

**IMPLEMENTASI *ZERO-SHOT LEARNING* UNTUK PREDIKSI
KALIMAT SOLUSI DARI MASALAH PADA ARTIKEL
ILMIAH MENGGUNAKAN *LARGE LANGUAGE MODEL*
(LLM)**

LAPORAN TUGAS AKHIR

Laporan ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan
program studi Teknik Informatika S-1 pada Fakultas Teknologi Industri
Universitas Islam Sultan Agung



Disusun Oleh:

ANITA SOFFIYUN NADA

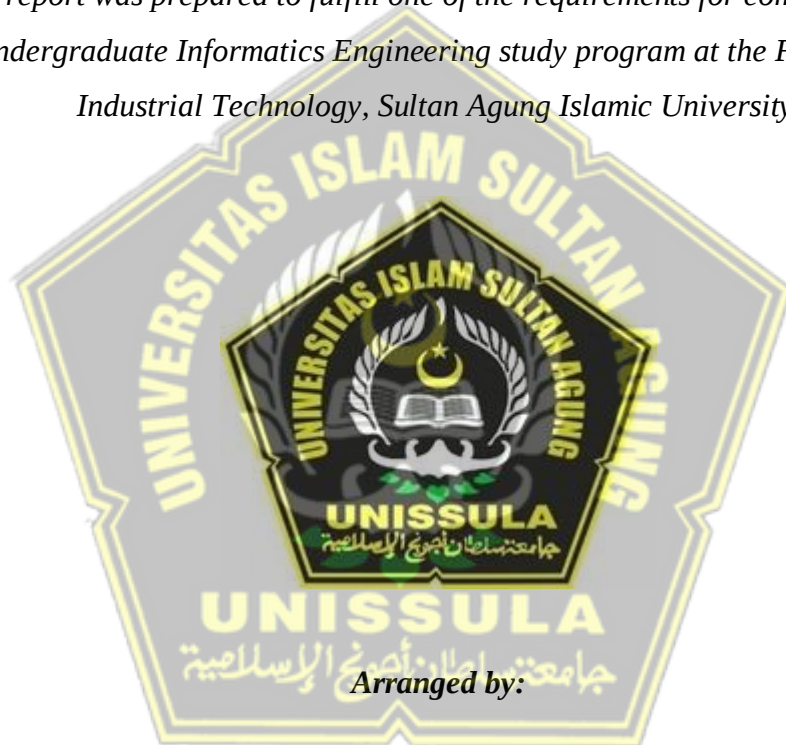
NIM 32602100030

**FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
2025**

**IMPLEMENTATION OF ZERO-SHOT LEARNING FOR
PREDICTION OF SOLUTION SENTENCES OF PROBLEMS IN
SCIENTIFIC ARTICLES USING LARGE LANGUAGE MODELS
(LLM)**

FINAL PROJECT

This report was prepared to fulfill one of the requirements for completing the Undergraduate Informatics Engineering study program at the Faculty of Industrial Technology, Sultan Agung Islamic University.



Arranged by:

ANITA SOFFIYUN NADA

NIM 32602100030

**MAJORING OF INFORMATICS ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
SULTAN AGUNG ISLAMIC UNIVERSITY
SEMARANG**

2025

LEMBAR PENGESAHAN

TUGAS AKHIR

**IMPLEMENTASI ZERO-SHOT LEARNING UNTUK PREDIKSI
KALIMAT SOLUSI DARI MASALAH PADA ARTIKEL ILMIAH
MENGUNAKAN *LARGE LANGUAGE MODELS* (LLM)**

**ANITA SOFFIYUN NADA
NIM 32602100030**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 25 Februari 2025.

TIM PENGUJI UJIAN SARJANA :

Arief Marwanto, ST, M.Eng, Ph.D
NIDN. 0628097501
(Ketua Penguji)

06-03-2025

Mustafa, ST., MM., M.Kom
NIDN. 0623117703
(Anggota Penguji)

11-03-2025

Sam Farisa C.H., ST., M.Kom
NIDN. 0628028602
(Pembimbing)

11-03-2025

Semarang, 11 Maret 2025.
Mengetahui,

Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch. Taufik, S.T., M.IT
NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Anita Soffiyun Nada

NIM : 32602100030

Judul Tugas Akhir : IMPLEMENTASI *ZERO-SHOT LEARNING* UNTUK
PREDIKSI KALIMAT SOLUSI DARI KALIMAT
MASALAH PADA ARTIKEL ILMIAH
MENGUNAKAN *LARGE LANGUAGE MODELS*
(LLM)

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapa pun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 13 Maret 2025

Yang Menyatakan



Anita Soffiyun Nada

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Anita Soffiyun Nada

NIM : 32602100030

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul :
IMPLEMENTASI *ZERO-SHOT LEARNING* UNTUK PREDIKSI KALIMAT
SOLUSI DARI KALIMAT MASALAH PADA ARTIKEL ILMIAH
MENGUNAKAN *LARGE LANGUAGE MODELS* (LLM)

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan
Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan
pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan
akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta.
Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari
terbukti ada pelanggaran Hak Cipta/Plagiatisme dalam karya ilmiah ini, maka
segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa
melibatkan Universitas Islam Sultan Agung.

Semarang, 13 Maret 2025

Yang Menyatakan



Anita Soffiyun Nada

KATA PENGANTAR

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan Tugas Akhir ini dengan judul “Implementasi *Zero-shot Learning* Untuk Prediksi Kalimat Solusi Dari Masalah Pada Artikel Ilmiah Menggunakan *Large Language Models* (LLM)” Tugas Akhir ini disusun sebagai salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung.

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada:

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu dikampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Novi Marlyana, S.T., M.T., IPU., ASEAN. Eng.
3. Dosen Pembimbing penulis Bapak Sam Farisa Chaerul Havana, ST., M.Kom yang telah meluangkan waktu, membimbing, dan memberi ilmu.
4. Orang tua penulis, Bapak Siswono dan Ibu Wantini, yang selalu memberikan restu, dukungan, serta Doa dalam menyelesaikan Tugas Akhir ini.
5. Terima kasih kepada SEVENTEEN atas musik dan kerja kerasnya yang menginspirasi serta memberi semangat dalam penyusunan laporan ini.
6. Para sahabat dan teman-teman atas bantuan, semangat, dan dukungannya.

Dengan rendah hati, penulis menyadari bahwa laporan masih memiliki banyak kekurangan dalam kualitas dan isi. karena itu, penulis mengharapkan kritikan dan saran yang membangun untuk penyempurnaan di masa depan.

Semarang, 24 Februari 2025

Anita Soffiyun Nada

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR	iv
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
ABSTRAK	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	2
1.3 Pembatasan Masalah	3
1.4 Tujuan	3
1.5 Manfaat	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori	8
2.2.1 Kalimat Solusi	8
2.2.2 <i>Natural Language Processing (NLP)</i>	9
2.2.3 <i>Large Language Model (LLM)</i>	10
2.2.4 <i>Zero-Shot Learning</i>	11
2.2.5 <i>Prompt</i>	11
2.2.6 Ollama	12
2.2.7 LLAMA3	12
BAB III METODE PENELITIAN	15
3.1 Metode Penelitian	15

3.1.1	<i>Studi Literatur</i>	15
3.1.2	Pengumpulan Dataset.....	16
3.1.3	Pemilihan dan Konfigurasi Model	18
3.1.4	<i>Zero-shot Learning</i> Model	19
3.1.5	Evaluasi Model.....	20
3.2	Analisis Sistem.....	22
3.3	Identifikasi Perangkat Lunak	23
3.4	Perancangan <i>User Interface</i>	26
BAB IV HASIL DAN ANALISIS PENELITIAN		28
4.1	Hasil Penelitian	28
4.1.1	Tampilan Halaman Utama	28
4.1.2	Halaman <i>Input</i> kalimat permasalahan.....	29
4.1.3	Halaman Hasil Prediksi.....	29
4.2	Analisa Penelitian.....	31
4.3	Hasil <i>Zero-shot Learning</i>	32
4.4	Hasil Evaluasi.....	33
BAB V KESIMPULAN DAN SARAN		37
5.1	Kesimpulan	37
5.2	Saran.....	37
DAFTAR PUSTAKA		

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur Llama	13
Gambar 3. 1 Alur Penelitian.....	15
Gambar 3. 2 Perancangan sistem	22
Gambar 3. 3 Halaman Utama.....	26
Gambar 3. 4 Halaman Prediksi	27
Gambar 4. 1 Halaman Utama.....	28
Gambar 4. 2 Halaman Input kalimat permasalahan	29
Gambar 4. 3 Contoh kalimat prediksi-solution.....	29
Gambar 4. 4 Contoh kalimat Tantangan-Jawaban.....	30
Gambar 4. 5 Contoh kalimat Peluang-Jawaban.....	30
Gambar 4. 6 Contoh Kalimat Kelemahan-Peningkatan.....	31



DAFTAR TABEL

Tabel 3. 1 Kalimat Permasalahan Per Kategori Dalam Dataset Evaluasi.....	17
Tabel 4. 1 Hasil Prediksi sistem dan reference	33
Tabel 4. 2 Hasil Evaluasi ROUGE.....	35
Tabel 4. 3 Hasil Evaluasi rata-rata ROUGE	35



ABSTRAK

Peningkatan jumlah publikasi ilmiah yang mencapai 2,6 hingga 3 juta artikel per tahun menimbulkan tantangan dalam mengekstraksi informasi penting, terutama dalam mengidentifikasi permasalahan dan solusinya yang sering kali tidak dinyatakan secara eksplisit. Proses manual yang memakan waktu dan rentan terhadap kesalahan manusia, sehingga mendorong perlunya solusi berbasis kecerdasan buatan. Penelitian ini mengusulkan pendekatan *Zero-Shot Learning* (ZSL) menggunakan model Llama 3 untuk memprediksi solusi dari kalimat masalah dalam artikel ilmiah tanpa memerlukan pelatihan tambahan. Model ini hanya mengandalkan pemrosesan berbasis *prompt* untuk menghasilkan solusi tanpa menggunakan dataset pelatihan. Evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap solusi referensi (*ground truth*) menggunakan metrik ROUGE. Hasil evaluasi menunjukkan bahwa skor ROUGE-1, ROUGE-2, dan ROUGE-L berada dalam rentang 3,5% hingga 22%, mengindikasikan bahwa meskipun kesamaan berbasis kata masih terbatas, model tetap mampu menghasilkan solusi yang relevan secara kontekstual. Dengan demikian, sistem berbasis *Zero-Shot Learning* dan *Large Language Model* ini diharapkan dapat mendukung analisis pola solusi dalam artikel ilmiah serta menjadi dasar bagi pengembangan model prediksi yang lebih akurat di masa depan.

Kata Kunci : *Zero-shot Learning*, Prediksi Solusi, *Large Language Model*, Artikel ilmiah

ABSTRACT

The increasing number of scientific publications reaching 2.6 to 3 million articles per year creates challenges in extracting important information, especially in identifying problems and solutions that are often not stated explicitly. Manual processes are time-consuming and prone to human error, driving the need for artificial intelligence-based solutions. This research proposes a *Zero-Shot Learning* (ZSL) approach using the Llama 3 model to predict solutions to problem sentences in scientific articles without additional training. This model relies solely on prompt-based processing to generate solutions without using a training dataset. Evaluation is carried out by comparing the model prediction results against the reference solution (*ground truth*) using the ROUGE metric. The evaluation results show that the ROUGE-1, ROUGE-2, and ROUGE-L scores are in the range of 3.5% to 22%, indicating that although word-based similarity is still limited, the model can still produce contextually relevant solutions. Thus, it is hoped that this *Zero-Shot Learning* and *Large Language Model*-based system can support the analysis of solution patterns in scientific articles and become the basis for developing more accurate prediction models in the future.

Keywords: *Zero-shot Learning*, Solution Prediction, *Large Language Model*, Scientific articles

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam perkembangan ilmu pengetahuan dan teknologi, artikel ilmiah merupakan sumber informasi yang penting. Jumlah artikel ilmiah yang diterbitkan terus meningkat secara signifikan setiap tahun. Menurut penelitian yang dilakukan oleh (Mendes-Da-Silva dan Leal, 2021), publikasi ilmiah mencapai sekitar 2,6 hingga 3 juta artikel setiap tahun, dengan tingkat pertumbuhan global sekitar 3,8 sejak tahun 2008. Peningkatan pesat ini menimbulkan tantangan baru untuk mengelola dan mengekstrak informasi penting dari artikel-artikel tersebut.

