

**PENERAPAN METODE TF-IDF DAN *DEEP NEURAL NETWORK* (DNN)
PADA CHATBOT LAYANAN INFORMASI AKADEMIK
FASILKOM-TI USU**

SKRIPSI

ANGGITRI SIHOMBING

181402128



**PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024**

PENERAPAN METODE TF-IDF DAN *DEEP NEURAL NETWORK* (DNN)
PADA CHATBOT LAYANAN INFORMASI AKADEMIK
FASILKOM-TI USU

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah
Sarjana Teknologi Informasi

ANGGITRI SIHOMBING

181402128



PROGRAM STUDI S1 TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS SUMATERA UTARA
MEDAN
2024

PERSETUJUAN

PERSETUJUAN

Judul : PENERAPAN METODE TF-IDF DAN DEEP NEURAL NETWORK (DNN) PADA CHATBOT LAYANAN INFORMASI AKADEMIK FASILKOM-TI USU

Kategori : Skripsi

Nama Mahasiswa : Anggiti Sihombing

Nomor Induk Mahasiswa : 181402128

Program Studi : Sarjana (S-1) Teknologi Informasi

Fakultas : Ilmu Komputer dan Teknologi Informasi

Universitas Sumatera Utara

Medan, 21 Mei 2024

Komisi Pembimbing:

Pembimbing 2,



Wati Jaya S.Si., M.Kom.

NIP. 198407072015041001

Pembimbing 1,



Dr. Marischa Elveny, S.Ti., M.Kom

NIP. 199003272017062001

Diketahui/disetujui oleh

Program Studi S-1 Teknologi Informasi

Ketua,



Dedy Arisandi, S.Ti., M.Kom.

NIP. 197908317100121002

PERNYATAAN

**PENERAPAN METODE TF-IDF DAN DEEP NEURAL NETWORK (DNN)
PADA CHATBOT LAYANAN INFORMASI AKADEMIK
FASILKOM-TI USU**

SKRIPSI

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, Maret 2024



ANGGITRI SIHOMBING

181402128

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan atas kasih karunia yang dari Allah Bapa Yang Maha Esa. Sebab karena berkat-Nya penulis dapat mengerjakan dan menyelesaikan skripsi yang berjudul **“Penerapan Metode TF-IDF dan Deep Neural Network (DNN) pada Chatbot Layanan Informasi Akademik Fasilkom-TI USU”** sebagai salah satu persyaratan untuk mendapatkan gelar Sarjana Komputer pada Prodi S1 Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi USU.

Penulis juga mengucapkan sangat banyak terimakasih kepada orang tua penulis, Parsaoran Sihombing dan Norma Nainggolan yang dengan senantiasa mendoakan dan mendukung penulis serta dengan sabar menunggu penulis menyelesaikan skripsi ini. Begitu juga untuk kakak penulis, Santri May Cathrin Sihombing dan abang penulis, Reinhard Vaber Sihombing yang selalu mendoakan dan memberikan semangat kepada penulis.

Selama pengerjaan penelitian skripsi ini, penulis telah mendapatkan banyak sekali bimbingan, dukungan, dan bantuan dari berbagai pihak untuk penulis baik selama masa perkuliahan maupun selama pengerjaan skripsi ini. Oleh karena itu penulis juga mengucapkan banyak sekali terima kasih kepada:

1. Bapak Dr Muryanto Amin, S.Sos, M.Si., yang merupakan Rektor Universitas Sumatera Utara
2. Ibu Dr. Maya Silvi Lydia B.Sc., M.Sc. selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara
3. Bapak Dedy Arisandi S.T., M.Kom. selaku Ketua Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
4. Ibu Dr. Marischa Elveny, S.TI., M.Kom. selaku Dosen Pembimbing I penulis yang telah memberikan banyak bantuan, bimbingan, motivasi, arahan, serta meluangkan waktu bagi penulis dalam pengerjaan skripsi ini.
5. Bapak Ivan Jaya S.Si., M.Kom. selaku Dosen Pembimbing II penulis yang telah memberikan masukan, saran, dan bimbingan selama penyusunan skripsi.
6. Bapak Prof. Dr. Drs. Opim Salim Sitompul M.Sc. dan Ibu Annisa Fadhillah Pulungan S.Kom., M.Kom. selaku dosen Penguji penulis yang telah banyak memberikan masukan dan saran untuk perbaikan skripsi saya.

7. Seluruh Bapak/Ibu dosen dan staff pegawai Program Studi Teknologi Informasi yang telah membimbing dan membantu penulis selama masa perkuliahan.
8. Sahabat-sahabat di grup Tydacc Sasdindah Manurung, Kristina Nainggolan, Jessica Gultom, Riris Gultom, Uli, Janrian, Andrian, Darius Simamora dan Albert yang senantiasa saling membantu, mendoakan, memberi semangat selama perkuliahan dan pengerjaan skripsi.
9. Sahabat selama perkuliahan Patrisia Tambunan, Jessica Wong, Alya Febriani Lubis, Monika Laurensia, Naomi Hutauruk, Putri Handayani, Putri Natasya, Mita Amelia yang senantiasa saling memberi dukungan dan doa.

Medan, Maret 2024



Penulis

ABSTRAK

Salah satu peran dari Lembaga Pendidikan adalah memenuhi kebutuhan informasi bagi para peserta didik. Ketersediaan informasi yang *valid* yang mudah diakses menjadi sangat vital dalam dunia pendidikan khususnya dalam lingkup universitas. Penyampaian informasi pada dalam lingkup Fasilkom-TI USU masih dilakukan dengan cara sederhana melalui pesan *whatsapp* yang biasanya dibagikan oleh komting. Cara ini membuat mahasiswa rentan ketinggalan informasi dan membuat informasi-informasi yang ada tidak terdokumentasikan dengan baik. Selain itu keterbatasan waktu dan jarak menjadi kendala untuk menanyakan informasi kepada pihak kampus. Oleh karena itu dibutuhkan adanya *chatbot* sebagai *platform* yang dapat menyediakan informasi yang dibutuhkan dan menjawab pertanyaan mahasiswa dengan cara yang interaktif dan dapat diakses dengan mudah dimanapun dan kapanpun. Pada penelitian ini digunakan *Term Frequency – Inverse Document Frequency* (TF-IDF) dan *Deep Neural Network* (DNN). TF-IDF digunakan sebagai pembobot kata untuk mendapatkan kata-kata kunci yang bermakna dari dataset untuk kemudian hasilnya diproses menggunakan metode *Deep Neural Network* sehingga *chatbot* kemudian dapat memprediksi jawaban atas pertanyaan *user*. Akurasi yang dihasilkan pada penelitian ini adalah sebesar 88.67% menunjukkan sistem *chatbot* yang dibangun dengan penerapan TF-IDF dan DNN pada penelitian ini memiliki performa yang cukup baik.

Kata kunci: Informasi, Sistem Interaktif, *Chatbot*, *Term Frequency – Inverse Document Frequency*, *Deep Neural Network*

**THE IMPLEMENTATION OF TF-IDF AND DEEP NEURAL NETWORK ON
ACADEMIC INFORMATION CHATBOT IN
FASILKOM-TI USU**

ABSTRACT

One of the roles of educational institutions is to meet the information needs of students. The availability of valid information that is easily accessible is very vital in the field of education, particularly in the university environment. Dissemination of information in the Fasilkom-TI USU is still carried out in a simple way via WhatsApp messages which are usually shared by komting. This method makes students vulnerable to missing information and makes existing information not properly documented. Apart from that, time and distance limitations are obstacles to asking the campus for information. Thus, there is a requirement for a chatbot as a platform that can provide the information needed and can answer students' questions in an interactive way that can be accessed easily anywhere and at any time. In this research, Term Frequency - Inverse Document Frequency (TF-IDF) and Deep Neural Network (DNN) were used. TF-IDF is used as a word weighter to get meaningful key words from the dataset and then the results are processed using the Deep Neural Network method so that the chatbot can then predict the answer to the user's question. The accuracy produced in this research was 88.67%, indicating that the chatbot system built using TF-IDF and DNN in this research had quite good performance.

Keywords: *Information, Interactive System, Chatbot, Term Frequency – Inverse Document Frequency, Deep Neural Network*

DAFTAR ISI

PERSETUJUAN	i
PERNYATAAN.....	ii
UCAPAN TERIMA KASIH	iii
ABSTRAK	v
ABSTRACT	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	xi
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian.....	4
1.4. Batasan Penelitian	4
1.5. Manfaat Penelitian.....	4
1.6. Metodologi Penelitian	5
1.7. Sistematika Penulisan.....	6
BAB 2 LANDASAN TEORI	7
2.1. <i>Natural Language Processing</i>	7
2.2. <i>Chatbot</i>	7
2.3. <i>Text Preprocessing</i>	9
2.4. <i>Spelling Corrector oleh Peter Norvig</i>	10
2.5. <i>Term Frequency-Inverse Document Frequency</i>	10
2.6. <i>Artificial Neural Network</i>	11
2.7. <i>Deep Neural Network</i>	11
2.8. Fungsi Aktivasi.....	14
2.8.1. Fungsi Aktivasi ReLU	14
2.8.2. Fungsi Aktivasi <i>Softmax</i>	15
2.9. <i>Optimizer</i>	15

2.10. Dropout	16
2.11. Penelitian Terdahulu.....	16
2.11.1. Perbedaan Penelitian	19
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	21
3.1. Data	21
3.2. Arsitektur Umum.....	25
3.2.1. Data Input.....	26
3.2.2. Preprocessing	26
3.2.3. Term Frequency – Image Document Frequency	32
3.2.4. Implementasi Deep Neural Network	34
3.2.5. Training Model DNN.....	35
3.2.6. Testing Model DNN	37
3.2.6. Output	37
3.2.6. Integrasi dengan Telegram.....	37
3.3. Pemodelan Sistem	37
3.3.1. Diagram Aktivitas	37
3.3.2. Flowchart Sistem.....	39
3.4. Perancangan Tampilan Sistem	40
3.4. Metode Evaluasi	41
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM	42
4.1. Kebutuhan Sistem.....	42
4.1.1. Perangkat keras	42
4.1.2. Perangkat lunak.....	42
4.2. Implementasi Sistem	43
4.2.1 Implementasi Tampilan Antarmuka Sistem	43
4.2.2. Pelatihan Model Deep Neural Network	44
4.3. Pengujian Chatbot	48
4.4. Evaluasi Chatbot.....	65
BAB 5 KESIMPULAN DAN SARAN	70
5.1. Kesimpulan.....	70
5.2. Saran	70
DAFTAR PUSTAKA	72

DAFTAR GAMBAR

Gambar 2.1 Model <i>Generative Chatbot</i> (Soufyane <i>et al</i> , 2021)	8
Gambar 2.2 Model <i>Retrieval-based Chatbot</i> (Soufyane <i>et al</i> , 2021).....	9
Gambar 2.3 Gambaran bentuk <i>Deep Neural Network</i> (Putra, 2020)	12
Gambar 2.4 Penurunan pada Persamaan <i>Deep Neural Network</i> (Putra, 2020).....	13
Gambar 2.5 Fungsi Aktivasi ReLU	15
Gambar 2.6 <i>Dropout Regularization</i> (Geron, 2022)	16
Gambar 3.1 Kaprodi dan Sekprodi Fasilkom-TI dari website Fasilkom-TI USU	22
Gambar 3.2 <i>Google Form</i> untuk Mendapatkan Variasi Pertanyaan	22
Gambar 3.3 <i>Pseudocode</i> data JSON.....	24
Gambar 3.4 Arsitektur Umum.....	25
Gambar 3.5 Daftar Kamus Kata	31
Gambar 3.6 <i>Index</i> untuk Setiap Token Katas	32
Gambar 3.7 <i>Snippet</i> Dataframe Hasil Perhitungan Bobot IDF	33
Gambar 3.8 <i>Snippet</i> Dataframe Hasil Perhitungan TF-IDF	33
Gambar 3.9 Hasil Normalisasi Bobot TF-IDF dengan <i>Euclidean Norm</i>	34
Gambar 3.10 Diagram Aktivitas <i>Chatbot</i>	38
Gambar 3.11 <i>Flowchart</i> Sistem <i>Chatbot</i>	39
Gambar 3.12 Rancangan Tampilan Antarmuka Sistem	41
Gambar 4.1 Halaman Profil <i>Chatbot</i> Fasilkom-TI USU.....	43
Gambar 4.2 Halaman Percakapan <i>Chatbot</i> dan <i>User</i>	44
Gambar 4.3 Matriks Vektor TF-IDF	45
Gambar 4.4 Perbandingan nilai akurasi beberapa nilai <i>batch size</i> dan <i>dropout</i>	46
Gambar 4.5 Gambar <i>Epoch Training</i>	47
Gambar 4.6 Grafik <i>Accuracy</i> dan <i>Loss</i>	47
Gambar 4.7 Grafik Ringkasan Penilaian <i>User</i> dari google form	66
Gambar 4.8 Grafik Ringkasan Penilaian <i>User</i> dari google form lanjutan	67

DAFTAR TABEL

Tabel 2.1 Proses Pelatihan <i>Deep Neural Network</i> dengan <i>Backpropagation</i>	14
Tabel 2.2 Penelitian Terdahulu	18
Tabel 2.2 Penelitian Terdahulu (lanjutan).....	19
Tabel 3.1 <i>Spreadsheet</i> Variasi Kalimat Pertanyaan	23
Tabel 3.2 Hasil <i>Case Folding</i>	27
Tabel 3.3 Hasil <i>Stopwords Removal</i>	29
Tabel 3.4 Hasil <i>Tokenizing</i>	29
Tabel 3.5 Hasil <i>Stemming</i>	30
Tabel 3.6 Hasil <i>Spell Checking</i>	31
Tabel 3.7 Tabel Hasil Perhitungan Bobot IDF.....	33
Tabel 3.8 Parameter Model DNN	36
Tabel 4.1 Spesifikasi perangkat keras	42
Tabel 4.2 Spesifikasi perangkat lunak.....	42
Tabel 4.3 Hasil pengujian model DNN	48
Tabel 4.4 Koreksi <i>Chatbot</i>	63
Tabel 4.5 Perbandingan <i>Chatbot</i> yang Menerapkan Peter Norvig dan Tidak.....	68

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Informasi terkait bidang akademik merupakan hal esensial yang sangat dibutuhkan oleh mahasiswa. Salah satu faktor keberhasilan Lembaga Pendidikan adalah apabila mampu memenuhi kebutuhan para pelaku pendidikan dalam hal informasi (Siburian, 2019). Informasi-informasi pada tingkat perguruan tinggi dapat meliputi lingkup universitas ataupun fakultas.

Universitas Sumatera (USU) adalah sebuah universitas negeri di Indonesia yang terletak di Medan, Sumatera Utara. USU memiliki sebanyak 15 fakultas dimana salah satunya adalah Fakultas Ilmu Komputer dan Teknologi Informasi. Mahasiswa Fakultas Ilmu Komputer dan Teknologi Informasi USU atau sering disingkat dengan Fasilkom-TI USU sangat memerlukan informasi akademik. Informasi-informasi tersebut disediakan oleh pihak fakultas. Beberapa bidang informasi akademik pada lingkup fakultas yang paling penting adalah informasi terkait layanan pembelajaran dan penyusunan tugas akhir.

Saat ini pengaksesan dan pencarian informasi di Fasilkom-TI USU diperkirakan dilakukan dengan menanyakan langsung ke bagian kemahasiswaan atau akademik yang berada di Gedung A Fasilkom-TI USU dan juga menanyakan kepada dosen pengajar dan staf akademik. Selain itu mahasiswa juga mendapatkan informasi melalui grup *WhatsApp* stambuk yang diteruskan oleh komandan tingkat (komting) dari pihak fakultas biasanya berupa surat edaran atau dalam bentuk pesan. Mahasiswa juga dapat mengakses informasi melalui website Fasilkom-TI USU <https://fasilkom-ti.usu.ac.id> dan website <https://satu.usu.ac.id/>. Namun, penyampaian informasi dengan cara-cara tersebut membuat informasi-informasi yang ada tidak terdokumentasikan dengan baik sehingga akan sulit untuk dicari kembali. Beberapa informasi juga tidak serta merta disebarkan secara langsung sehingga mahasiswa perlu untuk bertanya langsung ke bagian kemahasiswaan atau akademik.

Belajar dari masa pandemi yang telah kita alami dimana akses menuju kampus pada saat itu sangat terbatas menyebabkan mahasiswa kesulitan untuk memperoleh informasi yang dibutuhkan. Hal tersebut membuat penulis menyadari bahwa pengaksesan informasi yang dapat dilakukan dengan mudah tanpa terkendala jarak adalah hal yang esensial bagi mahasiswa Fasilkom-TI USU. Oleh karena itu, Fakultas Ilmu Komputer dan Teknologi Informasi USU membutuhkan suatu platform interaktif yang menyediakan informasi akademik dan dapat diakses dengan mudah oleh mahasiswa Fasilkom-TI USU, salah satu platform yang cocok untuk hal tersebut adalah *chatbot*.

Chatbot adalah suatu sistem yang bekerja dengan tujuan menyimulasikan percakapan layaknya manusia melalui perintah suara dan teks (Soufyane, 2021). Sejak dimulainya perkembangan AI dan *Computation Linguistic*, perancangan *chatbot* yang menyerupai manusia semakin berkembang diberbagai bidang industri salah satunya pada bidang akademik. Beberapa institusi pendidikan telah menggunakan *chatbot* untuk memberikan layanan bagi civitas akademika, namun *chatbot* yang ada masih cenderung kaku dan monoton dalam merespon pertanyaan pengguna sehingga kurang memberikan kesan *friendly* dan natural. Oleh karena itu, pada penelitian ini dibangun suatu sistem *chatbot* yang membantu pihak Fakultas Ilmu Komputer dan Teknologi Informasi dalam menjawab pertanyaan mahasiswa terkait informasi akademik dengan lebih dinamis dan natural sehingga lebih nyaman untuk digunakan.

Penelitian sebelumnya terkait perancangan *chatbot* sebagai layanan informasi akademik telah dilakukan sebelumnya dengan menggunakan metode *Artificial Intelligence Markup Language* (AIML) and *Latent Semantic Analysis* (LSA) oleh Ranoliya *et al.* (2017). Penelitian ini merancang suatu *chatbot* yang digunakan universitas untuk menjawab pertanyaan-pertanyaan mahasiswa. AIML digunakan untuk merespon pertanyaan-pertanyaan yang umum dan *template based* seperti salam dan sapaan. LSA digunakan untuk menjawab pertanyaan yang lebih kompleks dan *service based* dari user dimana LSA akan membandingkan kemiripan semantik antarkata yang direpresentasikan dalam bentuk vektor. Kandpal *et al.* (2022) dalam penelitiannya membangun *chatbot* sebagai layanan asisten kesehatan dengan pendekatan konsep *Natural Language Processing* (NLP) dengan *Deep Learning* menggunakan metode LSTM untuk mengklasifikasi *user query* dan *generate* respon dengan teknik *sequence to sequence*.

Penelitian lainnya dilakukan oleh Singh *et al.* (2018) “*Chatbot using TensorFlow for small Businesses*”. Penelitian ini bertujuan untuk membangun aplikasi layanan informasi untuk sebuah bisnis kecil sebagai pengganti layanan *customer support* dengan menerapkan model *Artificial Neural Network* (ANN) dan metode *Bag of Words* (BOW) sebagai fitur ekstraksi sebelum data di-input ke *training model*. Penelitian berhasil membangun *chatbot* dengan akurasi 68,51%. Penelitian oleh Kavitha *et al* (2019) berjudul “*Chatbot for healthcare system using Artificial Intelligence*” mengusulkan perancangan *chatbot* layanan informasi kesehatan publik menggunakan algoritma *N-grams* untuk pemberian *tag*, TF-IDF untuk melakukan pembobotan vektor dan *cosine similarity* untuk melakukan pengecekan kemiripan antara dua kalimat.

Sulthan *et al* (2021) mengusulkan penelitian berjudul “*Analisis Sentimen pada Tweet Bencana Alam Menggunakan Deep Neural Network dan Information Gain*” bertujuan untuk melakukan analisis sentimen dengan melakukan pembobotan kata oleh TF-IDF dan penyeleksian fitur oleh *Information Gain* yang selanjutnya hasilnya akan diklasifikasi dengan metode *Deep Neural Network* (DNN). Penelitian ini juga membandingkan hasil yang diperoleh dari metode yang diusulkan yaitu DNN + IG dengan beberapa metode lain yaitu *Naïve Bayes* (NB), *Random Forest* (RF), *Support Vector Machine* (SVM), *Decision Tree* (DT). Pada penelitian diperoleh hasil penggunaan metode DNN + IG dan TF-IDF sebagai seleksi fitur jika dibandingkan dengan metode-metode lain memberikan hasil akurasi terbaik yaitu sebesar 83.15%

Berdasarkan latar belakang yang telah dipaparkan, penerapan metode TF-IDF dan *Deep Neural Network* (DNN) dapat bekerja dengan baik pada sistem *chatbot* ataupun masalah klasifikasi. Sehingga pada penelitian ini diputuskan untuk menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan model *Deep Neural Network* (DNN) dengan pendekatan *Natural Language Processing* (NLP) untuk membangun sistem *chatbot* yang digunakan untuk menjawab pertanyaan-pertanyaan mahasiswa Fasilkom-TI USU seputar informasi akademik yang meliputi sub layanan pembelajaran dan tugas akhir. Oleh karena itu, penulis melakukan penelitian yang berjudul “PENERAPAN METODE TF-IDF DAN DEEP NEURAL NETWORK (DNN) PADA *CHATBOT* LAYANAN INFORMASI AKADEMIK FASILKOM-TI USU”.

1.2. Rumusan Masalah

Pengaksesan dan penyebaran informasi terkait akademik pada Fasilkom-TI USU masih dilakukan dengan satu arah melalui surat edaran atau pesan whatsapp yang diteruskan oleh komting dari pihak fakultas sedangkan untuk mendapatkan informasi tertentu yang spesifik, mahasiswa perlu bertanya langsung ke bagian akademik dan kemahasiswaan sehingga tidak efisien dan informasi tidak terdokumentasi dengan baik. Oleh karena itu, dibutuhkan suatu sistem interaktif yang dapat digunakan mahasiswa Fasilkom-TI USU untuk mengakses informasi akademik dengan mudah dari mana saja dan kapan saja.

