

BAB III

PELAKSANAAN KERJA

3.1 Kedudukan dan Koordinasi

3.1.1 Kedudukan

Kerja magang ini dilaksanakan di perusahaan PT Sampoerna Agro Tbk sebagai *Business Application and Solution* yang ditempatkan di divisi *IT Application* dan berada di bawah tim *IT Business Application*. Pada program ini, mahasiswa berperan sebagai asisten yang membantu tim *IT Business Application* dalam mengelola bisnis terutama dalam bidang data produksi sawit, serta pengembangan chatbot helpdesk SAP. Selama periode magang ini, terdapat pembimbing lapangan yang memberikan bimbingan serta arahan secara langsung untuk pekerjaan serta tugas yang dibutuhkan.

Pada divisi *IT Application* terutama tim *IT Business Application*, berperan penting dalam pengelolaan alur bisnis perusahaan terutama dalam hal proses bisnis yang menggunakan sistem *Enterprise Resource Planning* yaitu SAP (*System Application and Product in Data Processing*). Tim *IT Business Application* membantu dan mengelola penggunaan SAP di perusahaan dengan melakukan training, maintaining, serta arahan kepada divisi-divisi di perusahaan untuk keberlangsungan proses bisnis yang ada.

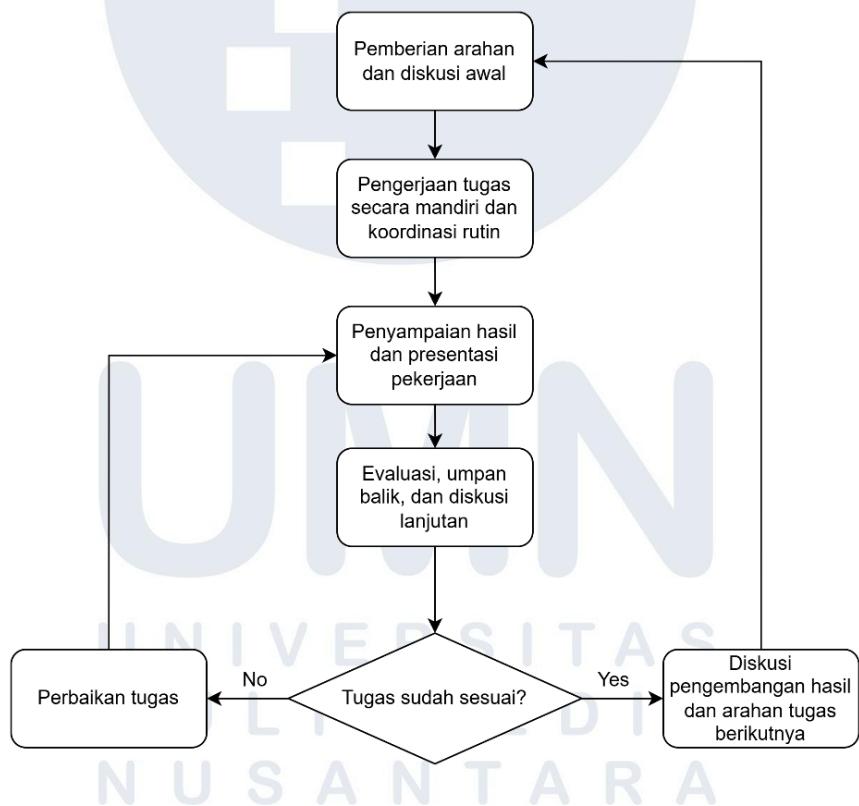
Salah satu proyek utama selama masa magang adalah pembuatan chatbot helpdesk SAP yang dapat membantu pengguna SAP dalam menggunakan modul-modul kompleks serta pencarian dokumen user guide yang sulit diakses dan memakan waktu yang lama. Projek ini juga dibuat atas permasalahan yang terjadi di tim *IT business application* yang seringkali menerima pertanyaan berulang dari user terkait penggunaan SAP. Dengan mengumpulkan data FAQ dan menggabungkan user guide SAP, chatbot helpdesk sap dibuat dengan tujuan untuk membantu pengguna SAP dalam mencari informasi seputar SAP

dan mengurangi beban kerja tim IT business application serta memanfaatkan data yang ada dalam perkembangan internal.

Selain fokus dalam pengerjaan projek chatbot helpdesk SAP, selama masa magang ini juga belajar ETL melalui Talend Studio untuk melihat dan memahami alur datamart perusahaan, serta merancang forecasting produksi kelapa sawit untuk ditambahkan dalam dashboard produksi kelapa sawit yang sudah ada untuk menambahkan insight tambahan.

3.1.2 Koordinasi

Terdapat alur kerja yang harus diikuti selama pelaksanaan program magang di Sampoerna agro yang dijelaskan melalui bagan sebagai berikut:



Gambar 3. 1 Alur Kerja Magang

Pada gambar 3.1 dijelaskan bagan alur kerja yang dilakukan selama program magang sehingga terdapat komunikasi yang lancar serta pemberian arahan dan masukan dari pembimbing lapangan. Penjelasan dari alur kerjanya dijelaskan sebagai berikut:

1. Arahan dan diskusi awal yang akan diberikan terkait tugas atau pekerjaan yang akan dikerjakan termasuk ruang lingkup pekerjaan, dan tujuan dari proyek ini agar pemahaman terhadap proyek dapat diperoleh secara menyeluruh
2. Pelaksanaan tugas akan dikerjakan secara mandiri dengan tetap didukung dengan ruang diskusi jika ditemukan kendala dalam penyelesaiannya dan dapat ditanyakan secara langsung.
3. Setelah tugas selesai dikerjakan akan melakukan presentasi atau diskusi terkait tugas yang dikerjakan berupa hasil analisis, rancangan solusi, atau output yang dihasilkan. Presentasi ini menjadi bentuk evaluasi sementara terhadap hasil dari pekerjaan yang dikerjakan.
4. Pembimbing akan memberikan umpan balik dan diskusi lanjutan mengenai aspek yang perlu di tingkatkan, kesesuaian dengan kebutuhan tim, serta kemungkinan pengembangan lebih lanjut dari pekerjaan yang telah di lakukan.
5. Jika masih ada hal yang perlu direvisi, maka tugas tersebut akan ditinjau ulang dan disesuaikan kembali berdasarkan masukan yang diberikan dan kemudian akan di diskusikan kembali.

Melalui alur kerja dan koordinasi tersebut, mahasiswa dapat memperoleh kesempatan untuk terlibat secara langsung dalam aktivitas operasional IT dan proyek strategis perusahaan. Pengalaman ini dapat memberikan pemahaman yang mendalam mengenai penerapan manajemen data, pengembangan aplikasi internal, dan pemecahan masalah berbasis teknologi di lingkungan agrobisnis. Keterlibatan dalam proyek-proyek nyata, terutama dalam pengembangan chatbot helpdesk dan analisis data produksi sawit, dapat memberikan kontribusi yang signifikan terhadap peningkatan wawasan dan kompetisi profesional.

3.2 Tugas dan Uraian Kerja Magang

Pelaksanaan program kerja magang di PT Sampoerna Agro Tbk. dilaksanakan pada posisi *Business Application and Solution* yang memiliki lingkup tanggung

jawab yang berfokus pada asistensi pengelolaan data operasional, proses bisnis, dan pengembangan solusi berbasis data pada divisi *IT Business Application*. Penugasan yang diberikan mencakup sejumlah kegiatan teknis seperti menambah forecasting produksi sawit, cek duplikasi data mingguan, cek alur *data mart*, dan pembuatan chatbot helpdesk.

Tugas-tugas yang diberikan disesuaikan menurut kebutuhan dan masalah yang ada di perusahaan dalam lingkup agrobisnis kelapa sawit. Pelaksanaan kerja magang ini, memberikan pengalaman kerja yang relevan dan menjadikan kontribusi peserta magang dalam mendukung pekerjaan melalui tugas-tugas yang dikerjakan. Berikut adalah rincian dari tugas dan uraian kerja magang selama di PT. Sampoerna Agro Tbk.:

Tabel 3. 1 Rincian Tugas Kerja Magang

No	Kegiatan	Tanggal	
		Mulai	Selesai
<i>On-Boarding Session</i>			
a	On Boarding dan Pengenalan Tim	25 Agustus 2025	25 Agustus 2025
b	Set up keperluan pekerjaan (<i>Internal Databases, Tableau connect, Accessibility, etc</i>)	25 Agustus 2025	25 Agustus 2025
<i>Data Management</i>			
a	Mengenal dan mempelajari database perusahaan	25 Agustus 2025	04 September 2025
b	Mengenal dan mempelajari Tableau Dashboard produksi perusahaan	25 Agustus 2025	04 September 2025
c	Membuat prediksi produksi kelapa sawit menggunakan time series	04 September 2025	19 September 2025
<i>Data Quality</i>			
a	Membantu mengecek data duplikat dan perbedaan database dan dashboard rutin	08 September 2025	24 Oktober 2025
b	Mempelajari alur kerja datamart perusahaan berdasarkan ETL Talend	22 September 2025	03 Oktober 2025
<i>Chatbot Project</i>			
a	Mengumpulkan requirement dan data yang dibutuhkan	04 Oktober 2025	03 Oktober 2025

No	Kegiatan	Tanggal	
		Mulai	Selesai
b	Mempelajari dan mencoba ekstraksi dokumen FAQ SAP kedalam CSV/dataset untuk pembuatan model chatbot	06 Oktober 2025	15 Oktober 2025
c	Membangun dan mengevaluasi model chatbot menggunakan beberapa model dan bentuk website	16 Oktober 2025	31 Oktober 2025
d	Mencoba mengekstrak data tambahan dari User Guide SAP dan beberapa fitur tambahan	04 November 2025	26 November 2025
e	Membangun chatbot helpdesk Infrastruktur berdasarkan model chatbot SAP yang sebelumnya dibuat	27 November 2025	17 Desember 2025

Tabel 3.1 berisi tugas-tugas yang telah dilaksanakan selama program kerja magang dari 25 Agustus hingga 17 Desember 2025, dimana sudah memenuhi persyaratan 640 jam kerja. Rincian dari tugas-tugas yang dilaksanakan dijelaskan sebagai berikut:

