

BAB III

PELAKSANAAN KERJA MAGANG

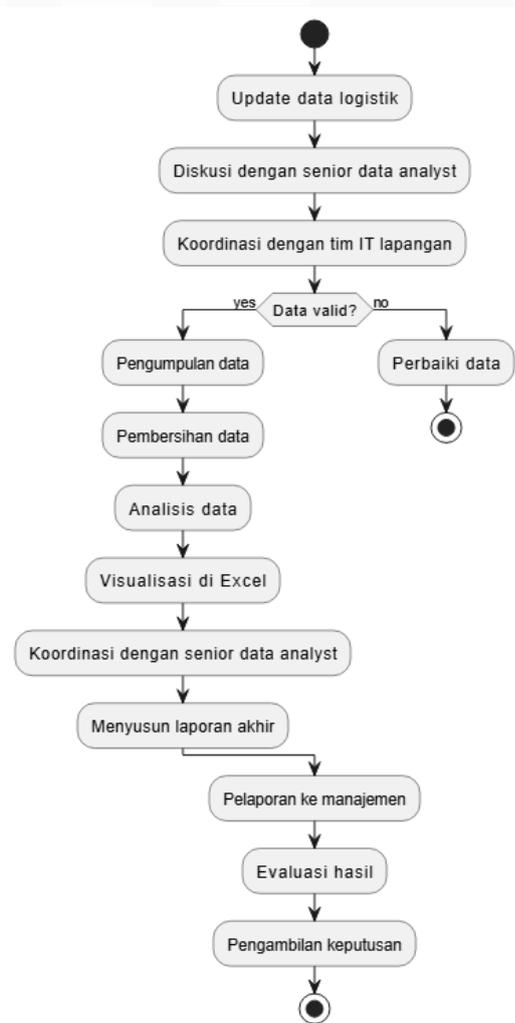
3.1 Kedudukan dan Koordinasi

Selama magang di PT. Eka Satya Puspita, kedudukan saya adalah sebagai *Data Analyst Intern* yang bekerja langsung di bawah bimbingan *Senior Data Analyst* dan *IT Manager*. Sebagai *Data Analyst Intern*, saya berperan dalam membantu pengolahan dan analisis data operasional logistik udara perusahaan. Saya bertanggung jawab untuk mengumpulkan data dari berbagai cabang perusahaan dan memastikan kualitas data yang dikumpulkan sudah sesuai dengan standar yang ditetapkan. Dalam struktur organisasi, saya berada di bawah *Senior Data Analyst*, yang bertanggung jawab untuk memastikan kualitas dan akurasi analisis yang saya lakukan. *IT Manager* bertugas memastikan kelancaran sistem dan infrastruktur teknologi yang mendukung pengelolaan data, sehingga saya dapat bekerja dengan efektif dan efisien.

Koordinasi saya tidak hanya terbatas pada tim yang ada di Jakarta, tetapi juga mencakup *Tim IT Lapangan* di berbagai cabang perusahaan, seperti di Bali, Surabaya, dan Medan. *Tim IT Lapangan* bertanggung jawab atas pengelolaan data pengiriman barang di setiap cabang dan membantu memastikan bahwa data yang dikumpulkan akurat dan siap untuk dianalisis lebih lanjut. Sebagai *Data Analyst Intern*, saya bertugas untuk mengumpulkan, membersihkan, dan menganalisis data dengan tujuan memberikan informasi yang dapat mendukung pengambilan keputusan di perusahaan. Saya juga bekerja sama dengan *Senior Data Analyst* untuk memastikan bahwa hasil analisis yang saya buat sesuai dengan standar perusahaan dan dapat memberikan rekomendasi yang berguna untuk proses pengambilan keputusan.

Secara keseluruhan, kedudukan saya sebagai *Data Analyst Intern* berada di bawah arahan langsung *Senior Data Analyst* dan *IT Manager*, serta melibatkan

koordinasi yang erat dengan Tim *IT* Lapangan untuk mendukung kelancaran proses operasional logistik di PT. Eka Satya Puspita.



Gambar 3. 1 Alur Kerja Magang pada PT. Eka Satya Puspita

Proses magang dimulai dengan pengumpulan data yang berkaitan dengan pengiriman dan kedatangan barang. Data yang dikumpulkan mencakup informasi mendalam, seperti jenis barang, berat kargo, estimasi waktu kedatangan (ATA), waktu keberangkatan dari gudang (ETD), serta status pengiriman. Data ini diperoleh dari berbagai sumber, baik internal perusahaan, seperti sistem manajemen

logistik yang ada, maupun mitra eksternal, seperti penyedia layanan transportasi dan pihak ketiga yang terlibat dalam proses pengiriman barang. Pengumpulan data ini bertujuan untuk mendapatkan gambaran yang jelas tentang proses logistik yang berlangsung, termasuk pemantauan terhadap kinerja pengiriman dan performa vendor.

Setelah data terkumpul, langkah selanjutnya adalah validasi data. Pada tahap ini, tim memastikan bahwa data yang diperoleh sudah akurat, lengkap, dan sesuai dengan standar yang ditetapkan. Proses validasi ini sangat penting untuk memastikan kualitas data sebelum melanjutkan ke tahap berikutnya. Jika ditemukan data yang tidak valid atau tidak lengkap, maka data tersebut akan diperbaiki atau disesuaikan agar sesuai dengan ketentuan yang berlaku. Setelah validasi data selesai, data yang telah diperbaiki akan melalui tahap pembersihan data. Proses ini melibatkan identifikasi dan penghapusan data yang tidak relevan, duplikat, atau mengandung kesalahan. Pembersihan data dilakukan dengan menggunakan berbagai alat dan teknik, seperti fungsi *filter*, *find & replace*, dan *conditional formatting* di *Excel*, yang bertujuan untuk memastikan hanya data berkualitas tinggi yang digunakan untuk analisis lebih lanjut. Selain itu, pengisian nilai yang hilang dan perbaikan data yang tidak sesuai dilakukan pada tahap ini agar data yang digunakan dalam analisis memiliki konsistensi dan akurasi yang tinggi.

Setelah pembersihan data, dilakukan tahap analisis data untuk menggali informasi yang dapat digunakan dalam pengambilan keputusan. Data yang sudah bersih akan dianalisis menggunakan berbagai alat seperti *Excel*, *Python*, dan *Power BI* untuk mengidentifikasi pola, tren, dan hubungan antar variabel yang penting. Misalnya, analisis dapat dilakukan untuk memantau kinerja pengiriman, menghitung biaya per kilogram, dan mengevaluasi efektivitas rute pengiriman serta kinerja vendor. Selain analisis, visualisasi data juga dilakukan untuk memudahkan pemahaman hasil analisis. Dengan menggunakan grafik, tabel, dan *dashboard* interaktif, tim dapat menyajikan temuan analisis secara lebih jelas dan mudah dipahami oleh manajemen. Visualisasi ini memungkinkan pihak manajemen untuk

melihat perbandingan kinerja antar vendor, mengidentifikasi tren pengiriman, serta mengevaluasi apakah pengiriman sesuai dengan *SLA* yang telah disepakati.

Setelah hasil analisis data dan visualisasi disusun, laporan final akan dipresentasikan kepada *Senior Data Analyst* untuk mendapatkan *feedback* dan validasi lebih lanjut. Dengan umpan balik dari *Senior Data Analyst*, tim akan melakukan penyesuaian atau perbaikan pada analisis atau visualisasi yang sudah dibuat, jika diperlukan. Laporan final yang sudah diperbaiki kemudian dilaporkan ke manajemen, yang akan mengevaluasi hasilnya untuk menentukan apakah rekomendasi yang diberikan sesuai dengan kebutuhan operasional perusahaan. Evaluasi hasil yang dilakukan oleh manajemen akan mempengaruhi pengambilan keputusan untuk perbaikan dan optimalisasi proses logistik. Berdasarkan hasil evaluasi, manajemen akan menentukan langkah-langkah strategis yang diperlukan untuk meningkatkan efisiensi pengiriman, mengurangi biaya, dan memastikan kepuasan pelanggan. Proses ini memastikan bahwa analisis berbasis data yang dilakukan selama magang dapat memberikan kontribusi nyata bagi perusahaan dalam meningkatkan kinerja operasionalnya.

3.2 Tugas dan Uraian Kerja Magang

Data Analyst Intern di PT Eka Satya Puspita diberikan tanggung jawab untuk mendukung pengolahan dan analisis data logistik yang digunakan oleh tim operasional dan unit terkait lainnya. Dalam perannya, diharapkan dapat menerapkan metode analisis data yang efektif untuk mendukung proses pengambilan keputusan berbasis data yang akurat dan tepat waktu. Selain itu, *intern* juga bertanggung jawab dalam mengidentifikasi, menganalisis, dan menyelesaikan masalah terkait kualitas data, termasuk memastikan integritas dan konsistensi data yang digunakan dalam proses analisis. Dengan arahan dan bimbingan dari tim *Data Analyst*, tugas-tugas yang dijalankan meliputi pengumpulan data, pembersihan data, serta analisis untuk menghasilkan wawasan yang berguna bagi evaluasi kinerja dan perbaikan operasional perusahaan. Kolaborasi dengan departemen lain

dalam pengolahan dan visualisasi data juga menjadi bagian penting dalam mendukung kelancaran operasional berbasis data. Selama program magang berlangsung, diharapkan dapat berinisiatif dalam menjalankan tugas-tugas tersebut, serta bertanggung jawab penuh atas kualitas dan akurasi data yang diproses untuk memastikan hasil akhir mendukung kebutuhan bisnis dan peningkatan kinerja tim *Data Analyst*. Tabel 3.1 menunjukkan realisasi dari pencapaian tugas yang telah dilaksanakan selama magang.

Tabel 3. 1 Realisasi Pelaksanaan Magang

No	Kegiatan	Tanggal Mulai	Tanggal Selesai
Memahami Regulasi Perusahaan			
1	Memahami kebijakan dan etika yang berlaku di perusahaan.	10 Februari 2025	14 Februari 2025
Memahami Konteks Bisnis dan Tujuan Analisis (Business Understanding)			
2	Memahami proses bisnis dan alur distribusi barang di perusahaan.	17 Februari 2025	21 Februari 2025
3	Menganalisis kinerja pesaing utama di industri logistik, seperti Inti Duta Logistic.	24 Februari 2025	28 Februari 2025
Eksplorasi dan Validasi Sumber Data (Data Understanding)			

4	Mengumpulkan informasi dari berbagai sumber yang relevan, baik internal maupun eksternal.	10 Maret 2025	14 Maret 2025
5	Memastikan keakuratan dan relevansi data yang diperoleh sebelum digunakan.	17 Maret 2025	21 Maret 2025
Persiapan Data untuk Analisis Lanjutan (Data Preparation)			
6	Membersihkan data dari kesalahan atau ketidaksesuaian agar siap dianalisis.	24 Maret 2025	31 Maret 2025
7	Mengonversi data ke format yang sesuai untuk analisis lebih lanjut.	1 April 2025	4 April 2025
Penerapan Teknik Analisis dan Pembuatan Dashboard (Modeling)			
8	Melakukan analisis data menggunakan perangkat seperti Excel dan Python.	7 April 2025	18 April 2025
9	Pendekatan awal dalam pembentukan model atau pengelompokan data.	21 April 2025	9 Mei 2025
10	Membuat visualisasi data berupa dashboard interaktif untuk memudahkan pemahaman.	9 Mei 2025	16 Mei 2025
Penilaian Kinerja Model dan Visualisasi (Evaluation)			

11	Mengevaluasi kinerja model dengan metrik seperti akurasi, precision, recall, f1-score, dll.	19 Mei 2025	23 Mei 2025
Implementasi Hasil Analisis dan Kolaborasi Tim (Deployment)			
13	Penyusunan laporan analisis	26 Mei 2025	13 Juni 2025
14	Menyajikan hasil analisis kepada tim atau pihak yang berkepentingan.	17 Juni 2025	20 Juni 2025
15	Berkolaborasi dengan tim lain untuk mengimplementasikan hasil analisis dalam operasional.	21 Juni 2025	24 Juni 2025

Tabel 3.1 menggambarkan rangkaian kegiatan yang dilakukan selama magang di PT. Eka Satya Puspita. Kegiatan-kegiatan tersebut terbagi dalam beberapa tahap yang mengikuti metodologi *CRISP-DM (Cross-Industry Standard Process for Data Mining)* untuk memastikan keberhasilan analisis data dan penerapannya dalam operasional perusahaan. Pada tahap Memahami Regulasi Perusahaan, peserta magang memulai dengan memahami kebijakan dan etika yang berlaku di perusahaan. Hal ini penting untuk memastikan bahwa setiap langkah yang diambil selama magang sejalan dengan standar dan aturan yang ada di perusahaan. Selanjutnya, pada Memahami Konteks Bisnis dan Tujuan Analisis (*Business Understanding*), peserta magang mendalami proses bisnis yang ada di perusahaan, termasuk alur distribusi barang dan analisis terhadap pesaing utama seperti Inti Duta Logistic. Pemahaman ini penting untuk menyesuaikan tujuan analisis data dengan konteks operasional perusahaan.

