

**PERANCANGAN *AUTOMATIC VISUALIZATION* DARI DATA TABEL
MENGUNAKAN *LARGE LANGUAGE MODELS* (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:

RAEHAN SUPRIANTONO PUTRI

NIM 32602100108

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

2025

***AUTOMATIC VISUALIZATION DESIGN FROM TABLE DATA 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:

RAEHAN SUPRIANTONO PUTRI

32602100108

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

2025

LEMBAR PENGESAHAN
TUGAS AKHIR

PERANCANGAN *AUTOMATIC VISUALIZATION* DARI DATA TABEL
MENGUNAKAN *LARGE LANGUAGE MODELS (LLM)*

RAEHAN SUPRIANTON PUTRI
NIM 32602100108

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 26 Februari 2025

TIM PENGUJI UJIAN SARJANA :

Mustafa, ST, MM, M.Kom

NIDN. 0623117703

(Ketua Penguji)



11-03-2025

Imam Much I.S, ST, MSc, PhD

NIDN. 0613037301

(Anggota Penguji)



12-03-2025

Sam Farisa C.H., ST., M.Kom

NIDN. 0628028602

(Pembimbing)



12-03-2025

Semarang, 13 Maret 2025

Mengetahui,

Kaprodik Teknik Informatika
Universitas Islam Sultan Agung



Mach Taufik, S.T., M.IT

NIDN. 0622037502

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Raehan Supriantono Putri

NIM : 32602100108

Judul Tugas Akhir : PERANCANGAN *AUTOMATIC VISUALIZATION* DARI
DATA TABEL MENGGUNAKAN *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 siapapun 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, 27 Februari 2025

Yang Menyatakan



Raehan Supriantono Putri

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Rachan Supriantono Putri

NIM : 32602100108

Program Studi : Teknik Informatika

Fakultas : Teknologi Industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul PERANCANGAN *AUTOMATIC VISUALIZATION* DARI DATA TABEL MENGGUNAKAN *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/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 27 Februari 2025

Yang menyatakan,



Rachan Supriantono Putri

KATA PENGANTAR

Dengan mengucapkan syukur Alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “Perancangan *Automatic Visualisation* Dari Data Tabel Menggunakan *Large Language Models* (LLM)” ini untuk memenuhi 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 Semarang.

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

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu di kampus ini
2. Dekan Fakultas Teknologi Industri Ibu Dr. Novi Marlyana, ST., M.T
3. Dosen pembimbing penulis Bapak Sam Farisa Chaerul Haviana, ST, M.Kom yang telah meluangkan waktu dan memberi ilmu
4. Orang tua penulis yang telah mengizinkan untuk menyelesaikan laporan ini
5. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan rendah hati, penulis menyadari bahwa laporan masih memiliki banyak kekurangan dalam hal kualitas, kuantitas, dan ilmu pengetahuan. Oleh karena itu, penulis mengharapkan kritikan dan saran yang membangun untuk membantu laporan ini menjadi lebih baik di masa depan.

Semarang, 27 Januari 2025

Raehan Supriantono Putri

DAFTAR ISI

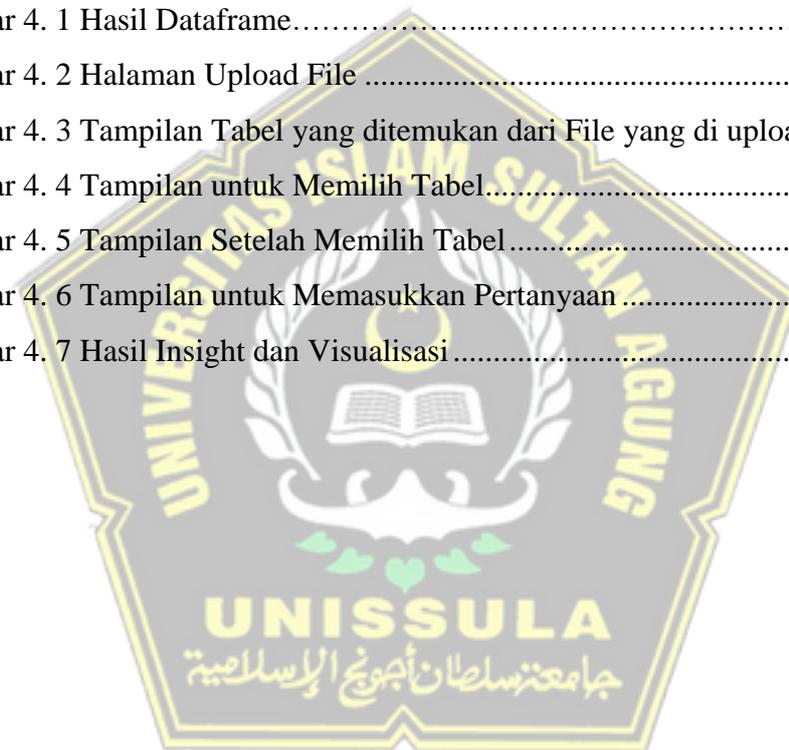
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
ABSTRAK	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	3
1.3 Pembatasan Masalah	3
1.4 Tujuan.....	3
1.5 Manfaat.....	3
1.6 Sistematika penulisan.....	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI	5
2.1 Tinjauan Pustaka	5
2.2 Dasar Teori.....	9
2.2.1 <i>Data Insight</i>	9
2.2.2 <i>Natural Language Precessing (NLP)</i>	10
2.2.3 <i>Large Language Models (LLM)</i>	11
2.2.4 <i>Table Parsing (TAPAS)</i>	12
BAB III METODE PENELITIAN	13
3.1 Tahapan Penelitian	13
3.2 Studi Literatur.....	13
3.3 Pengumpulan Data	13
3.4 Rancangan Sistem	14

3.4.1	Analisis Kebutuhan Sistem	14
3.4.2	<i>Flowchart</i> Sistem	15
3.5	Pembuatan Sistem	16
3.6	Evaluasi	16
BAB IV	HASIL DAN ANALISIS PENELITIAN	19
4.1	Hasil Analisa	19
4.1.1	Persiapan Data.....	19
4.1.2	Ekstraksi Tabel.....	19
4.1.3	Pembuatan <i>Dataframe</i>	20
4.1.4	Memproses <i>Dataframe</i> untuk LLM	21
4.2	Hasil Perancangan <i>User Interface</i>	22
4.3	Pembahasan.....	26
BAB V	KESIMPULAN DAN SARAN.....	27
5.1	Kesimpulan.....	27
5.2	Saran.....	27
DAFTAR PUSTAKA		



DAFTAR GAMBAR

Gambar 2. 1 Contoh Data Insight (Hernikawati, 2021).....	10
Gambar 2. 2 Arsitektur Table Parsing (TAPAS) (Herzig dkk., 2020).....	12
Gambar 3. 1 Flowchart Tahap Penelitian.....	13
Gambar 3. 2 Contoh Data Tabel	14
Gambar 3. 3 Flowchart Alur Sistem	15
Gambar 3. 4 Diagram Alir Kerja Sistem.....	16
Gambar 4. 1 Hasil Dataframe.....	21
Gambar 4. 2 Halaman Upload File	22
Gambar 4. 3 Tampilan Tabel yang ditemukan dari File yang di upload.....	23
Gambar 4. 4 Tampilan untuk Memilih Tabel.....	24
Gambar 4. 5 Tampilan Setelah Memilih Tabel.....	24
Gambar 4. 6 Tampilan untuk Memasukkan Pertanyaan.....	25
Gambar 4. 7 Hasil Insight dan Visualisasi.....	25



DAFTAR TABEL

Tabel 2. 1 Tinjauan Pustaka	5
Tabel 3. 1 Range Nilai MAPE.....	17
Tabel 3. 2 Akurasi Sistem	17