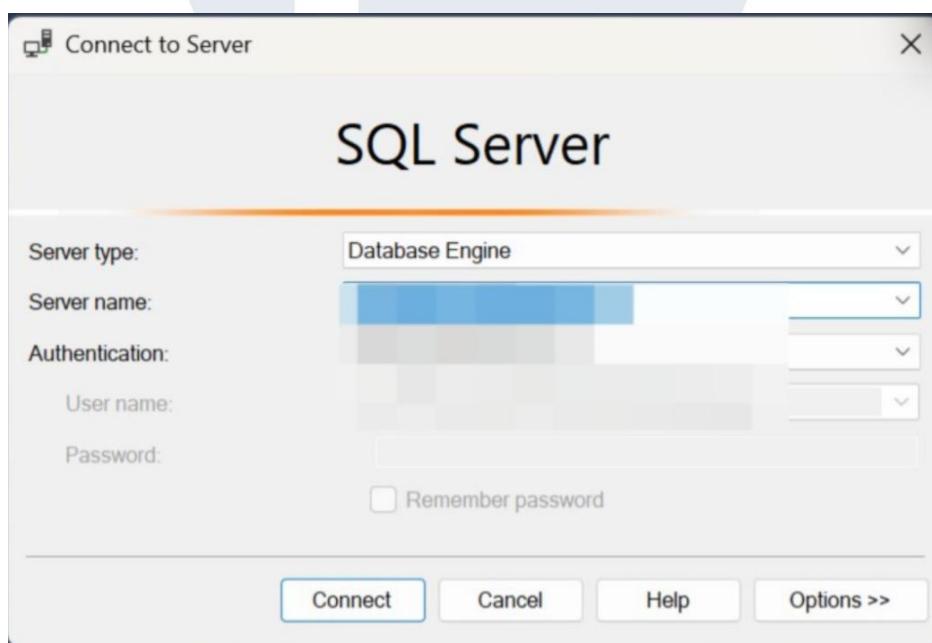
3.2.1 On-Boarding Session

Pada pelaksanaan kerja magang, dilaksanakan pertama kali pada tanggal 25 Agustus 2025 bertempat di gedung Sampoerna Strategic Square lantai 28, yang merupakan kantor pusat dari Sampoerna Agro Jakarta. Sesi *On-Boarding* ini dimulai dengan mengumpulkan mahasiswa peserta magang pada satu ruangan oleh salah satu HR di Sampoerna Agro, yang kemudian diberikan pembekalan terkait kegiatan dan peraturan di tempat magang, posisi yang akan di-isi, dan kemudian diantar ke-meja divisi masing-masing untuk selanjutnya di serahkan ke supervisi dan pembimbing lapangan.

Setelah sesi pembekalan dilakukan oleh tim HR, peserta kemudian diarahkan untuk bergabung ke divisi *Information Communication Technology (ICT)* dengan pembimbing lapangan. Peserta kemudian melakukan perkenalan ke masing-masing anggota divisi ICT baik *ICT Infrastructure*, *IT Cyber Security*, *IT Business Application*, dan *IT Application Development*. Setelah melakukan perkenalan diri kepada masing-masing anggota divisi, peserta

magang kemudian diarahkan untuk menunggu persiapan laptop untuk dipaparkan mengenai perangkat lunak dan sistem yang akan digunakan selama periode magang.

Peserta magang, diarahkan untuk melakukan sambungan koneksi SQL Server untuk mengakses database produksi perusahaan. Proses ini dimulai dengan pemasangan aplikasi *Microsoft SQL Server Management Studio* (SSMS) pada perangkat. Setelah proses instalasi selesai, yang dilanjutkan dengan konfigurasi *server name*, *authentication* menggunakan *Windows Authentication*. Setelah konfigurasi selesai, koneksi ke server database sudah bisa diakses dengan jaringan internal perusahaan saat menyambung ke *Wi-Fi* kantor. Ilustrasi keberhasilan koneksi server database ditampilkan pada Gambar 3.2 berikut:

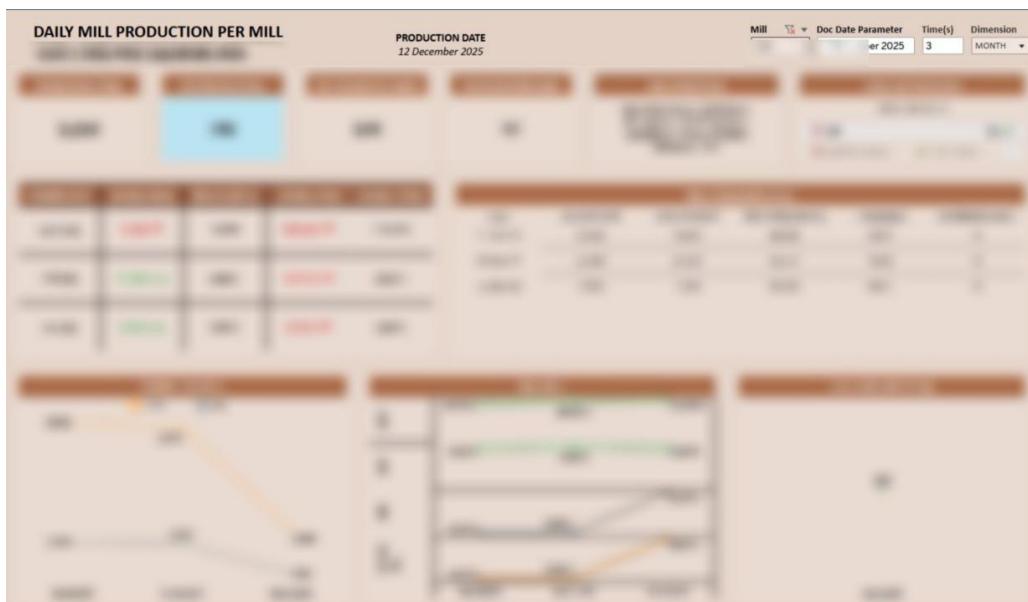


Gambar 3.2 Server SQL

Microsoft SQL Server dipilih menjadi server database utama perusahaan karena kemampuan untuk mengelola data berskala besar, menjaga konsistensi dan integritas data, serta dapat mendukung integrasi dengan berbagai aplikasi pendukung lainnya [6]. Pada saat pelaksanaan magang, *Microsoft SQL Server* digunakan untuk melakukan pengecekan kualitas data, khususnya terkait

duplicasi data, serta membandingkan kesesuaian data antara database dan *dashboard Tableau* produksi yang digunakan oleh perusahaan dalam melihat laporan produksi mingguan.

Selain akses ke *server* database perusahaan, dilakukan juga instalasi dan konfigurasi untuk aplikasi *Tableau Desktop*. Aplikasi ini digunakan sebagai alat Business Intellegent untuk menampilkan data produksi perusahaan dalam bentuk dashboard interaktif [5]. Dashboard produksi kelapa sawit yang sudah tersedia, dipelajari kembali melalui *Tableau Desktop* untuk memahami alur visualisasi datta, indikator kinerja yang digunakan, serta keterkaitan antara data sumber dan tampilan dashboard yang digunakan oleh managemen perusahaan. *Tableau* digunakan karena memiliki kemampuan visualisasi data yang interaktif dan mudah dipahami oleh pengguna non-teknis sehingga sangat sesuai dengan pelaporan manajemen perusahaan. Melalui *Tableau*, data yang kompleks akan disajikan dalam bentuk visual yang lebih informatid dan mendukung proses pengambilan keputusan, salah satu visualisasi dashboard produksi sawit adalah seperti gambar 3.3 berikut:



Gambar 3. 3 Visualisasi Dashboard Produksi

Selain itu, untuk memahami alur pengelolaan data mart perusahaan, dilakukan instalasi *Talend Data Studio* pada perangkat kerja. *Talend Data*

Studio digunakan sebagai *tools ETL* karena mampu mengintegrasikan data dari berbagai sumber secara terstruktur, mendukung proses transformasi data sesuai kebutuhan bisnis, serta mempermudah pemantauan alur data yang berjalan [4]. Dengan memahami penggunaan *Talend Data Studio*, pemahaman mengenai proses *Extract, Transform, dan Load* (ETL) yang diterapkan di perusahaan dapat diperoleh, sekaligus memberikan gambaran mengenai pentingnya kualitas dan konsistensi data dalam sistem pelaporan perusahaan. Proses pemahaman alur data yang berjalan di *Talend Data Studio* digunakan untuk memahami proses ETL yang terjadi, tetapi tidak menggunakan secara langsung untuk proses pengerjaan tugas. Gambar 3.4 berikut adalah gambaran salah satu alur datamart pada *Talend Data Studio*.



