

**RANCANG BANGUN OPTIMASI MODEL *CHATBOT***  
**DIGITAL HUB SINAR MAS LAND BERBASIS**  
***NATURAL LANGUAGE PROCESSING***



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

**SKRIPSI**

**Sherly**

**0000059189**

**PROGRAM STUDI SISTEM INFORMASI**  
**FAKULTAS TEKNIK DAN INFORMATIKA**  
**UNIVERSITAS MULTIMEDIA NUSANTARA**  
**TANGERANG**  
**2024**

**RANCANG BANGUN OPTIMASI MODEL *CHATBOT*  
DIGITAL HUB SINAR MAS LAND BERBASIS  
*NATURAL LANGUAGE PROCESSING***



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

**Sherly**

**0000059189**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS TEKNIK DAN INFORMATIKA  
UNIVERSITAS MULTIMEDIA NUSANTARA  
TANGERANG**

**2024**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Sherly  
Nomor Induk Mahasiswa : 00000059189  
Program Studi : Sistem Informasi

Skripsi dengan judul:

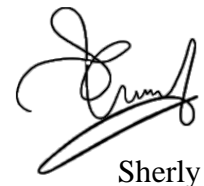
RANCANG BANGUN OPTIMASI MODEL *CHATBOT* DIGITAL HUB SINAR MAS LAND BERBASIS *NATURAL LANGUAGE PROCESSING*

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 14 Oktober 2024

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA



Sherly

## HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Sherly  
NIM : 00000059189  
Program Studi : Sistem Informasi  
Jenjang : S1  
Judul Karya Ilmiah : RANCANG BANGUN OPTIMASI  
MODEL *CHATBOT* DIGITAL HUB SINAR  
MAS LAND BERBASIS *NATURAL  
LANGUAGE PROCESSING*

Menyatakan dengan sesungguhnya bahwa saya bersedia\* (**pilih salah satu**):

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) \*\*.
- Lainnya, pilih salah satu:
  - Hanya dapat diakses secara internal Universitas Multimedia Nusantara
  - Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 14 Oktober 2024



Sherly

\* Pilih salah satu

\*\* Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

## KATA PENGANTAR

Terima kasih saya panjatkan kepada Tuhan Yang Maha Esa atas berkat dan karuniaNya, skripsi berjudul “Rancang Bangun Optimasi Model *Chatbot* Digital Hub Sinar Mas Land Berbasis *Natural Langange Processing*” selesai tepat waktu. Skripsi ini sulit diselesaikan tanpa adanya bantuan, bimbingan, dan dukungan dari berbagai pihak dalam perkuliahan. Maka dari itu, saya ingin mengucapkan terima kasih kepada:

1. Bapak Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo. S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti S.Kom., M.Kom., selaku Ketua Program Studi Sistem Informasi Universitas Multimedia Nusantara.
4. Ibu Wella, S.Kom., M.MSI., COBIT 5 sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya laporan ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Valen Claudia dan teman-teman yang telah menguatkan dan memberikan dukungan mental agar dapat menuntaskan tugas akhir ini.

Semoga karya ilmiah ini dapat memberikan dampak yang baik serta menjadi acuan penelitian di masa depan sehingga dapat bermanfaat bagi almamater, persada, dan sesama.

Tangerang, 21 November 2024



Sherly

# RANCANG BANGUN OPTIMASI MODEL *CHATBOT*

## DIGITAL HUB SINAR MAS LAND BERBASIS

### *NATURAL LANGUAGE PROCESSING*

Sherly

#### ABSTRAK

Perkembangan teknologi informasi, termasuk *Artificial Intelligence* (AI) dan *Machine Learning* (ML), telah meningkatkan sistem informasi dengan pengolahan data yang lebih cepat, akurat, dan adaptif. Adopsi AI dan ML terus meningkat, mencapai 57% pada 2021 menurut McKinsey, naik 45% dari tahun sebelumnya. Pasar *Natural Language Processing* (NLP) di Asia Pasifik juga tumbuh dari USD 2,7 miliar pada 2020 menjadi USD 4,2 miliar pada 2021, dengan proyeksi pertumbuhan signifikan hingga 2030. Dalam pemanfaatan NLP yang terus berkembang, penelitian ini berfokus pada rancang bangun dan optimasi model *chatbot* pada Digital Hub Sinar Mas Land dalam menjawab permasalahan perusahaan.

Penelitian ini menggunakan kerangka kerja CRISP-ML yang terdiri dari beberapa tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*. Dataset berupa teks pertanyaan dan jawaban diproses menggunakan teknik pra-pemrosesan seperti tokenisasi, penghapusan *stopword*, dan *stemming*, lalu ditransformasi menggunakan TF-IDF untuk ekstraksi fitur. Algoritma SVM dengan *kernel linear* diterapkan untuk membangun model klasifikasi yang mampu memahami dan merespon permintaan pengguna secara optimal. Model ini diimplementasikan menggunakan Streamlit untuk memudahkan akses oleh pengguna.