1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan *chatbot* layanan informasi akademik di Fasilkom-TI USU yang dapat digunakan mahasiswa Fasilkom-TI untuk mengakses informasi akademik dengan mudah dengan menerapkan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Deep Neural Network* (DNN).

1.4. Batasan Penelitian

Batasan-batasan masalah untuk mencegah perluasan ruang lingkup yang tidak diperlukan pada penelitian ini yaitu:

1. *Chatbot* hanya dapat berinteraksi dengan *user* menggunakan Bahasa Indonesia.
2. *Chatbot* hanya dapat berinteraksi dengan *user* melalui perintah teks.
3. *Chatbot* tidak menanggapi bahasa alay dan bahasa yang bercampur dengan angka .
4. Topik pembicaraan dibatasi hanya terkait informasi akademik pada Fasilkom-TI USU dengan fokus pada sub bagian layanan pembelajaran dan penyelesaian tugas akhir pada program sarjana.
5. Data yang digunakan sebagai sumber pengetahuan pada *chatbot* ini didapatkan secara manual dari website Fasilkom-TI USU, website Satu USU dan surat edaran.

1.5. Manfaat Penelitian

Pada penelitian ini terdapat beberapa manfaat yang hendak dicapai berdasarkan hasil penelitian yang diperoleh, yaitu:

1. Memudahkan pengaksesan informasi akademik pada Fasilkom-TI USU
2. Membantu mengembangkan platform yang menyediakan informasi akademik yang lebih lengkap dan sesuai dengan kebutuhan mahasiswa Fasilkom-TI USU

3. Membantu pengembangan aplikasi *chatbot* yang lebih interaktif dan dinamis dengan pengimplementasian NLP dengan TF-IDF dan DNN
4. Menjadi referensi untuk pengembangan *chatbot* Bahasa Indonesia yang berbasis *Natural Language Processing* dengan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Deep Neural Network* (DNN)

1.6. Metodologi Penelitian

Tahapan-tahapan yang dilakukan dalam proses pengerjaan penelitian ini yaitu:

1. Studi Literatur

Studi literatur dilakukan dengan melakukan pengumpulan data dan informasi dari berbagai referensi seperti jurnal, buku, website dan sumber informasi lain. Beberapa informasi terkait dengan penelitian ini yaitu mengenai *Chatbot*, *Natural Language Processing*, *Preprocessing*, *Spell Correction*, TF-IDF, ANN dan DNN.

2. Analisis Permasalahan

Selanjutnya informasi-informasi yang diperoleh hasil studi literatur dianalisis guna mendapatkan solusi penyelesaian permasalahan pada penelitian ini yaitu pengembangan *chatbot* informasi akademik Fasilkom-TI USU menggunakan metode TF-IDF dan *Deep Neural Network* (DNN).

3. Perancangan Sistem

Kemudian pada tahap ini dilakukan perancangan arsitektur umum, perancangan antarmuka *chatbot*, pengumpulan data, serta penentuan data latih dan data validasi.

4. Implementasi

Pada tahapan ini dilakukan pembangunan sistem informasi akademik Fasilkom-TI USU yang telah dirancang sebelumnya dengan mengimplementasikan NLP menggunakan TF-IDF dan *Deep Neural Network* (DNN).

5. Pengujian Sistem

Pada tahap ini dilakukan pengujian terhadap metode TF-IDF dan *Deep Neural Network* (DNN) yang telah diimplementasikan untuk mengetahui performa sistem dalam mengatasi permasalahan pada penelitian ini serta memastikan kinerja program

6. Penyusunan Laporan

Dalam tahap ini, penulis menyusun dokumentasi dan laporan yang berisi seluruh proses yang dilakukan dalam penelitian ini hingga diperoleh hasil akhir.

1.7. Sistematika Penulisan

Bab 1 Pendahuluan

Bab pertama pada penelitian ini menjabarkan latar belakang, rumusan masalah, batasan masalah, manfaat, dan tujuan dari penelitian. Selain itu juga terdapat metodologi penelitian dan sistematika penulisan pada penelitian.

Bab 2 Landasan Teori

Bab kedua merupakan landasan teori yang menjabarkan hal dasar terkait topik yang dibahas pada penelitian. Topik yang dibahas mencakup pemaparan dari *natural language processing*, *text preprocessing*, *chatbot*, metode yang digunakan, dan beberapa penelitian terkait.

Bab 3 Analisis dan Perancangan Sistem

Perancangan sistem dari penelitian yang diusulkan akan dijabarkan dan dijelaskan pada bab ini. Adapun beberapa hal yang akan dijabarkan adalah data yang akan digunakan, perancangan arsitektur umum dimulai dari data, *preprocessing*, pembangunan model, dan perancangan antarmuka sistem.

Bab 4 Implementasi dan Pengujian Sistem

Bab keempat menjabarkan hal terkait penerapan dari model dan sistem yang telah dirancang dan dibangun. Hasil dari pelatihan dan pengujian terhadap model dan sistem yang telah dibangun pada penelitian juga menjadi hal yang dipaparkan pada bab ini.

Bab 5 Kesimpulan dan Saran

Bab kelima berisi kesimpulan dan saran untuk pengembangan penelitian berikutnya yang diperoleh berdasarkan hasil dari penelitian yang telah dilakukan.

BAB 2

LANDASAN TEORI

2.1. Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah teknologi *modelling* yang memungkinkan komputer untuk memahami dan memanipulasi bahasa natural manusia (Tarwani *et al*, 2017). NLP berhubungan erat dengan bagaimana manusia dan komputer dapat berinteraksi menggunakan bahasa natural, yaitu dengan mempelajari struktur data dari pemrosesan data *natural language* dalam jumlah besar untuk membangun struktur pengetahuan. Struktur pengetahuan tersebut kemudian dapat digunakan untuk melakukan berbagai tugas seperti *text summarization* (meringkas teks), penerjemah bahasa, pengenalan *entity*, analisis semantik, analisis sentimen, *speech recognition* (identifikasi suara), dan *chatbot* (sistem yang mengotomasi percakapan tanya jawab pengguna dan mesin)

NLP bertugas untuk membuat agar manusia dapat berinteraksi dengan mesin menggunakan bahasa natural. Untuk dapat melaksanakan tugas tersebut, NLP dibangun dari 2 komponen utama yaitu NLU dan NLG. NLU atau *Natural Language Understanding* menangani permasalahan pemahaman bahasa manusia oleh mesin dari tata bahasa dan konteksnya dengan menganalisis konsep, entitas, emosi dan kata kunci (Khurana *et al*, 2022). NLU melibatkan pengkonversian bahasa natural ke dalam format yang dapat dibaca mesin. NLG atau *Natural Language Generation* merupakan proses pembentukan kalimat, frasa dan paragraf yang natural dan bermakna (Khurana *et al*, 2022).

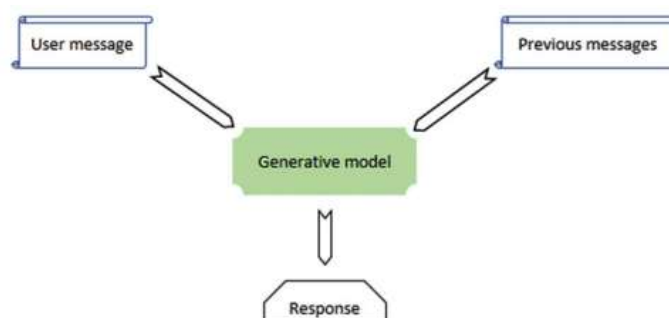
2.2. Chatbot

Chatbot adalah suatu sistem yang bekerja dengan tujuan menyimulasikan percakapan layaknya manusia melalui perintah suara dan teks (Soufyane, 2021). *Chatbot* memungkinkan manusia dan mesin untuk dapat saling berinteraksi dengan bahasa natural yaitu bahasa yang digunakan manusia berkomunikasi sehari-hari seperti Bahasa

Indonesia, Inggris, Mandarin, dan sebagainya. Prinsip kerja *chatbot* adalah dengan cara memindai kata kunci dari masukan yang diberikan pengguna, biasanya berupa pertanyaan, lalu kemudian memberikan respon berupa jawaban yang dipilih berdasarkan kata kunci yang paling cocok pada pola data di struktur pengetahuan (*knowledge base*).

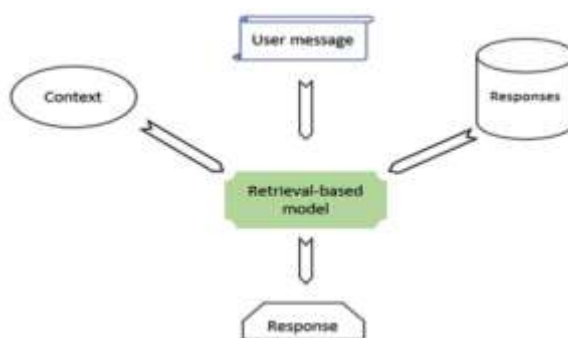
Salah satu aplikasi *chatbot* yang paling awal dikembangkan dan paling terkenal adalah ELIZA (Weizenbaum, 1966) yang dikembangkan untuk menjadi seperti seorang psikoterapis dimana chatbot ini akan merespon dengan melempar pertanyaan atas *input* yang diberikan pengguna. ELIZA menggunakan pendekatan NLP dengan metode pencocokan pola dimana dilakukan pengidentifikasian kata kunci (*keyword*) dari string *input* untuk memilih respon yang paling tepat dari aturan atau *template* yang telah dibuat sebelumnya. Contohnya, untuk *input* “*I am sad*”, kata ‘*sad*’ dipilih sebagai kata kunci dan jawaban yang diberikan adalah “*How long have you been ‘sad’?* “. *Turing test* yang diberikan pada ELIZA memberikan hasil bahwa ELIZA memiliki kecerdasan dan menyerupai manusia, tetapi gagal saat diberikan pertanyaan-pertanyaan yang kompleks dan rumit.

Pada tahun 2019, OpenAI mengenalkan Model Generative Pre-Training-2 (GPT-2) yang kemudian dikembangkan menjadi GPT-3, sebuah model *chatbot unsupervised generative* paling mutakhir yang telah memiliki kemampuan dalam pemodelan deep language, mengurangi anotasi manual dan menghindari pen-*training*-an model baru dari nol. Model *chatbot* pada GPT telah melampaui melakukan tugas *summarizing*, *reading comprehension*, *answering question*, dan *translation* dengan baik (Radford *et al*, 2019). Dengan pemodelan *unsupervised generative* ini GPT telah mampu menghasilkan respon jawaban yang sangat mirip dengan manusia. Gambaran model *chatbot generative* dapat dilihat pada **Gambar 2.1**



Gambar 2.1 Model *Generative Chatbot* (Soufyane *et al*, 2021)

(Soufyane *et al*, 2021) Berdasarkan *respons generated-based*, *chatbot* diklasifikasikan menjadi *retrieval-based model* dan *generative model* (model yang dimiliki GPT). *Chatbot* dengan model *retrieval-based* cenderung lebih sederhana dan lebih mudah diimplementasikan meskipun kita tidak bisa mendapatkan jawaban yang 100% akurat, tetapi kita bisa mengetahui respons yang memungkinkan dan memastikan *chatbot* tidak menghasilkan jawaban yang tidak sesuai. *Chatbot* dengan model *retrieval-based* mempertimbangkan konteks percakapan dan daftar pertanyaan-jawaban yang telah ditentukan sebelumnya untuk menghasilkan respon yang paling tepat (Soufyane *et al*, 2021). *Chatbot* yang dikembangkan pada penelitian ini adalah *chatbot* dengan model *retrieval-based*. Gambaran model *chatbot retrieval-based* dapat dilihat pada **Gambar 2.2**



Gambar 2.2 Model *Retrieval-based Chatbot* (Soufyane *et al*, 2021)

2.3. Text Preprocessing

Arsitektur *chatbot* memiliki 3 subproses yaitu *text preprocessing*, *semantic understanding* dan *response generation* (Maroengsit *et al*, 2019). Pemrosesan teks menjadi salah satu kunci penting dalam algoritma *machine learning* berbasis teks dan berpengaruh pada hasil akurasi. Pada penelitian *text preprocessing* yang akan diterapkan meliputi:

1. *Case folding* yaitu mengkonversi seluruh huruf yang ada pada data korpus menjadi huruf kecil. Tujuan *case folding* membantu peng-grupan *terms* yang memiliki makna yang sama dengan susunan huruf (*upper* atau *lowercase*) yang berbeda sehingga bisa mengurangi jumlah fitur yang perlu diolah oleh mesin (Uysal *et al*, 2014).
2. *Stopwords Removal* yaitu menghilangkan kata-kata *low-value* atau kata-kata yang tidak memiliki makna signifikan pada data korpus dengan tujuan meningkatkan efisiensi

sistem. *Stopwords removal* dapat meningkatkan akurasi pada beberapa penelitian yang telah dilakukan sebelumnya (Uysal *et al*, 2014).

3. *Tokenizing* yaitu prosedur memecah teks ke bentuk kata-kata, frasa, atau bagian-bagian yang bermakna yang disebut sebagai token.

4. *Stemming* yaitu prosedur untuk menghasilkan *stem* atau kata dasar dari sebuah kata dengan menghilangkan imbuhan.

2.4. Spelling Corrector oleh Peter Norvig

Penerapan *spell corrector* pada penelitian dilakukan karena mempertimbangkan kemungkinan *typo* yang lumrah terjadi saat *user* meng-input pertanyaan pada *chatbot*. Pada penelitian ini metode *spelling corrector* yang digunakan adalah Peter Norvig *Spell Checker*. Pada dasarnya cara kerja metode Peter Norvig adalah mencari setiap kemungkinan probabilitas kandidat untuk kata yang *typo* dengan melakukan operasi *delete* (menghapus satu karakter dari kata), *transpose* (menggeser satu karakter dari kata), *replaces* (mengganti satu karakter dari kata dengan huruf a-z) dan *insert* (menambahkan satu karakter pada kata). Kemudian sistem akan memilih kata dengan probabilitas paling tinggi sebagai kata dengan ejaan yang benar.

2.5. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF atau bisa dikenal dengan sebutan *Term Frequency – Inverse Document Frequency* adalah suatu metode algoritma untuk memberikan bobot setiap kata dengan mengukur keterhubungan kata terhadap dokumen. Metode ini meliputi dua poin penting yaitu frekuensi kemunculan suatu kata di dalam sebuah dokumen (*term frequency*) dan inverse frekuensi dari dokumen yang mengandung kata tersebut (*inverse document frequency*). Frekuensi kemunculan kata di dalam dokumen menunjukkan seberapa penting kata itu dalam dokumen. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot relevansi antar suatu kata dengan suatu dokumen akan tinggi apabila frekuensi kata itu tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut rendah pada kumpulan dokumen. Singkatnya *Term Frequency* (TF) adalah jumlah suatu kata didalam suatu dokumen (Hakim *et al*, 2014) dan *Inverse Document Frequency* (IDF) adalah komputasi dari log probabilitas *inverse* dari kata yang ditemukan dalam beberapa

dokumen (Hakim *et al*, 2014). Adapun nilai bobot TF-IDF adalah hasil perkalian dari nilai TF dan IDF yang ditunjukkan pada persamaan 2.1

$$tf\ idf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (2.1)$$

dengan

$$idf(t, D) = \log \frac{D}{df(t)} \quad (2.2)$$

Dimana

$tf\ idf(t, d, D)$ = bobot TF-IDF term t

$tf(t, d)$ = Jumlah term t dalam dokumen per total kata dalam dokumen

D = total dokumen

$df(t)$ = jumlah dokumen yang mengandung term t

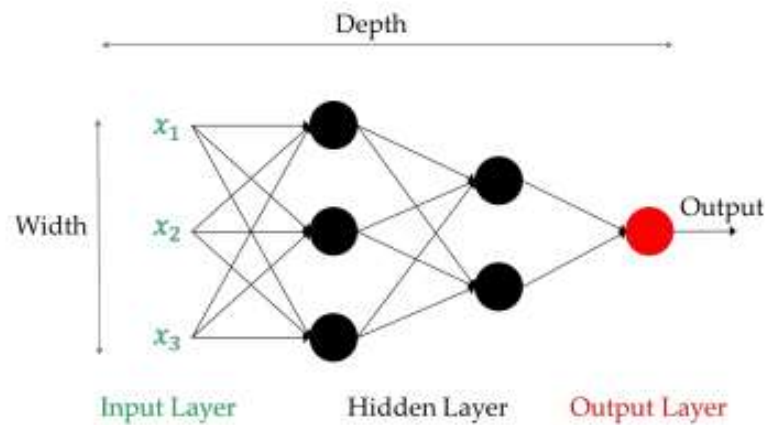
2.6. Artificial Neural Network (ANN)

García *et al* (2020) Artificial Neural Network adalah algoritma *machine learning* yang dibuat berdasarkan konsep sistem saraf pada manusia. Seperti halnya struktur saraf pada otak manusia, Artificial Neural Networks tersusun atas struktur *neuron* yang kompleks dan nonlinear dimana setiap *neuron* masing-masing terhubung satu sama lain dengan koefisien atau *weighted links* (Malekian *et al*, 2021). Artificial Neural Network merupakan model yang cukup baik untuk mengatasi masalah terkait prediksi, klasifikasi, dan pengenalan pola kompleks baik dalam format data teks ataupun gambar. Setiap proses pada model ANN yaitu penyimpanan dan analisis data, desain struktur jaringan, jumlah dari *hidden layers*, simulasi jaringan, dan *weights/bias trade-off* dimana semuanya dikomputasi melalui metode pembelajaran dan *training*.

2.7. Deep Neural Network (DNN)

Lim, *et al* (2022) Deep Neural Network (DNN) adalah salah satu algoritma *machine learning* berdasarkan Artificial Neural Network (ANN), yang cara kerjanya dengan meniru prinsip dan struktur jaringan saraf manusia. Pada masalah klasifikasi dan prediksi baik pada data *categorical*, *time series*, dan *image* akan mendapatkan hasil yang efektif dan efisien dengan menggunakan model dengan kombinasi *hidden layer* (Harumy *et al*, 2022). ANN tersusun atas *input layer*, *hidden layer*, dan *outer layer*. Sedangkan DNN sendiri tersusun atas lebih dari 3 *layers* (*input layer*, $N \geq 2$ *hidden layers*, dan *output layer*). Singkatnya Deep Neural Network (DNN) adalah varian

arsitektur ANN yang memiliki banyak layer. Ilustrasi dari gambaran bentuk DNN dapat dilihat pada **Gambar 2.3**



Gambar 2.3 Gambaran bentuk *Deep Neural Network* (Putra, 2020)

dimana:

Depth: kedalaman yang mengacu pada banyaknya *layer*

Width: lebar yang mengacu pada banyaknya unit pada *layer*

Biasanya, sebuah DNN akan mengaplikasikan fungsi aktivasi pada fungsi persamaan linear untuk menyelesaikan masalah nonlinear. Pada **Gambar 2.3** DNN tersusun atas 4 *layers* yaitu 1 buah *input layer*, 2 buah *hidden layer*, dan 1 buah *output layer* dimana untuk menghitung nilai outputnya dapat dilakukan dengan menggunakan persamaan 2.3

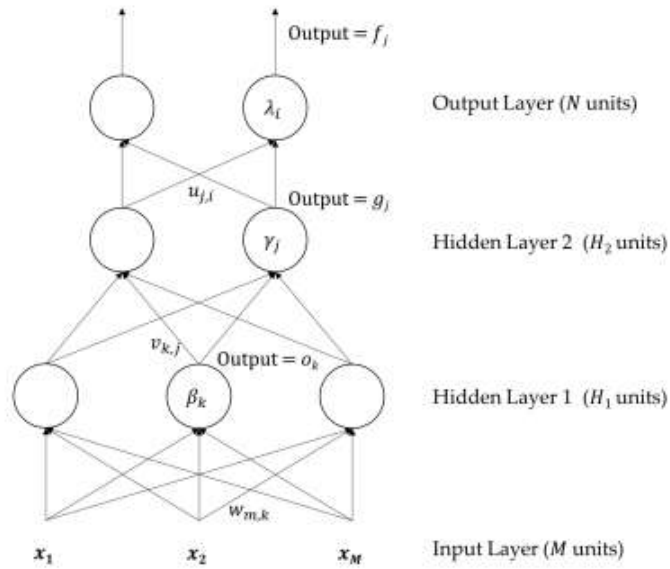
$$f_i = \sigma(\sum_{j=1}^{H_2} u_{j,i} \sigma(\sum_{k=1}^{H_1} v_{k,j} \sigma(\sum_{m=1}^M x_m w_{m,k} + \beta_k) + \gamma_j) + \lambda_i) \quad (2.3)$$

dimana:

β, γ, λ : *noise* atau *bias*

σ : fungsi aktivasi

Hasil penurunan dari persamaan 2.3 dapat dilihat pada **Gambar 2.4**



Gambar 2.4 Penurunan pada Persamaan *Deep Neural Network* (Putra, 2020)

Proses pelatihan pada *Deep Neural Network* untuk **Gambar 2.4** dapat dilakukan dengan *backpropagation* dengan persamaan-persamaan yang dapat dilihat pada **Tabel 2.1**

Tabel 2.1 Proses Pelatihan *Deep Neural Network* dengan *Backpropagation*

<p>(iii) <i>hidden layer 2 to Output</i></p> $f_i = \sigma \left(\sum_{j=1}^{H_2} g_j u_{j,i} + \lambda_i \right)$	<p>(iv) <i>output to hidden layer 1</i></p> $\delta_i = (y_i f_i) f_i (1 - f_i)$ $\Delta u_{j,i} = -\eta(t) \delta_i g_j$ $\Delta \lambda_i = -\eta(t) \delta_i$
<p>(ii) <i>hidden layer 1 to hidden layer 2</i></p> $g_j = \sigma \left(\sum_{k=1}^{H_1} o_k v_{k,j} + \gamma_j \right)$	<p>(v) <i>hidden layer 2 to hidden layer 1</i></p> $\varphi_j = \left(\sum_{i=1}^N \delta_i u_{j,i} \right) g_j (1 - g_j)$ $\Delta v_{k,j} = -\eta(t) \varphi_j o_k$ $\Delta \gamma_j = -\eta(t) \varphi_j$
<p>(i) <i>input to hidden layer 1</i></p> $o_k = \sigma \left(\sum_{m=1}^M x_m w_{m,k} + \beta_k \right)$	<p>(vi) <i>hidden layer 1 to input</i></p> $\mu_k = \left(\sum_{j=1}^{H_2} \varphi_j v_{k,j} \right) o_k (1 - o_k)$ $\Delta w_{m,k} = -\eta(t) \mu_k x_m$ $\Delta \beta_k = -\eta(t) \mu_k$

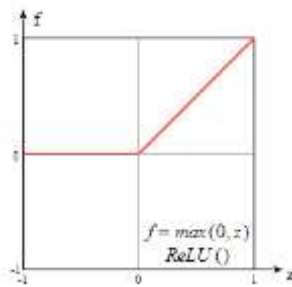
Deep Neural Network yang memiliki lebih banyak *layer* (*more depth*) dianggap lebih baik dibandingkan dengan *neural network* yang lebar (*more width*) dengan sedikit *layer* karena terjadi lebih banyak transformasi. Maksud dari lebih banyak transformasi yaitu kemampuan mengubah *input* menjadi suatu representasi (setiap *hidden layer* dapat dianggap sebagai salah satu bentuk representasi *input*) dengan langkah-langkah yang hierarkis. Meskipun terkesan kompleks, namun *Deep Neural Network* dapat bekerja dengan baik untuk mengatasi permasalahan yang praktis.