Kalimat solusi merupakan pernyataan yang menjawab atau penyelesaian masalah penelitian. Identifikasi masalah dan solusinya adalah dua komponen penting dalam sebuah artikel ilmiah, yang sering dicari oleh peneliti dan praktisi. Penelitian (Gao dkk., 2021), menemukan bahwa peneliti rata-rata menghabiskan 12 jam setiap minggu untuk membaca dan menganalisis artikel untuk penelitian mereka. Selain memakan waktu, proses manual juga rentan terhadap kesalahan manusia. Banyak artikel ilmiah tidak menampilkan masalah dan solusinya dengan jelas, membuat pembaca sulit merencanakan penelitian selanjutnya, terutama bagi peneliti pemula.

Perkembangan teknologi kecerdasan buatan, khususnya *Large Language Models* (LLM), menawarkan peluang baru dalam pengolahan dan analisis teks. Penelitian yang dilakukan oleh (Brown dkk., 2020), menemukan bahwa *Large Language Model* (LLM) memiliki kemampuan yang menjanjikan untuk memahami konteks dan struktur artikel ilmiah dengan akurasi 89%. Namun, penggunaan LLM konvensional menghadapi tantangan utama berupa kebutuhan data training yang besar dan *resource* komputasi yang intensif.

Zero-shot learning muncul sebagai potensial untuk mengatasi keterbatasan tersebut. (Badawi dkk., 2024), dalam publikasinya menunjukkan bahwa pendekatan *zero-shot learning* memungkinkan model untuk mengklasifikasikan data yang belum terlihat tanpa memerlukan pelatihan ulang yang kompleks. Hal ini mengatasi

keterbatasan metode tradisional yang bergantung pada dataset berlabel lengkap. Fleksibilitas ZSL membuatnya lebih efektif dalam menganalisis artikel ilmiah dan menemukan solusi untuk masalah yang tidak dijelaskan eksplisit dalam literatur.

Beberapa peneliti sebelumnya menunjukkan potensi penggunaan model bahasa besar (LLM) dalam konteks *Zero-shot Learning*. Misalnya, penelitian yang dilakukan oleh (Brown dkk., 2020), menunjukkan bahwa model GPT-2 dapat menghasilkan teks yang relevan dalam konteks yang belum pernah dilihat sebelumnya, menunjukkan potensi besar untuk aplikasi dalam prediksi kalimat solusi. Selain itu, penelitian oleh (Brown dkk., 2020) mengungkapkan bahwa model GPT-3 mampu melakukan berbagai tugas bahasa dengan sedikit atau tanpa pelatihan tambahan, yang menunjukkan keunggulan dalam fleksibilitas model dalam konteks *Zero-shot Learning*.

Berdasarkan tantangan dan peluang tersebut, penelitian ini mengusulkan implementasi *zero-shot learning* dengan model *Large Language Model* (LLM) untuk prediksi kalimat solusi dari masalah pada artikel ilmiah. Pendekatan ini diharapkan dapat mengatasi keterbatasan dataset berlabel serta mempertahankan akurasi yang tinggi dalam ekstraksi solusi. Salah satu model yang digunakan adalah Ollama, yang dirancang untuk memproses dan menganalisis teks artikel ilmiah serta menghasilkan kalimat solusi yang relevan berdasarkan pemahaman model terhadap masalah yang ada.

Implementasi sistem ini diharapkan dapat memberikan kontribusi signifikan dalam meningkatkan efisiensi penelitian ilmiah, mengoptimalkan pemanfaatan pengetahuan yang ada. Sistem berbasis *zero-shot learning* yang menggunakan model *Large Language Model* (LLM) seperti Ollama dapat menjadi dasar untuk pengembangan sistem yang lebih komprehensif dalam analisis artikel ilmiah dan prediksi solusi.

1.2 Perumusan Masalah

Bagaimana menerapkan LLM (*Large Language Model*) yang disediakan oleh Ollama untuk mengimplementasikan *Zero-Shot Learning* dalam memprediksi kalimat solusi dari masalah pada artikel ilmiah?

1.3 Pembatasan Masalah

Adapun batasan masalah dari penulisan proposal ini yaitu sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini hanya berupa kalimat dari artikel ilmiah berbahasa Indonesia yang diambil dari Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK). Dataset hanya digunakan untuk keperluan evaluasi model.
2. Penelitian ini hanya mencakup empat kategori permasalahan, yaitu problem-solution, tantangan-jawaban, peluang-jawaban, dan kelemahan-peningkatan.
3. Model yang digunakan dalam penelitian ini adalah *Large Language Model* (LLM) dengan pendekatan *Zero-Shot Learning* melalui platform Ollama.
4. Penelitian ini hanya memprediksi solusi dalam ruang lingkup bidang Teknik Informatika.
5. Dataset akan direview secara manual oleh peneliti dan divalidasi oleh dosen pembimbing untuk memastikan keakuratan serta kesesuaian data dengan kategori permasalahan yang telah ditentukan.

1.4 Tujuan

Membuat sistem untuk menghasilkan kalimat solusi dari masalah pada artikel ilmiah dengan mengimplementasikan *Zero-shot Learning* menggunakan *Large Language Model* (LLM).

1.5 Manfaat

Manfaat dari Tugas Akhir ini adalah sebagai berikut :

1. Membantu pembaca, peneliti, dan akademisi untuk mendapatkan solusi masalah dalam artikel ilmiah dengan lebih cepat dan efisien.
2. Memberikan kontribusi dalam pengembangan metode *Zero-Shot Learning* untuk pemrosesan artikel ilmiah berbahasa Indonesia dengan memanfaatkan teknologi LLM secara lokal melalui platform Ollama.
3. Memudahkan pembaca artikel ilmiah dalam menemukan kalimat solusi untuk masalah, tantangan, peluang, dan kelemahan yang dibahas, sehingga dapat mendukung pengambilan keputusan dalam penelitian.

1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan oleh penulis dalam sebuah laporan tugas akhir adalah sebagai berikut:

BAB I : PENDAHULUAN

Pada bab ini, penulis menjelaskan urgensi penelitian yang dilakukan, dimulai dengan penyusunan latar belakang, perumusan masalah, pembatasan masalah, serta penetapan tujuan dan manfaat yang diharapkan. Selain itu, sistematika penulisan juga disajikan untuk memberikan gambaran struktur penulisan secara keseluruhan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Bab ini memuat landasan teori yang mendukung penelitian serta referensi sebelumnya yang berkontribusi dalam perancangan sistem. Selain itu, bagian ini membantu penulis dalam memahami konsep yang berkaitan dengan *Zero-shot Learning*, *Large Language Model* (LLM), dan Ollama selama proses penelitian.

BAB III : METODE PENELITIAN

Bab ini menjelaskan tahapan penelitian, dimulai dari proses pengumpulan data hingga tahap pemodelan yang digunakan dalam penelitian. Serta membahas analisis terhadap proses sistem yang akan dikembangkan, termasuk perancangan sistem *website* serta desain antarmuka (*Interface Design*) yang dirancang.

BAB IV : HASIL DAN ANALISIS PENELITIAN

Bab ini menyajikan temuan penelitian, yaitu hasil implementasi *Zero-shot Learning* dalam memprediksi kalimat solusi dari permasalahan yang terdapat dalam artikel ilmiah menggunakan *Large Language Model* (LLM).

BAB V : KESIMPULAN DAN SARAN

Bab ini, penulis menyajikan kesimpulan dan saran dari penulis terhadap penelitian yang telah dilakukan.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Dalam sebuah penelitian yang dilakukan Riordan dan kawan-kawan mengatakan pada penerapan *zero-shot learning* (ZSL) menggunakan *large language model* (LLM) untuk menganalisis data teks terkait penggunaan zat adiktif, terutama dalam konteks media sosial. Dengan ZSL, model dapat mengklasifikasi data baru tanpa pelatihan khusus untuk tugas tertentu, sehingga memungkinkan analisis cepat dan fleksibel tanpa membutuhkan anotasi manual. Pendekatan ini, memungkinkan untuk analisis data dalam jumlah besar secara efisien dan akurat, serta memberikan peluang untuk menjawab pertanyaan penelitian baru secara cepat dan efektif (Riordan dkk., 2024).

Dalam penelitian yang kedua yang dilakukan Alhoshan dan kawan-kawan, dalam penelitian ini mereka menyelidiki potensi *Zero-Shot Learning* (ZSL) untuk klasifikasi kebutuhan, yang sering terhambat oleh kurangnya data berlabel dalam rekayasa kebutuhan. Menunjukkan hasil bahwa ZSL dapat mencapai kinerja yang diterima, dengan F1-skor antara 66% hingga 80% untuk tugas klasifikasi biner dan multi-kelas, tanpa memerlukan data pelatihan. Model bahasa umum yang dipadukan dengan konfigurasi label sederhana sering kali lebih unggul dibandingkan model bahasa khusus domain dengan konfigurasi label yang kompleks. Hal ini menunjukkan bahwa ZSL menawarkan solusi yang menjanjikan untuk masalah kelangkaan data dalam klasifikasi persyaratan, memungkinkan klasifikasi yang efektif bahkan dengan data pelatihan yang terbatas atau tanpa data pelatihan (Alhoshan dkk., 2023).

Dalam penelitian yang ketiga yang dilakukan Diky Dermawan, dengan menggunakan *prompt engineering* khusus pada *zero-shot* menunjukkan bahwa teknik ini memiliki potensi besar untuk membantu *Large Language Models* (LLMs) beradaptasi dengan tugas-tugas baru tanpa memerlukan pelatihan tambahan. *Zero-shot* memungkinkan LLMs untuk menjalankan berbagai tugas hanya berdasarkan instruksi *prompt* yang diberikan, meskipun model tersebut belum dilatih secara

khusus untuk tugas tersebut sebelumnya. *Zero-shot* masih menghadapi beberapa tantangan, seperti kecenderungan model untuk menghasilkan informasi yang salah, karena kurangnya pemahaman mendalam terhadap konteks. Hal ini menekankan perlunya strategi *prompt* yang dirancang dengan hati-hati untuk meminimalkan kesalahan dan memastikan hasil yang lebih akurat (Dermawan dan Herdianto, 2024).

Dalam penelitian yang keempat yang dilakukan Plaza dan kawan-kawan, mereka menunjukkan bahwa *Zero-Shot Learning* dengan *prompting* merupakan metode yang menjanjikan dalam mengidentifikasi ujaran kebencian, dengan kemampuan model untuk beradaptasi lintas bahasa dan mencapai hasil yang signifikan tanpa pelatihan spesifik. Studi ini menekankan pentingnya pemilihan model dan formulasi *prompt* yang tepat, serta potensi model bahasa berbasis instruksi *fine-tuning* dalam mendeteksi konten berbahaya, memberikan wawasan mendalam tentang tantangan dan peluang dalam klasifikasi ujaran kebencian menggunakan teknologi kecerdasan buatan (Plaza-Del-Arco dkk., 2023).

Dalam penelitian kelima yang dilakukan Rohit Bhuvaneshwar Mishra dan Hongbing Jiang, bahwa pada penelitian dengan menggunakan model *Machine Learning* (ML) dan *Deep Neural Networks* (DNN) sangat efektif untuk mengklasifikasikan string *problem* dan *solution* dalam teks ilmiah. Model yang baik digunakan, yaitu *Convolutional Neural Network* (CNN), mencapai akurasi 90% untuk klasifikasi *problem* dan 86% untuk klasifikasi *solution*. Hasil ini memperkuat bahwa CNN sangat cocok untuk menganalisis pola *problem-solution* (Mishra dan Jiang, 2021).

Dalam penelitian keenam yang dilakukan Nuzhul Citrasari Dewi dan Anita Qoiriah, menunjukkan bahwa kombinasi algoritma Jaro-Winkler Distance dan N-Gram untuk prediksi perbaikan kata bahasa Indonesia menghasilkan akurasi tertinggi sebesar 85.7% dan terendah 45%. Sistem ini efektif dalam mendeteksi kesalahan penulisan, meskipun terkadang juga mengidentifikasi kata yang sebenarnya tidak salah karena ketergantungannya pada korpus. Kualitas dan kuantitas korpus sangat mempengaruhi hasil deteksi, dimana semakin besar

kuantitas korpus/kamus n-gram, semakin baik sistem dapat menemukan prediksi kata yang sesuai dengan kata yang salah pada kalimat (Dewi dan Qoiriah, 2021).

