

## BAB III

### METODOLOGI PENELITIAN

#### 3.1. Objek dan Subjek Penelitian

Subjek dari penelitian adalah mahasiswa dan layanan akademik Institut Teknologi Telkom Purwokerto. Sedangkan untuk objek masalah yang akan diamati yaitu mahasiswa yang kesulitan mendapatkan informasi layanan akademik khusus program Merdeka Belajar Kampus Merdeka di Institut Teknologi Telkom Purwokerto.

#### 3.2. Alat dan Bahan Penelitian

Dalam melakukan penelitian membuat *AI Chatbot* untuk layanan akademik di Institut Teknologi Telkom Purwokerto membutuhkan beberapa peralatan dan bahan untuk menunjang poses penelitian.

##### 3.2.1. Alat

Pembuatan *AI Chatbot* ini membutuhkan *hardware* (perangkat keras) dan *software* (perangkat lunak) dalam sebuah komputer atau laptop. Berikut spesifikasi untuk kebutuhan utama dalam penelitian ini yang berupa tabel yaitu:

Tabel 3. 1 *Hardware* (Perangkat Keras)

Komponen	Spesifikasi
Prosesor	Intel(R) Core(TM) i5-8265U CPU
RAM	8192MB RAM
HDD	1 TB
Sistem Operasi	Windows 11 Home Single Language 64-bit

Tabel 3. 2 *Software* (Perangkat Lunak)

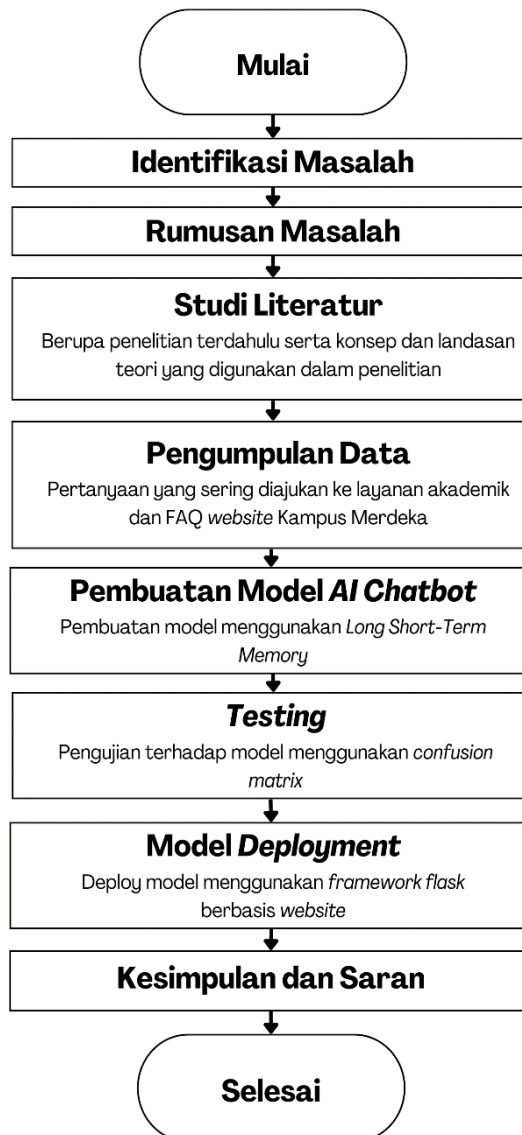
<b>Nama</b>	<b>Kegunaan</b>
Google Chrome	<i>Browser</i>
Python 3.8	Bahasa Pemrograman
Windows	Sistem Operasi
Anaconda	Pembuatan program
Github	<i>Version Control</i>
Jupyter Notebook	<i>Text Editor</i>
Tensorflow	<i>Library Deep Learning</i>
Visual Studio Code	<i>Text Editor</i>

### 3.2.2. Bahan

Bahan penelitian ini adalah data berupa pertanyaan umum tentang layanan akademik oleh mahasiswa di Institut Teknologi Telkom Puwokerto mengenai program Merdeka Belajar Kampus Merdeka dan *Frequently Asked Questions (FAQ) website* Kampus Merdeka. Dimana setiap pertanyaan itu akan disaring sesuai bagian pertanyaannya sehingga dapat dengan mudah dilakukan *preprocessing data*.