2.8. Fungsi Aktivasi

Jika kita merangkai beberapa persamaan linear maka hasil yang kita dapatkan adalah permasalahan linear. Sebagai contoh jika $f(x) = 2x + 3$ dan $g(x) = 5x - 1$, maka hubungan 2 persamaan tersebut akan menghasilkan persamaan linear pula yaitu $f(g(x)) = 2(5x - 1) + 3 = 10x + 1$. Oleh karena itu jika *neural network* tidak memiliki nilai nonlinear antara *layer* akan membuat *neural network* yang kompleks (*neural network* dengan banyak *layer*) sekalipun akan ekuivalen dengan *neural network* dengan satu *layer* sehingga tidak dapat digunakan untuk mengatasi permasalahan yang kompleks (Geron, 2022). Oleh karena itu dibutuhkan fungsi aktivasi yang akan ditambahkan pada neuron untuk memodifikasi *output*-nya (dengan memberikan nilai yang nonlinier) dimana nilai *output* yang telah dimodifikasi oleh fungsi aktivasi akan menjadi *input* untuk *layer* selanjutnya atau *output neural network*. Ada banyak sekali jenis fungsi aktivasi, pada penelitian ini yang digunakan ada dua yaitu fungsi aktivasi ReLU dan fungsi aktivasi softmax.

2.8.1. Fungsi aktivasi ReLU

Fungsi aktivasi ReLU (*Rectified Linear Unit*) bekerja dengan sangat baik pada *Deep Neural Network* (Geron, 2022). Hal ini dikarenakan mensaturasi nilai positif (dan dapat dikomputasi dengan cukup cepat). Pada fungsi aktivasi ReLU, jika nilai *output* lebih kecil dari 0 maka besar fungsi aktivasi ReLU menjadi 0. Fungsi aktivasi ReLU dapat dilihat pada **Gambar 2.5**



Gambar 2.5 Fungsi Aktivasi ReLU

Dilihat dari **Gambar 2.5** persamaan fungsi aktivasi ReLU adalah sebagai berikut.

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (2.4)$$

2.8.2. Fungsi aktivasi softmax

(Geron, 2022) Fungsi aktivasi softmax adalah fungsi aktivasi yang menyediakan estimasi probabilitas pada setiap kelas *output* dan memprediksi kelas dengan estimasi probabilitas paling tinggi (yaitu kelas dengan nilai tertinggi). Oleh karena dengan menggunakan fungsi aktivasi hanya bisa memprediksi 1 kelas dalam 1 waktu maka umumnya fungsi aktivasi ini digunakan untuk masalah *multiclass* dimana *neural network* hanya membutuhkan satu output yang benar. Persamaan fungsi aktivasi softmax adalah sebagai berikut

$$\sigma(s(x))_k = \frac{\exp(s_k(x))}{\sum_{j=1}^K (s_j(x))} \quad (2.5)$$

dimana:

K : jumlah *class*

S(x) : vektor yang memuat nilai dari setiap *class* untuk *instance* x

$\sigma(s(x))_k$: estimasi probabilitas yang *instance* x-nya adalah milik dari *class* K yang memberikan nilai dari setiap *class* untuk *instance* tersebut

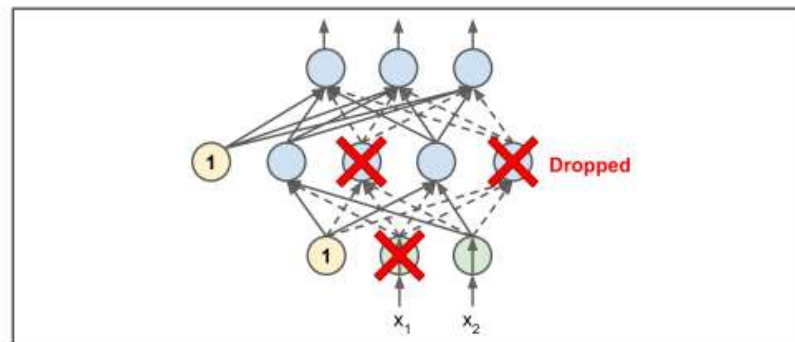
2.9. Optimizer

Cara kerja optimizer adalah mengubah nilai bobot dan bias saat proses backpropagation dengan tujuan menghasilkan nilai error yang minimum. Pada penelitian ini optimizer yang akan digunakan adalah optimizer Adam. Adam singkatan dari Adaptive moment estimation, menggabungkan ide *Momentum Optimization* dan *RMSprop*: sama seperti *Momentum Optimization* yang menyimpan riwayat rata-rata peluruhan eksponensial

gradien sebelumnya dan sama seperti *RMSprop* menyimpan riwayat rata-rata peluruhan eksponensial kuadrat gradien sebelumnya (Geron, 2022).

2.10. Dropout

Dropout adalah salah satu teknik regularisasi yang paling populer untuk *Deep Neural Network* yang pertama kali diperkenalkan oleh Geoffrey Hinton. *Dropout* adalah teknik untuk meningkatkan performa *neural network* dengan mengurangi *overfitting* (Srivastava *et al*, 2014). Cara kerja *dropout* menggunakan algoritma yang sederhana: Dimana pada setiap tahapan *training* pada setiap *neuron* (termasuk *input neuron* namun tidak termasuk *output neuron*) memiliki probabilitas p dihilangkan sementara (*dropped out*), artinya *neuron p* tersebut akan diabaikan sepenuhnya pada tahapan *training* tersebut, namun akan diaktifkan kembali pada tahap *training* selanjutnya. Ilustrasi *dropout* dapat dilihat pada **Gambar 2.6**. *Hyperparameter p* disebut sebagai *dropout rate* biasanya kisaran nilainya adalah 0.1 – 0.5.



Gambar 2.6 Dropout Regularization (Geron, 2022)

2.11. Penelitian Terdahulu

Penelitian terdahulu terkait pengembangan *chatbot* sebagai layanan informasi telah banyak dilakukan. Penelitian sebelumnya dilakukan oleh Ranoliya *et al.* (2017) *chatbot* dirancang sebagai *platform* layanan informasi akademik yang digunakan universitas untuk menjawab pertanyaan-pertanyaan mahasiswa. *Chatbot* dibangun menggunakan metode *Artificial Intelligence Markup Language* (AIML) and *Latent Semantic Analysis* (LSA). AIML digunakan untuk merespon pertanyaan-pertanyaan yang umum dan *template based* seperti salam dan sapaan. LSA digunakan untuk menjawab pertanyaan yang lebih kompleks dan *service based* dari user dimana LSA akan membandingkan kemiripan semantik antarkata yang direpresentasikan dalam bentuk vektor. AIML dan

LSA juga digunakan untuk memberikan jawaban yang bervariasi untuk merespon query yang sama sehingga jawaban yang diberikan tidak monoton.

Penelitian yang dilakukan Kandpal (2020) *chatbot* dibangun sebagai asisten kesehatan untuk memprediksi diagnosis pada beberapa masalah kesehatan dengan pendekatan metode LSTM. Pada penelitian ini LSTM digunakan untuk mengklasifikasi *user query* dan men-generate respon dengan teknik *sequence to sequence*.

Penelitian lainnya dilakukan oleh Singh *et al.* (2018) berjudul “*Chatbot using TensorFlow for small Businesses*”. Penelitian ini bertujuan untuk membangun aplikasi layanan *chatbot* untuk sebuah bisnis kecil sebagai pengganti layanan *customer support* menggunakan model *Artificial Neural Network* (ANN) dan menggunakan teknik *Bag of Words* (BoW) sebagai fitur ekstraksi sebelum data diinput ke *training model*. Penelitian berhasil membangun *chatbot* dengan akurasi 68,51%. Reza (2021) dalam penelitiannya membuat *chatbot* layanan informasi perbankan dengan menggunakan metode *Bag of Words* (BoW) untuk merepresentasikan kata kedalam bentuk vektor dan *Deep Neural Network* untuk melatih model. Penelitian ini memberikan akurasi 92.6%.

Penelitian oleh Kavitha *et al* (2019) berjudul “*Chatbot for healthcare system using Artificial Intelligence*” mengusulkan perancangan sistem *chatbot* layanan informasi kesehatan publik menggunakan algoritma *N-grams* untuk pemberian tag, TF-IDF untuk memberi bobot pada kata serta *cosine similarity* untuk melakukan pengecekan kemiripan antara dua kalimat. Selanjutnya Sulthan *et al* (2021) mengusulkan penelitian berjudul “Analisis Sentimen pada *Tweet* Bencana Alam Menggunakan *Deep Neural Network* dan *Information Gain*” untuk melakukan analisis sentimen dengan menggunakan TF-IDF untuk pembobotan kata dan *Information Gain* untuk penyeleksian fitur yang kemudian hasilnya akan diklasifikasi dengan metode *Deep Neural Network*. Penelitian ini juga membandingkan hasil dari metode yang diusulkan dengan metode-metode lain yaitu *Naïve Bayes* (NB), *Random Forest* (RF), *Support Vector Machine* (SVM), *Decision Tree* (DT). Hasil akurasi yang diperoleh dengan metode DNN + IG lebih tinggi 83.15% dibandingkan dengan metode lain dimana yang lain memperoleh akurasi *Naïve Bayes* 66.08%, *Random Forest* (RF) 63.86%, *Support Vector Machine* (SVM) 74.06, *Decision Tree* (DT) 75.17%.

Rangkuman dari penelitian-penelitian terdahulu yang telah dipaparkan diatas dapat dilihat pada **Tabel 2.2**

Tabel 2.2 Penelitian Terdahulu

No	Peneliti	Metode	Keterangan
1	Ranoliya et al. (2017)	<i>Artificial Intelligence Markup Language (AIML), Latent Semantic Analysis (LSA)</i>	Pengembangan <i>chatbot</i> untuk menjawab pertanyaan mahasiswa menggunakan metode AIML dan LSA. AIML untuk merespon pertanyaan yang umum dan LSA digunakan untuk menjawab pertanyaan yang lebih kompleks.
2	Kandpal et al. (2020)	<i>Long Short-Term Memory</i>	Pengembangan <i>chatbot</i> dengan menggunakan LSTM untuk mengklasifikasi <i>user query</i> dan <i>generate</i> respon dengan teknik <i>sequence to sequence</i>
3	Singh et al. (2018)	<i>Artificial Neural Network, Bag of Words</i>	Pengembangan <i>chatbot</i> dengan menggunakan ANN sebagai model dan BOW sebagai fitur ekstraksi dan memberikan akurasi 68,51%.
4	Reza. (2021)	<i>Bag of Words (BOW), Deep Neural Network</i>	Pengembangan <i>chatbot</i> dengan menggunakan BOW untuk representasi kata kedalam bentuk vektor dan DNN sebagai model untuk melakukan analisis prediktif. Penelitian ini memberikan akurasi 92,6%.
5	Kavitha et al (2019)	<i>N-grams, TF-IDF, Cosine similarity</i>	Penelitian ini merancang <i>chatbot</i> layanan informasi kesehatan publik menggunakan algoritma N-grams untuk pemberian tag, TF-IDF untuk memberi bobot pada kata, serta <i>cosine similarity</i> untuk pengecekan kemiripan antara dua kalimat.

Tabel 2.2 Penelitian Terdahulu (lanjutan)

No	Peneliti	Metode	Keterangan
6	Sulthan <i>et al</i> (2021)	TF-IDF, <i>Information Gain</i> , <i>Deep Neural Network</i> (DNN), <i>Naïve Bayes</i> (NB), <i>Random Forest</i> (RF), <i>Support Vector Machine</i> (SVM), <i>Decision Tree</i> (DT)	Penelitian ini melakukan analisis sentiment terhadap <i>tweet</i> Bahasa Indonesia terkait bencana alam dengan menggunakan TF-IDF untuk pembobotan kata, <i>Information Gain</i> untuk seleksi fitur dan <i>Deep Neural Network</i> (DNN) untuk klasifikasi kemudian membandingkan hasil yang diperoleh dengan metode – metode lain. Penelitian ini memberikan hasil akurasi yang diperoleh dengan metode DNN + IG adalah hasil yang paling tinggi sebesar 83.15% dibandingkan dengan metode-metode lain dimana NB 66.08%, RF 63.86%, SVM 74.06, DT 75.17%.

2.11.1. Perbedaan penelitian

Adapun perbedaan penelitian ini dengan penelitian terdahulu yaitu pada penelitian Ranoliya *et al.* (2017), *chatbot* informasi akademik dibangun menggunakan *Artificial Intelligence Markup Language* (AIML) dan *Latent Semantic Analysis* (LSA) sedangkan pada penelitian ini *chatbot* informasi akademik dibangun dengan menggunakan TF-IDF dan DNN. Lalu pada penelitian yang dilakukan oleh Kandpal (2020), *chatbot* digunakan dalam lingkup kesehatan untuk memahami gejala pasien dan mendapatkan diagnosis awal yang dibangun dengan konsep NLP dengan *Deep Learning* menggunakan metode LSTM kemudian pada penelitian yang dilakukan oleh Singh *et al* (2018), *chatbot* digunakan untuk lingkup bisnis dan menggunakan ANN dan BoW. Berbeda dari kedua penelitian yang disebutkan sebelumnya disebutkan, pada penelitian ini *chatbot* dibangun untuk lingkup pendidikan dan dibangun menggunakan metode TF-IDF dan *Deep Neural Network*.

Penelitian sebelumnya yang sama dengan penelitian ini yaitu menggunakan *Deep Neural Network* (DNN) untuk mengembangkan *chatbot* dilakukan oleh Reza (2021)

diterapkan pada *chatbot* layanan informasi perbankan dan menggunakan BoW untuk merepresentasikan kata kedalam bentuk vektor yang berbeda dengan yang dilakukan pada penelitian ini dimana menggunakan TF-IDF untuk pembobotan kata dan diterapkan pada layanan informasi akademik. Kemudian Kavitha *et al* juga menggunakan TF-IDF untuk pembobotan kata pada *chatbot* yang dibangun, Namun berbeda dengan penelitian ini, penelitian tersebut diterapkan untuk *chatbot* Bahasa Inggris dan menggunakan *cosine similarity* untuk pengecekan antara dua kalimat untuk mendapatkan jawaban atas pertanyaan *user*. Sedangkan pada penelitian ini, jawaban diperoleh dengan melakukan pembobotan vektor oleh TF-IDF dan menggunakan pemodelan DNN dan *chatbot* diterapkan untuk Bahasa Indonesia. Kemudian penelitian terdahulu yang sama dengan penelitian ini dimana TF-IDF dan DNN diterapkan pada Bahasa Indonesia dilakukan Sulthan *et al* (2021). Namun berbeda dengan penelitian ini, pada penelitian tersebut TF-IDF dan DNN digunakan untuk sentimen analisis sedangkan pada penelitian ini TF-IDF dan DNN digunakan untuk *chatbot* layanan informasi akademik.

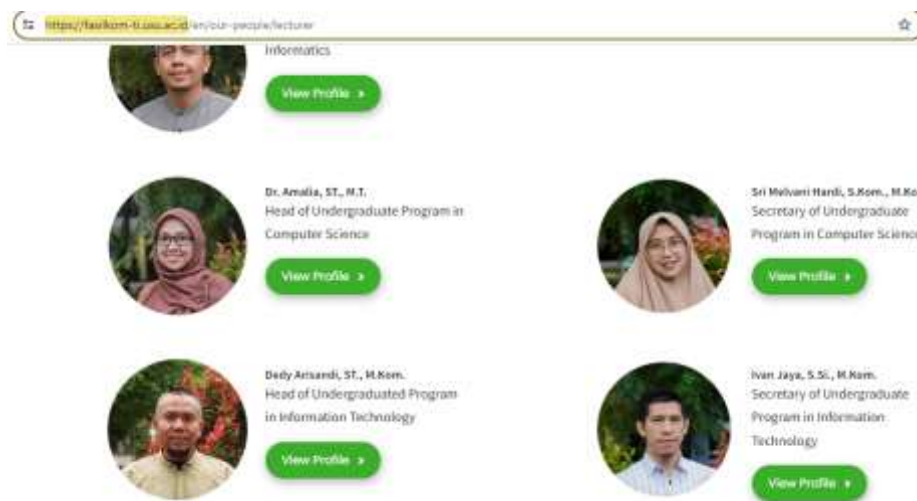
BAB 3

ANALISIS DAN PERANCANGAN SISTEM

3.1. Data

Dataset yang digunakan pada penelitian ini dikumpulkan secara manual dari berbagai sumber seperti surat edaran, e-learning usu, dan melalui *web* [https:// satu.usu.ac.id](https://satu.usu.ac.id) dan [web https://fasilkom-ti.usu.ac.id/](https://fasilkom-ti.usu.ac.id/) untuk mendapatkan informasi akademik sedangkan untuk variasi data pertanyaan dikumpulkan dari mahasiswa Fasilkom-TI USU menggunakan google form sehingga dihasilkan dataset variasi data pertanyaan dan jawaban. Data pertanyaan jawaban ini kemudian akan digunakan untuk membangun korpus data yang disimpan dalam format file *JavaScript Object Notation* (JSON). Setiap data *intents* disusun dari tiga bagian, yaitu *tag*, *patterns* dan *respons*. *Tag* yaitu label yang menjadi tolok ukur sistem dalam merespon, *patterns* berisi variasi kalimat-kalimat pertanyaan yang mungkin ditanyakan oleh pengguna, dan *respons* adalah jawaban atas pertanyaan yang mana pada penelitian ini dibuat lebih dari satu jenis jawaban.

