

BAB III

PELAKSANAAN KERJA

3.1 Kedudukan dan Koordinasi

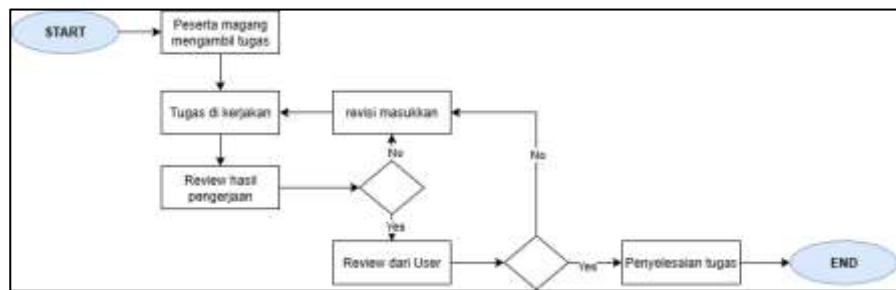
Posisi magang sebagai *Data Scientist* Intern berada di bawah Divisi *Research and Development* (R&D) PT. Mandala Prima Makmur. Perusahaan menempatkan peserta magang ke dalam proyek atau tim tertentu berdasarkan kompetensi dan kemampuan yang dimiliki oleh masing-masing intern. Penempatan ini bersifat fleksibel dan disesuaikan dengan kebutuhan proyek yang sedang berjalan.

Sebagai mahasiswa Program Studi Sistem Informasi dengan peminatan Big *Data Analytics* yang berfokus pada analisis dan pengolahan data, peserta magang menjalankan peran sebagai *Data Scientist* Intern yang dapat terlibat pada berbagai tim sesuai penugasan. Apabila peserta magang ditempatkan pada tim *Research and Development*, maka seluruh tugas yang diberikan pada tim tersebut akan dikerjakan hingga selesai sebelum berpindah ke tim lain apabila diperlukan.

Tugas-tugas yang dikerjakan berkaitan dengan pengembangan dan pemanfaatan *Artificial Intelligence* (AI) untuk kebutuhan internal perusahaan. Kedudukan posisi peserta magang dalam struktur Divisi *Research and Development* ditunjukkan pada Gambar 3.1

3.1.1 Koordinasi

Bagian ini menjelaskan alur koordinasi pekerjaan yang dilakukan peserta magang selama menjalani kegiatan magang di PT. Mandala Prima Makmur. Alur koordinasi tersebut dapat dilihat pada bagan alur kerja yang ditampilkan pada Gambar 3.1:



Gambar 3. 1. Bagan alur kordinasi

Alur kerja pada Gambar 3.1 dimulai dari pemberian tugas oleh Supervisor kepada peserta magang. Tugas tersebut umumnya berasal dari permintaan langsung pengguna (*user*) dan berkaitan dengan kegiatan riset maupun pengembangan sistem baru yang sedang dikembangkan oleh perusahaan. Pada tahap awal, Supervisor menjelaskan ruang lingkup pekerjaan, tujuan yang ingin dicapai, serta batasan-batasan yang perlu diperhatikan selama proses penggerjaan. Supervisor juga menyediakan referensi atau dokumentasi pendukung untuk membantu peserta magang memahami konteks dan kebutuhan tugas.

Setelah menerima arahan, peserta magang mulai mengerjakan tugas sesuai dengan instruksi yang diberikan. Selama proses penggerjaan, komunikasi dengan Supervisor tetap dilakukan secara berkala, terutama apabila terdapat kendala teknis atau kebutuhan klarifikasi tambahan. Koordinasi yang berkelanjutan ini bertujuan agar pekerjaan dapat berjalan secara efektif dan sesuai dengan ekspektasi.

Hasil pekerjaan yang telah diselesaikan kemudian melalui tahap review oleh Supervisor. Pada tahap ini, Supervisor mengevaluasi hasil kerja berdasarkan kualitas implementasi, kesesuaian solusi, serta kepatuhan

terhadap spesifikasi yang telah ditetapkan. Apabila hasil pekerjaan belum memenuhi standar, Supervisor akan memberikan masukan untuk dilakukan perbaikan. Peserta magang kemudian melakukan revisi hingga hasil dinyatakan layak.

Setelah mendapatkan persetujuan dari Supervisor, hasil pekerjaan dilanjutkan ke tahap review oleh pengguna (user). Pada tahap ini, pengguna menilai kesesuaian hasil dengan kebutuhan bisnis. Jika masih terdapat aspek yang perlu disesuaikan, peserta magang akan melakukan revisi lanjutan berdasarkan masukan yang diberikan. Tahap ini dapat mencakup penyempurnaan fitur, peningkatan performa sistem, maupun penyesuaian dokumentasi pendukung.

Apabila seluruh perbaikan telah dilakukan dan hasil akhir dinyatakan sesuai oleh Supervisor dan pengguna, maka tugas dinyatakan selesai. Selanjutnya, hasil pekerjaan didokumentasikan sebagai laporan akhir yang dapat digunakan sebagai referensi untuk pengembangan sistem di masa mendatang. Dengan demikian, kegiatan magang tidak hanya memberikan pengalaman praktis, tetapi juga menghasilkan kontribusi nyata bagi perusahaan.

3.2 Tugas yang Dilakukan

Selama pelaksanaan magang di PT. Mandala Prima Makmur, peserta magang sebagai *Data Scientist Intern* pada Divisi *Research and Development* melaksanakan berbagai tugas yang berkaitan dengan bidang *big data* dan *Artificial Intelligence*. Tugas-tugas tersebut berfokus pada pemanfaatan data dan teknologi AI untuk mendukung peningkatan efektivitas dan pengambilan keputusan perusahaan.

Peserta magang terlibat langsung dalam berbagai proyek yang diberikan oleh Supervisor dan tim R&D. Keterlibatan ini bertujuan untuk memberikan pengalaman praktis serta pemahaman yang lebih mendalam mengenai penerapan teknologi AI di lingkungan industri. Selama masa magang, tugas yang dikerjakan mencakup

proses pengolahan data, analisis, serta pengembangan solusi berbasis AI sesuai dengan kebutuhan perusahaan.

Seluruh tanggung jawab dan pekerjaan yang telah dilakukan selama periode magang dirangkum dalam Tabel 3.1, yang menampilkan daftar tugas peserta magang selama menjalani kegiatan magang di PT. Mandala Prima Makmur.

Tabel 3. 1. Detail Pekerjaan yang Dilakukan

No.	Kegiatan	Koordinasi	Pelaksanaan
1.	<i>Webscraping X on Cocoa Powder</i>	Marketing Manager	1 September 2025
1.1	<i>Analyzing web scraping data</i>		1-2 September 2025
1.2	<i>Prepare the web scraping data</i>		3-7 September 2025
1.3	<i>Data manipulation to have a good analysis</i>		8-10 September 2025
1.5	<i>Presenting the analysis that was done</i>		11-12 September 2025
2.	<i>Customer ChatBot</i>	Marketing Manager	15-16 September 2025
2.1	<i>Understanding what user needs</i>		17-19 September 2025
2.2	<i>Brainstorming ideas</i>		22-24 September 2025
2.3	<i>Knowledge Gathering</i>		25-30 September 2025
2.5	<i>Finetuning prompt</i>		6-8 October 2025
2.6	<i>Implementing knowledge to the LLM</i>		9-10 October 2025

No.	Kegiatan	Koordinasi	Pelaksanaan
3.	<i>Customer ChatBot - LLM</i>	Marketing Manager	
3.1	<i>Understanding what user needs</i>		13-14 October 2025
3.2	<i>Brainstorming ideas</i>		15-17 October 2025
3.3	<i>Researching the best LLM to use</i>		20-22 October 2025
3.4	<i>Implementing to python code</i>		23-31 October 2025
3.5	<i>Generating best prompt</i>		3-5 November 2025
3.6	<i>Inject knowledge to the LLM</i>		6-7 November 2025
3.7	<i>Make simple UI</i>		10-11 November 2025
4.	<i>Market 2025 Analysis</i>		
4.1	<i>Gathering sales data</i>	Marketing Manager	12 November 2025
4.2	<i>Understanding the data</i>		13 November 2025
4.3	<i>Preparing the data</i>		14 November 2025
4.4	<i>Data manipulation</i>		17-18 November 2025
4.5	<i>Data implementation</i>		19-20 November 2025
4.6	<i>Visualization</i>		21 November 2025

UNIVERSITAS
MULTIMEDIA
NUSANTARA

No.	Kegiatan	Koordinasi	Pelaksanaan
5.	Market Forecasting	Marketing Team and Sales Department	
5.1	<i>Collect and prepare historical sales data</i>		24-26 November 2025
5.2	<i>Structure monthly sales</i>		27-28 November 2025
5.3	<i>Clean and validate sales data</i>		1-3 December 2025
5.4	<i>Analyze sales using EDA</i>		4-5 December 2025
5.5	<i>Build time series forecasting model</i>		8-12 December 2025
5.6	<i>Generate 2026 market forecast and analyze trend & seasonality</i>		15–17 December 2025
5.7	<i>Generate visualization results and prepare analytical report</i>		18-24 December 2025
5.8	<i>Present forecasting insights</i>		25-31 December 2025

Tabel 3.1 berisi daftar tugas yang dikerjakan selama kegiatan magang di PT. Mandala Prima Makmur pada periode September hingga Desember 2025. Tugas-tugas tersebut meliputi kegiatan *web scraping* data pada platform X terkait *cocoa powder*, yang mencakup analisis data hasil *scraping*, persiapan data, manipulasi data, pembuatan visualisasi, serta penyampaian hasil analisis kepada pihak terkait.