### 3.3. Diagram Alir Penelitian

Alir penelitian yang dilakukan pada penelitian ini dijelaskan pada Gambar 3.1. Langkah-langkah yang dilakukan dalam penelitian ini dimulai dengan identifikasi masalah. Setelah itu perumusan masalah dan cari tinjauan pustaka yang mendukung penelitian. Kemudian mengumpulkan data yang diperlukan, seperti pertanyaan yang sering diajukan. *Chatbot* dirancang setelah pertanyaan diperoleh. Model *chatbot* dibuat menggunakan *long short-term memory* dengan bahasa pemrograman *python 3.8*. Setelah model *chatbot* dibuat, kemudian akan dilakukan pengujian terhadap model untuk melihat apakah model *chatbot* mempunyai performa yang bagus. Model yang sudah teruji bagus selanjutnya akan dideploy menggunakan *framework flask* berbasis *website*. Atas dasar temuan penelitian, rekomendasi kemudian dibuat.



Gambar 3. 1 Diagram alir penelitian

### 3.3.1. Studi Literatur

Selama fase ini, tinjauan pustaka dilakukan untuk membantu memecahkan masalah dan mencapai tujuan penelitian. Fase ini mengkaji beberapa referensi dari beberapa buku, jurnal, laporan penelitian, artikel, skripsi, serta pustaka dari hasil penelusuran di media masa. Referensi utama yang mendukung penelitian ini adalah tinjauan Pustaka tentang *chatbot*, *natural language processing*, *deep learning*, *long short-term memory*, *artificial intelligence*.

### 3.3.2. Sumber Data

Data bersumber dari pertanyaan yang sering diajukan mahasiswa terkait layanan akademik khusus program Merdeka Belajar Kampus Merdeka dan *FAQ* dari *website* Kampus Merdeka. Tabel 3.3 menunjukkan pertanyaan yang sering diajukan mahasiswa dan respon dari pertanyaan tersebut yang didapat dari jawaban *FAQ* dari *website* kampus merdeka.

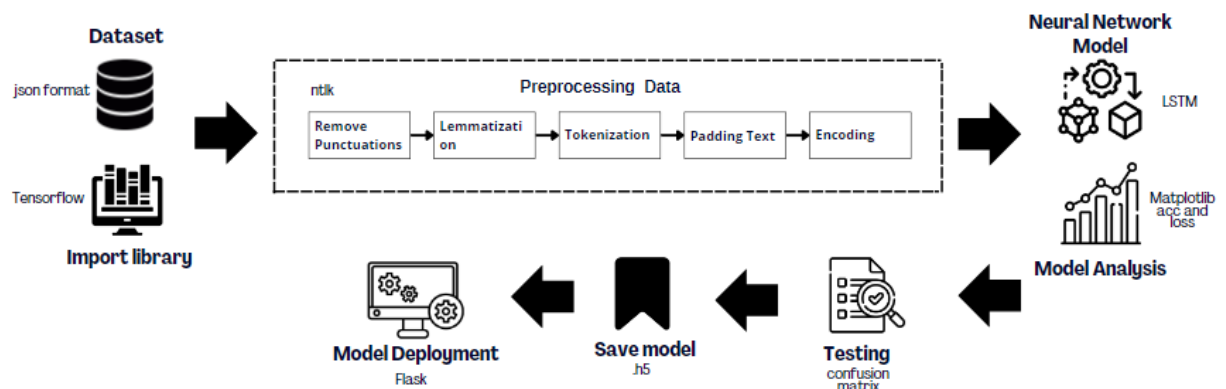
Tabel 3. 3 *Sample dataset AI Chatbot*

No	Pertanyaan	Respon
1	Apa itu Kampus Merdeka?	Kampus Merdeka merupakan kebijakan Menteri Pendidikan dan Kebudayaan Nadiem Makarim yang membebaskan mahasiswa untuk mengikuti kegiatan di luar program studinya selama 1 semester atau setara 20 sks dan paling lama 2 semester atau setara 40 sks.
2	Apa tujuan Kampus Merdeka?	Tujuan kebijakan Kampus Merdeka adalah memberikan kesempatan kepada mahasiswa untuk memilih mata kuliah yang akan mereka tempuh berdasarkan keinginan sendiri serta mendorong mahasiswa meningkatkan soft skills serta hard skills agar siap bersaing dalam dunia global.
3	Apa persyaratan mengikuti Kampus Merdeka?	Persyaratan mengikuti program Kampus Merdeka, yaitu merupakan mahasiswa yang berasal dari Program Studi terakreditasi dan terdaftar aktif pada PDDikti.
4	Apa saja program Kampus Merdeka?	Terdapat program Kampus Merdeka dari Kemendikbud, Kemensos, serta BUMN
5	Apa itu Magang Bersertifikat?	Salah satu program Kampus Merdeka Kemendikbud yang bertujuan untuk memberikan kesempatan kepada mahasiswa mendapatkan pengalaman kerja di industri/dunia profesi nyata untuk mendapatkan hard skills maupun soft skills.