ABSTRAK

Dalam era digital saat ini, data memiliki peran penting sebagai sumber informasi untuk mendukung pengambilan keputusan. Dalam menginterpretasikan suatu data tabel seringkali membutuhkan waktu yang lama, terutama jika data yang diolah memiliki kompleksitas yang tinggi dan berjumlah besar. Oleh karena itu diperlukan sistem yang dapat secara otomatis menginterpretasikan dan memvisualisasikan data tabel. Penelitian ini bertujuan untuk merancang Sistem *Automatic Visualization* dari data tabel menggunakan model berbasis *Large Language Models* (LLM) berupa *Table Parsing* (TAPAS). Sistem ini dirancang untuk mendukung visualisasi data secara efisien dengan mengunggah berbagai format file seperti CSV, *Excel*, maupun PDF. Proses sistem melibatkan parsing data tabel, interpretasi menggunakan LLM, dan visualisasi hasil berupa jawaban dan grafik. Berdasarkan hasil perhitungan Mean Absolute Percentage Error (MAPE), sistem menghasilkan rata-rata kesalahan sebesar 17.66%. Nilai ini menunjukkan bahwa sistem memiliki tingkat akurasi yang cukup baik, namun masih terdapat beberapa jawaban yang memiliki tingkat kesalahan tinggi, seperti yang ditunjukkan oleh nilai MAPE terbesar (121.48%). Hasil penelitian ini menunjukkan bahwa sistem mampu menghasilkan *insight* dan visualisasi yang sesuai dengan data yang diberikan, sehingga dapat membantu pengguna dalam memahami informasi secara lebih efisien. Sistem ini diimplementasikan dalam bentuk *website* yang mudah digunakan. Sistem ini diharapkan dapat memberikan kontribusi dalam meningkatkan efektifitas dalam menginterpretasikan data tabel.

Kata Kunci: *Large Language Models* (LLM), *Table Parsing* (TAPAS), *Automatic Visualization*, pemrosesan data tabel, visualisasi data.

ABSTRACT

In today's digital era, data plays an important role as a source of information to support decision making. Interpreting table data often takes a long time, especially if the data being processed is highly complex and large in number. Therefore, a system is needed that can automatically interpret and visualize table data. This study aims to design an Automatic Visualization System from table data using a Large Language Models (LLM) based model in the form of Table Parsing (TAPAS). This system is designed to support efficient data visualization by uploading various file formats such as CSV, Excel, and PDF. The system process involves parsing table data, interpreting using LLM, and visualizing the results in the form of answers and graphs. Based on the results of the Mean Absolute Percentage Error (MAPE) calculation, the system produces an average error of 17.66%. This value indicates that the system has a fairly good level of accuracy, but there are still some answers that have a high error rate, as indicated by the largest MAPE value (121.48%). The results of this study indicate that the system is able to produce insights and visualizations that are in accordance with the data provided, so that it can help users understand information more efficiently. This system is implemented in the form of an easy-to-use website. This system is expected to contribute to increasing the effectiveness of interpreting table data.

Keywords: Large Language Models (LLM), Table Parsing (TAPAS), Automatic Visualization, tabular data processing, data visualization.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital saat ini, data memiliki peran penting sebagai sumber informasi untuk mendukung pengambilan keputusan. Data merupakan kumpulan informasi yang dapat berupa teks, angka, ataupun simbol. Sedangkan data tabel merupakan kumpulan data yang disusun dalam format baris dan kolom yang sering digunakan untuk menyimpan informasi yang terstruktur, biasanya disimpan dalam format basis data maupun *spreadsheet*. Jumlah data setiap harinya selalu mengalami peningkatan, hal tersebut berpengaruh dalam menginterpretasikan data secara efektif. Data tabel sebenarnya dapat diberi sebuah pertanyaan yang dapat menghasilkan jawaban sesuai dengan isi pada tabel tersebut. Selain itu, interpretasi dari data tabel tidak selalu berupa angka, tetapi dapat berupa grafik maupun deksripsi.

Dalam menginterpretasikan suatu data tabel seringkali membutuhkan waktu yang lama, terutama jika data yang diolah memiliki kompleksitas yang tinggi dan berjumlah besar. Proses ini mencakup berbagai tahapan, mulai dari memahami struktur data hingga penarikan kesimpulan. Selain keterbatasan alat atau metode yang digunakan dapat menambah tantangan, sehingga memerlukan upaya dan waktu tambahan untuk memastikan hasil interpretasi akurat dan relevan. Oleh karena itu, diperlukan pendekatan yang lebih efisien, seperti pemanfaatan teknologi atau algoritma cerdas untuk mempercepat proses interpretasi.

Kemajuan teknologi kecerdasan buatan (*Artificial Intelligence*) dan pembelajaran mesin (*Machine Learning*) dapat dimanfaatkan untuk mengatasi tantangan ini. Salah satu metode yang dapat digunakan yaitu metode *Large Language Models* (LLM). *Large Language Models* (LLM) adalah jenis model AI yang menggunakan pendekatan pembelajaran mesin untuk menghasilkan teks yang serupa dengan bahasa manusia (Lubis dkk., 2024). *Large Language Models* (LLM) dilatih dengan jumlah data yang sangat besar sehingga memiliki kemampuan untuk

memahami konteks, pola, dan hubungan dalam data. *Large Language Models* (LLM) dapat digunakan untuk menghasilkan suatu keputusan tanpa memerlukan analisis manual (Tom dkk., 2020)

Selain itu, *Large Language Models* (LLM) juga dapat digunakan untuk menghasilkan keputusan atau prediksi berdasarkan data yang dianalisis. Hal ini dapat menghemat waktu dan juga mengurangi kemungkinan terjadinya kesalahan dalam analisis. Dengan kemampuannya untuk terus belajar dari data yang ada, *Large Language Models* (LLM) dapat terus meningkatkan kinerja dan akurasi seiring dengan bertambahnya jumlah data yang diproses. Namun, pada penelitian sebelumnya kurangnya perhatian terhadap integrasi antara *Large Language Models* (LLM) dan alat analisis data yang ada. Sebagian besar studi lebih fokus pada pengembangan model tanpa mempertimbangkan bagaimana model tersebut dapat diimplementasikan dalam praktik sehari-hari (Bommasani dkk., 2021).

Salah satu metode yang dapat digunakan untuk memahami dan memanipulasi tabel adalah metode *Table Parsing* (TAPAS). Metode *Table Parsing* (TAPAS) merupakan model *Large Language Models* (LLM) yang dapat menjawab pertanyaan bahasa alami melalui tabel biasanya dilihat sebagai tugas penguraian semantik (Herzig dkk., 2020). *Table Parsing* (TAPAS) memungkinkan pengolahan data tabel secara otomatis dengan algoritma yang dapat mengenali pola hubungan antar data. Dengan menggunakan model *Table Parsing* (TAPAS), tabel tidak lagi hanya sekedar angka dan teks, tetapi menjadi sumber informasi yang dapat memberikan jawaban dari pertanyaan yang relevan.

Dengan latar belakang ini, penelitian ini bertujuan untuk merancang sistem *automatic visualization* dari data tabel menggunakan LLM sehingga dapat menghasilkan interpretasi data tabel secara otomatis, menghasilkan *insight* yang relevan, dan disajikan dalam bentuk yang mudah dipahami sehingga memudahkan pengguna dalam menggali informasi berharga dari data tabel untuk mendukung pengambilan keputusan yang lebih baik. Penelitian ini juga berpotensi menjadi solusi inovatif yang lebih adaptif terhadap kebutuhan analisis data di era modern.

1.2 Perumusan Masalah

Rumusan masalah pada penelitian ini adalah bagaimana membaca secara otomatis data tabel untuk mendapatkan *insight* yang relevan?