Selain itu, peserta magang juga terlibat dalam pengembangan *Customer ChatBot*, yang meliputi proses memahami kebutuhan pengguna, *brainstorming* ide, pengumpulan *knowledge*, penyusunan dan penyempurnaan *prompt*, serta implementasi *knowledge* ke dalam model bahasa.

Pengembangan lanjutan *Customer ChatBot* berbasis *Large Language Model* (LLM) dilakukan melalui tahapan memahami kebutuhan pengguna, *brainstorming*, riset model LLM yang digunakan, implementasi ke dalam kode *Python*, penyusunan

prompt, integrasi *knowledge* ke dalam model, serta pembuatan antarmuka pengguna (*user interface / UI*) sederhana.

Tugas lainnya yang dikerjakan adalah analisis pasar tahun 2025, yang mencakup pengumpulan data penjualan, pemahaman data, persiapan dan manipulasi data, implementasi data untuk analisis, serta pembuatan visualisasi hasil analisis.

Selain itu, peserta magang juga melaksanakan *market forecasting*, yang meliputi pengumpulan dan persiapan data penjualan historis, penyusunan data penjualan bulanan, pembersihan dan validasi data, analisis data menggunakan *exploratory data analysis* (EDA), pembangunan model peramalan *time series*, pembuatan proyeksi pasar tahun 2026, visualisasi hasil peramalan, serta penyampaian hasil analisis kepada pihak terkait.

Seluruh kegiatan tersebut dilaksanakan secara bertahap dan dikoordinasikan dengan tim *Marketing*, *Sales Department*, dan *Research and Development*, sesuai dengan kebutuhan masing-masing proyek.

3.3 Uraian Pelaksanaan Kerja

Peserta magang melaksanakan praktik kerja magang selama total 640 jam di PT Mandala Prima Makmur, dengan kegiatan yang diarahkan dan dibimbing langsung oleh *Research and Development Manager* serta *Marketing team*. Selama periode magang, peserta mendapatkan beberapa tugas utama yang berkaitan langsung dengan pengembangan sistem berbasis *Artificial Intelligence (AI)* dan pengolahan data. Salah satu tugas yang menjadi bagian penting dalam proses pembelajaran dan kontribusi peserta magang adalah proyek *web scraping*, yang dijelaskan sebagai berikut:

3.3.1 Proses Pelaksanaan

Peserta magang melaksanakan praktik kerja magang di PT Mandala Prima Makmur dengan total durasi 640 jam. Selama pelaksanaan magang, peserta terlibat dalam berbagai proyek yang berkaitan dengan pengolahan data dan pengembangan sistem berbasis *Artificial Intelligence (AI)*, khususnya pada divisi *Research and Development*. Proyek-proyek yang

dikerjakan mencakup *web scraping*, pengembangan *Customer ChatBot*, analisis pasar, serta peramalan penjualan sebagai bentuk kontribusi terhadap kebutuhan analisis dan pengambilan keputusan perusahaan:

3.3.1.1 Web Scraping X on Cocoa Powder

Pada awal pelaksanaan magang, salah satu tugas utama yang dikerjakan adalah melakukan *web scraping* pada platform media sosial X (sebelumnya Twitter). Kegiatan ini dilaksanakan bekerja sama dengan tim Marketing PT. Mandala Prima Makmur untuk mendukung analisis pasar dan sentimen publik terhadap produk berbasis cocoa sebagai komoditas utama perusahaan.

Tujuan dari kegiatan ini adalah mengumpulkan data percakapan dan opini pengguna X yang berkaitan dengan kata kunci “cocoa”, “chocolate”, serta istilah relevan lainnya. Data yang diperoleh digunakan untuk mengidentifikasi persepsi publik, tren percakapan, dan potensi pasar global yang dapat dimanfaatkan dalam perumusan strategi pemasaran dan pengembangan produk.

Proses pengumpulan data dilakukan menggunakan *Tweet Harvest*, yaitu *open-source library* yang digunakan untuk mengekstraksi data publik dari platform X melalui *API endpoint* yang tersedia. Data yang diambil mencakup atribut utama seperti *username*, *timestamp*, *retweet count*, *like count*, dan isi teks cuitan. Seluruh data hasil ekstraksi disimpan dalam format CSV untuk memudahkan proses analisis lanjutan menggunakan Python dan Microsoft Excel.

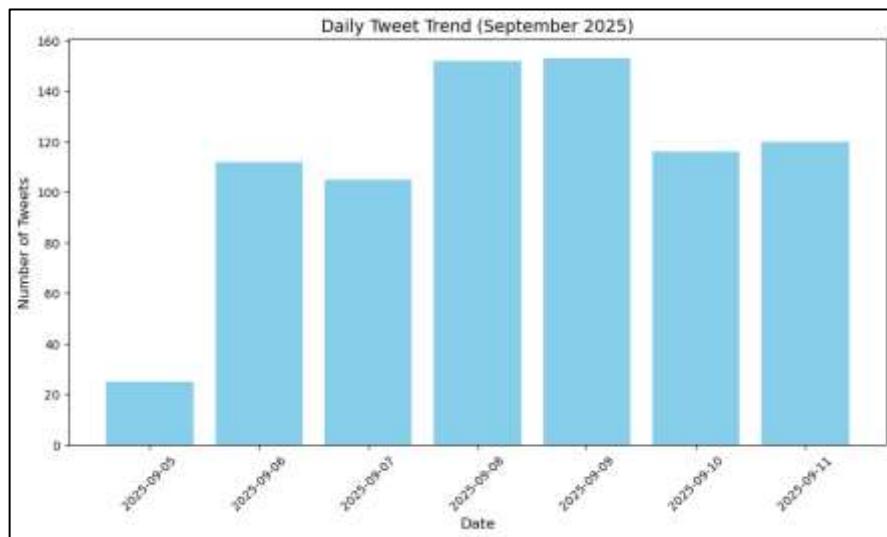
Berdasarkan arahan dari tim Marketing, target awal pengambilan data ditetapkan sebanyak 1.200 baris cuitan. Namun, akibat keterbatasan akses data publik dan pembatasan kuota dari API X, jumlah data yang berhasil dikumpulkan sebanyak 782 baris data yang valid. Data tersebut kemudian melalui tahap verifikasi dan *data*

cleaning untuk menghilangkan cuitan yang tidak relevan, duplikasi, serta konten bersifat *spam*.

Hasil *web scraping* yang telah dibersihkan digunakan sebagai dasar analisis tren pasar dan sentimen konsumen terhadap produk berbasis cocoa. Hasil analisis selanjutnya divisualisasikan dalam bentuk grafik dan tangkapan layar, sebagaimana ditunjukkan pada Gambar 3.2 hingga Gambar 3.6.

Gambar 3. 2. Data Used

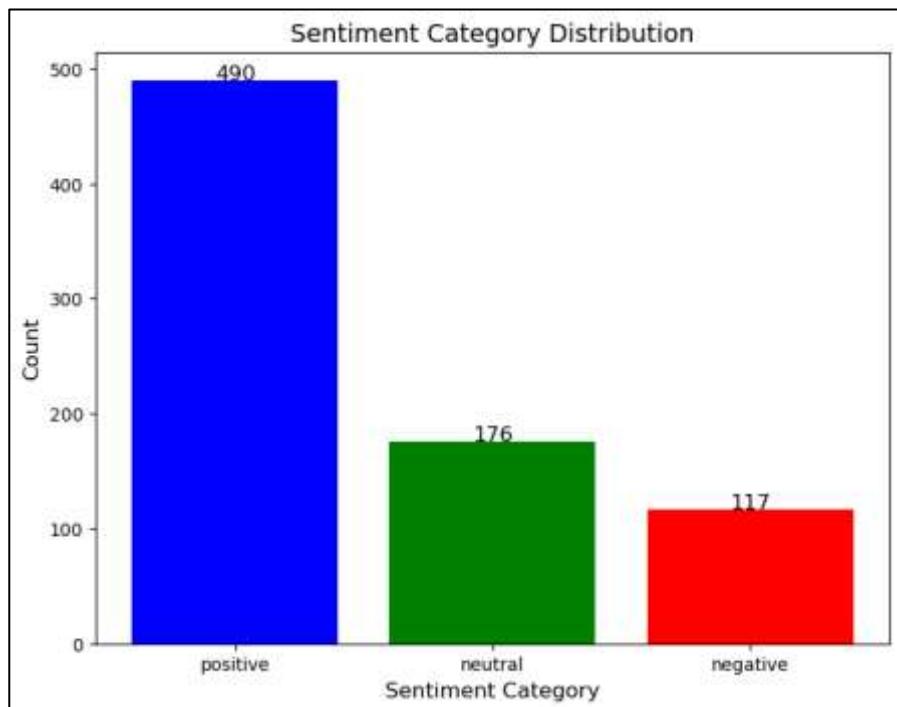
Gambar 3.2 menampilkan cuplikan data mentah hasil *web scraping* menggunakan *Tweet Harvest*. Dataset terdiri dari 782 baris data yang merepresentasikan cuitan unik pengguna X dengan mayoritas bahasa Inggris (*lang = en*). Kolom utama yang digunakan dalam analisis meliputi *conversation_id_str*, *created_at*, *favorite_count*, *full_text*, *tweet_url*, dan *user_id_str*. Konten cuitan didominasi oleh pembahasan minuman berbasis cokelat seperti *hot chocolate* dan *chocolate milk*.