Dalam penelitian ketujuh yang dilakukan oleh Tamer Shanableh, dalam penelitiannya mengembangkan metode pengenalan Bahasa isyarat Arab berbasis visi secara kontinu dengan pendekatan dua tahap: prediksi jumlah kata dalam kalimat dan pengenalan setiap kata. Dengan menggunakan kompensasi gerak dan *CNN transfer learning*, metode ini mencapai akurasi pengenalan kata 97,3% dan kalimat 92,6% pada dataset kalimat. Solusi ini menghilangkan kebutuhan pelabelan manual setiap frame video, cukup dengan mengetahui arti kata isyarat (Shanableh, 2023).

Dalam penelitian kedelapan yang dilakukan oleh Ke-Li Chiu dan kawan-kawan, dalam penelitiannya mampu mengeksplorasi kemampuan model Bahasa GPT-3 dalam ujaran kebencian melalui pendekatan *zero-shot*, *one-shot*, dan *few-shot learning*. Penelitian ini menggunakan dataset ETHOS (*haTe speech detectiOn dataset*), yang berisi teks daring yang dikategorikan berdasarkan muatan ujaran kebencian. Hasil penelitian menunjukkan bahwa performa model bervariasi tergantung pada pendekatan pembelajaran yang digunakan. Pada *zero-shot learning*, akurasi rata-rata yang dicapai adalah 56%, dengan F1-score sebesar 70%. Sementara itu, untuk *one-shot learning*, performa menurun dengan akurasi rata-rata sebesar 55% dan F1-score sebesar 55%. Sebaliknya, pada *few-shot learning*, performa meningkat dengan akurasi rata-rata sebesar 67% dan F1-score sebesar 62%. Hasil ini menyoroti bahwa GPT-3 dapat mengidentifikasi ujaran kebencian dengan efektivitas yang bervariasi berdasarkan jumlah informasi pelatihan yang diberikan (Chiu dkk., 2021).

Dalam penelitian kesembilan yang dilakukan oleh Hajime Sasaki dan kawan-kawan, dalam penelitian mereka mengembangkan model praktis berbasis model bahasa dengan mekanisme atensi untuk mengidentifikasi kalimat masalah dalam karya ilmiah, menghubungkan pengetahuan antara sains dan teknologi industri. Model ini menunjukkan kinerja yang lebih tinggi dibandingkan metode sebelumnya, terutama dalam artikel ilmiah. Penelitian ini mencakup pengelompokan kalimat masalah, perhitungan kemiripan semantik, dan analisis pasangan masalah dengan

kesamaan konteks, yang memungkinkan pemahaman lebih dalam tentang latar belakang masalah yang sulit diperoleh hanya melalui pencarian kata kunci. Dengan menggunakan data dari artikel ilmiah dalam bidang linguistik komputasional (*ACL Anthology*) dan dokumen paten (*Derwent Innovation*), penelitian ini menerapkan metode pengklasteran hierarkis dan *embedding* kata untuk menilai kesamaan masalah. Walaupun terdapat keterbatasan seperti ketidaksesuaian tahun publikasi dan panjang kalimat, penelitian ini berkontribusi pada identifikasi isu-isu penting dalam sains dan teknologi, serta memberikan wawasan baru untuk tren penelitian ilmiah dan target bisnis masa depan (Sasaki dkk., 2020).

Dalam penelitian selanjutnya yang dilakukan oleh Priya dan Sachin Gupta, dalam penelitiannya memanfaatkan model GPT-3 untuk mendeteksi ujaran kebencian dan mengklasifikasikan teks menjadi kalimat positif atau negatif. Pendekatan *zero-shot learning*, *one-shot learning*, *few-shot learning*. Hasil penelitian menunjukkan performa model yang berbeda pada setiap pendekatan. Skor akurasi yang diperoleh adalah 45% untuk *zero-shot learning*, meningkat menjadi 72% pada *one-shot learning*, dan mencapai 80% pada *few-shot learning*. Hasil ini mengindikasikan bahwa menambahkan jumlah contoh pelatihan secara signifikan meningkatkan kemampuan GPT-3 dalam memahami dan memprediksi teks berbasis ujaran kebencian (Priya dan Gupta, 2022).

2.2 Dasar Teori

2.2.1 Kalimat Solusi

Kalimat solusi merupakan bagian penting dalam artikel ilmiah karena menjelaskan langkah atau cara yang ditawarkan untuk mengatasi permasalahan yang dibahas. Kalimat solusi memberikan jawaban atas tantangan penelitian dalam bentuk tindakan atau strategi spesifik yang menunjukkan kontribusi penelitian dalam menyelesaikan masalah. Kalimat solusi biasanya ditemukan di bagian abstrak, pembahasan, atau kesimpulan, dan menjadi inti dari jawaban terhadap pernyataan yang diajukan. Namun, terkadang kalimat yang menyatakan solusi dari permasalahan tidak tertulis secara eksplisit atau bahkan tidak dibahas sama sekali dalam artikel ilmiah. Hal ini dapat menyebabkan pembaca kesulitan

memahami kontribusi penelitian dalam menyelesaikan masalah yang diangkat. Oleh karena itu, kalimat solusi dalam pernyataan masalah harus secara jelas mengartikulasikan manfaat dari solusi yang diusulkan, mengatasi tantangan penelitian melalui langkah-langkah atau metode tertentu yang secara langsung berkontribusi dalam menyelesaikan masalah yang diidentifikasi, sehingga meningkatkan kejelasan dan keefektifan pertanyaan secara keseluruhan (Rustipa dkk., 2023).

2.2.2 *Natural Language Processing (NLP)*

NLP merupakan cabang penting dari kecerdasan buatan yang fokusnya pada bagaimana komputer dapat memahami dan memproses bahasa manusia. NLP juga mempelajari cara bahasa alami berinteraksi dengan manusia dan komputer. Sejarah NLP dimulai pada tahun 1950 an, ketika penelitian ini menggunakan metode berbasis aturan untuk membangun sistem seperti terjemahan mesin (MT), tanya jawab (QA), dan analisis kata atau kalimat (Zhou dkk., 2020).

NLP terdiri dari dua bagian yaitu, Generasi Bahasa Alami, yang memiliki kemampuan untuk mengembangkan kemampuan untuk memahami dan menghasilkan teks, dan Generasi Bahasa Alami, yang memiliki kemampuan untuk analisis, interpretasi, dan manipulasi data bahasa alami dengan berbagai metode dan algoritma alat. Terjemahan mesin, deteksi spam email, ekstraksi informasi, peringkasan, menjawab pertanyaan, merupakan beberapa contoh bidang di mana pemrosesan bahasa natural dapat digunakan (Khurana dkk., 2023).

Perkembangan pesat dalam teknologi telah menjadikan NLP sebagai bidang penelitian yang semakin mendalam dan terus berkembang. Kemampuan NLP untuk mengolah dan memahami teks dalam bahasa manusia menjadi sangat efektif untuk diterapkan dalam berbagai aplikasi. NLP mencakup berbagai teknik dan metode yang digunakan untuk memproses teks agar lebih mudah dipahami oleh mesin. Pemahaman tentang perhitungan manual dan prinsip dasar metode sangat penting untuk mendapatkan hasil terbaik, meskipun banyak framework dan alat NLP secara *open source*. Dengan memanfaatkan NLP, developer bisa menyusun dan mengelola informasi untuk berbagai tujuan, seperti peringkasan dokumen atau analisis teks secara otomatis (Harumy dkk., 2024).

2.2.3 Large Language Model (LLM)

Large Language Model (LLM) adalah jenis kecerdasan buatan yang menggunakan pendekatan pembelajaran mesin untuk menghasilkan teks yang serupa dengan bahasa manusia. Model ini dilatih menggunakan sejumlah data teks yang besar dan menerapkan teknik *deep learning* untuk memahami pola dan struktur bahasa. Karena kemampuannya dalam memahami konteks dan menghasilkan teks yang relevan, LLM banyak digunakan dalam berbagai tugas berbasis bahasa, seperti pembuatan teks, terjemahan, ringkasan, menjawab pertanyaan, penulisan kode, dan analisis sentimen (Noor Kamala Sari dkk., 2024).

Selain itu, LLM membantu penyederhanaan model dengan memberikan pemahaman yang lebih baik terhadap konteks suatu kata. Hal ini menjadi komponen penting dalam analisis teks karena kombinasi antara metode pembersihan data dan penerapan LLM telah terbukti dapat meningkatkan hasil klasifikasi dalam berbagai sistem berbasis teks. Oleh karena itu, LLM menjadi alat yang sangat efektif untuk meningkatkan akurasi dan relevansi hasil analisis (Zulfikar dkk., 2025).

Salah satu contoh LLM yang paling terkenal adalah Chat GPT. Model ini menggunakan arsitektur transformer dan telah dilatih dengan jumlah data yang besar, sehingga dapat menghasilkan teks yang mirip dengan tulisan manusia (Topsakal dan Akinci, 2023). Oleh karena itu, model ini disebut sebagai model bahasa besar. Sebagian besar LLM dikembangkan oleh perusahaan teknologi, di mana banyak aspek latihannya, seperti pengumpulan dan pembersihan data. Model ini biasanya memiliki struktur transformer dengan banyak lapisan yang ditumpuk dalam jaringan saraf yang sangat dalam. Untuk meningkatkan efisiensi dan skalabilitas model, berbagai strategi paralel diterapkan secara bersamaan selama proses pelatihannya. Dalam hal ini, beberapa kerangka kerja optimasi, seperti *DeepSpeed* dan *Megatron-LM*, telah dikembangkan untuk mendukung pelatihan terdistribusi dan mempermudah implementasi dan penerapan algoritma paralel.

Selain arsitektur dan pelatihannya, teknik pengoptimalan juga sangat penting untuk performa model dan stabilitas pelatihan, seperti memulai ulang untuk mengatasi lonjakan kerugian pelatihan. Berdasarkan informasi terkini, GPT-4

menyarankan pembuatan infrastruktur dan teknik pengoptimalan khusus yang dapat dengan akurat memprediksi kinerja model yang jauh lebih kecil. Setelah melalui proses pelatihan menggunakan korpora berskala besar, LLM memiliki potensi untuk menyelesaikan berbagai tugas spesifik dengan tingkat akurasi yang lebih tinggi (Alhoshan dkk., 2023).

2.2.4 Zero-Shot Learning

Zero-Shot Learning adalah pembelajaran yang sedang berkembang dengan tujuan untuk menyelesaikan tugas tanpa memerlukan data pelatihan sebelumnya. Model Bahasa besar (LLM) merupakan pilihan yang tepat untuk mendukung *Zero-Shot Learning* (ZSL) karena memiliki kemampuan untuk berfungsi sebagai sumber informasi tambahan yang kuat. Dengan keakuratan dan pemahaman kontekstual yang dimiliki, LLM dapat langsung diterapkan untuk memprediksi data baru tanpa memerlukan pelatihan tambahan (Alhoshan dkk., 2023).

Zero-shot Learning memungkinkan pengklasifikasian kebutuhan baru berdasarkan label yang tersedia, menjadikan solusi yang menjanjikan untuk aplikasi yang membutuhkan efisiensi dan skalabilitas tinggi. Pendekatan ini memberikan efisiensi lebih baik dalam pengklasifikasian objek baru, sekaligus meningkatkan akurasi dan efektivitas dalam berbagai skenario aplikasi (Alhoshan dkk., 2023).

