

IMPLEMENTASI LONG SHORT-TERM MEMORY (LSTM)  
DAN INTEGER SEQUENCE MATCHING PADA SISTEM  
CHATBOT INFORMASI SAHAM INDONESIA

SKRIPSI

MICHAEL

191402059



PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2023

IMPLEMENTASI LONG SHORT-TERM MEMORY (LSTM)  
DAN INTEGER SEQUENCE MATCHING PADA SISTEM  
CHATBOT INFORMASI SAHAM INDONESIA

SKRIPSI

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

MICHAEL

191402059



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

MEDAN

2023

## PERSETUJUAN

Judul : IMPLEMENTASI LONG SHORT-TERM MEMORY (LSTM) DAN INTEGER SEQUENCE MATCHING PADA SISTEM CHATBOT INFORMASI SAHAM INDONESIA

Kategori : SKRIPSI

Nama : MICHAEL

Nomor Induk Mahasiswa : 191402059

Program Studi : SI TEKNOLOGI INFORMASI

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA

Kesepakatan Pembimbing :

Melalui, 10 Januari 2024

Pembimbing 1

Pembimbing 1

Prof. Dr. Drs. Opim Salim Sitompul M.Sc  
NIP. 196108171987011001

Dr. Erna Budhiarti Nababan M.IT  
NIP. 196210262017042001

Diketahui/disetujui oleh

Program Studi SI Teknologi Informasi



Dedy Arisandi, ST., M.Kom.  
NIP. 197908312009121002

**PERNYATAAN****IMPLEMENTASI LONG SHORT-TERMS MEMORYS (LSTM)  
DAN INTEGER SEQUENCE MATCHING PADA SISTEM  
CHATBOT INFORMASI SAHAM INDONESIA****SKRIPSI**

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

Medan, 22 Januari 2024

A handwritten signature in black ink, appearing to read 'Michael', with a horizontal line drawn underneath the name.

Michael

191402059

## UCAPAN TERIMA KASIH

Puji dan syukur penulis haturkan kepada Tuhan Yang Maha Esa yang dengan rahmat dan izin-Nya, memungkinkan penulis menuntaskan skripsi ini sebagai persyaratan meraih gelar Sarjana Komputer, di Program Studi S1 Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.

Skripsi ini kini penulis persembahkan kepada orang tua tercinta, Bapak Sjukiman Sjamsul dan Ibu Ely Ringo Djati, S.E, yang senantiasa menyertai dengan doa, semangat, dan kasih sayang. Semoga Tuhan Yang Maha Esa senantiasa melimpahkan kebahagiaan, kesehatan, dan umur yang panjang kepada mereka berdua. Terima kasih juga untuk adik penulis, Michelle, yang selalu memberikan dukungan dan semangat dalam perjalanan penulisan skripsi ini.

Penulis sadar betul bahwa perjalanan penelitian ini tidak akan mungkin terwujud tanpa bantuan dan dukungan berbagai pihak. Dengan penuh rendah hati, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Muryanto Amin, S.Sos., M.Si, sebagai Rektor Universitas Sumatera Utara.
2. Bapak Dr. Maya Silvi Lydia, B.Sc., M.Sc., yang menjabat sebagai Dekan Fasilkom-TI USU.
3. Bapak Dedy Arisandi, ST., M.Kom., selaku Ketua Program Studi S1 Teknologi Informasi di Universitas Sumatera Utara.
4. Bapak Ivan Jaya, S.Si., M.Kom., sebagai Sekretaris Program Studi S1 Teknologi Informasi di Universitas Sumatera Utara.
5. Ibu Dr. Erna Budhiarti Nababan M.IT, selaku Dosen Pembimbing I, yang telah rela meluangkan waktu, memberikan pemikiran, serta memberikan saran dan kritiknya untuk penulis dalam menyelesaikan skripsi ini.
6. Bapak Prof. Dr. Drs. Otim Salim Sitompul, M.Sc, sebagai Dosen Pembimbing II, yang juga senantiasa meluangkan waktu, memberikan pemikiran, saran, dan kritiknya untuk penulis dalam menyelesaikan skripsi ini.

7. Ibu Fanindia Purnamasari S.TL., M.IT, selaku Dosen Penguji I, yang telah meluangkan waktu, memberikan pemikiran, serta saran dan kritiknya untuk penulis dalam menyelesaikan skripsi ini.
8. Ibu Ade Sarah Huzaifah S.Kom., M.Kom, selaku Dosen Penguji II, yang telah meluangkan waktu, memberikan pemikiran, serta saran dan kritiknya kepada penulis dalam menyelesaikan skripsi ini.
9. Seluruh dosen dan staf pegawai di Program Studi Teknologi Informasi dan Fasilkom-TI USU, yang telah membantu kelancaran proses administrasi selama masa perkuliahan.
10. Keluarga besar penulis, yang tak henti-hentinya memberikan dukungan dan semangat.
11. Jason, Kevin Ricardo, dan Kenndy Susio, sebagai sahabat layaknya saudara penulis yang senantiasa mendukung penulis dalam berbagai kesempatan.
12. Serta semua pihak yang terlibat secara langsung maupun tidak langsung, yang tidak dapat kami sebutkan satu per satu, namun turut serta membantu menyelesaikan skripsi ini.

Semoga Tuhan Yang Maha Esa melimpahkan berkah kepada semua pihak yang telah memberikan bantuan, perhatian, dan dukungan kepada penulis dalam menyelesaikan skripsi ini. Penulis sadar akan adanya kekurangan dalam skripsi ini dan dengan rendah hati, penulis mengharapkan saran dan kritik yang membangun dari semua pihak untuk meningkatkan kualitas skripsi ini. Akhir kata, terima kasih atas segala dukungan yang diberikan.

Medan, 22 Januari 2024



Penulis

## ABSTRAK

Persentase jumlah masyarakat Indonesia yang berinvestasi saham di pasar modal Indonesia masih terhitung rendah dibandingkan dengan negara lain. Hal ini diakibatkan kurangnya Informasi terkait saham di kalangan masyarakat Indonesia. Masalah ini menjadi alasan diperlukannya sarana informasi saham Indonesia yang mudah diakses salah satunya adalah chatbot. Selain dapat diakses kapanpun dan dimanapun, chatbot juga merupakan sarana tanya jawab dua arah sehingga dapat lebih mempermudah penyaluran informasi. Penelitian implementasi *chatbot* saham dengan metode *Long Short-Term Memory* (LSTM) dan *Integer Sequence Matching* dapat membantu permasalahan tersebut dengan memberikan jawaban atas pertanyaan saham yang diberikan pengguna serta dengan fitur pendukung informasi saham Indonesia lainnya seperti modul belajar saham, informasi jenis sektor saham, informasi jenis sektor dan sahamnya, dan harga saham terkini dapat menarik minat investasi saham masyarakat Indonesia. Data yang digunakan didapatkan dari sekuritas bersertifikasi oleh OJK sehingga jawaban memiliki kualitas yang baik. Pengujian sistem menunjukkan bahwa *chatbot* yang dihasilkan memiliki tingkat akurasi sebesar 90% dengan rata-rata waktu respon sekitar 0.89 detik. Penilaian chatbot oleh masyarakat umum meliputi 30 orang yang belum pernah berinvestasi saham Indonesia, 30 orang yang pernah berinvestasi saham Indonesia, dan 4 pakar saham Indonesia mendapatkan penilaian yang baik dari sisi kemudahan pemahaman dan kebergunaan. Chatbot juga dapat meningkatkan minat 24 dari 30 pengguna yang belum pernah berinvestasi saham Indonesia untuk mencoba berinvestasi saham di bursa saham Indonesia.

Kata kunci : *Chatbot*, Investasi, Saham, *Long Short-Term Memory*, *Integer Sequence Matching*

# IMPLEMENTATION OF LONG SHORT-TERMS MEMORYS (LSTM) AND INTEGER SEQUENCE MATCHING IN INDONESIAN STOCK INFORMATION CHATBOT SYSTEM

## ABSTRACT

The percentage of Indonesians investing in the Indonesian stock market is still relatively low compared to other countries. This is due to a lack of information about stocks among the Indonesian population. This problem is a key reason for the need for accessible stock information tools, one of which is a chatbot. Besides being accessible anytime and anywhere, a chatbot also serves as a two-way question-and-answer platform, making the distribution of information much easier. The development of a stock chatbot using *Long Short-Term Memory* (LSTM) methods and *Integer Sequence Matching* can help address this issue by providing answers to user queries about stocks. It also includes supporting features such as a stock learning module, information on stock sectors, information on stock types per sector, and real-time stock prices, all of which can attract the interest of Indonesian investors. The data used is obtained from OJK-certified securities to ensure high-quality answers. Evaluation shows that the resulting chatbot has an accuracy rate of 90% with an average response time of around 0.89 seconds. The chatbot is positively assessed by the general public, including 30 persons who have never invested in Indonesian stocks, 30 persons who have invested in Indonesian stocks, and 4 Indonesian stock experts. It is considered to be easy to understand and bring benefit. It is assessed to enhance interest of 24 out of 30 persons who have never invested in Indonesian stocks to invest stock in Indonesia stock market.

Keywords: *Chatbot*, Investment, Stocks, *Long Short-Term Memory*, *Integer Sequence Matching*



## DAFTAR ISI

	Hal.
PERSETUJUAN .....	iii
PERNYATAAN .....	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK .....	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xiii
DAFTAR RUMUS .....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Batasan Masalah .....	2
1.4. Tujuan Penelitian .....	3
1.5. Manfaat Penelitian .....	3
1.6. Sistematika Penulisan .....	3
BAB II LANDASAN TEORI.....	5
2.1. Saham.....	5
2.2. Natural Language Processing.....	6
2.3. Chatbot .....	8
2.4. Algoritma Long Short-Term Memory (LSTM) .....	8
2.5. Application Programming Interface (API) .....	11
2.6. Penelitian Terdahulu .....	11
BAB III METODOLOGI PENELITIAN .....	15
3.1. Dataset.....	15
3.2. Arsitektur Umum .....	17
3.3. Activity Diagram Chatbot Informasi Saham Indonesia.....	31
3.4. Perancangan Interface Sistem .....	33

BAB IV IMPLEMENTASI DAN PENGUJIAN.....	35
4.1. Implementasi Sistem.....	35
4.1.1. Spesifikasi Perangkat Keras (Hardware) .....	35
4.1.2. Spesifikasi Perangkat Lunak (Software).....	35
4.2. Implementasi Interface.....	36
4.3. Training Model LSTM.....	44
4.4. Pengujian Model LSTM .....	47
4.5. Pengujian Sistem.....	54
BAB V KESIMPULAN DAN SARAN .....	67
5.1. Kesimpulan .....	67
5.2. Saran .....	68
DAFTAR PUSTAKA .....	69

## DAFTAR TABEL

	Hal.
<b>Tabel 2.1.</b> Penelitian Terdahulu .....	13
<b>Tabel 2.1.</b> Penelitian Terdahulu (Lanjutan).....	14
<b>Tabel 3.1.</b> Contoh Dataset Yang Digunakan.....	15
<b>Tabel 3.2.</b> Contoh <i>Case Folding</i> .....	19
<b>Tabel 3.3.</b> Contoh <i>Punctual Removal</i> .....	20
<b>Tabel 3.4.</b> Contoh <i>Normalization kata baku</i> .....	20
<b>Tabel 3.5.</b> Contoh Stopword Removal .....	20
<b>Tabel 3.6.</b> Contoh <i>Stemming</i> .....	21
<b>Tabel 3.7.</b> Contoh <i>Label Encoding</i> .....	21
<b>Tabel 3.8.</b> Pembagian Data Training, Data Validation, Data Testing & Data Corpus	22
<b>Tabel 3.9.</b> Contoh <i>Tokenisasi</i> .....	23
<b>Tabel 3.10.</b> Contoh <i>Model Vocabulary Word Index</i> .....	23
<b>Tabel 3.11.</b> Contoh <i>Model Integer Sequence</i> .....	24
<b>Tabel 3.12.</b> Contoh <i>Padding</i> .....	24
<b>Tabel 3.13.</b> Contoh <i>Tokenisasi</i> .....	25
<b>Tabel 3.14.</b> Contoh <i>Corpus Vocabulary Word Index</i> .....	25
<b>Tabel 3.15.</b> Contoh <i>Corpus Integer Sequence</i> .....	26
<b>Tabel 3.16.</b> Contoh <i>Integer Sequence Matching</i> .....	29
<b>Tabel 3.17.</b> Contoh <i>Integer Sequence Matching</i> kecocokan tertinggi yang sama .....	30
<b>Tabel 4.1.</b> Tabel Hasil Percobaan Parameter Tuning .....	46
<b>Tabel 4.2.</b> Tabel Hasil Percobaan Patience .....	46
<b>Tabel 4.3.</b> Tabel Pengujian Model Terhadap Data Testing.....	47
<b>Tabel 4.3.</b> Tabel Pengujian Model Terhadap Data Testing (Lanjutan).....	48
<b>Tabel 4.3.</b> Tabel Pengujian Model Terhadap Data Testing (Lanjutan).....	49
<b>Tabel 4.3.</b> Tabel Pengujian Model Terhadap Data Testing (Lanjutan).....	50
<b>Tabel 4.3.</b> Tabel Pengujian Model Terhadap Data Testing (Lanjutan).....	51
<b>Tabel 4.3.</b> Tabel Pengujian Model Terhadap Data Testing (Lanjutan).....	52
<b>Tabel 4.3.</b> Tabel Pengujian Model Terhadap Data Testing (Lanjutan).....	53

<b>Tabel 4.4.</b> Tabel Precision, Recall dan F1-Score Prediksi <i>Data Testing</i> .....	54
<b>Tabel 4.5.</b> Tabel Pengujian Sistem.....	54
<b>Tabel 4.5.</b> Tabel Pengujian Sistem (Lanjutan) .....	55

## DAFTAR GAMBAR

	Hal.
<b>Gambar 2.1.</b> Gambaran Umum Cara Kerja Chatbot .....	8
<b>Gambar 2.2.</b> Arsitektur <i>Long Short-Term Memory</i> .....	9
<b>Gambar 3.1.</b> Arsitektur Umum .....	18
<b>Gambar 3.2.</b> Arsitektur Model LSTM .....	28
<b>Gambar 3.3.</b> Activity Diagram Chatbot Informasi Saham Indonesia .....	32
<b>Gambar 3.4.</b> Interface Sistem .....	33
<b>Gambar 4.1.</b> Tampilan Halaman Profil .....	36
<b>Gambar 4.2.</b> Tampilan Halaman Pembuka .....	37
<b>Gambar 4.3.</b> Tampilan Halaman Menu Utama .....	38
<b>Gambar 4.4.</b> Tampilan Modul Saham .....	39
<b>Gambar 4.5.</b> Tampilan Daftar Sektor Dan Sahamnya .....	40
<b>Gambar 4.6.</b> Tampilan Informasi Perusahaan .....	41
<b>Gambar 4.7.</b> Tampilan Harga Saham .....	42
<b>Gambar 4.8.</b> Tampilan Tanya Jawab Saham .....	43
<b>Gambar 4.9.</b> Plot Model LSTM Yang Digunakan .....	44
<b>Gambar 4.10.</b> Histori Performa Akurasi Pelatihan Model .....	45
<b>Gambar 4.11.</b> Histori Performa Loss Pelatihan Model .....	45
<b>Gambar 4.12.</b> Confusion Matrix Data Testing .....	53
<b>Gambar 4.13.</b> Hasil Penilaian Fitur Tambahan Oleh Pihak Yang Belum Pernah Berinvestasi Saham .....	59
<b>Gambar 4.14.</b> Hasil Penilaian Fitur Tanya Jawab Oleh Pihak Yang Belum Pernah Berinvestasi Saham .....	61
<b>Gambar 4.15.</b> Hasil Penilaian Fitur Tambahan Oleh Pihak Yang Pernah Berinvestasi Saham .....	62
<b>Gambar 4.16.</b> Hasil Penilaian Fitur Tanya Jawab Oleh Pihak Yang Pernah Berinvestasi Saham .....	63
<b>Gambar 4.17.</b> Hasil Penilaian Fitur Tambahan Oleh Pakar Saham .....	64
<b>Gambar 4.18.</b> Hasil Penilaian Fitur Tanya Jawab Oleh Pakar Saham .....	65
<b>Gambar 4.19.</b> Timbulnya Minat Berinvestasi Saham .....	66

## DAFTAR RUMUS

	Hal.
<b>Rumus 2.1.</b> Rumus <i>Forget Gate Long Short-Term Memory</i> .....	9
<b>Rumus 2.2.</b> Rumus <i>Input Gate Long Short-Term Memory</i> .....	10
<b>Rumus 2.3.</b> Rumus <i>Cell State Long Short-Term Memory</i> .....	10
<b>Rumus 2.4.</b> Rumus <i>Updated Cell State Long Short-Term Memory</i> .....	10
<b>Rumus 2.5.</b> Rumus <i>Sigmoid Output Long Short-Term Memory</i> .....	11
<b>Rumus 2.6.</b> Rumus <i>Hidden State Long Short-Term Memory</i> .....	11
<b>Rumus 3.1.</b> Rumus <i>Kecocokan Integer Sequence</i> .....	29
<b>Rumus 4.1.</b> Rumus <i>Precision Confusion Matrix</i> .....	53
<b>Rumus 4.2.</b> Rumus <i>Recall Confusion Matrix</i> .....	54
<b>Rumus 4.3.</b> Rumus <i>F1 Score Confusion Matrix</i> .....	54
<b>Rumus 4.4.</b> Rumus <i>Accuracy Confusion Matrix</i> .....	54

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Kemajuan sektor teknologi di Indonesia menawarkan kemudahan dalam berinvestasi saham baik dari kemudahan bertransaksi sampai kemudahan mendapatkan informasi keadaan perusahaan yang sudah go public (Cahya & Kusuma, 2019). Akan tetapi, dari hasil penelitian oleh Lubis & Kusuma (2022) menunjukkan bahwa pada tahun 2022 masyarakat yang menjadi investor saham hanya 10.3 juta masyarakat saja. Dan dari hasil statistik oleh PT Kustadian Sentral Efek Indonesia (KSEI) yang merupakan sebuah Lembaga Penyimpanan serta Penyelesaian (LPP) di Bursa Efek Indonesia, hingga Agustus 2023 masyarakat Indonesia yang menjadi investor saham di Bursa Efek Indonesia meningkat menjadi 11.5 juta masyarakat dimana jumlah itu masih 4.1% dari keseluruhan 278.7 juta masyarakat yang ada di Indonesia sampai dengan Agustus 2023. Kecilnya persentase tersebut menurut hasil penelitian oleh Dewi & Purbawangsa (2018), dipengaruhi oleh kurangnya literasi mengenai investasi sehingga membuat individu ragu untuk menjadi seorang investor. Padahal menurut Sulistyowati & Rahmawati (2020), pasar saham sebagai sarana pendanaan usaha, memegang peranan penting dalam mendorong perekonomian negara.

Banyak pihak baik individu maupun instansi yang sudah melakukan pendekatan untuk memberikan informasi mengenai investasi saham kepada masyarakat Indonesia. Salah satu pendekatan yang banyak dilakukan yakni melalui konten media sosial. Akan tetapi, dari penelitian oleh Desai et al. (2022), informasi yang didapatkan melalui media sosial dapat merupakan informasi yang salah dan menjerumuskan. Hal ini didukung lagi oleh hasil penelitian Drahošová dan Balco (2017), dengan menyimpulkan bahwasannya informasi dari media sosial menyebabkan banjir informasi sehingga bukannya membuat mengerti melainkan menambah kebingungan. Adapula teknik untuk mendapatkan

informasi saham Indonesia melalui website. Akan tetapi, dari penelitian oleh Lalwani et al. (2018), hampir tidak memungkinkan untuk mendapatkan informasi lengkap yang diperlukan hanya dari satu halaman website. Dan dari hasil penelitiannya, disimpulkan bahwa chatbot dapat digunakan untuk menyelesaikan permasalahan tersebut.

Penelitian mengenai pengimplementasian chatbot sudah pernah dilakukan sebelumnya. Dan pada penelitian sebelumnya oleh Muangkammuen et al. (2018), dimana Chatbot yang dikembangkan menggunakan algoritma *RNN – LSTM* dapat menerima pertanyaan teks Thailand dan memberikan jawaban atas pertanyaan. Hasil dari pengujiannya menghasilkan chatbot yang dapat mengenali 86,36% label pertanyaan dengan akurasi jawaban sebesar 93,2%. Adapun penelitian lainnya dari Azni (2021) yang melakukan pengimplementasian *LSTM – Fuzzy String Matching* pada chatbot yang mana didapatkan keakuratan jawaban Chatbot sebesar 95%.

Dengan demikian, penulis mengajukan penelitian mengenai sarana informasi menggunakan algoritma *Long Short-Term Memory (LSTM)* dan *Integer Sequence Matching* dengan judul “Implementasi *Long Short-Term Memory (LSTM)* Dan *Integer Sequence Matching* Pada Sistem Chatbot Informasi saham Indonesia”.

## **1.2. Rumusan Masalah**

Sampai dengan Agustus 2023, masyarakat Indonesia yang menjadi investor saham Indonesia baru mencapai 4.1% dari keseluruhan jumlah masyarakat Indonesia. Kurangnya informasi dan pemahaman masyarakat mengenai investasi saham seperti istilah dalam dunia saham, strategi investasi saham, dan jenis saham yang ada menjadi beberapa alasan utama rendahnya partisipasi masyarakat Indonesia dalam berinvestasi saham di Indonesia. Oleh karena itu, pendekatan yang mampu menyediakan informasi tentang saham Indonesia yang mudah diakses dan bersifat dua arah cukup diperlukan.

## **1.3. Batasan Masalah**

Batasan masalah untuk penelitian ini, antara lain :

1. Chatbot hanya menerima input dan mengembalikan output dalam Bahasa Indonesia.
2. Data yang digunakan hanya berasal dari E-book saham bertipe PDF tahun 2019 – tahun 2023 yang dibagikan secara gratis oleh PT. *Stockbit Sekuritas* yang sudah berizin oleh Otoritas Jasa Keuangan (OJK) dan materi seminar saham resmi yang pernah diikuti oleh peneliti pada tahun 2022.



3. Informasi saham Indonesia yang diberikan oleh Chatbot hanya berfokus pada saham Indonesia.

#### **1.4. Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk mengimplementasikan Long Short-Term Memory dan Integer Sequence Matching dalam pembuatan Chatbot mengenai informasi saham Indonesia.

#### **1.5. Manfaat Penelitian**

Manfaat dari dilakukannya penelitian ini, yakni:

1. Mempermudah proses penyaluran informasi saham Indonesia Indonesia melalui layanan yang mudah diakses dan bersifat dua arah.
2. Meningkatkan literasi masyarakat tentang saham Indonesia.
3. Meningkatkan minat investasi saham masyarakat Indonesia di bursa saham Indonesia.

#### **1.6. Sistematika Penulisan**

Sistematika penulisan penelitian memiliki 5 bagian, yakni:

##### **Bab 1: Pendahuluan**

Bab 1 yakni pendahuluan memuat tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

##### **Bab 2: Landasan Teori**

Bab 2 yakni landasan teori berisi teori-teori pendukung dalam penelitian. Teori yang mendukung berupa teori Saham, Natural Language Processing (NLP), Chatbot, Long Short-Term Memory (LSTM), dan Application Programming Interface (API).

##### **Bab 3: Analisis Dan Perancangan Sistem**

Bab 3 yakni analisis dan perancangan sistem berisi analisis implementasi algoritma Long Short-Term Memory (LSTM) dan Integer Sequence Matching pada sistem chatbot informasi saham Indonesia. Bab ini juga berisi penjelasan data yang digunakan, arsitektur umum, hingga proses alur kerja chatbot.

##### **Bab 4: Implementasi Dan Pengujian**

Bab 4 yakni implementasi dan pengujian berisi implementasi dari perancangan sistem serta hasil dari pengujian sistem yang telah dibuat.