Gambar 3. 3. Data Tren

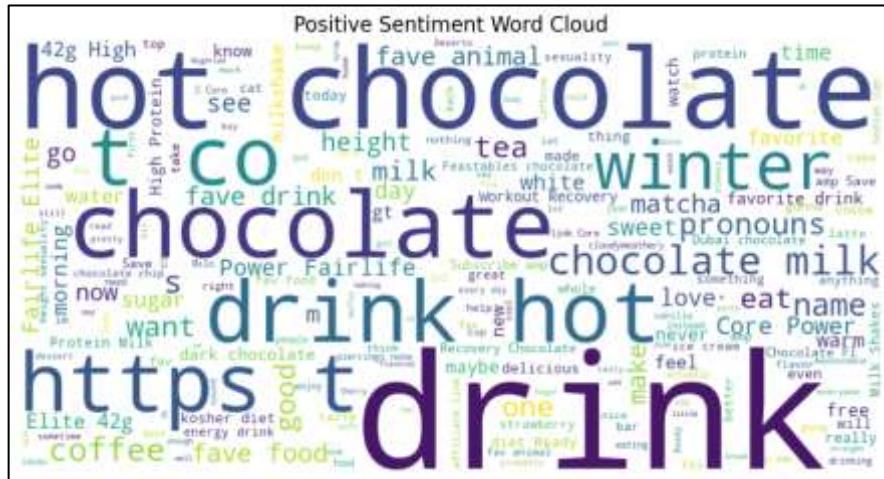
Gambar 3.3 menunjukkan tren jumlah cuitan harian selama periode pengambilan data pada September 2025. Aktivitas tertinggi tercatat pada 8 dan 9 September 2025 dengan sekitar 150 cuitan per hari, sedangkan aktivitas terendah terjadi pada 5 September 2025 dengan kurang dari 30 cuitan. Pola ini mengindikasikan bahwa diskusi publik terkait cocoa lebih aktif pada hari kerja dibandingkan akhir pekan.





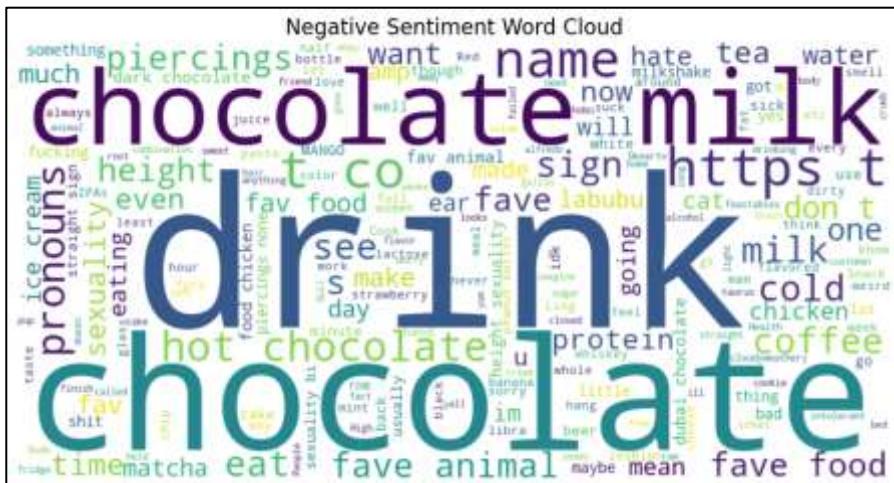
Gambar 3.4. Sentimen BarChart

Gambar 3.4 memperlihatkan hasil analisis sentimen terhadap 782 cuitan yang diklasifikasikan ke dalam tiga kategori, yaitu positif, netral, dan negatif. Hasil analisis menunjukkan dominasi sentimen positif sebanyak 490 cuitan, diikuti sentimen netral sebanyak 176 cuitan, dan sentimen negatif sebanyak 117 cuitan. Temuan ini mengindikasikan bahwa persepsi publik terhadap produk berbasis cocoa cenderung positif.



Gambar 3. 5. Positif WordCloud

Visualisasi *word cloud* pada Gambar 3.5 menampilkan kata-kata yang paling sering muncul pada cuitan dengan sentimen positif. Kata seperti “*drink*”, “*chocolate*”, “*hot*”, “*milk*”, dan “*winter*” muncul dengan frekuensi tinggi, yang menunjukkan bahwa persepsi positif terhadap cocoa banyak dikaitkan dengan konsumsi minuman cokelat hangat.



Gambar 3. 6. Negatif WordCloud

Gambar 3.6 menampilkan *word cloud* dari cuitan bersentimen negatif. Meskipun kata “*chocolate*” dan “*drink*” masih muncul, terdapat pula kata dengan konotasi negatif seperti “*sick*”, “*cold*”, dan “*hate*”. Namun, proporsi sentimen negatif relatif kecil

dibandingkan sentimen positif, sehingga tidak memberikan pengaruh signifikan terhadap persepsi pasar secara keseluruhan.

Secara umum, hasil *web scraping* ini menunjukkan bahwa persepsi publik terhadap produk berbasis cocoa di media sosial X cenderung positif. Temuan ini dapat dimanfaatkan oleh tim Marketing PT. Mandala Prima Makmur sebagai dasar pendukung dalam penyusunan strategi promosi, riset tren konsumen, serta pengembangan produk yang lebih sesuai dengan preferensi pasar.

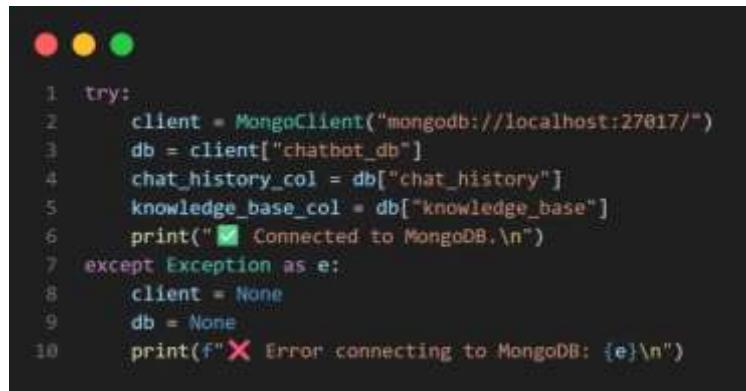
Kegiatan ini juga memberikan pengalaman praktis dalam penerapan konsep *data-driven marketing*, mulai dari proses *data acquisition*, *data preprocessing*, hingga *sentiment analysis*, serta menunjukkan bagaimana data media sosial dapat diolah menjadi *business insight* yang bernilai strategis bagi perusahaan.



3.3.1.2 Customer Chatbot - Knowlegde

Pada proyek kedua kegiatan magang, fokus utama diarahkan pada pembangunan *knowledge base* sebagai fondasi dari *Customer ChatBot* yang dikembangkan untuk mendukung proses layanan dan komunikasi di PT. Mandala Prima Makmur. Tahap ini bertujuan memastikan chatbot memiliki sumber informasi yang terstruktur, relevan, dan konsisten sesuai dengan kebutuhan perusahaan. Seluruh proses dilakukan melalui koordinasi dengan tim Marketing yang memberikan arahan terkait jenis informasi yang perlu dimasukkan ke dalam sistem.