1.3 Pembatasan Masalah

Dalam penelitian ini ada beberapa batasan masalah yang diterapkan. Berikut adalah batasan masalah dalam tugas akhir ini:

1. Data yang digunakan pada penelitian ini hanya berupa data tabel
2. Penggunaan *insight* dalam bahasa Indonesia
3. Output dari sistem berupa *insight* dan grafik

1.4 Tujuan

Penelitian ini bertujuan untuk membuat *automatic visualization* berbasis *Large Language Models* (LLM) dari data tabel.

1.5 Manfaat

Manfaat dari penelitian ini adalah:

1. Meningkatkan efisiensi interpretasi data tabel
2. Mempermudah pemahaman data tabel
3. Mendukung pengambilan keputusan data yang lebih akurat

1.6 Sistematika penulisan

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

BAB I : PENDAHULUAN

Pada Bab ini penulis mengutarakan latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Pada Bab ini menguraikan teori-teori yang relevan dan penelitian sebelumnya yang berkaitan dengan penelitian ini.

BAB III : METODE PENELITIAN

Pada Bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk pengumpulan data maupun perancangan sistem yang akan dilakukan.

BAB IV : HASIL DAN ANALISA

Pada bab ini berisi hasil dan pembahasan mengenai sistem automatic insight dari data tabel menggunakan metode LLM

BAB V : KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari seluruh bab dan saran penelitian selanjutnya.



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Tinjauan terhadap hasil penelitian sebelumnya merupakan elemen penting untuk mendukung penelitian yang akan dilakukan. Peninjauan ini menjadi bahan pertimbangan untuk membantu mengidentifikasi kelebihan dan kekurangan dari penelitian sebelumnya. Dengan melakukan tinjauan pustaka dapat memperkuat argument dalam penelitian yang dilakukan, sehingga memberikan wawasan yang lebih mendalam mengenai topik yang telah dikaji. Beberapa penelitian serupa dapat dilihat pada Tabel 2.1.

Tabel 2. 1 Tinjauan Pustaka

No	Judul	Penulis	Metode	Hasil	Kekurangan	Kelebihan
1	Large Language Models Meet NL2Code: A Survey	(Zandk., 2023)	Large Language Models (LLM)	Foundation models mampu menyelesaikan beberapa tugas data wrangling tanpa pelatihan tambahan	Model dapat menghadapi kendala dalam menangani dataset yang sangat besar atau kompleks	Mengurangi waktu dan usaha dalam membersihkan dan memformat data
2	Large Language Models are Complex	(Zhao dkk., 2023)	Large Language Models (LLM) GPT-3.5	Pendekatan yang digunakan secara signifikan melebihi metode	Penelitian ini mungkin tidak mencakup semua jenis tabel kompleks	Meningkatkan kemampuan pemahaman struktur hierarkis

No	Judul	Penulis	Metode	Hasil	Kekurangan	Kelebihan
	x Table Parsers			yang digunakan sebelumnya Pendekatan ini mencapai state-of-the-art (SOTA) performance.	atau konteks yang berbeda	oleh GPT-3.5
3	Table LLM: Enabling Tabular Data Manipulation by LLMs in Real Office Usage Scenarios	(Zhang dkk., 2024)	<i>Large Language Models (LLM)</i>	Table LLM secara signifikan lebih unggul dibandingkan dengan model LLM lainnya, baik yang bersifat umum maupun yang fokus pada data tabular	Memerlukan sumber daya komputasi yang besar untuk pelatihan dan penerapan.	Dirancang khusus untuk manipulasi data tabular, memungkinkan pemrosesan yang lebih efisien.
4	TABFACT: A LARGE	(Chen dkk., 2020)	TabFact transformer-	Model berbasis <i>transformer</i>	Keterbatasan dalam keragaman	Menyediakan baseline yang jelas

No	Judul	Penulis	Metode	Hasil	Kekurangan	Kelebihan
	-SCALE DATAS ET FOR TABLE BASED FACT VERIFI CATION N		based models	dapat mencapai hasil yang baik dalam verifikasi fakta dengan menggunakan informasi dari tabel	tipe tabel dan konteks	untuk evaluasi model-model baru dalam tugas verifikasi fakta berbasis tabel
5	Large Language Model for Table Processing: A Survey	(Ludk., 2024)	<i>Large Language Models (LLM)</i>	<i>Large Language Models (LLM)</i> dapat dioptimalkan untuk tugas pemrosesan tabel	pembahasan terlalu umum dan tidak mempertimbangkan variasi spesifik dalam konteks tabel yang berbeda	Metode yang digunakan mencakup berbagai jenis tugas pemrosesan tabel.
6	TableGPT2: A Large Multimodal Model with Tabular	(Sudk., 2024)	TableGPT2	Model ini mempertahankan kemampuan umum yang kuat sambil excel dalam tugas-tugas	Meskipun TableGPT2 lebih canggih, masih ada tantangan dalam menangani	Integrasi antara data teks dan tabel meningkatkan fleksibilitas model

No	Judul	Penulis	Metode	Hasil	Kekurangan	Kelebihan
	Data Integrati on			yang berkaitan dengan tabel.	skenario nyata yang sangat kompleks atau "dirty" data.	
7	Large Language Models for Tabular Data: Progresses and Future Directions	(Dong dan Wang, 2024)	<i>Large Language Models (LLM)</i>	LLM dapat dioptimalkan untuk tugas-tugas yang berkaitan dengan tabel, termasuk interpretasi, analisis, dan generasi data.	Metode ini lebih fokus pada data tabel yang terstruktur, sehingga kurang relevan untuk jenis data lain yang lebih kompleks atau tidak terstruktur.	Metode ini dirancang untuk mengoptimalkan kinerja LLM dalam konteks yang spesifik, seperti pengolahan dan analisis data tabular

Penelitian terkait *Large Language Models (LLM)* dalam pemrosesan data tabel telah berkembang pesat. Berbagai studi telah mengeksplorasi kemampuan *Large Language Models (LLM)* dalam menangani data tabular, termasuk dalam data *wrangling*, parsing tabel, verifikasi fakta, dan manipulasi data tabular. Berbagai tantangan masih muncul dalam menangani tabel kompleks terhadap berbagai jenis tabel. Penelitian oleh (Chen dkk., 2020), efektivitas model berbasis transformer dalam verifikasi fakta, namun masih terbatas pada keragaman tipe tabel. Penelitian ini mengambil metode *Table Parsing (TAPAS)* untuk mengatasi

keterbatasan tersebut untuk merancang Sistem *Automatic Visualization* untuk menginterpretasikan dan memvisualisasikan data tabel secara otomatis.

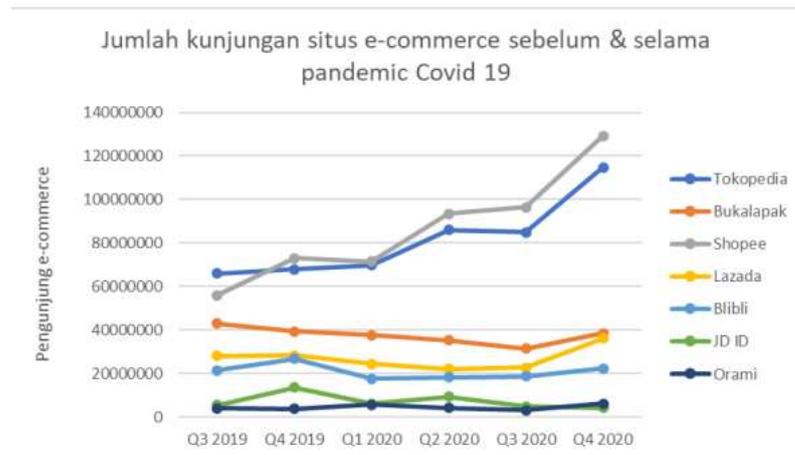
2.2 Dasar Teori

2.2.1 Data *Insight*

Data *insight* merupakan pemahaman mendalam yang didapatkan dari suatu data tertentu. Data *insight* bukan sekedar informasi mentah, melainkan hasil interpretasi yang memberikan makna yang lebih jelas dengan menyajikan gambaran situasi secara menyeluruh dalam memahami pola, tren, dan hubungan yang ada di dalam data, sehingga memudahkan dalam memahami situasi yang sedang dianalisis. Data *insight* dapat digunakan untuk mengambil keputusan yang lebih baik.