No	Pertanyaan	Respon
6	Apa saja persyaratan Magang Bersertifikat?	Tidak ada persyaratan khusus. Seluruh mahasiswa dari perguruan tinggi dan program studi apapun dapat mengikuti program sesuai persyaratan setiap mitra
7	Apa saja manfaat apabila mengikuti Magang Bersertifikat?	Manfaat yang didapatkan, yaitu konversi hingga 20 sks, dibimbing mentor staf profesional secara full-time, sertifikasi sesuai kinerja, dan uang saku serta biaya hidup selama magang
8	Apa itu program Bangkit?	Program Bangkit adalah program kesiapan karier yang didesain oleh Google untuk memberikan mahasiswa Indonesia paparan langsung dengan praktisi industri, serta mempersiapkan mahasiswa dengan keterampilan yang relevan untuk karier sukses di perusahaan teknologi terkemuka.
9	Apa saja jalur pembelajaran program Bangkit?	Bangkit menawarkan 3 pilihan jalur pembelajaran yaitu machine learning, mobile development, dan cloud computing. Tidak hanya keterampilan teknologi, Bangkit juga menyediakan pengembangan Bahasa Inggris dan soft skill yang penting untuk persiapan karir.

### 3.4. Tahapan Pembuatan *AI Chatbot*

Dalam pembuatan *AI Chatbot* menggunakan *Long Short Term Memory* terdapat beberapa tahapan seperti yang ditunjukkan pada Gambar 3.2 yang dikembangkan dari penelitian Ming-Hsiang Su *et al* [11].



Gambar 3. 2 Tahapan Pembuatan *AI Chatbot*

### 3.4.1. *Dataset*

*Dataset* berupa *file* dengan format *.json* yang berisi pertanyaan tentang layanan akademik mengenai program Merdeka Belajar Kampus Merdeka (MBKM). Terdapat 605 pasang pertanyaan dan respon dalam 60 kelas yang dikategorikan secara manual sesuai *FAQ* dari *website* Kampus Merdeka.

### 3.4.2. *Library*

*Library* yang dibutuhkan dalam pembuatan *chatbot* menggunakan *Long Shor Term Memory (LSTM)* seperti *NumPy* untuk komputasi matematika, *Matplotlib* untuk visualiasasi model data, *NLTK* untuk pengolahan teks, *Pandas* untuk membaca data, dan *Tensorflow* untuk model pada data menggunakan algoritme *LSTM*.

### 3.4.3. *Preprocessing*

Preprocessing data untuk manipulasi atau penghapusan data sebelum digunakan. Tahapan *preprocessing* diantaranya:

#### 3.4.3.1. *Remove Punctuations*

Pada tahap prapemrosesan awal untuk data teks, karakter khusus seperti tanda seru (!), tanda koma (,), tanda titik sebagai berhenti (.), tanda tanya (?) dan lainnya dihilangkan [37]. Tahapan ini akan mempermudah

pemrosesan data yang akan diolah. Tabel 3.4 menunjukkan perbedaan data setelah dilakukan penghapusan punctuation.

Tabel 3. 4 Proses *Remove Punctuations*

Sebelum	Sesudah
"Apa persyaratan mengikuti Kampus Merdeka?"	apa persyaratan mengikuti kampus merdeka

### 3.4.3.2. *Lemmatization*

*Lemmatization*, juga dikenal sebagai "lematisasi" adalah proses menghilangkan hanya akhiran infleksional dari kata-kata dan mengembalikan bentuk kamus (kata-kata yang ada dalam kamus) dari sebuah kata yang dikenal sebagai "*lemma*" [37]. Tabel 3.5 menunjukkan bagaimana mengidentifikasi bentuk kamus kata yang diubah berdasarkan makna yang dimaksudkan.