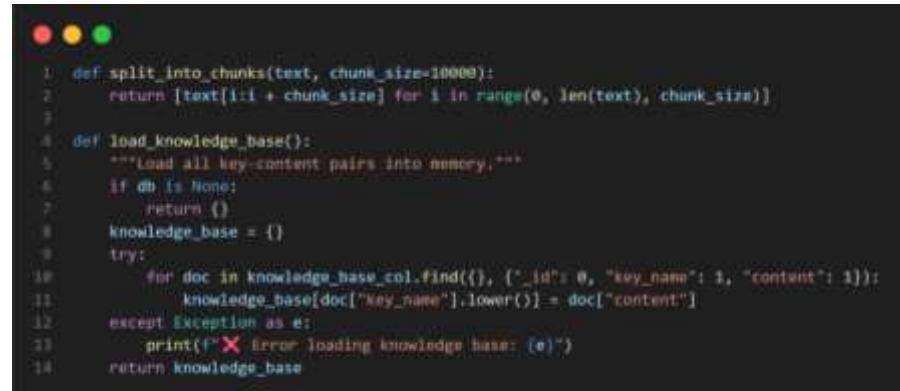
Kegiatan ini mencakup pengumpulan dan penyusunan data pengetahuan perusahaan yang meliputi profil perusahaan, spesifikasi produk *cocoa powder*, standar kualitas, prosedur pembelian, informasi teknis, serta daftar pertanyaan yang sering diajukan pelanggan (*frequently asked questions*). Informasi tersebut diperoleh dari dokumen internal, materi pemasaran, serta diskusi singkat dengan pihak terkait, kemudian disusun dalam format teks terstruktur agar dapat diproses secara konsisten oleh sistem berbasis kecerdasan buatan.



```
1 try:
2     client = MongoClient("mongodb://localhost:27017/")
3     db = client["chatbot_db"]
4     chat_history_col = db["chat_history"]
5     knowledge_base_col = db["knowledge_base"]
6     print("✅ Connected to MongoDB.\n")
7 except Exception as e:
8     client = None
9     db = None
10    print(f"❌ Error connecting to MongoDB: {e}\n")
```

Gambar 3. 7. Koneksi ke DB

Gambar 3.7 menunjukkan proses inisialisasi koneksi ke MongoDB sebagai media penyimpanan *knowledge base*. Pada tahap ini, dua koleksi utama disiapkan, yaitu *knowledge_base* dan *chat_history*, untuk memastikan struktur database telah siap digunakan sebelum proses pemuatan data dilakukan. MongoDB dipilih karena mendukung penyimpanan data semi-terstruktur dan fleksibel untuk pengembangan sistem berbasis LLM.

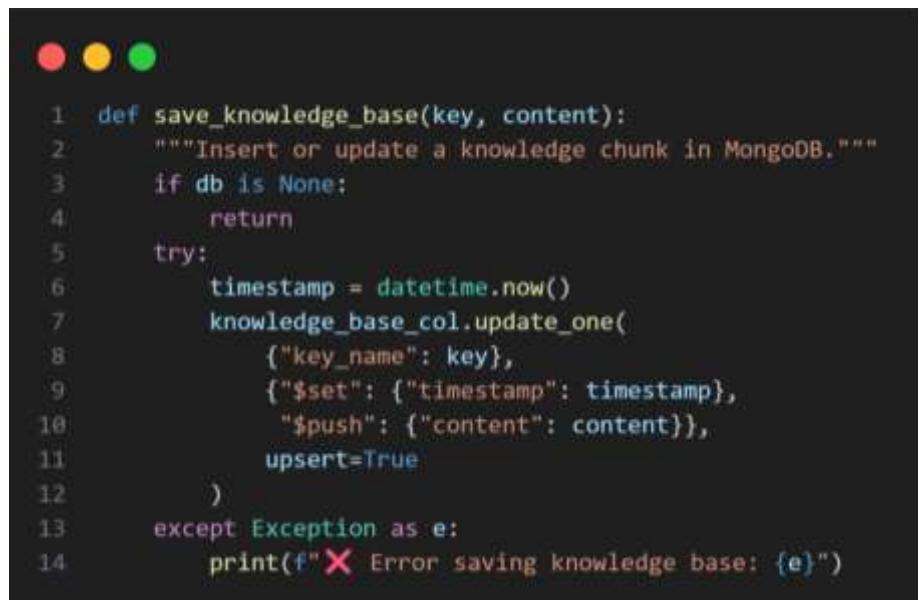


```
1 def split_into_chunks(text, chunk_size=10000):
2     return [text[i:i + chunk_size] for i in range(0, len(text), chunk_size)]
3
4 def load_knowledge_base():
5     """Load all key-content pairs into memory."""
6     if db is None:
7         return {}
8     knowledge_base = {}
9     try:
10         for doc in knowledge_base_col.find({}, {"_id": 0, "key_name": 1, "content": 1}):
11             knowledge_base[doc["key_name"].lower()] = doc["content"]
12     except Exception as e:
13         print(f"❌ Error loading knowledge base: {e}")
14     return knowledge_base
```

Gambar 3. 8. Penyimpanan ke DB

Gambar 3.8 menampilkan implementasi fungsi pemrosesan *knowledge base*, termasuk proses *chunking* teks. Pemecahan dokumen menjadi potongan kecil dilakukan agar data lebih efisien diproses oleh model bahasa. Selain itu, fungsi *load_knowledge_base*

digunakan untuk memuat data pengetahuan dari database ke dalam memori sistem.

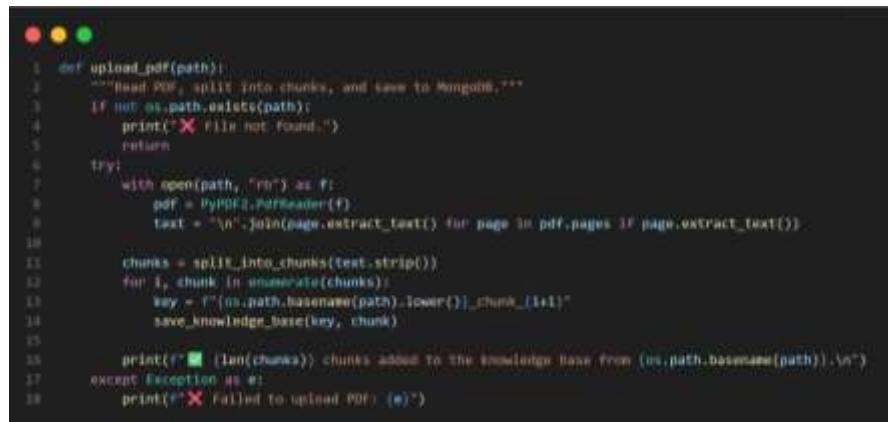


```
● ● ●
1 def save_knowledge_base(key, content):
2     """Insert or update a knowledge chunk in MongoDB."""
3     if db is None:
4         return
5     try:
6         timestamp = datetime.now()
7         knowledge_base_col.update_one(
8             {"key_name": key},
9             {"$set": {"timestamp": timestamp},
10              "$push": {"content": content}},
11              upsert=True
12          )
13     except Exception as e:
14         print(f"X Error saving knowledge base: {e}")
```

Gambar 3. 9. Penyimpanan Knowledge

Gambar 3.9 memperlihatkan proses penyimpanan *knowledge base* ke dalam MongoDB melalui fungsi *save_knowledge_base*. Setiap potongan teks disimpan sebagai satu entri dengan *key_name* dan *timestamp*, sehingga dokumen berukuran besar dapat dikelola secara terstruktur dan mendukung proses pencarian konteks yang lebih akurat.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



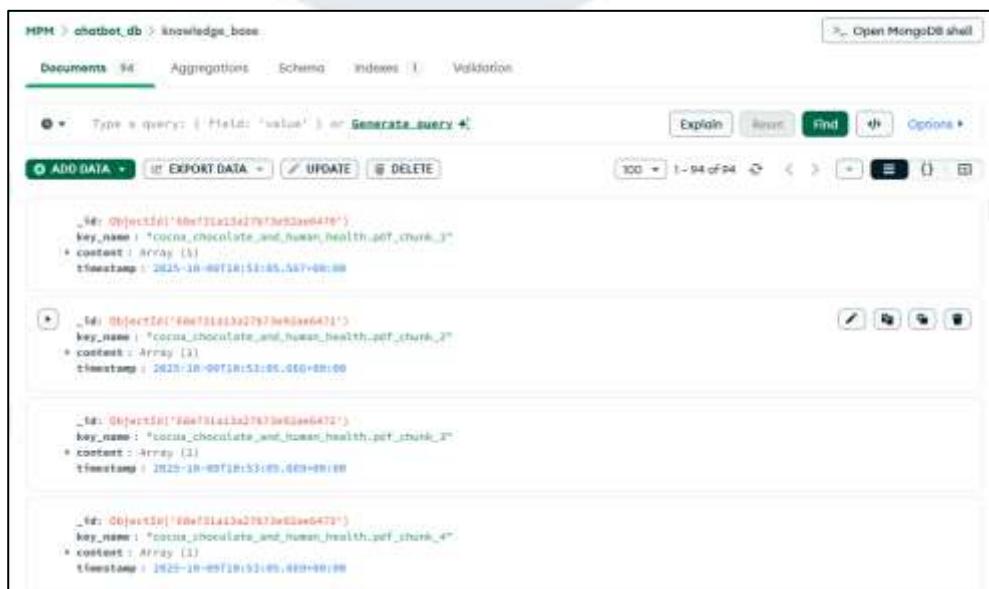
```

1 def upload_pdf(path):
2     """Read PDF, split into chunks, and save to MongoDB."""
3     if not os.path.exists(path):
4         print("File not found..")
5         return
6     try:
7         with open(path, "rb") as f:
8             pdf = PyPDF2.PdfReader(f)
9             text = "\n".join(page.extract_text() for page in pdf.pages if page.extract_text())
10
11         chunks = split_into_chunks(text.strip())
12         for i, chunk in enumerate(chunks):
13             key = f"{os.path.basename(path).lower()}_{chunk}_{i+1}"
14             save_knowledge_base(key, chunk)
15
16         print(f"> {len(chunks)} chunks added to the knowledge base from {os.path.basename(path)}.\n")
17     except Exception as e:
18         print(f"Failed to upload PDF: {e}")

```

Gambar 3. 10. Fungsi upload PDF

Gambar 3.10 menunjukkan mekanisme unggah dokumen PDF ke dalam sistem. Dokumen yang diunggah diproses melalui tahapan ekstraksi teks, pembersihan data, dan *chunking* sebelum disimpan ke database. Proses ini mempermudah integrasi dokumen perusahaan ke dalam *knowledge base* tanpa perlu input manual.