Gambar 3.4 Alur Data pada *Talend Data Studio*

Dalam mendukung aktivitas pengembangan aplikasi internal, khususnya pada proyek pengembangan chatbot helpdesk SAP, digunakan *Visual Studio Code* sebagai editor kode utama. *Visual Studio Code* dipilih karena bersifat ringan, fleksibel, serta mendukung berbagai bahasa pemrograman dan ekstensi yang dibutuhkan dalam pengembangan aplikasi berbasis web[7]. Penggunaan *Visual Studio Code* mempermudah proses penulisan, pengujian, dan pemeliharaan kode aplikasi yang dikembangkan selama periode magang. Pada projek ini, *Visual Studio Code* digunakan sebagai editor code untuk *deployment* model chatbot menggunakan *Flask Framework* dan didukung oleh *HTML* serta *CSS* untuk penampilan website chatbot helpdesk. Penggunaan *Visual Studio Code* dapat dilihat pada gambar 3.5 berikut ini.

```

    75     paths = {
    76         "chat": os.path.join(LOG_DIR, CFG["CHAT_LOG"]),
    77         "high": os.path.join(LOG_DIR, CFG["HIGH_CONF_LOG"]),
    78         "mid": os.path.join(LOG_DIR, CFG["AMBIGUOUS_LOG"]),
    79         "low": os.path.join(LOG_DIR, CFG["UNANSWERED_LOG"]),
    80     }
    81
    82     for key, path in paths.items():
    83         if not os.path.exists(path):
    84             with open(path, "w", encoding="utf-8", newline="") as f:
    85                 writer = csv.writer(f)
    86                 if key == "chat":
    87                     writer.writerow(["timestamp", "user_name", "email_user", "query", "module_faq", "topic", "confidence"])
    88                 elif key == "high":
    89                     writer.writerow(["timestamp", "user_name", "email_user", "query", "module_faq", "topic", "confidence"])
    90                 elif key == "mid":
    91                     writer.writerow(["timestamp", "user_name", "email_user", "query", "options"])
    92                 else:
    93                     writer.writerow(["timestamp", "user_name", "email_user", "query", "status", "answer"])
    94
    95     def log_entry(path, row):
    96         try:
    97             with open(path, "a", encoding="utf-8", newline="") as f:
    98                 writer = csv.writer(f)
    99                 writer.writerow(row)
    100
    101

```

Gambar 3. 5 Penggunaan Visual Studio Code

Selain *Visual Studio Code*, *Jupyter Notebook* digunakan sebagai lingkungan pengembangan interaktif untuk melakukan eksplorasi data dan pengolahan dataset baik untuk projek chatbot, maupun untuk pengelolaan data management produksi. *Jupyter Notebook* digunakan karena memungkinkan proses analisis data, pembersihan data, serta pengujian awal model dilakukan secara terstruktur dan terdokumentasi dengan baik [8]. Dalam pelaksanaan kerja magang, *Jupyter Notebook* dimanfaatkan untuk pembuatan *forecasting* produksi kelapa sawit untuk tambahan dashboard produksi, mengolah data FAQ dan dokumen *user guide* SAP, serta untuk melakukan pengujian awal terhadap model chatbot sebelum diimplementasikan ke dalam aplikasi web. Penggunaan *Jupyter Notebook* dapat dilihat pada gambar 3.6 berikut ini.

```
jupyter data prep-infra Last Checkpoint: 17 days ago
File Edit View Run Kernel Settings Help
+ X Markdown Trusted
JupyterLab Python 3 (ipykernel)
-----Data Prep-----
Read data from docs to csv
[2]: pip install tqdm
Collecting tqdm
  Downloading tqdm-4.67.1-py3-none-any.whl.metadata (57 kB)
Requirement already satisfied: colorama in c:\users\internship.ictapps3\appdata\local\programs\python\python314\lib\site-packages (from tqdm) (0.4.6)
  Downloading tqdm-4.67.1-py3-none-any.whl (78 kB)
Installing collected packages: tqdm
  Successfully installed tqdm-4.67.1
Note: you may need to restart the kernel to use updated packages.

[notice] A new release of pip is available: 25.2 -> 25.3
[notice] To update, run: C:\Users\internship.ictapps3\AppData\Local\Programs\Python\Python314\python.exe -m pip install --upgrade pip
[3]: import os
import re
import pandas as pd
from docx import Document
from tqdm import tqdm
import random
```

Gambar 3. 6 Penggunaan Jupyter Notebook

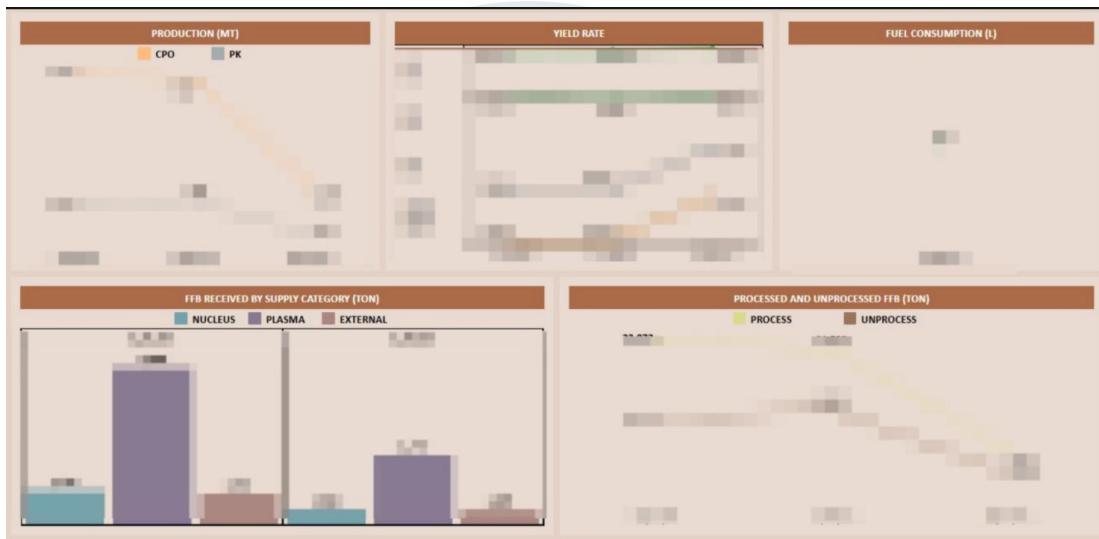
Melalui rangkaian kegiatan *On-Boarding* dan persiapan perangkat kerja tersebut, pemahaman awal mengenai lingkungan kerja, sistem, serta teknologi yang digunakan di PT Sampoerna Agro Tbk dapat diperoleh secara menyeluruh. Tahapan ini menjadi landasan penting bagi mahasiswa untuk dapat berkontribusi secara optimal dalam kegiatan pengelolaan data, analisis, dan pengembangan aplikasi pendukung SAP selama pelaksanaan kerja magang.

3.2.2 Data Management

Pada bagian Data Management, kegiatan kerja magang difokuskan pada pengenalan, pemahaman, dan pemanfaatan data perusahaan yang digunakan dalam mendukung proses bisnis dan pelaporan. Aktivitas yang dilakukan pada tahap ini bertujuan untuk memberikan pemahaman awal mengenai struktur data, alur penyajian informasi, serta pemanfaatan data dalam konteks operasional perusahaan.

Kegiatan awal yang dilakukan adalah mengenal dan mempelajari database perusahaan yang digunakan sebagai sumber utama data produksi. Database perusahaan diakses melalui *Microsoft SQL Server*, dan struktur tabel, relasi antar data, serta jenis data yang tersimpan dipelajari secara bertahap. Pada

tahap ini, pemahaman difokuskan pada data produksi kelapa sawit, seperti data hasil panen, periode waktu, dan atribut pendukung lainnya yang digunakan dalam proses pelaporan. Kegiatan ini dilakukan untuk memberikan gambaran mengenai bagaimana data operasional disimpan, dikelola, dan dimanfaatkan oleh perusahaan.



Gambar 3. 7 Dashboard Produksi Kelapa Sawit

Selain mempelajari database, dilakukan pula kegiatan pengenalan dan pemahaman terhadap dashboard produksi perusahaan yang dibangun menggunakan Tableau seperti pada gambar 3.7. Dashboard ini digunakan oleh manajemen dan tim terkait untuk memantau kinerja produksi kelapa sawit secara periodik. Pada tahap ini, struktur dashboard, jenis visualisasi yang digunakan, serta keterkaitan antara data sumber dan tampilan dashboard dipelajari. Melalui kegiatan ini, pemahaman mengenai bagaimana data mentah diolah menjadi informasi yang informatif dan mudah dipahami dapat diperoleh.

Sebagai bagian dari eksplorasi pemanfaatan data, dilakukan pula pembuatan prediksi produksi kelapa sawit menggunakan metode *time series*. Kegiatan ini bertujuan untuk mengeksplorasi potensi pemanfaatan data historis produksi dalam menghasilkan insight tambahan yang dapat mendukung analisis internal.

Data yang digunakan merupakan data produksi kelapa sawit yang tersimpan pada data mart perusahaan dan diakses melalui Microsoft SQL Server. Data historis produksi diambil pada rentang waktu tahun 2016 hingga 2021 sebagai data pelatihan (*training*), sedangkan data produksi pada periode 2022 hingga 2025 digunakan sebagai data pembanding (*testing*) seperti yang dapat dilihat pada gambar 3.8 berikut. Data produksi kemudian diagregasi dalam skala bulanan untuk menyesuaikan dengan pola musiman produksi kelapa sawit.

```
conn = pyodbc.connect(  
    'DRIVER={  
    'SERVER=S  
    'DATABASE  
    'Trusted  
}  
TRAIN_START = "2010-01-01"  
TRAIN_END   = "2020-12-31"  
FUTURE_YEARS = 2  
EVAL_FREQ   = "MS"  
AGG_METHOD  = "sum"  
LAG_RESID   = 12  
ROLLING_WINDOW = 60  
OUTPUT_CSV   = "hybrid_sarima_xgb_rolling.csv"  
EVAL_CSV     = "hybrid_eval_rolling.csv"
```

Gambar 3. 8 Akses data mart produksi

Model *forecasting* yang digunakan adalah dengan menggunakan metode *hybrid* SARIMA dan XGBoost. Model *Seasonal Arima* (SARIMA) digunakan pada tahapan awal untuk menangkap pola tren dan musiman pada data produksi historis. Parameter yang digunakan, ditentukan melalui proses *grid search* sederhana yang menangkap kombinasi parameter orde non-musiman dan musiman yang diterapkan periode sebesar 12 bulan untuk mempresentasikan pola tahunan produksi yang dapat dilihat pada gambar 3.9 berikut.

```

best_score, best_model, best_params = float("inf"), None, None
for params in param_grid:
    order = (params[0], params[1], params[2])
    seasonal_order = (params[3], params[4], params[5], params[6])
    try:
        model = SARIMAX(train_agg['y'],
                         order=order,
                         seasonal_order=seasonal_order,
                         enforce_stationarity=False,
                         enforce_invertibility=False)
        sarima_fit = model.fit(disp=False)
        if len(test_agg) > 0:
            forecast = sarima_fit.forecast(steps=len(test_agg))
            score = smape(test_agg['y'].values, forecast)
        else:
            score = sarima_fit.aic
        if score < best_score:
            best_score, best_params, best_model = score, (order, seasonal_order), sarima_fit
    except:
        continue

if best_model is None:
    continue