Data *insight* memiliki peran yang sangat penting dalam proses pengambilan keputusan, karena memberikan landasan yang kuat untuk membuat keputusan yang tepat. *Insight* yang diperoleh dari data memungkinkan untuk pengambilan keputusan yang berbasis data, yang pada akhirnya dapat meningkatkan efektivitas dalam mencapai tujuan yang diinginkan. Data yang dianalisis dengan baik dapat mengurangi resiko kesalahan dalam proses pengambilan keputusan.

Insight memberikan pemahaman tentang apa yang terjadi dalam suatu situasi. Misalnya, jumlah kunjungan situs e-commerce sebelum dan selama pandemi Covid-19. Maka dapat dilihat dengan menggunakan suatu grafik untuk melihat tren pengunjung selama Covid-19



Gambar 2. 1 Contoh Data *Insight* (Hernikawati, 2021)

Untuk menghasilkan data *insight* yang berkualitas, diperlukan pendekatan analitis yang tepat, salah satunya menggunakan teknologi *Natural Language Processing* (NLP). *Natural Language Processing* (NLP) memungkinkan komputer untuk memahami, menganalisis, dan menafsirkan bahasa manusia dan secara otomatis, sehingga data yang berbentuk teks seperti dokumen yang dapat diubah menjadi informasi yang bermakna. Model *Natural Language Processing* (NLP) dapat mengidentifikasi pola, sentiment, dan topik yang relevan dari data yang kompleks dan tidak terstruktur.

2.2.2 *Natural Language Processing* (NLP)

Natural Language Processing (NLP) salah satu bidang ilmu komputer, kecerdasan buatan dan bahasa (*linguistik*) yang berkaitan dengan interaksi antara komputer dengan bahasa alami manusia (Prasetyo, Benarkah dan Chrisintha, 2021). Fokus utama *Natural Language Processing* (NLP) yaitu membuat sistem komputer yang dapat memahami, memproses, dan menghasilkan bahasa manusia, sehingga manusia dapat berkomunikasi dengan mesin atau komputer layaknya berkomunikasi dengan manusia lain.

Natural Language Processing (NLP) dapat mempermudah interaksi antara manusia dengan mesin. Salah satunya adalah memungkinkan mesin untuk memahami dan merespon perintah atau pertanyaan dalam bahasa manusia baik yang berbentuk lisan maupun tulisan. *Natural Language Processing* (NLP) juga dapat mengubah teks menjadi bentuk lain, seperti mengkonversi teks menjadi suara,

atau menganalisis sentiment dalam teks untuk memahami opini atau emosi. Selain itu, *Natural Language Processing* (NLP) memiliki berbagai tugas utama, seperti tokenisasi, *stemming*, *lemmatization*, *mechine translation*, *text summarization*, dan *text classification*. *Natural Language Processing* (NLP) tidak hanya digunakan untuk memahami perintah, tetapi juga untuk menjawab pertanyaan, menerjemahkan bahasa, dan mengeskrak informasi dari teks secara otomatis.

Salah satu cabang *Natural Language Processing* (NLP) yang dapat digunakan untuk membuat data *insight* yaitu *Large Language Models* (LLM). Salah satu bidang ilmu komputer, kecerdasan buatan, dan bahasa (*linguistik*) adalah pemrosesan bahasa alami manusia (NLP). Bidang ini membahas bagaimana komputer berinteraksi dengan bahasa alami manusia.

2.2.3 Large Language Models (LLM)

Large Language Models (LLM) adalah jenis model *Artificial Intelligence* (AI) yang menggunakan pendekatan pembelajaran mesin untuk menghasilkan teks yang serupa dengan bahasa manusia (Lubis dkk., 2024). *Large Language Models* (LLM) dilatih menggunakan data dalam jumlah besar agar dapat memahami pola atau struktur bahasa manusia.

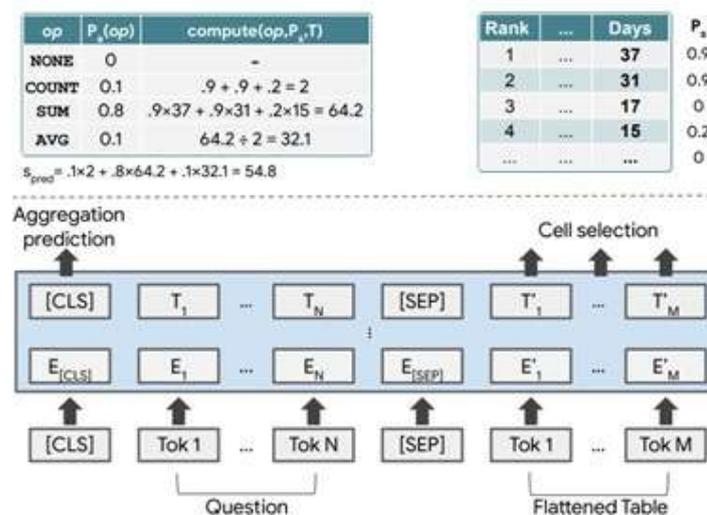
Large Language Models (LLM) memiliki kemampuan untuk memahami konteks dalam teks. Kemampuan ini memungkinkan LLM untuk dapat menangkap makna dan membuat respon dengan mempertimbangkan informasi dalam teks. Sehingga, LLM dapat memberikan jawaban yang sesuai dengan konteks yang sedang dibicarakan. Hal ini menjadikan *Large Language Models* (LLM) berguna dalam berbagai aplikasi baik untuk analisis teks hingga pengambilan keputusan.

Kemampuan *Large Language Models* (LLM) tidak hanya dapat memahami teks saja, tetapi juga dapat digunakan pada data terstruktur seperti tabel. Salah satu penerapannya adalah melalui model *Table Parsing* (TAPAS). Dengan menggunakan *Table Parsing* (TAPAS), *Large Language Models* (LLM) dapat menjawab pertanyaan berbasis tabel, mengidentifikasi pola data, dan mengidentifikasi pola data.

2.2.4 Table Parsing (TAPAS)

Table Parsing (TAPAS) merupakan pemrosesan bahasa alami yang dirancang untuk mengekstrak data dari tabel menggunakan model pembelajaran mesin. *Table Parsing* (TAPAS) memungkinkan mesin untuk memahami konteks tabel dan menjawab pertanyaan yang diajukan dalam bahasa alami dengan mengidentifikasi hubungan antar elemen tabel seperti kolom, baris, dan sel. Tidak hanya membaca tabel sebagai data mentah, *Table Parsing* (TAPAS) juga mampu mengolah tabel menjadi informasi yang terstruktur untuk menjawab pertanyaan yang diajukan dalam bahasa alami.

Table Parsing (TAPAS) dapat memprediksi penunjukan dengan memilih sel tabel dan secara opsional menerapkan operator agregasi yang sesuai dengan pilihan (Herzig dkk., 2020). Model ini memanfaatkan teknik penunjukan sel (*cell selection*) untuk menentukan bagian tabel yang relevan. Proses ini memungkinkan *Table Parsing* (TAPAS) untuk memberikan jawaban yang akurat tanpa memerlukan konversi tabel menjadi format teks biasa.