_id	key_name	content	timestamp
ObjectID('5be731a13a27873e6aa647e')	cocoa_chocolate_and_human_health.pdf_chunk_1	Array (1)	2025-10-09T18:51:05.537+00:00
ObjectID('5be731a13a27873e6aa647f')	cocoa_chocolate_and_human_health.pdf_chunk_2	Array (1)	2025-10-09T18:51:05.560+00:00
ObjectID('5be731a13a27873e6aa647g')	cocoa_chocolate_and_human_health.pdf_chunk_3	Array (1)	2025-10-09T18:51:05.603+00:00
ObjectID('5be731a13a27873e6aa647h')	cocoa_chocolate_and_human_health.pdf_chunk_4	Array (1)	2025-10-09T18:51:05.629+00:00

Gambar 3. 11. MongoDB knowledge

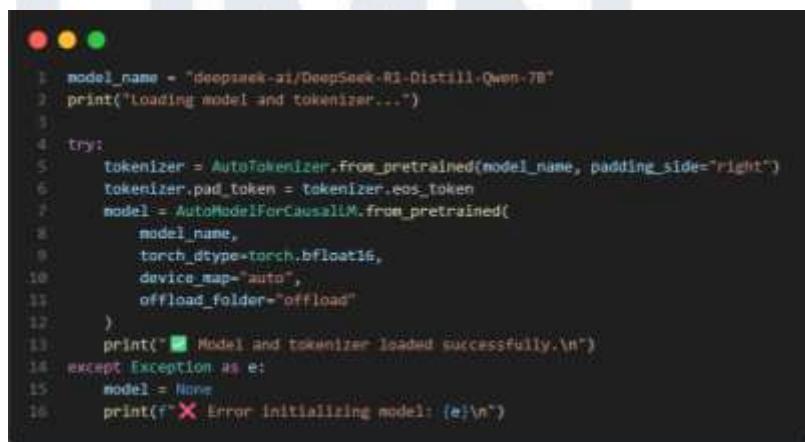
Gambar 3.11 menampilkan struktur akhir data *knowledge base* yang tersimpan di MongoDB. Setiap entri terdiri dari *key_name*, *timestamp*, dan *content* berupa potongan teks hasil ekstraksi. Struktur

ini menunjukkan bahwa data telah siap digunakan dalam mekanisme *retrieval-augmented generation* (RAG) pada tahap integrasi LLM.

Secara keseluruhan, tahap pembangunan *knowledge base* ini memastikan bahwa sistem chatbot memiliki sumber data yang terorganisasi dan mudah diakses. Proses ini memberikan dasar yang kuat bagi pengembangan chatbot berbasis LLM agar mampu menyampaikan informasi perusahaan secara konsisten dan terstruktur.

3.3.1.3 Customer ChatBot – LLM

Pada tahap ini, pengembangan *Customer ChatBot* memasuki fase integrasi *Large Language Model* (LLM) sebagai inti pemrosesan bahasa alami sistem. Fokus utama kegiatan ini adalah memastikan bahwa model bahasa dapat diinisialisasi dengan baik, terhubung dengan *knowledge base*, serta mampu memproses pertanyaan pengguna dan menyimpan hasil interaksi secara sistematis. Seluruh proses dikembangkan berdasarkan data dan struktur *knowledge base* yang telah disiapkan pada tahap sebelumnya.



```
1 model_name = "deepseek-ai/DeepSeek-R1-Distill-Qwen-7B"
2 print("Loading model and tokenizer...")
3
4 try:
5     tokenizer = AutoTokenizer.from_pretrained(model_name, padding_side="right")
6     tokenizer.pad_token = tokenizer.eos_token
7     model = AutoModelForCausalLM.from_pretrained(
8         model_name,
9         torch_dtype=torch.bfloat16,
10        device_map="auto",
11        offload_folder="offload"
12    )
13    print("✅ Model and tokenizer loaded successfully.\n")
14 except Exception as e:
15     model = None
16     print("❌ Error initializing model: {e}\n")
```

Gambar 3. 12. Integrasi LLM

Gambar 3.12 menampilkan proses pemanggilan dan inisialisasi model LLM menggunakan *DeepSeek-R1-Distill-Qwen-7B*. Model dan *tokenizer* dimuat melalui pustaka *transformers*, dengan

pengaturan *padding token* dan penggunaan tipe data *bfloat16* untuk efisiensi komputasi. Parameter *device_map="auto"* digunakan agar model dapat menyesuaikan penempatan perangkat secara optimal. Tahap ini memastikan model siap digunakan sebagai komponen utama sistem chatbot.

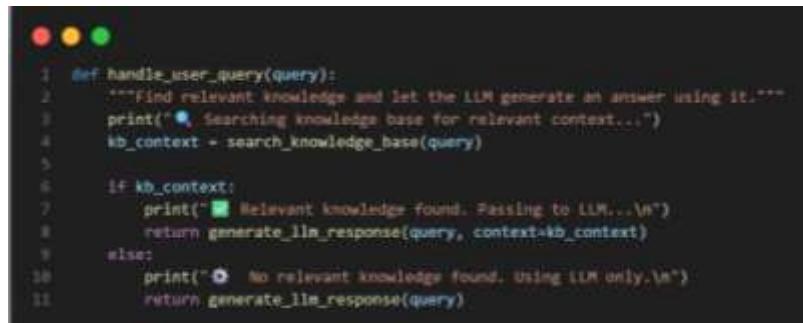


```
def generate_llm_response(query, context=None):
    """Generate a natural response using the LLM, with optional context."""
    try:
        # Build the augmented prompt
        if context:
            prompt = [
                f"You are a knowledgeable AI assistant.",
                f"(Use the information provided below to answer the user's question accurately.)",
                f"(Context){context}VM",
                f"(Question){query}VM",
                f"(Answer)VM"
            ]
            prompt = "\n".join(prompt)
        else:
            prompt = f"(User){query}VMassistant"

        print(f"\nGenerating LLM response... (This may take a few moments)\n")
        inputs = tokenizer(prompt, return_tensors="pt", truncation=True, padding=True)
        outputs = model.generate(
            **inputs,
            max_new_tokens=200,
            temperature=0.7,
            top_p=0.9
        )
    except Exception as e:
        print(f"An error occurred while generating the response: {e}")
        return None
    finally:
        del inputs, outputs, prompt
```

Gambar 3. 13. Fungsi Respon LLM

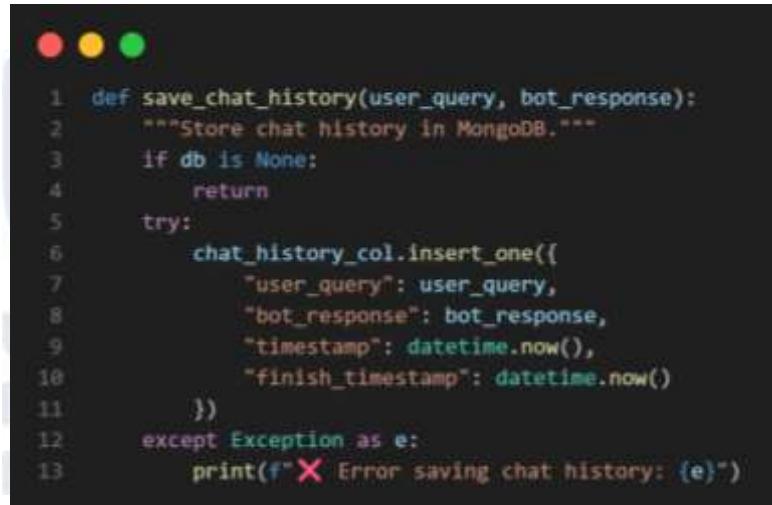
Gambar 3.13 menunjukkan fungsi *generate_llm_response* yang digunakan untuk menghasilkan respons berdasarkan input pengguna dan konteks tambahan. Fungsi ini menyusun *prompt* yang menggabungkan pertanyaan pengguna dengan konteks dari *knowledge base* sebelum diproses oleh model. Parameter generasi seperti *temperature* dan *top_p* diatur untuk menjaga konsistensi dan stabilitas respons yang dihasilkan oleh sistem.



```
1 def handle_user_query(query):
2     """Find relevant knowledge and let the LLM generate an answer using it."""
3     print("🔍 Searching knowledge base for relevant context...")
4     kb_context = search_knowledge_base(query)
5
6     if kb_context:
7         print("💡 Relevant knowledge found. Passing to LLM...")  
8         return generate_llm_response(query, context=kb_context)
9     else:
10        print("⚠️ No relevant knowledge found. Using LLM only.\n")
11        return generate_llm_response(query)
```