```

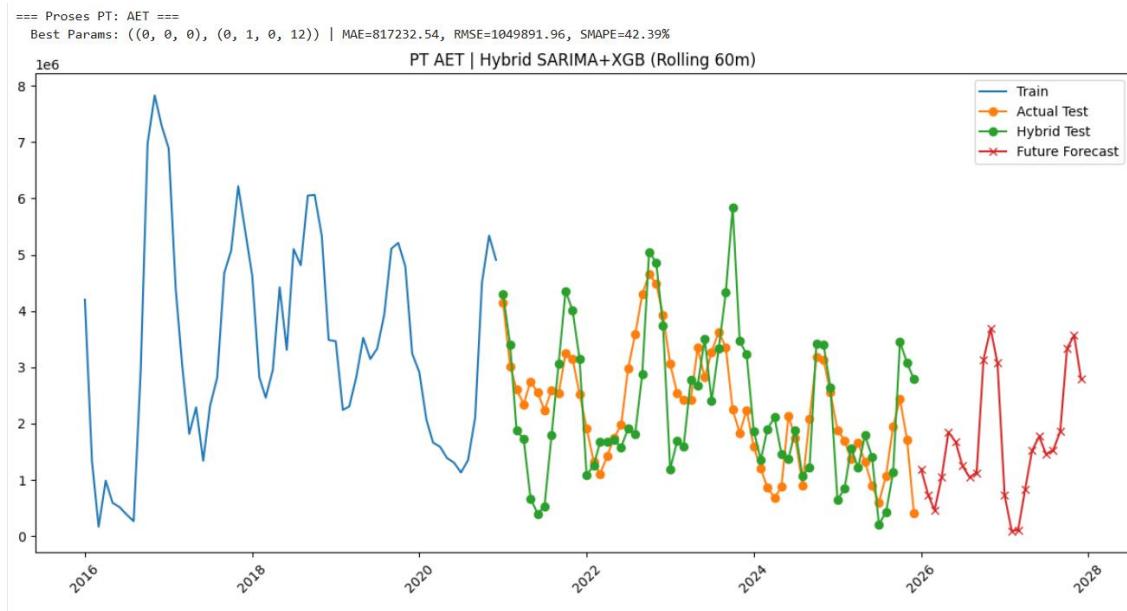
Gambar 3. 9 SARIMA Modelling

Setelah membentuk model SARIMA, kemudian dilakukan perhitungan residual antara nilai aktual dan produksi dengan nilai hasil prediksi SARIMA. Pada perhitungan ini dihitung menggunakan algoritma XGBoost Regression dengan pendekatan lag residual. Tujuan penggunaan algoritma ini adalah untuk menangkap pola non-linear dan fluktulasi data yang tidak sepenuhnya di modelkan oleh SARIMA. Hasil prediksi akhir diperoleh dengan menjumlahkan hasil prediksi SARIMA dan hasil prediksi residual dari XGBoost yang keduanya menghasilkan model hybrid.

Evaluasi performa model dilakukan dengan membandingkan hasil prediksi model hybrid terhadap data aktual pada periode pengujian. Metrik evaluasi yang digunakan meliputi MAE, RMSE, dan SMAPE. *Mean Absolute Error* (MAE), untuk mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual. *Root Mean Square Error* (RMSE), untuk mengukur besarnya kesalahan prediksi dengan penalti lebih besar pada error yang ekstrem. *Symmetric Mean Absolute Percentage Error* (SMAPE), untuk mengukur tingkat kesalahan dalam bentuk persentase yang lebih stabil terhadap skala data.

Hasil forecasting dibagi berdasarkan masing-masing PT yang berada di bawah naungan Sampoerna Agro, dan berikut adalah salah satu hasil

forecasting pada PT. AET menggunakan metode *hybrid SARIMA* dan *XGBoost*.



Gambar 3. 10 Hasil forecasting produksi sawit.

Berdasarkan visualisasi hasil *forecasting* pada Gambar 3.8, dapat dilihat perbandingan antara data historis produksi, data aktual pada periode pengujian, serta hasil prediksi yang dihasilkan oleh model *hybrid SARIMA* dan *XGBoost*. Grafik tersebut menunjukkan bahwa pola produksi kelapa sawit memiliki fluktuasi yang cukup tinggi dari waktu ke waktu, yang dipengaruhi oleh berbagai faktor operasional dan eksternal. Pada beberapa periode, hasil prediksi model mampu mengikuti arah tren data aktual, terutama dalam menangkap kecenderungan kenaikan dan penurunan produksi secara umum. Namun, pada periode tertentu, deviasi antara nilai prediksi dan data aktual masih terlihat cukup signifikan.

Visualisasi ini memberikan insight bahwa meskipun model prediksi belum mampu menghasilkan akurasi yang tinggi secara konsisten, pola musiman dan tren produksi jangka menengah masih dapat teridentifikasi. Informasi ini dapat dimanfaatkan sebagai gambaran awal mengenai potensi pergerakan produksi di masa mendatang, khususnya untuk kebutuhan analisis internal dan

pembelajaran. Selain itu, visualisasi ini juga menunjukkan bahwa variabilitas produksi kelapa sawit relatif tinggi, sehingga diperlukan data tambahan atau pendekatan lanjutan agar hasil prediksi dapat digunakan secara operasional.

Berdasarkan hasil evaluasi pada salah satu entitas perusahaan, yaitu PT AET, diperoleh nilai *Mean Absolute Error* (MAE) sebesar 817.232,54 dan *Root Mean Square Error* (RMSE) sebesar 1.049.891,96. Nilai MAE menunjukkan rata-rata selisih absolut antara hasil prediksi dan data aktual produksi, sedangkan nilai RMSE memberikan gambaran tingkat kesalahan yang lebih sensitif terhadap perbedaan nilai yang besar. Selain itu, diperoleh nilai *Symmetric Mean Absolute Percentage Error* (SMAPE) sebesar 42,39%, yang menunjukkan bahwa secara persentase, selisih antara nilai prediksi dan data aktual masih tergolong cukup tinggi.

Hasil evaluasi pada PT AET tersebut menunjukkan bahwa model prediksi belum mampu menghasilkan tingkat akurasi yang optimal untuk digunakan secara operasional. Namun demikian, hasil ini tidak sepenuhnya merepresentasikan performa model secara keseluruhan. Berdasarkan evaluasi yang dilakukan pada beberapa PT lain di bawah naungan PT Sampoerna Agro Tbk, tingkat akurasi prediksi berada pada kisaran 65% hingga 78%. Perbedaan performa ini dipengaruhi oleh variasi pola produksi, jumlah data historis, serta tingkat fluktuasi produksi pada masing-masing PT.

Dengan mempertimbangkan hasil evaluasi tersebut, forecasting produksi kelapa sawit yang dihasilkan dalam kegiatan kerja magang ini digunakan sebagai bahan eksplorasi dan pembelajaran analisis data, serta sebagai sumber insight tambahan yang bersifat non-operasional. Hasil prediksi ini tidak dijadikan sebagai komponen resmi pada dashboard produksi perusahaan, namun tetap memberikan pemahaman mengenai tantangan dalam melakukan prediksi produksi pada data dengan tingkat variabilitas yang tinggi.

Secara keseluruhan, kegiatan Data Management yang dilakukan selama kerja magang lebih difokuskan pada pemahaman alur data, pemanfaatan data

untuk analisis, serta eksplorasi potensi pengembangan insight berbasis data. Kegiatan ini memberikan pemahaman dasar yang penting mengenai peran data dalam mendukung proses bisnis dan pengambilan keputusan di perusahaan.

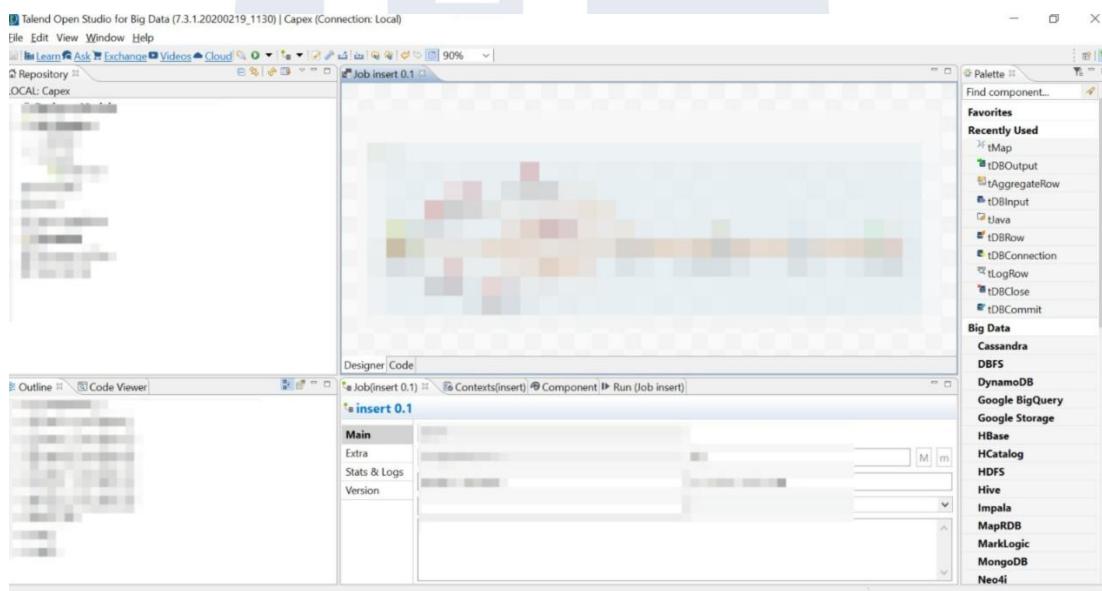
3.2.3 Data Quality

Pada bagian Data Quality, kegiatan kerja magang difokuskan pada upaya menjaga kualitas dan konsistensi data yang digunakan dalam proses pelaporan dan analisis. Kualitas data menjadi aspek penting dalam memastikan bahwa informasi yang dihasilkan dari sistem perusahaan dapat diandalkan dan sesuai dengan kondisi aktual di lapangan.