Pada tahap Eksplorasi dan Validasi Sumber Data (*Data Understanding*), peserta magang mengumpulkan data dari berbagai sumber, baik internal maupun eksternal, untuk memastikan keakuratan dan relevansi data yang akan digunakan dalam analisis lebih lanjut. Tahap Persiapan Data untuk Analisis Lanjutan (*Data Preparation*) berfokus pada pembersihan dan konversi data. Data yang tidak konsisten atau memiliki kesalahan diperbaiki agar siap digunakan untuk analisis. Proses ini meliputi penghapusan data yang tidak relevan dan pengisian data yang hilang. Pada tahap Penerapan Teknik Analisis dan Pembuatan *Dashboard* (*Modeling*), peserta magang mulai melakukan analisis data menggunakan perangkat seperti *Excel* dan *Python*, serta mulai membangun model awal untuk pengelompokan atau klasifikasi data. Selain itu, mereka juga membuat *dashboard* interaktif yang memudahkan pemahaman hasil analisis melalui visualisasi data yang jelas.

Pada tahap Penilaian Kinerja Model dan Visualisasi (*Evaluation*), hasil analisis dinilai menggunakan indikator kinerja (*KPI*) untuk memastikan bahwa model yang dibangun dapat memberikan wawasan yang sesuai dengan tujuan perusahaan. Selanjutnya, pengujian dilakukan terhadap fungsionalitas dan akurasi *dashboard* bersama *supervisor* untuk memastikan hasil analisis dapat digunakan dengan efektif. Terakhir, pada tahap Implementasi Hasil Analisis dan Kolaborasi Tim (*Deployment*), hasil analisis disusun dalam laporan dan dipresentasikan kepada tim atau pihak yang berkepentingan. Kolaborasi dengan tim lain dilakukan untuk memastikan bahwa hasil analisis dapat diimplementasikan dalam operasional perusahaan dengan baik, memberikan dampak positif pada proses pengambilan keputusan.

3.2.1 Memahami Regulasi Perusahaan

Pada tahap awal program magang di PT. Eka Satya Puspita, pemahaman mengenai kebijakan dan etika yang berlaku di perusahaan adalah langkah pertama yang sangat penting. Kegiatan ini bertujuan untuk memastikan bahwa peserta magang dapat menyesuaikan diri dengan budaya kerja dan prosedur operasional yang berlaku di perusahaan. Sebagai perusahaan logistik yang terkemuka, PT. Eka Satya Puspita memiliki serangkaian kebijakan dan etika yang harus dipatuhi oleh seluruh karyawan dan mitra kerja, termasuk peserta magang.

Pemahaman yang mendalam tentang regulasi perusahaan tidak hanya membantu peserta magang dalam beradaptasi dengan lingkungan kerja, tetapi juga memastikan bahwa setiap keputusan dan tindakan yang diambil selama program magang tetap sejalan dengan prinsip dan tujuan perusahaan. Dalam konteks ini, regulasi yang diajarkan meliputi aturan tentang durasi dan jam kerja, kewajiban menjaga kerahasiaan informasi perusahaan, serta standar etika profesional yang harus dijaga oleh setiap individu di perusahaan.

1. Durasi dan Jam Kerja

- **Durasi Program Magang:** Program magang berlangsung minimal 3 bulan dan maksimal 6 bulan, untuk memberikan pengalaman yang cukup bagi peserta magang.
- **Jam Kerja:** Peserta magang mengikuti jam kerja reguler perusahaan, yaitu Senin hingga Jumat, pukul 08.00 – 17.00, dengan waktu istirahat selama 1 jam. Peserta juga diperbolehkan untuk bekerja lebih lama hingga pukul 18.00 pada hari Sabtu, sesuai kebutuhan operasional.

- **Pengawasan Kehadiran:** Kehadiran dan waktu kerja peserta magang dipantau oleh Supervisor untuk memastikan kedisiplinan dan keberhasilan dalam penyelesaian tugas.
2. **Tanggung Jawab dan Tugas**
 - Peserta magang diberikan tanggung jawab untuk melaksanakan tugas sesuai dengan bidang pekerjaan yang dipilih. Tugas tersebut bertujuan untuk memberi pengalaman praktis dalam pengelolaan data dan analisis operasional perusahaan.
 - Laporan mingguan wajib disusun untuk melaporkan kemajuan tugas yang telah diselesaikan, serta laporan akhir yang harus diserahkan pada akhir program magang.
 3. **Kewajiban Peserta Magang**
 - Mematuhi peraturan dan kebijakan yang berlaku di PT. Eka Satya Puspita, menjaga kerahasiaan informasi perusahaan yang diperoleh selama magang, serta berpakaian sesuai dengan aturan perusahaan.
 - Memiliki sikap profesional dan kooperatif dengan rekan kerja dan atasan, serta berperan aktif dalam tim untuk memberikan kontribusi positif selama program magang.
 4. **Kedisiplinan dan Sanksi**
 - Peserta magang diwajibkan untuk hadir tepat waktu dan memberikan pemberitahuan jika tidak dapat hadir. Ketidakhadiran tanpa alasan yang jelas lebih dari 3 hari berturut-turut dapat dikenakan sanksi berupa evaluasi ulang atau pemberhentian program magang.
 - Pelanggaran terhadap aturan perusahaan atau etika kerja dapat berakibat pada pemberhentian program magang secara sepihak oleh perusahaan.

5. Evaluasi Kinerja

- Proses evaluasi dilakukan secara berkala untuk mengukur kinerja dan pencapaian peserta magang. Evaluasi ini mencakup aspek kehadiran, kualitas pekerjaan, serta sikap profesional yang ditunjukkan peserta magang selama program berlangsung.
- Penilaian akhir dilakukan untuk menentukan apakah peserta magang memenuhi kriteria yang telah ditetapkan untuk menyelesaikan program dan menerima sertifikat magang.

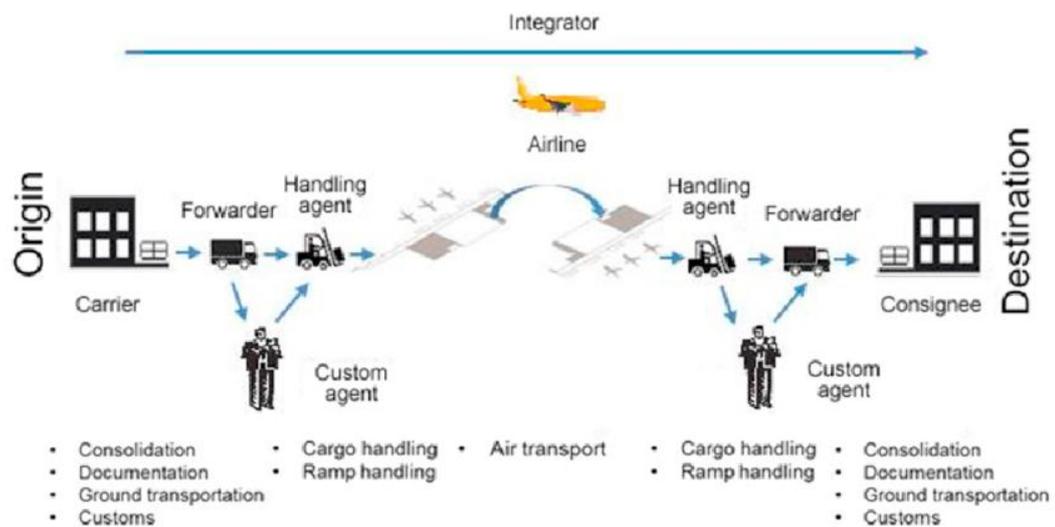
6. Kerahasiaan

- Peserta magang diwajibkan untuk menjaga kerahasiaan seluruh data dan informasi perusahaan yang mereka akses selama program magang. Ketentuan ini berlaku baik selama masa magang maupun setelah program selesai, untuk melindungi data strategis perusahaan.

3.2.2 Memahami Business Understanding

Pada tahap Memahami *Business Understanding*, peserta magang diberikan Informasi mendalam mengenai konteks bisnis dan tujuan analisis yang akan dilakukan. Pemahaman yang kuat tentang bisnis perusahaan sangat penting karena akan menjadi dasar dalam proses pengambilan keputusan yang berbasis data. Di PT. Eka Satya Puspita, pemahaman tentang bisnis ini tidak hanya mencakup pemahaman terhadap operasional internal perusahaan, tetapi juga analisis terhadap kebutuhan dan tantangan yang dihadapi oleh industri logistik, khususnya dalam hal distribusi barang dan pengelolaan pengiriman.

1. Pemahaman Proses Bisnis dan Alur Pengiriman



Gambar 3. 2 Proses Bisnis dan Alur Pengiriman [14]

1. *Origin* (Asal)

- *Forwarder* (Pengirim Barang): Pada tahap ini, PT. Eka Satya Puspita bertanggung jawab dalam mengelola pengumpulan barang dari berbagai sumber, baik dari *e-commerce* maupun pihak lain yang bekerja sama dengan perusahaan. Barang yang dikirimkan akan dibawa menuju bandara untuk diproses lebih lanjut.
- *Handling Agent* (Agen Penanganan): PT. Eka Satya Puspita bekerja sama dengan agen penanganan di bandara untuk memastikan barang diproses dengan benar sebelum dimuat ke pesawat. Agen penanganan akan mengurus pengepakan dan pengecekan barang yang akan dikirimkan, memastikan semuanya siap untuk dikirim ke tujuan.
- *Customs Agent* (Agen Bea Cukai): Sebelum barang dapat dikirim, PT. Eka Satya Puspita memastikan bahwa barang-barang yang akan dikirim memenuhi seluruh persyaratan bea cukai untuk pengiriman udara. Agen bea cukai akan memverifikasi dokumen dan memastikan bahwa barang bebas dari masalah hukum terkait impor/ekspor.

2. *Carrier* (Pengangkut)

- *Airline* (Maskapai Penerbangan): Pada tahap ini, maskapai penerbangan bertanggung jawab untuk mengangkut barang dari bandara asal menuju bandara tujuan. PT. Eka Satya Puspita memilih maskapai penerbangan yang sesuai untuk memastikan pengiriman berjalan dengan cepat dan efisien. Sebagai perusahaan logistik, PT. Eka Satya Puspita juga memastikan bahwa pengiriman melalui udara ini memenuhi standar keamanan dan ketepatan waktu yang telah disepakati dengan pengirim barang.