Tabel 3. 5 Proses *Lemmatization*

Sebelum	Sesudah
Apa persyaratan mengikuti Kampus Merdeka?	Apa syarat ikut Kampus Merdeka?

### 3.4.3.3. *Tokenization*

Menetapkan string karakter ke unit dokumen tertentu adalah proses tokenization. Tokenisasi adalah proses segmentasi kalimat menjadi "Token" dan menghilangkan komponen tertentu, termasuk tanda baca [38]. Tabel 3.6 menunjukkan bagaimana memecah kalimat menjadi bagian token.

Tabel 3. 6 Proses *Tokenization*

Kalimat	Hasil <i>Tokenization</i>
Apa persyaratan mengikuti Kampus Merdeka?	"Apa" "persyaratan" "mengikuti" "Kampus" "Merdeka"

Setiap kata pada kalimat yang sudah dipecah kedalam bentuk token, kata tersebut mempunyai nilai token unik numerik yang dapat dilihat pada Tabel 3.7

Tabel 3. 7 Sampel Indeks Setiap Kata pada Dataset

<i>Word</i>	<i>Index</i>	<i>Word</i>	<i>Index</i>
'program'	1	'dapat'	26
'apa'	2	'syarat'	27
'yang'	3	'dalam'	28
'mengikuti'	4	'saya'	29
'saja'	5	'di'	30
'magang'	6	'dengan'	31
'studi'	7	'pertukaran'	32
'independen'	8	'dipenuhi'	33
'kampus'	9	'mendaftar'	34
'dari'	10	'keuntungan'	35
'mahasiswa'	11	'periode'	36
'merdeka'	12	'bagaimana'	37
'manfaat'	13	'cara'	38
'kknt'	14	'siapa'	39
'untuk'	15	'defenisi'	40
'iisma'	16	'bisa'	41
'pmmb'	17	'mitra'	42
'apakah'	18	'pengertian'	43
'persyaratan'	19	'didapatkan'	44
'harus'	20	'itu'	45
'mengajar'	21	'mana'	46
'pejuang'	22	'tujuan'	47
'muda'	23	'bersertifikat'	48
'bangkit'	24	'lama'	49



<i>Word</i>	<i>Index</i>	<i>Word</i>	<i>Index</i>
'berapa'	25	'lowongan'	50

#### 3.4.3.4. *Padding Text*

Untuk membuat setiap urutan sama panjang, digunakan *padding*. Setiap *sequence* pada *padding* dibuat sama panjang dengan menambahkan nilai 0 melalui *suffix* atau *prefix* hingga mencapai panjang maksimum *sequence* [39]. Selain itu, padding dapat mempersingkat urutan ke panjang maksimumnya. Tabel 3.8 menunjukkan perbedaan setelah dilakukan *padding text*. Kalimat yang sudah ditokenisasi mempunyai panjang *sequence* yang berbeda. Jika memasukan *sequence* tersebut kedalam model, maka akan terjadi *error*. Menerapkan *padding* pada *sequence* dengan menambahkan nilai 0 diawal untuk membawa *sequence* yang pendek memiliki panjang yang sama sesuai dengan panjang maksimal *sequence*.

Tabel 3. 8 Proses *Padding Text*

<b>Sebelum</b>	<b>Sesudah</b>
[1 2 3]	[0 0 0 3 4 2]
[1 3 2 4 7 5]	[1 3 2 4 7 5]

#### 3.4.3.5. *Encoding*

*Encoding* adalah penerjemahan data kategorikal, seperti karakter atau teks, menjadi data numerik atau bilangan bulat sesuai dengan label data yang diterapkan. *Encoding* mengubah teks dalam kolom *tag* data dalam operasi ini menjadi data numerik menggunakan biner komputer, yang terdiri dari angka 0 dan 1 [39]. Setiap label akan di urutkan berdasarkan abjad dan diberikan token/angka pada masing-masing label mulai dari 0 sampai jumlah keseluruhan label. Tabel 3.9 menunjukkan sampel label yang sudah dilakukan encoding