Salah satu kegiatan utama yang dilakukan adalah membantu proses pengecekan data duplikat dan perbedaan data antara database dan dashboard. Kegiatan ini dilakukan berdasarkan permintaan (*request*) dari pengguna, baik dari pihak kebun maupun dari bagian produksi. Permintaan pengecekan data biasanya muncul ketika ditemukan perbedaan angka atau ketidaksesuaian data pada laporan atau dashboard yang digunakan. Pada kondisi tersebut, mahasiswa magang diarahkan oleh pembimbing lapangan untuk melakukan pengecekan data pada database menggunakan Microsoft SQL Server guna mengidentifikasi kemungkinan adanya data duplikat, kesalahan pencatatan, atau perbedaan periode waktu tertentu.

Proses pengecekan data dilakukan dengan menelusuri data berdasarkan tanggal, lokasi, dan atribut produksi yang relevan untuk menemukan sumber permasalahan. Setelah data duplikat atau perbedaan ditemukan, hasil pengecekan dilaporkan kepada pembimbing lapangan untuk dilakukan tindak lanjut, baik berupa perbaikan data maupun klarifikasi kepada pihak terkait. Kegiatan ini dilakukan secara berkala sesuai dengan kebutuhan dan permintaan dari pengguna, sehingga mahasiswa magang dapat memahami secara langsung bagaimana proses penanganan isu kualitas data dilakukan di lingkungan perusahaan.

Selain kegiatan pengecekan data, dilakukan pula kegiatan mempelajari alur kerja data mart perusahaan berdasarkan proses ETL yang dijalankan menggunakan Talend Data Studio seperti pada gambar 3.9. Pada tahap ini, mahasiswa magang mempelajari alur ekstraksi data dari sumber, proses transformasi data, serta pemuatan data ke dalam data mart yang digunakan untuk pelaporan. Kegiatan ini lebih difokuskan pada pemahaman alur kerja dan struktur proses ETL yang telah berjalan, tanpa melakukan perubahan atau refactoring terhadap alur data mart yang ada. Hal ini disebabkan oleh keterbatasan waktu dan kompleksitas sistem, serta padatnya aktivitas operasional tim yang terlibat dalam pengelolaan data mart.



Gambar 3. 11 Talend Open Studio Workspace

Melalui kegiatan pembelajaran alur ETL tersebut, pemahaman mengenai pentingnya proses integrasi data dan pengendalian kualitas data dalam sistem pelaporan perusahaan dapat diperoleh. Secara keseluruhan, kegiatan Data Quality yang dilakukan selama kerja magang memberikan pengalaman langsung dalam menangani permasalahan kualitas data serta memahami peran data mart dalam mendukung konsistensi dan keandalan informasi di PT Sampoerna Agro Tbk.

3.2.4 Project Chatbot

Pengembangan chatbot ini bertujuan untuk mengembangkan sebuah aplikasi atau sistem pendukung yang ditujukan untuk membantu pengguna dalam mengakses informasi terkait penggunaan sistem SAP dan layanan IT internal perusahaan. Kegiatan kerja magang untuk mengembangkan sebuah aplikasi pendukung ini, dilakukan sebagai upaya peningkatan efisiensi layanan *IT Business Application* serta pemanfaatan data dan dokumentasi yang telah dimiliki oleh perusahaan.

1. Persiapan Data dan Pembuatan Model Chatbot FAQ SAP

Pada tahap persiapan data dan pembuatan model, pengembangan chatbot helpdesk SAP difokuskan pada pemanfaatan data *Frequently Asked Questions* (FAQ) yang telah dikumpulkan oleh tim *IT Business Application* PT Sampoerna Agro Tbk. Dataset yang digunakan terdiri dari 407 pasangan pertanyaan dan jawaban yang mencakup 11 modul utama SAP, seperti *Material Management (MM)*, *Plant Maintenance (PM)*, *Sales and Distribution (SD)*, serta modul pendukung lainnya dengan distribusi modul seperti pada 3.10. Setiap pasangan pertanyaan dan jawaban merepresentasikan satu permasalahan penggunaan SAP yang umum dihadapi oleh pengguna internal perusahaan. Dalam konteks pengujian dan evaluasi, digunakan asumsi bahwa satu query pengguna memiliki satu jawaban benar yang terdapat di dalam basis data FAQ.

FAQ Modul SAP PS (Project System)	76
SAP Material Management (MM)	66
SAP FI - FAQ Lengkap	61
SAP SD & QM	47
SAP Plant Maintenance (PM)	44
SAP CO - FAQ	29
SAP Checkroll (CH)	27
SAP Estate Management (EM)	23
SAP Production Planning (PP)	19
SAP Investment Management (IM)	9

Gambar 3. 12 Distribusi pertanyaan dalam setiap modul

Data FAQ tersebut kemudian dipelajari dan dipersiapkan sebagai basis pengetahuan utama chatbot. Sebelum digunakan dalam tahap pemodelan, data melalui proses preprocessing untuk meningkatkan kualitas dan konsistensi representasi teks. Proses ini meliputi penggabungan informasi modul, topik, pertanyaan, jawaban, serta kode transaksi SAP ke dalam satu kolom konteks (*context_clean*). Selain itu, dilakukan normalisasi teks seperti pengubahan huruf menjadi huruf kecil, penghapusan spasi berlebih, serta pembersihan karakter yang tidak relevan. Tahapan ini bertujuan agar setiap entri data mampu merepresentasikan konteks penuh dari pasangan pertanyaan dan jawaban SAP secara konsisten, sehingga dapat digunakan secara optimal oleh model *retrieval*.

Pengembangan chatbot dilakukan dengan menggunakan pendekatan retrieval-based chatbot, di mana sistem tidak menghasilkan jawaban baru, melainkan mencari dan menampilkan jawaban paling relevan dari data FAQ yang tersedia. Pada tahap ini, digunakan dua pendekatan utama, yaitu TF-IDF dan SBERT. Model TF-IDF digunakan sebagai metode pencarian berbasis leksikal karena efektif dalam menangkap terminologi teknis SAP yang bersifat tetap, seperti kode transaksi (*t-code*), nama modul, dan frasa baku yang sering muncul pada dokumentasi SAP. Sementara itu, model SBERT digunakan sebagai metode pencarian berbasis semantik yang mampu memahami kesamaan makna antar kalimat, sehingga dapat menangani variasi bahasa alami pengguna yang tidak selalu identik dengan teks pada data FAQ.

Kedua model tersebut tidak dilatih sebagai *model supervised*, melainkan digunakan untuk membentuk representasi fitur dari seluruh dataset FAQ. Seluruh korpus digunakan secara penuh tanpa pembagian data latih dan data uji, karena tidak terdapat proses pelatihan parameter terhadap target tertentu. Pendekatan ini dipilih untuk menjaga kesederhanaan arsitektur sistem serta menyesuaikan dengan ukuran dataset FAQ yang relatif terbatas.

Untuk menggabungkan keunggulan dari kedua pendekatan tersebut, dikembangkan model *Hybrid Retrieval* dengan mengombinasikan skor

kemiripan dari TF-IDF dan SBERT menggunakan bobot alpha (α). Penentuan nilai alpha dilakukan melalui manual grid search eksploratif dengan menguji beberapa kombinasi bobot, yaitu (0.5 : 0.5), (0.6 : 0.4), dan (0.7 : 0.3) untuk kontribusi SBERT dan TF-IDF seperti pada tabel 3.2 berikut. Setiap kombinasi diuji secara internal menggunakan metrik Top-k Accuracy untuk menilai kestabilan dan relevansi hasil pencarian.

Tabel 3. 2 Hasil Manual Grid Search

No	Bobot SBERT (α)	Bobot TF-IDF	Top-1 Accuracy	Top-3 Accuracy	Keterangan
1	0.5	0.5	86.42%	91.03%	Performa cukup stabil, namun masih kurang optimal pada query berbasis konteks
2	0.6	0.4	88.70%	92.87%	Performa paling stabil dan optimal, dipilih sebagai konfigurasi final
3	0.7	0.3	87.15%	91.54%	Performa semantik meningkat, tetapi menurun pada pencocokan istilah teknis

Berdasarkan hasil pengujian eksploratif tersebut, kombinasi $\alpha = 0.6$ untuk SBERT dan 0.4 untuk TF-IDF menghasilkan performa yang paling stabil, terutama pada metrik *Top-3 Accuracy*, yang dinilai lebih sesuai dengan kebutuhan chatbot helpdesk. Dalam konteks penggunaan chatbot, tujuan utama sistem adalah membantu pengguna menemukan informasi dengan cepat dan mengurangi pertanyaan berulang kepada tim IT, sehingga penyajian beberapa kandidat jawaban terbaik dianggap lebih realistik dibandingkan hanya mengandalkan satu jawaban teratas. Oleh karena itu, model *Hybrid Retrieval* dengan konfigurasi tersebut dipilih sebagai model utama dalam pengembangan chatbot helpdesk SAP karena mampu menyeimbangkan pencocokan

terminologi teknis dan pemahaman konteks semantik dari pertanyaan pengguna.

2. Evaluasi dan Penerapan Model Chatbot FAQ SAP

Keseluruhan evaluasi model *TF-IDF*, *SBERT*, dan *Hybrid Retrieval* dijabarkan pada tabel 3.3 berikut.

Tabel 3.3 Evaluasi Matrik Performa Model

Model	Top-1 Accuracy	Top-3 Accuracy	Top-5 Accuracy	MRR	Precision@5	Recall@5
TF-IDF	93,12%	95,09%	95,82%	0.941	0.321	0.958
SBERT	51,35%	76,17%	81,57%	0.622	0.213	0.816
Hybrid Retrieval	88,70%	92,87%	93,86%	0.905	0.315	0.939

Keseluruhan evaluasi performa model *TF-IDF*, *SBERT*, dan *Hybrid Retrieval* ditampilkan pada Tabel 3.3. Evaluasi dilakukan menggunakan beberapa metrik, antara lain *Top-1 Accuracy*, *Top-3 Accuracy*, *Top-5 Accuracy*, *Mean Reciprocal Rank (MRR)*, *Precision@5*, dan *Recall@5*, guna memberikan gambaran yang komprehensif terhadap kemampuan masing-masing model dalam konteks sistem chatbot berbasis retrieval.