### **Bab 5: Kesimpulan Dan Saran**

Bab 5 yakni kesimpulan dan saran berisi kesimpulan dari rancangan yang telah dibuat serta saran yang diberikan untuk melanjutkan penelitian kedepannya.

## **BAB II**

### **LANDASAN TEORI**

Pada Bab ini berisi teori penelitian terdahulu sebagai pedoman dalam penelitian penerapan Long Short-Term Memory (LSTM) pada sistem chatbot informasi saham Indonesia.

#### **2.1. Saham**

Saham merupakan tanda penyertaan atau kepemilikan seseorang atau badan dalam suatu perusahaan atau perseroan terbatas. Saham sebagai salah satu alat untuk mencari tambahan dana, memberikan implikasi bahwa investor yang menanamkan modal tersebut mendapat kepemilikan perusahaan dimana pemilik saham disebut juga pemegang saham (*shareholder* atau *stockholder*) (Darmadji & Fakhrudin, 2015). Saham memiliki jangka waktu tidak terbatas dan siklus bisnis yang bersifat fluktuatif sehingga pendapatannya tidak pasti tergantung relatif pasar. Terdapat dua keuntungan yang didapat investor dalam membeli saham yaitu deviden dan capital gain. *Deviden* merupakan pendapatan (*earning*) yang diperoleh selama saham masih dimiliki dan biasanya dibagikan setiap akhir tahun dalam kebijakan perusahaan. Sedangkan, *capital gain* merupakan keuntungan yang diperoleh pada saat saham yang dimiliki tersebut dijual kembali pada harga yang lebih mahal (Samsul, 2015).

Menurut Ash-Shidiq & Setiawan (2015), harga saham adalah nilai suatu saham yang mencerminkan kekayaan perusahaan yang mengeluarkan saham tersebut, dimana perubahan atau fluktuasinya sangat ditentukan oleh kekuatan permintaan dan penawaran yang terjadi di pasar bursa (pasar sekunder). Semakin banyak investor yang ingin membeli atau menyimpan saham, harganya semakin naik. Sebaliknya semakin banyak investor yang ingin menjual atau melepaskan suatu saham, maka harganya semakin bergerak turun.

Salah satu karakteristik yang paling dikenal dari saham adalah risiko yang tinggi namun potensi pengembalian yang juga tinggi. Artinya, saham memiliki peluang untuk menghasilkan keuntungan yang besar, tetapi juga memiliki risiko kegagalan yang tinggi. Pemilik saham dapat mengalami keuntungan (capital gain) yang signifikan dalam waktu yang relatif singkat. Namun, fluktuasi harga saham yang besar juga dapat menyebabkan kerugian yang signifikan bagi pemilik saham dalam waktu yang singkat.

Dalam analisis saham, terdapat dua metode yang umum digunakan, yaitu analisis fundamental dan analisis teknikal. Analisis fundamental adalah pendekatan yang mempelajari faktor-faktor dasar atau fundamental, seperti tingkat kepemilikan, suku bunga, rasio keuangan, dan lain-lain, untuk mengevaluasi kinerja perusahaan dan nilai sahamnya. Di sisi lain, analisis teknikal mencoba menggunakan data historis harga saham untuk membuat prediksi tentang pergerakan harga di masa depan, dengan tujuan melakukan transaksi beli atau jual di pasar saham.

Analisis fundamental membahas kondisi perusahaan, termasuk manajemen, organisasi, sumber daya manusia, dan keuangan, yang tercermin dalam kinerja keuangan. Pendekatan ini dianggap lebih cocok untuk keputusan investasi jangka panjang. Analisis ini berusaha memperkirakan harga saham di masa depan dengan mengevaluasi faktor-faktor fundamental yang memengaruhi harga saham. Sedangkan, analisis teknikal berorientasi pada aspek harga (pembukaan, penutupan, tertinggi, dan terendah) suatu instrumen investasi dalam kerangka waktu tertentu. Analisis ini mengamati pola perilaku pasar yang direpresentasikan dalam grafik historis harga, dengan tujuan memproyeksikan pergerakan harga di masa depan. Analisis teknikal berusaha meramalkan fluktuasi harga dalam jangka pendek, memberikan rekomendasi kapan sebaiknya melakukan pembelian atau penjualan saham (Putri & Shabri, 2022).

Dengan menggunakan kedua metode analisis ini, investor dapat memperoleh pemahaman yang lebih baik tentang saham, mengidentifikasi peluang investasi, dan membuat keputusan yang lebih informasi mengenai perdagangan saham

## **2.2. Natural Language Processing**

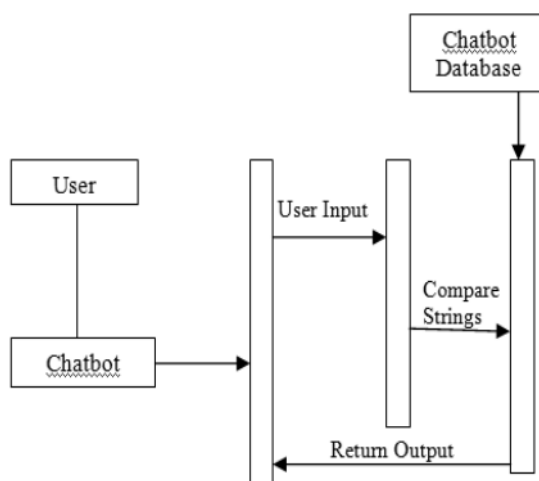
Natural Language Processing (NLP) merupakan metode untuk menganalisis dan merepresentasikan teks yang secara alami pada satu atau lebih tingkat analisis linguistik untuk memproses bahasa dengan gaya yang mirip dengan manusia (Liddy, 2001).

Menurut Pustejovsky & Stubbs (2012), tujuan penerapan Natural Language Processing (NLP) adalah untuk memfasilitasi komunikasi antara manusia dan mesin. Pustejovsky & Stubbs (2012) menambahkan bahwa Natural Language Processing (NLP) dapat diterapkan dalam berbagai bidang studi, termasuk:

1. Sistem Penjawab Pertanyaan (Question Answering Systems): Sistem ini memiliki potensi untuk menjawab pertanyaan yang diajukan oleh pengguna melalui penggunaan bahasa alami. Dengan menggunakan sistem penjawab pertanyaan, pengguna dapat mengajukan pertanyaan menggunakan bahasa sehari-hari, seperti bahasa Inggris, Mandarin, atau Indonesia, dan mendapatkan jawaban melalui mesin pencarian dengan memasukkan frasa kunci.
2. Ringkasan (Summarization): Bidang ini berkaitan dengan pembuatan rangkuman dari teks yang panjang dengan cepat. Perangkat lunak ringkasan membantu pengguna dalam mengubah konten teks yang besar menjadi ringkas. Dengan adanya teknologi ini, pengguna dapat dengan mudah mengakses informasi penting tanpa harus membaca seluruh teks yang panjang.
3. Terjemahan Mesin (Machine Translation): Terjemahan mesin melibatkan pengembangan perangkat lunak yang dapat memahami dan menerjemahkan teks dari satu bahasa ke bahasa lain. Contoh yang terkenal adalah Google Translate, yang terus meningkat dalam kemampuannya dalam menerjemahkan bahasa. BabelFish juga merupakan contoh sistem terjemahan mesin yang dapat menerjemahkan bahasa secara real-time.
4. Pengenalan Ucapan (Speech Recognition): Bidang ini melibatkan pengembangan teknologi yang kompleks untuk mengenali ucapan manusia. Metode pengenalan ucapan digunakan dalam sistem telepon atau komputer yang dapat merespons dan memahami instruksi atau pertanyaan dalam bahasa lisan.
5. Klasifikasi Dokumen (Document Classification): Bidang penelitian NLP ini berfokus pada pengembangan perangkat lunak yang dapat mengklasifikasikan dokumen ke dalam kategori yang sesuai. Fungsi utama dari klasifikasi dokumen adalah untuk menentukan area terbaik di mana sebuah dokumen baru harus ditempatkan dalam sistem. Hal ini sangat berguna dalam penyaringan email spam, klasifikasi artikel berita, dan aplikasi penilaian film.

### 2.3. Chatbot

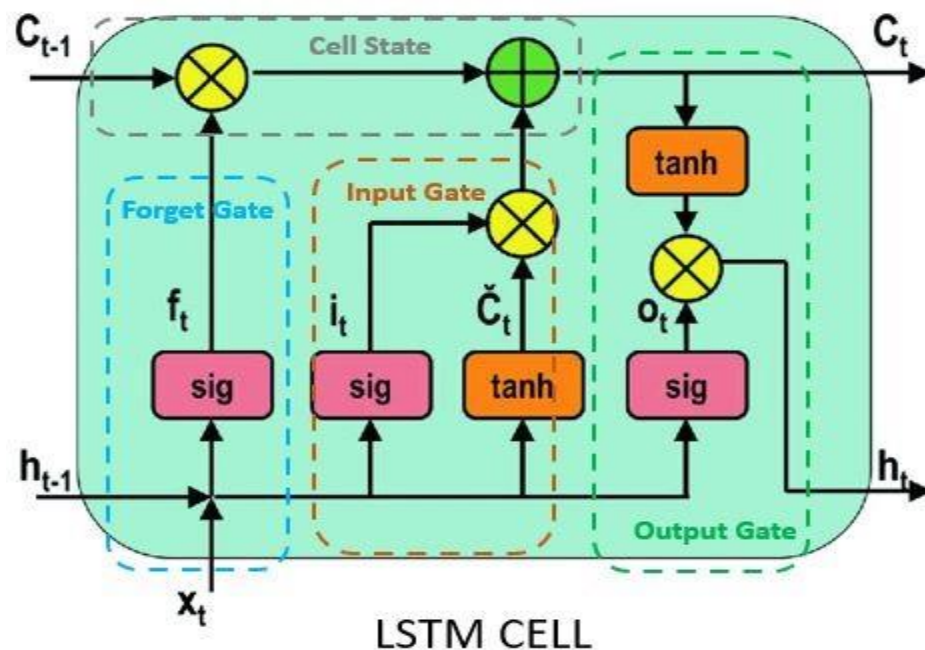
Chatbot merupakan sistem yang memungkinkan komunikasi antara user dengan sistem. Cara kerja Chatbot adalah chatbot menjawab pertanyaan yang diberikan oleh pengguna, tanggapan yang dikembalikan oleh chatbot kepada pengguna berasal dari data yang telah tersimpan dalam knowledge base (Dahiya, 2017). Untuk mendapatkan jawaban atas pertanyaan pengguna maka Chatbot perlu melakukan matching dari pertanyaan pengguna ke knowledge base yang tersedia (Migunani & Aditama, 2020). Matching pertanyaan merupakan metode matching pola yang digunakan untuk memeriksa urutan token yang ada di dalam suatu kalimat. Contoh dari metode matching pertanyaan yaitu Fuzzy String Matching (Ribalta, 2014). Gambaran umum cara kerja chatbot dapat dilihat pada Gambar 2.1 berikut:



**Gambar 2.1.** Gambaran Umum Cara Kerja Chatbot (Dahiya, 2017)

### 2.4. Algoritma Long Short-Term Memory (LSTM)

Algoritma Long Short-Term Memory (LSTM) adalah sebuah jaringan saraf tiruan yang mana dapat menangkap keterkaitan dari sebuah urutan data yang panjang untuk menentukan bagaimana memanfaatkan informasi yang tersimpan di dalam memory cell (Olah, 2015). Memory cell yang terdapat dalam Algoritma LSTM juga memungkinkan penyimpanan masukan untuk waktu yang lama dan menjadi solusi untuk mengatasi terjadinya *vanishing gradient* (*gradient* menjadi sangat kecil selama proses *backpropagation* sehingga menyebabkan bobot tidak banyak diperbarui) saat memproses data sequential yang panjang (Manaswi, 2018). Berikut adalah Gambar 2.2. mengenai arsitektur dari algoritma Long Short-Term Memory (LSTM):



**Gambar 2.2.** Arsitektur Long Short-Term Memory (Olah, 2015)

Dalam ilustrasi di atas, terlihat bahwa setiap neuron LSTM memiliki proses memory cells yang berfungsi. Pada gambar tersebut, terlihat pula bahwa unit pintu (gates units) memiliki empat proses aktivasi pada setiap input di neuron tersebut. Keempat proses pada gates units tersebut adalah output gates, forget gates, cell gates, dan input gates.

Proses forget gates merupakan tahap di mana informasi dalam setiap data input dikelola dan difilter untuk menentukan apakah informasi tersebut disimpan atau dibuang dalam memory cells. Forget gates menggunakan fungsi aktivasi sigmoid, yang menghasilkan output bernilai 1 jika tidak 0. Jika output bernilai 1, proses tersebut menyimpan semua data. Namun, jika output bernilai 0, maka semua data dibuang. Proses ini dihitung menggunakan Rumus 2.1. yakni rumus *forget gate* berikut:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (2.1)$$

Dimana :

- $f_t$  = forget gate
- $\sigma$  = sigmoid activation function
- $W_f$  = weight
- $h_{t-1}$  = previous hidden state

$x_t$  = input data  
 $b_i$  = bias

Untuk input gates, terdapat dua gates pelaksanaan yang terlibat. Pertama, menggunakan fungsi aktivasi sigmoid untuk menentukan nilai mana yang diperbarui. Setelah itu, dilakukan pembuatan vektor nilai baru dengan menggunakan fungsi aktivasi tanh, yang kemudian disimpan di memory cell. Proses ini dihitung menggunakan Rumus 2.2. yakni rumus *input gate* dan Rumus 2.3. yakni rumus *cell state* berikut:

$$i_t = \sigma( W_f \cdot [h_{t-1}, x_t] + b_i ) \quad (2.2)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c ) \quad (2.3)$$

Dalam Cell gates, terjadi proses penggantian memory cell baru yang terjadi ketika nilai yang diperoleh dari penggabungan nilai yang ada di forget gate dan input gate. Hasilnya dihitung menggunakan Rumus 2.4. yakni rumus *updated cell state* berikut:

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (2.4)$$

Dimana :

$i_t$  = input gate  
 $\tilde{c}_t$  = cell state  
 $c_t$  = updated cell state  
 $c_{t-1}$  = previous cell state  
 $\sigma$  = sigmoid activation function  
 $\tanh$  = tanh activation function  
 $W_f$  = weight  
 $W_c$  = weight cell state  
 $h_{t-1}$  = previous hidden state  
 $x_t$  = input data  
 $b_i$  = bias  
 $b_c$  = nilai bias pada cell state  
 $f_t$  = forget gate

Sementara itu, output gates terdiri dari dua gates yang berfungsi untuk menentukan nilai dari memory cell yang dikeluarkan melalui fungsi aktivasi sigmoid.



Selanjutnya, dengan menggunakan fungsi aktivasi tanh, nilai tersebut ditempatkan kembali ke dalam memory cell. Proses terakhir adalah mengalikan dua gates tersebut untuk mendapatkan nilai output yang dihasilkan ( $h_t$ ). Proses ini menggunakan Rumus 2.5. yakni rumus *sigmoid output* dan Rumus 2.6 yakni rumus *hidden state* berikut:

$$o_t = \sigma( W_o \cdot [h_{t-1}, x_t] + b_o ) \quad (2.5)$$

$$h_t = o_t * \tanh(c_t) \quad (2.6)$$

Dimana:

- $o_t$  = sigmoid output
- $\sigma$  = sigmoid activation function
- $W_o$  = weight of the output
- $h_{t-1}$  = previous hidden state
- $x_t$  = input data
- $b_o$  = output bias
- $h_t$  = hidden state
- $\tanh$  = tanh activation function
- $c_t$  = updated cell state

## 2.5. Application Programming Interface (API)

Application Programming Interface (API) merupakan antarmuka penghubung layanan atau data yang disediakan oleh aplikasi perangkat lunak melalui serangkaian sumber daya yang telah ditentukan sebelumnya, seperti metode, objek, atau URI (Stylos et al., 2009). Dengan menggunakan API, aplikasi lain dapat mengakses data atau layanan tanpa harus mengimplementasikan objek dan prosedur yang mendasarinya. API memegang peranan utama dalam banyak arsitektur perangkat lunak modern, karena API menyediakan antarmuka yang memfasilitasi tugas pemrograman, mendukung desain aplikasi perangkat lunak yang terdistribusi dan modular, serta penggunaan ulang kode (Robillard, 2009).

## 2.6. Penelitian Terdahulu

Khairani, (2021) pada penelitiannya yang berjudul “*Aplikasi Chatbot Tanya Jawab tentang Kesehatan Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma Knuth Morris Pratt*” mengajukan penelitian tentang chatbot untuk membantu dalam pengdiagnostikan kesehatan. Proses tanya jawab pada chatbot ini menggunakan 2

metode yakni metode stemming dalam mencari akar kata dan metode matching kata. Hasil yang ditunjukkan dalam penelitian ini membuktikan bahwa Chatbot dengan menggunakan *Knuth Morris Pratt* mendapatkan tingkat akurasi sebesar 85,36 % dengan jawaban yang benar sebanyak 35 dari 41 data uji.

Selanjutnya adalah penelitian oleh Sihotang, (2019) yang berjudul “*Chatbot Tanya Jawab Islamic APP Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma Fuzzy String Matching*”. Penelitian Chatbot ini bertujuan untuk membantu kemudahan pencarian ilmu Islamic. Penelitian yang dilakukan ini menerapkan dua method yaitu stemming dan string matching. Data uji penelitian ini berasal dari responden pengguna Chatbot, Hasil dari pengujian data uji menunjukkan bahwa Chatbot dengan menerapkan Fuzzy String Matching kurang efisien karena menghasilkan tingkat akurasi 70,37 % dengan jawaban yang benar sebanyak 19 dari 27 responden.

Penelitian berikutnya berasal dari Muangkammuen et al., (2018) yang berjudul “*AutomatedThai-FAQ Chatbot using RNN-LSTM*”. Cara kerja chatbot ini adalah dengan menerima pertanyaan teks bahasa Thailand dan memberikan jawaban selayaknya seorang customer service. Hasil dari pengujian menunjukkan bahwa chatbot dapat mengenali 86,36% pertanyaan dan jawaban dengan akurasi 93,2%.

Selanjutnya ada penelitian oleh Nugraha, (2020) yakni “*Aplikasi Pemesanan Travel Menggunakan Chatbot dengan Artificial Intelligence Markup Language*”. Chatbot yang dikembangkan dari penelitian ini dapat mempermudah pengguna dalam mendapatkan informasi seputar pemesanan tiket. Algoritma yang digunakan dalam penelitian ini adalah algoritma. Hasil dari pengujian membuktikan Chatbot mampu menjawab pertanyaan mengenai jadwal, harga, fasilitas, dan lain-lain dengan rata-rata akurasi yang didapatkan sebesar 83%.

*Artificial Intelligence Markup Language (AIML)* merupakan bahasa markup yang digunakan untuk mengembangkan sistem berbasis aturan untuk kecerdasan buatan, khususnya dalam konteks pembangunan chatbot. AIML menggunakan konsep pengenalan pola untuk memfasilitasi percakapan antara manusia dan chatbot dengan pendekatan pola-respons. AIML menonjol dengan sintaksis sederhana, kemudahan pemahaman, fleksibilitas terhadap varian input, dan skalabilitas untuk chatbot

sederhana hingga kompleks, didukung oleh komunitas pengembang yang aktif. Namun, kelemahan AIML terletak pada keterbatasan dalam pemrosesan bahasa alami yang kompleks, menjadikannya lebih cocok untuk aplikasi dengan tingkat kompleksitas rendah hingga menengah, sementara pengelolaan dan pemeliharaan aturan AIML dapat menjadi rumit seiring meningkatnya kompleksitas chatbot (Barbosa et al., 2013).

Penelitian selanjutnya oleh Azni, (2021) berjudul “*Implementasi Natural Language Processing Pada Sistem Chatbot Informasi saham Indonesia dengan Algoritma Long Short-Term Memory (LSTM) dan Fuzzy String Matching*”. Pada penelitian ini, penulis mengimplementasikan NLP pada Chatbot yang dapat memberikan informasi informasi seputar saham. Penelitian ini menerapkan algoritma Long Short-Term Memory dan Fuzzy String Matching sebagai algoritma pendukungnya dalam implementasi dan pengembangannya. Hasil pengujian oleh penulis menghasilkan sebuah model dengan accuracy 98.58% dan loss sebesar 5.9%. Dengan menerapkan LSTM dan Fuzzy String Matching penulis berhasil menghasilkan Chatbot yang memiliki jawaban akurat dengan tingkat akurasi 95%. Ringkasan dari penelitian terdahulu dapat dilihat pada Tabel 2.1. berikut:

**Tabel 2.1.** Tabel Penelitian Terdahulu

No	Nama	Judul	Metode	Keterangan
1.	Muangkammuen, Narong Intiruk, dan Kanda Runapongsa Saikaew (2018)	Automated Thai-FAQ Chatbot using RNN-LSTM	Recurrent Neural Network dan Long Short-Term Memory	Chatbot dapat mengenali 86,36% pertanyaan dengan akurasi jawaban 93,2%.
2.	Muhammad Tohir Sihotang (2019)	Chatbot Tanya Jawab Islamic APP Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma Fuzzy String Matching	Algoritma Enhanced Confix Stripping dan Algoritma Fuzzy String Matching	Hasil Chatbot dengan Fuzzy String Matching kurang efisien dengan tingkat akurasi 70,37%.

**Tabel 2.1.** Tabel Penelitian Terdahulu (Lanjutan)

No	Nama	Judul	Metode	Keterangan
3.	Adhindya Trigha Nugraha (2020)	Aplikasi Pemesanan Travel Menggunakan Chatbot dengan AIML	Artificial Intelligence Markup Language	Hasil dari pengujian membuktikan Chatbot mampu menjawab pertanyaan mengenai jadwal, harga, fasilitas, dan lain-lain dengan rata-rata akurasi yang didapatkan sebesar 83%.
4.	Fitri Khairani (2021)	Aplikasi Chatbot Tanya Jawab tentang Kesehatan Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma KMP	Algoritma Enhanced Confix Stripping dan Algoritma Knuth Morris Pratt	Hasil yang ditunjukkan dalam penelitian ini membuktikan bahwa Chatbot mampu mendapatkan tingkat akurasi sebesar 85,36 %.
5.	Azmitha Azni (2021)	Implementasi Natural Language Processing Pada Sistem Chatbot Informasi saham Indonesia dengan Algoritma LSTM dan Fuzzy String Matching	Algoritma Long Short- Term Memory (LSTM) dan Fuzzy String Matching	Hasil pengujian menghasilkan sebuah model dengan accuracy 98.58% dan loss sebesar 5.9%. Dengan menerapkan LSTM dan Fuzzy String Matching, penulis berhasil menghasilkan chatbot yang memiliki tingkat akurasi sebesar 95%.

## **BAB III**

### **METODOLOGI PENELITIAN**

Pada bab ini membahas tentang data yang digunakan dan arsitektur chatbot informasi saham Indonesia menggunakan algoritma Long Short-Term Memory (LSTM) dan Integer Sequence Matching.