Pada penelitian ini jumlah data yang digunakan ada sebanyak 252 *tag* label dengan masing-masing tag memiliki 10-16 variasi pertanyaan (*patterns*) dan 1-3 variasi jawaban (*responses*). Topik yang dimuat pada korpus data adalah terkait informasi akademik pada Fasilkom-TI USU dengan fokus pada sub bagian layanan pembelajaran dan penyelesaian tugas akhir. Contoh sumber informasi yang digunakan pada penelitian ini adalah dari web fasilkom ti usu yang memuat informasi terkait kaprodi dan sekprodi pada **Gambar 3.1**. Pengumpulan data variasi pertanyaan dilakukan dengan membagikan *google form* kepada mahasiswa Fasilkom-TI USU dengan meminta mereka menginput pertanyaan untuk mendapatkan jawaban terkait Fasilkom-TI USU yang diberikan. Bentuk *Google form* yang dibagikan dapat dilihat pada **Gambar 3.2**. Tabel *Spreadsheet* variasi kalimat pertanyaan dari beberapa *user* dapat dilihat pada **Tabel 3.3** atau link <https://s.id/249zP> dan <https://s.id/249Ap>. Hasil informasi terkait Fasilkom-TI USU yang telah dikumpulkan kemudian dimasukkan kedalam file JSON dengan bentuk *snippet* data JSON dapat dilihat pada **Gambar 3.3**



Gambar 3.1 Informasi Kaprodi dan Sekprodi Fasilkom-TI USU dari website Fasilkom-TI USU

Contoh variasi pertanyaan

Untuk mengetahui kriteria judul yang tepat untuk diajukan ke dosen pembimbing. Pertanyaan seperti apa yang mungkin kamu ajukan?

boleh ngga si mentuin judul yang sama persis dengan penelitian terdahulu

Nama kamu *

Jawaban Anda

Angkatan *

☐ 2018
☐ 2019
☐ 2020
☐ 2021
☐ 2022
☐ 2023
☐ Opal 7

Untuk mengetahui apa saja matkul yang diambil untuk semester 6 prodi TI. Pertanyaan yang mungkin kamu ajukan... *

Jawaban Anda

Untuk mengetahui cara mengurus cuti atau penundaan kegiatan akademik pertanyaan yang kamu ajukan yaitu... *

Jawaban Anda

Gambar 3.2 Google Form untuk Mendapatkan Variasi Pertanyaan

Tabel 3.1 Spreadsheet Variasi Kalimat Pertanyaan

Nama kamu	Angkatan	Untuk mengetahui kriteria judul yang tepat untuk diajukan ke dosen pembimbing/ apa tahap awal yang harus kamu ambil tugas : mahasiswa bebas ga sih itu form nya di isi sesuai mau nyari contoh-cont	Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info	Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info	Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info	Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info Untuk mendapatkan info
Kris G.S' Gutom	2018	Tidak bisakah judul tugas akhir langsung diberikan oleh dosen pembimbing?	2018	Tidak bisakah judul tugas akhir langsung diberikan oleh dosen pembimbing?	2018	Tidak bisakah judul tugas akhir langsung diberikan oleh dosen pembimbing?
Darius Simamora	2018	Apakah syarat penulisan judul untuk tugas akhir	2018	Apakah syarat penulisan judul untuk tugas akhir	2018	Apakah syarat penulisan judul untuk tugas akhir
Janrian Nahom Simbolor	2018	Apakah judul penelitian saya dapat mengandung unsur bidang ilmu lain?	2018	Apakah judul penelitian saya dapat mengandung unsur bidang ilmu lain?	2018	Apakah judul penelitian saya dapat mengandung unsur bidang ilmu lain?
Kristina	2018	apakah saya dapat menggunakan metode yang sudah sering digunakan oleh bagaimana memilih judul apa saja bidang penelitian bagaimana cara mendapat bagaimana saya bisa menda referensi penelitian	2018	apakah saya dapat menggunakan metode yang sudah sering digunakan oleh bagaimana memilih judul apa saja bidang penelitian bagaimana cara mendapat bagaimana saya bisa menda referensi penelitian	2018	apakah saya dapat menggunakan metode yang sudah sering digunakan oleh bagaimana memilih judul apa saja bidang penelitian bagaimana cara mendapat bagaimana saya bisa menda referensi penelitian
Alya Febriani Lubis	2018	Boleh gk metode yg saya pakai sama dengan penelitian terdahulu tapi objek	2018	Boleh gk metode yg saya pakai sama dengan penelitian terdahulu tapi objek	2018	Boleh gk metode yg saya pakai sama dengan penelitian terdahulu tapi objek
Felix Purba	2018	Apakah bisa saya mengajukan judul tidak berdasarkan permasalahan yg ser	2018	Apakah bisa saya mengajukan judul tidak berdasarkan permasalahan yg ser	2018	Apakah bisa saya mengajukan judul tidak berdasarkan permasalahan yg ser
Jessica wong	2018	kriteria judul seperti apa yang boleh diambil untuk mengajukan tugas akhir	2018	kriteria judul seperti apa yang boleh diambil untuk mengajukan tugas akhir	2018	kriteria judul seperti apa yang boleh diambil untuk mengajukan tugas akhir
Ayu Maulina	2018	Apakah judul yang umum tetapi belum pernah diterapkan di lokasi ini bisa p	2018	Apakah judul yang umum tetapi belum pernah diterapkan di lokasi ini bisa p	2018	Apakah judul yang umum tetapi belum pernah diterapkan di lokasi ini bisa p
Albert	2018	apakah judul saya sudah memenuhi kriteria	2018	apakah judul saya sudah memenuhi kriteria	2018	apakah judul saya sudah memenuhi kriteria
Sasindiah Imoeltz	2018	Bagaimana cara mencari judul yang sedang banyak diminati	2018	Bagaimana cara mencari judul yang sedang banyak diminati	2018	Bagaimana cara mencari judul yang sedang banyak diminati
Patrisia Tambunan	2018	Bagaimana ya kalau saya mengajukan judul yang mirip dengan punya senior	2018	Bagaimana ya kalau saya mengajukan judul yang mirip dengan punya senior	2018	Bagaimana ya kalau saya mengajukan judul yang mirip dengan punya senior
Putri Handayani Malik Pz	2018	Bagaimanakah kriteria judul yang bakal diterima oleh dosen?	2018	Bagaimanakah kriteria judul yang bakal diterima oleh dosen?	2018	Bagaimanakah kriteria judul yang bakal diterima oleh dosen?
Arif Pafi Lubis	2018	Berikan daftar judul skripsi tentang chatbot AI	2018	Berikan daftar judul skripsi tentang chatbot AI	2018	Berikan daftar judul skripsi tentang chatbot AI
Farhan Al Zuhri Nasution	2018	apakah judul yang saya ajukan sudah sesuai dengan yang ibu inginkan?	2018	apakah judul yang saya ajukan sudah sesuai dengan yang ibu inginkan?	2018	apakah judul yang saya ajukan sudah sesuai dengan yang ibu inginkan?
Karvin	2020	Matkul semester 6 apa aja?	2020	Matkul semester 6 apa aja?	2020	Matkul semester 6 apa aja?
Christine Amanda	2020	Matkul apa saja yang harus saya ambil untuk semester 6?	2020	Matkul apa saja yang harus saya ambil untuk semester 6?	2020	Matkul apa saja yang harus saya ambil untuk semester 6?
Fidrah Zata	2020	mata kuliah semester 6 TI USU	2020	mata kuliah semester 6 TI USU	2020	mata kuliah semester 6 TI USU
Fadhil	2021	apa aja matkul semester 6 TI?	2021	apa aja matkul semester 6 TI?	2021	apa aja matkul semester 6 TI?
Jeco	2022	daftar matkul semester 6	2022	daftar matkul semester 6	2022	daftar matkul semester 6

```

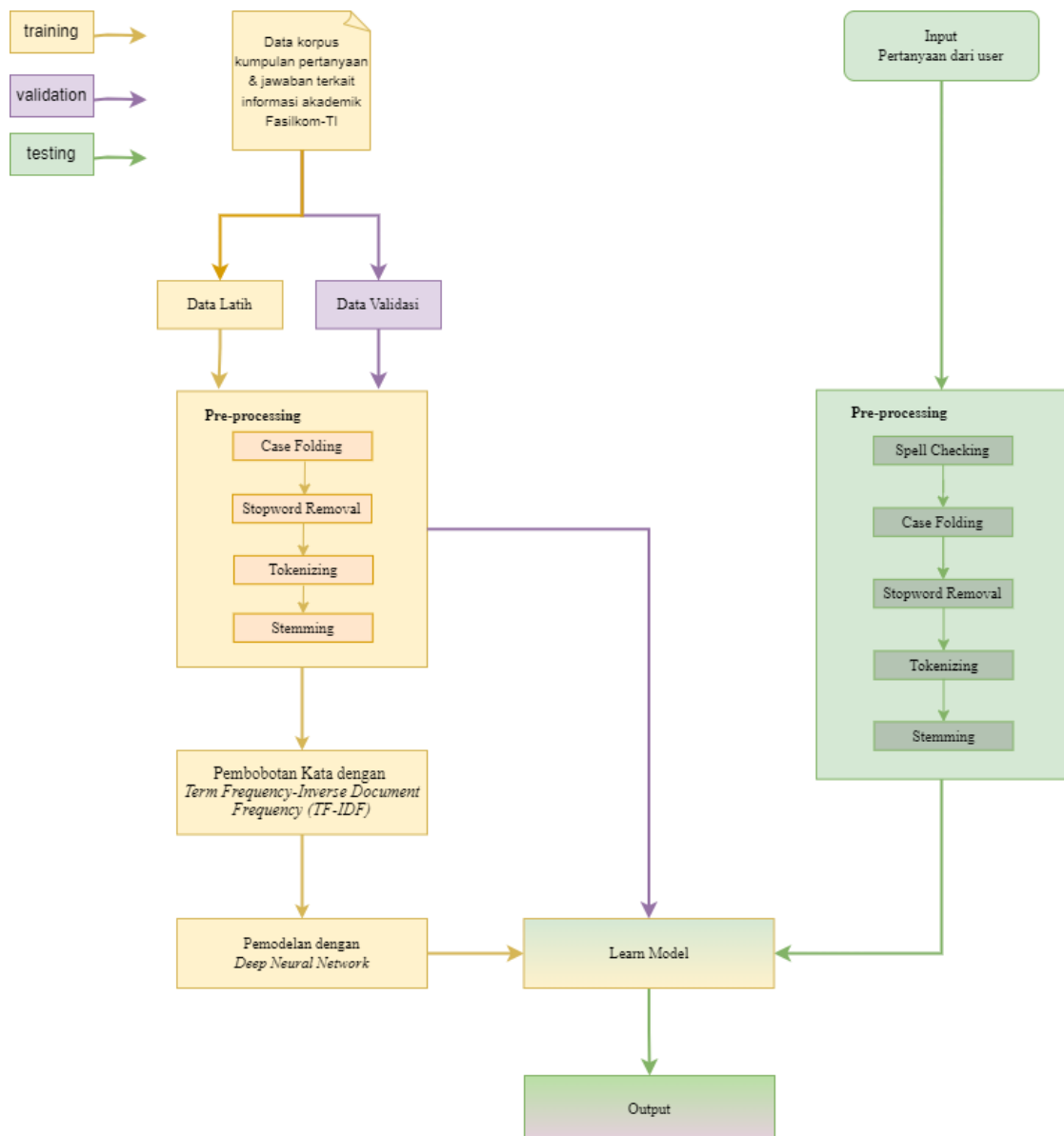
{"intents": [
  {
    "tag": "kaproditi",
    "patterns": [
      "siapa kaprodi teknologi informasi",
      "siapa kaprodi ti sekarang",
      "siapa kepala prodi teknologi informasi",
      "siapa kepala prodi ti",
      "bisa beritahu siapa kaprodi ti",
      "siapa kepala prodi teknologi informasi",
      "nama siapa kaprodi ti",
      "NIP pak dedy arisandi",
      "bisa beritahu NIP pak dedy arisandi",
      "NIP pak dedy"],
    "responses": [
      "Ketua prodi Teknologi Informasi adalah Dedy Arisandi, S.T.,
M.Kom (NIP : 197908312009121002) untuk periode 2021-2026"]
    },
  {
    "tag": "apaituexsum",
    "patterns": ["apa itu exsum",
      "bisa jelaskan terkait executivesummary",
      "exsum adalah",
      "pengertian exsum",
      "penjelasan exsum",
      "apa itu executive summary",
      "bisa jelaskan terkait exsum",
      "executive summary adalah",
      "pengertian executive summary",
      "penjelasan executive summary"],
    "responses": "Executive summary atau sering disingkat exsum
adalah ringkasan berupa poin-poin penting dari suatu dokumen. Pada exsum
tugas akhir poin-poin yang perlu kamu isi adalah sebagai berikut \n Judul,
Latar Belakang, Penelitian Terdahulu (minimal 5), Rumusan Masalah,
Metodologi, dan Referensi secara ringkas dan jelas"]
    },
  {
    "tag": "terimakasih",
    "patterns": ["terimakasih", "terimakasih sudah membantu",
      "makasih", "makasih ya", "makasih atas bantuannya", "thank you"],
    "responses": ["sama-sama", "senang bisa membantu", "terimakasih
kembali"]
    },
  ...
]

```

Gambar 3.3 *Snippet* data JSON

Pada penelitian ini akan dilakukan pembagian data dengan perbandingan 85:15 dimana sebanyak 85% dari data korpus akan menjadi data latih dan sebanyak 15% akan menjadi data validasi. Jadi dari sebanyak 4.032 data pertanyaan jawaban akan di-*split* secara acak sehingga didapatkan 3.427 data pertanyaan jawaban sebagai data latih dan 605 data pertanyaan jawaban akan digunakan sebagai data validasi.

3.2. Arsitektur Umum



Gambar 3.4 Arsitektur Umum

Rancangan arsitektur umum pada penelitian ini dapat dilihat pada **Gambar 3.4**. Arsitektur Umum pada penelitian ini dibagi menjadi beberapa tahapan yang dimulai dari tahap pengumpulan data korpus berupa *list* pertanyaan jawaban seputar informasi akademik Fasilkom-TI USU. Selanjutnya dataset akan melalui tahap *pre-processing* untuk membersihkan data yang meliputi *case folding* untuk mengkonversi semua huruf penyusun data menjadi huruf kecil, *stopwords removal* untuk menghilangkan *stopwords*, *tokenizing* untuk memecah teks pertanyaan menjadi token-token dan *stemming* untuk mengembalikan token-token tersebut kedalam bentuk dasar dengan menghilangkan

imbuhan pada kata. Berbeda dari *pre-processing* pada data korpus, pada kalimat pertanyaan masukan *user* untuk tahap *pre-processing*-nya diawali dengan *spell checking* untuk pengecekan kesalahan pengejaan kata pada kalimat.

Kemudian data hasil *pre-processing* akan dilakukan pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) yang akan memberikan hasil berupa vektor-vektor dari kata. Selanjutnya vektor-vektor tersebut akan dipelajari oleh sistem menggunakan library TFLearn melalui pemodelan DNN (*Deep Neural Network*). Model DNN ini akan melatih data korpus dengan menjadikan bobot vektor hasil TF-IDF sebagai input. Hasil dari model DNN inilah yang akan memprediksi maksud pertanyaan *user* dengan mengukur nilai bobot dan mencari mana yang paling cocok dari data yang sudah dilatih. Setelah ditemukan sistem akan memberikan jawaban yang sesuai. Jawaban sudah tersedia di daftar jawaban pada korpus data. Jawaban yang telah ditemukan selanjutnya akan dikirimkan kepada *user* melalui aplikasi perpesanan *telegram*.

3.2.1. Data input

Pada penelitian ini ada dua jenis data input yaitu dataset korpus JSON terkait Fasilkom-TI dan data input berupa kalimat pertanyaan terkait FASILKOM-TI yang diinput sendiri oleh *user*. Dataset korpus penelitian yang digunakan berupa data JSON berisi 252 *tag* label informasi, 15-16 variasi pertanyaan pada *patterns* dan 1-3 variasi jawaban pada *responses* dengan total ada sebanyak 4.032 variasi pertanyaan sebagai dataset. Perbandingan data latih dan data validasi 85:15 dimana data latih sebesar 85% yaitu sebanyak 3.427 data latih dan data validasi sebesar 15% yaitu sebanyak 605 data validasi. Sedangkan untuk pengujian dilakukan dengan menggunakan data input *user* yaitu berupa pertanyaan-pertanyaan yang di-input sendiri oleh *user*. Oleh karena *chatbot* dalam penelitian ini diintegrasikan dengan *platform chat* Telegram maka *user* mengirimkan pertanyaannya melalui aplikasi perpesanan ini. Pertanyaan tidak harus menggunakan bahasa yang baku, *user* dapat berinteraksi dengan *chatbot* menggunakan bahasa nonformal yang biasa digunakan sehari-hari. Pada penelitian ini, data *input* yang diterima sistem hanya data teks Bahasa Indonesia.

3.2.2. Preprocessing

Pre-processing dilakukan pada dua data yaitu data korpus, data format JSON yang berisi kumpulan pertanyaan-jawaban dan data input, kalimat pertanyaan *user* yang diinputkan kedalam aplikasi perpesanan. *Pre-processing* adalah tahapan memproses data yang tidak terstruktur dan mengubahnya menjadi data terstruktur sehingga memudahkan untuk pemrosesan selanjutnya. *Pre-processing* yang dilakukan pada penelitian ini dilakukan dengan beberapa tahap dimulai dari *case folding* kemudian *stopword removal* dilanjutkan dengan *tokenizing* dan terakhir *stemming*. Penjelasan tahapan-tahapan tersebut diuraikan sebagai berikut:

3.2.2.1. Case folding

Case Folding adalah tahapan menyamaratakan seluruh teks pada data dimana semua huruf diubah menjadi huruf kecil. Proses ini bertujuan untuk mempercepat perbandingan saat proses pengindeksan dan mencegah misinterpretasi dan misrepresentasi dari kata. Adapun implementasi *case folding* pada penelitian ini dapat dilihat pada algoritma berikut

Define Function to *Case Folding* yang menampung parameter *input text*:
Return hasil konversi input text ke huruf kecil dengan *.lower() method*

Adapun beberapa contoh hasil *case folding* yang diterapkan pada penelitian ini dapat dilihat pada **Tabel 3.2**.

Tabel 3.2 Contoh Penerapan *Case Folding*

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
Siapa Kaprodi TI?	siapa kaprodi ti?
Bagaimana alur pembuatan proposal	bagaimana alur pembuatan proposal
Syarat minimal sks untuk sidang?	syarat minimal sks untuk sidang?

3.2.2.2. Stopwords removal

Stopwords adalah kata-kata yang sangat sering muncul dari data teks dan bukan merupakan suatu kata kunci sehingga dapat dihilangkan karena tidak memiliki makna yang berarti. Hasil yang diperoleh melalui proses penghilangan *stopword* adalah

mendapatkan data yang hanya memuat kata-kata kunci saja sehingga memudahkan pekerjaan mesin untuk memahaminya. Salah satu *library* yang menyediakan daftar *stopword* untuk Bahasa Indonesia adalah *library* Sastrawi. Pada praktiknya, biasanya pengembang akan mendefinisikan sendiri *stopword* untuk sistem yang dikembangkan tergantung dengan kebutuhan. Pada penelitian ini penulis menggunakan daftar *stopwords* dari *library* Sastrawi dan menggabungkannya dengan daftar *stopword* yang didefinisikan sendiri oleh penulis. Berikut adalah daftar *stopwords* yang digunakan pada penelitian ini.

```
[ 'yang', 'untuk', 'pada', 'ke', 'antara', 'dia', 'ia', 'seperti',
'jika', 'sehingga', 'kembali', 'dan', 'ini', 'karena', 'kepada',
'oleh', 'saat', 'harus', 'sementara', 'setelah', 'kami', 'sekitar',
'bagi', 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal',
'ketika', 'adalah', 'itu', 'dalam', 'bisa', 'bahwa', 'atau', 'hanya',
'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka', 'sudah', 'saya',
'terhadap', 'secara', 'agar', 'lain', 'anda', 'begitu', 'mengapa',
'kenapa', 'yaitu', 'yakni', 'daripada', 'itulah', 'lagi', 'maka',
'tentang', 'demi', 'pula', 'sambil', 'sebelum', 'sesudah', 'supaya',
'apakah', 'kecuali', 'sebab', 'selain', 'seolah', 'seraya',
'seterusnya', 'tanpa', 'agak', 'boleh', 'dapat', 'dsb', 'dst', 'dll',
'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'nggak',
'mari', 'nantinya', 'melainkan', 'oh', 'ok', 'seharusnya', 'sebetulnya',
'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'ya',
'walau', 'tolong', 'tentu', 'amat', 'apalagi', 'bagaimanapun', 'ke',
'para', 'namun', 'menurut', 'tahu', 'beri tahu', 'beritahu', 'kekmana',
'sih', 'yaa', 'dong', 'kamu', 'boleh', 'yg', 'cemana', 'yaaa' ]
```

Adapun implementasi *stopwords removal* pada penelitian ini dapat dilihat pada algoritma dibawah ini.

```
Import StopWordRemoverFactory() method dari library Sastrawi
Initialize variable untuk menampung daftar stopwords
Define Function to Remove Stopwords yang menampung parameter input text:
    For setiap kata dari input text hilangkan kata jika kata tersebut ada pada
    daftar stopwords // dengan method StopWordRemoverFactory()
    Return hasil input text yang telah dihilangkan stopwords-nya
```

Contoh hasil penerapan *Stopwords Removal* pada penelitian ini dapat dilihat pada **Tabel 3.3**

Tabel 3.3 Hasil *Stopwords Removal*

Sebelum <i>Stopwords Removal</i>	Setelah <i>Stopwords Removal</i>
----------------------------------	----------------------------------

siapa kaprodi ti?	siapa kaprodi ti?
bagaimana alur pembuatan proposal?	bagaimana alur pembuatan proposal?
syarat minimal sks untuk sidang?	syarat minimal sks sidang?

3.2.2.3. Tokenizing

Tokenizing adalah proses dimana seluruh *string* pada data dipecah menjadi token-token berdasarkan kata yang menyusun *string* tersebut, yang artinya setiap kalimat akan dipecah menjadi kata-kata. Pada tahap ini juga dilakukan pembuangan tanda baca. Tujuan dilakukannya tokenisasi adalah untuk menghindari token-token yang tidak diproses dengan baik contohnya: *usu? ; ti?* dan menghasilkan token-token yang *clear* yaitu: *usu ; ti*. Adapun implementasi *tokenizing* pada penelitian ini dapat dilihat pada algoritma berikut.

```

Import word_tokenize() method dari library nltk
Initialize variable untuk menampung daftar stop_token (tanda baca)
Define Function to tokenizing yang menampung parameter input text:
    For setiap kata dari input text pecah menjadi sebuah token // dengan
    method word_tokenize()
        For token jika token tersebut tidak ada pada daftar stop_token
            Return hasil token-token hasil pemecahan input text

```

Contoh hasil penerapan *tokenizing* pada penelitian ini dapat dilihat pada **Tabel 3.4**.

Tabel 3.4 Hasil *Tokenizing*

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
siapa kaprodi ti?	'siapa', 'kaprodi', 'ti'
bagaimana alur pembuatan proposal?	'bagaimana', 'alur', 'pembuatan', 'proposal'
syarat minimal sks sidang?	'syarat', 'minimal', 'sks', 'sidang'

3.2.2.4. Stemming

Pada tahapan *stemming* dilakukan proses mengubah seluruh kata pada data menjadi kata dasar dengan menghilangkan imbuhan pada setiap kata. Tujuan dari proses *stemming*

adalah untuk menghindari mesin memproses 2 token yang memiliki makna yang sama seperti ‘bagaimanakah’ dan ‘bagaimana’ sehingga membuat proses pencarian informasi lebih cepat dan efisien. Proses *stemming* Bahasa Indonesia pada penelitian ini dilakukan dengan menggunakan *library* Sastrawi. Adapun implementasi *stemming* pada penelitian ini dapat dilihat pada algoritma berikut.

Import Stemmer Factory dari library Sastrawi
Define Function *to Remove Stopwords* yang menampung parameter *tokens*:
Initialize variable untuk menampung *method create_stemmer* yang dipanggil dari Stemmer Factory
For setiap token dari *tokens* ubah ke bentuk dasar // dengan *method stem()* dari *create_stemmer*
Return hasil token yang telah di-*stemming* dan di-*filter* dari tanda baca

Beberapa contoh hasil *stemming* yang diterapkan pada penelitian ini dapat dilihat pada **Tabel 3.5**.

Tabel 3.5 Hasil *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
‘siapa’, ‘kaprodi’, ‘ti’	‘siapa’, ‘kaprodi’, ‘ti’
‘‘bagaimana’, ‘alur’, ‘pembuatan’, ’proposal’	‘bagaimana’, ‘alur’, ‘buat’, ‘proposal’
‘syarat’, ‘minimal’, ‘sks’, sidang’	‘syarat’, ‘minimal’, ‘sks’, sidang’

3.2.2.5. *Spell checking*

Spell checking adalah proses memperbaiki kesalahan ejaan kata pada kalimat masukan *user*. Proses ini dilakukan pada kalimat pertanyaan yang diinput oleh *user* pada aplikasi perpesanan. Pada penelitian ini proses ini akan dilakukan menggunakan algoritma *Peter Norvig*. Adapun algoritma dari *Spelling Corrector* oleh Peter Norvig dapat dilihat pada algoritma berikut.