Berdasarkan hasil evaluasi, model *TF-IDF* menunjukkan performa tertinggi pada metrik *Top-1 Accuracy* sebesar 93,12%. Hal ini mengindikasikan bahwa *TF-IDF* sangat efektif dalam mengembalikan jawaban yang tepat pada peringkat pertama ketika pertanyaan pengguna memiliki kesesuaian terminologi yang tinggi dengan dataset FAQ SAP. Namun, pendekatan berbasis pencocokan kata ini memiliki keterbatasan dalam menangani variasi bahasa alami pengguna, seperti penggunaan sinonim atau struktur kalimat yang berbeda dari data latih.

Model SBERT menunjukkan performa *Top-1 Accuracy* yang lebih rendah dibandingkan TF-IDF, namun memiliki keunggulan dalam memahami kesamaan semantik antar kalimat. Hal ini tercermin dari peningkatan performa pada metrik *Top-3 dan Top-5 Accuracy* dibandingkan Top-1, yang menunjukkan kemampuan model dalam menangkap konteks pertanyaan meskipun tidak identik secara leksikal.

Model *Hybrid Retrieval*, yang menggabungkan pendekatan TF-IDF dan SBERT, dipilih sebagai model utama dalam pengembangan chatbot. Meskipun nilai *Top-1 Accuracy Hybrid* (88,70%) lebih rendah dibandingkan TF-IDF, model ini menunjukkan performa yang lebih seimbang dan konsisten pada metrik Top-k, khususnya *Top-3 Accuracy* sebesar 92,87% serta nilai MRR dan *Recall@5* yang tinggi. Dalam konteks chatbot, metrik Top-k lebih relevan karena sistem menampilkan beberapa kandidat jawaban terbaik kepada pengguna. Oleh karena itu, stabilitas performa pada Top-k menjadi pertimbangan utama dalam pemilihan model, dibandingkan hanya mengandalkan *Top-1 Accuracy* semata. Dengan demikian, *Hybrid Retrieval* dinilai lebih sesuai untuk kebutuhan chatbot helpdesk SAP karena mampu mengakomodasi variasi bahasa pengguna sekaligus mempertahankan tingkat relevansi jawaban yang tinggi.

Chatbot helpdesk SAP ini kemudian diaplikasikan kedalam website internal perusahaan yang dapat digunakan oleh pengguna SAP di Sampoerna Agro dan dapat membantu mengurangi frekuensi pertanyaan berulang yang biasanya diterima oleh tim IT Business Application. Chatbot dirancang menggunakan Flask Framework yang dapat membuat website sederhana untuk diaplikasikan dalam pengembangan. Tampilan awal chatbot dapat dilihat pada gambar 3.11, dimana chatbot SAP Helpdesk ini menyambut pengguna sebelum masuk ke percakapan.



Gambar 3. 13 Tampilan Chatbot FAQ SAP Helpdesk

Pada tahapan ini, chatbot digunakan oleh pengguna SAP di Sampoerna Agro dan didapatkan beberapa hasil evaluasi dari pengguna. Chatbot dirancang untuk menjawab pertanyaan user berdasarkan *similarity check* dan juga dijawab menggunakan *confidence level* yang sebelumnya sudah ditentukan. Sebagai bagian dari evaluasi fungsional chatbot helpdesk SAP yang dikembangkan selama kerja magang, dilakukan uji coba penggunaan (*user testing*) oleh pengguna internal SAP di PT Sampoerna Agro Tbk. Uji coba ini melibatkan 6 orang anggota tim SAP internal yang memiliki pengalaman langsung dalam penggunaan sistem SAP sehari-hari. Pengujian dilakukan selama periode awal implementasi chatbot, dengan total 7 pertanyaan uji yang merepresentasikan pertanyaan umum, teknis, ambigu, serta pertanyaan di luar cakupan basis data FAQ.

Chatbot dirancang untuk memberikan jawaban berdasarkan hasil pencarian similarity menggunakan model hybrid retrieval, dengan mekanisme pengelompokan respons ke dalam tiga tingkat keyakinan (*confidence level*), yaitu *high confidence*, *mid confidence*, dan *low confidence*. Pembagian ini

bertujuan untuk menjaga relevansi jawaban sekaligus meminimalkan risiko penyampaian informasi yang kurang tepat kepada pengguna.

Tabel 3. 4 Hasil dari evaluasi user

Query User	Jawaban Chatbot	Level Confidence	Keterangan
Create order maintenance tco denya apa?	Berikut jawaban yang paling relevan dari modul SAP Plant Maintenance (PM): 📌 Maintenance Order Gunakan T-Code IW34. Isi data seperti Order Type, Notification, Work Center, Equipment, Priority, Dates, Operations, Components, dan WBS Element. Setelah semua data lengkap, klik Save untuk menyimpan.	High Confidence	Jawaban sesuai FAQ
Tcode hapus PR ?	Baik, berikut jawabannya dari modul SAP Material Management (MM): 📌 ME52N – Delete PR T-Code ME52N → input nomor PR → pilih item → Delete Item → Save. Nomor PR tetap, item terhapus.	High Confidence	Jawaban sesuai FAQ
Apa tcode untuk membuat po local?	Berikut jawaban dari modul SAP Material Management (MM): 📌 ME21N – Create PO Local T-Code ME21N → Doc Type ZLOC, Vendor One-Time → isi Header (Org/Group/Company) → item (Material, Qty, Price, Plant) → isi data vendor di header (Name/Address) → Invoice (Tax) → Save.	High Confidence	Jawaban sesuai FAQ
Apa tcode so laboratory?	Ini langkah yang bisa dicoba di modul SAP SD & QM: 📌 VA01 – Create Sales Order of Laboratory	High Confidence	Jawaban sesuai FAQ

	<p>Service Gunakan tcode VA01, pilih Order Type Z005, masukkan data organisasi sales, pelanggan, detail kontrak (jika ada), lalu isi material, kuantitas, UoM, dan parameter uji di tab Texts. Setelah data lengkap, simpan dokumen.</p>		
Create sales order tcodenya apa?	<p>Pertanyaan Anda masih terlalu umum. Apakah maksud Anda salah satu berikut?</p> <p>VA01 - Create Sales Order Seedling (SAP SD & QM) VA01 - Create Sales Order (Umum) (SAP SD & QM) VA01 - Create Sales Order Layanan Agronomi (SAP SD & QM)</p> <p>Silakan pilih angka 1/2/3.</p>	Mid Confidence	Memberikan saran topik yang mirip
Vra order confirm?	<p>Pertanyaan Anda masih terlalu umum. Apakah maksud Anda salah satu berikut?</p> <p>1. SAP Plant Maintenance (PM) (VRA Order (Vehicle/Equipment Related Activity)) 2. SAP Plant Maintenance (PM) (Confirmation (Workshop & VRA))</p> <p>Silakan pilih angka 1/2/3.</p>	Mid Confidence	Memberikan saran topik yang mirip
Apa itu sgro?	<p>✗ Maaf, saya belum menemukan jawaban relevan.</p> <p>Laporkan ke PIC</p>	Low Confidence	Tidak menemukan jawaban relevan dari data FAQ, diarahkan ke PIC SAP

Berdasarkan hasil uji coba yang dirangkum pada Tabel 3.4, sebanyak 4 dari 7 pertanyaan ($\pm 57\%$) diklasifikasikan ke dalam kategori *high confidence*. Pada

kategori ini, chatbot mampu memberikan jawaban yang sesuai dengan data FAQ SAP yang tersedia, khususnya untuk pertanyaan yang mengandung terminologi teknis SAP yang jelas, seperti *t-code* dan modul tertentu. Seluruh jawaban pada kategori high confidence dinilai relevan oleh pengguna karena memiliki kesesuaian langsung dengan kebutuhan informasi yang dicari.

Selanjutnya, 2 dari 7 pertanyaan ($\pm 29\%$) berada pada kategori *mid confidence*. Pertanyaan pada kategori ini bersifat lebih umum atau ambigu, sehingga chatbot tidak langsung memberikan satu jawaban tunggal, melainkan menyajikan beberapa opsi topik atau konteks yang relevan untuk dipilih oleh pengguna. Pendekatan ini dinilai cukup efektif dalam menjaga pengalaman interaktif pengguna sekaligus menghindari pemberian jawaban yang tidak tepat.

Adapun 1 dari 7 pertanyaan ($\pm 14\%$) termasuk dalam kategori *low confidence*, yaitu pertanyaan yang tidak memiliki padanan yang jelas dalam basis data FAQ. Pada kondisi ini, chatbot secara otomatis menampilkan opsi untuk melaporkan pertanyaan tersebut kepada PIC SAP melalui fitur pelaporan. Mekanisme ini berfungsi sebagai bentuk *feedback loop* agar tim SAP dapat meninjau dan menambahkan pertanyaan baru ke dalam basis data di masa mendatang.

Secara keseluruhan, hasil user testing menunjukkan bahwa chatbot helpdesk SAP mampu memberikan respon yang relevan pada sebagian besar skenario penggunaan, terutama untuk pertanyaan teknis yang sering muncul. Selain itu, pembagian tingkat confidence membantu menjaga kualitas jawaban dan memberikan mekanisme eskalasi yang jelas ketika chatbot belum mampu menjawab pertanyaan pengguna.

Riwayat Percakapan Chatbot SAP

Nama: chyntia
Email: [REDACTED]
Waktu: [REDACTED]

Peran	Pesan
User	chyntia
Bot	Halo chyntia! Boleh tahu alamat email kamu agar saya bisa kirim hasil percakapan nanti?
User	[REDACTED]
Bot	Terima kasih chyntia! Silakan ketik pertanyaanmu tentang SAP 😊
User	[REDACTED]

Gambar 3. 14 Hasil pengiriman riwayat percakapan

Pada gambar 3.12 dapat terlihat fitur pengiriman riwayat chatbot kepada user melalui notification email dalam server kantor. Pada pengiriman riwayat ini dapat digunakan sebagai arsip pribadi user yang bisa dilihat pada lain waktu. Selain itu, chatbot juga menyimpan hasil percakapan user kedalam log chat dalam bentuk CSV. Data dari riwayat percakapan chatbot ini, kemudian digunakan sebagai data untuk menganalisis penggunaan SAP di perusahaan, membantu analisis kekurangan hasil training, serta dapat menjadi data untuk pengembangan berikutnya.