2.2.5 Prompt

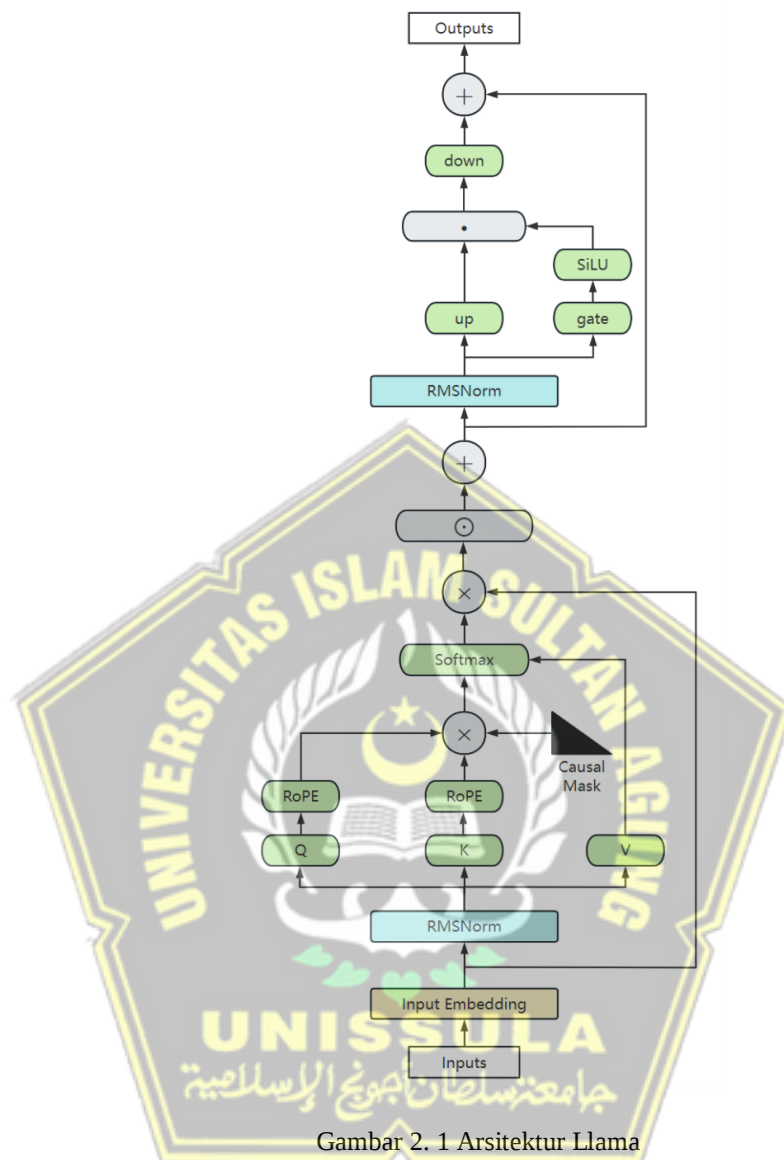
Prompt merupakan serangkaian intruksi yang diberikan kepada LLM untuk mengarahkan dan meningkatkan kemampuannya (Liu dkk., 2023). Dengan memberikan garis besar aturan awal untuk percakapan LLM, *prompt* dapat mempengaruhi interaksi selanjutnya dengan dihasilkannya keluaran dari LLM. Secara khusus, *prompt* menetapkan konteks percakapan dan memberitahu LLM informasi apa yang penting dan bentuk keluaran serta konten yang diinginkan. Misalnya, *prompt* dapat menetapkan bahwa LLM hanya boleh membuat kode dengan gaya pengkodean pemrograman tertentu. Hal untuk mendapatkan bahwa LLM harus memberikan penjelasan tambahan tentang kata kunci tertentu. Dengan menggunakan ini maka *output* yang diberikan terstruktur (White dkk., 2023).

2.2.6 Ollama

Ollama merupakan sebuah perangkat lunak yang digunakan untuk menjalankan model kecerdasan buatan berbasis LaMA (*Large Model for Language Applications*) secara langsung melalui terminal komputer, dengan kemampuan utama untuk mengeksekusi model bahasa besar LLM secara lokal (Peña dan Ortega-Castro, 2024). Ollama bertujuan untuk menjadikan perantara antara kompleksitas teknologi LLM dan kebutuhan pengguna akan solusi AI (*Artificial Intelligence*) yang mudah diakses dan dapat disesuaikan. Ollama dapat digunakan untuk mengekstrak dan mengatur pengetahuan dari sumber data yang tidak terstruktur, seperti artikel, jurnal, atau laporan penelitian, yang dapat memudahkan peneliti dalam mengakses informasi yang relevan.

2.2.7 LLAMA3

LLM telah mengembangkan kemampuan untuk berinteraksi dan memproses banyak data teks. Dengan menerjemahkan bahasa, menganalisis dokumen, dan menghasilkan teks yang mirip dengan teks manusia, Model-model ini telah menunjukkan kemampuan yang sangat luar biasa. Munculnya model terbuka seperti Llama3 menunjukkan perubahan besar dalam penggunaan. dengan LLM lokal yang menggunakan paket Ollama sebagai antara muka untuk pengambilan model. Dari 24 uji coba eksperimen, ditemukan 11 kasus *output* terstruktur berhasil sepenuhnya dan menghasilkan jawaban. Karena Llama3 memiliki kinerja yang sebanding dengan pesaingnya, ini menunjukkan kemandirian model (Shorten dkk., 2024). Gambar 2.1 merupakan arsitektur Llama.



Gambar 2. 1 Arsitektur Llama

Llama 3 memperkenalkan mekanisme *Group Query Attention* untuk mengoptimalkan penggunaan memori, sehingga secara signifikan mengurangi kebutuhan memori dan komputasi serta meningkatkan efisiensi operasional. Setelah proses *self-attention*, terdapat *residual connection* yang menghubungkan kembali keluaran yang dimasukkan. Selanjutnya, *feedforward neural network* (FFN), yang terdiri dua lapisan linear dan fungsi aktivasi (misal SiLU), memanfaatkan hasilnya untuk meningkatkan deteksi fitur non-linear. Proses ini berakhir dengan normalisasi RMSNorm Final. Menghasilkan *array* probabilitas login untuk prediksi token berdasarkan pemilihan strategis (Chen dkk., 2024).

Arsitektur model Llama unggul dalam desainnya yang ringkas dan efisien karena berfokus pada tugas-tugas generatif melalui kerangka kerja *Decoder-only*. Dengan demikian, pelatihan menjadi lebih mudah dan prosesnya lebih cepat. Penggunaan teknik positional encoding yang inovatif dan RMSNorm meningkatkan stabilitas dan kemampuan ekspresif model, memungkinkan generalisasi yang lebih baik dan efisiensi dalam pemrosesan dataset berskala besar. Model ini sangat efektif dalam tugas-tugas pemrosesan bahasa alami (NLP) karena memiliki mekanisme perhatian dan jaringan *feedforward* yang dirancang dengan cermat. ini membuatnya mampu menangkap dan memahami fitur bahasa yang kompleks (Chen dkk., 2024).

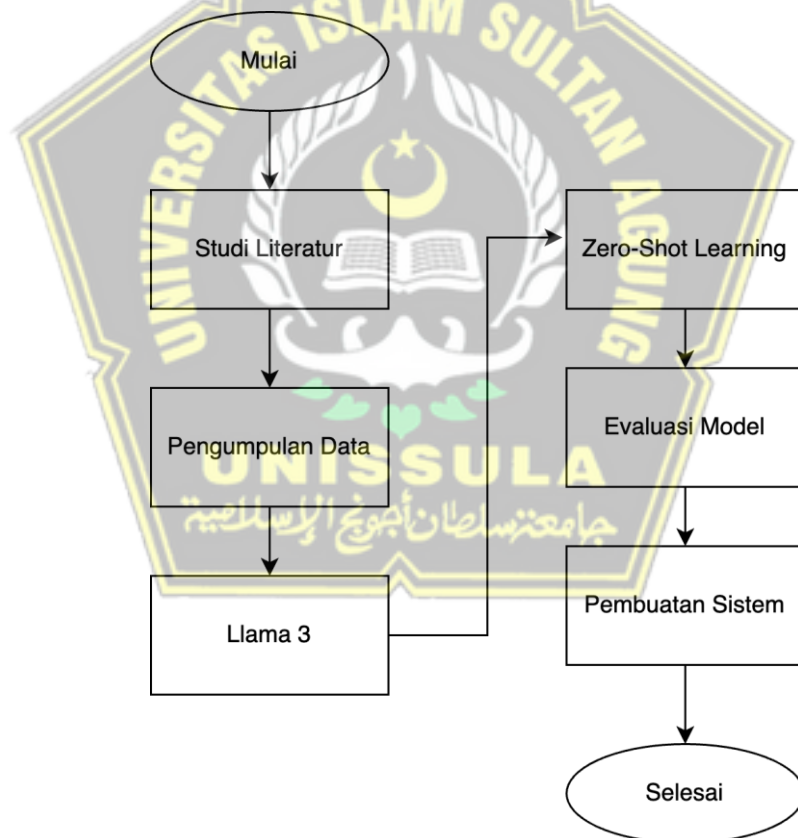


BAB III

METODE PENELITIAN

3.1 Metode Penelitian

Pada penelitian ini, metode yang digunakan adalah *Zero-shot learning* dan *Large Language Model* (LLM). Dengan menggabungkan kedua metode tersebut, penelitian ini bertujuan untuk memprediksi kalimat solusi dari masalah yang ditemukan dalam artikel ilmiah. Pendekatan *Zero-shot learning* memanfaatkan kemampuan bawaan model *Large Language Model* (LLM) tanpa memerlukan pelatihan tambahan, yang memungkinkan prediksi dilakukan secara langsung. Adapun tahapan penelitian yang dilakukan sebagai berikut :



Gambar 3. 1 Alur Penelitian

3.1.1 Studi Literatur

Peneliti akan meninjau berbagai literatur, termasuk artikel, tesis dan skripsi, serta sumber dari berbagai situs *website* terkait dengan topik yang relevan. Studi

literatur ini bertujuan untuk mempelajari teori terkait dengan *Zero-Shot Learning* (ZSL), *Large Language Model* (LLM), teknik prediksi berbasis teks, serta penerapan model pembelajaran mesin dalam menganalisis solusi dari berbagai masalah pada artikel ilmiah.

3.1.2 Pengumpulan Dataset

Pengumpulan dataset dilakukan secara manual karena belum tersedia dataset yang sesuai dengan kebutuhan penelitian. Dataset ini dibentuk dengan mereview 100 artikel dari Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK), yang diberikan oleh dosen pembimbing. Dataset ini terdiri dari 20 data yang akan digunakan dalam proses evaluasi penelitian.

Tahap pengumpulan Data:

a. Identifikasi Permasalahan dalam Artikel

- Proses dimulai dengan mengkaji permasalahan yang terdapat pada artikel yang telah direview.
- Setiap kalimat yang mengandung permasalahan dan solusinya akan dicatat untuk membentuk dataset evaluasi.

b. Kategori Data

Dataset dikelompokkan ke dalam beberapa kategori utama:

- *Problem-solution*
- Tantangan-jawaban
- Peluang-jawaban
- Kelemahan-peningkatan

c. Pengambilan Kalimat Permasalahan dan Solusi

- Jika dalam artikel terdapat lebih dari satu kalimat yang menyatakan permasalahan, semua kalimat tersebut akan diambil dan dimasukkan ke dalam dataset.
- Jika dalam artikel terdapat kalimat permasalahan beserta solusinya, keduanya akan diambil dan dimasukkan ke dalam dataset.
- Jika ditemukan kalimat yang menyatakan permasalahan tanpa solusi di artikel tersebut, maka solusi dibuat berdasarkan asumsi yang relevan dan mengarah

pada penyelesaian masalah. Asumsi ini akan divalidasi oleh dosen pembimbing sebelum dimasukkan ke dalam dataset evaluasi.

d. Validasi Dataset

Evaluasi Setelah seluruh data terkumpul, dataset akan divalidasi oleh dosen pembimbing untuk memastikan kualitasnya sesuai dengan kebutuhan penelitian.

e. Link Akses Artikel

https://drive.google.com/drive/folders/1XsB0msl-S9PWNlgRFd-tzN0FX1YPn6dF?usp=drive_link.

Berikut ini adalah contoh kalimat permasalahan berdasarkan masing-masing kategori untuk data evaluasi.

Tabel 3. 1 Kalimat Permasalahan Per Kategori Dalam Dataset Evaluasi

No	Kategori	<i>Problem</i>	<i>Solution</i>
1	<i>Problem-Solution</i>	Proses pengelolaan data sensus harian rawat inap masih menggunakan metode manual, yang mengakibatkan ketidakakuratan data dan kesulitan pencarian data	Untuk mengatasi hal tersebut, dikembangkan sistem automatic indikator rumah sakit berbasis VB.NET yang mempermudah pengolahan data secara otomatis dan mengurangi kesalahan pencatatan
2	Tantangan-Jawaban	Transisi ke rekam medis elektronik menjadi langkah penting, namun juga menghadirkan tantangan besar dalam hal pengelolaan data	Dengan menerapkan metode Agile, rumah sakit memiliki kesempatan untuk mengatasi masalah operasional, meningkatkan ketepatan data, dan menyederhanakan proses pelayanan pasien
3	Peluang-Jawaban	E-commerce yang berkembang pesat membuka peluang bagi	Dengan meningkatnya penggunaan layanan pembayaran digital, penyedia fintech dapat

		fintech untuk memudahkan pembayaran digital di masa mendatang	memanfaatkan tren ini untuk menarik lebih banyak pengguna dengan menyediakan layanan yang aman dan efisien
4	Kelemahan-Peningkatan	Implementasi algoritma Multinomial Naive Bayes membutuhkan data besar untuk hasil yang akurat.	Perlu memastikan ketersediaan data yang cukup sebelum mengimplementasikan metode ini.