Tabel 3. 9 Sampel Label *Encoding*

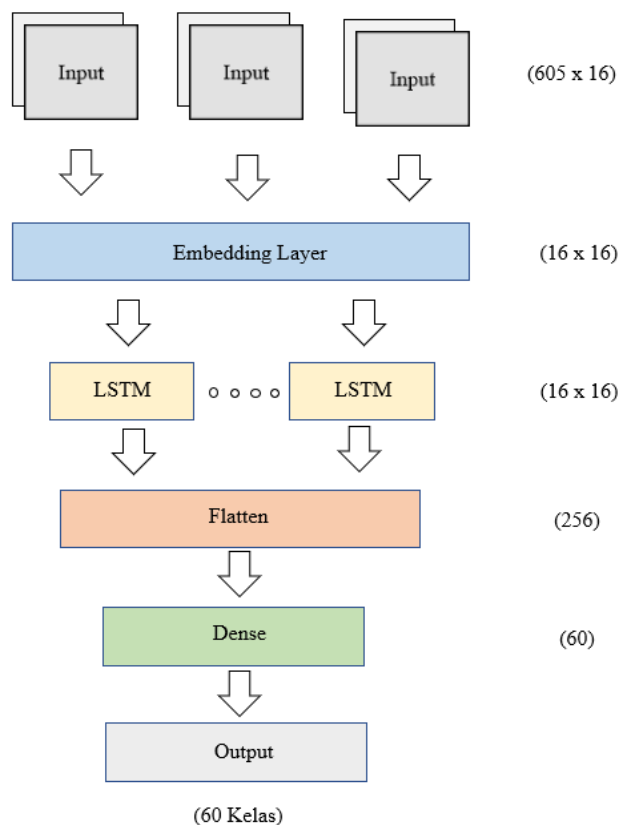
No	Label	Encoding
1	ITTPBot	0
2	biaya_studi_independen	1
3	daftar_lowongan_magang	2
4	gaji_magang	3
5	goodbye	4
6	greeting	5
7	jangka_waktu_kknt	6
8	jenis_kampus_merdeka	7
9	jurusan_magang	8
10	jurusan_studi_independen	9
11	lokasi_kknt	10
12	lowongan_magang	11
13	manfaat_IISMA	12
14	manfaat_PMMB	13
15	manfaat_bangkit	14
16	manfaat_kampus_mengajar	15
17	manfaat_kknt	16
18	manfaat_magang	17
19	manfaat_pejuang_muda	18
20	manfaat_pertukaran_mahasiswa	19

#### 3.4.4. Long Short Term Memory Model

Pada tahap ini dilakukan perancangan model algoritme *LSTM* untuk mengklasifikasikan pertanyaan berdasarkan kelasnya. Dataset yang sudah dilakukan *preprocessing* data untuk menyiapkan data yang terstruktur sehingga dapat diolah menggunakan algoritme *LSTM*. Perancangan arsitektur *LSTM* dilakukan dengan *training* model menggunakan 4, 8, 16, 32, 64, dan 128 unit *LSTM*.

Pada Gambar 3.3 menunjukkan arsitektur model *LSTM* dengan menggunakan unit *LSTM* 16. Terdapat 605 data *input* dengan panjang *sequence* 16 kemudian digunakan *embedding layer* untuk mengubah *input*

yang berukuran 2 dimensi menjadi 3 dimensi karena *LSTM layer* membutuhkan *input* berukuran 3 dimensi. *Flatten layer* digunakan untuk mengubah *input* yang memiliki banyak dimensi menjadi satu dimensi yang akan digunakan pada *dense layer*. Terdapat 60 kelas sebagai *output* dari klasifikasi dengan menggunakan *softmax* karena merupakan klasifikasi *multiclass*.



Gambar 3. 3 Arsitektur Model *LSTM*

Setelah dilakukan *training* dan *testing* terhadap model, tahap selanjutnya adalah menguji performa dari model menggunakan *confusion matrix*. Hasil pengujian model akan dianalisis apakah mempunyai kinerja yang baik atau tidak.

#### 3.4.5. *Model Analysis*

Setelah menjalankan pelatihan model dengan algoritme *LSTM* serta telah mengetahui hasil akurasi pada step terakhir. Maka, tahapan selanjutnya adalah menganalisa model dengan visualisasi plot akurasi dan loss untuk melihat hasil akurasi dari algoritme pelatihan model *LSTM* tersebut [40].