Gambar 3. 14. Fungsi query user

Gambar 3.14 menampilkan fungsi *handle_user_query* yang berperan dalam pengelolaan alur pertanyaan pengguna. Sistem melakukan pencarian konteks relevan melalui fungsi *search_knowledge_base* sebelum meneruskan permintaan ke LLM. Jika konteks ditemukan, informasi tersebut digunakan sebagai tambahan input; jika tidak, sistem tetap memproses pertanyaan tanpa konteks. Pendekatan ini menerapkan konsep *retrieval-augmented generation* (RAG) untuk mendukung pemrosesan jawaban berbasis data internal.



```
1 def save_chat_history(user_query, bot_response):
2     """Store chat history in MongoDB."""
3     if db is None:
4         return
5     try:
6         chat_history_col.insert_one({
7             "user_query": user_query,
8             "bot_response": bot_response,
9             "timestamp": datetime.now(),
10            "finish_timestamp": datetime.now()
11        })
12    except Exception as e:
13        print(f"🔴 Error saving chat history: {e}")
```

Gambar 3. 15. Fungsi penyimpanan chat

Gambar 3.15 menunjukkan fungsi *save_chat_history* yang digunakan untuk menyimpan riwayat percakapan ke dalam MongoDB. Setiap interaksi dicatat dalam koleksi *chat_history* yang

mencakup pertanyaan pengguna, respons sistem, serta *timestamp*. Mekanisme ini bertujuan untuk memastikan seluruh interaksi terdokumentasi dan dapat digunakan untuk keperluan pemeliharaan serta pengembangan sistem.



```
_id: ObjectId("64b794f8a2339a30a21")
user_query: "What is racism?"
text_response: "Racism is a complex mixture of economic and oligemonic phenomena. Future aspects should be focused on the possible interaction between flaxans in the gastrointestinal milieu and their synergic effects on their bioavailability."
timestamp: 2023-10-09T11:07:28.779Z
finish_timestamp: 2023-10-09T11:07:28.779Z
ObjectID
string
string
Date
Date
CANCEL UPDATE
```

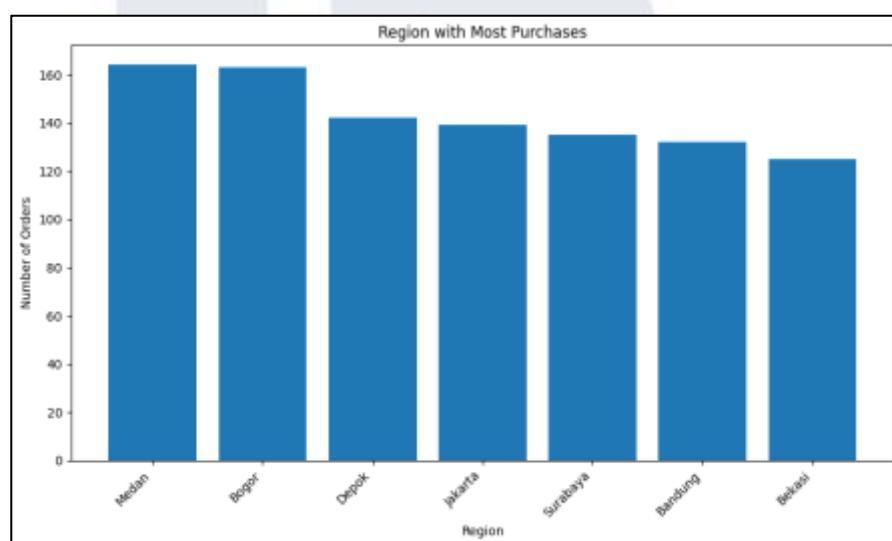
Gambar 3. 16. Contoh respon LLM yang disimpan di MongoDB

Gambar 3.16 memperlihatkan contoh struktur data percakapan chatbot yang tersimpan di MongoDB. Contoh ini digunakan untuk memverifikasi bahwa proses penyimpanan data berjalan dengan baik, mulai dari penerimaan input pengguna hingga pencatatan respons sistem. Visualisasi ini menegaskan bahwa integrasi antara LLM, *knowledge base*, dan sistem penyimpanan telah berfungsi secara end-to-end.



3.3.1.4 Market 2025 Analysis

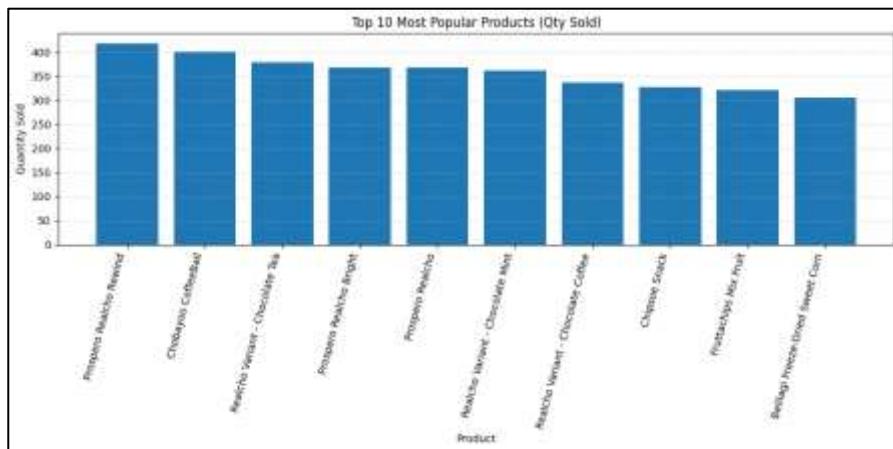
Pada proyek ketiga, kegiatan magang difokuskan pada analisis pasar tahun 2025 yang berkaitan dengan produk cocoa dan turunannya. Analisis ini dilakukan untuk mendukung proses pengambilan keputusan strategis PT. Mandala Prima Makmur, khususnya dalam perencanaan pemasaran, pengembangan produk, dan penyesuaian strategi bisnis. Data yang digunakan berasal dari data penjualan internal perusahaan yang dianalisis menggunakan pendekatan deskriptif dan visualisasi data.



Gambar 3. 17. Pembelian berdasarkan regional

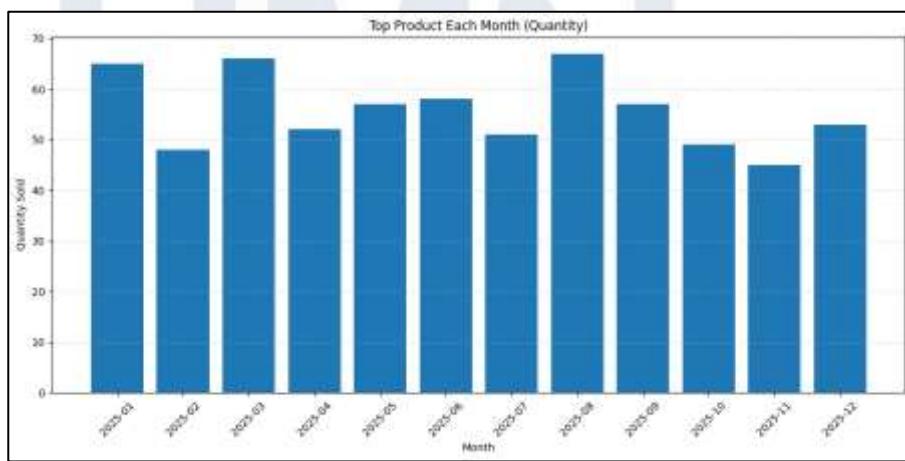
Gambar 3.17 menampilkan distribusi jumlah pembelian berdasarkan wilayah. Data menunjukkan bahwa Medan dan Bogor merupakan wilayah dengan volume pembelian tertinggi, masing-masing mencatat lebih dari 160 pesanan sepanjang tahun 2025. Wilayah lain seperti Depok, Jakarta, Surabaya, Bandung, dan Bekasi menunjukkan volume pembelian yang relatif stabil meskipun ada di bawah dua wilayah utama. Pola ini memberikan gambaran awal