Hasil eksperimen menunjukkan bahwa model yang dikembangkan memiliki tingkat akurasi sebesar 95,77% menggunakan *kernel linear*. Penelitian ini berkontribusi dalam pengembangan teknologi chatbot berbasis NLP yang efisien dan efektif. Hasil *closed beta testing* juga menunjukkan bahwa *chatbot* yang diuji oleh peneliti memiliki performa dengan rata-rata 87,50%. Saran pengembangan di masa depan mencakup perluasan dataset, penggunaan *feature selection* untuk meningkatkan akurasi, serta pengintegrasian fitur *generative-based* untuk meningkatkan fleksibilitas *chatbot*.

**Kata kunci:** *Chatbot*, Digital Hub, NLP, SVM, TF-IDF.

**DESIGN AND DEVELOPMENT OF OPTIMIZATION MODEL  
CHATBOT DIGITAL HUB SINAR MAS LAND BASED ON  
NATURAL LANGUAGE PROCESSING**

Sherly

**ABSTRACT (English)**

*The rapid development of information technology, particularly Artificial Intelligence (AI) and Machine Learning (ML), has significantly advanced information systems, enabling faster, more accurate, and adaptive data processing. The adoption of AI and ML continues to exhibit substantial growth, with McKinsey reporting a 57% adoption rate in 2021, an increase of 45% from the previous year. Furthermore, the Natural Language Processing (NLP) market in the Asia-Pacific region expanded from USD 2.7 billion in 2020 to USD 4.2 billion in 2021, with projections indicating continued robust growth through 2030. In leveraging the evolving potential of NLP, this study focuses on the design and optimization of a chatbot model at Digital Hub Sinar Mas Land to address the company's operational challenges.*

*The research adopts the CRISP-ML framework, encompassing several stages: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The dataset, consisting of text-based questions and answers, underwent pre-processing techniques such as tokenization, stopword removal, and stemming before being transformed using TF-IDF for feature extraction. The SVM algorithm with a linear kernel was implemented to develop a classification model capable of accurately understanding and responding to user queries. The model was deployed using Streamlit to facilitate accessibility for end users.*

*The experimental results demonstrate that the developed model achieves an accuracy rate of 95.77% using a linear kernel. This study contributes to the advancement of efficient and effective NLP-based chatbot technologies. Additionally, the results from the closed beta testing indicate that the chatbot, as evaluated by the researchers, performs with an average score of 87.50%. Future development recommendations include expanding the dataset, employing feature selection techniques to improve accuracy, and integrating generative-based features to enhance the chatbot's flexibility and overall performance.*

**Keywords:** *Chatbot, Digital Hub, NLP, SVM, TF-IDF.*

## DAFTAR ISI

<b>HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH</b> .....	iii
<b>KATA PENGANTAR</b> .....	iv
<b>ABSTRAK</b> .....	v
<b>ABSTRACT (English)</b> .....	vi
<b>DAFTAR ISI</b> .....	vii
<b>DAFTAR TABEL</b> .....	x
<b>DAFTAR GAMBAR</b> .....	xi
<b>DAFTAR LAMPIRAN</b> .....	xii
<b>BAB I PENDAHULUAN</b> .....	1
<b>1.1 Latar Belakang</b> .....	1
<b>1.2 Rumusan Masalah</b> .....	6
<b>1.3 Batasan Masalah</b> .....	6
<b>1.4 Tujuan dan Manfaat Penelitian</b> .....	7
<b>1.4.1 Tujuan Penelitian</b> .....	7
<b>1.4.2 Manfaat Penelitian</b> .....	7
<b>1.5 Sistematika Penulisan</b> .....	8
<b>BAB II LANDASAN TEORI</b> .....	11
<b>2.1 Penelitian Terkait</b> .....	11
<b>2.2 Tinjauan Teori</b> .....	17
<b>2.2.1 Artificial Intelligence (AI)</b> .....	17
<b>2.2.2 Machine Learning (ML)</b> .....	17
<b>2.2.3 Natural Language Processing (NLP)</b> .....	18
<b>2.2.4 Chatbot</b> .....	19
<b>2.3 Framework/Algoritma yang digunakan</b> .....	21
<b>2.3.1 Framework CRISP-ML</b> .....	21
<b>2.3.2 Algoritma Support Vector Machine (SVM)</b> .....	23
<b>2.3.3 Metode TF-IDF</b> .....	26
<b>2.3.4 Metode Evaluasi Metrik Akurasi</b> .....	28
<b>2.3.5 Metode Closed Beta Testing</b> .....	29
<b>2.4 Tools yang digunakan</b> .....	30