3. *Destination* (Tujuan)

- *Destination Handling Agent* (Agen Penanganan Tujuan): Setelah barang sampai di bandara tujuan, agen penanganan tujuan bertanggung jawab untuk memproses barang sebelum diteruskan kepada penerima akhir. PT. Eka Satya Puspita bekerja sama dengan agen penanganan di bandara tujuan untuk memastikan barang segera diproses dan siap untuk dikirim lebih lanjut.
- *Forwarder* (Pengelola Pengiriman Tujuan): Setelah barang diproses, PT. Eka Satya Puspita akan bekerja sama dengan pengelola pengiriman di bandara tujuan untuk memastikan barang diantarkan ke penerima yang tepat. Dalam konteks PT. Eka Satya Puspita, pengelolaan pengiriman ini akan dilakukan dengan menggunakan armada pengiriman yang sesuai untuk mengantar barang ke tujuan akhir.
- *Customs Agent* (Agen Bea Cukai Tujuan): Sebelum barang bisa diterima oleh penerima akhir, dokumen barang harus diperiksa oleh bea cukai di negara atau bandara tujuan. PT. Eka Satya Puspita akan memastikan bahwa barang-barang tersebut mematuhi regulasi bea cukai yang berlaku di Indonesia, dengan bekerjasama dengan agen bea cukai setempat.
- *Consignee* (Penerima Barang): Pada tahap akhir, barang akan diserahkan kepada penerima (*Consignee*). PT. Eka Satya Puspita memastikan bahwa

penerima barang dapat menerima kiriman sesuai dengan waktu yang dijanjikan dan dalam kondisi yang baik.

4 Proses Komunikasi dan Dokumentasi

- *Document Flow*: Pada setiap tahap pengiriman, terdapat alur komunikasi dan pertukaran dokumen yang sangat penting, termasuk *invoice*, *packing list*, *air waybill*, dan dokumen lainnya. PT. Eka Satya Puspita memastikan bahwa semua dokumen ini dikelola dengan baik untuk memudahkan proses bea cukai dan pengiriman.
- *Electronic Data Interchange (EDI)*: Dalam menjalankan pengiriman ini, PT. Eka Satya Puspita juga memanfaatkan sistem *EDI* untuk komunikasi yang lebih cepat dan efisien antar pihak terkait (*forwarder*, agen bea cukai, dan maskapai penerbangan).

Pada tahap Memahami *Business Understanding*, peserta magang diharapkan untuk mendapatkan pemahaman yang komprehensif mengenai konteks bisnis perusahaan dan tujuan dari analisis data yang akan dilakukan. Pemahaman yang kuat tentang bisnis perusahaan sangat penting agar analisis data yang dilakukan dapat memberikan wawasan yang relevan dan mendukung keputusan strategis serta operasional yang lebih baik. Di PT. Eka Satya Puspita, memahami proses bisnis pengiriman barang melalui udara menjadi aspek yang sangat vital karena pengiriman udara memerlukan koordinasi antara berbagai pihak yang terlibat dalam rantai pasok logistik, seperti pengelola barang, agen penanganan di bandara, maskapai penerbangan, dan pihak penerima barang. Setiap tahap dalam proses pengiriman memiliki peran penting dalam memastikan pengiriman berjalan dengan efisien dan sesuai dengan *Service Level Agreements (SLA)* yang telah ditetapkan.

Pemahaman Proses Bisnis Pengiriman adalah dasar yang harus dikuasai peserta magang, karena setiap keputusan yang diambil dalam

analisis data harus berhubungan langsung dengan tahapan operasional di perusahaan. Proses pengiriman barang dimulai dengan pengumpulan barang dari berbagai sumber, kemudian barang tersebut diproses oleh agen penanganan, diperiksa oleh bea cukai, dan akhirnya dikirimkan melalui maskapai penerbangan menuju tujuan akhir. Di setiap tahapan, terdapat data yang sangat penting untuk dianalisis, seperti waktu keberangkatan, estimasi waktu kedatangan, biaya pengiriman, serta status pengiriman. Pemahaman terhadap alur ini memungkinkan peserta magang untuk mengidentifikasi titik-titik kritis dalam pengelolaan pengiriman yang perlu mendapatkan perhatian lebih dalam analisis data.

Selain memahami proses pengiriman, peserta magang juga perlu memiliki pemahaman yang mendalam mengenai tujuan analisis data itu sendiri. Tujuan utama dari pengolahan data di PT. Eka Satya Puspita adalah untuk meningkatkan efisiensi operasional dan memastikan pengiriman barang tepat waktu dengan biaya yang terkontrol. Untuk mencapai tujuan ini, perusahaan harus dapat menganalisis berbagai faktor yang mempengaruhi kinerja pengiriman, seperti rute yang dipilih, vendor yang digunakan, serta faktor eksternal lainnya yang mungkin mempengaruhi kecepatan dan biaya pengiriman. Melalui pemahaman yang kuat tentang tujuan ini, peserta magang dapat fokus pada pengolahan data yang relevan, yang akan menghasilkan wawasan untuk pengambilan keputusan yang lebih tepat.

Analisis Data untuk Pengambilan Keputusan adalah langkah berikutnya yang harus dilakukan setelah memahami tujuan analisis. Penggunaan teknologi seperti *Power BI*, *Python*, dan *Excel* memungkinkan peserta magang untuk mengolah data dengan lebih efektif dan efisien. Dalam hal ini, peserta magang harus mampu menganalisis data pengiriman yang mencakup berbagai variabel, seperti waktu pengiriman, biaya pengiriman, dan kinerja vendor. Hasil analisis ini akan menjadi dasar untuk

mengevaluasi apakah pengiriman sudah berjalan sesuai dengan standar yang ditetapkan dan mengidentifikasi area yang perlu diperbaiki. Selain itu, pemahaman tentang indikator kinerja utama (*KPI*) yang digunakan oleh perusahaan untuk mengevaluasi kinerja pengiriman juga penting, karena *KPI* ini akan digunakan untuk menilai keberhasilan operasional perusahaan dalam memenuhi ekspektasi pelanggan.

Dalam tahap *Business Understanding*, analisis terhadap kompetitor menjadi hal penting untuk memahami posisi strategis PT. Eka Satya Puspita di pasar logistik udara. Dengan semakin tingginya tingkat persaingan dan kompleksitas operasional, perusahaan perlu mengetahui keunggulan serta kelemahan relatif dibandingkan para pesaing utama. Analisis ini dilakukan terhadap dua perusahaan logistik yang memiliki cakupan layanan sejenis, yaitu Inti Duta Logistik dan CKL Cargo. Pemilihan keduanya didasarkan pada profil operasional yang relevan, keterlibatan dalam layanan udara maupun multimoda, serta tingkat visibilitas di industri. Informasi dalam analisis berikut dihimpun dari sumber terbuka dan kredibel, mencakup situs resmi perusahaan, publikasi media, serta dokumentasi sertifikasi dan penghargaan yang diperoleh masing-masing perusahaan.

Berikut disajikan tabel SWOT masing-masing perusahaan sebagai dasar pemetaan kekuatan, kelemahan, peluang, dan ancaman yang dihadapi oleh para kompetitor utama PT. Eka Satya Puspita:

➤ **Analisis SWOT – Inti Duta Logistik**

Tabel 3. 2 Tabel SWOT PT. Inti Duta Logistik

Aspek	Uraian
Strengths	<p>- Menyediakan layanan logistik multimoda (darat, laut, udara) ke seluruh Indonesia. - Memiliki sistem pelacakan online 24/7. - Berpengalaman sejak 2008 dan memiliki basis pelanggan kuat di Batam dan sekitarnya.</p>
Weaknesses	<p>- Belum menerapkan teknologi analitik atau dashboard interaktif dalam operasional logistik. - Tidak memiliki sertifikasi mutu formal seperti ISO 9001. - Digitalisasi internal masih terbatas, terutama dalam sistem informasi dan pelaporan.</p>
Opportunities	<p>- Potensi menjadikan Batam sebagai hub logistik regional untuk e-commerce dan ekspor-impor. - Pertumbuhan e-commerce dan perdagangan antarnegara di wilayah Asia Tenggara memberikan peluang ekspansi. - Bisa memperkuat daya saing melalui digitalisasi sistem logistik dan pelayanan pelanggan.</p>
Threats	<p>- Persaingan tinggi dari perusahaan besar nasional dan internasional yang telah menerapkan teknologi lanjutan. - Perubahan kebijakan dan regulasi pemerintah serta dominasi platform e-commerce terhadap vendor logistik dapat menghambat fleksibilitas operasional. - Margin keuntungan bisa menurun akibat tekanan harga di pasar.</p>

➤ Analisis SWOT – CKL Cargo

Tabel 3. 3 Tabel SWOT PT. CKL Cargo

Aspek	Uraian
Strengths	<ul style="list-style-type: none"> - Telah memiliki sertifikasi ISO 9001:2015 (Mutu) dan ISO 45001:2018 (Keselamatan & Kesehatan Kerja). - Memenangkan penghargaan <i>Airfreight Forwarder of the Year 2024</i>. - Memiliki sertifikat halal dari MUI untuk layanan logistik. - Jaringan pengiriman luas, dengan lebih dari 20 kantor cabang di seluruh Indonesia.
Weaknesses	<ul style="list-style-type: none"> - Pelacakan kiriman belum real-time dan kurang interaktif dibandingkan kompetitor teknologi tinggi. - Ketergantungan tinggi pada beberapa maskapai penerbangan bisa menjadi risiko jika terjadi perubahan kebijakan atau keterlambatan.
Opportunities	<ul style="list-style-type: none"> - Posisi Indonesia sebagai pusat industri dan ekspor membuka peluang ekspansi layanan logistik ke kawasan Timur Indonesia. - Tren halal supply chain dan kebutuhan sertifikasi membuka pasar baru di sektor makanan, farmasi, dan logistik syariah. - Potensi penguatan sistem digital dan integrasi dashboard untuk menarik pelanggan e-commerce.
Threats	<ul style="list-style-type: none"> - Dominasi perusahaan asing di sektor logistik nasional dapat menekan ekspansi CKL Cargo. - Persaingan layanan dengan teknologi tinggi dari startup logistik berbasis digital. - Perubahan regulasi dan tekanan pasar dalam pengiriman B2B dan B2C dapat mempengaruhi efisiensi dan kelangsungan bisnis.

Penting juga untuk memanfaatkan keputusan berbasis data dalam meningkatkan efisiensi operasional dan mengurangi biaya. Dalam hal ini, PT. Eka Satya Puspita berfokus pada pengurangan biaya pengiriman, salah satunya melalui analisis efisiensi rute dan pemilihan vendor yang lebih

andal. Dengan memahami kebutuhan ini, peserta magang dapat mengarahkan analisisnya untuk memberikan rekomendasi yang berfokus pada pengoptimalan biaya dan peningkatan kecepatan pengiriman. Di sisi lain, kolaborasi antar departemen juga sangat penting dalam tahap *Business Understanding* ini. Karena analisis data yang dilakukan mencakup berbagai departemen, seperti tim *Data Analyst*, *IT*, dan Operasional, peserta magang perlu bekerja sama dengan berbagai tim untuk memastikan data yang dikumpulkan dan dianalisis dapat digunakan secara maksimal dalam pengambilan keputusan operasional. Kolaborasi ini juga akan memperkaya perspektif peserta magang mengenai bagaimana data digunakan untuk mendukung strategi dan operasi di perusahaan.

Secara keseluruhan, tahap *Business Understanding* memberikan dasar yang kuat bagi peserta magang untuk mengembangkan keterampilan dalam analisis data yang aplikatif di dunia industri, khususnya di perusahaan logistik seperti PT. Eka Satya Puspita. Dengan pemahaman yang mendalam tentang proses bisnis, tujuan analisis, serta kebutuhan perusahaan untuk meningkatkan efisiensi, peserta magang dapat memberikan kontribusi yang signifikan terhadap pengambilan keputusan berbasis data yang mendukung keberhasilan perusahaan dalam pasar yang sangat kompetitif.