Gambar 2. 2 Arsitektur *Table Parsing* (TAPAS) (Herzig dkk., 2020)

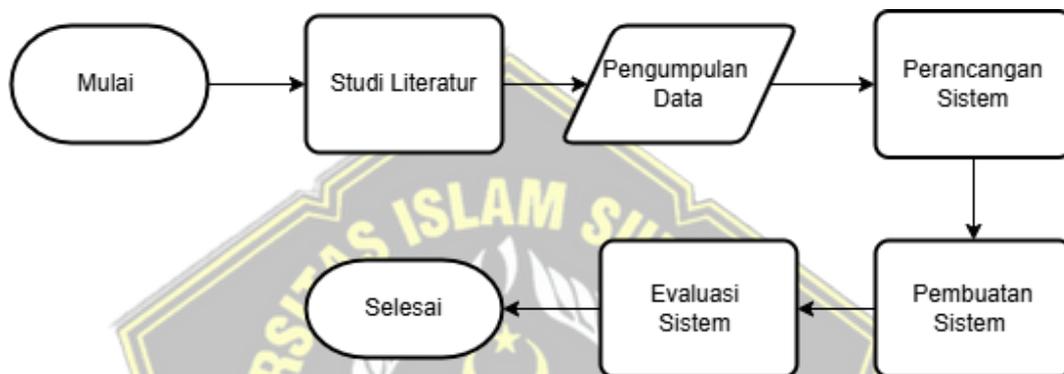
Dengan integrasi *Table Parsing* (TAPAS), pengguna dapat dengan mudah berinteraksi dengan data tabel melalui antarmuka berbasis alami, tanpa perlu pengetahuan mendalam tentang struktur tabel. *Table Parsing* (TAPAS) dapat membantu pengambilan keputusan dan eksplorasi data tabel.

BAB III

METODE PENELITIAN

3.1 Tahapan Penelitian

Dalam penelitian ini langkah-langkah yang akan dilakukan yaitu Studi Literatur, Pengumpulan Data, Perancangan Sistem, Pembuatan Sistem, Testing Sistem, dan Evaluasi. Seperti yang terlihat pada Gambar 3.1.



Gambar 3. 1 *Flowchart* Tahap Penelitian

3.2 Studi Literatur

Dalam penelitian ini akan dilakukan tinjauan terhadap beberapa *e-book*, makalah, jurnal, teks, dan skripsi terdahulu akan diulas selain mengunjungi berbagai situs web. Tujuan dari tinjauan ini adalah untuk mempelajari teori di balik konsep seperti *Table Parsing* (TAPAS) dengan pendekatan Tabular.

Dalam studi literatur ini dilakukan pengumpulan informasi dari berbagai sumber yang relevan dengan topik penelitian yaitu tentang penggunaan *Table Parsing* (TAPAS). Studi ini mencakup identifikasi konsep dasar, metode yang telah dikembangkan, perbandingan dengan penelitian sebelumnya, dan evaluasi kekurangan kelebihan metode yang digunakan dalam penelitian.

3.3 Pengumpulan Data

Tahap selanjutnya dalam penelitian ini adalah melakukan pengumpulan data. Data yang digunakan merupakan data tabel yang didapat dari jurnal ilmiah. Data

yang digunakan merupakan berbagai kumpulan data survei yang diambil dari jurnal bersumber dari Google Scholar. Jumlah data yang dikumpulkan sebanyak 10 dataset. Dataset ini nantinya akan digunakan sebagai bahan pengujian sistem untuk memastikan bahwa sistem yang telah dibuat dapat memberikan hasil yang akurat dan relevan.

Dusun	Laki-Laki	Perempuan	Jumlah
1. Jurgang	390	391	781
2. Muncar	227	225	452
3. Laok Lorong	226	250	476
Jumlah	843	866	1709

Gambar 3. 2 Contoh Data Tabel

Gambar 3.2 merupakan salah satu contoh data tabel yang digunakan untuk menguji sistem *Automatic Visualization* pada penelitian ini.

3.4 Rancangan Sistem

Perancangan implementasi sistem merupakan proses perencanaan langkah-langkah untuk membuat sistem. Tahap ini mencakup berbagai aspek, mulai dari persiapan kebutuhan sistem seperti perangkat keras dan perangkat lunak, hingga pembuatan gambaran alur kerja sistem.

3.4.1 Analisis Kebutuhan Sistem

1. *Software* (perangkat lunak)

Software (perangkat lunak) yang akan digunakan pada tahap perancangan sistem terdiri dari sebagai berikut:

- a. Browser: Chrome
- b. Bahasa Pemrograman: Python
- c. *Tools*: Google Colab

2. *Hardware* (perangkat keras)

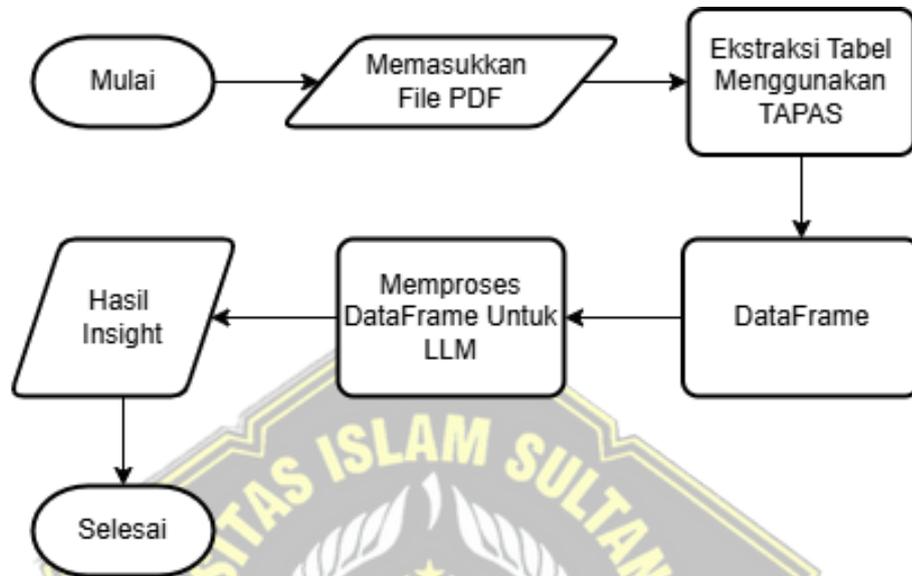
Hardware (perangkat keras) yang akan digunakan pada tahap perancangan sistem terdiri sebagai berikut:

- a. Pcesssor: 11th Gen Intel(R) Core(TM) i5-1135G7
- b. RAM: 8GB

c. Harddisk: 1000GB

3.4.2 Flowchart Sistem

Diagram alir kerja sistem akan seperti Gambar 3.3.