2.4.1	Python .....	30
2.4.2	Jupyter Notebook .....	30
2.4.3	Visual Studio Code .....	31
2.4.4	Streamlit .....	31
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>33</b>
3.1	<b>Gambaran Umum Objek Penelitian .....</b>	<b>33</b>
3.1.1	Sejarah Perusahaan .....	33
3.1.2	Struktur Organisasi Perusahaan .....	34
3.2	<b>Metode Penelitian .....</b>	<b>35</b>
3.2.1	Metode Machine Learning .....	35
3.2.2	Metode Vektorisasi Teks .....	37
3.2.3	Alur Penelitian .....	40
3.3	<b>Teknik Pengumpulan Data .....</b>	<b>42</b>
3.3.1	Metode Pengumpulan Data .....	42
3.3.2	Narasumber .....	43
3.3.3	Periode Pengambilan Data .....	43
<b>BAB IV ANALISIS DAN HASIL PENELITIAN .....</b>		<b>45</b>
4.1	<b>Business &amp; Data Understanding .....</b>	<b>45</b>
4.2	<b>Data Preparation .....</b>	<b>47</b>
4.2.1	<i>Import, Read, and Info Dataset</i> .....	47
4.2.2	<i>Case Folding</i> .....	49
4.2.3	<i>Tokenizing</i> .....	50
4.2.4	<i>Stopword Removal</i> .....	51
4.2.5	<i>Stemming</i> .....	52
4.2.6	<i>Fitur TF-IDF (Term Frequency-Inverse Document Frequency)</i> ..	53
4.3	<b>Modeling .....</b>	<b>55</b>
4.4	<b>Evaluation .....</b>	<b>59</b>
4.5	<b>Deployment .....</b>	<b>59</b>
4.6	<b>Pengujian <i>Closed Beta Testing</i> .....</b>	<b>62</b>
4.7	<b>Hasil dan Diskusi .....</b>	<b>65</b>
<b>BAB V SIMPULAN DAN SARAN .....</b>		<b>70</b>
5.1	<b>Simpulan .....</b>	<b>70</b>

<b>5.2 Saran</b> .....	71
<b>DAFTAR PUSTAKA</b> .....	73
<b>LAMPIRAN</b> .....	79



**UMN**  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu .....	11
Tabel 3.1 Perbandingan CRISP-DM dan CRISP-ML .....	35
Tabel 3.2 Perbandingan Vektorisasi Teks.....	37
Tabel 4.1 Contoh Dataset <i>chatbot</i> Digital Hub.....	45
Tabel 4.2 Hasil dari implementasi <i>chatbot</i> Digital Hub Sinar Mas Land.....	62



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR GAMBAR

Gambar 1.1 Grafik Asia Pasific Natural Language Processing Market[3].....	1
Gambar 2.1 CRISP-ML Life Cycle Process[37].....	21
Gambar 2.2 Hyperplane pada SVM [42]. .....	24
Gambar 2.3 Detail of Hyperplane pada SVM [42]. .....	25
Gambar 2.4 Linear SVM vs Non-linear SVM [43].....	25
Gambar 3.1 Logo Sinar Mas Land .....	33
Gambar 3.2 Logo Digital Hub .....	34
Gambar 3.3 Struktur Perusahaan Digital Hub by Sinar Mas Land .....	34
Gambar 3.4 Alur Penelitian.....	42
Gambar 3.5 Dataset chatbot Digital Hub .....	43
Gambar 4.1 Read Dataset.....	48
Gambar 4.2 Info Dataset .....	49
Gambar 4.3 Case Folding untuk Pertanyaan dan Jawaban .....	50
Gambar 4.4 Tokenizing untuk Pertanyaan dan Jawaban .....	51
Gambar 4.5 Stopword Removal untuk Pertanyaan dan Jawaban .....	52
Gambar 4.6 Proses Stemming Pertanyaan .....	53
Gambar 4.7 Fitur TF-IDF.....	54
Gambar 4.8 Membagi data latih & data uji.....	56
Gambar 4.9 Model <i>Chatbot</i> Menggunakan <i>Kernel RBF</i> .....	56
Gambar 4.10 Model <i>Chatbot</i> Menggunakan <i>Kernel Sigmoid</i> .....	56
Gambar 4.11 Model <i>Chatbot</i> Menggunakan <i>Kernel Linear</i> .....	57
Gambar 4.12 Pembuatan model menggunakan algoritma SVM.....	57
Gambar 4.13 Uji Coba Menggunakan Deep Learning (Transformer & BERT)...	58
Gambar 4.14 Evaluasi model SVM menggunakan metrik evaluasi .....	59
Gambar 4.15 Instalasi Streamlit pada Anaconda Prompt .....	60
Gambar 4.16 Konfigurasi lingkungan untuk implementasi chatbot .....	60
Gambar 4.17 Cara Membuka Digibot.....	61
Gambar 4.18 Halaman Awal Rancang Bangun Digibot .....	61
Gambar 4.19 Contoh implementasi chatbot menggunakan Streamlit.....	61

UNIVERSITAS  
MULTIMEDIA  
NUSANTARA

## DAFTAR LAMPIRAN

Lampiran A Turnitin Similarity Report.....	79
Lampiran B Form Konsultasi Bimbingan.....	87
Lampiran C Surat Pernyataan Wawancara untuk Keperluan Dataset Chatbot.....	88



UMN  
UNIVERSITAS  
MULTIMEDIA  
NUSANTARA