mengenai konsentrasi pasar dan wilayah aktivitas penjualan tertinggi



Gambar 3. 18. Pembelian setiap produk

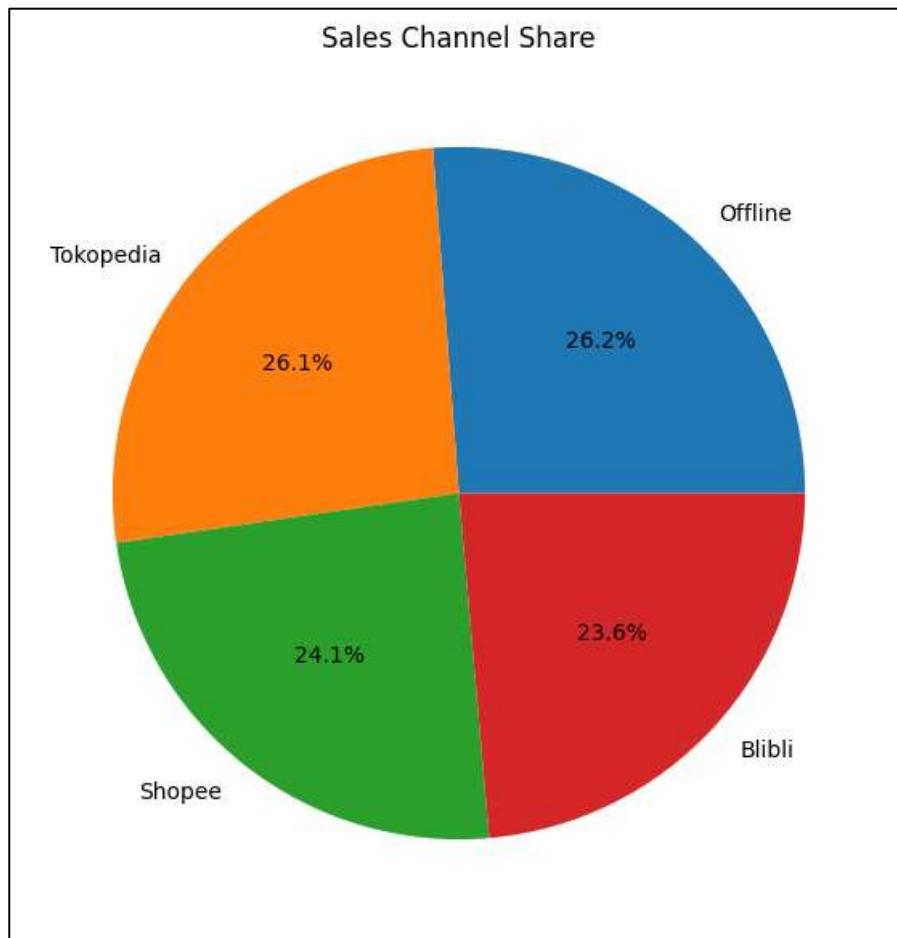
Gambar 3.18 menyajikan sepuluh produk dengan jumlah penjualan tertinggi berdasarkan kuantitas. Hasil visualisasi menunjukkan bahwa varian Prospero dan Realcho mendominasi penjualan, dengan produk *Prospero Realcho Rewind* sebagai produk terlaris, diikuti oleh *Chobayoo CoffeeBall* dan beberapa varian Realcho lainnya. Informasi ini mencerminkan preferensi konsumen terhadap produk tertentu dalam portofolio perusahaan.



Gambar 3. 19. Penjualan tiap bulan

Gambar 3.19 menggambarkan perubahan jumlah penjualan produk terlaris pada setiap bulan sepanjang tahun 2025. Data menunjukkan adanya fluktuasi permintaan, dengan beberapa bulan

seperti Januari, Maret, dan Agustus mencatat volume penjualan yang lebih tinggi dibandingkan bulan lainnya. Pola ini mengindikasikan adanya kecenderungan musiman (*seasonality*) yang perlu diperhatikan dalam perencanaan produksi dan pengelolaan persediaan.



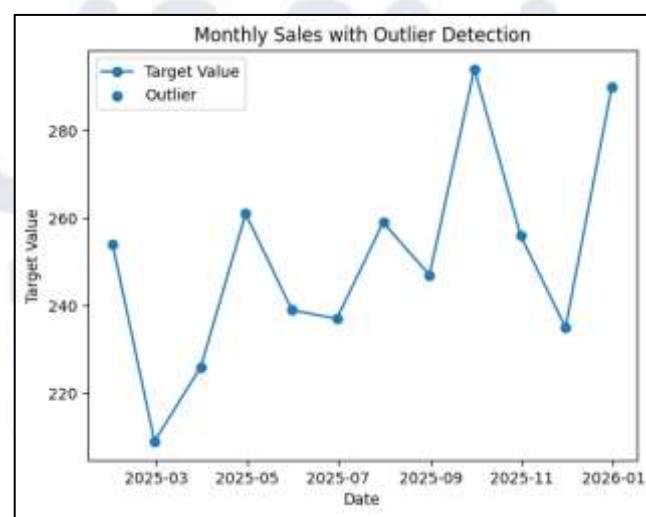
Gambar 3. 20. Piechart penjualan

Gambar 3.20 menunjukkan proporsi penjualan berdasarkan saluran distribusi. Penjualan offline memiliki kontribusi terbesar sebesar 26,2%, diikuti oleh Tokopedia (26,1%), Shopee (24,1%), dan Blibli (23,6%). Distribusi yang relatif seimbang antar saluran menunjukkan bahwa pendekatan *multi-channel* yang diterapkan perusahaan berjalan secara konsisten dan seluruh platform memberikan kontribusi yang signifikan terhadap total penjualan.

3.3.1.5 Market Forecasting

Pada tahap lanjutan analisis pasar, dilakukan peramalan (*forecasting*) tren pasar produk cocoa untuk tahun 2026. Kegiatan ini bertujuan memberikan estimasi awal mengenai arah permintaan pasar di masa mendatang sebagai bahan pendukung perencanaan perusahaan, khususnya dalam aspek produksi, pemasaran, dan distribusi. Proses peramalan dilakukan menggunakan pendekatan *time series* dengan memanfaatkan model Prophet berdasarkan data penjualan bulanan tahun 2025.

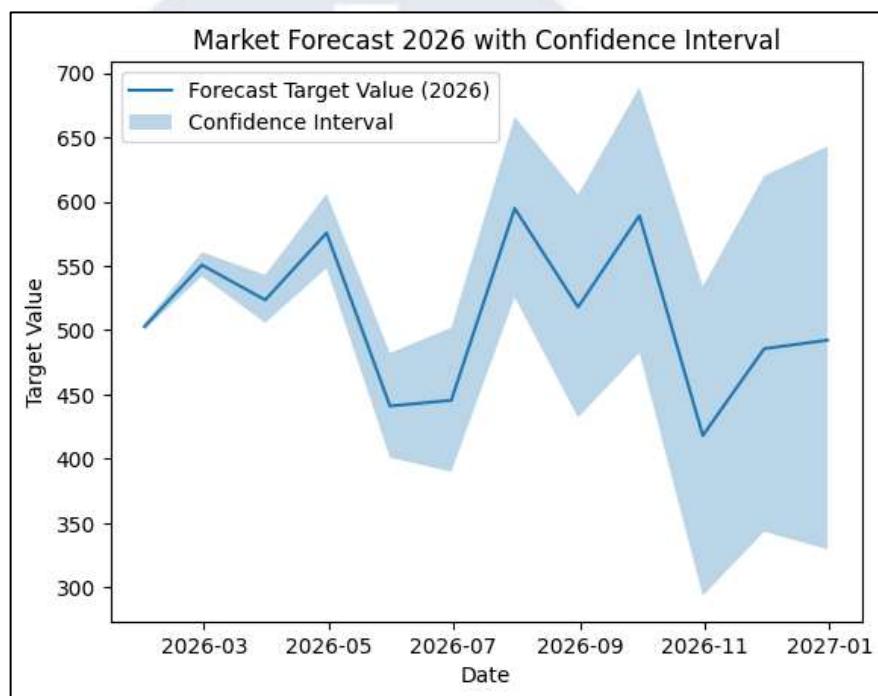
Dalam pemodelan *time series* menggunakan Prophet, data penjualan disusun ke dalam format standar yang terdiri dari dua variabel utama, yaitu *ds* dan *y*. Variabel *ds* merepresentasikan dimensi waktu (*date stamp*), sedangkan variabel *y* merepresentasikan nilai target yang diprediksi, yaitu total penjualan produk cocoa pada setiap periode. Format ini digunakan agar model dapat mempelajari pola tren dan musiman secara optimal.



Gambar 3. 21. Tren Penjualan Bulanan dan Deteksi *Outlier* Tahun 2025

Gambar 3.21 menampilkan tren penjualan bulanan produk cocoa selama tahun 2025 yang disertai dengan proses deteksi *outlier*.

Data menunjukkan adanya fluktuasi penjualan antar bulan, dengan beberapa titik berada di luar rentang normal berdasarkan metode *Interquartile Range* (IQR). Identifikasi *outlier* ini dilakukan untuk memahami karakteristik data historis serta mengantisipasi potensi anomali yang dapat memengaruhi proses peramalan.