Function to Generate sebuah kamus kata

Function to Return perbandingan jumlah munculnya kata dan total keseluruhan kata

Function to Return kandidat kata dengan probabilitas paling tinggi

Function to Check jika sebuah kata ada pada kamus kata

Function to Generate iterasi semua kemungkinan kata dengan perubahan 1 huruf

Function to Return kata dengan ejaan yang benar yang probabilitasnya paling tinggi

Kamus kata pada penelitian ini didefinisikan sendiri oleh penulis berdasarkan kebutuhan yaitu kata-kata yang dianggap penting dan rentan *typo* dapat dilihat pada

Gambar 3.5

```
kamuskata = words(open('kosakata.txt').read())
print(kamuskata)

['proposal', 'seminar', 'skripsi', 'exsum', 'bagaimana', 'pembimbing', 'bimbingan', 'judul', 'semhas', 'dosen', 'sempro',
'lab', 'penguji', 'doping', 'nilai', 'cara', 'akhir', 'tugas', 'kapan', 'berapa', 'memerlukan', 'menuju', 'lulus', 'tulis',
'program', 'akhir', 'form', 'teliti', 'pembagian', 'kepala', 'proses', 'turnitin', 'digrade', 'akreditasi', 'menggunakan',
'ilmu', 'plagiarisme', 'zoom', 'sekarang', 'kaprodi', 'executive', 'summary', 'presentasi', 'fasilkom', 'file', 'link',
'alat', 'harus', 'bidang', 'format', 'ilkom', 'sekprodi', 'ngapain', 'jadual', 'informasi', 'prodi', 'huruf', 'daftar', 'w
aktu', 'kalab', 'username', 'virtual', 'background', 'kertas', 'laboratorium', 'komputer', 'teknologi', 'exum', 'tahap',
'dekan', 'dokumen', 'umum', 'slur', 'syarat', 'ngerjain', 'cover', 'smpul', 'selamat', 'computer', 'komputer', 'ilmu', 'wa
rna', 'sekretaris', 'vision', 'video', 'jenis', 'ulang', 'akademik', 'tepat', 'diskusi', 'ppt', 'plagiarism', 'plagiarisme',
'rektor', 'belakang', 'biasa', 'calon', 'tabel', 'nomor', 'spasi', 'iot', 'chat', 'wajib', 'nim', 'drive', 'revisi', 'c
heck', 'mulai', 'kalender', 'perangkat', 'margin', 'tulang', 'paragraf', 'elearning', 'fakultas', 'multimedia', 'barcode',
'scan', 'mengupload', 'poin', 'kunci', 'selesai', 'mendownload', 'jilid', 'panjang', 'gambar', 'referensi', 'fasilkomti']
```

Gambar 3.5 Daftar Kamus Kata

Beberapa contoh hasil penerapan *Spell Checking* pada penelitian ini dapat dilihat pada

Tabel 3.6

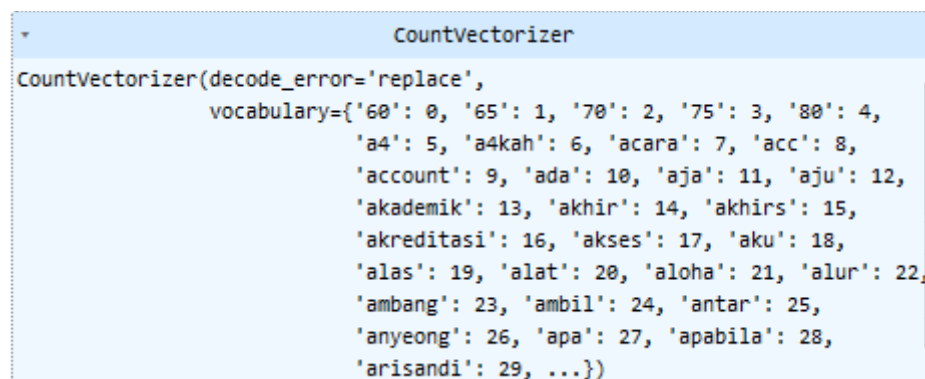
Tabel 3.6 Hasil *Spell Checking*

Sebelum <i>Spell Checking</i>	Setelah <i>Spell Checking</i>
Siapa 'kprodi' TI?	Siapa 'kaprodi' TI?
Bagaiman alur pembuatan 'priposal'?	Bagaimana alur pembuatan 'proposal'?
Berapa jumlah 'skss' minimal yang harus diambil untuk pengajuan sidang?	Berapa jumlah 'sks' minimal yang harus diambil untuk pengajuan sidang?

3.2.3. Term Frequency – Inverse Document Frequency (TF-IDF)

Setelah melalui tahapan *preprocessing* dilakukan akan dihasilkan data yang memuat token-token kata yang telah diubah menjadi kata-kata dasar. Hasil tersebut kemudian akan dihitung bobot TF-IDFnya. Tahapan ini disebut juga sebagai *feature extraction* yaitu perepresentasian fitur-fitur dalam bentuk vektor agar dapat dipahami oleh mesin. Pada penelitian ini pembobotan TF-IDF berfungsi untuk mencari kata-kata penting atau *keyword* dari setiap *tag* dan mencari kata-kata yang tidak penting yang perlu diabaikan sehingga tidak mengganggu efektivitas *Neural Network*.

Cara kerja TF-IDF adalah pertama-tama dengan melakukan pemberian *index* pada token-token kata dasar yang unik hasil *pre-processing* dimana pada penelitian ini terdapat 665 kata unik yang masing-masing akan diberikan *index* dengan urutan *ascending*. Hasil dari proses pemberian *index* (*Counter Vectorizer*) dapat dilihat pada **Gambar 3.6**. Pada penelitian ini setiap kalimat pertanyaan (*patterns*) pada penelitian ini akan menjadi *document* yang akan dihitung nilai vektor atau bobotnya.



```
CountVectorizer
CountVectorizer(decode_error='replace',
               vocabulary={'60': 0, '65': 1, '70': 2, '75': 3, '80': 4,
                          'a4': 5, 'a4kah': 6, 'acara': 7, 'acc': 8,
                          'account': 9, 'ada': 10, 'aja': 11, 'aju': 12,
                          'akademik': 13, 'akhir': 14, 'akhirs': 15,
                          'akreditasi': 16, 'akses': 17, 'aku': 18,
                          'alas': 19, 'alat': 20, 'aloha': 21, 'alur': 22,
                          'ambang': 23, 'ambil': 24, 'antar': 25,
                          'anyong': 26, 'apa': 27, 'apabila': 28,
                          'arisandi': 29, ...})
```

Gambar 3.6 Index untuk Setiap Token Kata

Selanjutnya adalah dilakukan perhitungan vektor untuk masing masing kalimat pertanyaan pada *patterns / documents*, berikut adalah demonstrasi kalkulasi TF-IDF pada 3 kalimat pertanyaan pada penelitian ini:

Vektor 494 : *docs* [494] bagaimana alur buat proposal

Vektor 1840 : *docs* [1840] syarat minimal sks sidang

Vektor 2330 : *docs* [2330] siapa kaprodi ti

Feature index dari vektor 494 adalah {'bagaimana': 42, 'alur': 22, 'buat': 79, 'proposal': 464}

Feature index dari vektor 1840 adalah {'syarat': 562, 'minimal': 345, 'sks': 541, 'sidang': 535}

Feature index dari vektor 2330 adalah {'siapa': 533, 'kaprodi': 244, 'ti': 598}

Selanjutnya dilakukan perhitungan nilai TF pada masing-masing 3 kalimat pertanyaan yang telah disebutkan sebelumnya. Pada **Gambar 3.7** dapat dilihat sebagian *dataframe* hasil perhitungan TF pada penelitian ini yang mengandung token-token kata penyusun 3 kalimat pertanyaan diatas.

	alur	bagaimana	buat	kaprodi	minimal	proposal	siapa	sidang	sks	syarat	ti
494	1		1	1	0	0	1	0	0	0	0
1480	0		0	0	0	0	0	0	0	0	0
2330	0		0	0	1	0	0	1	0	0	1

Gambar 3.7 Snippet *Dataframe* Hasil Perhitungan TF

Selanjutnya dilakukan perhitungan bobot IDF pada masing-masing token-token kata sebelumnya (*features name*) yang dapat dilihat pada **Tabel 3.7**

Tabel 3.7 Tabel Hasil Perhitungan Bobot IDF

	feature_name	idf_weights
22	alur	5.870540
42	bagaimana	4.108194
79	buat	4.462739
244	kaprodi	6.233445
345	minimal	6.703449
464	proposal	3.206942
533	siapa	3.906168
535	sidang	3.733035
541	sks	5.540298
562	syarat	5.870540
598	ti	4.962983

Selanjutnya adalah perhitungan TF-IDF yang dihasilkan dengan mengalikan nilai TF dan IDF yang dapat dilihat hasilnya pada **Gambar 3.8**

	alur	bagaimana	buat	kaprodi	minimal	proposal	siapa	sidang	sks	syarat	ti
494	5.87054	4.108194	4.462739	0.000000	0.000000	3.206942	0.000000	0.000000	0.000000	0.000000	0.000000
1840	0.000000	0.000000	0.000000	0.000000	6.703449	0.000000	0.000000	3.733035	5.540298	5.87054	0.000000
2330	0.000000	0.000000	0.000000	6.233445	0.000000	0.000000	3.906168	0.000000	0.000000	0.000000	4.962983

Gambar 3.8 Snippet *Dataframe* Hasil Perhitungan

Oleh karena pada penelitian ini proses TF-IDF dilakukan dengan menggunakan *library sklearn* dimana pada *library* ini dilakukan normalisasi terhadap hasil nilai TF-IDF sebelumnya menggunakan *Euclidean Norm* dimana hasil normalisasinya dapat dilihat pada **Gambar 3.9**. Hasil dari akhir proses pembobotan TF-IDF ini akan berupa vektor-vektor yang kemudian menjadi input untuk model *Deep Neural Network*.

	alur	bagaimana	buat	kaprodi	minimal	proposal	siapa \
494	0.650115	0.454949	0.494212	0.000000	0.000000	0.355143	0.000000
1840	0.000000	0.000000	0.000000	0.000000	0.601915	0.000000	0.000000
2330	0.000000	0.000000	0.000000	0.702451	0.000000	0.000000	0.440189

	sidang	sks	syarat	ti
494	0.000000	0.000000	0.000000	0.000000
1840	0.335196	0.497473	0.527126	0.000000
2330	0.000000	0.000000	0.000000	0.559282

Gambar 3.9 Hasil Normalisasi Bobot TF-IDF dengan *Euclidean Norm*

3.2.4. Implementasi Deep Neural Network (DNN)

Data korpus yang telah direpresentasikan menjadi bentuk bobot vektor hasil TF-IDF selanjutnya akan dijadikan sebagai masukan untuk model *Deep Neural Network* untuk kemudian dilatih dengan menggunakan model ini. DNN tersusun atas *input layer*, beberapa *hidden layer*, dan *output layer*. Berikut adalah tahapan-tahapan yang dilakukan dalam pengimplementasian model *Deep Neural Network*:

1. Membuat Sequential-type Model

Untuk merancang arsitektur model *deep learning* dilakukan dengan pemodel tipe *sequential* yaitu tumpukan banyak layer dengan 1 *input* dan 1 *output*. Pengimplentasian model ini dilakukan dengan memanggil fungsi *Sequential()*

2. Input Layer

Selanjutnya adalah mendefinisikan jumlah *n* input node untuk input layer dengan menginisiasi nilai *input shape* yang mana nilainya equivalen dengan jumlah data variabel vektor dari hasil pemrosesan TF-IDF yang dilakukan sebelumnya.

3. Hidden Layer

Selanjutnya adalah menentukan jumlah *hidden layer* yang akan dilatih dengan data dan mendefinisikan jumlah unit pada setiap *hidden layer*. *Hidden layer* pada penelitian ini akan menggunakan fungsi aktivasi ReLU yang akan menambahkan nilai non-linear pada model untuk meningkatkan performa komputasi.

4. *Output Layer*

Selanjutnya adalah menentukan jumlah unit pada *output layer* yang mana besarnya adalah jumlah kategori untuk *output*. Pada penelitian ini adalah jumlah *tag* pada data korpus *json* yaitu sebanyak 252 buah tag informasi.

5. *Dropout Layer*

Dropout layer berfungsi untuk mencegah terjadinya *overfitting* pada saat pelatihan model. Cara kerjanya adalah dengan melepaskan sementara hubungan *backwards* dan *forward* dengan *node* yang ditandai sebagai *dropped node*. Biasanya penentuan *dropout* dalam kisaran angka antara 0 dan 1. Untuk model pada penelitian ini dilakukan beberapa percobaan menggunakan beberapa nilai *dropout* yaitu 0.2, 0.3, 0.4, dan 0.5 pada *input* dan *hidden layer* 1 dan 2.

6. *Compile Model*

Selanjutnya adalah tahapan *meng-compile* semua layer yang telah didefinisikan sebelumnya. Pada bagian ini didefinisikan beberapa *hyperparameter* yaitu *loss* (*categorical_crossentropy*), *optimizer* (*Adam*), dan *metrics*(*accuracy*).

Arsitektur Model yang dibangun pada penelitian ini adalah sebuah model *Deep Neural Network* yang memiliki 1 buah *input layer* dengan 64 *node*, 2 buah *hidden layer* dengan masing-masing memiliki 128 *node* dan 256 *node* dan 1 buah *output layer* dengan 252 *node* (jumlah variasi *output* atau *tag*).

Pada penelitian ini juga digunakan fungsi *early stopping* yaitu salah satu teknik regularisasi untuk mencegah *overfitting*. Monitor *early stopping* yang digunakan pada penelitian ini adalah *accuracy_val* yang artinya akan dilakukan pengecekan pada nilai *accuracy* untuk validasi dimana model disimpan pada saat nilai *accuracy* validasi paling optimal tercapai.

3.2.5. *Training model DNN*

Selanjutnya adalah melakukan pelatihan model (*training model*) sehingga model DNN memiliki pemahaman terhadap data yang dilatih. Proses ini dimulai dari memberikan data *input* pada model *Deep Neural Network* untuk kemudian dipelajari dengan proses perulangan sehingga model dapat menemukan pola dari data. Perulangan saat *training model* disebut sebagai *epoch* dan pada tiap akhir *epoch* model DNN akan melakukan validasi. Pada penelitian ini nilai *epoch* yang digunakan sebesar 150 artinya dilakukan perulangan sebanyak 150 kali untuk mempelajari data latih. Pada proses

training model DNN penelitian ini digunakan sebanyak 4.032 variasi pertanyaan sebagai data dengan perbandingan data latih sebesar 85% yaitu sebanyak 3.427 data latih dan data validasi sebesar 15% yaitu sebanyak 605 data validasi. Tujuan digunakannya data validasi adalah untuk memastikan kemampuan model dan kualitas data.

Selain *epoch*, pada proses *training* didefinisikan pula jumlah *batch size* dan *dropout* yang digunakan. *Batch size* adalah besar *minibatch* dalam satu step *feedforwarding* dan *backprogragation*. Misal ada 8000 data training, jika menggunakan *batch_size* 64 maka akan menghasilkan 125 *minibatch* sehingga dalam 1 *epoch* akan terjadi sebanyak 125 *step* atau iterasi dimana pada tiap iterasi akan dilakukan *feedforwarding* dan *backpropagation* pada setiap *minibatch* yang ukurannya 64. Sedangkan nilai *dropout* adalah besar probabilitas atau persentase *node* yang dihilangkan pada setiap iterasi (*batch*) dengan tujuan untuk mencegah *overfitting*. Pada penelitian ini pelatihan model akan dilakukan *hyperparameter tuning* dengan mencoba beberapa nilai *batch size* (16, 32, 64) dan *dropout* (0.2, 0.3, 0.4, 0.5) untuk mencari model yang performanya paling optimal. Adapun parameter-parameter yang digunakan pada pelatihan model DNN dengan proses *hyperparameter tuning* pada penelitian ini dapat dilihat pada **Tabel 3.8**

Tabel 3.8 Parameter Model DNN

Parameter	Keterangan
Fungsi Aktivasi	<i>Input Layer</i> : ReLU <i>Hidden Layer 1 dan Hidden Layer 2</i> : ReLU <i>Output Layer</i> : Softmax
<i>Batch Size</i>	16, 32, dan 64
<i>Dropout</i>	0.2, 0.3, 0.4 dan 0.5
<i>Optimizer</i>	Adam
<i>Learning Rate</i>	0.001
<i>Loss</i>	<i>categorical_crossentropy</i>
<i>Metrics</i>	<i>accuracy</i>
<i>Epoch</i>	150

Setelah proses *training model* selesai dilakukan dimana *epoch* terakhir selesai dijalankan maka model akan tersimpan dalam sebuah file dengan ekstensi h.5. File ini berisi *learn model* yaitu model yang telah belajar dan siap digunakan untuk *testing*.

3.2.6. Testing model DNN

Terdapat dua jenis *testing* model DNN yang dilakukan pada penelitian ini, yaitu:

1. Pengujian training

Hal yang diamati pada proses training adalah hasil akurasi dari data yang telah dimodelkan menggunakan algoritma Deep Neural Network

2. Pengujian testing

Hal yang diamati adalah performa dari kemampuan model dalam menjawab input berupa pertanyaan user,

3.2.7. Output

Ada dua jenis output yang dihasilkan pada penelitian ini, *output* pertama adalah besar nilai akurasi dari sistem dalam memprediksi jawaban yang sesuai atas *input* pertanyaan *user*. *Output* kedua adalah respon atau jawaban atas *input* pertanyaan *user* yang berhasil diprediksi oleh sistem.

3.2.8. Integrasi dengan Telegram

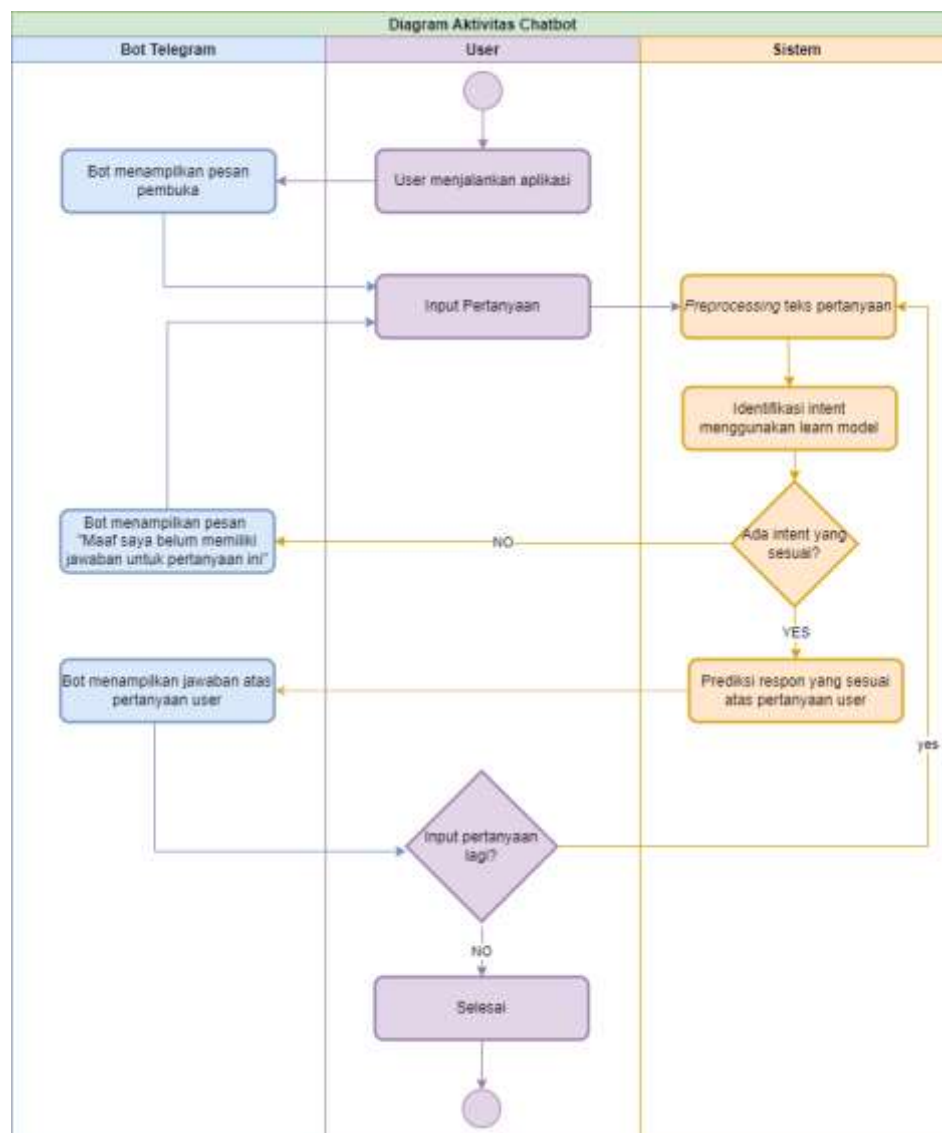
Pada penelitian ini sistem *chatbot* akan diintegrasikan dengan Telegram menggunakan API menggunakan token unik yang didapat dari fitur BotFather pada Telegram yang memungkinkan sistem terhubung dengan HTTP API Telegram. Melalui aplikasi Telegram inilah *user* dan sistem bisa saling berinteraksi dimana *user* akan mengirimkan input pertanyaan terkait Fasilkom-TI USU dan sistem akan mengirimkan hasil prediksi jawaban atas pertanyaan *user* melalui telegram. Proses interaksi ini terjadi pada ruang obrolan dengan *channel* akun Fasilkom-TI USU.