3.1.3 Pemilihan dan Konfigurasi Model

Pada penelitian ini, model yang digunakan adalah model berbasis *Large Language Model* (LLM) dengan pendekatan *Zero-shot Learning*. Model ini dipilih karena kemampuannya dalam memahami konteks Bahasa alami tanpa perlu pelatihan ulang dengan dataset khusus. Dengan menggunakan Teknik *prompt engineering*, model dapat memberikan instruksi secara langsung untuk menghasilkan solusi berdasarkan permasalahan yang diberikan. Model yang digunakan dalam penelitian ini adalah Llama 3, yang diakses oleh Ollama framework. Konfigurasi model mencakup pengaturan nilai temperature sebesar 0.7, yang memungkinkan model menghasilkan jawaban yang cukup bervariasi tetapi tetap relevan dengan konteks permasalahan.

Untuk memastikan keluaran yang dihasilkan sesuai dengan kebutuhan penelitian, dibuat sebuah template prompt yang mendefinisikan format *input* dan *output* model. Template ini mencakup pernyataan masalah, permintaan solusi singkat dalam Bahasa Indonesia, serta klasifikasi permasalahan ke dalam empat kategori utama, yaitu Problem-Solution, Kelemahan-Peningkatan, Peluang-Jawaban, Tantangan-Jawaban. Selain itu, hasil dari model diproses lebih lanjut agar hanya mengambil bagian yang berisi solusi tanpa informasi tambahan yang tidak diperlukan. Dengan konfigurasi ini, model diharapkan mampu memberikan solusi logis, relevan, dan sesuai dengan bidang Teknologi Informatika.

3.1.4 Zero-shot Learning Model

Zero-shot Learning memungkinkan model untuk memahami dan mengklasifikasikan teks tanpa perlu pelatihan khusus pada dataset tertentu. Menurut penelitian yang dilakukan oleh (Sivarajkumar dan Wang, 2022), metode ini dapat diterapkan dalam berbagai bidang, termasuk pemrosesan bahasa alami (*Natural Language Processing*), untuk memahami teks yang belum pernah ditemukan sebelumnya. Dalam penelitian ini, ZSL diterapkan menggunakan LLM untuk prediksi solusi dari permasalahan dalam artikel ilmiah. Model yang digunakan adalah Llama 3, yang telah dikonfigurasi agar dapat memberikan solusi yang singkat dan relevan dalam Teknik Informatika.

Sistem model ini dikembangkan melalui beberapa tahapan utama. Pertama, sistem menerima *input* berupa kalimat permasalahan dari artikel ilmiah, kemudian diformat menggunakan template *prompt* untuk memastikan model dapat memahami konteks masalah dan menghasilkan solusi yang sesuai. Setelah itu, model Llama 3 dipanggil melalui *framework* Ollama, dengan parameter *temperature* yang telah dikonfigurasi agar hasil prediksi tetap relevan. Model akan memproses *input* dan mengembalikan solusi dalam bentuk teks singkat, yang kemudian diproses lebih lanjut untuk memastikan hanya bagian solusi yang diambil.

Selain itu juga mengelompokkan permasalahan ke dalam empat kategori utama, yaitu Problem-Solution, Kelemahan-Peningkatan, Peluang-Jawaban, Tantangan-Jawaban. *Output* yang dikeluarkan sistem terdiri dari dua komponen utama, yaitu solusi yang diprediksi dan kategori permasalahan yang diidentifikasi oleh model. Untuk mengevaluasi performa model, hanya prediksi dibandingkan dengan dataset evaluasi yang telah dikumpulkan sebelumnya, menggunakan metrik ROUGE guna mengukur kesesuaian solusi yang dihasilkan dengan solusi referensi.

Dengan pendekatan *Zero-shot Learning*, model mampu menghasilkan solusi tanpa perlu pelatihan ulang pada dataset tertentu, sehingga meningkatkan efisiensi dalam proses analisis teks ilmiah. Sistem ini secara otomatis memberikan solusi atas permasalahan yang terdapat dalam artikel ilmiah.

3.1.5 Evaluasi Model

Setelah model diakses menggunakan pendekatan *zero-shot learning* melalui Ollama, dilakukan validasi menggunakan metrik ROUGE untuk mengukur kualitas prediksi kalimat solusi yang dihasilkan. Validasi ini bertujuan untuk mengevaluasi seberapa relevan prediksi model dengan kalimat solusi yang diharapkan berdasarkan dataset yang telah divalidasi sebelumnya.

ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) adalah model evaluasi yang terdiri dari sejumlah matriks yang digunakan untuk evaluasi *automatic text summarization* dan *machine translation* dalam NLP. ROUGE digunakan untuk melakukan evaluasi berbasis sistem untuk mengukur tingkat akurasi hasil pengujian. Cara kerja ROUGE menggunakan metode intrinsik, yaitu dengan membandingkan hasil prediksi solusi yang dihasilkan oleh mesin dengan dataset. Dataset ini disebut juga dengan *ground truth summarization* (Ayu Irhani, 2023).

ROUGE menyediakan beberapa jenis metrik yang digunakan sebagai standar untuk menilai keakuratan prediksi suatu sistem. Berikut ini adalah metrik utama yang sering digunakan:

1. ROUGE-1 (unigram): mengukur kesamaan kata tunggal (n-gram) antara prediksi dan referensi. Ini dapat mengukur apakah kata-kata penting dalam solusi yang diprediksi cocok dengan yang ada pada solusi yang diharapkan.

$$precision = \frac{jumlah\ unigram\ kata\ sama}{jumlah\ total\ unigram\ dalam\ hasil} \quad (1)$$

$$recall = \frac{jumlah\ unigram\ kata\ sama}{jumlah\ total\ unigram\ dalam\ referensi} \quad (2)$$

$$f1 - score = 2 \times \frac{precision\ (ROUGE-1) \times recall\ (ROUGE-1)}{precision\ (ROUGE-1) + recall\ (ROUGE-1)} \quad (3)$$

2. ROUGE-2: Mengukur kesamaan pasangan kata (bigram) antara prediksi dan referensi. ROUGE-2 berguna untuk menilai seberapa baik pasangan kata dalam kalimat prediksi cocok dengan pasangan kata yang ada dalam solusi yang

diharapkan. Ini mengukur kekonsistenan dan kelengkapan konteks dalam solusi yang diprediksi.

$$precision = \frac{\text{jumlah bigram kata sama}}{\text{jumlah total unigram dalam hasil}} \quad (4)$$

$$recall = \frac{\text{jumlah bigram kata sama}}{\text{jumlah total unigram dalam referensi}} \quad (5)$$

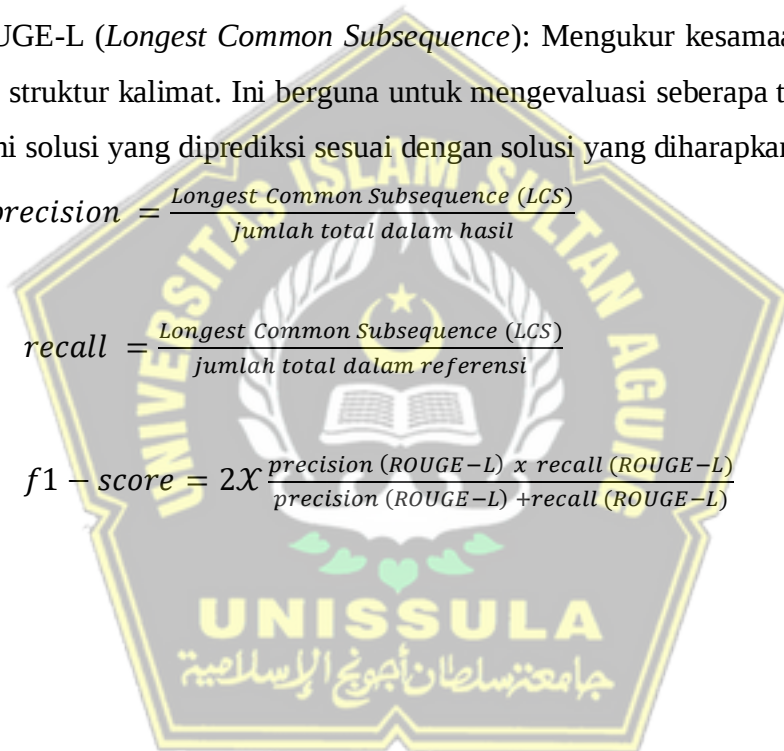
$$f1 - score = 2 \times \frac{precision (ROUGE-2) \times recall (ROUGE-2)}{precision (ROUGE-2) + recall (ROUGE-2)} \quad (6)$$

3. ROUGE-L (*Longest Common Subsequence*): Mengukur kesamaan urutan kata atau struktur kalimat. Ini berguna untuk mengevaluasi seberapa terstruktur dan alami solusi yang diprediksi sesuai dengan solusi yang diharapkan.

$$precision = \frac{\text{Longest Common Subsequence (LCS)}}{\text{jumlah total dalam hasil}} \quad (7)$$

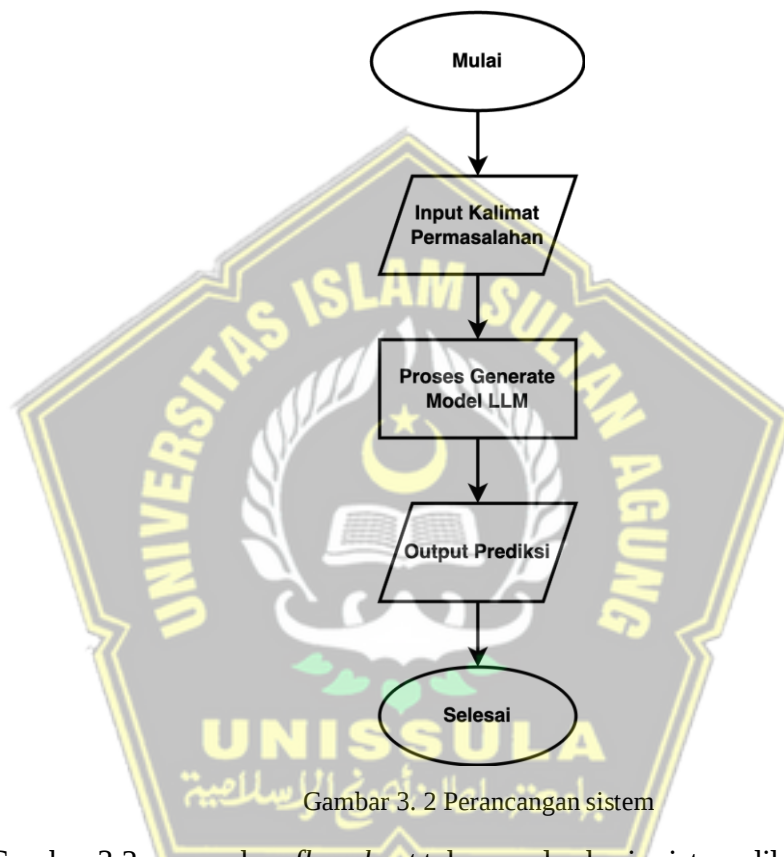
$$recall = \frac{\text{Longest Common Subsequence (LCS)}}{\text{jumlah total dalam referensi}} \quad (8)$$

$$f1 - score = 2 \times \frac{precision (ROUGE-L) \times recall (ROUGE-L)}{precision (ROUGE-L) + recall (ROUGE-L)} \quad (9)$$



3.2 Analisis Sistem

Dalam penelitian ini, penulis akan mengembangkan sistem prediksi solusi berbasis *web* yang dirancang untuk menawarkan sistem prediksi solusi bagi peneliti, akademisi dan semua orang yang membutuhkan. Merancang alur sistem, diperlukan *flowchart* yang menunjukkan Langkah-langkah bagaimana sistem berjalan. Pada gambar 3.1 merupakan *flowchart* rancangan alur sistem yang akan dibangun :



Gambar 3.2 merupakan *flowchart* tahapan alur kerja sistem dilakukan dengan tahapan dibawah:

1. Mulai

Pengguna terlebih dahulu mengakses halaman utama antar muka untuk menampilkan proses kerja dari aplikasi sistem prediksi solusi.