#### **3.1. Dataset**

Pada penelitian ini, terdapat 350 baris dataset yang dikutip dan digunakan sebagai corpus dari beberapa sumber yakni E-book saham dan seminar saham. Sekitar 200 baris dataset dari E-book saham (tahun 2019 – tahun 2023) yang disediakan untuk umum secara gratis oleh PT. *Stockbit Sekuritas* yang sudah berizin oleh Otoritas Jasa Keuangan (OJK), 150 baris dataset lainnya berasal dari gabungan beberapa materi seminar saham yang diikuti oleh penulis pada tahun 2022. Dataset yang dikumpulkan kemudian diberi label terlebih dahulu dengan bantuan pakar saham Indonesia. Jenis pelabelan didasarkan pada penelitian sebelumnya oleh Azni, (2021) dimana pada penelitiannya menggunakan 13 pelabelan yakni ‘Saham’, ‘Strategi’, ‘Investasi’, ‘Sekuritas’, ‘Broker’, ‘Pasar’, ‘Indeks’, ‘Sektor’, ‘Trading’, ‘Analisis’, ‘Perusahaan’, ‘Reksadana’, dan ‘Istilah’. Berdasarkan penelitian oleh Chi et al. (2022) jumlah label yang terlalu banyak dapat mempengaruhi performa model. Oleh karena itu, beberapa pelabelan yang terlalu spesifik disatukan dengan bantuan pakar saham Indonesia sehingga menyisakan 5 jenis pelabelan utama yakni: ‘Istilah’ sebanyak 230 baris data, ‘Investasi’ sebanyak 52 baris data, ‘Trading’ sebanyak 12 baris data, ‘Strategi’ sebanyak 46 baris data, dan ‘IPO’ sebanyak 10 baris data. Dimana IPO (*Initial Public Offering*) merupakan suatu peristiwa penawaran saham yang dilakukan oleh perusahaan (emiten) kepada masyarakat umum (investor) untuk pertama kalinya (Sunariyah, 2003). Pelabelan dibantu oleh seorang pakar saham Indonesia bersertifikat dalam bidang saham bernama *Fransiskus Wiguna, S.AB* yang merupakan seorang broker sekuritas &

mentor komunitas saham Indonesia. Berikut adalah Tabel 3.1 yang berisi beberapa dataset yang sudah dilabeli dengan bantuan pakar saham Indonesia yang digunakan dalam penelitian:

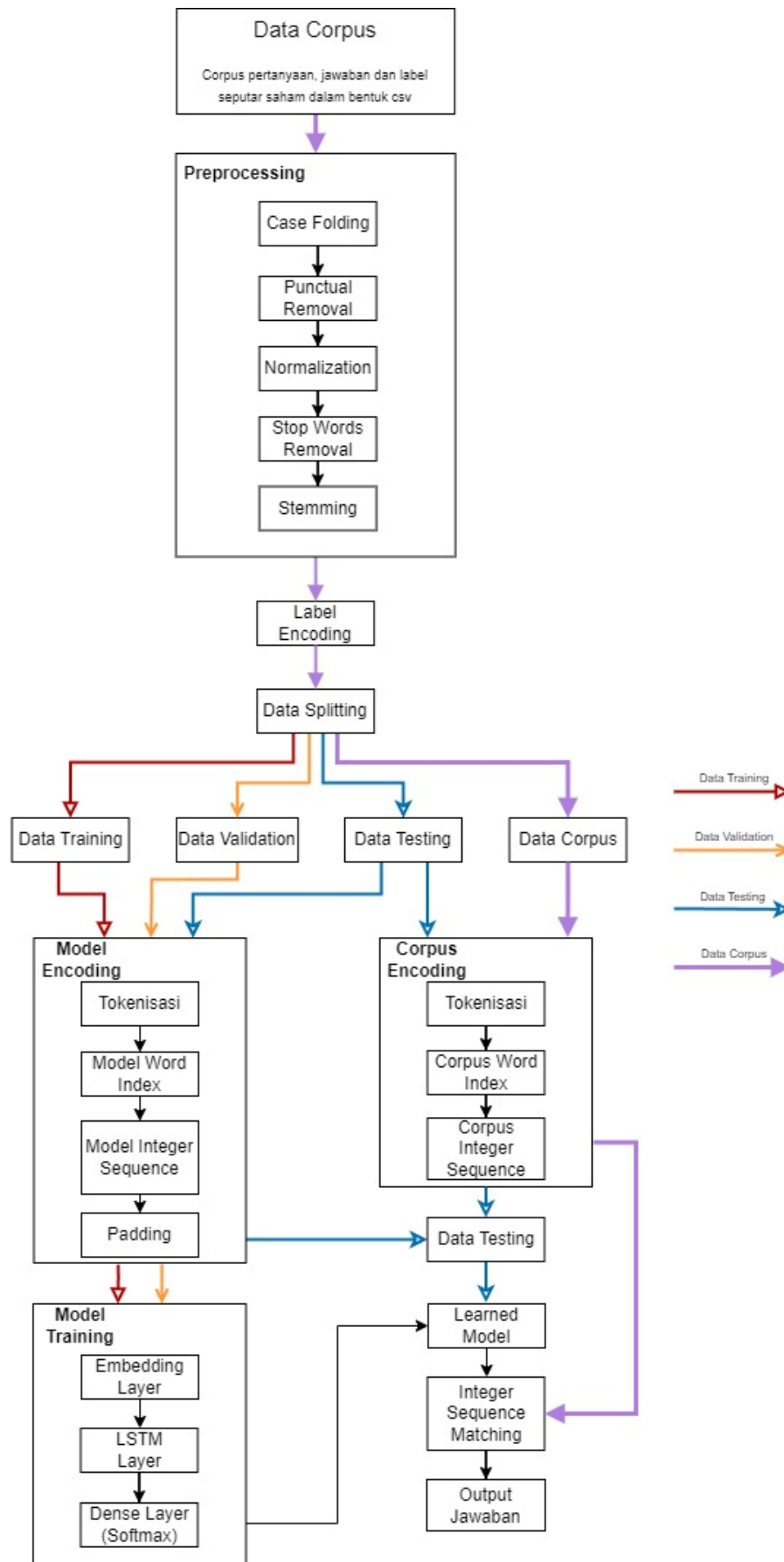
**Tabel 3.1.** Contoh Dataset Yang Digunakan

<b>Pertanyaan</b>	<b>Jawaban</b>	<b>Label</b>
Apa itu saham?	Saham adalah instrumen keuangan yang mewakili kepemilikan dalam suatu perusahaan.	Istilah
Apa itu IPO?	IPO adalah proses di mana perusahaan pertama kali menawarkan sahamnya kepada publik. Ini adalah saat di mana saham perusahaan dapat dibeli oleh investor umum di pasar.	IPO
Bagaimana cara membeli saham?	Anda dapat membeli saham melalui perusahaan sekuritas atau broker saham. Anda perlu membuka akun di perusahaan tersebut dan melakukan transaksi pembelian saham melalui platform mereka.	Strategi
Bagaimana cara menjual saham?	Untuk menjual saham Anda dapat menggunakan platform perdagangan saham yang disediakan oleh perusahaan sekuritas atau broker saham. Anda dapat memilih saham yang ingin dijual dan mengeksekusinya.	Strategi
Apa yang harus dilakukan saat pasar saham mengalami volatilitas?	Saat pasar saham mengalami volatilitas penting untuk tetap tenang dan tidak panik. Disarankan untuk mempertahankan portofolio yang terdiversifikasi melakukan riset tambahan dan berkonsultasi dengan penasihat keuangan jika diperlukan.	Strategi
Tujuan trading saham?	Tujuan trading adalah untuk mengejar keuntungan maksimal dalam waktu yang singkat dengan strategi dan analisis yang bagus untuk menghindari kerugian ketika melakukan kegiatan trading. Dalam trading, trader harus berfokus pada sentimen dan kondisi pasar.	Trading
Apa perbedaan antara investasi jangka pendek dan investasi jangka panjang?	Investasi jangka pendek biasanya berlangsung dalam waktu kurang dari satu tahun sementara investasi jangka panjang dilakukan untuk periode yang lebih lama biasanya bertahun-tahun. Tujuan dan strategi investasi dapat berbeda antara keduanya.	Investasi
Apa yang harus diperhatikan saat memilih saham untuk investasi?	Saat memilih saham untuk investasi perhatikan kinerja keuangan perusahaan tren industri manajemen perusahaan dan faktor-faktor lain yang dapat mempengaruhi harga saham.	Investasi

### 3.2. Arsitektur Umum

Arsitektur umum dimulai dengan *preprocessing* pada *data corpus*. *Data corpus* yang sudah melalui proses *preprocessing*, dilakukan *label encoding* untuk mengubah *categorical* label menjadi label *integer encoded* sehingga dapat dipahami dan diproses model. Setelahnya, *data corpus* dipisah menjadi *data training*, *data validation*, *data testing*, dan *data corpus* dengan rasio 70%:10%:20%:100%. *Data training* dan *data validation* melalui proses *model encoding* yang meliputi tokenisasi dan *data training* dijadikan *model word index* sebagai pedoman dalam menghasilkan *model integer sequence* yang terpaded untuk melatih model, *data corpus* melalui proses *corpus encoding* yang meliputi tokenisasi dan dijadikan *model word index* sebagai pedoman dalam menghasilkan *model integer sequence* untuk digunakan dalam *integer sequence matching*, sedangkan *data testing* melalui kedua proses tersebut yakni *model encoding* dan *corpus encoding* dan menghasilkan *model integer sequence* dan *corpus integer sequence* untuk proses pengujian model dalam memprediksi label dan pengujian *integer sequence matching*.

*Model integer sequence* milik *data training* dan *data validation* digunakan dalam pelatihan model *LSTM* untuk menghasilkan model *LSTM* yang memiliki performa yang optimal. Model yang telah dilatih kemudian diuji terhadap *data testing* menggunakan *model integer sequencenya* untuk menghasilkan prediksi label. *Data testing* yang telah diprediksi labelnya dilakukan *integer sequence matching* menggunakan *corpus integer sequencenya* terhadap *corpus integer sequence* milik *data corpus* dengan label yang sama. Jawaban dengan tingkat kecocokan *corpus integer sequence* tertinggi diambil dan dijadikan output. Berikut adalah Gambar 3.1 yang merupakan arsitektur umum dalam penelitian ini:



Gambar 3.1. Arsitektur Umum



Penjelasan tentang Arsitektur Umum pada Gambar 3.1:

### 1. *Preprocessing*

*Preprocessing* adalah tahap awal dimana, data pertanyaan corpus disesuaikan untuk mempermudah pelatihan model. *Preprocessing* meliputi tahap *Case Folding*, *Punctual Removal*, *Normalization*, *Stopword Removal*, dan *Stemming*.

#### 1.1. *Case Folding*

*Case Folding* adalah salah satu langkah dalam tahap preprocessing dalam pengolahan teks. Proses *Case Folding* dilakukan untuk menyeragamkan karakter dalam data dengan mengubah seluruh huruf menjadi huruf kecil. Dalam proses ini, setiap karakter 'A'-'Z' yang terdapat dalam data diubah menjadi karakter 'a'-'z'.

Tujuan dari *Case Folding* adalah untuk menghilangkan perbedaan antara huruf besar (uppercase) dan huruf kecil (lowercase) dalam teks. Dengan mengubah semua huruf menjadi huruf kecil, sehingga dapat memperlakukan kata-kata yang memiliki huruf besar dan huruf kecil yang sama secara seragam. Sehingga kata "Saham" dan "saham" dianggap sama setelah dilakukan *Case Folding* dan memudahkan dalam proses pemrosesan dan analisis teks. Contoh penerapan *Case Folding* dapat dilihat pada Tabel 3.2:

**Tabel 3.2.** Contoh *Case Folding*

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
Bgm cara membeli saham?	bgmn cara membeli saham?

#### 1.2. *Punctual Removal*

*Punctual Removal*, yang dikenal sebagai penghapusan tanda baca, merupakan tahap dalam *Preprocessing* yang melibatkan eliminasi tanda baca dari teks tersebut. Tanda baca yang termasuk di antaranya adalah titik, koma, tanda tanya, tanda seru, tanda hubung, dan sejenisnya. Contohnya adalah: [!'"#\$%&'()\*+,-./:;<=>?@[\\]^\_`{|}~].

Tujuan utama dari punctual removal adalah untuk membersihkan teks dari tanda baca yang tidak relevan dalam beberapa tugas pemrosesan bahasa alami. Dengan menghapus tanda baca tersebut, teks menjadi lebih bersih dan dapat diproses lebih efektif oleh model nantinya. Contoh penerapan *Punctual Removal* dapat dilihat pada Tabel 3.3:

**Tabel 3.3.** Contoh *Punctual Removal*

Sebelum <i>Punctual Removal</i>	Setelah <i>Punctual Removal</i>
bgmn cara membeli saham?	bgmn cara membeli saham

### 1.3. *Normalization*

*Normalization* adalah proses mengubah kata-kata menjadi bentuk baku atau standar yang umum digunakan. Tujuannya adalah untuk mengatasi variasi dalam penulisan kata yang dapat muncul dalam teks yang berbeda. *Normalization* kata baku membantu meningkatkan keseragaman dan konsistensi dalam teks, serta memudahkan pemrosesan dan analisis bahasa alami. Dengan menggunakan proses ini, model NLP dapat menghasilkan hasil yang lebih akurat dan konsisten. Contoh penerapan *Normalization* dapat dilihat pada Tabel 3.4:

**Tabel 3.4.** Contoh *Normalization Kata Baku*

Sebelum <i>Normalization</i>	Setelah <i>Normalization</i>
bgmn cara membeli saham	bagaimana cara membeli saham

### 1.4. *Stopword Removal*

*Stopword Removal* adalah proses dalam *preprocessing* yang melakukan penghapusan kata-kata yang sering muncul dan cenderung tidak memberikan informasi penting. Tujuan utama dari penghapusan stopwords adalah untuk membersihkan teks dari kata yang tidak penting sehingga fokus dapat diarahkan pada kata-kata yang lebih informatif dan penting. Beberapa contoh *Stopword* adalah kata: adalah, dan, atau, yang, di, dalam, pada, dengan, untuk, dari, itu, ini, atas, bawah, seperti, saja, sebelum, sesudah, oleh, karena, maka, hingga, jika, pada, tentang, bagaimana, apa. Contoh penerapan *Stopword Removal* dapat dilihat pada Tabel 3.5:

**Tabel 3.5.** Contoh *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Setelah <i>Stopword Removal</i>
bagaimana cara membeli saham	membeli saham

### 1.5. *Stemming*

*Stemming* adalah proses dalam *preprocessing* yang melibatkan eliminasi akhiran atau awalan kata, misal: ke-,me-,mem-,ber- dan sejenisnya untuk

mendapatkan bentuk dasar atau akar kata. Fungsinya adalah untuk mengurangi variasi dalam bentuk kata yang memiliki akar yang sama. Dalam proses stemming, digunakan aturan heuristik yang sederhana untuk mengenali dan menghapus akhiran atau awalan kata tersebut. Pendekatan ini tidak mempertimbangkan konteks atau makna kata-kata tersebut, melainkan hanya fokus pada struktur kata itu sendiri. Contoh penerapan *Stemming* dapat dilihat pada Tabel 3.6. berikut:

**Tabel 3.6.** Contoh *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
membeli saham	beli saham

## 2. *Label Encoding*

*Label Encoding* merupakan tahap setelah *preprocessing* dimana label dari setiap baris data diubah dari bentuk *categorical* labels menjadi bentuk *integer encoded* labels. Proses *Label Encoding* ini dilakukan agar model dapat menerima dan memproses label yang ada untuk diprediksi. Hal ini dikarenakan model hanya dapat memahami data bertipe integer sedangkan dalam bertipe string tidak dapat diproses oleh model. *Label Encoding* dilakukan dengan menggunakan module *LabelEncoder* yang tersedia di library *Sklearn*. Tabel 3.7. menunjukkan *Label Encoding* label dari data yang dimiliki:

**Tabel 3.7.** Contoh *Label Encoding*

Sebelum <i>Label Encoding</i>	Setelah <i>Label Encoding</i>
IPO	0
Investasi	1
Istilah	2
Strategi	3
Trading	4

## 3. *Data Splitting*

*Data Splitting* merupakan tahap setelah *preprocessing* dimana setelah data disesuaikan dan label tiap data sudah di encoding maka *corpus* dipecah menjadi *data training*, *data validation*, *data testing*, dan *data corpus* dengan perbandingan 70%:10%:20%:100%. Dimana 70% dari keseluruhan corpus sebagai *data training*,

10% dari keseluruhan corpus diluar *data training* sebagai *data validation*, dan 20% dari keseluruhan corpus diluar *data training* dan *data validation* sebagai *data testing* dan 100% dari keseluruhan corpus sebagai *data corpus*. *Data Training* digunakan untuk melatih menghasilkan model yang memiliki akurasi yang tinggi. *Data Validation* digunakan untuk proses validasi model dan mencegah *overfitting*. *Data Testing* digunakan untuk *testing* model, sebagai simulasi penggunaan model pada data yang tidak pernah dilihat oleh model sebelumnya. Sedangkan *data Corpus* digunakan untuk Integer Sequence Matching input pertanyaan terhadap *Integer Sequence* pertanyaan *corpus* untuk diambil jawaban yang paling sesuai berdasarkan kecocokan tertinggi. Tabel 3.8. berikut menunjukkan jumlah *data training*, *data validation*, *data testing*, dan *data corpus* setelah di splitting:

**Tabel 3.8.** Pembagian Data Training, Data Validation, Data Testing & Data Corpus

Data Training	Data Validation	Data Testing	Data Corpus
245 baris data	35 baris data	70 baris data	350 baris data

#### 4. Model Encoding

*Model Encoding* merupakan tahap dimana *pertanyaan* dari *data training* dan *data testing* ditransformasi dari sequence bertipe string menjadi sequence bertipe integer (*Integer Sequence*) yang mana dapat diterima dan diproses oleh model. Dalam tahap *Model Encoding* ini, *data training* dijadikan sebagai pedoman *Vocabulary Word Index*. Tahap *data encoding* ini meliputi proses *tokenization*, *Model Vocabulary Word Index*, *model Integer Sequence*, dan *padding*.

##### 4.1. Tokenisasi

*Tokenisasi* merupakan proses pemisahan kalimat menjadi potongan-potongan yang disebut token, yang dapat berupa kata, angka, simbol, tanda baca, atau entitas penting lainnya. Dalam konteks *Natural Language Processing* (NLP) ini, token mengacu pada kata-kata pecahan dari kalimat pertanyaan. Dalam kasus lain, tokenisasi juga dapat dilakukan pada tingkat paragraf atau kata. Proses tokenisasi memungkinkan analisis lebih lanjut terhadap kalimat, dimana setiap tokennya dapat dianggap sebagai unit terpisah yang memiliki makna atau relevansi dalam konteksnya. Contoh penerapan *Tokenisasi* dapat dilihat pada Tabel 3.9. berikut:

**Tabel 3.9.** Contoh *Tokenisasi*

Sebelum <i>Tokenisasi</i>	Setelah <i>Tokenisasi</i>
beli saham	['beli','saham']

#### 4.2. *Model Vocabulary Word Index*

*Model Vocabulary Word Index* merupakan tahap pembuatan daftar indeks yang digunakan untuk merepresentasikan kata-kata atau token dalam teks dalam bentuk numerik. Pada tahap *Model Encoding* ini, *data training* yang dijadikan pedoman dalam pembuatan daftar indeks. Pemberian penomoran indeks didasarkan pada urutan frekuensi kemunculan. Semakin sering kata itu muncul maka semakin tinggi urutannya. Kata <OOV> (Out Of Vocabulary) juga diberikan indexing untuk mempersiapkan kata-kata asing agar tetap bisa diolah. Contoh *Model Vocabulary Word Index* dapat dilihat pada Tabel 3.10. berikut:

**Tabel 3.10.** Contoh *Model Vocabulary Word Index*

Kata	Index
<OOV>	1
saham	2
beli	3
jual	4
analisis	5
investasi	6
.....	.....
turnover	335

#### 4.3. *Model Integer Sequence*

*Model Integer Sequence* adalah proses dimana tiap baris *sequence* pertanyaan yang sudah di tokenisasi tadinya dijadikan sebagai *Integer Sequence* berdasarkan *Model Vocabulary Word Index* yang sudah dibuat sebelumnya. Proses ini mengubah *sequence* pertanyaan yang semula bertipe *string* menjadi *sequence* pertanyaan bertipe *integer*. Alasan dilakukannya proses ini sama seperti alasan dilakukannya proses *Label Encoding* yakni dikarenakan model tidak dapat mencerna data bertipe *string* sehingga harus diubah dahulu menjadi *integer* agar dapat dimengerti model. Contoh penerapan *Model Integer Sequence* dapat dilihat pada Tabel 3.11. berikut:

**Tabel 3.11.** Contoh *Model Integer Sequence*

Sebelum <i>Model Integer Sequence</i>	Setelah <i>Model Integer Sequence</i>
['beli','saham']	[3,2]

#### 4.4. *Padding*

*Padding* adalah proses dimana panjang tiap *Integer Sequence* disamakan. Proses *padding* ini memiliki tujuan yang sangat penting untuk *Embedding Layer*. Dalam pemrosesan teks, setiap urutan kata atau sequence harus memiliki panjang yang sama saat diinput ke *Embedding Layer*. Panjang maksimal yang ditetapkan untuk tiap *Integer Sequence* adalah sama dengan panjang sequence terpanjang dalam kumpulan *Integer Sequence* yang dimiliki.

Dalam corpus, panjang sequence data terpanjang yang dimiliki adalah 8. Oleh karena itu, setiap *Integer Sequence* pola pertanyaan diubah menjadi matriks dengan panjang 8 digit dengan cara menambahkan nilai khusus yang menandakan bahwa itu adalah bagian yang tidak relevan dari urutan kata (biasanya menggunakan angka 0). Dengan demikian, semua urutan kata memiliki panjang yang seragam. Contoh penerapan *Padding* dapat dilihat pada Tabel 3.12. berikut:

**Tabel 3.12.** Contoh *Padding*

Sebelum <i>Padding</i>	Setelah <i>Padding</i>
[3,2]	[0,0,0,0,0,0,3,2]

#### 5. *Corpus Encoding*

*Corpus Encoding* merupakan tahap yang memiliki kemiripan dengan *model encoding*. Dilakukannya *corpus encoding* ini adalah untuk menghasilkan *Integer Sequence* yang digunakan dalam proses *Integer Sequence Matching*. *Integer Sequence* hasil proses *corpus encoding* berbeda dengan *Integer Sequence* hasil *model encoding* dikarenakan *corpus encoding* menggunakan *Corpus Vocabulary Word Index* yang berbeda dengan *Model Vocabulary Word Index*. *Corpus Vocabulary Word Index* menggunakan keseluruhan *data corpus* sebagai pedoman *indexing* berbeda dengan *Model Vocabulary Word Index* yang menggunakan *data training* sebagai pedoman *indexing*. Tahap-tahap dalam *corpus encoding* ini pun sedikit berbeda dengan *model encoding* karena hanya meliputi proses *tokenization*, *Corpus Vocabulary Word Index*, dan *corpus Integer Sequence*.

### 5.1. Tokenisasi

*Tokenisasi* merupakan proses pemisahan kalimat menjadi potongan-potongan yang disebut token, yang dapat berupa kata, angka, simbol, tanda baca, atau entitas penting lainnya. Dalam konteks *Natural Language Processing* (NLP) ini, token mengacu pada kata-kata pecahan dari kalimat pertanyaan. Dalam kasus lain, tokenisasi juga dapat dilakukan pada tingkat paragraf atau kata. Proses tokenisasi memungkinkan analisis lebih lanjut terhadap kalimat, dimana setiap tokennya dapat dianggap sebagai unit terpisah yang memiliki makna atau relevansi dalam konteksnya. Contoh penerapan *Tokenisasi* dapat dilihat pada Tabel 3.13. berikut:

**Tabel 3.13.** Contoh *Tokenisasi*

Sebelum <i>Tokenisasi</i>	Setelah <i>Tokenisasi</i>
beli saham	['beli', 'saham']

### 5.2. Corpus Vocabulary Word Index

*Corpus Vocabulary Word Index* merupakan tahap pembuatan daftar indeks yang digunakan untuk merepresentasikan kata-kata atau token dalam teks dalam bentuk numerik. Pada tahap *Corpus Encoding* ini, *data corpus* yang dijadikan pedoman dalam pembuatan daftar indeks. Pemberian penomoran indeks didasarkan pada urutan frekuensi kemunculan. Semakin sering kata itu muncul maka semakin tinggi urutannya. Kata <OOV> (Out Of Vocabulary) juga diberikan indexing untuk mempersiapkan kata-kata asing agar tetap bisa diolah. Contoh *Corpus Vocabulary Word Index* dapat dilihat pada Tabel 3.14. berikut:

**Tabel 3.14.** Contoh *Corpus Vocabulary Word Index*

Kata	Index
<OOV>	1
saham	2
harga	3
trading	4
beli	5
jual	6
.....	.....
gorengan	376

### 5.3. Corpus Integer Sequence

*Corpus Integer Sequence* adalah proses dimana tiap baris sequence pertanyaan yang sudah di tokenisasi tadinya dijadikan sebagai *Integer Sequence* berdasarkan *Corpus Vocabulary Word Index* yang sudah dibuat sebelumnya. Proses ini mengubah sequence pertanyaan yang semula bertipe string menjadi sequence pertanyaan berbentuk *integer*. Alasan dilakukannya proses ini sama seperti alasan dilakukannya proses *Label Encoding* yakni dikarenakan model tidak dapat mencerna data berbentuk string sehingga harus diubah dahulu menjadi integer agar dapat dimengerti model. Contoh penerapan *Corpus Integer Sequence* dapat dilihat pada Tabel 3.15. berikut:

**Tabel 3.15.** Contoh *Corpus Integer Sequence*

Sebelum <i>Corpus Integer Sequence</i>	Setelah <i>Corpus Integer Sequence</i>
['beli','saham']	[5,2]

## 6. Model Training Implementasi Algoritma LSTM

Dalam pembuatan model *LSTM* yang diimplementasikan dalam penelitian ini, terdapat beberapa layer yang digunakan, yakni:

6.1. *Embedding Layer*: Layer pertama dari model menggunakan layer embedding untuk merepresentasikan setiap kata dalam teks. Vector yang digunakan memiliki panjang 100 untuk menggambarkan setiap kata menurut kedekatan makna antar kata. Kata yang memiliki kedekatan makna yang dekat memiliki vektor yang berdekatan pula. Fungsi *Embedding layer* adalah untuk mempelajari hubungan antar kata agar dapat diproses dengan lebih efektif.