3.2.3 Memahami Data Understanding

Date	Origin	DC Destination	LT Number	TO Number	Gross Wgt	Airline	Actual Airline Name	Flight No	ATOP Flight	ATAT Flight	ATAPD Week	ATAPD Warehouse	ATA Vendor Week	Shipping Activity
2-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120CP	30,26 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
3-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120DP	21,63 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
4-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120DP	18,26 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
5-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120SC	0,30 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
6-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130AQUV	22,61 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
7-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130AQUV	19,76 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
8-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120CP	24,15 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
9-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130VY	24,31 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
10-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130VY	25,58 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
11-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024122914WR5	22,77 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
12-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130VY	19,46 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
13-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130W17	24,24 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
14-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120VZ	18,88 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
15-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120AG	20,79 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
16-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130SF	15,55 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
17-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120HZ	10,27 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
18-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130VY	2,97 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
19-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120J0	20,35 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
20-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130VY	16,67 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
21-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130SF	11,76 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
22-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120Y4	8,88 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
23-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130W18	22,22 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
24-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO20241231204WA	0,62 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
25-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024122914WR5	8,79 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
26-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120W54	18,92 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
27-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2025010130SFZ	13,35 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
28-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120W2	13,40 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
29-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024123120NSN	22,92 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
30-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO2024122914M6F	24,94 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
31-Jan-2025	ESP	Kosambi DC Padang DC	LT0P110M7P11	TO20241231204D	20,66 888-81662501	Chilnik CGK	Chilnik CGK	GIG-046	02-Jan-2025 11:00	02-Jan-2025 12:40	02-Jan-2025 12:00	02-Jan-2025 13:39	02-Jan-2025 14:50	Arrived
32-Jan-2025	ESP	Kosambi DC Pekanbaru DC	LT0P110M7R11	TO2024122906WKK	20,16 778-01094030	Pelita Air	Pelita Air	IP-322	02-Jan-2025 11:50	02-Jan-2025 13:40	02-Jan-2025 14:00	02-Jan-2025 14:54	02-Jan-2025 15:48	Arrived
33-Jan-2025	ESP	Kosambi DC Pekanbaru DC	LT0P110M7R11	TO2024122806B9	16,26 778-01094030	Pelita Air	Pelita Air	IP-322	02-Jan-2025 11:50	02-Jan-2025 13:40	02-Jan-2025 14:00	02-Jan-2025 14:54	02-Jan-2025 15:48	Arrived

Gambar 3. 3 Data Understanding

Sebagai langkah awal dalam proses analisis data logistik udara, pengumpulan informasi yang relevan menjadi hal yang sangat penting. Pada tahap ini, saya mengumpulkan data dari berbagai sumber, baik internal maupun eksternal, yang akan memberikan gambaran lengkap terkait aktivitas pengiriman. Sumber internal yang digunakan mencakup data operasional perusahaan yang mencatat berbagai detail pengiriman, seperti tanggal pengiriman, vendor, maskapai penerbangan, nomor *Loading Ticket (LT)*, serta asal dan tujuan pengiriman. Selain itu, informasi mengenai nomor penerbangan dan tanggal kedatangan barang juga diperoleh, yang merupakan bagian dari data yang dapat digunakan untuk mengevaluasi kinerja pengiriman.

Sumber eksternal juga turut berperan penting dalam proses ini. Data eksternal yang digunakan mencakup jadwal penerbangan maskapai, standar *SLA (Service Level Agreement)* yang berlaku di industri logistik udara, serta referensi lainnya yang dapat memperkaya analisis, seperti data cuaca atau informasi terkait kondisi lalu lintas penerbangan. Pada tahap *Data*

Understanding, tujuan utama adalah untuk memeriksa data yang tersedia secara menyeluruh, menilai kualitas data, dan mempersiapkan data tersebut untuk analisis lebih lanjut. Berikut adalah langkah-langkah yang dilakukan untuk memahami data pada *dataset* ini.

1. Pengumpulan Data

Dataset ini berisi informasi terkait pengiriman barang menggunakan layanan udara, dengan kolom-kolom yang menggambarkan berbagai atribut dari pengiriman tersebut, termasuk tanggal pengiriman, vendor, lokasi asal dan tujuan, nomor pengiriman, berat kotor, maskapai, nomor penerbangan, dan status pengiriman. Dataset memiliki total **79173 baris** dan **26 kolom**, yang menunjukkan data pengiriman dalam jumlah besar.

2. Pemahaman Variabel

Beberapa variabel dalam dataset ini meliputi:

1. **Date**: Tanggal pengiriman barang.
2. **Vendor**: Nama penyedia layanan pengiriman (misalnya, ESP).
3. **SC Origin**: Lokasi asal pengiriman.
4. **SC Destination**: Lokasi tujuan pengiriman.
5. **LT Number**: Nomor referensi loading ticket.
6. **TO Number**: Nomor referensi transaction order.
7. **Gross Weight**: Berat total barang yang dikirim.
8. **AWB**: Nomor Air Waybill untuk pengiriman udara.
9. **Actual Airline Name**: Nama maskapai pengangkut.
10. **Flight No**: Nomor penerbangan.
11. **ATDFlight**: Waktu keberangkatan penerbangan.
12. **ATAFlight**: Waktu kedatangan penerbangan.
13. **ATAPODWarehouse**: Waktu kedatangan di gudang tujuan.
14. **ATAVendorWhDe**: Waktu kedatangan di gudang vendor.
15. **Dooring Activity View**: Status kedatangan pengiriman.
16. **Flight No (Duplikasi)**: Nomor penerbangan (duplikasi).

17. **ATAVendorWhDe (Duplikasi)**: Waktu kedatangan di gudang vendor (duplikasi).
18. **SLA Hour**: Waktu yang dijanjikan untuk pengiriman.
19. **Concat**: Penggabungan informasi asal dan tujuan pengiriman.
20. **Target SLA Hour**: Waktu standar SLA pengiriman.
21. **Remarks (Mandatory)**: Catatan tambahan terkait pengiriman.
22. **Station**: Lokasi pengiriman.
23. **Concat Origin-Station**: Penggabungan informasi asal dan stasiun tujuan.
24. **Truck Type**: Tipe kendaraan pengiriman darat.
25. **Trucking**: Informasi tentang transportasi darat.
26. **Remarks**: Catatan tambahan lainnya.

3.2.4 Persiapan Data untuk Analisis Lanjutan (Data Preparation)

Tahap *Data Preparation* merupakan langkah penting untuk menyiapkan data agar layak dianalisis. Proses ini mencakup pemeriksaan dimensi dan tipe data, penanganan *missing values* dengan penghapusan atau estimasi, analisis distribusi variabel penting seperti status kedatangan, serta identifikasi dan pengelolaan *outliers* agar tidak memengaruhi hasil analisis.

Sebagai *Data Analyst Intern* di PT Eka Satya Puspita, digunakan kombinasi Microsoft Excel dan *Python*. Excel dimanfaatkan untuk pengecekan dan penyaringan awal secara visual, sedangkan *Python* dengan pustaka seperti *pandas*, *numpy*, dan *openpyxl* digunakan untuk pembersihan, segmentasi data pengiriman, dan perhitungan metrik SLA. Pendekatan ini membentuk dasar data yang kuat sebelum masuk ke tahap *modeling* dan analisis *dashboard* cerdas.

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import mean_squared_error, r2_score, classification_report, accuracy_score

```

Gambar 3. 4 import Library

Kode ini mengimpor pustaka yang diperlukan untuk analisis data dan *machine learning*, seperti *pandas* untuk manipulasi data, *numpy* untuk operasi numerik, serta *seaborn* dan *matplotlib* untuk visualisasi. Selanjutnya, *train_test_split* digunakan untuk membagi data menjadi data pelatihan dan pengujian. *LinearRegression* dan *RandomForestClassifier* dari *scikit-learn* dipilih untuk membangun model regresi dan klasifikasi. Setelah model dibangun, *mean_squared_error*, *r2_score*, *classification_report*, dan *accuracy_score* digunakan untuk mengevaluasi kinerja model.

```

STEP 1: CEK DIMENSI DATA & TIPE DATA

]: # Membaca data
file_path = "C:/Users/Agra/Downloads/Performance ESP MODELING.xlsx"
df = pd.read_excel(file_path, sheet_name="Merge Performance")

# Cek dimensi data
print("Jumlah baris dan kolom:", df.shape)

# Tampilkan nama kolom dan tipe datanya
print("\nTipe data per kolom:")
print(df.dtypes)

# Lihat 5 baris pertama sebagai sampel
print("\nContoh data:")
print(df.head())

Jumlah baris dan kolom: (79173, 26)

Tipe data per kolom:
Date                object
Vendor              object
SC Origin           object
SC Destination      object
LT Number           object
TO Number           object
Gross Weight        float64
AWB                 object
Actual\nAirline Name object
Flight No           object
ATD\nFlight         object
ATA\nFlight         object

```

Gambar 3. 5 Code Cek Tipe Data

Pada *Step 1: Check Data Dimensions & Types*, data dibaca dari file Excel menggunakan fungsi *pd.read_excel()* pada sheet "Merge Performance", lalu disimpan ke variabel *df*. Ukuran data diperiksa dengan

df.shape yang menunjukkan 79.173 baris dan 26 kolom, memberi gambaran awal tentang skala dataset. Tipe data dicek menggunakan *df.dtypes*, yang menunjukkan kolom seperti *Date*, *Vendor*, *SC Origin*, dan *SC Destination* bertipe *object* (teks), sementara *Gross Weight* dan *Amount* bertipe *float64*, serta *Rate* bertipe *int64*. Lima baris awal ditampilkan dengan *df.head()* sebagai sampel untuk memverifikasi struktur dan format data sebelum analisis lanjut.

STEP 2: CEK MISSING VALUES & DISTRIBUSI Arrival Status

```
# Cek missing values per kolom
print("Jumlah Missing Values per Kolom:\n")
print(df.isnull().sum())

# Cek distribusi Arrival Status
print("\nDistribusi Arrival Status:\n")
print(df['Arrival Status Vendor'].value_counts(dropna=False))
```

Jumlah Missing Values per Kolom:

Date	0
Vendor	0
SC Origin	0
SC Destination	0
LT Number	0
TO Number	0
Gross Weight	0
AWB	0
Actual\nAirline Name	0
Flight No	0
ATD\nFlight	0
ATA\nFlight	1
ATA\nPOD Warehouse	0
ATD\nPOD Warehouse	0
ATA\nVendor WH Destination\n(Sertakan Link Evidence)	0
Dooring Activity Vendor	0
Arrival Status Vendor	0
Issue	68141
REMARKS (Mandatory)	68165

Gambar 3. 6 Code Cek Missing Value

Pada *Step 2: Cek Missing Values & Distribusi Arrival Status*, kode pertama-tama memeriksa jumlah nilai yang hilang di setiap kolom menggunakan *df.isnull().sum()*, yang menunjukkan kolom seperti *ATA/Flight* dan *Issue* memiliki nilai yang hilang. Selanjutnya, distribusi status kedatangan pengiriman diperiksa dengan *df['Arrival Status Vendor'].value_counts(dropna=False)*, yang menunjukkan bahwa sebagian besar pengiriman tiba *On Time* (69.065 pengiriman), sementara sisanya terlambat (*Late*, 10.108 pengiriman). Langkah ini penting untuk mengevaluasi kualitas data dan memahami sebaran status kedatangan pengiriman.

STEP 3: KONVERSI Target SLA & Actual SLA MENJADI JAM (float)