2. Input Kalimat Permasalahan

Kemudian pengguna memasukkan kalimat permasalahan yang terdapat dalam artikel ilmiah, setelah itu klik tombol “Submit” yang terdapat di tengah untuk memasukkan kalimat *input* teks.

3. Permasalahan Proses

Setelah pengguna memasukkan kalimat permasalahan, sistem akan memproses *input* tersebut dengan menggunakan model Llama 3. Kalimat permasalahan yang dimasukkan akan diformat menjadi *prompt* yang sesuai dengan model dapat memahami konteksnya dengan baik. Kemudian menganalisis *prompt* tersebut dan menghasilkan prediksi solusi berdasarkan pengetahuan yang dimiliki.

4. Output Prediksi

Setelah diproses prediksi selesai, sistem akan menampilkan hasil berupa solusi yang dihasilkan oleh model Llama 3. Hasil ini ditampilkan pada antarmuka pengguna dalam bentuk teks yang mudah dipahami.

5. Selesai

Setelah selesai ditampilkan, proses selesai. Pengguna dapat kembali ke tahap awal untuk memasukkan kalimat permasalahan lain.

3.3 Identifikasi Perangkat Lunak

Pada tahap analisis kebutuhan, peneliti menganalisis apa saja perangkat lunak yang dibutuhkan untuk mengembangkan aplikasi ini sehingga proses *input* hingga mengeluarkan hasil *output* sesuai dengan yang diinginkan. Adapun perangkat lunak yang digunakan untuk mengembangkan sistem ini adalah:

1. Python 3.12.0

Python merupakan salah satu bahasa pemrograman komputer yang dapat digunakan untuk membuat situs web, *software* atau aplikasi dan analisis data. Bahasa pemrograman yang digunakan dalam penelitian ini adalah Bahasa python dengan versi 3.12.0 digunakan dikarenakan Bahasa ini memiliki library yang lengkap dan *open source*. Selain itu, python dikenal dengan sintaks yang sederhana serta mudah dipahami, sehingga mempermudah proses pengembangan. Dengan dukungan komunitas yang luas dan berbagai pustaka bawaan maupun eksternal, python menjadi pilihan yang fleksibel dan efisien untuk berbagai jenis pengembangan perangkat lunak.

2. Ollama

Ollama dipilih sebagai platform untuk menjalankan model LLaMa 3 secara lokal. Ollama memiliki kemampuannya dalam mengelola model AI dengan efisien, termasuk dalam hal manajemen sumber daya dan kemudahan integrasi dengan berbagai aplikasi. Llama 3 merupakan model dengan kemampuan pemrosesan Bahasa alami yang dapat menghasilkan teks relevan sesuai dengan konteksnya. Model ini mampu memahami dan merespons berbagai jenis pertanyaan dengan tingkat akurasi yang tinggi, sehingga dapat digunakan dalam berbagai aplikasi berbasis kecerdasan buatan. Dengan dukungan dari Ollama, Llama 3 dapat dijalankan secara lokal tanpa memerlukan koneksi ke server eksternal, sehingga meningkatkan efisiensi dan keamanan dalam penggunaannya.

3. LangChain

LangChain digunakan untuk mengelola interaksi dengan model Bahasa besar. Dalam penelitian ini, LangChain membantu dalam pembuatan PromptTemplate, yang digunakan untuk membangun format input yang sesuai sebelum dikirim ke model. Selain itu, LangChain memfasilitasi eksekusi model melalui metode `invoke()` untuk menghasilkan solusi berdasarkan input masalah yang diberikan.

4. Visual Studio Code

Visual Studio Code dipilih sebagai text editor pada pembagian aplikasi dalam penelitian ini, *Visual Studio Code* dipilih karena mendukung banyak bahasa pemrograman dan framework, mulai platform, performa yang sangat cepat, mempunyai banyak *extensions* yang dapat mempercepat dan mempermudah proses pengembangan *website*. Selain itu, editor ini juga memiliki fitur sebagai IntelliSense untuk melengkapi kode secara otomatis, debugging bawaan, serta integrasi dengan sistem control versi seperti Git, yang sangat membantu dalam pengelolaan proyek. Dengan tampilan yang ringan dan fleksibel, *Visual Studio Code* menjadi pilihan yang ideal untuk pengembangan perangkat lunak dalam penelitian ini.

5. Pandas

Pandas digunakan untuk mengolah dan menganalisis data dalam bentuk tabel seperti DataFrame, sehingga memudahkan pembacaan, pembersihan, dan pengolahan data, termasuk file CSV. Dalam evaluasi model pandas digunakan untuk membaca dataset, memastikan kelengkapan kolom, serta mengolah data sebelum dianalisis menggunakan metode ROUGE.

6. ROUGE

ROUGE digunakan untuk mengevaluasi kualitas hasil prediksi model dengan membandingkan output yang dihasilkan model dengan referensi dari dataset evaluasi. Dalam penelitian ini, metrik ROUGE digunakan untuk mengukur tingkat kesesuaian antara solusi yang diprediksi oleh model dengan solusi yang terdapat dalam dataset evaluasi. Rouge dapat memberikan gambaran sejauh mana model mampu menghasilkan keluaran yang mendekati referensi, baik dari segi kemiripan kata, frasa, maupun struktur kalimat. Dengan demikian, metrik ini alat penting dalam menilai efektivitas model dalam menghasilkan solusi yang akurat dan relevan sejauh mana solusi prediksi model dengan solusi dataset evaluasinya.

7. Framework Flask

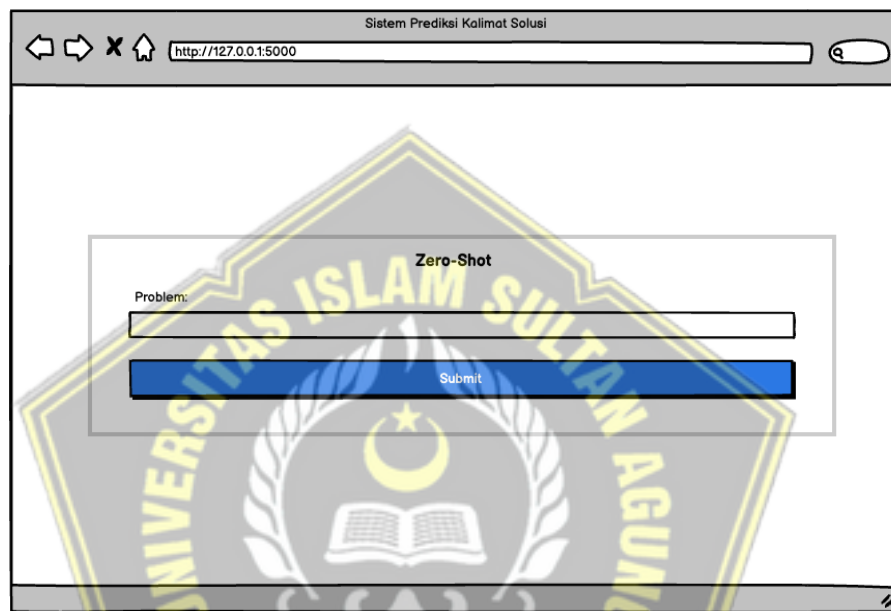
Framework Flask dipilih sebagai framework dalam pengembangan website dikarenakan dengan menggunakan flask penulis bisa membuat *website* yang menggunakan bahasa *python* sehingga akan lebih mudah dan praktis karena core dari sistem yang akan dikembangkan sendiri berbahasa python. Selain itu, Flask bersifat ringan dan fleksibel, memungkinkan pengembangan untuk mengelola rute HTTP dengan sederhana serta memungkinkan input dari pengguna melalui format web.

3.4 Perancangan *User Interface*

Berikut ini merupakan rancangan desain dari sistem yang akan digunakan pada penelitian ini:

1. Halaman Utama

Halaman utama ini merupakan desain antar muka untuk halaman utama dimana halaman ini akan dilihat pertama kali oleh pengguna

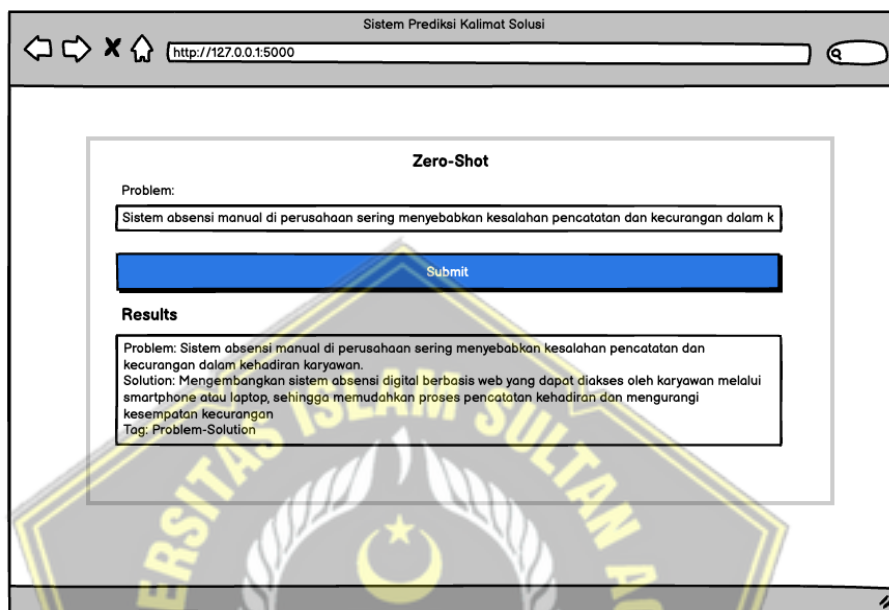


Gambar 3. 3 Halaman Utama

Gambar 3. 3 merupakan halaman tampilan utama, pengguna pertama kali akan melihat halaman utama ini terdapat sebuah label "Problem" yang diikuti oleh kotak teks *input*, dimana pengguna dapat memasukkan pertanyaan tentang kalimat masalah dan dibawahnya terdapat sebuah tombol submit yang berfungsi untuk mengirimkan *input* ke sistem agar diproses dan akan mengeluarkan *output* berupa solusi dari masalah yang diinput.

2. Halaman Prediksi

Halaman prediksi merupakan halaman yang menampilkan hasil dari *input* masalah dari pengguna dan jawaban yang diberikan dari sistem prediksi tersebut, seperti terlihat pada Gambar 3. 4



Gambar 3. 4 Halaman Prediksi

Pada Gambar 3. 4 merupakan halaman yang menampilkan isi dari masalah yang diinput antara pengguna dengan sistem prediksi. Dimana pada tahapan ini pengguna dapat menginput kalimat masalah dan akan diproses oleh sistem, maka sistem akan memberikan *responses* dengan cara memberikan jawaban dan menampilkannya di bagian utama pada halaman sistem.

BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Hasil Penelitian

Setelah model berhasil dijalankan, langkah selanjutnya adalah mengimplementasikan aplikasi dalam bentuk *website*. *Website* ini dirancang agar pengguna cukup memasukkan kalimat permasalahan pada kolom Problem, kemudian sistem akan memproses *input* tersebut dan menampilkan hasil prediksi solusi pada halaman hasil. Dengan memanfaatkan *Flask*, aplikasi ini dapat dijalankan secara lokal atau diunggah ke server, sehingga dapat diakses dengan lebih mudah oleh pengguna lain.

4.1.1 Tampilan Halaman Utama



Gambar 4. 1 Halaman Utama

Pada Gambar 4.1 menunjukkan tampilan halaman utama *website*, yang merupakan halaman pertama yang ditampilkan kepada pengguna saat mereka mengakses sistem. Pada halaman ini, pengguna dapat memasukkan kalimat masalah ke dalam kolom yang berlabel “Problem:”. Setelah itu, pengguna dapat menekan tombol “Submit” untuk mengirimkan input tersebut agar diproses lebih lanjut oleh sistem.