6.2. *LSTM Layer*: Layer selanjutnya adalah LSTM layer dengan 32 unit neuron.

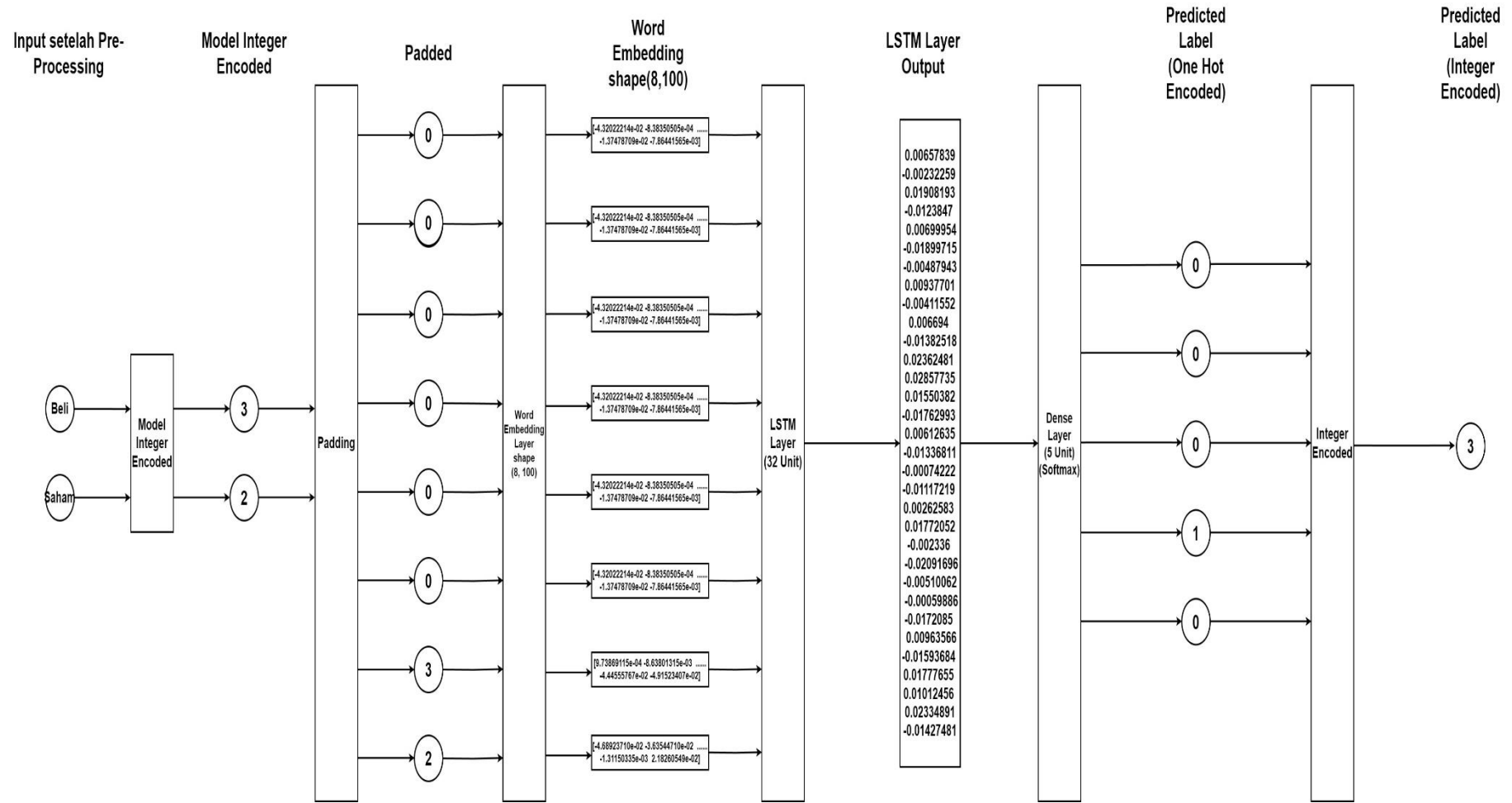
6.3. *Dense Layer*: Layer terakhir dari model adalah Dense Layer dengan jumlah neuron sebanyak jumlah kelas atau label yang ada yakni sebanyak 5 unit. Layer ini menggunakan aktivasi softmax, yang menghasilkan probabilitas untuk setiap kelas. Hal ini digunakan untuk melakukan klasifikasi multiclass.

## 7. Prediksi Data Testing

Data testing yang telah melalui tahap model encoding sebelumnya, diambil model Integer Sequencenya dan dimasukkan kedalam model untuk menguji model yang telah dilatih dalam memprediksi label dari data testing yang ada. Tingginya jumlah prediksi label data testing yang benar menunjukkan kelayakan model untuk digunakan.



Gambar 3.2. berikut merupakan arsitektur LSTM yang menunjukkan alur proses pemrediksian *label* pertanyaan menggunakan model *LSTM* mulai dari input pertanyaan, layer-layer yang dilalui, hasil output tiap layernya dan output akhirnya yakni *predicted label*:



Gambar 3.2. Arsitektur Model LSTM

### 8. *Integer Sequence Matching*

Setelah *predicted label* didapatkan maka dapat dilakukan matching pertanyaan terhadap pertanyaan corpus yang memiliki label *predicted label* menggunakan metode *Integer Sequence Matching*. *Integer Sequence Matching* merupakan metode matching yang peneliti kembangkan dengan modifikasi dari *Fuzzy String Matching*. Dimana pada *Fuzzy String Matching*, dilakukan perbandingan satu persatu huruf dalam keseluruhan kalimat untuk menghitung perbedaan berdasarkan penyisipan, penghapusan, atau penggantian huruf (Azni, 2021). Sedangkan dalam *Integer Sequence Matching* input pertanyaan pengguna dalam bentuk *corpus Integer Sequence* dicocokkan dengan *data corpus* berbentuk *corpus Integer Sequence* yang berlabel sama dengan label hasil prediksi untuk menemukan jawaban yang sesuai. Dengan demikian, matching tidak dilakukan per huruf seperti pada *Fuzzy String Matching* melainkan per kata dalam bentuk *Integer Sequence* sehingga dinamakan *Integer Sequence Matching*. *Integer Sequence Matching* menghasilkan kecocokan yang berbeda-beda dan yang dipilih adalah jawaban dari matching yang mempunyai kecocokan tertinggi. Untuk menghitung kecocokan, digunakan Rumus 3.1. yakni rumus kecocokan *Integer Sequence* berikut:

$$\text{Kecocokan} = (\text{Jumlah } Integer Sequence Corpus \text{ Input Pengguna yang terdapat di dalam } Integer Sequence Corpus \text{ Pertanyaan Corpus} / \text{Panjang } Integer Sequence Corpus \text{ Input Pengguna}) * 100\% \quad (3.1)$$

Contoh penerapan *Integer Sequence Matching* dapat dilihat pada Tabel 3.16. berikut:

**Tabel 3.16.** Contoh *Integer Sequence Matching*

Token input pertanyaan user ['beli','saham']		<i>Integer Sequence Corpus</i> Input pertanyaan user [5,2]		Predicted Label : 'Strategi'	
Baris	Pertanyaan Corpus	<i>Integer Sequence Corpus</i>	Jawaban Corpus	Corpus Label	Kecocokan
1)	['jual','saham']	[6,2]	Cara menjual saham: 1. Login ke Akun Trading.....	'Strategi'	$= \frac{1}{2} * 100\%$ $= 50\%$
2)	['beli','saham']	[5,2]	Cara Membeli Saham: 1. Pilih Perusahaan.....	'Strategi'	$= \frac{2}{2} * 100\%$ $= 100\%$

Dikarenakan *Integer Sequence Corpus* data baris kedua yang memiliki persen kecocokan yang paling tinggi maka *jawaban* yang diambil adalah *jawaban* dari baris data kedua yakni “Anda dapat membeli saham melalui perusahaan sekuritas atau broker saham”. Apabila terdapat beberapa pertanyaan dengan nilai kecocokan tertinggi yang sama maka jawaban yang diambil adalah jawaban dari pertanyaan corpus dengan panjang terpendek. Contoh penerapan *Integer Sequence Matching* dengan kecocokan tertinggi yang sama dapat dilihat pada Tabel 3.17. berikut:

**Tabel 3.17.** Contoh *Integer Sequence Matching* Kecocokan Tertinggi Yang Sama

Token input pertanyaan user ['beli', 'saham']		<i>Integer Sequence Corpus</i> Input pertanyaan user [5,2]		Predicted Label : ‘Strategi’	
Baris	Pertanyaan Corpus	<i>Integer Sequence Corpus</i>	Jawaban Corpus	Label Corpus	Kecocokan
1)	['beli', 'saham']	[5,2]	Cara Membeli Saham: 1. Pilih Perusahaan Sekuritas: Pilih perusahaan sekuritas yang terdaftar di OJK sebagai perantara .....	‘Strategi’	$= \frac{2}{2} * 100\%$ $= 100\%$
2)	['beli', 'saham', 'gorengan']	[5,2,376]	Cara Membeli Saham Gorengan: 1. Pilihlah saham gorengan yang memiliki volume tinggi dan .....	‘Strategi’	$= \frac{2}{2} * 100\%$ $= 100\%$

Dikarenakan kedua pertanyaan corpus memiliki kecocokan tertinggi yang sama yakni “100%” maka yang diambil ada pertanyaan baris ke 1 ['beli', 'saham'] dikarenakan memiliki panjang yang lebih pendek yakni “2” dibandingkan dengan pertanyaan baris kedua dengan panjang “3”. Tujuannya adalah untuk menyaring jawaban yang lebih tepat dari jawaban-jawaban yang sama-sama memiliki kecocokan tertinggi.

### 9. API Data Harga Saham

Untuk menampilkan harga saham harian digunakan API dari <https://goapi.id/> yang menyediakan API penarikan data harga saham gratis dengan delay harga selama 3 menit. Request url yang digunakan untuk melakukan penarikan data harga saham adalah `requests.get("https://api.goapi.id/v1/stock/idx/prices?api_key="+key_goAPI+"&symbols="+kode_saham)` yang mengembalikan data harga saham harian berupa *tanggal*, *kode saham*, *harga buka*, *harga tertinggi*, *harga terendah* dan *harga penutupan*. Data dikembalikan oleh API dalam bentuk JSON dan diolah untuk di tampilkan kepada pengguna.

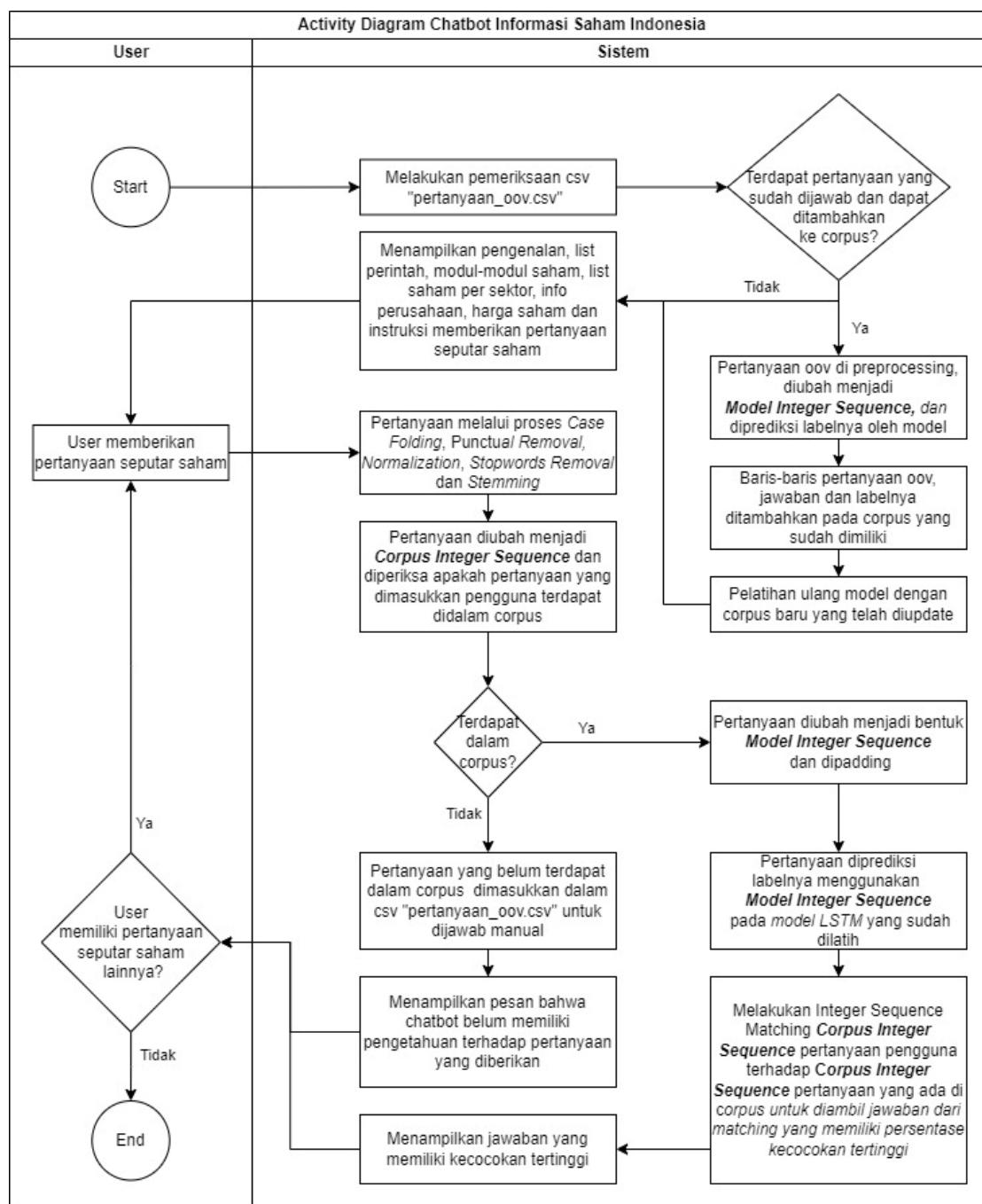
### 10. Output

*Output* merupakan tahap terakhir dimana jawaban terhadap pertanyaan pengguna yang sudah ditemukan dikembalikan dan ditampilkan kepada pengguna.

## 3.3. Activity Diagram Chatbot Informasi Saham Indonesia

Activity diagram *chatbot* informasi saham Indonesia dimulai dengan pengaktifan chatbot. Setelah diaktifkan *chatbot* mulai melakukan pemeriksaan csv "pertanyaan\_oov.csv" dan apabila terdapat pertanyaan oov yang sudah dijawab dan dapat ditambahkan, maka pertanyaan oov akan di preprocessing kemudian diprediksi label pertanyaan oov tersebut untuk ditambahkan ke corpus kemudian dilakukan pelatihan model ulang. Setelah proses ini selesai, maka chatbot dapat mulai digunakan oleh user. Dimulai dengan menampilkan ucapan selamat datang serta list perintah, modul-modul saham, daftar sektor dan sahamnya, info perusahaan, harga saham dan instruksi memberikan pertanyaan seputar saham sesuai perintah yang diberikan. Pertanyaan yang diberikan oleh pengguna masuk ke sistem dan dilakukan proses *preprocessing*. Setelah pertanyaan melewati proses *preprocessing* maka pertanyaan di ubah menjadi bentuk *corpus integer sequence* dan diperiksa apakah corpus memiliki pengetahuan atas pertanyaan tersebut. Apabila tidak, maka pertanyaan pengguna akan dimasukkan ke "pertanyaan\_oov.csv" untuk ditambahkan ke corpus nantinya setelah memiliki jawaban. Chatbot kemudian mengembalikan pesan bahwa chatbot belum memiliki pengetahuan atas pertanyaan tersebut. Apabila pertanyaan pengguna terdapat dalam corpus, maka pertanyaan yang sudah di *preprocessing* diubah menjadi *model integer sequence* serta di *padding*. Pertanyaan dalam bentuk *model integer sequence* kemudian diprediksi *label*-nya oleh *model LSTM* yang sudah dilatih. *Label* hasil

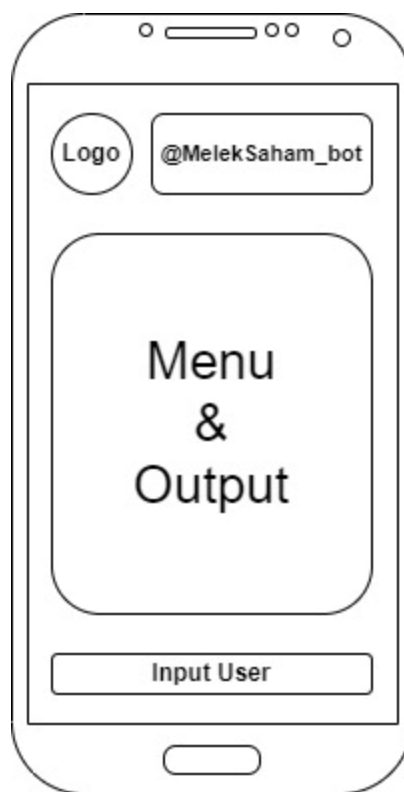
prediksi ini digunakan sebagai pedoman dalam *Integer Sequence Matching* dengan mencocokkan *Integer Sequence* pertanyaan terhadap *Integer Sequence* corpus yang berlabel sama dengan label hasil prediksi pertanyaan. Pertanyaan dalam corpus yang memiliki tingkat kecocokan tertinggi diambil jawabannya dan dikembalikan kepada user. Gambar 3.3 berikut menunjukkan activity diagram *chatbot* informasi saham Indonesia:



**Gambar 3.3.** Activity Diagram Chatbot Informasi Saham Indonesia

### 3.4. Perancangan Interface Sistem

*Interface Sistem* yang diterapkan dalam penelitian ini merupakan *interface* dari *Telegram*. Gambar 3.4. menunjukkan *rancangan interface sistem* yang dipakai:



**Gambar 3.4.** Interface Sistem

Penjelasan Gambar 3.4. mengenai komponen dari *Interface Sistem*:

1. *Logo*

*Logo MelekSaham* terletak disini.

2. *MelekSaham\_bot*

Username dari chatbot informasi saham Indonesia yang dibuat adalah *MelekSaham\_bot*. Pengguna yang ingin mengakses chatbot ini hanya perlu membuka telegram baik melalui aplikasi telegram langsung maupun dari website application telegram. Pengguna bisa mencari *MelekSaham\_bot* di kolom search dan *MelekSaham\_bot* dapat digunakan oleh pengguna.

3. *Menu & Output*

*Menu & Output* menampilkan menu perintah dan output dari perintah yang diberikan oleh pengguna yakni instruksi penggunaan, modul-modul saham yang

dapat dipelajari pengguna, list sektor dan sahamnya, informasi perusahaan, harga saham dan jawaban atas pertanyaan yang diberikan pengguna.

4. *Input User*

*Input User* merupakan kolom tempat pengguna memasukkan perintah yang diinginkan. Kolom ini juga merupakan tempat user memberikan input pertanyaan seputar saham yang diproses sistem untuk mengembalikan jawaban yang sesuai.



## **BAB IV**

### **IMPLEMENTASI DAN PENGUJIAN**

Pada bab ini membahas tentang implementasi Model LSTM dan *Integer Sequence Matching* pada sistem chatbot informasi saham Indonesia sesuai dengan penjabaran pada Bab 3 serta pengujian sistem yang dihasilkan.