3. Pengembangan Lanjutan Model Chatbot FAQ SAP

Seiring dengan berjalannya pengembangan chatbot ini, dilakukan pengembangan lanjutan dengan menambahkan sumber data tambahan berupa dokumen *user guide* SAP. Dokumen *user guide* tersebut sebelumnya bersifat statis dan tersebar dalam berbagai file, sehingga relatif sulit diakses oleh pengguna ketika dibutuhkan. Mengangkat permasalahan ini, diajukan solusi untuk menggabungkan *data knowledge* dari FAQ SAP yang sudah ada dengan dokumen *user guide* SAP sehingga dapat meningkatkan hasil jawaban yang diberikan oleh chatbot.

Pada tahap ini, dilakukan proses ekstraksi dan parsing data dari dokumen *user guide* untuk mengambil informasi tekstual serta menyesuaikannya dengan

gambar langkah-langkah penggunaan (*step by step*) yang terdapat di dalam dokumen. Keseluruhan dokumen *user guide* SAP dibaca dan ekstrak dengan bantuan *Python Library* berupa *docx document* untuk ekstraksi data teks dari dokumen, dan *library PIL* untuk mengambil gambar. Dokumen *user guide* SAP, terdiri dari beberapa bagian yang dapat diambil sebagai data yang dapat digunakan sebagai *knowledge* baru diantaranya adalah tujuan, latar belakang, prasyarat, jalur menu, kode transaksi, petunjuk tambahan, prosedur, dan hasil. Data teks dokumen ini dibaca berdasarkan *section key* yang sudah di definisikan sebelumnya seperti pada gambar 3.13 berikut ini.

```
import os
import re
import json
from io import BytesIO
from docx import Document
from docx.oxml.ns import qn
from PIL import Image
import pandas as pd
from tqdm import tqdm

# ----- CONFIG -----
BASE_DIR = r"D:\FAQ\SAP"
OUT_IMG_DIR = r"D:\FAQ\SAP\Extracted_Images_Final2"
OUT_CSV = "userguides_with_procedures_images.csv"
MIN_IMG_WIDTH = 80
MIN_IMG_HEIGHT = 30
# ----- 

SECTION_KEYS = {
    "purpose": ["Tujuan", "Purpose"],
    "trigger": ["Latar Belakang", "Trigger", "Background"],
    "prerequisite": ["Prasyarat", "Prerequisite"],
    "menu_path": ["Jalur Menu", "Menu Path"],
    "transaction_code": ["Kode Transaksi", "Transaction Code"],
    "petunjuk_tambahan": ["Petunjuk Tambahan", "Additional Notes", "Helpful hints"],
    "result": ["Hasil", "Result"]
}
```

Gambar 3. 15 Parsing Teks dan Image Dokumen User Guide

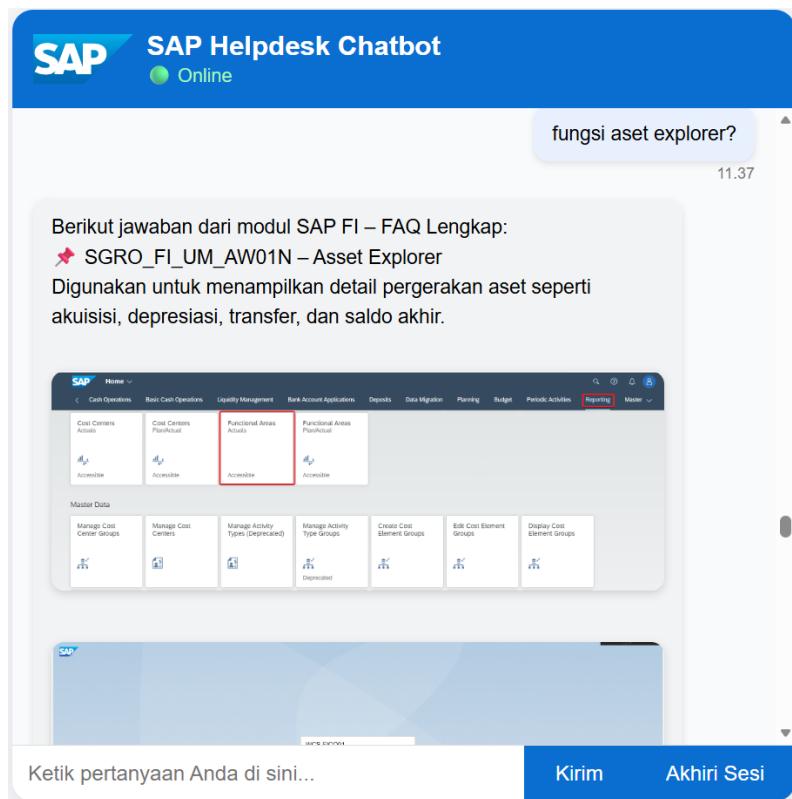
Dokumen *user guide* SAP di Sampoerna Agro terdiri dari total 618 dokumen, yang dibagi berdasarkan modul SAP masing-masing proses bisnis. Dari hasil ekstraksi teks dokumen ini didapatkan dataset berisi data teks dokumen, beserta kolom prosedur dan image dalam bentuk nested json untuk mempermudah pengambilan data image. Detail dari isi dataset ini ditampilkan pada gambar 3.14 berikut ini.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 618 entries, 0 to 617
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   file_path        618 non-null    object  
 1   file_name        618 non-null    object  
 2   purpose          615 non-null    object  
 3   trigger          610 non-null    object  
 4   prerequisite     380 non-null    object  
 5   menu_path        573 non-null    object  
 6   transaction_code 615 non-null    object  
 7   petunjuk_tambahan 610 non-null    object  
 8   result           489 non-null    object  
 9   procedure_count  618 non-null    int64  
 10  images_total     618 non-null    int64  
 11  tables           618 non-null    object  
 12  procedures_json  618 non-null    object  
 13  image_dict       618 non-null    object  
dtypes: int64(2), object(12)
memory usage: 67.7+ KB
```

Gambar 3. 16 Dataset hasil parsing teks dan image dokumen user guide

Dengan penambahan data ini, chatbot tidak hanya mampu memberikan jawaban berbasis teks, tetapi juga dapat menampilkan panduan visual berupa gambar yang relevan dengan pertanyaan pengguna. Pengembangan ini bertujuan untuk meningkatkan pemahaman pengguna SAP, khususnya dalam mengikuti prosedur penggunaan sistem secara lebih jelas dan terstruktur seperti pada gambar 3.15 berikut.





Gambar 3. 17 Tampilan Chatbot dengan Gambar

Selain pengembangan sumber data, dilakukan pula eksplorasi pengembangan model chatbot melalui proses *fine-tuning* terhadap model yang digunakan. *Fine-tuning* dilakukan dengan memanfaatkan log percakapan (*chat log*) pengguna yang dikumpulkan selama kurang lebih satu bulan penggunaan chatbot sebelumnya. Data log tersebut digunakan untuk menyesuaikan model agar lebih relevan dengan pola pertanyaan dan bahasa yang digunakan oleh pengguna internal perusahaan.

```
num_epochs = 2
warmup_steps = int(len(train_dataloader) * num_epochs * 0.1)

model_save_path = "models/sbert_finetuned_sap"
model.fit(
    train_objectives=[(train_dataloader, train_loss)],
    epochs=num_epochs,
    warmup_steps=warmup_steps,
    output_path=model_save_path,
    show_progress_bar=True
)
```

Gambar 3. 18 Train Finetuned SAP

Pada pengembangan *finetuning* ini, dilakukan pendekatan training data log pertanyaan dan jawaban yang sebelumnya sudah di rapihkan dengan menggunakan model SBERT dasar. Model "*paraphrase-multilingual-MiniLM-L12-v2*", digunakan sebagai dasar untuk pengaturan fine tuning ini agar tetap memahami multi bahasa yang ada berdasarkan model dasar.

Berdasarkan hasil pengujian, peningkatan performa model yang dihasilkan dari proses fine-tuning belum menunjukkan perbedaan yang signifikan dibandingkan dengan model sebelumnya. Meskipun demikian, kegiatan ini memberikan pemahaman tambahan mengenai potensi pemanfaatan data interaksi pengguna untuk pengembangan model chatbot di masa mendatang.

4. Pengembangan Model Chatbot IT Infrastructure Helpdesk

Sebagai pengembangan lanjutan, konsep chatbot helpdesk yang telah dibangun juga diterapkan pada kebutuhan layanan IT Infrastruktur. Pada tahap ini, dilakukan pengembangan chatbot helpdesk IT Infrastruktur dengan memanfaatkan data Frequently Asked Questions (FAQ) yang disediakan oleh tim ICT Infrastructure. Data FAQ tersebut mencakup berbagai permasalahan umum yang sering dihadapi oleh pengguna dalam aktivitas kerja sehari-hari, seperti kendala jaringan Wi-Fi, permasalahan printer, akses perangkat, serta isu teknis infrastruktur lainnya.

Dokumen FAQ IT Infrastruktur yang digunakan berisi 27 pasang pertanyaan dan jawaban yang dikumpulkan oleh tim ICT Infrastructure berdasarkan pertanyaan yang paling sering diajukan oleh karyawan. Dokumen ini disusun dalam format teks dengan pola pertanyaan dan jawaban (Q and A). Untuk dapat dimanfaatkan sebagai basis pengetahuan chatbot, dilakukan proses ekstraksi data dari dokumen tersebut. Proses ekstraksi dilakukan dengan memanfaatkan library Python Docx untuk membaca dokumen dan mengambil isi teks berdasarkan pola pertanyaan dan jawaban yang teridentifikasi. Hasil ekstraksi kemudian disusun ke dalam bentuk dataset terstruktur yang berisi pasangan pertanyaan dan jawaban, sebagaimana ditampilkan pada Gambar 3.16.