3.3. Pemodelan Sistem

3.3.1. Diagram aktivitas

Diagram aktivitas *chatbot* menggambarkan aktivitas yang dilakukan saat melakukan interaksi dengan *chatbot*. Pertama-tama saat user membuka tampilan percakapan bot Fasilkom-TI USU di telegram, *chatbot* akan merespon dengan memberikan salam pembuka dan mempersilakan *user* bertanya. Selanjutnya *user* dapat menginput pertanyaan terkait informasi di Fasilkom-TI USU. Selanjutnya sistem akan melakukan *preprocessing* terhadap teks input *user* untuk mendapatkan data yang lebih *clean*

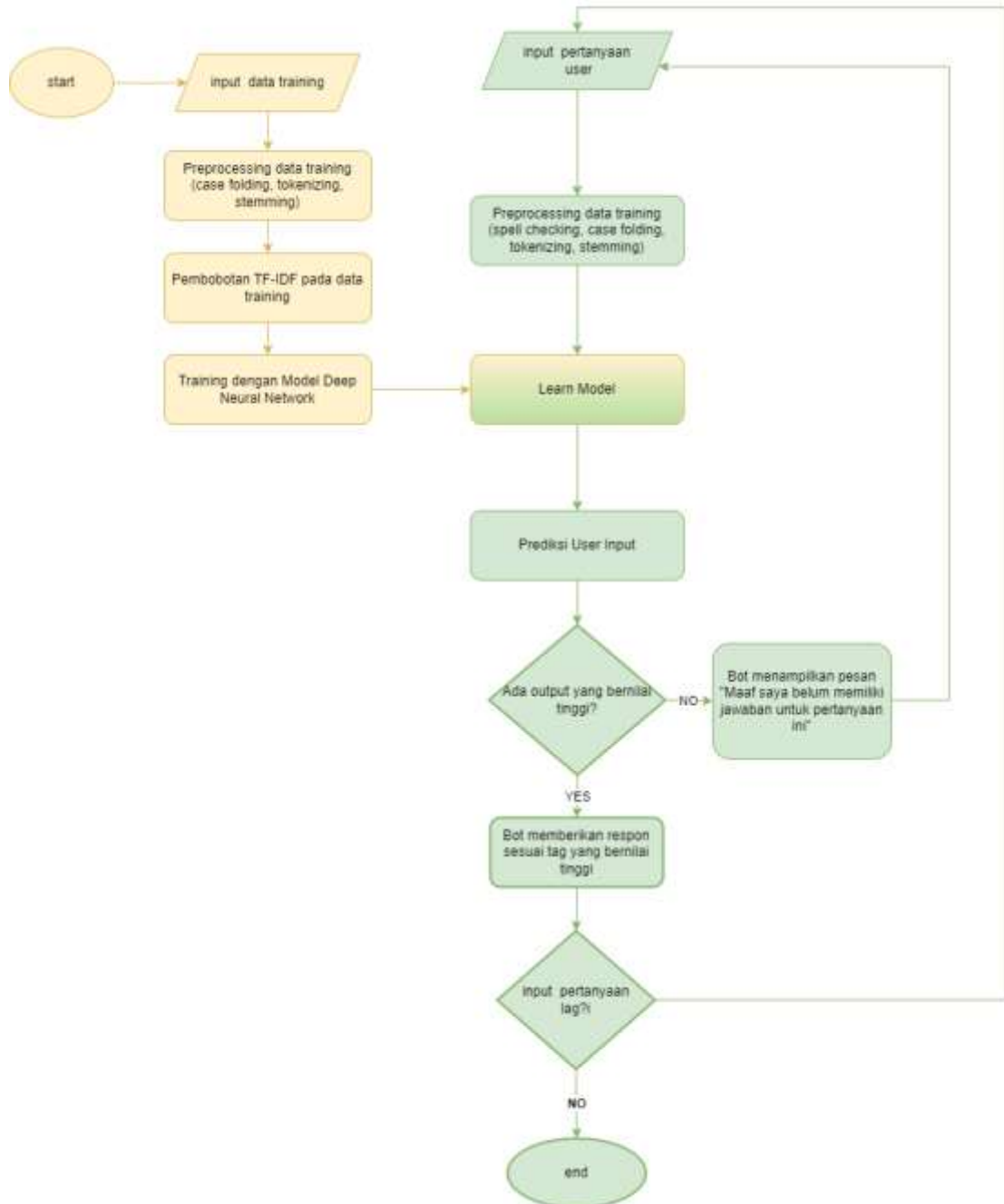
sehingga lebih mudah dimengerti oleh mesin. Selanjutnya hasil *preprocessing* input user akan direpresentasikan kedalam bentuk vektor dengan TF-IDF dan selanjutnya diidentifikasi apakah ada inten yang sesuai dengan pertanyaan *user* melalui *learn model*. *Learn model* adalah model hasil training yang dilakukan oleh algoritma *Deep Neural Network*. Jika ada inten yang sesuai dengan input pertanyaan *user*, maka *chatbot* akan menampilkan salah satu isi *responses* dari data korpus sebagai jawaban. Namun, jika tidak ada inten yang sesuai, maka artinya sistem tidak mampu memahami maksud pertanyaan *user* sehingga *chatbot* akan menampilkan jawaban pesen “Maaf jawaban atas pertanyaan ini belum tersedia”. Diagram aktivitas *chatbot* dapat dilihat pada **Gambar 3.10**



Gambar 3.10 Diagram Aktivitas *Chatbot*

3.3.2. Flowchart sistem

Flowchart Sistem adalah diagram yang menjelaskan alur proses yang dilakukan pada sistem. *Flowchart* sistem pada penelitian ini dilihat pada **Gambar 3.11**



Gambar 3.11 Flowchart Sistem Chatbot

Adapun alur proses yang dilakukan oleh sistem pada penelitian ini adalah sebagai berikut:

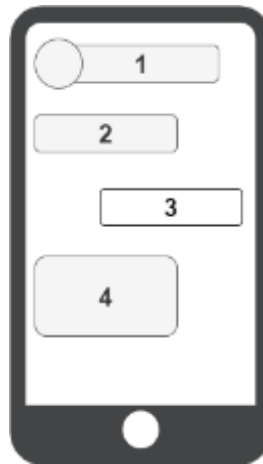
1. Pertama-tama melakukan input data korpus untuk dilakukan *training neural network* pada data. Data korpus yang diinput adalah dalam bentuk file json berisi

inten-inten pertanyaan jawaban seputar Fasilkom-TI USU dengan format *tag, pattern, responses*.

2. Selanjutnya data korpus yang diinput akan melalui tahap *pre-processing* yaitu menyamaratakan seluruh teks ke dalam huruf kecil (*case folding*), menghilangkan *stopword* (*stopwords removal*), memecah seluruh teks ke dalam bentuk token-token kata (*tokenizing*), mengubah seluruh token-token kata ke dalam bentuk dasar (*stemming*).
3. Lalu hasil data hasil *pre-processing* kemudian akan dihitung setiap bobot katanya dan direpresentasikan kedalam bentuk vektor dengan algoritma TF-IDF
4. Data dalam bentuk vektor tadi kemudian di-*traning* dengan model *Deep Neural Network* untuk dipelajari sistem.
5. Hasil traning data korpus akan disimpan kedalam *learn model neural network* yang kemudian akan digunakan untuk memprediksi jawaban untuk pertanyaan user
6. Kemudian *user* dapat berinteraksi dengan *chatbot* dengan mengi-*nput* kalimat pertanyaan seputar Fasilkom-TI USU
7. Kalimat *input user* akan melalui tahap pre-processing
8. Selanjutnya sistem akan melakukan melakukan prediksi jawaban terhadap *input* pertanyaan *user*. Jika ada *output* (probabilitas) yang bernilai tinggi > 0.7 maka sistem akan memberikan jawaban yang diambil dari salah satu kalimat *responses* yang telah disimpan. Namun, jika tidak ada *output* yang bernilai tinggi maka *chatbot* akan memberikan pesan “Maaf jawaban atas pertanyaan ini belum tersedia”. Ambang *threshold* yang diambil 0.7 untuk mengeliminasi pertanyaan yang tidak relevan dari *user* mencegah *chatbot* memberikan jawaban yang tidak tepat yang tidak ada relasinya dengan pertanyaan *user*.
9. Program berakhir jika *user* telah selesai berinteraksi dengan *chatbot*.

3.4. Perancangan Tampilan Sistem

Pada bagian ini dijelaskan bagaimana rancangan tampilan antarmuka sistem dari penelitian yang dilakukan. Rancangan tampilan sistem *chatbot* untuk penelitian ini dapat dilihat pada **Gambar 3.12**



Gambar 3.12 Rancangan Tampilan Antarmuka Sistem

Dengan keterangan sebagai berikut:

1. Avatar dan Nama Akun Bot Telegram

Nama akun bot telegram pada penelitian ini adalah Fasilkom-TI USU dengan username @fasilkom-ti_bot yang dapat diakses *user* melalui aplikasi telegram dengan mencari username atau nama bot di kolom pencarian.

2. Salam Pembuka dan Deskripsi Bot

Jika *user* pertama kali membuka ruang percakapan dengan bot, maka akan ditampilkan salam pembuka dan deskripsi dari bot Fasilkom-TI USU

3. Input Pertanyaan *User*

User dapat mengirimkan pertanyaan seputar Fasilkom-TI USU kepada bot

4. Jawaban *Chatbot*

Chatbot akan memberikan jawaban atas pertanyaan *user*.

3.5. Metode Evaluasi

Diperlukan sebuah metode evaluasi untuk mengukur performa model DNN *chatbot* yang telah di-*training*. Pada penelitian ini metode evaluasi yang digunakan adalah *matric accuracy*. Perhitungan nilai akurasi hasil pengujian model dilakukan dengan menggunakan rumus pada persamaan 3.1.

$$\text{Accuracy} = \frac{\text{Jumlah hasil analisis benar}}{\text{Total data}} \times 100\% \quad (3.1)$$

Selain *accuracy*, juga digunakan evaluasi terhadap *user satisfaction* dimana datanya dikumpulkan melalui kuisisioner yang diberikan kepada *user* setelah melakukan *usability testing*

BAB 4

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Kebutuhan Sistem

4.1.1. Spesifikasi Perangkat Keras

Spesifikasi perangkat keras yang digunakan untuk mendukung penelitian ini dapat dilihat pada **Tabel 4.1**

Tabel 4.1 Spesifikasi perangkat keras

No	Komponen	Keterangan
1	Laptop	HP 14s-cf2xxx
2	Processor	Intel® Core™ i5-10210U CPU @ 1.60GHz 2.11 GHz
3	GPU	Intel® UHD Graphics
4	Memori	12.0 GB
5	Penyimpanan	500 GB SSD
6	Sistem Operasi	Windows 11 Home Single Language

4.1.2. Spesifikasi Perangkat Lunak

Spesifikasi perangkat lunak yang digunakan untuk mendukung penelitian ini dapat dilihat pada **Tabel 4.2**

Tabel 4.2 Spesifikasi perangkat lunak

No	Komponen	Versi
1	Python	3.10.7
2	Visual Studio Code	1.85.2(user setup)
3	Telegram	-
5	Anaconda	22.9.0
6	Draw.io online	-

4.2. Implementasi Sistem

4.2.1. Implementasi tampilan antarmuka sistem

1. Halaman Profil

Pada halaman profil diberikan deskripsi akun *chatbot* Fasilkom-TI USU yaitu avatar, nama *chatbot* (Fasilkom-TI USU) dan *username chatbot* (@fasilkomti_bot). *User* dapat mengetikkan nama atau *username* tersebut pada kolom pencarian telegram untuk dapat mengaksesnya. Tampilan halaman profil dapat dilihat pada **Gambar 4.1**



Gambar 4.1 Halaman Profil *Chatbot* Fasilkom-TI USU

2. Halaman Percakapan *Chatbot* dan *User*

Halaman ini adalah halaman tempat *user* dapat berinteraksi dengan *chatbot* Fasilkom-TI USU. *User* dapat mengirimkan berbagai pertanyaan seputar Fasilkom-TI USU dan akan menerima jawaban dari *chatbot* Fasilkom-TI.



Gambar 4.2 Halaman Percakapan *Chatbot* dan *User*

4.2.2. Pelatihan model *Deep Neural Network (DNN)*

Pada penelitian ini data korpus yang diinput adalah data berupa pertanyaan-jawaban terkait Fasilkom-TI USU dengan format json (*tag*, *patterns*, dan *responses*). Jumlah tag yang berhasil dikumpulkan adalah sebanyak 252 label informasi (*tag*) dengan masing-masing *tag* memiliki 10-16 variasi pertanyaan (*patterns*) dan 1-3 variasi jawaban (*responses*). Selanjutnya data korpus akan melalui tahapan *pre-processing* (*case folding*, *stopwords removal*, *tokenizing*, dan *stemming*). Hasil dari proses *pre-processing* adalah berupa token-token kata yang sudah disamaratakan semuanya kedalam huruf kecil, diubah kebentuk dasar dan bukan merupakan *stopwords*. Token-token kata tersebut kemudian akan diberikan bobot dengan algoritma TF-IDF yang mana hasilnya adalah berupa representasi vektor dari masing-masing token-token kata tersebut dalam bentuk *matrix* [4032 rows x 665 columns] yang dapat dilihat pada **Gambar 4.3**

	60	65	70	75	80	a4	a4kah	acara	acc	account	...	wakil	\
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
...	
2993	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
2994	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
2995	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
2996	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	
2997	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	

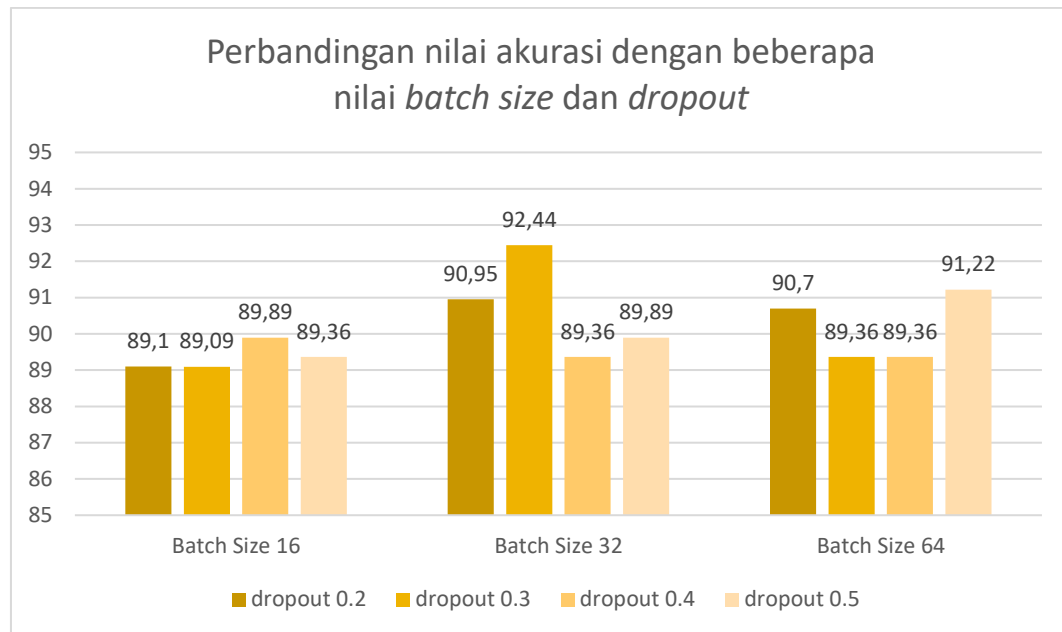
	waktu	wanita	warna	web	website	whatsapp	wisuda	ya	zoom
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
2993	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2994	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2995	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2996	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2997	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Gambar 4.3 Matriks Vektor TF-IDF

Hasil representasi vektor TF-IDF inilah yang kemudian selanjutnya akan dilakukan proses pelatihan *neural network* dengan menggunakan model DNN. Dari representasi vektor TF-IDF yang dihasilkan dari data korpus akan dibagi dengan perbandingan 85:15 dimana dari total 4.032 data sebanyak 85% data yaitu sejumlah 3.427 data akan menjadi data latih dan 15% data yaitu sejumlah 605 data akan digunakan untuk validasi saat proses pelatihan model DNN.

Model DNN pada penelitian ini memiliki 1 buah *input layer*, 2 buah *hidden layer*, dan 1 buah *output layer*. Fungsi aktivasi ReLU digunakan pada *input* dan *hidden layer* dan fungsi aktivasi *softmax* digunakan pada *output layer*. Proses pelatihan *neural network* bergantung pada pendefinisian nilai *epoch*, *batch size*, *dropout* dan *learning rate*. Pada penelitian nilai *epoch* yang diberikan sebesar 150 *epoch*, *learning rate* 0.001 menggunakan *optimizer Adam*. Pada penelitian ini akan dilakukan *hyperparameter tuning* dengan beberapa skenario pada tahap *training*, pertama menggunakan beberapa nilai *batch size* yaitu 16, 32 dan 64 dan dikombinasikan dengan beberapa nilai *dropout* yaitu 0.2, 0.3, 0.4 dan 0.5. Hasil nilai akurasi dari tahap *training* dengan menggunakan kombinasi beberapa nilai *batch size* dan beberapa nilai *dropout* dapat dilihat pada

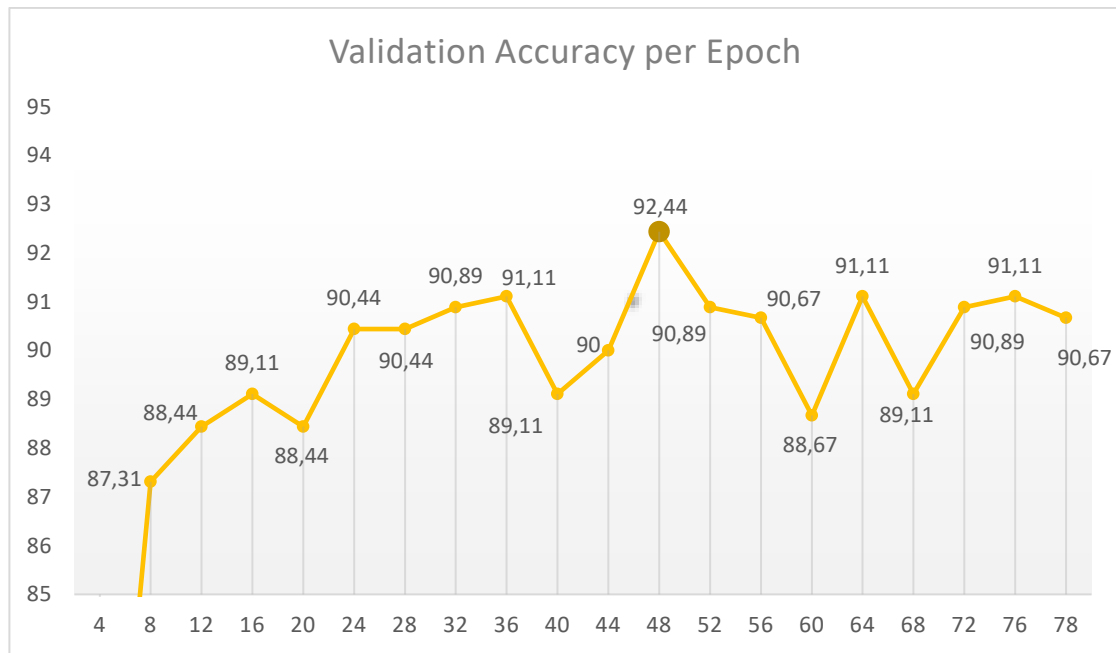
Gambar 4.4



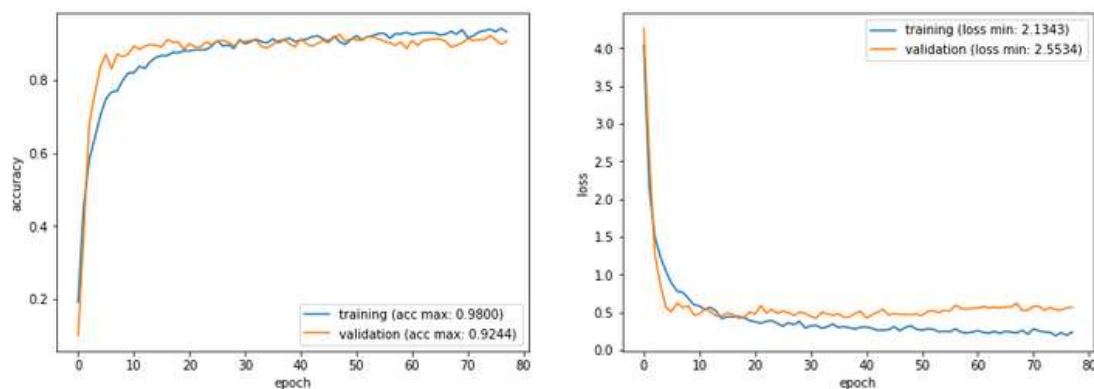
Gambar 4.4 Perbandingan nilai akurasi dengan beberapa nilai *batch size* dan *dropout*

Berdasarkan diagram perbandingan hasil akurasi dengan percobaan *training* menggunakan beberapa nilai *batch size* dan *dropout* dapat disimpulkan bahwa *training* data dengan performa paling baik dilakukan dengan menggunakan *batch size* dengan nilai 32 dan *dropout* dengan nilai 0.3 yang menghasilkan nilai akurasi sebesar 92,44%.

Proses pelatihan data pada penelitian ini menggunakan *early stopping* dengan monitor *val_accuracy* dan *patience 30* untuk mencegah *overfitting*. Dari grafik *epoch training* pada **Gambar 4.5** menunjukkan bahwa proses pelatihan data pada penelitian ini tidak mencapai *epoch* ke-150 dan berhenti diangka *epoch* ke-78. Hal ini berarti model tidak lagi mengalami peningkatan akurasi selama 30 kali setelah mencapai nilai akurasi yang paling optimal yaitu sebesar 92.44% pada saat *epoch* ke-48. Grafik *accuracy* dan *loss* selama proses *training* ini dapat dilihat pada **Gambar 4.6**



Gambar 4.5 Grafik *Epoch Training*



Gambar 4.6 Grafik *Accuracy* dan *Loss*

Model yang disimpan dalam file dengan ekstensi h5 adalah *history training* saat model memiliki akurasi paling tinggi sebesar 92.44% pada *epoch* ke-48. Pada tahap ini model *Deep Neural Network* telah dilatih untuk memahami makna dan konteks pertanyaan terkait Fasilkom-TI dan memiliki kemampuan untuk memberikan respon jawaban atas pertanyaan. Hal ini juga dibantu dengan penggunaan TF-IDF yang melakukan pembobotan vektor kata.