Gambar 3. 3 Flowchart Alur Sistem

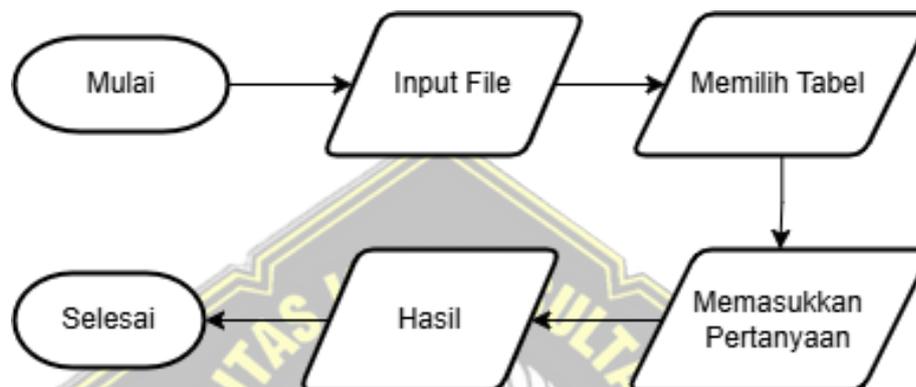
Diagram Alir Sistem pada gambar terdiri dari beberapa langkah, yaitu :

1. Memasukkan data tabel. Langkah pertama adalah memasukkan data tabel ke dalam sistem. Data tabel dapat berformat *excel*, pdf, maupun csv.
2. Ekstraksi Tabel. Langkah selanjutnya yaitu sistem membaca dan mengekstrak elemen-elemen dalam tabel. Ekstraksi ini bertujuan untuk memperoleh informasi dari tabel sehingga data dapat diproses lebih lanjut oleh sistem.
3. Membuat *Dataframe*. Data yang telah diekstraksi dari tabel diubah menjadi bentuk *dataframe*, yaitu struktur data berbentuk tabular yang digunakan dalam pemrograman sehingga data dapat diolah dan dianalisis dengan mudah.
4. Memproses *Datafame* untuk LLM. Data dari *dataframe* diubah menjadi format yang dapat dipahami oleh LLM. Konversi ini bertujuan agar LLM dapat mengenali pola, hubungan, dna konteks dari data tabel.

5. Hasil *Insight*. Setelah data diproses oleh LLM, sistem menghasilkan *output* berupa *insight* berdasarkan data yang telah diunggah. *Output* sistem berupa *insight* dan grafik.

3.5 Pembuatan Sistem

Diagram alir kerja sistem seperti Gambar 3.4.



Gambar 3. 4 Diagram Alir Kerja Sistem

Diagram alir kerja sistem pada gambar terdiri dari beberapa langkah, yaitu :

1. Input file. Tahap pertama yaitu menginputkan *file* yang memiliki format tabel.
2. Memilih tabel. Setelah *file* di inputkan, kemudian memilih tabel yang akan di visualisasikan
3. Memasukkan pertanyaan. Setelah tabel di pilih, masukkan pertanyaan untuk tabel yang ingin di ketahui *insight* dan divisualisasikan
4. Hasil. Setelah memasukkan pertanyaan, akan keluar hasil jawaban berupa *insight* dan visualiasi sesuai dengan pertanyaan yang telah diajukan

3.6 Evaluasi

Pada tahap evaluasi model ini menggunakan *Mean Absolute Percentage Error* (MAPE). *Mean Absolute Percentage Error* (MAPE) adalah metrik yang digunakan untuk mengukur tingkat kesalahan dalam suatu prediksi atau estimasi dibandingkan dengan nilai sebenarnya. MAPE sering digunakan dalam analisis data dan evaluasi model prediksi karena memberikan hasil dalam bentuk persentase, sehingga mudah untuk diinterpretasikan.

Rumus MAPE :

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

Di mana:

y_i = Jawaban benar (Ground Truth)

\hat{y}_i = Jawaban sistem

N = Jumlah pertanyaan

Tabel 3. 1 Range Nilai MAPE

Range MAPE	Arti
< 10%	Kemampuan model sangat baik (akurat)
10% - 20%	Kemampuan model cukup baik
20% - 50%	Kemampuan model kurang akurat
>50%	Kemampuan model tidak akurat

Tabel 3. 2 Akurasi Sistem

No	Pertanyaan	Jawaban Benar (Ground Truth)	Jawaban Sistem	MAPE (%)
1.	Berapa jumlah laki-laki di dusun Jurgang?	390	390	0.00%
2.	Berapa jumlah perempuan di dusun Jurgang?	391	866	121.48 %
3.	Berapa jumlah penduduk di dusun Jurgang?	781	781	0.00%
4.	Berapa jumlah laki-laki di dusun Muncar?	227	227	0.00%
5.	Berapa jumlah perempuan di dusun Muncar?	225	225	0.00%

6.	Berapa jumlah penduduk di dusun Muncar?	452	452	0.00%
7.	Berapa jumlah laki-laki di dusun Laok Lorong?	226	226	0.00%
8.	Berapa jumlah perempuan di dusun Laok Lorong?	250	476	90.4%
9.	Berapa jumlah penduduk di dusun Laok Lorong?	476	476	0.00%
10.	Berapa jumlah laki-laki di seluruh dusun?	843	843	0.00%
11.	Berapa jumlah perempuan di seluruh dusun?	866	866	0.00%
12.	Berapa jumlah penduduk di seluruh dusun?	1709	1709	0.00%
Rata-rata MAPE				17.66%

Berdasarkan Tabel 3.1 sistem *automatic visualization* data tabel menggunakan *Large Language Models* (LLM) dengan metode *Table Parsing* (TAPAS), sistem memiliki kemampuan yang cukup baik dalam menghasilkan jawaban dan visualisasi bersarkan pertanyaan yang diajukan. Berdasarkan hasil perhitungan *Mean Absolute Percentage Error* (MAPE), sistem menghasilkan rata-rata kesalahan sebesar 17.66%. Nilai ini menunjukkan bahwa sistem memiliki tingkat akurasi yang cukup baik, namun masih terdapat beberapa jawaban yang memiliki tingkat kesalahan tinggi, seperti yang ditunjukkan oleh nilai MAPE terbesar (121.48%).

Hasil evaluasi ini digunakan untuk mengukur akurasi sistem dalam memahami pertanyaan dan mengekstraksi informasi yang relevan dari tabel. Semakin rendah skor nilai MAPE semakin baik kinerja sistem dalam menjawab pertanyaan berbasis data tabel secara otomatis.

BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Hasil Analisa

4.1.1 Persiapan Data

Pada proses ini dilakukan proses persiapan data yang akan digunakan dalam sistem. Data yang digunakan berupa *file* dalam format CSV, *Excel*, dan PDF yang berisi tabel. Persiapan data dilakukan dengan memastikan bahwa file yang akan diunggah memiliki format tabel yang jelas agar tabel dapat dikenali oleh sistem. Setelah data di siapkan, *file* tersebut kemudian diunggah ke sistem untuk di uji, mulai dari proses pembacaan tabel, pemilihan tabel, hingga jawaban yang dihasilkan oleh sistem.

4.1.2 Ekstraksi Tabel

Proses ekstraksi tabel bertujuan untuk mengambil data yang berbentuk tabel dari format CSV, *Excel*, dan PDF agar dapat diolah lebih lanjut oleh sistem.

1. Ekstraksi dari CSV

Sistem membaca file CSV menggunakan pustaka *pandas*

```
df = pd.read_csv(uploaded_file, dtype=str)
```

2. Ekstraksi dari *Excel*

Sistem membaca seluruh lembar kerja (*sheet*) yang ada di dalam file. Setiap *sheet* yang berisi tabel di ekstrak dan di konversi menjadi *dataframe*

```
df = pd.read_excel(xls, sheet_name=sheet, dtype=str)
```

3. Ekstraksi dari PDF

Sistem menggunakan pustaka *pdfplumber* untuk mengekstrak tabel dari setiap halaman dalam dokumen. Data yang diperoleh kemudian dikonversi ke dalam format *dataframe*.

```
with pdfplumber.open(uploaded_file) as pdf:
```

```
    for page in pdf.pages:
```

```
        table = page.extract_table()
```

```
        if table and len(table) > 1:
```

```
            headers = table[0]
```

```
headers = [str(col).strip() if col else f"Kolom_{i}" for i,
col in enumerate(headers)]
```

4.1.3 Pembuatan *Dataframe*

Pada tahap ini, sistem membaca isi file dan mengidentifikasi struktur tabel menggunakan pustaka *pandas* untuk *CSV/Excel*, serta *pdfplumber* untuk *PDF*. Setelah data diekstrak, sistem mengkonversinya ke dalam bentuk *Dataframe* menggunakan `pd.DataFrame()`.