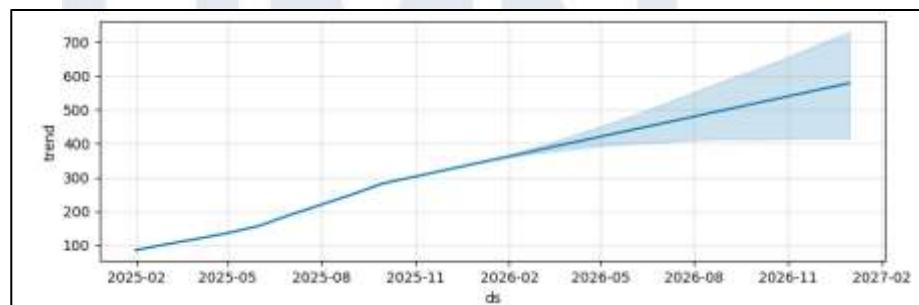
Gambar 3. 22. Hasil Peramalan Tren Pasar Produk Cocoa Tahun 2026 dengan *Confidence Interval*

Gambar 3.22 menyajikan hasil peramalan penjualan produk cocoa untuk tahun 2026 menggunakan model Prophet. Garis utama merepresentasikan nilai estimasi penjualan (*forecast*), sedangkan area bayangan di sekelilingnya menunjukkan *confidence interval* sebagai indikator tingkat ketidakpastian prediksi. Visualisasi ini memberikan gambaran mengenai proyeksi arah permintaan produk cocoa pada periode mendatang.



Gambar 3. 23. Perbandingan Penjualan Aktual Tahun 2025 dan Hasil *Forecast* Tahun 2026

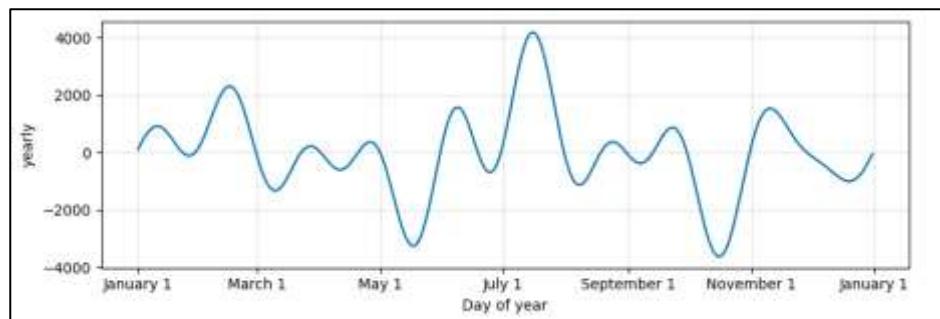
Gambar 3.23 memperlihatkan perbandingan antara data penjualan aktual tahun 2025 dan hasil peramalan penjualan tahun 2026. Grafik ini menunjukkan transisi dari data historis menuju nilai prediksi berdasarkan dimensi waktu (*ds*) dan nilai target penjualan (*y*), sehingga memudahkan pemahaman terhadap perubahan pola penjualan secara keseluruhan.



Gambar 3. 24. Komponen Tren (*Trend*) Model Prophet

Gambar 3.24 menampilkan komponen tren (*trend*) yang dihasilkan oleh model Prophet. Komponen ini merepresentasikan perubahan jangka panjang dari nilai penjualan terhadap waktu tanpa dipengaruhi oleh fluktuasi musiman atau *noise*. Visualisasi tren ini

membantu dalam memahami kecenderungan umum perkembangan penjualan produk cocoa.



Gambar 3. 25. Komponen Musiman (*Seasonality*) Model Prophet

Gambar 3.25 menunjukkan komponen musiman (*seasonality*) yang menggambarkan pola fluktuasi penjualan dalam satu siklus tahunan. Pola ini mengindikasikan adanya periode tertentu di mana permintaan produk cocoa cenderung meningkat maupun menurun. Informasi ini memberikan gambaran mengenai karakteristik musiman pasar yang dapat dipertimbangkan dalam perencanaan operasional perusahaan.

3.3.2 Kendala yang Ditemukan

Selama menjalani kegiatan magang sebagai *Data Scientist Intern* di Divisi *Research and Development* (R&D) PT. Mandala Prima Makmur, terdapat beberapa kendala yang dihadapi dalam proses adaptasi maupun pelaksanaan proyek. Kendala tersebut terutama berkaitan dengan kompleksitas teknis, pengolahan data, serta pemahaman konteks bisnis pada proyek yang dikerjakan.

Kendala awal yang dihadapi adalah keterbatasan pemahaman terhadap istilah teknis yang umum digunakan dalam lingkungan R&D, khususnya yang berkaitan dengan *machine learning*, sistem berbasis AI, dan *Large Language Model* (LLM). Istilah seperti *embedding*, *prompt engineering*, *vector search*,

knowledge retrieval, dan *model fine-tuning* memerlukan waktu adaptasi sebelum dapat dipahami dan diterapkan secara optimal dalam proyek.

Pada proyek *Webscraping X on Cocoa Powder*, tantangan utama berasal dari karakteristik data eksternal yang tidak terstruktur dan mengandung *noise*. Data hasil ekstraksi dari platform X sering kali mencakup duplikasi, entri yang tidak relevan, serta perbedaan format, sehingga memerlukan proses *data cleaning* yang lebih mendalam sebelum dapat digunakan sebagai dasar analisis.

Kendala berikutnya muncul pada aspek pemahaman konteks bisnis perusahaan, terutama saat mengerjakan proyek *Market 2025 Analysis* dan *Market Forecasting*. Data yang dianalisis tidak hanya merepresentasikan nilai numerik, tetapi juga mencerminkan dinamika permintaan pasar, pola penjualan, serta strategi pemasaran perusahaan, sehingga membutuhkan pemahaman lintas fungsi agar hasil analisis tetap relevan.

Dalam pengembangan *Customer ChatBot* dan *Customer ChatBot – LLM*, tantangan teknis muncul akibat keterbatasan pengalaman sebelumnya dalam membangun sistem berbasis *knowledge* menggunakan LLM. Proses integrasi *knowledge base*, implementasi *Retrieval-Augmented Generation* (RAG), serta penyusunan *prompt* yang stabil membutuhkan pemahaman arsitektur model yang lebih mendalam.

Selain itu, ritme kerja tim R&D yang cepat dan berbasis eksperimen menuntut kemampuan manajemen waktu yang baik. Peserta magang perlu menyesuaikan diri dengan pola kerja yang menekankan iterasi cepat, pengujian solusi, serta penyelesaian tugas sesuai tenggat waktu yang ditetapkan

3.3.3 Solusi atas Kendala yang Ditemukan

Selama pelaksanaan kegiatan magang di Divisi Research and Development PT. Mandala Prima Makmur, peserta magang menghadapi beberapa kendala yang berkaitan dengan aspek teknis, pemahaman bisnis, serta pola kerja tim.

1. Keterbatasan pemahaman awal terhadap istilah dan konsep teknis di bidang data science dan kecerdasan buatan yang digunakan dalam berbagai proyek R&D.
2. Kompleksitas dan ketidakteraturan data hasil *web scraping*, seperti adanya data duplikat, format data yang tidak konsisten, serta data yang kurang relevan untuk keperluan analisis.
3. Keterbatasan pemahaman terhadap konteks dan kebutuhan bisnis perusahaan, sehingga hasil analisis pada tahap awal belum sepenuhnya selaras dengan tujuan strategis perusahaan.
4. Keterbatasan pengalaman praktis dalam pengembangan *customer chatbot* berbasis *Large Language Model*, khususnya dalam pemilihan model, pengaturan parameter, serta integrasi *knowledge base* ke dalam sistem chatbot.
5. Tantangan dalam beradaptasi dengan ritme kerja tim Research and Development yang dinamis, di mana beberapa proyek berjalan secara paralel dengan tenggat waktu yang relatif ketat.

Untuk mengatasi berbagai kendala tersebut, diterapkan beberapa solusi yang disesuaikan dengan kebutuhan teknis, pemahaman bisnis, serta pola kerja tim Research and Development.

1. Melakukan pembelajaran mandiri melalui dokumentasi resmi, literatur ilmiah, serta sumber pembelajaran daring seperti dokumentasi open-

source dan video pembelajaran teknis, yang didukung dengan diskusi rutin bersama supervisor dan anggota tim R&D.

2. Menerapkan proses *data preprocessing* secara sistematis yang mencakup pembersihan data, penghapusan duplikasi, standarisasi format, serta verifikasi relevansi data sebelum dilakukan tahap analisis lebih lanjut.
3. Melakukan diskusi lintas divisi dengan tim Marketing dan Sales Department untuk memahami konteks dan tujuan penggunaan data, sehingga hasil analisis pasar dan peramalan yang dihasilkan dapat lebih relevan dalam mendukung pengambilan keputusan perusahaan.
4. Melakukan eksplorasi terhadap berbagai model LLM open-source, eksperimen *prompt engineering* dengan penyesuaian parameter seperti *temperature* dan *max token*, serta mempelajari referensi implementasi *Retrieval-Augmented Generation* sebagai dasar integrasi *knowledge base* perusahaan ke dalam sistem *customer chatbot*.
5. Menyusun perencanaan kerja yang lebih terstruktur melalui penentuan prioritas tugas dan penjadwalan aktivitas secara berkala guna menyesuaikan diri dengan ritme kerja tim Research and Development yang dinamis.