#### **4.1. Implementasi Sistem**

Pengimplementasian sistem chatbot informasi saham Indonesia dengan algoritma Long Short Term Memory dan *Integer Sequence Matching* menggunakan perangkat keras serta perangkat lunak, diantaranya:

##### **4.1.1. Spesifikasi Perangkat Keras (Hardware)**

Spesifikasi perangkat keras yang digunakan pada penelitian yakni:

1. Processor Ryzen 5 3500U
2. Kapasitas penyimpanan Drive Solid-State (SSD) 512 GB.
3. Memory RAM 6 GB.

##### **4.1.2. Spesifikasi Perangkat Lunak (Software)**

Spesifikasi perangkat lunak yang digunakan pada penelitian yakni:

1. Sistem operasi Windows 10 Home 64-bit.
2. Python version 3.9
3. JetBrains PyCharm Community Edition 2020.2.4.
4. Google Colaboratory Jupyter Notebook
5. Telegram version 10.2.9

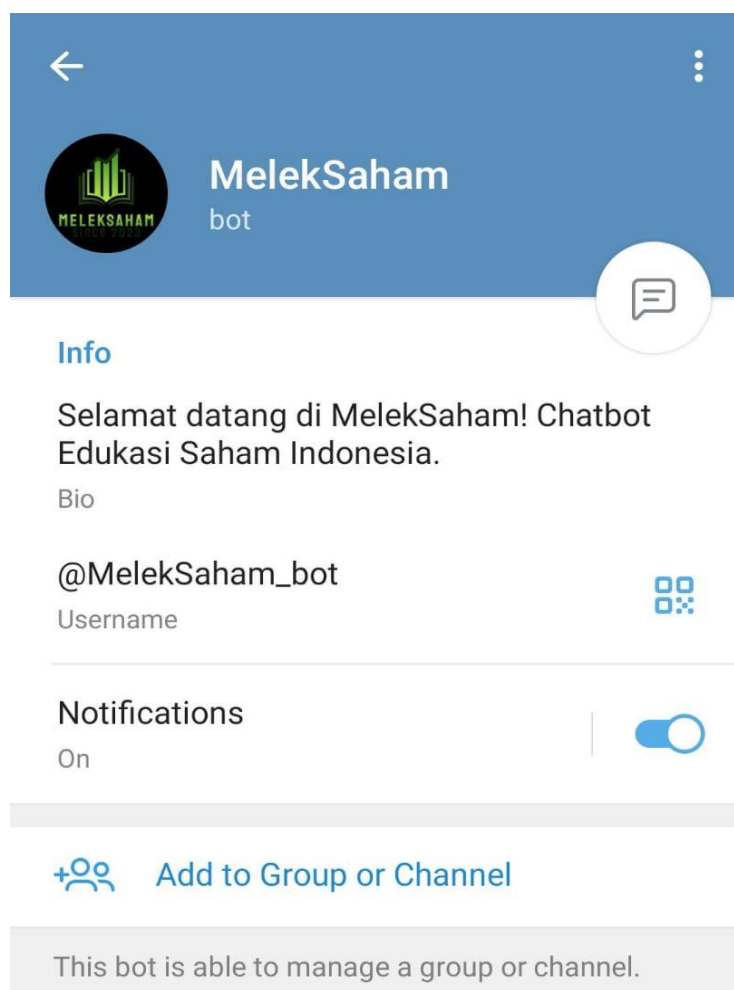
6. Library *Pandas* v.1.2.2, *Numpy* v.1.24.2, *Nltk* v.3.7, *Sastrawi* v.1.0.1, *Sklearn* v.0.0, *Tensorflow* v.2.10.0, *Seaborn* v.0.11.1, *Matplotlib* v.3.3.4, *Requests* v.2.25.1, *Aiogram-telegram-bot* v.2.25.1

## 4.2. Implementasi Interface

Implementasi rancangan *Interface Sistem* yang ditunjukkan pada Bab 3 sebelumnya adalah sebagai berikut:

### 1. Tampilan Halaman Profil

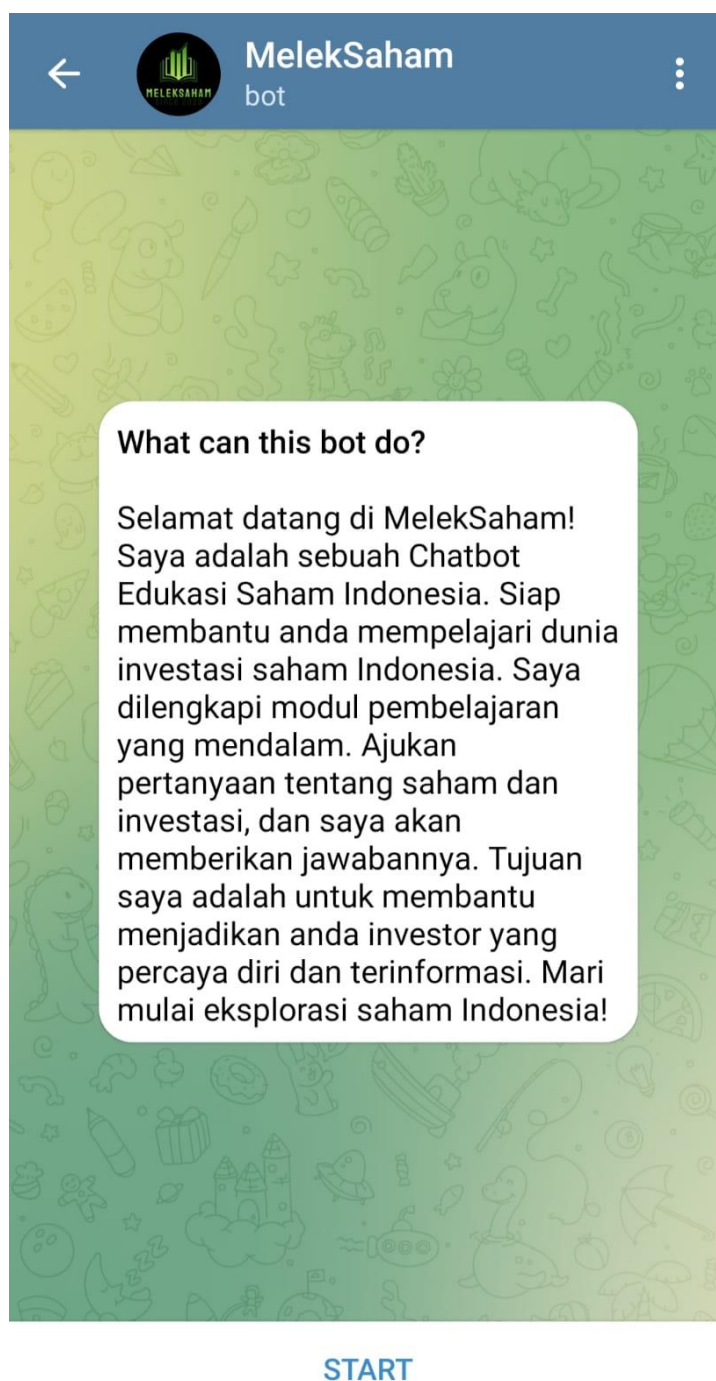
Pada halaman profil, ditampilkan informasi mengenai *chatbot* informasi saham Indonesia bernama “MelekSaham”. Informasi yang terdapat ppada halaman profil yakni *logo chatbot*, nama *chatbot*, informasi pendek tentang *chatbot*, dan *username chatbot* yang dapat dicari pengguna untuk menggunakan *chatbot*. Gambar 4.1 berikut menunjukkan tampilan halaman profil *chatbot*:



**Gambar 4.1.** Tampilan Halaman Profil

## 2. Tampilan Halaman Pembuka

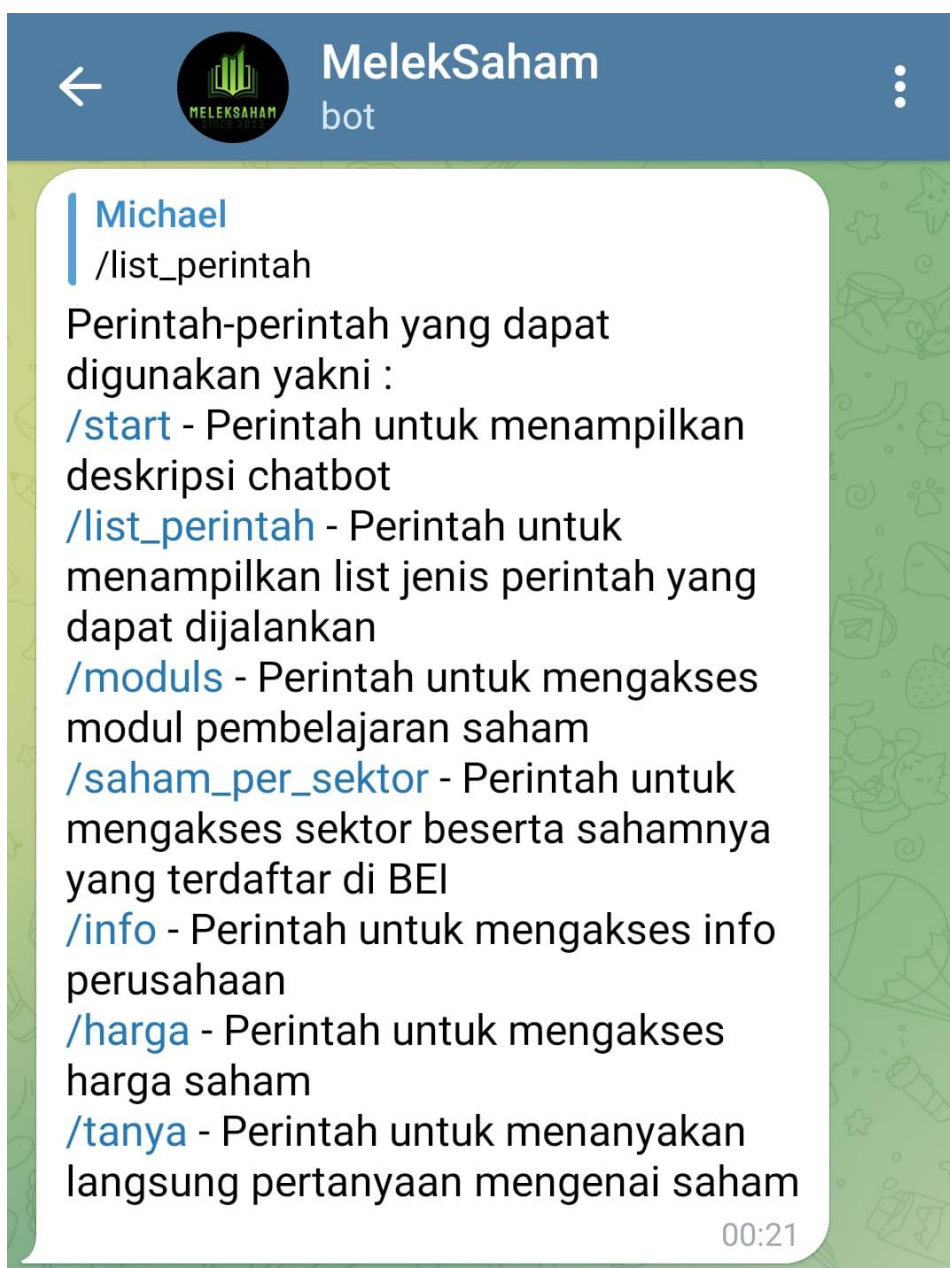
Pada halaman pembuka, saat pengguna pertama kali menggunakan *chatbot* maka ditampilkan deskripsi umum tentang chatbot dan tombol start untuk mulai memberikan perintah kepada *chatbot*. Gambar 4.2 berikut menunjukkan tampilan halaman pembuka *chatbot*:



**Gambar 4.2.** Tampilan Halaman Pembuka

### 3. Tampilan Halaman Menu Utama

Pada tampilan halaman menu utama, setelah chatbot dijalankan atau setelah perintah /list\_perintah diberikan, ditampilkan pilihan-pilihan perintah yang dapat dipilih untuk dijalankan chatbot. Gambar 4.3 berikut menunjukkan tampilan halaman menu utama:

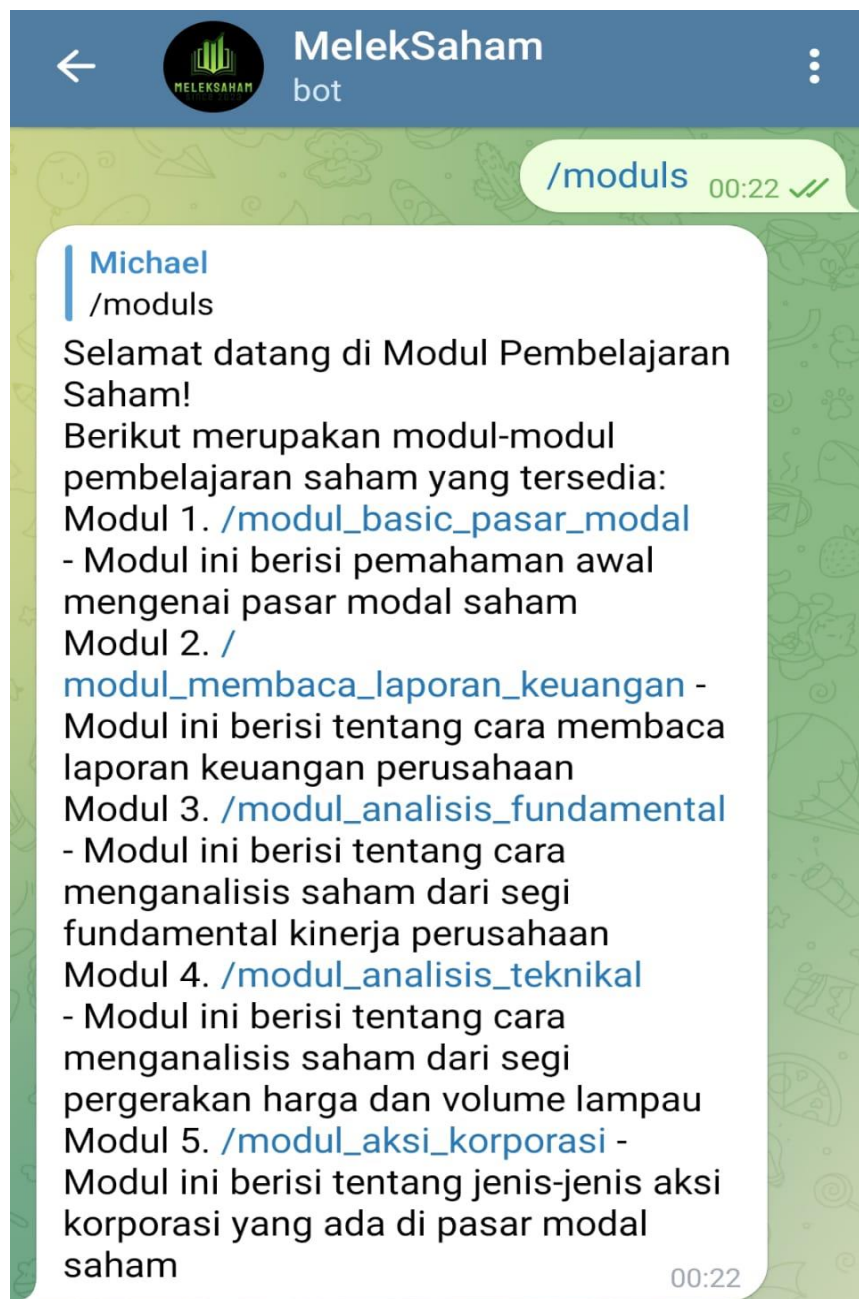


**Gambar 4.3.** Tampilan Halaman Menu Utama

Adapula pilihan perintah-perintah pada halaman menu utama yang dapat dijalankan *chatbot* yakni:

a. /moduls

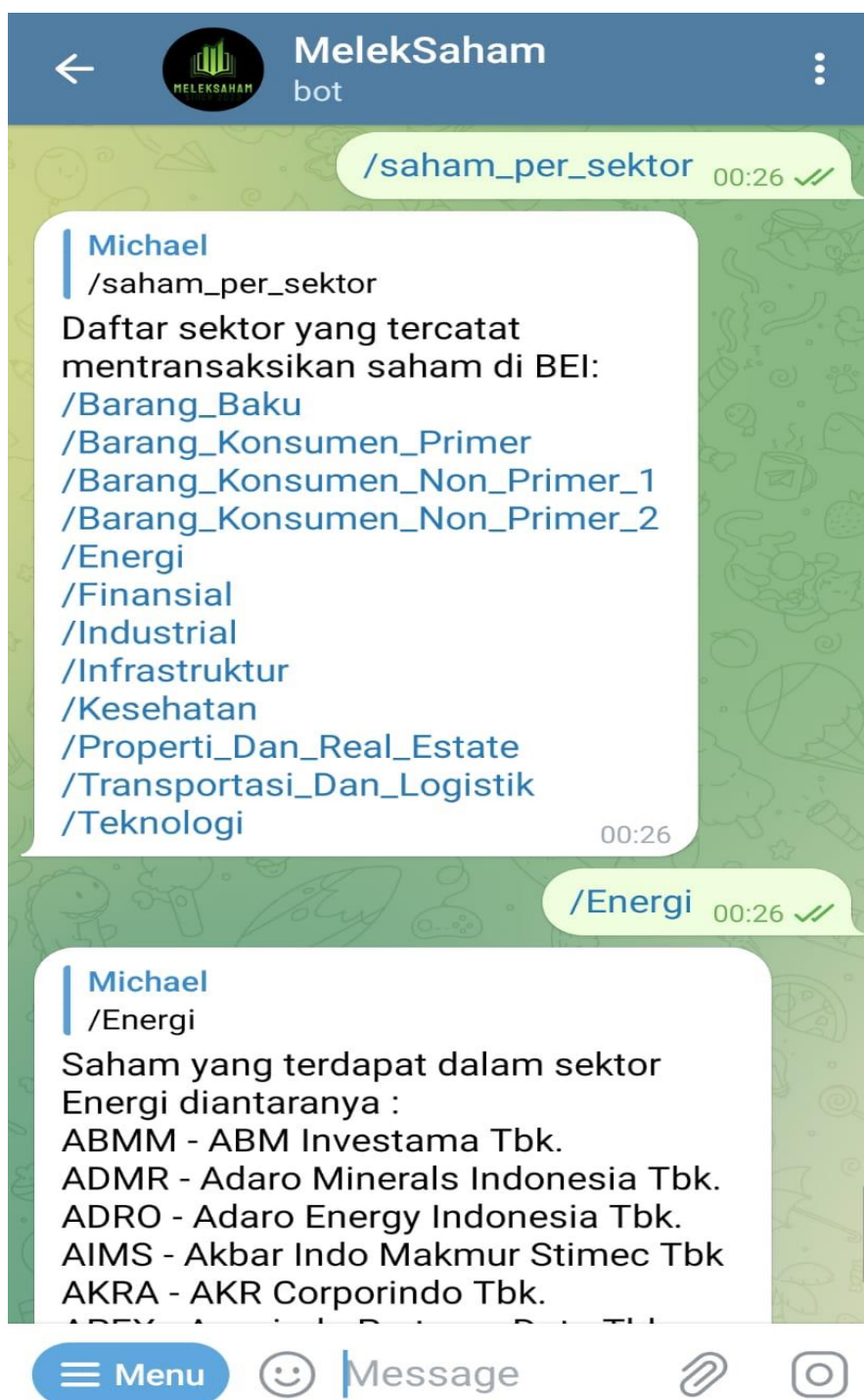
/moduls adalah perintah untuk memunculkan modul-modul pembelajaran saham mulai dari modul pengenalan saham sampai modul analisa saham. Modul yang disusun berujuk pada materi-materi pembelajaran dari pihak yang berpengalaman yakni sekuritas dan seminar saham. Gambar 4.4 berikut menunjukkan tampilan modul saham:



**Gambar 4.4.** Tampilan Modul Saham

b. /saham\_per\_sektor

/saham\_per\_sektor adalah perintah untuk menampilkan 11 sektor saham beserta sahamnya yang terdaftar pada Bursa Efek Indonesia (BEI). Gambar 4.5 berikut menunjukkan tampilan sektor dan saham didalamnya:

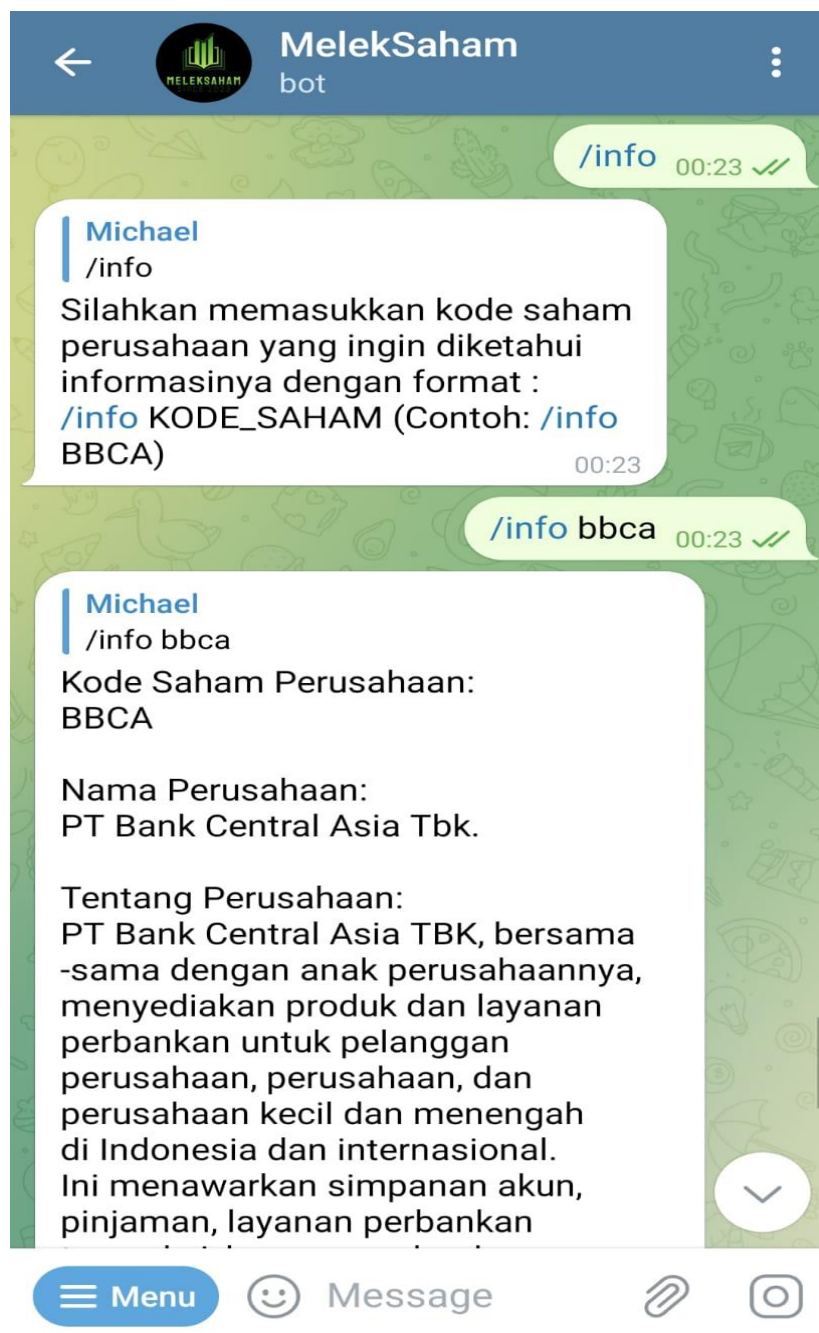


**Gambar 4.5.** Tampilan Daftar Sektor Dan Sahamnya



c. /info

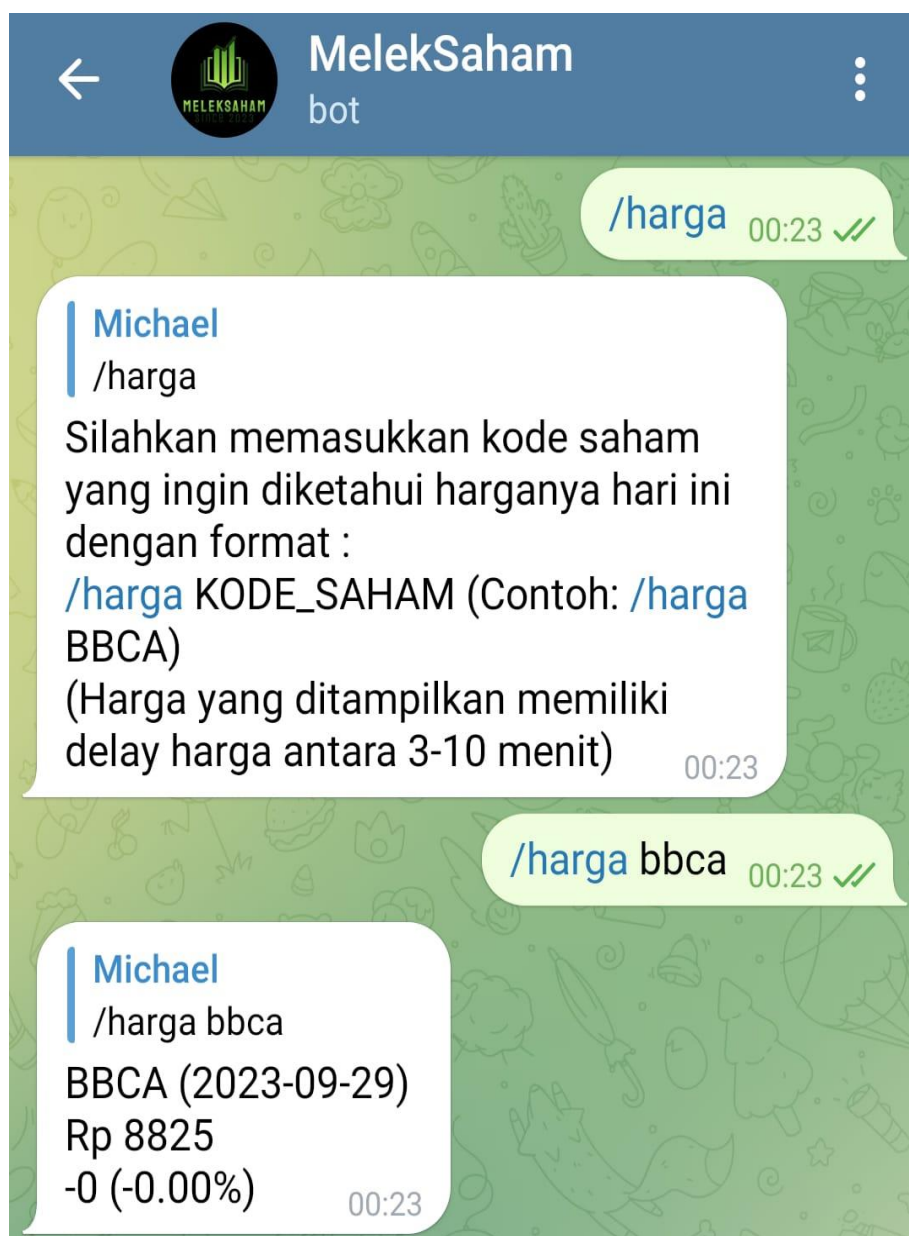
/info adalah perintah untuk menampilkan informasi detail mengenai perusahaan berdasarkan kode sahamnya diberikan dan kode saham tersebut terdaftar di Bursa Efek Indonesia (BEI). Gambar 4.6. berikut menunjukkan tampilan informasi detail perusahaan berdasarkan kode saham yang diberikan:



**Gambar 4.6.** Tampilan Informasi Perusahaan

## d. /harga

/harga merupakan perintah untuk menampilkan harga saham dari kode sahamnya yang diberikan dan kode saham tersebut terdaftar di Bursa Efek Indonesia (BEI). Harga saham diambil dari API saham Indonesia GoAPI.id, API tersebut merupakan API gratis sehingga terdapat delay keterlambatan harga 3 – 10 menit. Gambar 4.7. berikut menunjukkan tampilan harga saham berdasarkan kode saham yang diberikan:

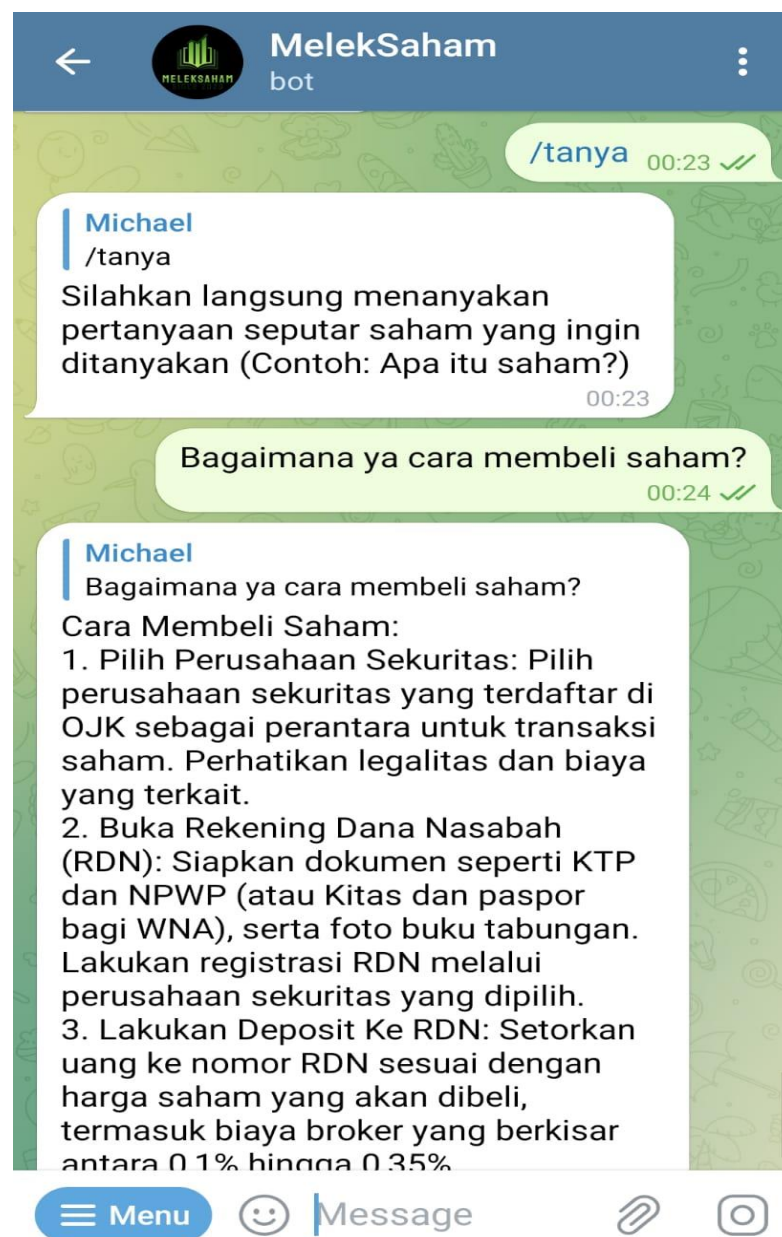


**Gambar 4.7.** Tampilan Harga Saham



e. /tanya

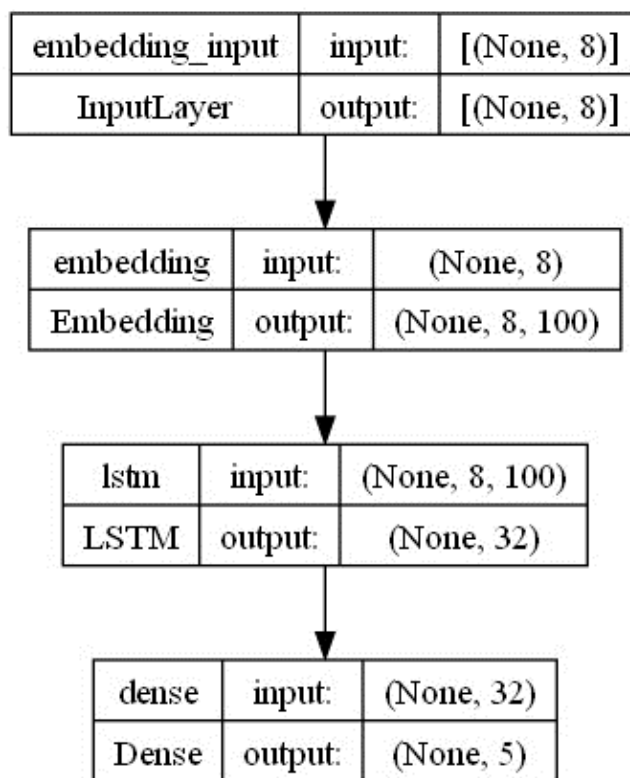
/tanya merupakan perintah untuk menanyakan hal seputar saham. Perintah /tanya bekerja sedikit berbeda dengan perintah lain dimana ketika perintah /tanya maka memberikan instruksi cara memberi pertanyaan. Dan untuk memberikan pertanyaan hanya perlu langsung memasukkan pertanyaan tanpa perlu menggunakan perintah. Adapula pertanyaan yang dapat diberikan merupakan pertanyaan seputar istilah, cara, *trading*, *investasi* dan *IPO*. Gambar 4.8 berikut menunjukkan tampilan tanya jawab saham:



**Gambar 4.8.** Tampilan Tanya Jawab Saham

### 4.3. Training Model LSTM

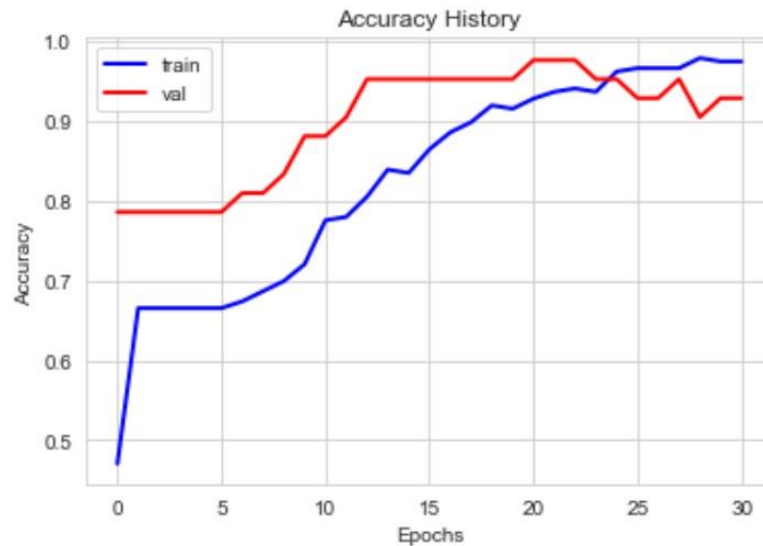
Model LSTM yang dilatih dan diimplementasikan ke *chatbot* terdiri dari 4 layer yakni *input layer*, *embedding layer*, *LSTM layer* dan *dense layer*. Pertanyaan data training dan validation dalam bentuk *Model Integer Sequence* yang sudah di padding menjadi panjang 8 (panjang pertanyaan terpanjang dalam corpus) memasuki input layer. Dari input layer, pertanyaan data training dan validation selanjutnya memasuki embedding layer. Embedding layer yang digunakan memiliki *vocab\_size* sebesar 341 (jumlah word index data training) dan *output\_dim* 100. Layer selanjutnya adalah *layer LSTM* dengan 32 *unit* dan diakhiri dengan *dense layer* aktivasi *softmax* dengan 5 *unit* yang mana merupakan jumlah label yang ada. *Plot model LSTM* yang digunakan ditunjukkan pada Gambar 4.9 berikut:



**Gambar 4.9.** Plot Model LSTM Yang Digunakan

Pelatihan model menggunakan *loss function categorical\_crossentropy*, *optimizer adam* dan *metrics accuracy*. Untuk mencegah overfitting, *callback earllystopping* di terapkan pada pelatihan model dengan *patience* 3 untuk memonitor *validation\_loss*. Sehingga apabila *validation\_loss* tidak mengalami perbaikan selama 3 *epoch* berturut maka pelatihan *model* dianggap selesai meskipun *epoch* belum mencapai *epoch* maksimum. Dalam pelatihan model *LSTM* untuk *chatbot MelekSaham*,

terjadi 31 kali *epoch*. Histori performa *akurasi* pelatihan model dan *loss* pelatihan model terhadap data *training* dan data *validation* ditunjukkan oleh gambar 4.10 dan 4.11 berikut:



**Gambar 4.10.** Histori Performa Akurasi Pelatihan Model



**Gambar 4.11.** Histori Performa Loss Pelatihan Model

Dari grafik histori performa *akurasi* dan *loss* pelatihan *model* yang telah dilakukan didapatkan hasil performa untuk data *training* berupa akurasi sebesar 97.46% dan *loss* sebesar 12.36%. Dan untuk data *validation* didapatkan akurasi sebesar 92.86% dan *loss* sebesar 18.62% dari 31 epoch.

Pemilihan ukuran split data *training-testing-validation* didasarkan pada penelitian oleh (Tolevea, 2021) dimana dari hasil penelitiannya disimpulkan ukuran

split yang optimal adalah 60% / 70% / 80% untuk data training, 10% / 20% untuk data validation, and 10% / 20% untuk data testing. Sedangkan parameter *embedding unit*, *LSTM unit*, serta *patience* untuk *earlystopping* didasarkan pada beberapa penelitian terdahulu terkait LSTM model yang disertai *trial and error* untuk menghasilkan *train accuracy*, *train loss*, *validation accuracy* dan *validation loss* terbaik sebagai tolak ukur performa model sebagaimana yang disebutkan oleh Chi et al. (2022) dalam penelitian mereka. Tabel 4.1 dan tabel 4.2 menunjukkan hasil trial dan error tersebut dalam menghasilkan model dengan performa terbaik:

**Tabel 4.1.** Tabel Hasil Percobaan Parameter Tuning

Train Size	Validation Size	Test Size	Embedding Unit	LSTM Unit	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
80%	10%	10%	100	32	0.1241	0.9685	0.285	0.875
<b>70%</b>	<b>10%</b>	<b>20%</b>	<b>100</b>	<b>32</b>	<b>0.1236</b>	<b>0.9746</b>	<b>0.1862</b>	<b>0.9286</b>
70%	20%	10%	100	32	0.1647	0.948	0.2732	0.8667
60%	20%	20%	100	32	0.1449	0.9446	0.4004	0.8644
80%	10%	10%	100	16	0.1572	0.9746	0.1884	0.9048
70%	10%	20%	100	16	0.1261	0.9725	0.2666	0.8929
70%	20%	10%	100	16	0.2266	0.948	0.2816	0.9222
60%	20%	20%	100	16	0.1921	0.9722	0.258	0.9222
80%	10%	10%	50	16	0.2044	0.9534	0.1949	0.9048
70%	10%	20%	50	16	0.1344	0.9775	0.2702	0.8929
70%	20%	10%	50	16	0.2715	0.94	0.2446	0.9556
60%	20%	20%	50	16	0.1656	0.9788	0.2909	0.8814
80%	10%	10%	50	32	0.1347	0.9703	0.1936	0.9048
70%	10%	20%	50	32	0.1322	0.973	0.3518	0.8928
70%	20%	10%	50	32	0.1783	0.948	0.256	0.8667
60%	20%	20%	50	32	0.1311	0.9703	0.3501	0.8814

**Tabel 4.2.** Tabel Hasil Percobaan Patience

Patience	Epoch	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
<b>3</b>	<b>31</b>	<b>0.1236</b>	<b>0.9746</b>	<b>0.186</b>	<b>0.9286</b>
5	34	0.0989	0.9746	0.227	0.9286
10	38	0.0628	0.9915	0.251	0.9286
20	55	0.0343	0.9915	0.2959	0.9286
30	61	0.0309	0.9915	0.2872	0.9048
40	77	0.0196	0.9915	0.2909	0.881
50	78	0.018	0.9915	0.291	0.8667

Dari hasil percobaan *parameter tuning* dan *patience* pada tabel 4.1 dan 4.2, train test split dengan rasio 70% *data training* 10% *data validation* dan 20% *data testing*, 100 *embedding unit*, 32 *LSTM unit*, dan *patience* sebesar 3 memiliki *train loss*, *train accuracy*, *validation loss* dan *validation accuracy* terbaik. Dimana *Train Accuracy*

adalah persentase akurasi prediksi terhadap *data training*, sementara *train loss* mengindikasikan sejauh mana model mampu memprediksi label pada *data training* dengan meminimalkan nilai loss. Sebaliknya, *validation accuracy* mengevaluasi kinerja model pada *data validation* yang tidak terlibat dalam training model, dan *validation loss* mengukur loss pada data validasi.

#### 4.4. Pengujian Model LSTM

Setelah pelatihan model selesai, maka dilakukan pengujian model *LSTM* terhadap 70 baris *data testing* yang sudah dipisahkan terlebih dahulu sebelum pelatihan *model* untuk menguji performa model terhadap data yang belum dikenalnya. Pengujian model LSTM terhadap *data testing* ditunjukkan pada Tabel 4.3 berikut:

**Tabel 4.3.** Tabel Pengujian Model Terhadap Data Testing

Baris	Pertanyaan	Stemmed	Model Sequence	Predicted Label	True Label
1	Bagaimana memilih saham yang cocok investasi berjangka?	pilih saham cocok investasi jangka	[16, 2, 1, 3, 33]	Investasi	Investasi
2	Bagaimana pengaruh berita dan peristiwa ekonomi harga saham?	pengaruh berita peristiwa ekonomi harga saham	[77, 315, 1, 306, 5, 2]	Istilah	Strategi
3	Apa itu BEI	bei	[290]	Istilah	Istilah
4	Bagaimana memahami indikator teknikal trading saham?	paham indikator teknikal trading saham	[32, 103, 22, 9, 2]	Trading	Trading
5	Apa itu dividen?	dividen	[15]	Istilah	Istilah
6	Apa itu Saham small cap?	saham small cap	[2, 1, 243]	Istilah	Istilah
7	Apa itu short selling?	short selling	[129, 130]	Istilah	Istilah
8	Apa itu Bearish	bearish	[1]	Istilah	Istilah

**Tabel 4.3.** Tabel Pengujian Model Terhadap Data Testing (Lanjutan)

Baris	Pertanyaan	Stemmed	Model Sequence	Predicted Label	True Label
9	Apa itu IDX Small-Mid Cap?	idx small mid cap	[11, 1, 1, 243]	Istilah	Istilah
10	Apa yang dimaksud dengan Earning per Share (EPS) dalam dunia saham?	earning share eps dunia saham	[270, 229, 1, 12, 2]	Istilah	Istilah
11	Bagaimana cara membaca dan memahami laporan laba rugi perusahaan?	baca paham lapor laba rugi usaha	[106, 32, 19, 40, 1, 13]	Strategi	Strategi
12	Bagaimana diversifikasi portofolio?	diversifikasi portofolio	[134, 135]	Istilah	Strategi
13	Apa itu IDX Property?	idx property	[11, 1]	Istilah	Istilah
14	Bagaimana cara menjual saham?	jual saham	[61, 2]	Strategi	Strategi
15	Apa itu broker?	broker	[264]	Istilah	Istilah
16	Bagaimana cara mengenali saham yang mengalami manipulasi pasar?	nali saham alami manipulasi pasar	[43, 2, 66, 1, 10]	Strategi	Strategi
17	Analisis teknikal yang berkelanjutan dalam trading saham?	analisis teknikal lanjut trading saham	[4, 22, 1, 9, 2]	Trading	Trading
18	Apa itu saham high beta?	saham high beta	[2, 265, 1]	Istilah	Istilah
19	Apa itu IDX Agriculture?	idx agriculture	[11, 1]	Istilah	Istilah
20	Apa itu saham preferen?	saham preferen	[2, 69]	Istilah	Istilah
21	Apa itu IDX Syariah?	idx syariah	[11, 1]	Istilah	Istilah
22	Apa saja itu jenis utang berdasarkan karakteristiknya ?	jenis utang dasar karakteristi k	[39, 63, 105, 145]	Strategi	Istilah

**Tabel 4.3.** Tabel Pengujian Model Terhadap Data Testing (Lanjutan)

Baris	Pertanyaan	Stemmed	Model Sequence	Predicted Label	True Label
23	Bagaimana cara menggunakan analisis sentimen dalam investasi saham?	analisis sentimen investasi saham	[4, 286, 3, 2]	Investasi	Investasi
24	Apa itu IHSG (Indeks Harga Saham Gabungan)?	ihsg indeks harga saham gabung	[284, 53, 5, 2, 1]	Istilah	Istilah
25	Apa yang dimaksud dengan analisis teknikal dalam investasi saham?	analisis teknikal investasi saham	[4, 22, 3, 2]	Istilah	Investasi
26	Bagaimana cara memahami laporan keuangan perusahaan?	paham lapor uang usaha	[32, 19, 21, 13]	Strategi	Strategi
27	Apa itu Profitability Ratio?	profitability ratio	[1, 7]	Istilah	Istilah
28	Dari mana saja pendapatan bisa diperoleh untuk mencapai kebebasan finansial?	dapat oleh capai bebas finansial	[1, 1, 101, 102, 44]	Strategi	Strategi
29	Apa itu saham cyclical?	saham cyclical	[2, 1]	Istilah	Istilah
30	Apa yang dimaksud dengan indeks saham?	indeks saham	[53, 2]	Istilah	Istilah
31	Bagaimana cara mengukur kinerja investasi saham?	ukur kerja investasi saham	[25, 107, 3, 2]	Strategi	Investasi
32	Apa yang dimaksud dengan analisis fundamental dalam investasi saham?	analisis fundamental investasi saham	[4, 28, 3, 2]	Istilah	Investasi

**Tabel 4.3.** Tabel Pengujian Model Terhadap Data Testing (Lanjutan)

Baris	Pertanyaan	Stemmed	Model Sequence	Predicted Label	True Label
33	Apa yang dimaksud dengan analisis teknikal dalam investasi saham?	analisis teknikal investasi saham	[4, 22, 3, 2]	Istilah	Investasi
34	EPS dikatakan baik apabila?	eps	[1]	Istilah	Investasi
35	Bagaimana cara mengenali saham dengan potensi pertumbuhan yang tinggi dalam investasi saham?	nali saham potensi tumbuh investasi saham	[43, 2, 56, 93, 3, 2]	Investasi	Investasi
36	Apa yang harus dipertimbangkan oleh investor sebelum membeli saham dalam IPO?	timbang investor beli saham ipo	[90, 55, 23, 2, 8]	IPO	IPO
37	Apa itu saham IPO lock-up?	saham ipo lock up	[2, 8, 1, 316]	IPO	Istilah
38	Apa arti dari kebebasan finansial?	arti bebas finansial	[1, 102, 44]	Strategi	Istilah
39	Bagaimana membedakan saham blue chip mid cap dan small cap?	beda saham blue chip mid cap small cap	[114, 2, 14, 17, 1, 243, 1, 243]	Istilah	Istilah
40	Apa saja komponen laporan keuangan dalam investasi?	komponen lapor uang investasi	[84, 19, 21, 3]	Istilah	Investasi
41	Apakah ada metode strategi untuk mengidentifikasi saham yang sedang dijual dengan harga diskon?	metode strategi identifikasi saham jual harga diskon	[1, 108, 26, 2, 61, 5, 1]	Strategi	Strategi



**Tabel 4.3.** Tabel Pengujian Model Terhadap Data Testing (Lanjutan)

Baris	Pertanyaan	Stemmed	Model Sequence	Predicted Label	True Label
42	Apa itu Swing Trading?	swing trading	[1, 9]	Istilah	Istilah
43	Apa itu trading saham?	trading saham	[9, 2]	Istilah	Istilah
44	Bagaimana pengaruh berita dan rumor terhadap harga saham dalam investasi saham?	pengaruh berita rumor harga saham investasi saham	[77, 315, 1, 5, 2, 3, 2]	Investasi	Investasi
45	Apa yang dimaksud dengan analisis sektoral?	analisis sektoral	[4, 1]	Istilah	Istilah
46	Apa yang dimaksud dengan saham dividend yield (DY)?	saham dividend yield dy	[2, 24, 125, 210]	Istilah	Istilah
47	Apa itu Bandar	bandar	[1]	Istilah	Istilah
48	Apa itu IPO (Initial Public Offering)?	ipo initial public offering	[8, 50, 29, 38]	IPO	Istilah
49	Apa yang dimaksud dengan dividen?	dividen	[15]	Istilah	Istilah
50	Berapa jenis opini auditor yang ada dan manakah yang wajib dipertimbangkan saat berinvestasi?	jenis opini auditor mana wajib timbang investasi	[39, 1, 269, 1, 1, 90, 3]	Investasi	Investasi
51	Untuk apa melakukan trading saham?	trading saham	[9, 2]	Istilah	Trading
52	Apa yang menjadi fokus dari analisis fundamental dalam berinvestasi saham?	fokus analisis fundamental investasi saham	[1, 4, 28, 3, 2]	Istilah	Investasi

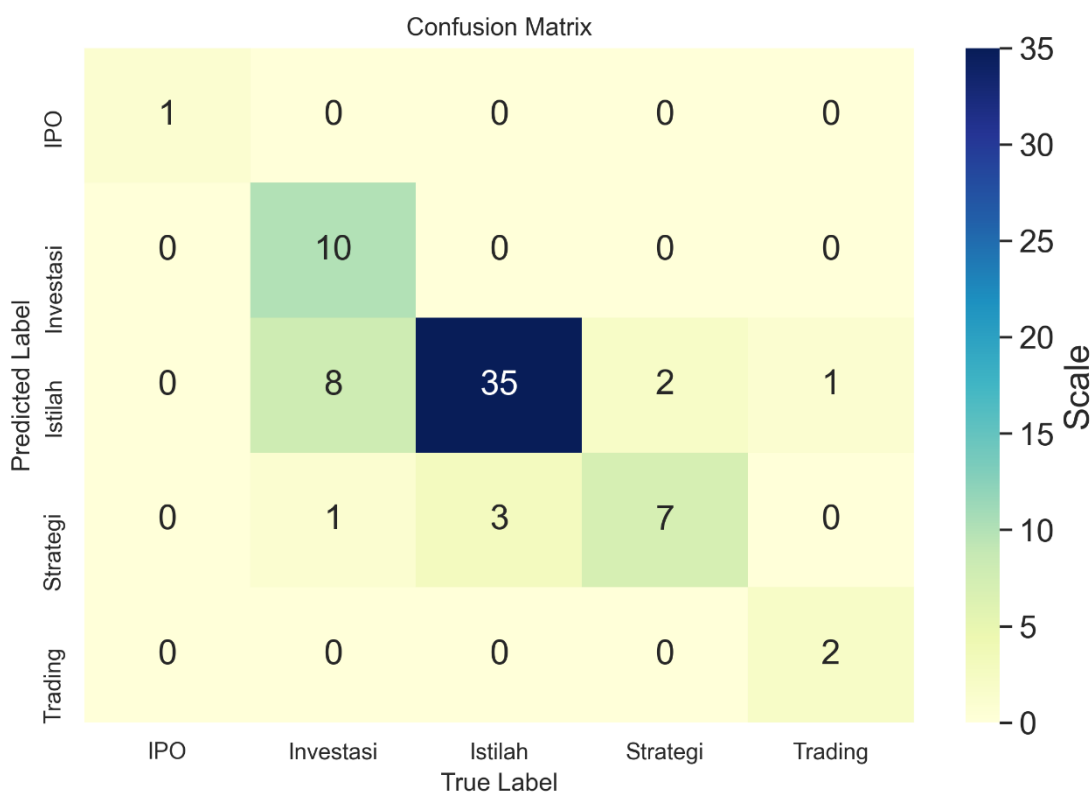
**Tabel 4.3.** Tabel Pengujian Model Terhadap Data Testing (Lanjutan)