#### 3.4.6. *Testing*

Setelah mengetahui hasil dari akurasi dan loss pada model yang telah ditetapkan dengan algoritme *LSTM*. Maka, tahapan selanjutnya adalah menguji atau testing model pada *chatbot* yang telah dilatih sebelumnya menggunakan *multi-confusion matrix* [41] dengan parameter *y\_test* dan *prediction*. *Multi-confusion matrix* digunakan untuk mengukur performa klasifikasi machine learning dimana *output* berupa multi kelas.

#### 3.4.7. *Save Model*

Setelah pengujian *Chatbot* telah disesuaikan dengan kalimat dan jawabannya. Maka, model *chatbot* bisa disimpan dengan format *.h5* atau *.pkl* (*pickle*) untuk penggunaan aplikasi *AI Chatbot* dengan *website* atau sistem Android [42]. Penyimpanan file model bisa langsung secara *transient* atau bisa taruh di *Google Drive*.

#### 3.4.8. *Model Deployment*

Gambar 3.4 menunjukkan tampilan awal dari *chatbot* yang sudah di *deploy* berbasis *website*. Proses *deployment* dibangun menggunakan *framework flask* menggunakan bahasa pemrograman python beserta pembuatan *website* menggunakan HTML, CSS, PHP, dan JavaScript.

Bahasa pemrograman *PHP* diperlukan selama tahap pengembangan karena digunakan untuk aplikasi *web*. *Application Programming Interface* (*API*) diperlukan untuk menghubungkan PHP ke *Flask* yang memungkinkan komunikasi melalui permintaan HTTP. *API* ini memungkinkan PHP untuk

mengirimkan permintaan ke titik akhir Flask untuk mengambil dan mengirim data. Permintaan data dapat dikirim menggunakan metode *GET* atau *POST* dengan memberikan informasi yang diperlukan dalam permintaan.



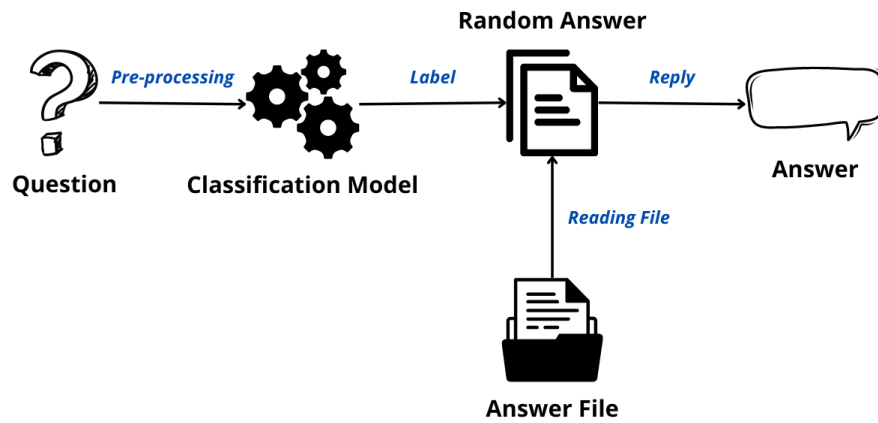
Gambar 3. 4 Tampilan Awal *Chatbot*

### 3.5. Skema Sistem *Chatbot*

Gambar 3.5 menunjukkan bagaimana sistem *chatbot* bekerja. Untuk menjawab pertanyaan pengguna, *chatbot* mengubah teks bahasa Indonesia menjadi daftar bilangan bulat yang kemudian digunakan sebagai *input* dalam model klasifikasi. Bilangan bulat akan dihasilkan oleh model klasifikasi yang diberi label dengan kumpulan kueri yang digunakan untuk membangun model. Sistem akan dapat memberikan tanggapan yang benar dengan mengidentifikasi serangkaian pertanyaan yang tepat.

Model klasifikasi bekerja dengan menerima *input* berupa data *sequence* berukuran 2 dimensi hasil dari *preprocessing* data. Pada *embedding layer*, inputan tersebut dirubah menjadi vektor berukuran 3 dimensi sebagai inputan dari *LSTM layer*. Kemudian *flatten* akan mengubah vektor dari *output lstm* menjadi 1 dimensi yang selanjutnya diteruskan ke *dense layer* untuk menghasilkan probabilitas untuk setiap kelas. Hasil probabilitas terbesar pada indeks array akan menjadi output

dari model yang mana indeks tersebut merepresentasikan label yang sudah dilakukan *encoding*.



Gambar 3.5 Chatbot Overview [15]