4.3. Pengujian *Chatbot*

Pengujian *chatbot* pada penelitian ini dilakukan melalui *usability testing* yang dilakukan oleh 30 orang *user* dengan total jumlah pertanyaan yang diuji adalah sebanyak 150 pertanyaan. Beberapa pertanyaan dari hasil pengujian yang dilakukan pada tahap ini dijabarkan pada **Tabel 4.3** dan koreksi kesalahan jawaban *chatbot* saat dilakukan pengujian dapat dilihat pada **Tabel 4.4**.

Tabel 4.3 Hasil Pengujian Model DNN

No	Input Pertanyaan	Respon <i>Chatbot</i>	Hasil	Probability
1	Siapa <i>Kprodi</i> TI?	Ketua Prodi Teknologi Informasi saat ini adalah bapak Dedy Arisandi, S.T., M.Kom. (NIP: 197908312009121002)	Sesuai	0.9977
2	Tahu nggak <i>akreditasi</i> Teknologi Informasi saat ini?	Teknologi Informasi USU terakreditasi dengan peringkat UNGGUL	Sesuai	0.7770
3	Tolong share kalender akademik dong?	Kamu dapat melihat kalender akademik pada https://www.usu.ac.id/id/kalender-akademik	Sesuai	0.9999
4	Proses penentuan doping gimana ya?	Berikut tahapan-tahapan yang dilakukan untuk mendapatkan dosen pembimbing 1. Mahasiswa wajib mengisi mengirimkan form calon dosen pembimbing yang dapat didownload di kelas.usu.ac.id 2. Isi bagian berikut: a. Nama b. NIM	Sesuai	0.9966

	<p>c. IPK</p> <p>d. Foto</p> <p>e. Bidang Ilmu</p> <p>3. Nama calon dosen pembimbing 1 dan 2 dikosongkan saja</p> <p>4. Adapun bidang ilmu diisi dengan 2 pilihan bidang ilmu yang kemungkinan akan kamu jadikan topik untuk penelitianmu. Opsi bidang ilmunya yaitu:</p> <p>a. Data Science and Intelligent System</p> <p>b. Computer System</p> <p>c. Computer Graphics and Vision</p> <p>d. Multimedia.</p> <p>5. Form ini bisa didapatkan dari kelas.usu.ac.id pada course Tugas Akhir Teknologi Informasi > Calon Peming > Calon dosen Pembimbing dan dapat disubmit pada tempat yang sama.</p> <p>6. Selanjutnya tinggal menunggu pihak kampus menginformasikan daftar dosen pembimbing tugas akhir pada mahasiswa</p>		
5	<p>Kapan boleh menghubungi dosen doping?</p> <p>Kamu dapat meghubungi pada hari dan jam kerja (Senin-Jumat pukul 08.00 – 16.00).</p> <p>Jangan menghubungi dosen dihari weekend atau hari libur nasional ya.</p> <p>Jangan lupa untuk menggunakan Bahasa yang baik dan sopan.</p>	Sesuai	0.8855

6	Bisa jelasin proses pengajuan judul gimana	<p>Berikut proses pemilihan judul yang dapat kamu lakukan:</p> <ol style="list-style-type: none"> 1. Cari tahu bidang yang diminati dari 4 bidang berikut: Data Science and Intelligent, Computer System, Computer Graphics and Vision, dan Multimedia. 2. Selanjutnya kamu bisa mencari tahu apa saja permasalahan-permasalahan yang bisa kamu selesaikan terkait dengan bidang yang kamu minati tadi 3. Kamu bisa mencari referensi dengan membaca jurnal-jurnal terbaru (sebaiknya 5 tahun terakhir) di situs jurnal kredibel seperti https://www.sciencedirect.com/, https://www.elsevier.com/, https://ieeexplore.ieee.org/, https://scholar.google.com 4. Setelah kamu mengetahui permasalahan apa yang ingin kamu selesaikan dan mencari tahu metode apa yang bisa kamu gunakan untuk menyelesaikan permasalahan tersebut kamu sudah bisa membuat judul berdasarkan dua hal tersebut 5. Selanjutnya kamu bisa mendisukikan judul yang sudah kamu buat dengan dosen pembimbing 	Sesuai	0.9999
---	--	--	--------	--------

7	Apa itu Exsum?	Executive summary atau sering disingkat exsum adalah ringkasan berupa poin-poin penting dari suatu dokumen. Pada exsum tugas akhir poin-poin yang perlu kamu isi adalah sebagai berikut Judul, Latar Belakang, Penelitian Terdahulu (minimal 5), Rumusan Masalah, Metodologi, dan Referensi secara ringkas dan jelas	Sesuai	0.9999
8	Nama file untuk exsum?	File exsum dikumpulkan dalam format pdf dengan nama file exsum_NIM.pdf	Sesuai	0.9999
9	Berapa nilai minimum exsum supaya bisa lanjut ke tahap proposal?	Jika nilai exsum ≥ 70 maka kamu bisa lanjut ke tahap proposal. Jika nilai < 70 kamu perlu membuat exsum ulang atau melakukan perbaikan sesuai dengan feedback comment yang diberikan oleh Kepala Lab	Sesuai	0.9999
10	Bisa jelaskan alur proses buat proposal nggak?	Berikut proses pembuatan proposal: 1. Setelah lulus tahap exsum, kamu sudah mulai bisa menulis proposalmu berdasarkan exsum yang sudah dibuat 2. Adapun isi proposal adalah pengembangan dari isi exsum yang terdiri dari Lembar Cover, Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Metodologi	Sesuai	0.9999

Penelitian, Landasan Teori,
Penelitian Terdahulu, Daftar Pustaka

3. Anda masih boleh melakukan perubahan dari isi exsum ke proposal, misal menggunakan algoritma yang berbeda atau menggunakan scope data yang berbeda tetapi sebelumnya harus mendiskusikannya dengan dosen pembimbing anda.

4. Setelah kamu menyelesaikan proposal selanjutnya kamu bisa mengechat dosen pembimbingmu untuk menginformasikan bahwa kamu telah menyelesaikan proposal sekaligus mengirimkan file pdf atau word proposalmu dan tanyakan kapan kira-kira dosen pembimbingmu punya waktu untuk bimbingan proposal

5. Bimbingan proposal bisa dilakukan secara online dan offline tergantung dari dosen pembimbing. Setelah melakukan bimbingan proposal dengan dosen pembimbing 1 dan telah mendapatkan revisi. Kamu bisa mengerjakan revisi yang kamu dapat.

6. Jika kamu sudah menyelesaikan revisi silahkan kirim kembali proposal yang telah kamu revisi kepada dosen pembimbing 1 dan tanyakan juga apakah masih perlu

		<p>untuk melakukan bimbingan langsung. Jika dosen pembimbing 1 sudah setuju dengan proposalmu selanjutnya kirim proposalmu ke dosen pembimbing 2. Kerjakan revisi jika ada revisi dari dosen pembimbing 2 dan kirim kembali hasil revisi ke doping 2.</p> <p>7. Setelah mendapatkan persetujuan dari doping 1 dan doping 2, kamu selanjutnya meminta tanda tangan untuk File Form Persetujuan Seminar Proposal.</p> <p>8. Selanjutnya adalah pengajuan seminar proposal.</p>		
11	tidak lulus proposal biasanya karena apa ya	<p>Ada kemungkinan untuk tidak lulus seminar proposal. Jika kamu tidak menjelaskan proposalmu dengan baik atau mungkin ada kesalahan-kesalahan fatal dari proposal yang kamu ajukan kamu bisa diminta untuk melakukan seminar proposal ulang</p>	Sesuai	0.9956
12	Apa isi ppt untuk sempro?	<p>Poin-poin yang harus kamu presentasikan saat seminar proposal dan harus dimuat dalam ppt proposalmu, yaitu:</p> <ol style="list-style-type: none"> 1. Judul penelitianmu, sebaiknya tulis nama dan nim pada lembar judul ppt 2. Latar Belakang 3. Rumusan Masalah 	Sesuai	0.9999

	4. Tujuan Penelitian			
	5. Batasan Masalah			
	6. Manfaat Penelitian			
	7. Metodologi penelitian yang memuat arsitektur umum dan contoh data			
	8. Penelitian Terdahulu dan disarankan membuat penjelasan apa perbedaan penelitian yang kamu buat dengan penelitian terdahulu			
	Ketujuh outline dibuat berdasarkan dari proposal yang kamu buat. Buatlah poin-poin presentasi yang singkat dan jelas sehingga dalam waktu 10-15 menit kamu bisa menjelaskan tentang proposal penelitianmu kepada Bapak/Ibu dosen			
13	Berapa menit waktu untuk presentasi seminar proposal?	Lama waktu yang diberikan kepada mahasiswa untuk mempresentasikan proposal penelitian skripsinya adalah sekitar 10-15 menit	Sesuai	0.9999
14	Boleh nggak menyaksikan seminar proposal lain?	Anda boleh menyaksikan seminar proposal orang lain	Sesuai	0.9999
15	Sempro adalah...	Sempro atau Seminar Proposal adalah saat dimana kamu mempresentasikan proposal penelitian yang sudah kamu susun	Sesuai	0.9999

dan sudah disetujui oleh kedua dosen pembimbingmu. Seminar proposal akan dilihat oleh dosen-dosen lain dan terbuka untuk mahasiswa-mahasiswa lain. Saat seminar proposal, setelah mahasiswa selesai mempresentasikan penelitiannya, dosen-dosen dipersilahkan untuk bertanya kepada mahasiswa penyaji terkait proposal yang dipresentasikannya. Dosen-dosen juga dipersilahkan untuk memberikan masukan-masukan kepada mahasiswa penyaji untuk penelitian mahasiswa penyaji tersebut kedepannya Mahasiswa penyaji perlu mencatat masukan-masukan dari dosen-dosen untuk dipertimbangkan dan didiskusikan dengan dosen pembimbing.

- 16 Cara minjam alat iot ke Lab Penelitian Jaringan yang gimana ya? Kamu bisa meminjam alat penelitian ke Lab Penelitian Jaringan yang berada di Gedung Ilmu Komputer Lantai. Sebelumnya kamu perlu menanyakan kesediaan alat ke Kepala Lab Penelitian Jaringan (melalui chat). Lalu untuk mendapatkan alat silahkan jumpain Kepala Lab Penelitian Jaringan di ruangannya (ruangan kepala Lab Penelitian Jaringan, Gedung Ilkom lantai 2)
-

Sesuai 0.9996

17	Apa saja alat penelitian yang tersedia di lab untuk dipinjam?	Kamu boleh meminjam alat-alat penelitian yang tersedia di lab penelitian untuk digunakan dalam melakukan penelitian seperti kamera, raspberry pi, berbagai alat sensor seperti sensor gerak, sensor suara, sensor gps, dan lain-lain.	Sesuai	0.9996
18	proses ngerjain skripsi gimana ya?	<p>Berikut gambaran yang bisa saya berikan untuk proses pengerjaan skripsi:</p> <p>1. Setelah kamu selesai seminar proposal dan mendapatkan masukan-masukan dari dosen-dosen saat sempro silahkan pertimbangkan masukan-masukan tersebut untuk skripsimu. Kamu bisa membaca lebih banyak jurnal penelitian lain untuk mendapatkan ide yang lebih banyak untuk penelitianmu. Catat ide-ide yang kamu dapat dan pertanyaan pertanyaan yang kamu punya untuk kamu diskusikan dengan dosen pembimbing</p> <p>2. Selanjutnya silahkan menghubungi dosen pembimbing 1 dan dosen pembimbing 2 untuk melakukan bimbingan terkait hasil seminal proposal. Kamu bisa menanyakan terkait masukan-masukan yang telah diberikan dosen-dosen lain saat seminar proposal, ide-ide yang kamu miliki untuk</p>	Sesuai	0.9965

penelitianmu ataupun pertanyaan-pertanyaan yang kamu perlu tanyakan

3. Hasil dari bimbingan dengan dosen pembimbing 1 dan dosen pembimbing 2 bisa menjadi modal untuk kamu mulai mengerjakan skripsi

4. Selama proses pengerjaan skripsi mulai dari pengumpulan data, membangun model, perancangan sistem antarmuka, evaluasi pengujian sistem dan penulisan akhir kamu sebaiknya melakukan bimbingan dengan dosen pembimbing dalam prosesnya. Kamu bisa mendiskusikan dengan dosen pembimbing terkait kendala yang kamu alami, pertanyaan yang kamu punya serta progres yang sudah kamu kerjakan

19	Sumber data skripsi didapat darimana ya?	Data skripsi dapat kamu ambil dari berbagai sumber seperti Kaggle dimana biasanya beberapa peneliti terdahulu menyediakan open access untuk data yang mereka gunakan dalam penelitian disini, website atau sosmed dengan cara scrapping, mengambil data langsung dengan memfoto (untuk image processing), melalui google form jika butuh data berupa kuisisioner (untuk text	Sesuai	0.9999
----	--	--	--------	--------

		processing), melakukan wawancara langsung ke narasumber atau ahli		
20	Ngerjain skripsi enakya dimana ya...	Kamu boleh mengerjakan skripsi di lab computer vision dan lab multimedia lantai 2 gedung TI atau lab jaringan lantai 2 gedung Ilkom. Disana tersedia PC dengan spesifikasi yang tinggi (RAM, VGA, Processor) sehingga mempercepat proses training sehingga sangat membantu untuk melakukan training jika membutuhkan waktu yang lama.	Sesuai	0.9999
21	di lab nggak boleh ngapain?	Pengguna Lab DILARANG untuk 1. Makan dan Minum di dalam ruangan Lab 2. Menghapus aplikasi atau data yang ada pada komputer lab 3. Melakukan aktivitas yang dapat membuat keributan di dalam ruangan lab 4. Memindahkan atau mencabut perangkat keras lab 6. Meninggalkan komputer lab dalam keadaan hidup Saat meninggalkan lab tanpa izin Kepala Lab 7. Menggunakan fasilitas Lab untuk aktivitas di luar kebutuhan akademik	Sesuai	0.9999
22	Boleh skripsian di lab sampai jam berapa?	Biasanya lab dibuka dari jam 09.00 sampai jam 18.00	Sesuai	0.9999

23	Apa itu semhas?	Semhas atau seminar hasil adalah seminar untuk mahasiswa mempresentasikan hasil penelitian yang dilakukannya. Adapun yang dipresentasikan adalah mulai dari latar belakang penelitian sampai penelitian terdahulu dan terakhir video uji program. Pada seminar hasil, dosen penguji 1 dan dosen penguji 2 berhak memberikan pertanyaan/menguji mahasiswa terkait hasil penelitian yang telah dipresentasikannya. Dosen penguji juga akan memberikan revisi terkait penelitian mahasiswa yang wajib dikerjakan oleh mahasiswa sebelum lanjut ke tahap sidang	Sesuai	0.9999
24	Apa itu uji program?	<p>Saat uji program, mahasiswa menunjukan atau mempresentasikan program/sistem yang sudah dibuat untuk penelitiannya. Berikut tata cara yang dilakukan untuk melakukan uji program</p> <ol style="list-style-type: none"> 1. Mahasiswa menghubungi pembimbing 1 atau 2 untuk menentukan jadwal uji program dan tempat melaksanakan uji program (jika offline) atau platform yang akan dipakai untuk presentasi (jika online) 2. Mahasiswa perlu mendownload dan mengeprint Form penilaian uji program dari kelas.usu.ac.id > Form 	Sesuai	0.9999

Tugas Akhir > 7. Form Penilaian Uji
Program - Form 2C.docx

3. Mahasiswa melakukan uji program dan Dosen pembimbing melakukan penilaian sesuai kriteria pada form penilaian Uji Program.

4. Mahasiswa membuat video uji program yang nanti akan disubmit di kelas.usu.ac.id. Video kemudian diupload di kelas.usu.ac.id

5. Setelah selesai melakukan uji program, mahasiswa diwajibkan memberikan submission pada kelas.usu.ac.id > Link Video Demo Program pada bagian online text.

6. Dosen akan menginput nilai yang telah ia miliki pada submission milik mahasiswa (modul 1. Uji Program wajib digrade) yang bersangkutan sebagai syarat untuk dapat melanjutkan Seminar Hasil.

25	Uji program yang dinilai apa aja sih?	Berikut Komponen Penilaian Uji Program :		
		1. Kemampuan dasar pemrograman (0-40)		
		2. Kecocokan metode/algoritma dengan sintaks program (0-10)	Sesuai	0.9994
		3. Penguasaan pemrograman berdasarkan kasus pada skripsi (0-20)		
		4. Penguasaan pembuatan User Interface (0-10)		
		5. Validasi output program (0-20)		

26	Apa isi video uji program?	Video uji program yang menampilkan uji sistem aplikasi dari penelitian yang telah kamu buat (dibuat dengan merecord layar laptop saat menge-test sistem aplikasi yang kamu buat)	Sesuai	1.0
27	kemana kemana ku harus mengurus kkm kemana?	KKM adalah transkrip penilaian mahasiswa untuk mengurus KKM silakan datang ke layanan akademik di Gedung A Fasilkom-TI pada jam kerja.	Sesuai	1.0
28	syarat minimal sks untuk bisa sidang skripsi?	SKS minimal yang harus diambil untuk bisa sidang skripsi adalah 144 sks	Sesuai	0.9666
29	info seragam sidang donggg	Pria: 1. Kemeja lengan panjang warna putih 2. Memakai jas lengan Panjang berwarna hitam 3. Celana berbahan kain berwarna hitam 4. Dasi berwarna hitam 5. Sepatu pantofel (formal) berwarna hitam 6. Rambut rapi 7. Wajib rapi Wanita: 1. Kemeja lengan Panjang warna putih	Sesuai	0.9999

			2. Jas/blazer lengan Panjang warna hitam 3. Rok berbahan kain berwarna hitam 4. Memakai rok berbahan kain warna hitam (Panjang dibawah lutut) 5. Bagi yang memakai jilbab berwarna hitam 6. Memakai flat shoes (formal) 7. Rambut rapi		
30	Margin skripsi?	1. Margin Awal Bab Margin atas-bawah-kiri-kanan : 50-25-38- 25 (mm) 2. Margin Keseluruhan Naskah Margin atas-bawah-kiri-kanan : 30-25-38- 25 (mm)	sesuai	0.9999	
31	Ada info terkait jurusan psikologi	Maaf saat ini jawaban atas pertanyaan ini belum tersedia	diluar topik cakupan chatbot	0.4429	
32	ada template chat minta bimbingan terkait exum	Assalamualaikum pak/bu Selamat pagi pak/bu Maaf mengganggu waktunya Saya [nama mahasiswa] , NIM [nim mahasiswa]. Ingin memberitahukan bahwa bapak/ibu sebagai dosen pembimbing I saya, oleh karena itu saya ingin melampirkan exum untuk pengajuan judul skripsi saya pak/bu, mohon bimbingan dan arahnya pak/bu apakah saya dapat	Sesuai	0.9998	

		melanjutkan ke penulisan proposal pak/bu.		
33	Batas pengisian KRS sampai kapan ya?	Batas pengisian KRS 16 Januari 2024 – 4 Februari 2024	Sesuai	0.8907
34	btw sampek tgl brp ya masih bis4 krsan? ada info g sih?	Maaf saat ini jawaban atas pertanyaan ini belum tersedia	Tidak Sesuai	0.4587
35	Tahu apa itu neural network	Pak Seniman, S.Kom., M.Kom. (NIP : 198705252014041001) adalah Kepala Laboratorium bidang Computer Systems, Network, and Security	Tidak Sesuai	0.9605

Tabel 4.4 Koreksi *Chatbot*

No	Kesalahan	Penjelasan
1	<p>Pertanyaan:</p> <p>Ada info terkait jurusan psikologi?</p> <p>Respon Chatbot:</p> <p>Maaf saat ini jawaban atas pertanyaan ini belum tersedia</p> <p>Keterangan:</p> <p>diluar topik cakupan <i>chatbot</i></p> <p>Probabilitas:</p> <p>0.4429</p>	<p><i>Chatbot</i> memberikan respon “Maaf saat ini jawaban atas pertanyaan ini belum tersedia” karena pertanyaan diluar topik <i>chatbot</i> atau tidak terkait Fasilkom-TI</p>

2	<p>Pertanyaan: Chatbot memberikan respon “Maaf saat ini jawaban btw sampek tgl brp ya atas pertanyaan ini belum tersedia” padahal masih bis4 krsan? ada seharusnya respon yang diberikan adalah Batas info g sih? pengisian KRS 16 Januari 2024 – 4 Februari 2024 .</p> <p>Respon Chatbot: Hal ini terjadi karena probabilitas yang diberikan Maaf saat ini jawaban tidak memenuhi ambang threshold yang atas pertanyaan ini didefinisikan yaitu sebesar 0.7.</p> <p>belum tersedia</p> <p>Keterangan:</p> <p>Tidak Sesuai</p> <p>Probabilitas:</p> <p>0.4587</p>	
3	<p>Pertanyaan: Meskipun probabilitas tinggi tapi jawaban yang Tahu apa itu neural diberikan oleh chatbot tidak sesuai. Hal ini terjadi network karena informasi terkait neural network tidak tersedia pada data korpus. Keyword yang paling</p> <p>Respon Chatbot: dianggap penting oleh sistem adalah neural network Pak Seniman, S.Kom., dan kata network ada dimuat pada intent dengan tag M.Kom. (NIP : 198705252014041001) ‘kepalabjaringan’ sehingga memberikan adalah Kepala probabilitas yang cukup tinggi oleh karena itu Laboratorium bidang respon yang diberikan adalah kepala lab jaringan Computer Systems, Network, and Security</p> <p>Keterangan:</p> <p>Tidak Sesuai</p> <p>Probabilitas:</p> <p>0.9605</p>	<p>Meskipun probabilitas tinggi tapi jawaban yang diberikan oleh chatbot tidak sesuai. Hal ini terjadi karena informasi terkait neural network tidak tersedia pada data korpus. Keyword yang paling dianggap penting oleh sistem adalah neural network dan kata network ada dimuat pada intent dengan tag ‘kepalabjaringan’ sehingga memberikan probabilitas yang cukup tinggi oleh karena itu respon yang diberikan adalah kepala lab jaringan</p> <pre> sentence_words_checked: apa itu neural network ? sentence_words_cleaned: ['apa', 'neural', 'network'] 1/1 [=====] - is 550ms/step results [[78, 0.96057606]] processed_tags [{"intent": 'kepalabjaringan', 'probability': '0.96057606'}] message_class: [{"intent": 'kepalabjaringan', 'probability': '0.96057606'}] {"intent": 'kepalabjaringan', 'probability': '0.96057606'} Kepala Lab untuk bagian Computer Systems, Network, and Security adalah 14041001) message_response: Kepala Lab untuk bagian Computer Systems, Network, and Security (NIP : 198705252014041001) </pre> <pre> { "tag": "kepalabjaringan", "patterns": ["siapa kepala laboratorium bagian computer systems", "siapa kepala laboratorium bagian jaringan", "siapa kepala lab bagian security", "siapa kalab untuk computer systems dan network", "bisa beritahu nama kalab bidang network", "NIP pak Seniman",] } </pre>