```
df_clean.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27 entries, 0 to 26
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   topic       27 non-null    object  
 1   question    27 non-null    object  
 2   answer      27 non-null    object  
 3   context     27 non-null    object  
 4   question_clean  27 non-null  object  
 5   answer_clean 27 non-null    object  
 6   context_clean 27 non-null    object  
dtypes: object(7)
memory usage: 1.6+ KB
```

Gambar 3. 19 Dataset FAQ IT Infrastruktur

Setelah proses ekstraksi data dasar selesai dilakukan, jumlah data yang tersedia dinilai masih terbatas untuk mendukung variasi pertanyaan yang mungkin diajukan oleh pengguna. Dalam praktiknya, pengguna sering kali menyampaikan pertanyaan dengan variasi bahasa yang berbeda, meskipun memiliki maksud yang sama. Oleh karena itu, dilakukan proses augmentasi data untuk memperkaya variasi pertanyaan tanpa mengubah makna jawaban yang tersedia. Augmentasi data dilakukan dengan tujuan untuk meningkatkan kemampuan chatbot dalam memahami berbagai bentuk pertanyaan yang memiliki konteks serupa.

```

SYNONYM = {
    "printer": ["mesin printer", "alat printer", "device printer"],
    "wifi": ["wi-fi", "jaringan wifi", "koneksi wireless"],
    "scan": ["pemindaian", "scan dokumen"],
}

TEMPLATE_QUESTIONS = [
    "apa solusi jika {q}?", 
    "Bagaimana cara mengatasinya {q}?", 
    "Kenapa bisa terjadi {q}?", 
    "apa penyebab {q}?", 
]

PREFIX = [
    "bagaimana ", 
    "apa yang harus dilakukan jika ", 
    "apa penyebab ", 
]

def generate_augmented(df):
    rows = []

    for _, row in tqdm(df.iterrows(), total=len(df)):
        topic = row["topic"]
        q = row["question"]
        qc = row["question_clean"]
        ans = row["answer"]

        for prefix in PREFIX:
            for template in TEMPLATE_QUESTIONS:
                rows.append({
                    "topic": topic,
                    "question": f"{prefix}{template} {q}",
                    "answer": ans,
                    "source": "original"
                })

```

Gambar 3. 20 Augmentasi FAQ IT Infra

Proses augmentasi dilakukan dengan beberapa pendekatan seperti pada gambar 3.17, antara lain dengan menambahkan variasi sinonim pada kata-kata tertentu yang sering digunakan dalam konteks permasalahan IT Infrastruktur, seperti istilah jaringan, printer, dan perangkat keras. Selain itu, dilakukan pula penambahan variasi kalimat dengan menggunakan awalan (prefix) dan template pertanyaan umum, seperti pertanyaan yang diawali dengan “bagaimana cara”, “apa penyebab”, atau “apa yang harus dilakukan jika”. Dengan pendekatan ini, satu pertanyaan dasar dapat dikembangkan menjadi beberapa variasi pertanyaan yang tetap mengarah pada jawaban yang sama.

	topic	question	answer	source	context	question_clean	context_clean
0	Printer	Driver printer sudah ada tapi saya print tidak...	Cek jaringan Wi-Fi apa yang digunakan, jika ti...	original	Pertanyaan: Driver printer sudah ada tapi saya print tidak...	driver printer sudah ada tapi saya print tidak...	pertanyaan: driver printer sudah ada tapi saya...
1	Printer	driver mesin printer sudah ada tapi saya print...	Cek jaringan Wi-Fi apa yang digunakan, jika ti...	synonym	Pertanyaan: driver mesin printer sudah ada tapi...	driver mesin printer sudah ada tapi saya print...	pertanyaan: driver mesin printer sudah ada tap...
2	Printer	driver alat printer sudah ada tapi saya print ...	Cek jaringan Wi-Fi apa yang digunakan, jika ti...	synonym	Pertanyaan: driver alat printer sudah ada tapi...	driver alat printer sudah ada tapi saya print ...	pertanyaan: driver alat printer sudah ada tapi...
3	Printer	driver device printer sudah ada tapi saya prin...	Cek jaringan Wi-Fi apa yang digunakan, jika ti...	synonym	Pertanyaan: driver device printer sudah ada ta...	driver device printer sudah ada tapi saya prin...	pertanyaan: driver device printer sudah ada ta...
4	Printer	bagaimana driver printer sudah ada tapi saya p...	Cek jaringan Wi-Fi apa yang digunakan, jika ti...	prefix	Pertanyaan: bagaimana driver printer sudah ada...	bagaimana driver printer sudah ada tapi saya p...	pertanyaan: bagaimana driver printer sudah ada...

Gambar 3. 21 Dataset FAQ IT Infra

Hasil dari proses augmentasi ini adalah dataset FAQ IT Infrastruktur yang lebih beragam dan representatif terhadap cara pengguna menyampaikan permasalahan mereka. Dataset hasil augmentasi seperti pada gambar 3.18, kemudian digunakan sebagai basis pengetahuan chatbot helpdesk IT Infrastruktur. Data ini kemudian diterapkan kepada model TF-IDF, SBERT, dan *Hybrid Retrieval* chatbot sama seperti pembuatan untuk chatbot SAP.



Gambar 3. 22 Tampilan Chatbot Helpdesk IT Infra

Konsep dan alur chatbot IT Infrastruktur dirancang serupa dengan chatbot helpdesk SAP, namun dengan basis pengetahuan yang disesuaikan dengan kebutuhan dan konteks layanan infrastruktur IT. Chatbot IT Infrastruktur ini juga ditampilkan dengan bantuan *framework Flask Python* pada website internal perusahaan yang ditunjukkan pada gambar 3.19. Chatbot ini dirancang untuk membantu karyawan internal yang membutuhkan bantuan seputar IT Infrastruktur dengan cepat. Pengembangan model chatbot ini menunjukkan bahwa chatbot yang dibangun memiliki potensi untuk dikembangkan lebih lanjut dan diterapkan pada berbagai kebutuhan layanan IT internal perusahaan.

Secara keseluruhan, pengembangan proyek chatbot helpdesk selama kerja magang dilakukan secara bertahap dan berfokus pada pemanfaatan data yang tersedia, pengembangan fitur pendukung, serta eksplorasi peningkatan kualitas layanan. Meskipun tidak seluruh pengembangan diarahkan pada implementasi teknis yang mendalam seperti pada penelitian skripsi, kegiatan ini memberikan kontribusi nyata dalam bentuk pengembangan aplikasi pendukung SAP dan layanan IT internal. Proyek chatbot helpdesk ini diharapkan dapat membantu meningkatkan efisiensi layanan IT, mempercepat akses informasi bagi pengguna, serta mengurangi beban kerja tim IT Business Application dan ICT Infrastructure.

3.3 Kendala yang Ditemukan

Pelaksanaan kerja magang pada Sampoerna Agro sebagai Business Application and Solution Intern berjalan secara lancar dalam pelaksanaannya, tetapi terdapat beberapa kendala atau kesulitan yang ditemukan sehingga menjadikan hambatan dalam pengerjaan proyek, diantaranya yaitu:

1. Kegiatan magang dilaksanakan secara WFO (*Work From Office*) dengan jam kerja dari 08.30 WIB hingga 17.30 WIB, sehingga menimbulkan kendala seperti manajemen waktu yang ketat untuk tiba di kantor setiap hari dengan jarak yang jauh, serta manajemen waktu untuk mengerjakan laporan baik untuk laporan magang maupun laporan skripsi.
2. Keterbatasan informasi dan pengalaman dalam informasi bidang agrobisnis terutama bagian kelapa sawit sehingga kesulitan dalam memahami data produksi untuk melakukan analisis lebih lanjut.
3. Perbedaan data produksi sawit setiap tahun menjadikan sulit untuk melakukan analisis forecasting disertai kurangnya data pendukung seperti curah hujan dan faktor-faktor lain sehingga kesulitan untuk mendapatkan hasil forecasting yang sesuai dengan yang diinginkan.
4. Kesulitan untuk memahami serta mencari cara untuk mengekstrak data knowlegde base FAQ dan User Guide sebagai data utama yang digunakan

dalam pengembangan sistem chatbot helpdesk SAP sehingga sering menimbulkan revisi ulang dalam proses pengambilan datanya.

5. Spesifikasi laptop yang diberikan perusahaan untuk bekerja kurang sesuai dengan kebutuhan dan seringkali menyebabkan kinerja yang terhambat karena performa laptop kurang membantu.

3.4 Solusi atas Kendala yang Ditemukan

Proses kerja magang ini dapat diselesaikan dengan baik setelah mengidentifikasi masalah dan kendala yang muncul selama proses magang sebagai Business Application and Solution Intern di Sampoerna Agro, dengan rincian solusi sebagai berikut:

1. Menggunakan *time table* sebagai acuan dalam kegiatan sehari-hari dan menetapkan deadline pribadi dalam penyusunan laporan magang dan skripsi agar tidak kewalahan dalam manajemen waktu.
2. Inisiatif menggali informasi baik informasi dari internet, dokumen user guide perusahaan, serta bertanya secara langsung kepada pembimbing dan senior yang ada di Tim IT Business Application untuk lebih memahami lingkup kerja seputar agrobisnis.
3. Melakukan diskusi lebih lanjut dengan pembimbing dan mendapatkan solusi untuk mencoba forecasting dengan data produksi dari tahun-tahun sebelumnya sehingga time series yang didapatkan hasil akurasi sebesar 70-80% per pabrik.
4. Berdiskusi secara langsung dengan pembimbing lapangan dan melakukan percobaan beberapa teknik dan alur ekstraksi data hingga mendapatkan hasil yang memuaskan.
5. Menghindari penggunaan aplikasi atau hal yang tidak diperlukan sehingga tidak membebankan kinerja laptop serta melakukan back up rutin terhadap data pekerjaan terutama untuk project pengembangan chatbot yang di upload kedalam repository github agar lebih aman.