4.1.2 Halaman *Input* kalimat permasalahan



Gambar 4. 2 Halaman Input kalimat permasalahan

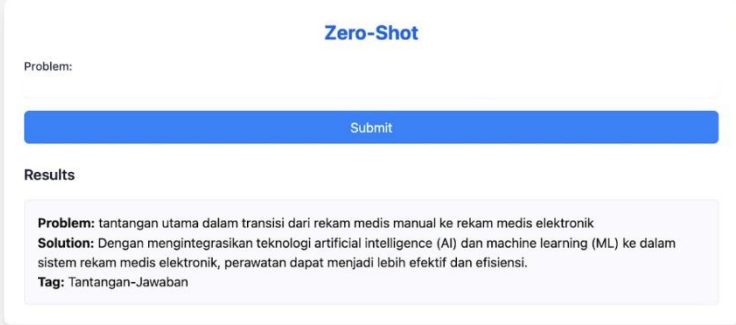
Pada Gambar 4.2 merupakan tampilan halaman saat pengguna memasukkan kalimat masalah pada sistem. Setelah pengguna menginput teks pada kolom yang tersedia, sistem akan memprosesnya ketika tombol “submit” diklik. Halaman ini menunjukkan bagaimana interaksi awal pengguna dengan sistem sebelum mendapatkan hasil atau respons dari proses yang dijalankan.

4.1.3 Halaman Hasil Prediksi



Gambar 4. 3 Contoh kalimat prediksi-solution

Gambar 4.3 merupakan contoh kalimat permasalahan yang diinput dalam jenis kalimat problem-solution. Aplikasi kemudian memproses *input* tersebut dan menghasilkan prediksi solusi yang sesuai dengan permasalahan yang diberikan



Zero-Shot

Problem:

Submit

Results

Problem: tantangan utama dalam transisi dari rekam medis manual ke rekam medis elektronik

Solution: Dengan mengintegrasikan teknologi artificial intelligence (AI) dan machine learning (ML) ke dalam sistem rekam medis elektronik, perawatan dapat menjadi lebih efektif dan efisiensi.

Tag: Tantangan-Jawaban

Gambar 4. 4 Contoh kalimat Tantangan-Jawaban

Gambar 4. 4 merupakan contoh kalimat permasalahan yang diinput dalam jenis kalimat tantangan-jawaban. Aplikasi kemudian memproses *input* tersebut dan menghasilkan prediksi solusi yang sesuai dengan permasalahan yang diberikan.



Zero-Shot

Problem:

Submit

Results

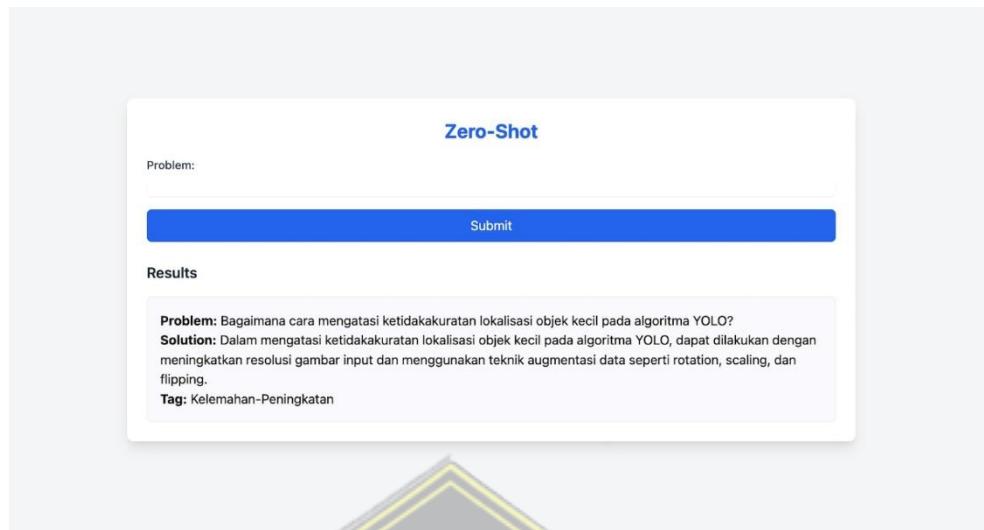
Problem: Bagaimana cara fintech dapat memanfaatkan tren pembayaran digital untuk menarik lebih banyak pengguna?

Solution: Fintech dapat memanfaatkan tren pembayaran digital dengan mengembangkan aplikasi mobile yang mudah digunakan, aman, dan memberikan reward kepada pengguna setelah melakukan transaksi.

Tag: Peluang-Jawaban

Gambar 4. 5 Contoh kalimat Peluang-Jawaban

Gambar 4. 5 merupakan contoh kalimat permasalahan yang diinput dalam jenis kalimat peluang-jawaban. Aplikasi kemudian memproses *input* tersebut dan menghasilkan prediksi solusi yang sesuai dengan permasalahan yang diberikan.



Gambar 4. 6 Contoh Kalimat Kelemahan-Peningkatan

Gambar 4. 6 merupakan contoh kalimat permasalahan yang diinput dalam jenis kalimat kelemahan-peningkatan. Aplikasi kemudian memproses *input* tersebut dan menghasilkan prediksi solusi yang sesuai dengan permasalahan yang diberikan.

4.2 Analisa Penelitian

Penelitian ini mengembangkan sistem berbasis *Large Language Model* (LLM) dengan metode *Zero-Shot Learning* untuk memprediksi kalimat solusi dari permasalahan dalam artikel ilmiah. Sistem ini tidak menggunakan dataset untuk pelatihan, melainkan sepenuhnya mengandalkan *prompt engineering* untuk menghasilkan solusi berdasarkan pemahaman model terhadap konteks yang diberikan.

Dalam prosesnya, pengguna memasukkan kalimat permasalahan dari artikel ilmiah sebagai *input*. Model Llama 3, yang dijalankan menggunakan Ollama, kemudian memproses *input* tersebut melalui mekanisme inferensi berbasis *prompt*. Dengan pendekatan ini, model tidak sekadar menyalin informasi, tetapi mampu membangun solusi yang relevan secara kontekstual.

Evaluasi sistem dilakukan dengan menggunakan metrik ROUGE-1, ROUGE-2, dan ROUGE-L, yang membandingkan hasil prediksi model dengan solusi referensi (*ground truth*). Hasil pengujian menunjukkan bahwa meskipun skor

ROUGE berada dalam rentang 3,5% hingga 22%, model tetap mampu menghasilkan solusi yang masuk akal dan relevan dengan permasalahan yang diberikan. Hal ini menunjukkan bahwa model tidak hanya mencocokkan teks, tetapi juga mampu menyusun solusi baru berdasarkan pemahaman terhadap *prompt* yang diberikan.

Tantangan utama dalam pendekatan ini adalah bagaimana memastikan model menghasilkan solusi yang sesuai dengan ekspektasi tanpa dataset pelatihan. Oleh karena itu, optimasi *prompt* menjadi faktor kunci dalam meningkatkan akurasi dan relevansi solusi yang dihasilkan. Beberapa strategi yang diterapkan meliputi penyusunan struktur *prompt* yang jelas dan eksplisit agar model memahami tugas dengan lebih baik, penggunaan instruksi yang spesifik untuk membimbing model dalam memberikan jawaban yang relevan, serta eksperimen dengan berbagai variasi *prompt* guna menemukan pola terbaik yang menghasilkan prediksi optimal.

Dengan hasil yang diperoleh, penelitian ini menunjukkan bahwa *Zero-Shot Learning* dengan Llama 3 memiliki potensi besar dalam membantu pencarian solusi dalam artikel ilmiah tanpa perlu dataset tambahan untuk pelatihan. Melalui pengaturan *prompt* yang tepat, model dapat lebih optimal dalam memahami permasalahan dan menghasilkan solusi yang lebih sesuai dengan kebutuhan pengguna.

4.3 Hasil *Zero-shot Learning*

Hasil implementasi *Zero-Shot Learning* (ZSL) dalam penelitian ini menunjukkan bahwa model mampu memberikan prediksi solusi berdasarkan permasalahan yang diberikan tanpa memerlukan dataset selama proses prediksi. Model bekerja dengan memanfaatkan *prompt* yang dirancang secara spesifik agar dapat memahami dan menyusun jawaban yang sesuai dengan konteks permasalahan di bidang teknik informatika.

Hasil pengujian menunjukkan bahwa model dapat menghasilkan solusi yang relevan dengan mempertahankan struktur jawaban yang telah ditetapkan dalam *prompt*. Selain itu, model juga mampu mengklasifikasikan jenis permasalahan ke dalam kategori tertentu, seperti problem-solusi, tantangan-jawaban, kelemahan-

peningkatan, dan peluang-jawaban. Meskipun model tidak dilatih menggunakan dataset khusus, hasil yang diperoleh menunjukkan bahwa model tetap dapat menyusun solusi yang sesuai berdasarkan pemahaman kontekstualnya terhadap teks masukan.

Model juga menunjukkan kemampuan dalam menyesuaikan jawaban sesuai dengan pola yang diberikan dalam *prompt*. Dengan pendekatan ini, model tidak hanya menghasilkan solusi secara generik, tetapi juga mampu memberikan jawaban yang lebih terarah sesuai dengan kebutuhan pengguna. Hasil prediksi yang diperoleh tetap berada dalam ruang lingkup teknik informatika dan mengikuti batasan yang telah ditentukan dalam perancangan sistem.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan berbasis ZSL dapat digunakan untuk memprediksi solusi permasalahan dalam artikel ilmiah tanpa melalui proses pelatihan dengan dataset khusus. Evaluasi lebih lanjut diperlukan untuk mengukur efektivitas model dengan membandingkan hasil prediksi terhadap solusi referensi.

4.4 Hasil Evaluasi

Untuk memastikan bahwa model yang digunakan mampu menghasilkan jawaban yang relevan, evaluasi dilakukan menggunakan metrik ROUGE yang mengukur kesamaan solusi model dengan referensi dalam dataset.

Tabel 4.1 menampilkan hasil perbandingan antara jawaban sistem dan referensi, yang diperoleh dari dataset evaluasi. Selain itu, hasil evaluasi menggunakan metrik ROUGE juga ditampilkan dalam Tabel 4.2.

Tabel 4. 1 Hasil Prediksi sistem dan reference

No	Kalimat masalah	Referensi Jawaban	Prediksi Jawaban Sistem
1	Pedagang keliling saat ini mengalami kesulitan mendapatkan pelanggan atau pembeli akibat persaingan dengan penyedia jasa	Mereka dapat memanfaatkan media sosial dan platform online untuk mempromosikan produk mereka serta menawarkan layanan	Pedagang keliling dapat meningkatkan kualitas pelayanan dan menawarkan harga kompetitif untuk mempertahankan pelayanan. Selain itu, mereka juga dapat

	pemesanan makanan secara online	pesan antar yang lebih fleksibel	berkolaborasi dengan penyedia jasa pemesanan makanan online untuk menciptakan sistem pemesanan <i>hybrid</i> yang terintegrasi.
2	Dengan meningkatnya penyebaran <i>hate speech</i> di internet, tantangan besar muncul dalam upaya menciptakan lingkungan internet yang lebih positif dan aman.	mengembangkan sistem deteksi ujaran kebencian secara otomatis berbasis web pada komentar berita daring untuk membantu dan menggantikan moderasi konten yang dilakukan secara manual.	Salah satu cara untuk mengatasi <i>hate speech</i> di internet adalah menggunakan algoritma berbasis <i>machine learning</i> yang dapat mendeteksi dan memblokir konten tidak sesuai.
3	Penerapan sistem otomatis indikator rumah sakit membuka peluang untuk mengurangi kesalahan pencatatan dan mempercepat waktu pengolahan data	Membangun sistem yang mampu memberikan informasi secara <i>real-time</i> , memungkinkan pengambilan keputusan yang lebih cepat dan tepat untuk meningkatkan layanan rumah sakit	Implementasi sistem otomatis indikator rumah sakit yang berbasis <i>artificial intelligence (AI)</i> untuk meningkatkan akurasi pencatatan dan mengoptimalkan waktu pengolahan data.
4	Dalam menghadapi tantangan dari era digital dan konektivitas yang cepat, Pusdatik harus memiliki kesiapan untuk mengatasi gangguan layanan dan operasional yang disebabkan oleh bencana dan ancaman.	Salah satu langkah untuk meminimalkan dampak dari gangguan tersebut dengan melalui penyusunan <i>Disaster Recovery Plan (DRP)</i>	Pusdatik dapat menggunakan teknologi Big Data dan Analytics untuk mengidentifikasi gangguan layanan dan operasional yang disebabkan oleh bencana dan ancaman, serta mengembangkan sistem <i>early warning</i> untuk mendapatkan informasi sebelum terjadi gangguan.