Baris	Pertanyaan	Stemmed	Model Sequence	Predicted Label	True Label
53	DY dikatakan baik apabila?	dy	[210]	Istilah	Investasi
54	Contoh Indeks Saham dalam investasi saham?	contoh indeks saham investasi saham	[1, 53, 2, 3, 2]	Istilah	Investasi
55	Bagaimana cara mengukur tingkat risiko saham dalam investasi saham?	ukur tingkat risiko saham investasi saham	[25, 1, 6, 2, 3, 2]	Investasi	Investasi
56	Apa itu saham blue chip?	saham blue chip	[2, 14, 17]	Istilah	Istilah
57	Apa itu EPS (Earning Per Share)?	eps earning share	[1, 270, 229]	Istilah	Istilah
58	Apa itu depresiasi?	depresiasi	[1]	Istilah	Istilah
59	Apa itu Dividen Properti?	dividen properti	[15, 1]	Istilah	Istilah
60	Apa itu berinvestasi saham?	investasi saham	[3, 2]	Istilah	Istilah
61	Apa itu saham blue chip consortium?	saham blue chip consortium	[2, 14, 17, 1]	Istilah	Istilah
62	Apa perbedaan antara investasi jangka pendek dan jangka panjang?	beda investasi jangka pendek jangka panjang	[114, 3, 33, 1, 33]	Investasi	Investasi
63	Bagaimana cara mengidentifikasi saham overvalued (bernilai terlalu tinggi)?	identifikasi saham overvalued nila	[26, 2, 1, 1]	Strategi	Strategi
64	DER dikatakan baik apabila?	der	[144]	Istilah	Investasi
65	Saham adalah?	saham	[2]	Istilah	Istilah
66	Apa itu lot dalam perdagangan saham di BEI?	lot dagang saham bei	[1, 81, 2, 290]	Istilah	Istilah

**Tabel 4.3.** Tabel Pengujian Model Terhadap Data Testing (Lanjutan)

Baris	Pertanyaan	Stemmed	Model Sequence	Predicted Label	True Label
67	Apa itu Lot?	lot	[1]	Istilah	Istilah
68	Apa keuntungan berinvestasi dalam reksa dana?	untung investasi reksa dana	[100, 3, 112, 62]	Investasi	Investasi
69	Apa yang harus dilakukan jika terjadi kegagalan dalam investasi saham?	gagal investasi saham	[1, 3, 2]	Investasi	Investasi
70	Apa itu IDX80?	idx80	[1]	Istilah	Istilah

Kinerja model dalam memprediksi label *data testing* disajikan dalam bentuk confusion matrix pada Gambar 4.12 berikut:

**Gambar 4.12.** Confusion Matrix Data Testing

Dengan menggunakan rumus 4.1 yakni rumus *precision*, rumus 4.2 yakni rumus *recall*, rumus 4.3 yakni rumus *F1-score*, dan rumus 4.4 yakni rumus *accuracy* berikut:

#### 1. Precision

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4.1)$$

## 2. Recall

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4.2)$$

## 3. F1-Score

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.3)$$

## 4. Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.4)$$

Dihasilkan nilai precision, recall dan F1-score untuk tiap prediksi label *data testing* yang terdapat dalam *confusion matrix* seperti yang ditunjukkan pada Tabel 4.2 berikut:

**Tabel 4.4.** Tabel Precision, Recall Dan F1-Score Prediksi *Data Testing*

Label	Precision	Recall	F1-Score
IPO	1	1	1
Investasi	1	0.55	0.71
Istilah	0.76	0.92	0.83
Strategi	0.7	0.77	0.73
Trading	1	0.66	0.8

Dilihat dari tabel 4.4 performa chatbot dalam memprediksi data *testing* mendapatkan rata-rata *accuracy* sebesar 80%, rata-rata *precision* sebesar 89%, rata-rata *recall* sebesar 78% dan rata-rata *F1-Score* sebesar 82%.

#### 4.5. Pengujian Sistem

Pada tahap pengujian sistem, dilakukan pengujian keakuratan dan kecepatan sistem dalam mengembalikan jawaban yang tepat terhadap 10 pertanyaan acak selayaknya dalam penelitian Azni, (2021). Hasil pengujian sistem implementasi *Long Short-Term Memory* dan *Integer Sequence Matching* ditunjukkan pada Tabel 4.5 berikut:

**Tabel 4.5.** Tabel Pengujian Sistem

Nomor	Pertanyaan	Jawaban	Kecocokan	Waktu
1	Apa itu saham?	Saham adalah instrumen investasi berupa surat kepemilikan ...	Cocok	0.77 Detik
2	Bagaimana cara membeli saham?	Cara Membeli Saham: 1. Pilih Perusahaan Sekuritas: ...	Cocok	0.75 Detik

**Tabel 4.5.** Tabel Pengujian Sistem (Lanjutan)

Nomor	Pertanyaan	Jawaban	Kecocokan	Waktu
3	Tujuan IPO adalah?	Tujuan utama dari sebuah IPO adalah untuk mengumpulkan dana ...	Cocok	1.12 Detik
4	Apa perbedaan antara investasi jangka pendek dan jangka panjang?	Investasi jangka pendek biasanya berlangsung ...	Cocok	0.89 Detik
5	Bagaimana cara mengenali saham yang mengalami breakout dalam trading saham?	Mengenali saham yang mengalami breakout ...	Cocok	1.17 Detik
6	Apa itu moving average?	Moving Average (MA) adalah indikator teknikal yang menghitung ...	Cocok	1.03 Detik
7	Apa itu dividen?	Dividen adalah bagian laba perusahaan pada ...	Cocok	0.78 Detik
8	Bagaimana cara mengelola emosi saat beli saham?	Untuk mengelola emosi saat berinvestasi saham penting untuk memiliki rencana yang ...	Cocok	0.78 Detik
9	Apa yang harus dilakukan jika terjadi kejatuhan harga saham?	Jika terjadi kejatuhan pasar saham penting untuk tetap tenang dan tidak panik. Evaluasi ulang ...	Cocok	0.81 Detik
10	Bagaimana harga saham dalam IPO ditentukan?	Menentukan waktu yang tepat untuk membeli atau menjual saham...	Tidak Cocok	0.83 Detik

Hasil pengujian pada 10 pertanyaan menghasilkan 9 pertanyaan yang dijawab dengan benar dan 1 pertanyaan yang dijawab dengan salah. Sehingga didapatkan akurasi pengujian sistem sebesar  $\frac{9}{10} \times 100\%$  yakni 90%. Dan dari pengujian sistem didapatkan rata-rata kecepatan sistem dalam menghasilkan jawaban adalah 0.893 Detik. Terjadinya ketidak cocokan dalam pengembalian jawaban baris ke 10 dikarenakan kesalahan dalam prediksi label yang seharusnya “IPO” akan tetapi diprediksi sebagai “Strategi”.

Pengujian sistem untuk mengetahui seberapa baik dan sesuai sistem yang dibuat dilakukan terhadap pihak umum sebagai pengguna utama dari chatbot “MelekSaham” melalui penggunaan chatbot dan pengisian kuisioner penilaian. Pengujian sistem yang dilakukan meliputi pengujian dua fitur yakni fitur tambahan dan fitur tanya jawab pertanyaan seputar saham yang merupakan hasil implementasi model *Long Short-Term Memory* dan *Integer Sequence Matching*.

Fitur tambahan terdiri dari:

1. Instruksi penggunaan
2. Modul saham
3. List sektor beserta sahamnya
4. Informasi perusahaan
5. Harga saham

Adapula teknik sampling yang digunakan adalah teknik *purposive sampling*. Dimana menurut Fauzy (2014), *purposive sampling* merupakan metode penentuan sampel yang dilakukan dengan memilih sampel berdasarkan pengetahuan peneliti terkait tujuan atau masalah penelitian. Sampel yang dipilih dianggap sesuai untuk penelitian dan dapat memberikan informasi yang relevan karena mempunyai karakteristik, ciri, kriteria, atau sifat tertentu. Dengan demikian, pengambilan sampelnya dilakukan tidak secara acak.

Jenis *purposive sampling* yang digunakan adalah *purposive sampling homogen* karena teknik ini melakukan sampling pada kelompok yang memiliki karakteristik sama, dalam penelitian ini berupa karakteristik belum pernah berinvestasi saham Indonesia, dan karakteristik pernah berinvestasi saham Indonesia. Dengan memilih sampel *homogen*, kita dapat lebih mendalam memahami pandangan, sikap, dan faktor-faktor yang memengaruhi orang yang belum pernah berinvestasi saham dan orang yang pernah berinvestasi saham (Sugiyono, 2014). Pemilihan jumlah responden orang yang

belum pernah berinvestasi saham dan orang yang pernah berinvestasi saham didasarkan pada penelitian oleh Sugiyono (2014) yang menyebutkan jumlah sampel minimal yang layak dalam *purposive sampling homogen* adalah 30 orang.

Untuk mendapatkan perspektif yang mendalam dan informasi yang lebih kredibel terkait topik penelitian maka diperlukan pakar (Sugiyono, 2014). Dalam pemilihan pakar saham, digunakan teknik *purposive sampel ahli (Expert Sampling)* dimana peneliti memilih responden yang dianggap sebagai ahli atau pakar dalam bidang tertentu yang menjadi fokus penelitian. Pemilihan responden *expert sampling* dilakukan berdasarkan pengetahuan, keahlian, dan pengalaman responden ahli yang dianggap dapat memberikan wawasan mendalam dan kualitatif terkait dengan tujuan penelitian. Pemilihan jumlah responden *expert sampling* didasarkan pada penelitian oleh Sugiyono (2014), yang menyebutkan jumlah sampel minimal yang layak dalam *expert sampling* adalah 1 orang.

Sehingga pihak umum yang terlibat dalam pengujian dan penilaian, yakni :

1. 30 orang yang belum pernah berinvestasi saham Indonesia
2. 30 orang yang pernah berinvestasi saham Indonesia
3. 4 orang pakar saham Indonesia

Dan pakar saham yang terlibat dalam pengujian dan penilaian, yakni :

1. Fransiskus Wiguna (pialang Semesta sekuritas & mentor komunitas saham Indonesia)
2. Adhim Syafiq (pialang Mirae sekuritas dan mentor komunitas saham Indonesia)
3. Om Silent (pialang MNC sekuritas dan mentor komunitas saham Indonesia)
4. Jasni (pialang Trimegah sekuritas dan mentor komunitas saham Indonesia)

Para pengisi kuisioner diminta untuk mencoba chatbot “MelekSaham” dan diminta untuk memberikan penilaian berdasarkan skala *likert* dimana skala *likert* melakukan penilaian dengan skala dari skala 1 (sangat negatif) sampai 5 (sangat positif) dimana menurut Sugiyono (2014), skala *likert* digunakan untuk mengukur sikap, pendapat dan persepsi seorang atau kelompok orang terhadap sesuatu hal. Penilaian dilakukan pada dua aspek yakni *understandability* dan *functionality* (Ribalta, 2014). Penjelasan aspek *understandability* dan aspek *functionality* adalah sebagai berikut:

1. *Understandability*

*Understandability* merupakan aspek yang menunjukkan tingkat pemahaman pengguna terhadap output yang dikembalikan oleh chatbot. Adapun terdapat

lima skala penilaian yang digunakan aspek *understandability* pada penilaian yakni:

- Skala 1 : menyatakan output kembalian *chatbot* sangat sulit dipahami.
- Skala 2 : menyatakan output kembalian *chatbot* sulit dipahami.
- Skala 3 : menyatakan output kembalian *chatbot* cukup mudah dipahami.
- Skala 4 : menyatakan output kembalian *chatbot* mudah dipahami.
- Skala 5 : menyatakan output kembalian *chatbot* sangat mudah dipahami.

## 2. *Functionality*

*Functionality* merupakan aspek yang menunjukkan tingkat kegunaan output yang dikembalikan oleh chatbot kepada pengguna. Adapun terdapat lima skala penilaian yang digunakan aspek *functionality* pada penilaian yakni:

- Skala 1 : menyatakan output kembalian *chatbot* sangat tidak berguna.
- Skala 2 : menyatakan output kembalian *chatbot* tidak berguna.
- Skala 3 : menyatakan output kembalian *chatbot* cukup berguna.
- Skala 4 : menyatakan output kembalian *chatbot* berguna.
- Skala 5 : menyatakan output kembalian *chatbot* sangat berguna.

Adapula pertanyaan yang diberikan kepada pengguna untuk melakukan penilaian fitur tambahan dan fitur tanya jawab yang dinilai dari aspek *understandability* dan *functionality* adalah sebagai berikut:

1. Pertanyaan penilaian fitur tambahan berdasarkan aspek *understandability* yakni:
  - a) Apakah instruksi penggunaan chatbot mudah dimengerti?
  - b) Apakah modul informasi saham yang disediakan chatbot mudah dimengerti?
  - c) Apakah list sektor dan sahamnya yang diberikan chatbot mudah dimengerti?
  - d) Apakah informasi perusahaan yang diberikan chatbot mudah dimengerti?
  - e) Apakah informasi harga saham yang diberikan chatbot mudah dimengerti?
2. Pertanyaan penilaian fitur tambahan berdasarkan aspek *functionality* yakni:
  - a) Apakah instruksi penggunaan chatbot berguna?
  - b) Apakah modul informasi saham yang disediakan chatbot berguna dalam meningkatkan pengetahuan anda tentang saham?



- c) Apakah list sektor dan sahamnya yang diberikan chatbot berguna dalam pemilihan saham?
- d) Apakah informasi perusahaan yang diberikan chatbot berguna dalam pemilihan saham?
- e) Apakah informasi harga saham yang diberikan chatbot berguna dalam pemilihan saham?

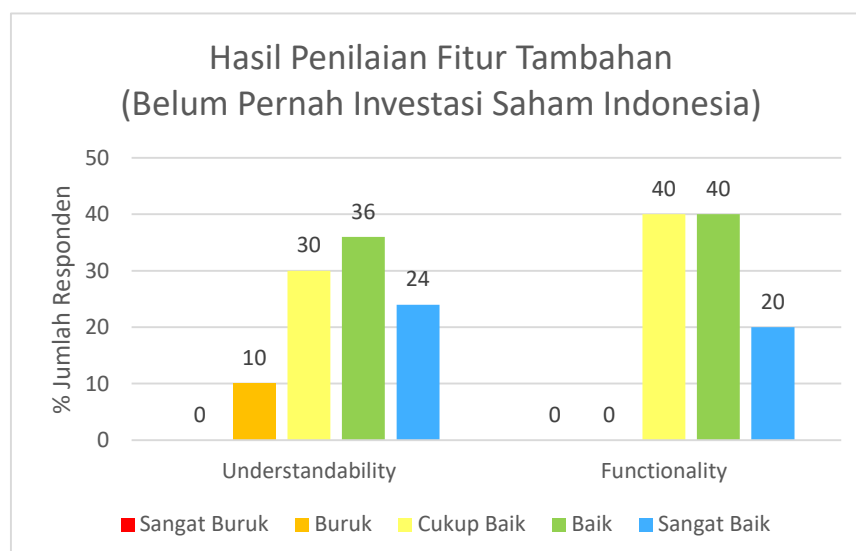
3. Pertanyaan penilaian fitur tanya jawab berdasarkan aspek *understandability* yakni:

Apakah jawaban atas pertanyaan Anda yang dikembalikan oleh fitur tanya jawab chatbot mudah dimengerti?

4. Pertanyaan penilaian fitur tanya jawab berdasarkan aspek *functionality* yakni:

Apakah fitur tanya jawab chatbot berguna dalam meningkatkan pengetahuan anda tentang saham?

Gambar 4.13 dan Gambar 4.14 menunjukkan hasil penilaian kedua aspek penilaian terhadap fitur tambahan dan fitur tanya jawab chatbot “MelekSaham” yang dinilai oleh 30 orang yang belum pernah berinvestasi saham Indonesia. Setelah itu Gambar 4.15 dan Gambar 4.16 menunjukkan hasil penilaian hal yang sama yang dinilai oleh 30 orang yang pernah berinvestasi saham Indonesia. Dan pada Gambar 4.17 dan Gambar 4.18 menunjukkan hasil penilaian hal yang sama yang dinilai oleh 4 pakar saham Indonesia.



**Gambar 4.13.** Hasil Penilaian Fitur Tambahan Oleh Pihak Yang Belum Pernah Berinvestasi Saham Indonesia

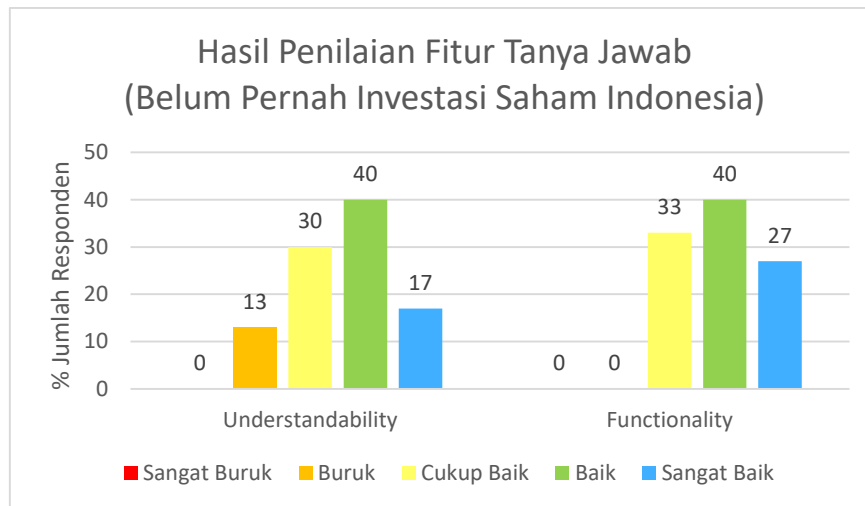
Gambar 4.13 menunjukkan penilaian fitur tambahan yang meliputi fitur modul saham, list sektor dan sahamnya, informasi perusahaan dan harga saham yang dinilai oleh 30 orang yang belum pernah berinvestasi saham Indonesia. Hasil dari penilaian pada Gambar 4.13 dapat dianalisis sebagai berikut:

1. Hasil penilaian aspek *understandability* fitur tambahan oleh pihak yang belum pernah berinvestasi saham Indonesia menunjukkan bahwa 10% (3 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sulit dipahami, 30% (9 orang) menyatakan output fitur tambahan yang dikembalikan chatbot cukup mudah dipahami, 36% (11 orang) menyatakan output fitur tambahan yang dikembalikan chatbot mudah dipahami, dan 24% (7 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat mudah dipahami. Untuk menghitung persentase akhir dari setiap aspek penilaian, penelitian ini menggunakan metode perhitungan dengan Persamaan 4.5 dimana persamaan ini dirujuk pada penelitian oleh Hengky (2017) dimana persentase akhir merupakan hasil penjumlahan aspek penilaian aspek skala 3 sampai dengan 5.

$$\text{Persentase Aspek (\%)} = \sum_{i=3}^5 \text{Persentase Skala } i \quad (4.5)$$

Sehingga berdasarkan Persamaan 4.5, didapatkan bahwa dari aspek *understandability* fitur tambahan chatbot dapat dipahami dengan mudah oleh 90% (27 orang) yang belum pernah berinvestasi saham Indonesia.

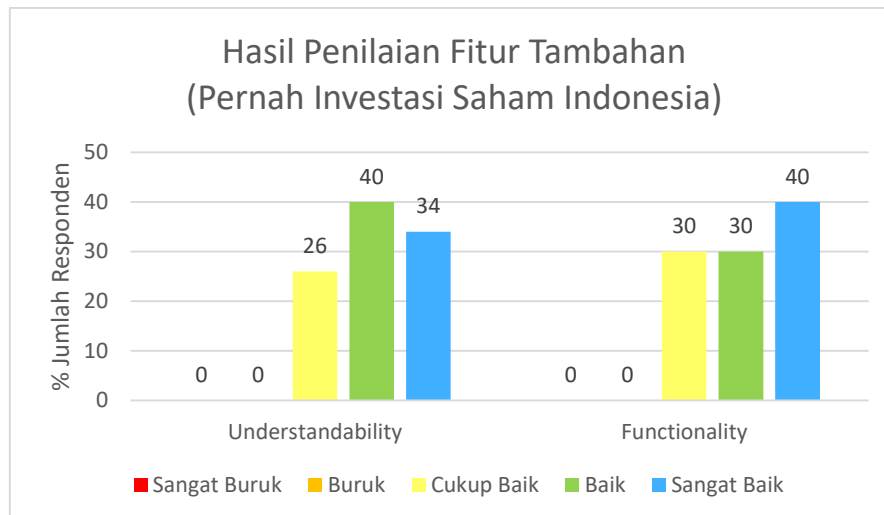
2. Hasil penilaian aspek *functionality* fitur tambahan oleh pengguna yang belum pernah berinvestasi saham Indonesia menunjukkan bahwa 40% (12 orang) menyatakan output fitur tambahan yang dikembalikan chatbot cukup berguna, 40% (12 orang) menyatakan output fitur tambahan yang dikembalikan chatbot berguna, dan 20% (6 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat berguna. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *functionality*, output fitur tambahan yang dikembalikan chatbot dinilai berguna oleh 100% (30 orang) yang belum pernah berinvestasi saham Indonesia.



**Gambar 4.14.** Hasil Penilaian Fitur Tanya Jawab Oleh Pihak Yang Belum Pernah Berinvestasi Saham Indonesia

Gambar 4.14 menunjukkan penilaian fitur tanya jawab pertanyaan seputar saham yang dinilai oleh 30 orang yang belum pernah berinvestasi saham Indonesia. Hasil dari evaluasi pada Gambar 4.14 dapat dianalisis sebagai berikut:

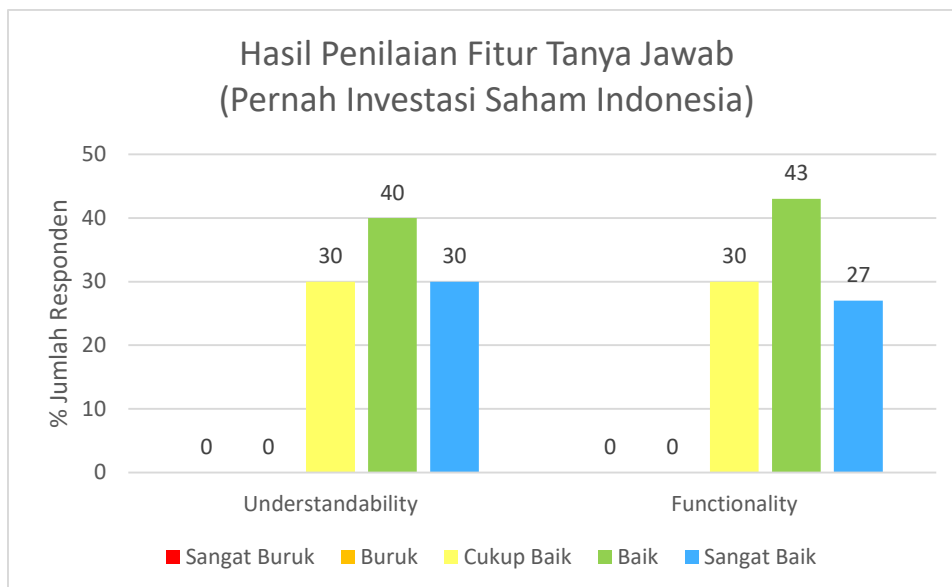
1. Hasil penilaian aspek *understandability* fitur tanya jawab oleh pihak yang belum pernah berinvestasi saham Indonesia menunjukkan bahwa 13% (4 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot sulit dipahami, 30% (9 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot cukup mudah dipahami, 40% (12 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot mudah dipahami, dan 17% (5 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot sangat mudah dipahami. Sehingga berdasarkan Persamaan 4.5, didapatkan bahwa dari aspek *understandability*, output fitur tanya jawab yang dikembalikan chatbot dapat dipahami dengan mudah oleh 87% (26 orang) yang belum pernah berinvestasi saham Indonesia.
2. Hasil penilaian aspek *functionality* fitur tanya jawab oleh pengguna yang belum pernah berinvestasi saham Indonesia menunjukkan bahwa 33% (10 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot cukup berguna, 40% (12 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot berguna, dan 27% (8 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat berguna. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *functionality*, output fitur tanya jawab yang dikembalikan chatbot dinilai berguna oleh 100% (30 orang) yang belum pernah berinvestasi saham.