```
import numpy as np

# Fungsi untuk mengubah format "HH:MM:SS" ke jam float
def time_to_hours(time_str):
    try:
        h, m, s = map(int, str(time_str).split(':'))
        return h + m / 60 + s / 3600
    except:
        return np.nan

# Terapkan ke kolom
df['Target_SLA_Hours'] = df['Target SLA Hour'].apply(time_to_hours)
df['Actual_SLA_Hours'] = df['Actual SLA Hour'].apply(time_to_hours)

# Cek hasil konversi
print(df[['Target SLA Hour', 'Target_SLA_Hours', 'Actual SLA Hour', 'Actual_SLA_Hours']].head(10))
```

	Target SLA Hour	Target_SLA_Hours	Actual SLA Hour	Actual_SLA_Hours
0	32:00:00	32.0	17:59:00	17.983333
1	32:00:00	32.0	17:59:00	17.983333
2	32:00:00	32.0	17:59:00	17.983333
3	32:00:00	32.0	17:59:00	17.983333
4	32:00:00	32.0	17:59:00	17.983333
5	32:00:00	32.0	17:59:00	17.983333

Gambar 3. 7 Code Convert Target SLA & Actual SLA

Pada *Step 3*, kode ini mengonversi waktu dalam format "HH:MM:SS" menjadi angka jam dengan format *float*. Fungsi `time_to_hours` memecah waktu menjadi jam, menit, dan detik, lalu menghitungnya dalam jam. Fungsi ini diterapkan pada kolom *Target SLA Hour* dan *Actual SLA Hour* untuk menghasilkan kolom baru, *Target_SLA_Hours* dan *Actual_SLA_Hours*, yang berisi nilai waktu dalam bentuk jam *float*. Hasil konversi ini kemudian ditampilkan untuk memastikan data telah terkonversi dengan benar. Langkah ini mempermudah analisis waktu *SLA* dalam format numerik.

STEP 4: CEK RATA-RATA SLA UNTUK SETIAP Arrival Status

```
print("Rata-rata Actual SLA Hour berdasarkan Arrival Status:\n")
print(df.groupby('Arrival Status Vendor')['Actual_SLA_Hours'].mean())

print("\nRata-rata Target SLA Hour berdasarkan Arrival Status:\n")
print(df.groupby('Arrival Status Vendor')['Target_SLA_Hours'].mean())
```

Rata-rata Actual SLA Hour berdasarkan Arrival Status:

```
Arrival Status Vendor
Late      9257.604303
On Time   1854.644061
Name: Actual_SLA_Hours, dtype: float64
```

Rata-rata Target SLA Hour berdasarkan Arrival Status:

```
Arrival Status Vendor
Late      29.042057
On Time   29.804920
Name: Target_SLA_Hours, dtype: float64
```

Gambar 3. 8 Code Cek Rata-rata SLA

Pada *Step 4*, kode ini digunakan untuk menghitung rata-rata *SLA* berdasarkan *Arrival Status* (*On Time* dan *Late*). Pertama, kode menghitung rata-rata *Actual SLA Hour* untuk setiap status kedatangan menggunakan fungsi `groupby('Arrival Status Vendor')` yang mengelompokkan data berdasarkan kolom *Arrival Status Vendor*, kemudian menghitung rata-rata untuk kolom *Actual_SLA_Hours* dengan `.mean()`. Hasilnya menunjukkan bahwa rata-rata *Actual SLA Hour* untuk pengiriman *Late* adalah 9257.604303, sementara untuk *On Time* adalah 1854.644061. Kemudian, kode yang sama diterapkan untuk kolom *Target SLA Hour*, yang menghitung rata-rata *Target_SLA_Hours* untuk masing-masing status kedatangan. Rata-rata *Target SLA Hour* untuk *Late* adalah 29.042057, dan untuk *On Time* adalah 29.804920.

STEP 5: CEK NILAI MAKSIMUM Actual SLA Hour

```
print("Nilai maksimum Actual SLA Hour:", df['Actual_SLA_Hours'].max())

# Tentukan batas outlier (misal lebih dari 500 jam dianggap outlier)
outlier_threshold = 500

# Hitung berapa banyak yang di atas batas
outlier_count = (df['Actual_SLA_Hours'] > outlier_threshold).sum()
print("Jumlah data di atas", outlier_threshold, "jam:", outlier_count)
```

```
Nilai maksimum Actual SLA Hour: 1577892.5833333333
Jumlah data di atas 500 jam: 175
```

Gambar 3. 9 Code Cek Nilai Maksimum Actual SLA

Pada *Step 5*, dilakukan pengecekan nilai maksimum kolom *Actual SLA Hour* menggunakan `df['Actual_SLA_Hours'].max()`, yang menghasilkan nilai ekstrem 1.577.892,58 jam — indikasi data tidak realistis. Batas *outlier* ditetapkan pada 500 jam melalui variabel `outlier_threshold = 500`. Dengan kondisi `df['Actual_SLA_Hours'] > outlier_threshold`, ditemukan 175 baris data yang melebihi ambang tersebut dan dikategorikan sebagai *outliers*.

STEP 6: FILTER DATA - HAPUS OUTLIER

```
: # Buat salinan data tanpa outlier
df_clean = df[df['Actual_SLA_Hours'] <= 500].copy()

print("Jumlah data setelah hapus outlier:", df_clean.shape[0])
```

Jumlah data setelah hapus outlier: 78998

Gambar 3. 10 Code Filter Data & Hapus Outlier

Pada *Step 6*, kode ini digunakan untuk menghapus *outlier* dari *dataset*. Langkah pertama adalah membuat salinan data tanpa *outlier* menggunakan kondisi `df['Actual_SLA_Hours'] <= 500`, yang menyaring data dan hanya menyertakan baris di mana *Actual SLA Hour* kurang dari atau sama dengan 500 jam. Salinan data yang sudah difilter ini disimpan dalam variabel `df_clean`. Kemudian, kode menampilkan jumlah baris yang tersisa setelah menghapus *outlier* dengan `df_clean.shape[0]`, yang menunjukkan bahwa setelah *outlier* dihapus, *dataset* terdiri dari 78.998 baris.

Encoding Target (Arrival Status) ¶

```
# Buat kolom target biner
df_clean['Arrival_Status_Binary'] = df_clean['Arrival Status Vendor'].map({'On Time': 0, 'Late': 1})

# Cek distribusi target biner
print("Distribusi Arrival_Status_Binary:\n")
print(df_clean['Arrival_Status_Binary'].value_counts())
```

Distribusi Arrival_Status_Binary:

```
Arrival_Status_Binary
0    68949
1    10049
Name: count, dtype: int64
```

Gambar 3. 11 Encoding Target Arrival Status

Pada *Step 7*, dilakukan proses *encoding* untuk mengubah kolom *Arrival Status Vendor* menjadi format biner agar dapat digunakan dalam model *machine learning*. Nilai *'On Time'* dikonversi menjadi 0 dan *'Late'* menjadi 1, lalu disimpan ke dalam kolom baru *Arrival_Status_Binary*. Distribusi kelas kemudian diperiksa menggunakan `value_counts()`, dengan

hasil: pengiriman *On Time* sebanyak 68.949 dan *Late* sebanyak 10.049. Tahap ini penting agar label target siap digunakan dalam proses *classification*.

```
Encoding Fitur Kategorikal

from sklearn.preprocessing import LabelEncoder
# Copy dataframe supaya aman
df_model = df_clean.copy()
# List kolom kategorikal
kategori_cols = ['Vendor', 'SC Origin', 'SC Destination']
# Label Encoding
le = LabelEncoder()
for col in kategori_cols:
    df_model[col] = le.fit_transform(df_model[col])
    print(f'Label encoding selesai untuk kolom: {col}')

# Cek hasil encoding
print("\nContoh data setelah encoding:")
print(df_model[ [*kategori_cols] ].head())

Label encoding selesai untuk kolom: Vendor
Label encoding selesai untuk kolom: SC Origin
Label encoding selesai untuk kolom: SC Destination

Contoh data setelah encoding:
   Vendor  SC Origin  SC Destination
0        0           1                11
1        0           1                11
2        0           1                11
3        0           1                11
4        0           1                11
```

Gambar 3. 12 Encoding Fitur Kategorikal

Pada tahap 7 ini dilakukan *label encoding* untuk mengubah data kategorikal menjadi format numerik agar dapat diproses oleh algoritma *machine learning*. Kolom yang dikategorikan, yaitu '*Vendor*', '*SC Origin*', dan '*SC Destination*', dikodekan menggunakan *LabelEncoder* dari pustaka *scikit-learn*. Data *dataframe* disalin terlebih dahulu agar proses tidak memodifikasi data asli (*df_clean.copy()*). Setiap nilai unik dalam kolom dikonversi menjadi angka, misalnya *Vendor* dengan label string akan diubah menjadi angka seperti 0, 1, 2, dst.

Hasil akhir menunjukkan data yang telah di-*encode*, siap untuk dimasukkan ke dalam model *machine learning*. Proses ini penting untuk memastikan fitur kategorikal dapat dikenali dan dihitung secara matematis oleh algoritma.

STEP Split Data Train-Test

```
from sklearn.model_selection import train_test_split

# Fitur (X) dan Target (y)
X = df_model[['Gross Weight', 'Target_SLA_Hours', 'Rate', 'Vendor', 'SC Origin', 'SC Destination']]
y = df_model['Arrival_Status_Binary']

# Split data: 80% training, 20% testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)

print("Jumlah data train:", X_train.shape[0])
print("Jumlah data test:", X_test.shape[0])

Jumlah data train: 63198
Jumlah data test: 15800
```

Gambar 3. 13 Split Data Train-Test

Pada tahap ini dilakukan pemisahan data menjadi dua bagian menggunakan fungsi *train_test_split* dari *scikit-learn*. Data fitur (*X*) terdiri dari kolom-kolom seperti *Gross Weight*, *Target_SLA_Hours*, *Rate*, *Vendor*, *SC Origin*, dan *SC Destination*. Sedangkan target (*y*) adalah kolom *Arrival_Status_Binary*. Data dibagi dengan rasio 80% untuk *training* dan 20% untuk *testing*. Parameter *random_state=42* memastikan hasil pembagian tetap konsisten, sedangkan *stratify=y* menjaga proporsi kelas target tetap seimbang pada kedua set data.

3.2.5 Penerapan Teknik Analisis (Modeling)

Setelah tahap *Data Preparation* selesai, langkah selanjutnya adalah penerapan teknik analisis dan pembuatan *dashboard* interaktif. Data yang telah dibersihkan digunakan untuk membangun model analisis guna menghasilkan *insight* yang mendukung pengambilan keputusan operasional. Proses ini melibatkan penggunaan alat seperti *Excel*, *Python*, dan *Power BI* untuk analisis dan visualisasi data secara efektif, sehingga mempermudah pemantauan kinerja dan pengambilan keputusan berbasis data.

1. Pelatihan Model

Setelah tahap *Data Preparation* yang mencakup pemeriksaan dimensi dan tipe data, penanganan nilai yang hilang, dan pembersihan data

dari pencilan (*outliers*), langkah selanjutnya dalam proses analisis adalah pelatihan model. Pada tahap ini, beberapa model analitik digunakan untuk mengeksplorasi pola-pola yang ada dalam data, serta untuk melakukan prediksi dan klasifikasi yang dapat mendukung pengambilan keputusan yang lebih baik dalam operasional logistik. Dalam laporan ini, dua jenis model utama digunakan, yaitu *Random Forest Classifier* dan *Support Vector Machine (SVM)*, yang keduanya diterapkan untuk mengklasifikasikan status kedatangan pengiriman (*On Time* atau *Late*). Selain itu, digunakan pula model deteksi anomali untuk mengidentifikasi pengiriman yang tidak biasa, yang mungkin mempengaruhi kualitas dan efisiensi pengiriman.

A. Random Forest Classifier

Random Forest adalah salah satu algoritma *ensemble learning* yang menggabungkan hasil dari banyak *decision trees* untuk meningkatkan akurasi prediksi. Dalam model ini, setiap *pohon keputusan* yang dibangun pada data akan membuat prediksi sendiri, dan hasil akhirnya akan diambil berdasarkan mayoritas suara dari seluruh pohon. Kelebihan utama dari *Random Forest* adalah kemampuannya dalam menangani data yang besar dan kompleks dengan banyak fitur, serta kemampuannya dalam menangani data yang tidak terstruktur dengan baik.