1. Pembuatan *Dataframe* dari file *CSV*

```
if uploaded_file.name.endswith(".csv"):
    df = pd.read_csv(uploaded_file, dtype=str)
    df = clean_table(df)
    tables[f"Tabel {table_count}"] = df
    table_count += 1
```

2. Pembuatan *Dataframe* dari file *Excel*

```
elif uploaded_file.name.endswith(".xlsx"):
    xls = pd.ExcelFile(uploaded_file)
    for sheet in xls.sheet_names:
        df = pd.read_excel(xls, sheet_name=sheet, dtype=str)
        df = clean_table(df)
        tables[f"Tabel {table_count}"] = df
        table_count += 1
```

3. Pembuatan *Dataframe* dari file *PDF*

```
elif uploaded_file.name.endswith(".pdf"):
    with pdfplumber.open(uploaded_file) as pdf:
        for page in pdf.pages:
            table = page.extract_table()
            if table and len(table) > 1:
                headers = table[0]
                headers = [str(col).strip() if col else f"Kolom_{i}" for i,
col in enumerate(headers)]
                df = pd.DataFrame(table[1:], columns=headers, dtype=str)
                df = clean_table(df)
                tables[f"Tabel {table_count}"] = df
                table_count += 1
```

Setelah tabel di ekstraksi, kemudian dilakukan pembersihan data agar data dapat diproses oleh *Table Parsing* (TAPAS) dengan maksimal. Proses *cleaning* yang dilakukan yaitu menghapus kolom duplikat, menghapus baris kosong, dan mengkonversi semua data menjadi data *string*.

```
def clean_table(df):
    df = df.loc[:, ~df.columns.duplicated()].dropna(how="all")
    df.columns = [col.strip() for col in df.columns] # Hapus
    spasi di nama kolom
    df = df.astype(str).fillna("") # Konversi semua data ke
    string agar kompatibel dengan TAPAS
    return df
```

Hasil pembuatan dari *Dataframe* bisa dilihat pada Gambar 4.1.

	Dusun	Laki-Laki	Perempuan	Jumlah
0	1. Jurgang	390	391	781
1	2. Muncar	227	225	452
2	3. Laok Lorong	226	250	476
3	Jumlah	843	866	1709

Gambar 4. 1 Hasil *Dataframe*

4.1.4 Memproses *Dataframe* untuk LLM

Setelah tabel diekstraksi dan diubah menjadi *Dataframe*, sistem mengkonversi *Dataframe* tersebut ke dalam format yang dapat diproses oleh model *Table Parsing* (TAPAS). Proses ini melibatkan beberapa tahapan utama:

1. Konversi *Dataframe* ke format *string*

Data dalam *Dataframe* diubah menjadi *string* agar kompatibel dengan model TAPAS.

2. Tokenisasi input menggunakan TAPAS tokenizer

Model TAPAS membutuhkan format input tertentu, sehingga tabel dan pertanyaan dikodekan menggunakan TAPAS tokenizer.

```
inputs = tokenizer(table=table, queries=[question],
padding="max_length", return_tensors="pt")
```

`table=table`: Memasukkan *Dataframe* sebagai tabel yang akan diproses.

`queries=[question]`: Memasukkan pertanyaan pengguna.

`padding="max_length"`: Mengatur *padding* agar panjang input sesuai dengan model.

`return_tensors="pt"`: Mengubah hasil tokenisasi menjadi tensor PyTorch agar bisa digunakan dalam model

Proses pemahaman *Dataframe* ke LLM dilakukan dengan tokenisasi tabel menggunakan `TapasTokenizer`. Kemudian hasil tokenisasi dikirim ke model *Table Parsing* (TAPAS) untuk menghasilkan jawaban berdasarkan tabel.

4.2 Hasil Perancangan *User Interface*

Tahap ini merupakan implementasi pada platform *website* sistem yang dibuat berupa *automatic visualization* dari data tabel. Berikut adalah hasil *website* yang telah dibuat



Gambar 4. 2 Halaman *Upload File*

Gambar 4.2 merupakan tampilan antarmuka dari *website Automatic Visualization Data Tabel with TAPAS*. Halaman ini merupakan area untuk mengunggah *file* yang mendukung format CSV, Excel, dan PDF. Pengguna dapat mengunggah *file* dengan menggunakan tombol “*Browse Files*” untuk memilih *file*

dari perangkat. Batas ukuran maksimal *file* yang di unggah adalah 200MB. Setelah *file* berhasil di unggah, sistem akan secara otomatis menampilkan daftar tabel yang terdapat dalam *file* tersebut.

Tabel 1

	Dusun	Laki-Laki	Perempuan	Jumlah
0	1. Jurgang	390	391	781
1	2. Muncar	227	225	452
2	3. Laok Lorong	226	250	476
3	Jumlah	843	866	1709

Tabel 2

	No	Nama Dusun	Jumlah Penduduk Menurut Kelor
0	None	None	0-1
1	1	Jurgang	11
2	2	Muncar	2
3	3	Laok Lorong	9
4		Jumlah	22

Gambar 4. 3 Tampilan Tabel yang ditemukan dari *File* yang di *upload*

Gambar 4.3 merupakan tampilan yang dihasilkan setelah mengupload *file*. Setelah *file* di *upload*, sistem akan mengekstraksi seluruh tabel yang terdapat dalam *file* yang diunggah. Semua tabel yang berhasil diekstraksi kemudian ditampilkan

pada sistem. Jika *file* yang diupload memiliki lebih dari satu tabel, sistem akan menampilkan semua tabel yang tersedia dengan memberikan label sesuai urutan.

Pilih tabel untuk dianalisis:



Gambar 4. 4 Tampilan untuk Memilih Tabel

Gambar 4.4 merupakan tampilan untuk memilih tabel yang ingin divisualisasikan. Jika *file* yang di upload memiliki lebih dari satu tabel, dapat memilih tabel mana yang ingin di visualisasikan.

Pilih tabel untuk dianalisis:



Tabel yang Dipilih:

	Dusun	Laki-Laki	Perempuan	Jumlah
0	1. Jurgang	390	391	781
1	2. Muncar	227	225	452
2	3. Laok Lorong	226	250	476
3	Jumlah	843	866	1709

Gambar 4. 5 Tampilan Setelah Memilih Tabel

Gambar 4.5 merupakan tampilan setelah memilih tabel yang ingin di visualisasikan. Setelah memasukkan pilihan tabel yang divisualisasikan, sistem akan menampilkan tabel yang telah dipilih. Dengan tampilan tersebut, maka tidak perlu menscroll untuk melihat kembali tabel yang ingin divisualisasikan.

Ajukan pertanyaan tentang tabel ini:

Insight

Gambar 4. 6 Tampilan untuk Memasukkan Pertanyaan

Gambar 4.6 merupakan halaman untuk memasukkan pertanyaan berdasarkan tabel yang telah di pilih untuk di visualisasikan. Pertanyaan yang dimasukkan dapat berupa berbagai jenis, seperti perbandingan, *sorting*, dan berbagai jenis pertanyaan lain berdasarkan tabel. Setelah pertanyaan dimasukkan klik “*insight*” untuk mendapatkan jawaban.