Berdasarkan Tabel 4.1, hasil evaluasi yang dilakukan terhadap 4 sampel menunjukkan bahwa prediksi solusi yang dihasilkan oleh model sudah relevan dengan *input* permasalahan yang diberikan oleh pengguna. Pengujian ini dilakukan dengan menginputkan 4 kalimat permasalahan yang diambil dari 20 dataset evaluasi. Hasil prediksi solusi kemudian dibandingkan dengan referensi solusi yang terdapat dalam dataset evaluasi menggunakan metrik ROUGE, yang mencakup ROUGE-1, ROUGE-2, dan ROUGE-L. Evaluasi ini bertujuan untuk mengukur tingkat kesamaan antara solusi yang dihasilkan oleh model dengan solusi referensi.

Tabel 4. 2 Hasil Evaluasi ROUGE

ROUGE	PRECISION	RECALL	F1-score
ROUGE-1	0.202	0.251	0.220
ROUGE-2	0.033	0.039	0.035
ROUGE-L	0.144	0.182	0.150

Tabel 4. 3 Hasil Evaluasi rata-rata ROUGE

ROUGE-1	ROUGE-2	ROUGE-L
0.220	0.035	0.158

Berdasarkan hasil evaluasi menggunakan metrik ROUGE yang disajikan dalam Tabel 4.3, dapat dilihat bahwa kinerja sistem dalam metrik ROUGE masih belum optimal.

- ROUGE-1 yang diperoleh adalah 0.220 atau setara dengan 22%, menunjukkan bahwa hanya 22% dari kata-kata dalam jawaban sistem yang cocok dengan referensi.
- ROUGE-2 sebesar 0.035 atau setara dengan 3,5%, yang sangat rendah menunjukkan bahwa sistem kurang mampu menangkap kombinasi dua kata (bigram) yang mirip dengan referensi.
- ROUGE-L sebesar 0.158 atau setara dengan 15,8%, menunjukkan bahwa kesamaan struktur kalimat dengan referensi masih terbatas.

Dari hasil evaluasi ini, keakuratan sistem dalam memberikan jawaban berada dalam rentang nilai 3,5% hingga nilai tertinggi 22%. Meskipun skor ROUGE rendah, jawaban yang dihasilkan tetap relevan dengan masalah yang diberikan. Hal

ini disebabkan karena ROUGE hanya mengukur kesamaan kata dan tidak menilai pemahaman semantik.

Keterbatasan ROUGE dalam menilai kesesuaian makna dan konteks membuat evaluasi manual oleh pakar menjadi penting. Sistem mungkin menghasilkan jawaban yang berbeda dalam pemilihan kata, tetapi tetap relevan dengan inti permasalahan. Oleh karena itu, metode evaluasi tambahan, seperti penilaian manual oleh ahli, dapat digunakan untuk memastikan bahwa solusi yang dihasilkan tidak hanya secara struktural mirip, tetapi juga sesuai dengan konteks masalah yang diberikan.



BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil evaluasi, sistem prediksi solusi dalam artikel ilmiah menggunakan *Zero-Shot Learning* dengan model LLM seperti Ollama mampu menghasilkan solusi yang relevan terhadap permasalahan yang diberikan. Meskipun model tidak dilatih menggunakan dataset spesifik, hasil prediksi tetap menunjukkan keterkaitan dengan permasalahan yang diinput pengguna.

Evaluasi menggunakan metrik ROUGE menunjukkan bahwa skor ROUGE-1, ROUGE-2, dan ROUGE-L berada dalam rentang 3,5% hingga 22%. Nilai ini relatif rendah karena ROUGE lebih berfokus pada kesamaan berbasis n-gram, bukan pemahaman semantik. Oleh karena itu, meskipun skor ROUGE rendah, model tetap mampu memberikan solusi yang sesuai dengan konteks permasalahan.

Penelitian ini menunjukkan bahwa pendekatan *Zero-Shot Learning* memiliki potensi dalam menghasilkan solusi tanpa memerlukan pelatihan dengan dataset spesifik. Namun, kualitas prediksi sangat bergantung pada perancangan *prompt* yang digunakan. Oleh karena itu, strategi optimasi *prompt engineering* menjadi faktor penting dalam meningkatkan akurasi dan relevansi solusi yang dihasilkan.

Untuk penelitian selanjutnya, dapat dipertimbangkan pendekatan *Few-Shot Learning* atau *Fine-Tuning* guna meningkatkan akurasi model. Selain itu, metode evaluasi tambahan seperti penilaian berbasis pakar dapat digunakan untuk menilai kualitas solusi secara lebih komprehensif.

5.2 Saran

Untuk meningkatkan kinerja model dalam memprediksi kalimat solusi dari permasalahan pada artikel ilmiah, beberapa tahapan yang bisa dilakukan antara lain:

1. Meningkatkan model dengan menggunakan LLM yang lebih canggih atau melakukan *fine-tuning* pada model Ollama untuk meningkatkan akurasi prediksi.
2. Mengatasi rendahnya skor ROUGE yang disebabkan oleh model yang belum melalui proses *fine-tuning* dengan dataset spesifik.

3. Mengoptimalkan strategi *prompt engineering* agar hasil prediksi lebih sesuai dengan ekspektasi pengguna.
4. Mengeksplorasi pendekatan *Few-shot Learning* atau *Fine-Tuning* guna meningkatkan akurasi model.
5. Dengan menambahkan metode evaluasi tambahan, seperti penilaian manual oleh pakar, untuk menilai kualitas prediksi secara lebih komprehensif.



DAFTAR PUSTAKA

- Alhoshan, W., Ferrari, A., & Zhao, L. (2023). Zero-shot learning for requirements classification: An exploratory study. *Information and Software Technology*, 159. <https://doi.org/10.1016/j.infsof.2023.107202>
- Ayu Irhani, N. (2023). *Representasi Matan Jauharah Tauhid Menggunakan Metode Recurrent Neural Network Dengan Model Arsitektur Long Short-Term Memory Fakultas Sains Dan Teknologi Universitas Islam Negeri Ar-Raniry Banda Aceh 2023 M/ 1444 H Menggunakan Metode Recurrent Neural Ne.*
- Badawi, M., Abushanab, M., Bhat, S., & Maier, A. (2024). *Review of Zero-Shot and Few-Shot AI Algorithms in The Medical Domain.*
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 2020-Decem.
- Chen, S., Li, H., Zhang, Z., Sun, Z., & Cheng, Y. (2024). OMAI: A Specialized Large Language Model for Operational Maintenance in Institute of High Energy Physics OMAI: A Specialized Large Language Model for Operational Maintenance. <https://pos.sissa.it/>
- Chiu, K.-L., Collins, A., & Alexander, R. (2021). *Detecting Hate Speech with GPT-3. March*, 1–29.
- Dermawan, R. D., & Herdianto. (2024). Meningkatkan Kinerja Output ChatGPT Melalui Teknik Prompt Engineering Yang Dapat Dikustomisasi. *Journal Of Social Science Research*, 4(1), 10646–10664.
- Dewi, N. C., & Qoiriah, A. (2021). Implementasi Algoritma Jaro-Winkler Distance dan N-Gram untuk Deteksi dan Prediksi Perbaikan Kesalahan Penulisan Kata Bahasa Indonesia pada Karya Tulis Ilmiah Mahasiswa. *Journal of Informatics and Computer Science (JINACS)*, 2(03), 169–177. <https://doi.org/10.26740/jinacs.v2n03.p169-177>

- Gao, T., Fisch, A., & Chen, D. (2021). Making pre-trained language models better few-shot learners. *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3816–3830. <https://doi.org/10.18653/v1/2021.acl-long.295>
- Harumy, T. H. F., Dewi Sartika Br Ginting, & Fuzy Yustika Manik. (2024). *Kecerdasan Buatan (Teori Dan Implementasi)*. 1–141.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9). <https://doi.org/10.1145/3560815>
- Mendes-Da-Silva, W., & Leal, C. C. (2021). Salami Science in the Age of Open Data: Déjà lu and Accountability in Management and Business Research. *Revista de Administração Contemporânea*, 25(1), 1–12. <https://doi.org/10.1590/1982-7849rac2021200194>
- Mishra, R. B., & Jiang, H. (2021). Classification of problem and solution strings in scientific texts: Evaluation of the effectiveness of machine learning classifiers and deep neural networks. *Applied Sciences (Switzerland)*, 11(21). <https://doi.org/10.3390/app11219997>
- Noor Kamala Sari, N., Irawan, J., & Handrianus Pranatawijaya, V. (2024). Implementasi Gemini API untuk Generatif Teks Deskripsi Karya Otomatis dalam Aplikasi Pameran Berbasis Web dengan Metode Waterfall. *Jurnal Teknologi Informasi dan Terapan (J-TIT)*, 11(1), 2580–2291.
- Peña, I. P. A., & Ortega-Castro, J. C. (2024). Implementation and Evaluation of an Anti-Fraud Prototype Based on Generative Artificial Intelligence for the Ecuadorian Financial Sector. *Revista de Gestão Social e Ambiental*, 18(9), e08601. <https://doi.org/10.24857/rgsa.v18n9-162>

- Plaza-Del-Arco, F. M., Nozza, D., & Hovy, D. (2023). Respectful or Toxic? Using Zero-Shot Learning with Language Models to Detect Hate Speech. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 60–68. <https://doi.org/10.18653/v1/2023.woah-1.6>
- Priya, & Gupta, S. (2022). Hate Speech Detection using OpenAI and GPT-3. *International Journal of Emerging Technology and Advanced Engineering*, 12(5), 132–138. https://doi.org/10.46338/ijetae0522_15
- Riordan, B., Bonela, A. A., He, Z., Nibali, A., Anderson-Luxford, D., & Kuntsche, E. (2024). How to apply zero-shot learning to text data in substance use research: An overview and tutorial with media data. *Addiction*, 119(5), 951–959. <https://doi.org/10.1111/add.16427>
- Rustipa, K., Purwanto, S., & Rozi, F. (2023). Rhetorical Structures, Strategies, and Linguistic Features of Problem Statement to Promote a Teaching Writing Model. *Studies in English Language and Education*, 10(2), 575–597. <https://doi.org/10.24815/siele.v10i2.30855>
- Sasaki, H., Agchbayar, A., Enkhbayasgalan, N., & Yamamoto, S. (2020). Extracting Problem Linkages to Improve Knowledge Exchange between Science and Technology Domains using an Attention-based Language Model. *Engineering, Technology and Applied Science Research*, 10(4), 5903–5913. <https://doi.org/10.48084/etasr.3598>
- Shanableh, T. (2023). Two-Stage Deep Learning Solution for Continuous Arabic Sign Language Recognition Using Word Count Prediction and Motion Images. *IEEE Access*, 11(November), 126823–126833. <https://doi.org/10.1109/ACCESS.2023.3332250>
- Shorten, C., Pierse, C., Smith, T. B., Cardenas, E., Sharma, A., Trengrove, J., Van, B., & Weaviate, L. (2024). *StructuredRAG: JSON Response Formatting with Large Language Models*.
- Sivarajkumar, S., & Wang, Y. (2022). *HealthPrompt: A Zero-shot Learning Paradigm for Clinical Natural Language Processing*.
- Topsakal, O., & Akinci, T. C. (2023). Creating Large Language Model Applications Utilizing LangChain: A Primer on Developing LLM Apps Fast.

International Conference on Applied Engineering and Natural Sciences, 1(1), 1050–1056. <https://doi.org/10.59287/icaens.1127>

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*.

Zhou, M., Duan, N., Liu, S., & Shum, H. Y. (2020). Progress in Neural NLP: Modeling, Learning, and Reasoning. *Engineering*, 6(3), 275–290. <https://doi.org/10.1016/j.eng.2019.12.014>

Zulfikar, M. F., Budiyanto, U., Informasi, F. T., Informatika, T., Luhur, U. B., Geografis, S. I., Naive, M., & Levenshtein, A. J. (2025). *Sistem Monitoring Keadaan Darurat Berdasarkan Caption Instagram Menggunakan Naïve Bayes Dengan Jarak Levenshtein*. 8, 1–12.