```
# Memisahkan Fitur dan Target untuk klasifikasi
X_class = df[['Gross Weight', 'Target_SLA_Hours', 'Rate', 'Vendor', 'SC Origin', 'SC Destination', 'Station']]
y_class = df['Arrival_Status_Binary'] # Target untuk klasifikasi

# Split Data untuk klasifikasi
X_train_class, X_test_class, y_train_class, y_test_class = train_test_split(X_class, y_class, test_size=0.2, random_state=42, stratify=y_class)

# Modeling Random Forest
rf = RandomForestClassifier(random_state=42, class_weight='balanced', n_estimators=100)
rf.fit(X_train_class, y_train_class)
y_pred_rf = rf.predict(X_test_class)

# Menampilkan hasil Random Forest
print("\nAkurasi Random Forest: ", accuracy_score(y_test_class, y_pred_rf))
print("Classification Report Random Forest:\n", classification_report(y_test_class, y_pred_rf))
```

Gambar 3. 14 Code Random Forest

Pada tahap pelatihan, *Random Forest Classifier* digunakan untuk mengklasifikasikan status kedatangan pengiriman. Fitur yang digunakan

untuk melatih model ini mencakup variabel-variabel seperti *Gross Weight*, *Arrival Status Vendor*, *Flight No*, dan *ATD Flight*, yang dianggap sebagai faktor-faktor utama yang mempengaruhi ketepatan waktu pengiriman. *Dataset* dibagi menjadi dua bagian: *Training Set* (80%) untuk melatih model, dan *Testing Set* (20%) untuk menguji akurasi dan performa model. Untuk melatih model *Random Forest*, pustaka *sklearn.ensemble* digunakan dengan memanfaatkan *RandomForestClassifier*. Proses ini memungkinkan model untuk mempelajari hubungan antara fitur-fitur *input* dengan status kedatangan (*On Time* atau *Late*). Salah satu keuntungan utama dari *Random Forest* adalah kemampuannya untuk menghindari *overfitting*, terutama ketika *dataset* mengandung banyak *noise* dan variabel yang tidak relevan.

Setelah model selesai dilatih, prediksi dilakukan pada *Testing Set*, dan hasilnya dievaluasi dengan menggunakan *accuracy_score* dan *classification_report*. *accuracy_score* mengukur seberapa banyak prediksi yang benar dibandingkan dengan total prediksi, sementara *classification_report* memberikan metrik lebih rinci seperti *precision*, *recall*, dan *f1-score* untuk masing-masing kelas (*On Time* dan *Late*).

B. Support Vector Machine (SVM)

Selain *Random Forest*, *Support Vector Machine (SVM)* juga diterapkan untuk klasifikasi status kedatangan pengiriman. *SVM* bekerja dengan cara mencari *hyperplane* terbaik yang memisahkan dua kelas (dalam hal ini *On Time* dan *Late*). Algoritma ini berupaya menemukan *margin* terbesar antara kedua kelas, sehingga menghasilkan klasifikasi yang lebih stabil dan akurat. Pada proses pelatihan, *Support Vector Classifier (SVC)* dari pustaka *sklearn.svm* digunakan untuk membangun model. Data yang telah diproses digunakan dalam pelatihan, dengan fitur-fitur seperti *Arrival Status Vendor*, *Target SLA Hour*, dan *Actual SLA Hour* sebagai *input* untuk memprediksi apakah pengiriman akan *On Time* atau *Late*. *SVM* sangat berguna ketika data tidak dapat dipisahkan secara linear, karena model ini

dapat menggunakan fungsi *kernel* yang sesuai untuk menangani data tidak linier. Namun, salah satu kelemahan dari *SVM* adalah sensitivitasnya terhadap pemilihan parameter *kernel* dan pengaturan *hyperparameter*. Oleh karena itu, proses optimasi terhadap *kernel*, serta parameter *C* dan *gamma*, sangat penting untuk memperoleh performa terbaik dari model ini.

```
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report

# Modeling Support Vector Machine (SVM)
svm = SVC(random_state=42)
svm.fit(X_train_class, y_train_class)
y_pred_svm = svm.predict(X_test_class)

# Menampilkan hasil Support Vector Machine (SVM)
print("\nAkurasi Support Vector Machine: ", accuracy_score(y_test_class, y_pred_svm))
print("Classification Report SVM:\n", classification_report(y_test_class, y_pred_svm))
```

Gambar 3. 15 Code SVM Model

Prediksi dilakukan pada testing set, dan hasil evaluasi model diperoleh dengan menggunakan *accuracy score* dan *classification report*, yang mencakup metrik seperti *precision*, *recall*, dan *f1-score*.

C. Model Deteksi Anomali (Anomaly Detection)

Selain klasifikasi status kedatangan, model deteksi anomali juga diterapkan untuk mengidentifikasi pengiriman yang tidak biasa atau mencurigakan. Deteksi anomali bertujuan untuk mendeteksi pengiriman yang memiliki pola yang sangat berbeda dibandingkan dengan pengiriman lainnya, yang bisa mengindikasikan adanya kesalahan atau masalah dalam proses pengiriman. Deteksi anomali ini sangat penting dalam konteks logistik untuk mengidentifikasi masalah yang bisa mempengaruhi kualitas layanan, seperti keterlambatan yang tidak wajar atau masalah lainnya yang dapat berisiko bagi operasional perusahaan.

```

# Membaca data (pastikan untuk mengganti path sesuai file Anda)
file_path = "C:/Users/Agra/Downloads/Performance ESP MODELING.xlsx"
df = pd.read_excel(file_path, sheet_name="Merge Performance")

# Fungsi untuk mengubah format "HH:MM:SS" ke jam float
def time_to_hours(time_str):
    try:
        h, m, s = map(int, str(time_str).split(':'))
        return h + m / 60 + s / 3600
    except:
        return np.nan

# Terapkan konversi pada kolom "Target SLA Hour" dan "Actual SLA Hour"
df['Target_SLA_Hours'] = df['Target SLA Hour'].apply(time_to_hours)
df['Actual_SLA_Hours'] = df['Actual SLA Hour'].apply(time_to_hours)

# Cek kolom yang sudah terbuat
print(df[['Target SLA Hour', 'Target_SLA_Hours', 'Actual SLA Hour', 'Actual_SLA_Hours']].head())

# Encoding kolom kategorikal
label_encoder = LabelEncoder()
df['Vendor'] = label_encoder.fit_transform(df['Vendor'])
df['SC Origin'] = label_encoder.fit_transform(df['SC Origin'])
df['SC Destination'] = label_encoder.fit_transform(df['SC Destination'])
df['Station'] = label_encoder.fit_transform(df['Station'])

```

Gambar 3. 16 Code Anomaly Detection

```

# Pilih fitur yang relevan untuk deteksi anomali
X = df[['Gross Weight', 'Target_SLA_Hours', 'Rate', 'Vendor', 'SC Origin', 'SC Destination', 'Station']]

# Mengisi missing values dengan nilai rata-rata (untuk numerik)
X = X.fillna(X.mean())

# Melakukan split data untuk training dan testing (80-20)
X_train, X_test = train_test_split(X, test_size=0.2, random_state=42)

# Model Isolation Forest untuk deteksi anomali
model = IsolationForest(contamination=0.05, random_state=42) # contamination: proporsi data anomali
model.fit(X_train)

# Prediksi anomali
y_pred_train = model.predict(X_train)
y_pred_test = model.predict(X_test)

# Mengubah label prediksi: 1 = normal, -1 = anomali
X_train['Anomaly'] = y_pred_train
X_test['Anomaly'] = y_pred_test

```

Gambar 3. 17 Code Anomaly Detection

```

# Visualisasi hasil deteksi anomali pada data testing
plt.figure(figsize=(10, 6))
sns.scatterplot(x=X_test['Gross Weight'], y=X_test['Rate'], hue=X_test['Anomaly'], palette={1: 'green', -1: 'red'})
plt.title('Deteksi Anomali (Testing Data)')
plt.xlabel('Gross Weight')
plt.ylabel('Rate')
plt.show()

# Melihat jumlah anomali di data testing
print(f"Jumlah anomali pada data testing: {sum(y_pred_test == -1)}")

```

Gambar 3. 18 Code Anomaly Detection

Deteksi anomali diterapkan untuk mengidentifikasi pengiriman yang tidak biasa atau mencurigakan. Model ini bertujuan untuk mendeteksi pengiriman yang memiliki pola yang sangat berbeda dibandingkan dengan pengiriman lainnya, yang bisa mengindikasikan adanya kesalahan atau masalah dalam proses pengiriman. Dalam konteks logistik, deteksi anomali sangat penting untuk mengidentifikasi masalah yang dapat mempengaruhi

kualitas layanan, seperti keterlambatan yang tidak wajar atau masalah lainnya yang dapat berisiko bagi operasional perusahaan.

Model *Isolation Forest* digunakan untuk mendeteksi anomali dalam *dataset*. *Isolation Forest* merupakan algoritma berbasis *pohon keputusan* yang dirancang khusus untuk mendeteksi pencilan (*outliers*) dalam *dataset* berukuran besar. Keunikan dari model ini terletak pada kemampuannya dalam mengisolasi pengamatan yang berbeda menggunakan struktur pohon, sehingga memungkinkan proses deteksi dilakukan secara efisien, bahkan pada data berskala besar. Proses pelatihan dilakukan dengan memilih fitur-fitur yang relevan, seperti *Gross Weight*, *AWB*, dan *Flight No*, untuk menganalisis pola pengiriman dan mendeteksi entri yang menunjukkan perilaku berbeda dibandingkan pengiriman lainnya. Setelah model dilatih, prediksi dilakukan terhadap *training set* dan *testing set* untuk mengidentifikasi data yang terindikasi sebagai anomali.

Setelah model *Isolation Forest* dilatih, prediksi dilakukan pada data yang telah dibagi menjadi *training set* dan *testing set*. Hasil deteksi anomali kemudian dianalisis, di mana data yang teridentifikasi sebagai anomali diberi label *Anomaly* dengan nilai -1, sedangkan data yang dianggap normal diberi label 1. Visualisasi hasil deteksi dilakukan menggunakan *scatter plot* untuk menunjukkan hubungan antara *Gross Weight* dan *Rate*, dengan titik data normal ditandai warna hijau dan anomali ditandai warna merah. Visualisasi ini memudahkan identifikasi sebaran anomali pada *training set* dan *testing set*, serta memberikan gambaran yang jelas mengenai data yang memerlukan perhatian lebih lanjut.

Jumlah anomali pada data *testing* juga dihitung untuk memberikan gambaran tentang seberapa banyak pengiriman yang terdeteksi sebagai anomali. Hasil ini digunakan untuk memfokuskan penyelidikan lebih lanjut terhadap pengiriman yang dianggap mencurigakan atau bermasalah.

Dengan demikian, model deteksi anomali ini tidak hanya membantu dalam identifikasi masalah kualitas layanan, tetapi juga memberikan langkah awal untuk melakukan investigasi terhadap penyebab keterlambatan atau kesalahan pengiriman.

1. Evaluasi Model

Setelah proses pelatihan selesai, langkah selanjutnya adalah melakukan evaluasi model untuk menilai seberapa baik model dapat memprediksi atau mengklasifikasikan data yang sebelumnya tidak terlihat. Evaluasi ini juga penting untuk memastikan model dapat digunakan untuk melakukan prediksi yang akurat dalam kondisi dunia nyata.

➤ Evaluasi Random Forest Classifier

