetapkan. Dengan demikian, dapat disimpulkan bahwa dalam pengujian Tanya Jawab ini, Chatbot menunjukkan kemampuan yang sangat baik dalam memahami topik pertanyaan yang diajukan oleh pengguna.

4. Kesimpulan

Penelitian ini telah menghasilkan sebuah sistem Chatbot informasi penerimaan mahasiswa baru untuk Fakultas Sains dan Informatika Universitas Jenderal Achmad Yani menggunakan metode *fastText* untuk merepresentasi kata ke dalam bentuk vektor, dan metode LSTM untuk mengklasifikasikan teks berupa pertanyaan yang dimasukkan oleh pengguna Chatbot.

Dari hasil *confusion matrix*, diperoleh evaluasi performa model yang telah diuji dengan menunjukkan performa yang baik. Model yang dibuat tanpa menggunakan *casefolding* mendapatkan akurasi lebih besar yaitu 90% dibandingkan dengan model yang menggunakan *casefolding* yang mendapatkan akurasi sebesar 89%.

Hasil eksperimen *threshold confidence* menunjukkan bahwa ambang batas 0.65 memberikan performa terbaik bagi model. 30 pertanyaan yang diuji, model berhasil mengklasifikasikan dengan akurat sebanyak 28 pertanyaan, sehingga hanya 2 pertanyaan yang menghasilkan klasifikasi yang tidak sesuai.

Berdasarkan hasil pengujian menggunakan metode *Black Box* menunjukkan kemampuan Chatbot sangat baik dalam memahami topik pertanyaan yang diajukan oleh pengguna.

Daftar Rujukan

- [1] E. Adamopoulou and L. Moussiades, *An Overview of Chatbot Technology An Overview of Chatbot Technology*, no. June. Springer International Publishing, 2020. doi: 10.1007/978-3-030-49186-4.
- [2] D. W. Wibowo and J. D. Kristanto, "Penerapan Metode TF-IDF Untuk Chatbot Pada Sistem Informasi Pelayanan Percetakan Online," 2020.
- [3] M. Mittal *et al.*, "Web-based chatbot for Frequently Asked Queries (FAQ) in Hospitals," *J Taibah Univ Med Sci*, vol. 16, no. 5, pp. 740–746, 2021, doi: 10.1016/j.jtumed.2021.06.002.
- [4] P. Song, C. Geng, and Z. Li, "Research on Text Classification Based on Convolutional Neural Network," *2019 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pp. 229–232, 2019, doi: 10.1109/ICCNEA.2019.00052.
- [5] S. Data *et al.*, "Clustering Berita Menggunakan Algoritma Tf-Idf Dan K-Means Dengan Memanfaatkan Sumber Data Crawling Pada Situs Detik.Com," vol. 3, no. 1, 2022.
- [6] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis," *Expert Syst Appl*, vol. 110, pp. 298–310, 2018, doi: 10.1016/j.eswa.2018.06.022.
- [7] T. Yao, "Text Classification Model Based on fastText," pp. 154–157, 2020.
- [8] Y. Yu, X. Si, C. Hu, and J. Zhang, "A Review of Recurrent Neural Networks : LSTM Cells and Network Architectures," vol. 1270, pp. 1235–1270, 2019, doi: 10.1162/neco.
- [9] A. Ait-Mlouk and L. Jiang, "KBot: A Knowledge Graph Based ChatBot for Natural Language Understanding over Linked Data," *IEEE Access*, vol. 8, pp. 149220–149230, 2020, doi: 10.1109/ACCESS.2020.3016142.
- [10] K. Kumar, N. Nikhil, K. Singh, and V. Shivshanker, *A survey on providing customer and public administration based services using AI: chatbot*. Springer US, 2022. doi: 10.1007/s11042-021-11458-y.
- [11] A. Gatt, "Survey of the State of the Art in Natural Language Generation : Core tasks , applications and evaluation," vol. 65–170, 2018.