**Gambar 4.15.** Hasil Penilaian Fitur Tambahan Oleh Pihak Yang Pernah Berinvestasi Saham Indonesia

Gambar 4.15 menunjukkan penilaian fitur tambahan yang meliputi fitur modul saham, list sektor dan sahamnya, informasi perusahaan dan harga saham yang dinilai oleh 30 orang yang pernah berinvestasi saham Indonesia. Hasil dari evaluasi pada Gambar 4.15 dapat dianalisis sebagai berikut:

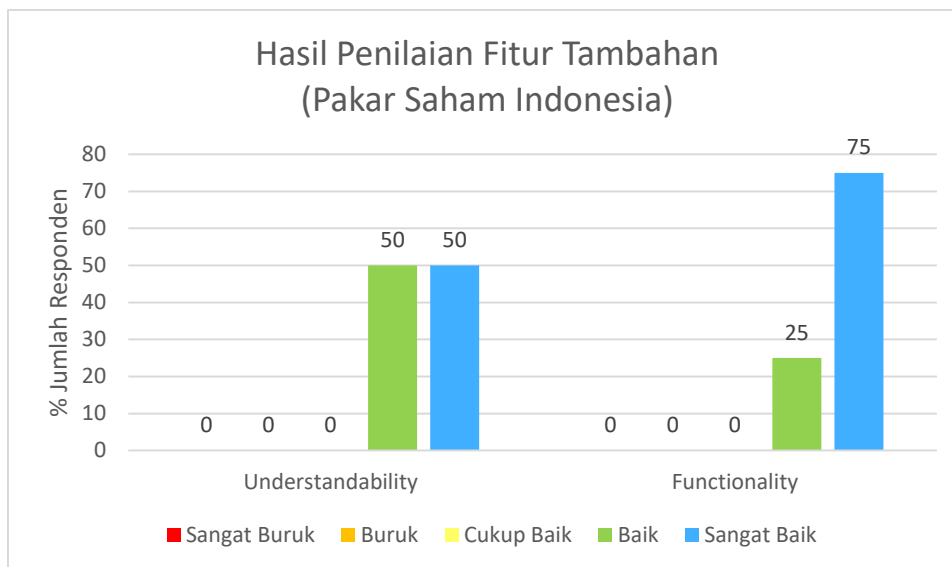
1. Hasil penilaian aspek *understandability* fitur tambahan oleh pihak yang pernah berinvestasi saham Indonesia menunjukkan bahwa 26% (8 orang) menyatakan output fitur tambahan yang dikembalikan chatbot cukup mudah dipahami, 40% (12 orang) menyatakan output fitur tambahan yang dikembalikan chatbot mudah dipahami dan 34% (10 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat mudah dipahami. Sehingga berdasarkan Persamaan 4.5, didapatkan bahwa dari aspek *understandability*, output fitur tambahan yang dikembalikan chatbot dapat dipahami dengan mudah oleh 100% (30 orang) yang pernah berinvestasi saham Indonesia.
2. Hasil penilaian aspek *functionality* fitur tambahan oleh pengguna yang pernah berinvestasi saham Indonesia menunjukkan bahwa 30% (9 orang) menyatakan output fitur tambahan yang dikembalikan chatbot cukup berguna, 30% (9 orang) menyatakan output fitur tambahan yang dikembalikan chatbot berguna, dan 40% (12 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat berguna. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *functionality*, output fitur tambahan yang dikembalikan chatbot dinilai berguna oleh 100% (30 orang) yang pernah berinvestasi saham Indonesia.



**Gambar 4.16.** Hasil Penilaian Fitur Tanya Jawab Oleh Pihak Yang Pernah Berinvestasi Saham Indonesia

Gambar 4.16 menunjukkan penilaian fitur tanya jawab pertanyaan seputar saham yang dinilai oleh 30 orang yang pernah berinvestasi saham Indonesia. Hasil dari evaluasi pada Gambar 4.16 dapat dianalisis sebagai berikut:

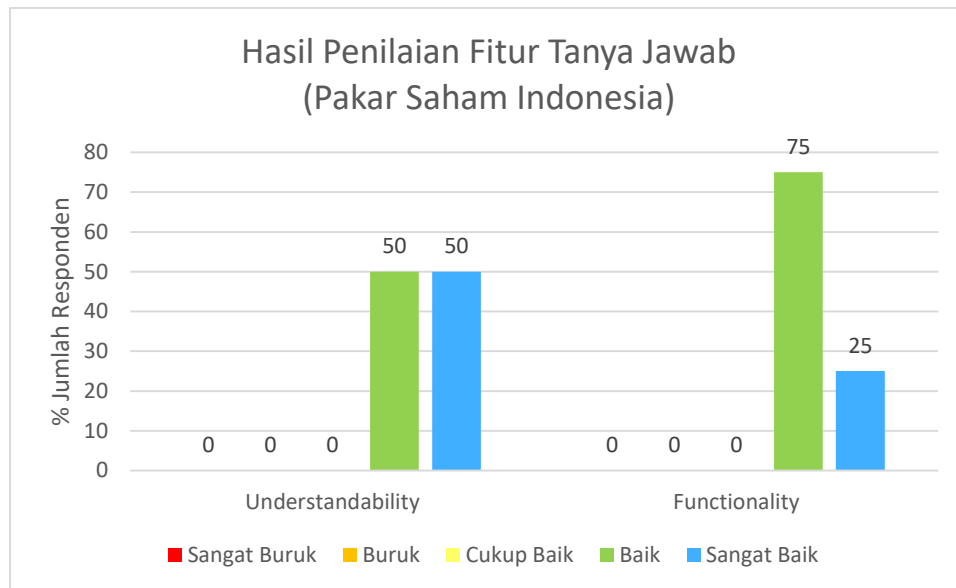
1. Hasil penilaian aspek *understandability* fitur tanya jawab oleh pihak yang pernah berinvestasi saham Indonesia menunjukkan bahwa 30% (9 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot cukup mudah dipahami, 40% (12 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot mudah dipahami, dan 30% (9 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot sangat mudah dipahami. Sehingga berdasarkan Persamaan 4.5, didapatkan bahwa dari aspek *understandability*, output fitur tanya jawab yang dikembalikan chatbot dapat dipahami dengan mudah oleh 100% (30 orang) yang pernah berinvestasi saham Indonesia.
2. Hasil penilaian aspek *functionality* fitur tanya jawab oleh pengguna yang pernah berinvestasi saham Indonesia menunjukkan bahwa 30% (9 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot cukup berguna, 43% (13 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot berguna, dan 27% (8 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat berguna. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *functionality*, output fitur tanya jawab yang dikembalikan chatbot dinilai berguna oleh 100% (30 orang) yang pernah berinvestasi saham Indonesia.



**Gambar 4.17.** Hasil Penilaian Fitur Tambahan Oleh Pakar Saham Indonesia

Gambar 4.17 menunjukkan penilaian fitur tambahan yang meliputi fitur modul saham, list sektor dan sahamnya, informasi perusahaan dan harga saham yang dinilai oleh 4 pakar saham Indonesia. Hasil dari evaluasi pada Gambar 4.17 dapat dianalisis sebagai berikut:

1. Hasil penilaian aspek *understandability* fitur tambahan oleh pakar saham Indonesia menunjukkan bahwa 50% (2 orang) menyatakan output fitur tambahan yang dikembalikan chatbot mudah dipahami dan 50% (2 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat mudah dipahami. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *understandability*, output fitur tambahan yang dikembalikan chatbot dapat dipahami dengan mudah oleh 100% (4 orang) pakar saham Indonesia.
2. Hasil penilaian aspek *functionality* fitur tambahan oleh pakar saham Indonesia menunjukkan bahwa 25% (1 orang) menyatakan output fitur tambahan yang dikembalikan chatbot berguna dan 75% (3 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat berguna. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *functionality*, output fitur tambahan yang dikembalikan chatbot dinilai berguna oleh 100% (4 orang) pakar saham Indonesia.



**Gambar 4.18.** Hasil Penilaian Fitur Tanya Jawab Oleh Pakar Saham Indonesia

Gambar 4.18 menunjukkan penilaian fitur tanya jawab pertanyaan seputar saham yang dinilai oleh 4 orang pakar saham Indonesia. Hasil dari evaluasi pada Gambar 4.18 dapat dianalisis sebagai berikut:

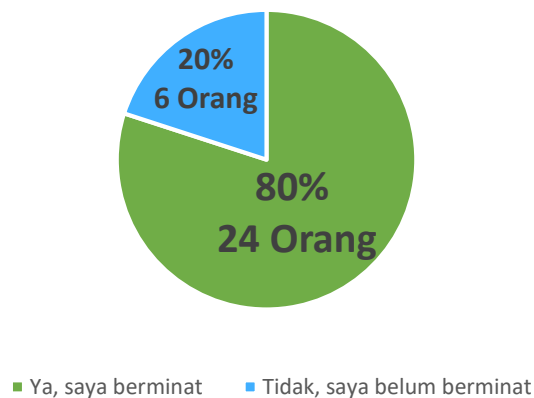
1. Hasil penilaian aspek *understandability* fitur tanya jawab oleh pakar saham Indonesia menunjukkan bahwa 50% (2 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot mudah dipahami dan 50% (2 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot sangat mudah dipahami. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *understandability*, output fitur tanya jawab yang dikembalikan chatbot dapat dipahami dengan mudah oleh 100% (4 orang) pakar saham Indonesia.
2. Hasil penilaian aspek *functionality* fitur tanya jawab oleh pakar saham Indonesia menunjukkan bahwa 75% (3 orang) menyatakan output fitur tanya jawab yang dikembalikan chatbot berguna dan 25% (1 orang) menyatakan output fitur tambahan yang dikembalikan chatbot sangat berguna. Sehingga berdasarkan persamaan 4.5, didapatkan bahwa dari aspek *functionality*, output fitur tanya jawab yang dikembalikan chatbot dinilai berguna oleh 100% (4 orang) pakar saham Indonesia.

Berdasarkan hasil penilaian chatbot oleh 30 orang yang belum pernah berinvestasi saham Indonesia, 30 orang yang pernah berinvestasi saham Indonesia dan 4 pakar saham Indonesia dimana penilaian menggunakan skala *likert* dan didasarkan pada dua aspek yakni aspek *understandability* dan aspek *functionality* terhadap fitur

tambahan dan fitur tanya jawab. Dapat disimpulkan bahwa chatbot yang dihasilkan pada penelitian ini memiliki kemudahan pemahaman dan fungsionalitas yang baik.

Pada kuisioner untuk 30 orang yang belum pernah berinvestasi saham Indonesia terdapat pertanyaan tambahan untuk menilai penumbuhan minat setelah penggunaan chatbot “MelekSaham”. Pertanyaan tambahan yang diberikan adalah “Apakah setelah penggunaan chatbot, anda menjadi berminat berinvestasi saham di pasar saham Indonesia?”. Hasil jawaban tersebut ditampilkan pada Gambar 4.19 berikut:

Tumbuhnya Minat Berinvestasi Saham  
(Belum Pernah Berinvestasi Saham Indonesia)



**Gambar 4.19.** Timbulnya Minat Berinvestasi Saham

Dari Gambar 4.19 didapatkan bahwa dari 30 orang yang belum pernah berinvestasi saham Indonesia. 80% (24 orang) menjadi berminat untuk berinvestasi saham di pasar saham Indonesia setelah menggunakan chatbot “MelekSaham”, sedangkan 20% (6 orang) masih belum berminat untuk berinvestasi saham di pasar saham Indonesia setelah menggunakan chatbot “MelekSaham”. Alasan yang diungkapkan oleh 6 orang yang masih belum berminat untuk berinvestasi saham di pasar saham Indonesia adalah dikarenakan kondisi keuangan yang masih belum memungkinkan untuk berinvestasi saham Indonesia.

Dilihat dari 80% (24 orang) yang berminat untuk berinvestasi saham di pasar saham Indonesia setelah menggunakan chatbot “MelekSaham” dan 20% (6 orang) yang belum berminat untuk berinvestasi saham di pasar saham Indonesia setelah menggunakan chatbot “MelekSaham” dikarenakan masalah keuangan yang belum memungkinkan untuk investasi. Dapat disimpulkan bahwa chatbot berhasil dalam menumbuhkan minat berinvestasi saham di pasar saham Indonesia.



## **BAB V**

### **KESIMPULAN DAN SARAN**

Bab kesimpulan dan saran membahas tentang kesimpulan yang didapatkan dari implementasi algoritma *Long Short-Term Memory* dan *Integer Sequence Matching* dalam pembuatan chatbot informasi saham Indonesia serta saran-saran untuk penelitian selanjutnya.

#### **5.1. Kesimpulan**

Dari hasil pengujian sistem pada Bab 4, diperoleh kesimpulan sebagai berikut:

1. Pada pelatihan model, dihasilkan akurasi sebesar 97.46% dan loss sebesar 12.36% untuk data *training* serta untuk data *validation* dihasilkan akurasi sebesar 92.86% dan *loss* sebesar 18.62%.
2. Pada pengujian model, dihasilkan accuracy sebesar 80% dengan rata rata precision sebesar 89%, rata rata recall sebesar 78% dan rata rata F1-Score sebesar 82%.
3. Pada pengujian sistem, implementasi algoritma *Long Short-Term Memory* dan *Integer Sequence Matching* pada chatbot informasi saham Indonesia berhasil menghasilkan jawaban dengan tingkat akurasi sebesar 90% serta rata-rata kecepatan menghasilkan jawaban sebesar 0.893 detik.
4. Berdasarkan hasil penilaian chatbot oleh 30 orang yang belum pernah berinvestasi saham Indonesia, 30 orang yang pernah berinvestasi saham Indonesia dan 4 pakar saham Indonesia dimana penilaian menggunakan skala *likert* dan didasarkan pada dua aspek yakni aspek *understandability* dan aspek *functionality* terhadap fitur tambahan dan fitur tanya jawab. Dapat disimpulkan bahwa chatbot yang dihasilkan pada penelitian ini memiliki kemudahan pemahaman dan fungsionalitas yang baik.

5. Chatbot berhasil dalam menumbuhkan minat 24 dari 30 orang yang belum pernah berinvestasi saham Indonesia untuk mencoba berinvestasi saham di pasar saham Indonesia.

## **5.2. Saran**

Adapun saran untuk pengembangan penelitian selanjutnya yakni sebagai berikut:

1. Diharapkan pada penelitian selanjutnya ditambahkan sistem yang dapat menambahkan secara otomatis jawaban atas pertanyaan pengguna yang belum ada di dalam corpus.
2. Penelitian selanjutnya dapat menambahkan sistem untuk membantu pengguna dalam melakukan analisis saham baik secara teknikal maupun secara fundamental.

## DAFTAR PUSTAKA

- Al-Ash, H. S., Anki, P., Bustamam, A., & Sarwinda, D. (2021). Intelligent Chatbot Adapted from Question and Answer System Using RNN-LSTM Model. *Journal of Physics: Conference Series* 1844(1). <https://doi.org/10.1088/1742-6596/1844/1/012001>.
- Ash-Shidiq, H., & Setiawan, A. B. (2015). Analisis Pengaruh Suku Bunga SBI, Uang Beredar, Inflasi Dan Nilai Tukar Terhadap Indeks Harga Saham Jakarta Islamic Index (JII) Periode 2009-2014. *Jurnal Ekonomi dan Perbankan Syariah* 3(2): 25-46.
- Azni, A. (2021). Implementasi *Natural Language Processing* Pada Sistem Chatbot Informasi Saham dengan Algoritma *Long Short-Term Memory* (LSTM) dan *Fuzzy String Matching*. Skripsi. Universitas Sumatera Utara.
- Barbosa, G., Franca, R. D. S., Marietto, M., Pimentel, E. P., Silva, V. L. D., Aguiar, R. V. D., Botelho, W. T. (2013). Artificial Intelligence Markup Language: A Brief Tutorial. *International Journal of Computer Science & Engineering Survey* 4(3). <https://doi.org/10.5121/ijcses.2013.4301>.
- Cahya, B., & Kusuma, W. N. (2019). Pengaruh Motivasi dan Kemajuan Teknologi Terhadap Minat Investasi Saham. *Al-Masharif: Jurnal Ilmu Ekonomi dan Keislaman* 7(2): 192-207. <https://doi.org/10.24952/masharif.v7i2.2182>.
- Chi, J., Liu, Y., Wang, V., & Yan, J. (2022). Performance Analysis of Three kinds of Neural Networks in the Classification of Multi Classes. *Journal of Physics: Conference Series* 2181(1). <https://doi.org/10.1088/1742-6596/2181/1/012032>.
- Dahiya, M. (2017). A Tool Of Conversation : Chatbot. *International Journal of Computer Sciences and Engineering* 5(5): 158-161.
- Darmadji, T. & Fakhruddin, H. M. (2015). Pasar saham Indonesia. Edisi Ketiga. Jakarta: Salemba Empat.
- Desai, A. N., Ruidera, D., Steinbrink, J. M., Granwehr, B., & Lee, D. H. (2022). Misinformation and Disinformation: The Potential Disadvantages of Social Media in Infectious Disease and How to Combat Them. *Clinical infectious diseases : an official journal publication of the Infectious Diseases Society of America* 74(3): 34-39. <https://doi.org/10.1093/cid/ciac109>.
- Dewi, I. M., & Purbawangsa, I. (2018). Pengaruh Literasi Keuangan, Pendapatan Serta Masa Bekerja Terhadap Perilaku Keputusan Investasi. *E-Jurnal Ekonomi Dan Bisnis Universitas Udayana*: 1867-1894. <https://doi.org/10.24843/EEB.2018.v07.i07.p04>.
- Dipietro, R., & Hager, G. D. (2019). Deep learning: RNNs and LSTM. In book of Medical Image Computing. *Computer Assisted Intervention*. <https://doi.org/10.1016/B978-0-12-816176-0.00026-0>.
- Drahošová, M., & Balco, P. (2017). The analysis of advantages and disadvantages of use of social media in European Union. *Proceeding Of Computer Science* 109: 1005-1009. <https://doi.org/10.1016/j.procs.2017.05.446>.
- Fauzy, A. (2014). Konsep Dasar Teori Sampling. *Journal Of Antimicrobial Agents and Chemotherapy* 58(12).

- Harani, N. H., Khaqiqi, M.I.T., & Prianto, C. (2023). Performance Analysis and Development of QnA Chatbot Model Using LSTM in Answering Questions. *Journal of Computer Science* 12(3). <https://doi.org/10.33022/ijcs.v12i3.3249>.
- Joshi, S., Rathod, S., Sagar, N., Shinde, P., & Shinde, S. (2023). Chatbot Using Deep Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. <https://doi.org/10.32628/cseit2390148>.
- Khairani, F. (2021). Aplikasi Chatbot Tanya Jawab tentang Kesehatan Menggunakan Algoritma *Enhanced Confix Stripping* dan Algoritma *Knuth Morris Pratt*. Skripsi. Universitas Sumatera Utara.
- Lalwani, T., Bhalotia, S., Pal, A., Bisen, S., & Rathod, V. (2018). Implementation of a Chatbot System using AI and NLP. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)* 6(3). <https://doi.org/10.21276/ijircst.2018.6.3.2>.
- Liddy, E. D. (2001). *Natural Language Processing*. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.
- Lubis, M. H. R., & Kusuma, I. G. K. C. B. A. (2022). Analisis Pertumbuhan Investor Ritel Pada Masa Pandemi Dan Implikasi Pajak Penghasilan Final Atas Penjualan Saham Di Bursa. *Jurnal Pajak Indonesia (Indonesian Tax Review)* 6(2): 245–264. <https://doi.org/10.31092/jpi.v6i2.1854>.
- Manaswi, N. K., & John, S. (2018). *Deep learning with applications using Python*. pp. 31-43. Apress: Berkeley.
- Migunani, & Aditama, K. (2020). Pemanfaatan Natural Language Processing Dan Pertanyaan Matching Dalam Pembelajaran Melalui Guru Virtual. *Jurnal Elektronika dan Komputer Universitas STEKOM* 13(1): 121-131.
- Muangkammuen, P., Intiruk, N., & Saikaew, K. R. (2018). Automated Thai-FAQ Chatbot using RNN-LSTM. *22nd International Computer Science and Engineering Conference (ICSEC)*. <https://doi.org/10.1109/ICSEC.2018.8712781>.
- Nugraha, A. T. (2020). Aplikasi Pemesanan *Travel* Menggunakan *Chatbot* Dengan *Artificial Intelligence Markup Language*. Tesis. Universitas Islam Negeri Maulana Malik Ibrahim.
- Olah, C. 2015. Understanding LSTM Networks. (Online) <http://colah.github.io/2015-08-Understanding-LSTMs> (29 September 2023).
- Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning*. pp.3-4. O'Reilly Media, Inc: California.
- Putri, M., Shabri, S. (2022). Analisis Fundamental Dan Teknikal Saham. *Al-Bank: Journal Islamic Banking and Finance* 2(1).
- Ribalta, R. (2014). Rancang Bangun Sistem Pembelajaran Menggunakan *Chatbot* dengan Algoritma *Boyer-Moore*. Skripsi. Universitas Islam Negeri Sultan Syarif Kasim.
- Samsul, M. (2015). Pasar saham Dan Manajemen Portofolio. Edisi Kedua. Erlangga: Jakarta.
- Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena* 404. <https://doi.org/10.1016/j.physd.2019.132306>.
- Sihotang, M. T. (2019). *Chatbot Tanya Jawab Islamic APP* Menggunakan Algoritma *Enhanced Confix Stripping* dan Algoritma *Fuzzy String Matching*. Skripsi. Universitas Sumatera Utara.

- Stylos, J., Faulring, A., Yang, Z., & Myers, B. A. (2009). Improving API Documentation Using API Usage Information. *2009 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pp. 119-126.
- Sugiyono. (2014). *Metode Penelitian Kuantitatif Kualitatif dan R&D*. Alfabeta: Bandung.
- Sulistyowati, & Rahmawati, M. (2020). Analisis Pengaruh Pasar Saham Terhadap Pertumbuhan Ekonomi Di Negara Berkembang (Suatu Kajian Literatur). *Research Fair Unisri*. <https://doi.org/4.10.33061/rsfu.v4i1.3392>.
- Sunariyah. (2006). Pengantar Pengetahuan Pasar Modal. *UPP STIM YKPN* 5(1).
- Toleva, B. (2021). The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems. *Business Systems Research Journal* 12(1). <https://doi.org/10.2478/bsrj-2021-0015>.