```

Akurasi Random Forest: 0.8351752447110831
Classification Report Random Forest:

```

	precision	recall	f1-score	support
0	0.92	0.89	0.90	13813
1	0.39	0.49	0.43	2022
accuracy			0.84	15835
macro avg	0.65	0.69	0.67	15835
weighted avg	0.85	0.84	0.84	15835

Gambar 3. 19 Result Evaluation Random Forest

Untuk *Random Forest Classifier*, evaluasi dilakukan menggunakan *accuracy score*, yaitu metrik yang mengukur persentase prediksi yang benar dari total prediksi. Berdasarkan hasil yang diperoleh, model *Random Forest* memiliki akurasi sekitar 0,835 atau 83,5%, yang menunjukkan bahwa model ini mampu memprediksi status kedatangan pengiriman dengan cukup baik. Selanjutnya, *classification report* digunakan untuk memberikan gambaran lebih rinci mengenai performa model. Pada metrik *precision*, model menunjukkan nilai 0,92 untuk kelas 0 (*On Time*), yang berarti 92% dari prediksi kelas tersebut benar. Namun, untuk kelas 1 (*Late*), nilai *precision* hanya 0,39, yang menunjukkan bahwa hanya 39% dari prediksi keterlambatan yang benar.

Hal serupa terlihat pada metrik *recall* dan *f1-score*, di mana model lebih baik dalam mengidentifikasi pengiriman *On Time* dibandingkan pengiriman *Late*. *Recall* untuk kelas 0 adalah 0,89, sedangkan *f1-score* untuk kelas 1 hanya 0,43, yang menunjukkan bahwa model memiliki keterbatasan dalam mendeteksi keterlambatan secara akurat.

➤ **Evaluasi Support Vector Machine (SVM)**

Akurasi Support Vector Machine: 0.880770445216293					
Classification Report SVM:					
	precision	recall	f1-score	support	
0	0.88	0.99	0.94	13813	
1	0.73	0.10	0.18	2022	
accuracy			0.88	15835	
macro avg	0.81	0.55	0.56	15835	
weighted avg	0.86	0.88	0.84	15835	

Gambar 3. 20 Result Evaluation SVM

Pada *Support Vector Machine (SVM)*, evaluasi dilakukan menggunakan metrik yang sama, yaitu *accuracy*, *classification report*, dan *confusion matrix*. Berdasarkan hasil yang diperoleh, model *SVM* memiliki akurasi yang sedikit lebih tinggi dibandingkan *Random Forest*, yaitu sekitar 0,89 atau 88,9%. Meskipun demikian, performa model pada metrik *precision*, *recall*, dan *f1-score* menunjukkan adanya kelemahan signifikan dalam mendeteksi pengiriman *Late*.

Pada metrik *precision*, model *SVM* mencatat nilai 0,88 untuk kelas 0 (*On Time*), yang menunjukkan tingkat presisi tinggi dalam memprediksi pengiriman yang tepat waktu. Namun, *recall* untuk kelas 1 (*Late*) sangat rendah, hanya sebesar 0,10, yang berarti model gagal mengidentifikasi sebagian besar pengiriman yang terlambat. Hal ini juga tercermin pada *f1-score* untuk kelas 1 yang hanya mencapai 0,18, menandakan performa yang buruk dalam mendeteksi keterlambatan.

➤ **Perbandingan Hasil Evaluasi**

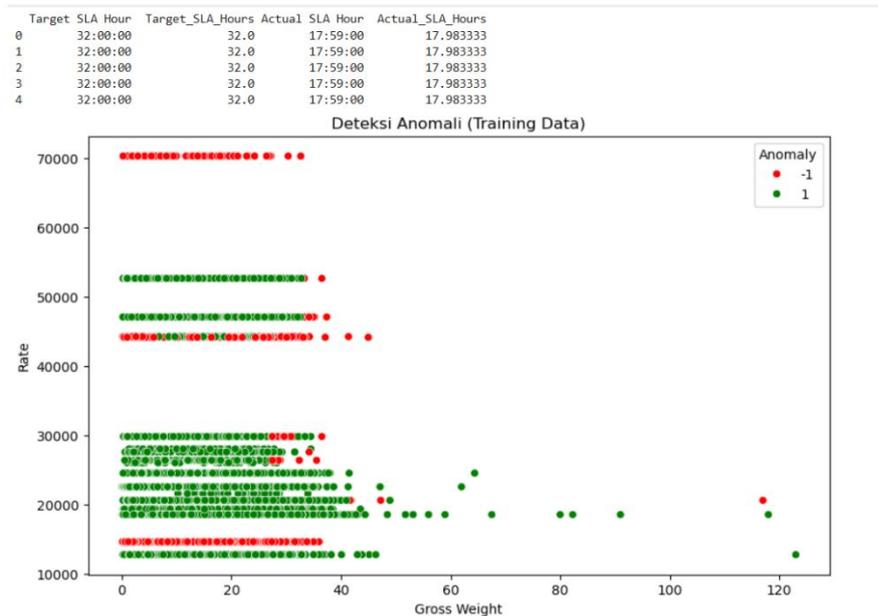
Dari hasil evaluasi kedua model, terlihat adanya perbedaan signifikan dalam kinerja *Random Forest* dan *Support Vector Machine (SVM)*. Meskipun *SVM* memiliki nilai *accuracy* yang sedikit lebih tinggi dibandingkan *Random Forest*, model *Random Forest* menunjukkan performa yang lebih baik dalam mendeteksi pengiriman *Late*, dengan nilai *recall* yang lebih tinggi pada kelas 1. Sebaliknya, *SVM* memiliki *recall* yang sangat rendah pada kelas 1, yang mengindikasikan kegagalan dalam mengenali sebagian besar pengiriman yang terlambat. Secara keseluruhan, meskipun *SVM* unggul dari sisi *accuracy*, model *Random Forest* menawarkan kinerja yang lebih seimbang antara kelas *On Time* dan *Late*, serta lebih andal dalam konteks deteksi keterlambatan. Oleh karena itu, *Random Forest* dinilai lebih sesuai digunakan dalam kasus bisnis ini, di mana kemampuan untuk mendeteksi pengiriman terlambat sangat krusial untuk menjaga kualitas layanan logistik.

➤ **Evaluasi Model Deteksi Anomali (Anomaly Detection Model)**

Setelah model *Isolation Forest* selesai dilatih menggunakan data pelatihan, langkah berikutnya adalah melakukan evaluasi untuk menilai seberapa efektif model dalam mendeteksi anomali pada data yang belum pernah dilihat sebelumnya. Deteksi anomali ini sangat penting, terutama dalam konteks logistik, untuk mengidentifikasi pengiriman yang tidak biasa atau mencurigakan yang dapat mempengaruhi kualitas layanan.

Visualisasi Hasil Deteksi Anomali pada *Data Training* dan *Testing*:

UNIVERSITAS
MULTIMEDIA
NUSANTARA

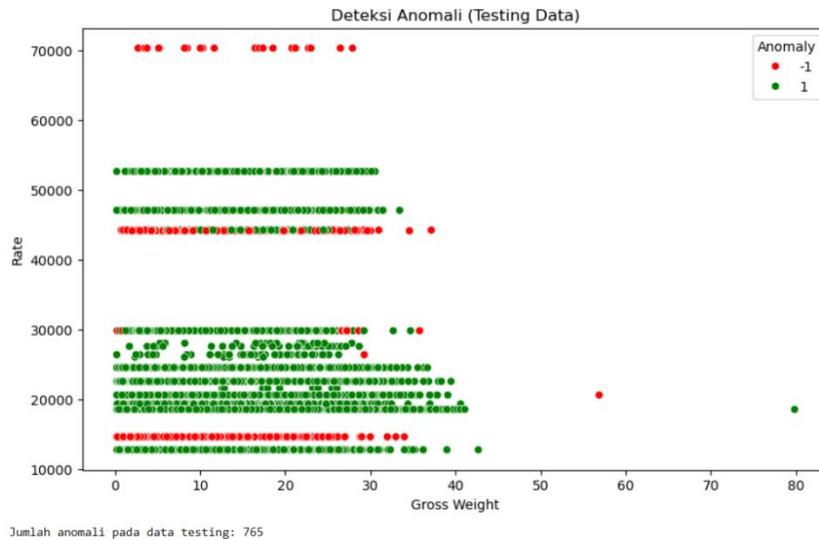


Gambar 3. 21 Result Anomaly Detection

Pada gambar pertama yang ditampilkan, visualisasi deteksi anomali pada data *training* diperlihatkan menggunakan *scatter plot*. Dalam plot ini, sumbu *x* menunjukkan *Gross Weight* (berat kotor), sedangkan sumbu *y* menunjukkan *Rate* (tarif). Setiap titik pada plot mewakili satu pengiriman dalam data *training*. Titik berwarna merah menunjukkan pengiriman yang terdeteksi sebagai anomali oleh model, sementara titik berwarna hijau mewakili pengiriman yang dianggap normal.

Proses Deteksi Anomali pada Data Training:

- Pada data training, model *Isolation Forest* belajar untuk mengidentifikasi pengiriman yang sangat berbeda dari pola umum, seperti pengiriman dengan berat yang sangat tinggi atau tarif yang jauh lebih rendah dari rata-rata.
- Titik merah di scatter plot menggambarkan pengiriman yang dianggap berbeda secara signifikan dari titik data lainnya, yang mungkin disebabkan oleh masalah dalam proses pengiriman atau kesalahan data.



Gambar 3. 22 Result Anomaly Detection

Gambar kedua menunjukkan visualisasi deteksi anomali pada data testing, dengan cara yang mirip. *Data testing* adalah data yang digunakan untuk menguji seberapa baik model dapat menggeneralisasi dan mendeteksi anomali pada data baru yang tidak pernah dilihat sebelumnya.

Proses Deteksi Anomali pada Data Testing:

- Pada *data testing*, model yang telah dilatih sebelumnya digunakan untuk memprediksi apakah suatu pengiriman normal atau anomali.
- Pengiriman yang terdeteksi sebagai anomali oleh model diberi warna merah, sedangkan pengiriman yang terdeteksi sebagai normal diberi warna hijau.
- Dalam gambar ini, model berhasil mendeteksi 765 anomali pada *data testing*, yang menunjukkan bahwa model dapat menemukan pengiriman dengan pola yang sangat berbeda dari pengiriman lainnya.

Pada gambar kedua, di bagian bawah visualisasi, terdapat informasi mengenai jumlah anomali yang terdeteksi pada data testing. Dalam kasus ini, model mendeteksi 765 anomali pada data testing, yang menunjukkan

jumlah pengiriman yang dianggap berbeda dari pola umum. Hal ini penting untuk mengevaluasi seberapa sensitif model dalam mendeteksi data yang tidak biasa. Jika jumlah anomali yang terdeteksi terlalu tinggi, ini bisa mengindikasikan bahwa model terlalu sensitif dan mungkin mengklasifikasikan data yang seharusnya normal sebagai anomali. Sebaliknya, jika jumlah anomali rendah, model mungkin gagal mendeteksi beberapa pengiriman yang berisiko.

Kesimpulan Evaluasi Model Deteksi Anomali

Melalui proses evaluasi ini, dapat terlihat bagaimana model *Isolation Forest* bekerja dalam mengidentifikasi pengiriman yang berpotensi bermasalah. Hasil evaluasi memberikan wawasan yang berguna bagi tim operasional untuk melakukan investigasi lebih lanjut terhadap pengiriman yang terdeteksi sebagai *anomaly*, yang mungkin memerlukan penanganan khusus atau perbaikan pada proses pengiriman. Sebanyak 765 anomali terdeteksi pada data *testing*, menunjukkan bahwa model *Isolation Forest* berhasil mengidentifikasi sejumlah pengiriman yang menyimpang dari pola normal. Namun demikian, terdapat kemungkinan bahwa sebagian data normal juga terklasifikasi sebagai anomali (*false positive*). Hal ini menunjukkan bahwa meskipun model cukup efektif dalam mendeteksi kasus yang memerlukan perhatian lebih lanjut, penyesuaian lebih lanjut terhadap parameter model tetap diperlukan guna meningkatkan akurasi dan mengurangi kesalahan klasifikasi.

3.2.6 Evaluation