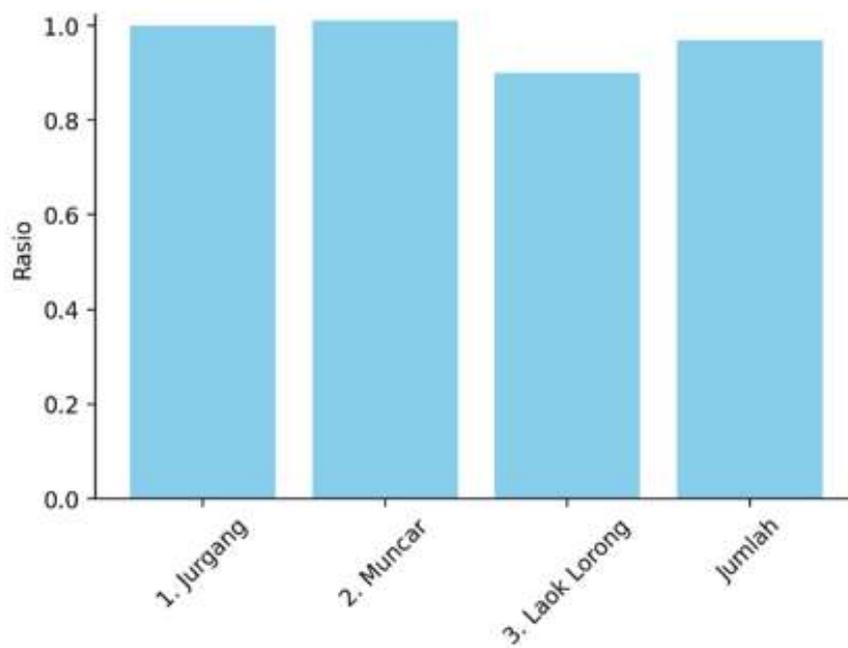
Ajukan pertanyaan tentang tabel ini:

berapa perbandingan jumlah laki-laki dan perempuan di seluruh dusun?

Insight

 **Insight:**

Dusun Laki-Laki Perempuan Perbandingan 1. Jurgang 390 391 1.00 2. Muncar 227 225 1.01 3. Laok Lorong 226 250 0.90
Jumlah 843 866 0.97



Gambar 4. 7 Hasil *Insight* dan Visualisasi

Gambar 4.7 merupakan hasil jawaban berdasarkan pertanyaan yang telah dimasukkan. Setelah memasukkan pertanyaan terkait tabel, sistem akan memproses pertanyaan menggunakan model *Table Parsing* (TAPAS) berbasis LLM untuk menghasilkan jawaban yang relevan. Jawaban yang dihasilkan berupa *insight* yang didapat dari tabel dan visualisasi berbentuk diagram batang.

4.3 Pembahasan

Berdasarkan hasil dari sistem yang telah dibuat dengan menggunakan model *Table Parsing* (TAPAS), sistem memiliki performa yang cukup baik dalam mengekstraksi data tabel dan menjawab pertanyaan berbasis tabel. Model ini mampu memahami struktur tabel dan menghasilkan jawaban yang relevan sesuai dengan pertanyaan yang diberikan.

Dalam menjawab pertanyaan *Table Parsing* (TAPAS), sangat bergantung pada struktur tabel dan kompleksitas pertanyaan. Untuk pertanyaan sederhana seperti nilai pada suatu kolom, model dapat memberikan jawaban dengan tingkat akurasi yang tinggi. Namun untuk pertanyaan yang kompleks seperti melibatkan perbandingan atau pengurutan, model terkadang memberikan hasil yang kurang akurat. Hal ini dikarenakan *Table Parsing* (TAPAS) memiliki keterbatasan dalam memahami hubungan antar kolom. Sehingga untuk pertanyaan yang mengandung banyak variabel *Table Parsing* (TAPAS) bekerja kurang maksimal.

Selain itu, kecepatan dalam menghasilkan jawaban dan visualisasi relatif cepat. Hasil jawaban pertanyaan tidak hanya berupa *insight* tetapi juga dalam bentuk diagram batang. Diagram batang digunakan untuk mempermudah dalam memahami data, sehingga informasi menjadi lebih mudah dipahami.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian ini, dapat diambil kesimpulan tentang Sistem *Automatic Visualization* Dari Data Tabel Menggunakan *Large Language Models* (LLM), sistem yang dihasilkan mampu bekerja dengan baik dalam memberikan jawaban *insight* dan visualisasi. Sistem memungkinkan untuk mengunggah *file* yang berisi tabel, kemudian hasil dari pertanyaan yang diberikan akan memperoleh jawaban berupa *insight* dalam bentuk teks dan visualisasi grafik. Hal ini ditunjukkan dengan Mean Absolute Percentage Error (MAPE) rata-rata sebesar 17.66%. Meskipun demikian, terdapat jawaban dengan tingkat kesalahan tinggi, seperti nilai MAPE terbesar yang mencapai 121.48%. Hasil evaluasi ini menekankan pentingnya penurunan MAPE untuk meningkatkan kinerja sistem dalam menjawab pertanyaan berbasis data tabel secara otomatis.

5.2 Saran

Saran untuk pengembangan sistem lebih lanjut adalah untuk meningkatkan kemampuan sistem dalam mengekstraksi tabel yang terpotong halaman, menjawab pertanyaan yang lebih spesifik, mengembangkan berbagai bentuk visualisasi, dan melakukan evaluasi sistem dengan berbagai metrik evaluasi. Pengambilan jawaban berdasarkan struktur tabel, sehingga tabel dengan struktur header yang tidak rapi akan mempengaruhi hasil *insight* dan visualisasi. Kualitas jawaban dan visualisasi juga masih memerlukan peningkatan dan dapat menggunakan model yang lebih bervariasi agar mendapatkan hasil yang lebih baik.

DAFTAR PUSTAKA

- Bommasani, R. *dkk.* (2021) “On the Opportunities and Risks of Foundation Models.”
- Chen, W. *dkk.* (2020) “Tabfact: a Large-Scale Dataset for Table-Based Fact Verification,” *8th International Conference on Learning Representations, ICLR 2020*.
- Dong, H. dan Wang, Z. (2024) “Large Language Models for Tabular Data: Progresses and Future Directions,” *SIGIR 2024 - Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, hal. 2997–3000. doi: 10.1145/3626772.3661384.
- Hernikawati, D. (2021) “Analisis Dampak Pandemi COVID-19 terhadap Jumlah Kunjungan pada Situs E-Commerce di Indonesia Menggunakan Uji T Berpasangan,” *Jurnal Studi Komunikasi dan Media*, 25(2), hal. 191. doi: 10.31445/jskm.2021.4389.
- Herzig, J. *dkk.* (2020) “T A P AS : Weakly Supervised Table Parsing via Pre-training,” hal. 4320–4333.
- Lu, W. *dkk.* (2024) “Large Language Model for Table Processing: A Survey,” 0(0), hal. 1–32.
- Lubis, A. T. U. B. *dkk.* (2024) “Question Answering System pada Chatbot Telegram Menggunakan Large Language Models (LLM) dan Langchain (Studi Kasus UU Kesehatan),” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), hal. 955–964. doi: 10.57152/malcom.v4i3.1378.
- Prasetyo, V. R., Benarkah, N. dan Chrisintha, V. J. (2021) “Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya Implementation of Natural Language Processing in Creating Chatbots at Information Technology Program , University of Surabaya,” 10(2), hal. 114–121. doi: 10.34148/teknika.v10i2.370.
- Su, A. *dkk.* (2024) “TableGPT2: A Large Multimodal Model with Tabular Data

Integration,” hal. 1–32.

Tom *dkk.* (2020) “Language Models are Few-Shot Learners,” (Juli).

Zan, D. *dkk.* (2023) “Large Language Models Meet NL2Code: A Survey,”
Proceedings of the Annual Meeting of the Association for Computational Linguistics, 1, hal. 7443–7464. doi: 10.18653/v1/2023.acl-long.411.

Zhang, X. *dkk.* (2024) “TableLLM: Enabling Tabular Data Manipulation by LLMs in Real Office Usage Scenarios.”

Zhao, B. *dkk.* (2023) “Large Language Models are Complex Table Parsers,”
EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings, hal. 14786–14802. doi: 10.18653/v1/2023.emnlp-main.914.