	<pre> "tahu NIP pak Seniman berapa", "siapa kepala lab bagian penelitian yang berhubungan dengan iot", "siapa kalab untuk penelitian terkait internet of things", "bisa beritahu nama kalab terkait robotik", "siapa nama kalab jaringan", "kepala lab jaringan", "kalab jaringan", "kepala laboratorium jaringan", "ka. laboratorium jaringan", "nama kepala lab jaringan"], "responses": [...] }</pre>
--	---

4.4. Evaluasi *Chatbot*

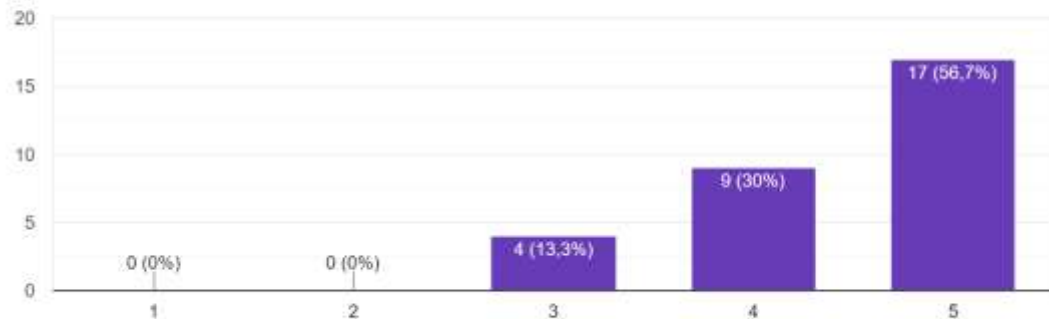
Metode evaluasi pada pengujian *chatbot* pada penelitian ini akan dilakukan melalui *usability testing* oleh 30 orang *user* yaitu mahasiswa Fasilkom-TI. Prosesnya adalah setiap *user* masing-masing memberikan 5 pertanyaan terkait Fasilkom-TI pada *chatbot* Fasilkom-TI yang telah dibangun. Selanjutnya untuk mendapatkan nilai kepuasan *user* terhadap pengalamannya menggunakan *chatbot* Fasilkom-TI, *user* diminta untuk mengisi kuisisioner *google form user satisfaction* dengan beberapa pertanyaan dengan penilaian skala 1-5. Adapun isi pertanyaan yang diberikan kepada *user* untuk melakukan evaluasi terhadap *chatbot* yaitu:

1. Dari 5 pertanyaan yang kamu ajukan berapa banyak yang berhasil dijawab dengan benar oleh Chatbot Fasilkom-TI?
2. Seberapa mudah Chatbot Fasilkom-TI untuk diakses dari platform yang disediakan saat ini?
3. Seberapa mudah jawaban yang disediakan oleh Chatbot Fasilkom-TI untuk dimengerti?
4. Berapa nilai kepuasan Anda saat menggunakan Chatbot Fasilkom-TI

Adapun ringkasan penilaian *user* yang didapatkan dari kuisisioner *google form* dapat dilihat pada **Gambar 4.7** dan **Gambar 4.8**

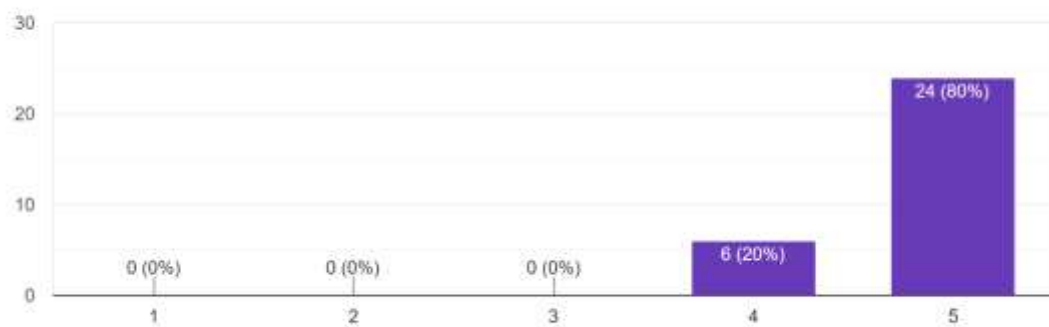
Dari 5 pertanyaan yang kamu ajukan berapa banyak yang berhasil dijawab dengan benar oleh Chatbot Fasilkom-TI?

30 jawaban



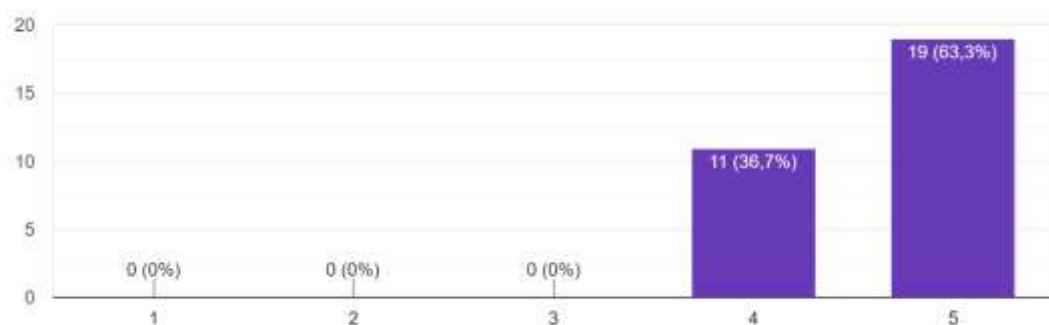
Seberapa mudah Chatbot Fasilkom-TI untuk diakses dari platform yang disediakan saat ini?

30 jawaban



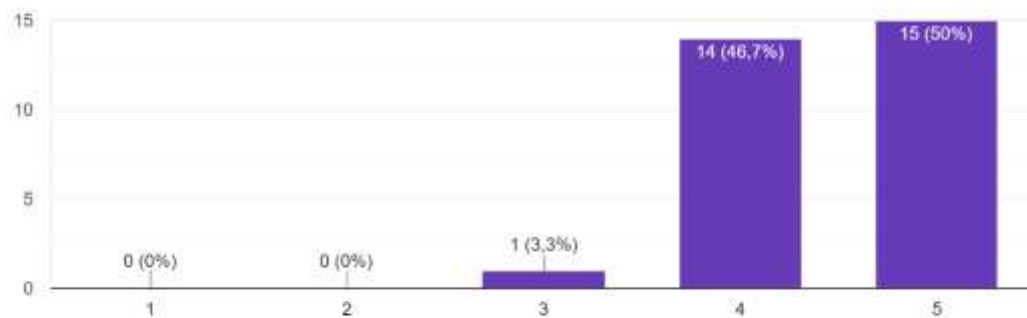
Seberapa mudah jawaban yang disediakan oleh Chatbot Fasilkom-TI untuk dimengerti?

30 jawaban



Gambar 4.7 Grafik Ringkasan Penilaian *User* dari google form

Berapa nilai kepuasan Anda saat menggunakan Chatbot Fasilkom-TI
30 jawaban



Gambar 4.8 Grafik Ringkasan Penilaian *User* dari google form lanjutan

Adapun perhitungan komponen-komponen *customer satisfaction* yang didapatkan dari hasil pengujian yaitu:

$$\begin{aligned}
 Accuracy &= \frac{\text{Jumlah hasil analisis benar}}{\text{Total data}} \times 100\% \\
 &= \frac{4(3)+9(4)+17(5)}{30(5)} \times 100\% \\
 &= \frac{128}{150} \times 100\% \\
 &= 88.67\%
 \end{aligned}$$

Penilaian kemudahan pengaksesan *chatbot* melalui *platform* yang disediakan saat ini dilakukan dengan meminta penilaian user menggunakan skala 1-5 dan dihasilkan:

$$\begin{aligned}
 &= \frac{6(4)+24(5)}{30} \\
 &= 4.80 / 5.0
 \end{aligned}$$

Penilaian kemudahan jawaban yang disediakan untuk dimengerti dilakukan dengan meminta penilaian user menggunakan skala 1-5 dan dihasilkan:

$$\begin{aligned}
 &= \frac{11(4)+19(5)}{30} \\
 &= 4.63 / 5.0
 \end{aligned}$$

Penilaian kepuasan terhadap *Chatbot* Fasilkom-TI yang saat ini dibangun dilakukan dengan meminta penilaian user menggunakan skala 1-5 dan dihasilkan:

$$\begin{aligned}
 &= \frac{1(3)+14(4)+15(5)}{30(5)} \\
 &= 4.67 / 5.0
 \end{aligned}$$

Evaluasi *chatbot* terkait kemampuan *chatbot* dalam memahami pertanyaan yang mengandung kata *typo* dilakukan dengan membandingkan jawaban yang dihasilkan *chatbot* yang menerapkan *spelling corrector* Peter Norvig dengan *chatbot* yang tidak menerapkan *spelling correcter*. Hasilnya dapat dilihat pada **Tabel 4.5**

Tabel 4.5 Perbandingan Chatbot yang Menerapkan Peter Norvig dan Tidak

No	Pertanyaan mengandung kata <i>typo</i>	Jawaban <i>chatbot</i> yang menerapkan Peter Norvig	Jawaban <i>chatbot</i> yang tidak menerapkan Peter Norvig
1	siapa kprodi ti?	Ketua Prodi TI adalah Dedy Arisandi, S.T., M.Kom. (NIP : 197908312009121002)	Maaf ya... Jawaban atas pertanyaan ini belum tersedia!
2	karender usu	Kamu dapat melihat kalender akademik pada https://www.usu.ac.id/id/kalender-akademik	Saat ini USU berakreditasi A
3	margon skripsi	1. Margin Awal Bab : atas-bawah-kiri-kanan : 50-25-38-25 (mm) 2. Margin Keseluruhan Naskah : atas-bawah-kiri-kanan : 30-25-38- 25 (mm)	Maaf ya... Jawaban atas pertanyaan ini belum tersedia!
4	kapan boleh menghubungi doing	Kamu dapat meghubungi pada hari dan jam kerja (Senin-Jumat pukul 08.00 – 16.00). Jangan menghubungi dosen dihari weekend atau hari libur nasional ya. Jangan lupa untuk menggunakan Bahasa yang baik dan sopan.	Batas pengisian KRS 12 Februari 2024 – 16 Februari 2024

5	eksum adalah	Executive summary atau sering disingkat exsum adalah suatu ringkasan yang berisi poin-poin penting terkait isi suatu dokumen. Pada exsum tugas akhir poin-poin yang perlu kamu isi adalah sebagai berikut Judul, Latar Belakang, Penelitian Terdahulu (minimal 5), Rumusan Masalah, Metodologi, dan Referensi secara ringkas dan jelas	Maaf ya... Jawaban atas pertanyaan ini belum tersedia!
6	jawdal semhas	Jadwal Seminar Hasil dan Daftar Penyaji Hasil (dalam bentuk undangan semhas pdf) akan dibagikan sekitar satu sampai dua hari sebelum hari pelaksanaan (biasanya hari Selasa dan Kamis atau bisa dilihat di elearning bagian jadwal seminar proposal, seminar hasil dan sidang meja hijau) dan dibagikan di grup angkatan atau grup tugas akhir perangkatan oleh bang Fadly Jadwal Semhas terdekat pada Selasa 26 Maret 2024	Maaf ya... Jawaban atas pertanyaan ini belum tersedia!
7	kalab komputer vission	Kepala Lab untuk bagian Computer Vision and Multimedia adalah bapak Jos Timanta Tarigan, S.Kom., M.Sc. (NIP : 198501262015041001)	Maaf ya... Jawaban atas pertanyaan ini belum tersedia!

BAB 5

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Sesuai dengan hasil yang diperoleh pada penelitian penerapan metode TF-IDF dan DNN pada *chatbot* layanan informasi akademik Fasilkom-TI USU yang telah dilakukan, dapat diperoleh beberapa kesimpulan berikut ini:

1. Pada penelitian *chatbot* yang dibangun dengan model TF-IDF dan *Deep Neural Network* terhadap data korpus yang mempunyai 252 tag informasi dengan 10-16 variasi pertanyaan dan 1-3 variasi jawaban dengan total data pertanyaan sebanyak 4.032 buah menghasilkan tingkat akurasi sebesar 92.44% dengan 150 *epoch*.
2. Dengan algoritma TF-IDF dan model DNN *chatbot* mampu memprediksi jawaban yang sesuai dengan pertanyaan *user*
3. Hasil *usability testing* yang dilakukan terhadap 30 orang *user* memberikan akurasi sebesar 88.67%
4. Algoritma Peter Norvig yang digunakan untuk *spelling corrector* dapat bekerja dengan baik untuk meng-*handle* 7 pertanyaan *user* yang mengandung *typo* yaitu *kprodi*, *karender*, *margon*, *menghubugni doing*, *eksum*, *jawdal*, dan *komputer vission*
5. Model *chatbot* yang dibangun dengan TF-IDF dan DNN belum mampu menangani pertanyaan *user* yang mengandung singkatan atau kata-kata gaul.

5.2. Saran

Berikut adalah beberapa saran yang bisa diterapkan untuk pengembangan selanjutnya ini dapat diterapkan:

1. Dataset pertanyaan dan jawaban yang digunakan sebagai data pelatihan perlu ditambahkan variasi pertanyaan jawaban. Semakin banyak variasi pertanyaan maka semakin baik pula akurasi sistem dalam menjawab pertanyaan *user*

khususnya untuk pertanyaan-pertanyaan yang mengandung kata-kata singkatan ataupun kata-kata gaul.

2. Menerapkan *chatbot* dengan metode *multilabel* dikarenakan sebuah pertanyaan bisa saja memiliki lebih dari satu kata kunci
3. Membuat cakupan lingkup informasi yang lebih luas, untuk penelitian ini *chatbot* diterapkan hanya dalam lingkup Fasilkom-TI untuk penelitian selanjutnya diharapkan *chatbot* dapat digunakan mahasiswa dari fakultas-fakultas lain.

DAFTAR PUSTAKA

- García AE, Zarazúa GM, Ayala MT, Araiza ER, Barrios AG. 2020. *Applications of Artificial Neural Networks in Greenhouse Technology and Overview for Smart Agriculture Development. Applied Sciences*
- Geron, A. 2022. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow Third Edition*. O'Reilly Media, Inc.
- Hakim, A. A., Erwin, A., Eng, K. I., Galinium, M., & Muliady, W. 2014. *Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach*. 2014 6th ICITEE
- Harumy T. H. B, Ginting, D. S., & Manik, F. Y. 2022. *Deep Neural Network: Implementasi dalam Berbagai Kasus dan Disertai dengan Struktur Coding*. Eureka Media Aksara.
- Kandpal, P., Jasnani, K., Raut, R., & Bhorge, S. 2020. *Contextual Chatbot for Healthcare Purposes (using Deep Learning)*. 2020 *Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*.
- Kavitha, B. R., & Murthy, C. R. 2019. *Chatbot for healthcare system using Artificial Intelligence*. *Int J Adv Res Ideas Innov Technol*, 5, 1304-1307.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. 2023. *Natural Language Processing: State of The Art, Current Trends and Challenges*. *Multimedia Tools and Applications*, 82(3)
- Lim, J., Jeong, S., Lim, S., Cho, H., Shim, J. Y., Hong, S. I., ... & Kim, J. 2022. *Development of Dye Exhaustion Behavior Prediction Model Using Deep Neural Network*. In *Computer aided chemical engineering (Vol. 49, pp. 1825-1830)*. Elsevier.
- Malekian, A., & Chitsaz, N. 2021. *Concepts, Procedures, and Applications Of Artificial Neural Network Models in Streamflow Forecasting*. In *Advances in Streamflow Forecasting (pp. 115-147)*. Elsevier.
- Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. 2019. *A Survey on Evaluation Methods for Chatbots*. *Proceeding of the 2019 7th International Conference on Information and Education Technology - ICIET*

- Putra, J. W. G. 2020. Pengenalan Konsep Pembelajaran Mesin dan Deep Learning. <https://wiragotama.github.io/resources/ebook/intro-to-ml-secured> (diakses 10 Januari 2024)
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. 2019. Language models are unsupervised multitask learners. OpenAI blog
- Ranoliya, B. R., Raghuwanshi, N., & Singh, S. 2017. *Chatbot for university related FAQs. In 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE.
- Reza, M. 2021. *Question Answering System Berbasis Natural Language Processing sebagai Media Informasi Perbankan*. Skripsi. Universitas Sumatera Utara
- Siburian, T.F. 2019. Pentingnya Sistem Informasi Manajemen dalam Pendidikan. *Jurnal Ilmiah Skylandsea*
- Singh, R., Paste, M., Shinde, N., Patel, H., & Mishra, N. 2018. *Chatbot using TensorFlow for small Businesses*. 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT).
- Soufyane, A., Abdelhakim, B.A., Ahmed, M.B. (2021). *An Intelligent Chatbot Using NLP and TF-IDF Algorithm for Text Understanding Applied to the Medical Field*. In: Ben Ahmed, M., Mellouli, S., Braganca, L., Anouar Abdelhakim, B., Bernadetta, K.A. (eds) *Emerging Trends in ICT for Sustainable Development. Advances in Science, Technology & Innovation*. Springer, Cham.
- Sulthan, M. B., Wahyudi, I., & Suhartini, L. 2021. Analisis Sentimen Pada Bencana Alam Menggunakan *Deep Neural Network* dan *Information Gain*. *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, 2(2), 65-71.
- Tarwani, K. & Edem, S. 2017. *Survey on Recurrent Neural Network in Natural Language Processing*. *International Journal of Engineering Trends and Technology (IJETT)*. Volume: 48
- Uysal, A. K., & Gunal, S. 2014. *The impact of preprocessing on text classification*. *Information processing & management*, 50(1), 104-112.
- Weizenbaum, J. 1966. *ELIZA—a computer program for the study of natural language communication between man and machine*. *Communications of the ACM*, 9(1), 36-45



**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET, DAN TEKNOLOGI**
UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

Jalan Universitas No. 9A Gedung A, Kampus USU Medan 20155, Telepon: (061) 821007
Laman: <http://Fasikomti.usu.ac.id>

KEPUTUSAN
DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI
NOMOR : 1563/UN5.2.14.D/SK/SPB/2024
DEKAN FAKULTAS ILMU KOMPUTER
DAN TEKNOLOGI INFORMASI UNIVERSITAS SUMATERA UTARA

- Membaca** : Surat Permohonan Mahasiswa Fasilkom-TI USU tanggal 17 Mei 2024 perihal permohonan ujian skripsi
Nama : ANGGITRI SIHOMBING
NIM : 181402128
Program Studi : Sarjana (S-1) Teknologi Informasi
Judul Skripsi : Penerapan Metode TF-IDF dan Deep Neural Network (DNN) Pada Chatbot Layanan Informasi Akademik Fasilkom-TI USU
- Memperhatikan** : Bahwa Mahasiswa tersebut telah memenuhi kewajiban untuk ikut dalam pelaksanaan Meja Hijau Skripsi Mahasiswa pada Program Studi Sarjana (S-1) Teknologi Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara TA 2023/2024.
- Menimbang** : Bahwa permohonan tersebut diatas dapat disetujui dan perlu ditetapkan dengan surat keputusan
- Mengingat** : 1. Undang-undang Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional.
2. Peraturan Pemerintah Nomor 17 tahun 2010 tentang pengelolaan dan penyelenggara pendidikan.
3. Keputusan Rektor USU Nomor 03/UN5.1.R/SK/SPB/2021 tentang Peraturan Akademik Program Sarjana Universitas Sumatera Utara.
4. Surat Keputusan Rektor USU Nomor 1876/UN5.1.R/SK/SDM/2021 tentang pengangkatan Dekan Fasilkom-TI USU Periode 2021-2026
- MEMUTUSKAN**
- Menetapkan** :
- Pertama** : Membentuk dan mengangkat Tim Penguji Skripsi mahasiswa sebagai berikut:
Ketua : Prof. Dr. Drs. Opim Salim Sitompul M.Sc
NIP: 196108171987011001
Sekretaris : Annisa Fadhillah Puhungan S.Kom, M.Kom
NIP: 199308092020012001
Anggota Penguji : Dr. Maricha Elvery S.TI, M.Kom
NIP: 199003272017062001
Anggota Penguji : Ivan Jaya S.Si, M.Kom.
NIP: 198407072015041001
Modemtor : -
Panitea : -
- Kedua** : Segala biaya yang diperlukan untuk pelaksanaan kegiatan ini dibebankan pada Dana Penerimaan Bukan Pajak (PNPB) Fasilkom-TI USU Tahun 2024.
- Ketiga** : Keputusan ini berlaku sejak tanggal ditetapkan dengan ketentuan bahwa segala sesuatunya akan diperbaiki sebagaimana mestinya apabila dikemudian hari terdapat kekeliruan dalam surat keputusan ini.

Tembusan :
1. Ketua Program Studi Sarjana (S-1) Teknologi Informasi
2. Yang bersangkutan
3. Arsip

Medan, 20 Mei 2024
Ditandatangani secara elektronik oleh:
Dekan



Maya Silvi Lydia
NIP 197401272002122001