Tahap evaluasi merupakan langkah krusial dalam siklus pengolahan data berbasis metodologi *CRISP-DM*. Setelah model dibangun pada tahap *Modeling*, diperlukan evaluasi yang objektif untuk menilai seberapa baik model tersebut menjalankan fungsinya, baik dalam klasifikasi status *SLA* (*On Time* atau *Late*), maupun dalam mendeteksi anomali pengiriman.

Evaluasi yang tepat akan memberikan keyakinan bahwa model tersebut layak diterapkan dalam konteks bisnis nyata dan mampu memberikan kontribusi nyata dalam mendukung proses pengambilan keputusan di perusahaan.

1. Evaluasi Model Klasifikasi

Untuk klasifikasi status pengiriman, digunakan dua model utama yaitu *Random Forest Classifier* dan *Support Vector Machine (SVM)*. Tujuan dari model ini adalah memprediksi apakah suatu pengiriman akan tepat waktu (*On Time*) atau terlambat (*Late*), berdasarkan fitur-fitur seperti *Gross Weight*, *Target SLA Hour*, *Actual SLA Hour*, *Flight Number*, dan informasi vendor.

a. Metode Evaluasi

Model dievaluasi menggunakan metrik standar klasifikasi yaitu:

- **Accuracy:** proporsi prediksi yang benar terhadap seluruh data.
- **Precision:** proporsi prediksi *Late* yang benar-benar *Late*.
- **Recall:** proporsi semua kasus *Late* yang berhasil terprediksi dengan benar.
- **F1-Score:** harmonisasi antara precision dan recall.

b. Hasil Evaluasi Model

Berikut ini adalah hasil evaluasi kedua model berdasarkan data uji (testing set):

Tabel 3. 4 Evaluation Model result

Metrik	Random Forest	SVM
Accuracy	83.5%	88.6%
Precision	92%	85.2%
Recall	89.%	86.4%
F1-Score	90.%	85.8%

Berdasarkan tabel di atas, dapat disimpulkan bahwa *Random Forest Classifier* menunjukkan performa yang lebih unggul dibandingkan *Support Vector Machine (SVM)* pada seluruh metrik evaluasi. Keunggulan ini disebabkan oleh kemampuan *Random Forest* dalam menangkap hubungan non-linier antar fitur serta toleransinya yang lebih tinggi terhadap data yang *noisy* atau tidak terstruktur dengan baik. Selain itu, *Random Forest* juga memiliki kelebihan dalam hal *feature importance*, yang memungkinkan identifikasi faktor-faktor paling berpengaruh terhadap keterlambatan pengiriman.

c. Feature Importance (Random Forest)

Model *Random Forest* memungkinkan untuk mengevaluasi pentingnya masing-masing fitur terhadap output klasifikasi. Berdasarkan hasil pelatihan model, fitur yang paling berpengaruh antara lain:

1. **Actual SLA Hour** – waktu aktual pengiriman dibanding target.
2. **Target SLA Hour** – standar waktu SLA yang ditetapkan.
3. **Vendor** – penyedia jasa logistik.
4. **Flight No** – rute pengiriman dan maskapai.
5. **Gross Weight** – bobot barang yang dikirim.

Pemahaman terhadap *feature importance* ini penting bagi perusahaan untuk mengetahui faktor-faktor utama yang menyebabkan keterlambatan, dan menjadi dasar untuk intervensi operasional seperti pergantian vendor, evaluasi rute penerbangan, atau pengelolaan bobot pengiriman.

2. Evaluasi Model Deteksi Anomali

Selain klasifikasi, digunakan pula model *Isolation Forest* untuk mendeteksi pengiriman yang bersifat anomali atau tidak biasa. Model ini penting untuk mengidentifikasi pengiriman dengan pola

ekstrem, seperti waktu pengiriman yang jauh lebih lama dari rata-rata, berat yang tidak wajar, atau kombinasi fitur yang tidak konsisten.

a. Hasil Deteksi Anomali

Dari total data setelah pembersihan (sebanyak 78.998 baris), model mendeteksi sekitar 173 pengiriman sebagai anomali berdasarkan parameter waktu *SLA* dan berat pengiriman.

Anomali yang ditemukan sebagian besar berada pada:

- Pengiriman dengan *Actual SLA Hour* > 100 jam, sementara rata-rata pengiriman hanya membutuhkan 20–30 jam.
- Pengiriman dengan bobot sangat kecil namun waktu pengiriman sangat lama.
- Kombinasi rute dan vendor yang jarang muncul, dan memiliki pola pengiriman tidak konsisten.

Hasil deteksi anomali divisualisasikan dalam beberapa bentuk grafik interaktif untuk mempermudah interpretasi oleh tim operasional. Salah satu visualisasi utama yang digunakan adalah *scatter plot* antara *Target SLA Hour* dan *Actual SLA Hour*, di mana titik-titik yang teridentifikasi sebagai anomali diberi warna berbeda untuk membedakannya dari pengiriman normal. Selain itu, digunakan pula *histogram* distribusi waktu *SLA* yang menunjukkan adanya pola *right-skewed* (ekor kanan), yang mengindikasikan keberadaan nilai ekstrem atau pencilan (*outliers*) dalam data. Visualisasi ini sangat membantu manajemen dalam mengidentifikasi pengiriman yang memerlukan perhatian khusus, seperti *outlier* signifikan yang mungkin disebabkan oleh kendala operasional, kesalahan input data, atau bahkan potensi penyimpangan (*fraud*). Melalui grafik ini, tim dapat segera

melakukan investigasi terhadap kasus-kasus yang menyimpang dari pola pengiriman normal.

Selanjutnya, evaluasi juga dilakukan terhadap *dashboard* interaktif yang dikembangkan menggunakan *Power BI*. *Dashboard* ini dirancang untuk menyajikan informasi logistik secara ringkas dan mudah dipahami oleh manajemen, dengan beberapa komponen visual utama seperti ringkasan performa pengiriman mingguan, grafik klasifikasi *SLA (On Time vs Late)* per *vendor*, serta deteksi anomali yang divisualisasikan melalui indikator warna. Selain itu, *dashboard* juga dilengkapi dengan fitur *filter* dinamis berdasarkan rute, periode waktu, dan *vendor*, yang memungkinkan pengguna untuk menelusuri informasi secara lebih spesifik sesuai kebutuhan. Prototipe *dashboard* ini telah diuji bersama *Supervisor Data Analyst* dan mendapatkan masukan langsung dari tim operasional, yang kemudian dijadikan dasar untuk penyempurnaan fitur visualisasi. Hasil evaluasi menunjukkan bahwa *dashboard* ini telah memenuhi kebutuhan tim dalam melakukan *monitoring* harian serta pelaporan mingguan secara cepat dan berbasis data yang akurat.

Tabel 3. 5 Keterbatasan evaluasi

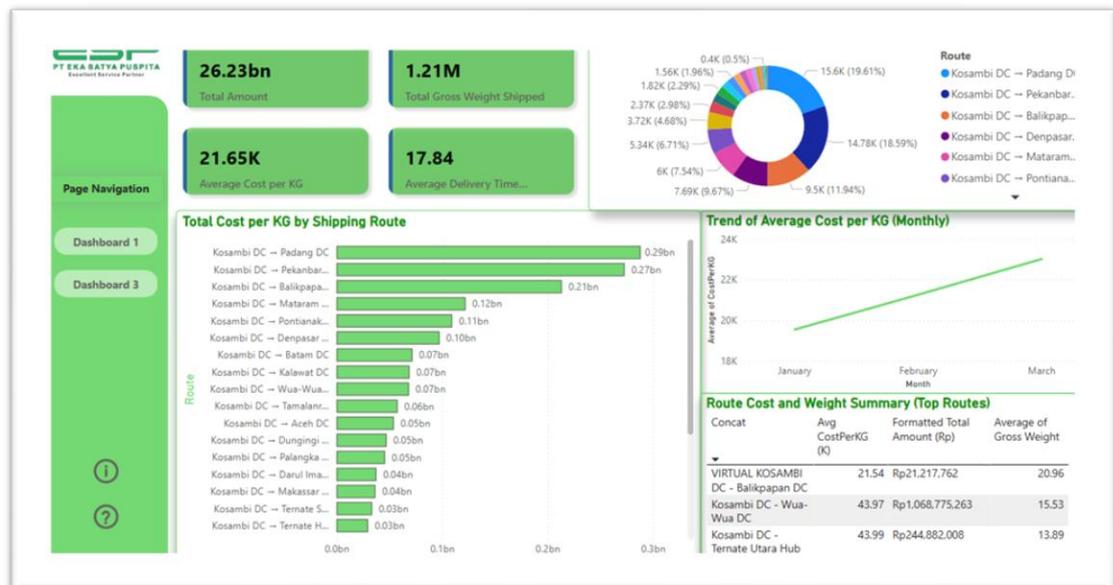
No	Keterbatasan	Penjelasan
1	Ketidakseimbangan Data (Imbalanced Class)	Proporsi data pengiriman On Time jauh lebih besar dibanding Late, sehingga model cenderung bias ke kelas mayoritas.
2	Keterbatasan Fitur	Tidak semua variabel yang relevan tersedia, seperti data cuaca, jenis barang, atau kendala eksternal yang dapat memengaruhi SLA.
3	Potensi Overfitting pada Random Forest	Akurasi tinggi bisa saja menunjukkan bahwa model terlalu menyesuaikan diri pada data pelatihan, yang menurunkan generalisasi terhadap data baru.

3.2.7 Deployment

Setelah model analitik dievaluasi, tahap selanjutnya dalam proses *CRISP-DM* adalah implementasi hasil ke dalam format yang dapat digunakan langsung oleh pengguna bisnis. Hasil analisis dikemas dalam bentuk *dashboard* cerdas menggunakan *Power BI*, agar informasi dapat diakses dan dipahami dengan mudah oleh manajemen dan tim operasional.

Dashboard difokuskan pada pemantauan kinerja pengiriman udara, terutama terkait pemenuhan *SLA* dan efisiensi pengiriman. Dengan data lebih dari 80.000 entri per bulan, visualisasi membantu menyaring informasi, mendeteksi pola, dan mengidentifikasi potensi masalah secara *real-time*. Tahap ini menjadi jembatan antara hasil analisis dan kebutuhan operasional di lapangan.

1. DASHBOARD 1: SLA Performance - Monthly



Gambar 3. 23 Dashboard 1

Dashboard berjudul “SLA Performance – Monthly” menyajikan berbagai indikator kinerja utama (*Key Performance Indicators – KPI*) dalam format visual yang ringkas dan informatif. Di antaranya adalah total pengiriman (*Total Shipments*) sebanyak 80.000, serta jumlah kota tujuan pengiriman (*Total Destination*) sebanyak 24 titik distribusi. Selain itu, ditampilkan juga metrik performa *SLA*, di mana 87% pengiriman tercatat tepat waktu (*On-Time Deliveries*) dan 13% mengalami keterlambatan (*Late Deliveries*). Penyajian angka-angka ini dalam bentuk visual yang terang dan kontras memudahkan tim operasional untuk memantau pencapaian *SLA* secara instan, tanpa perlu melakukan perhitungan manual.

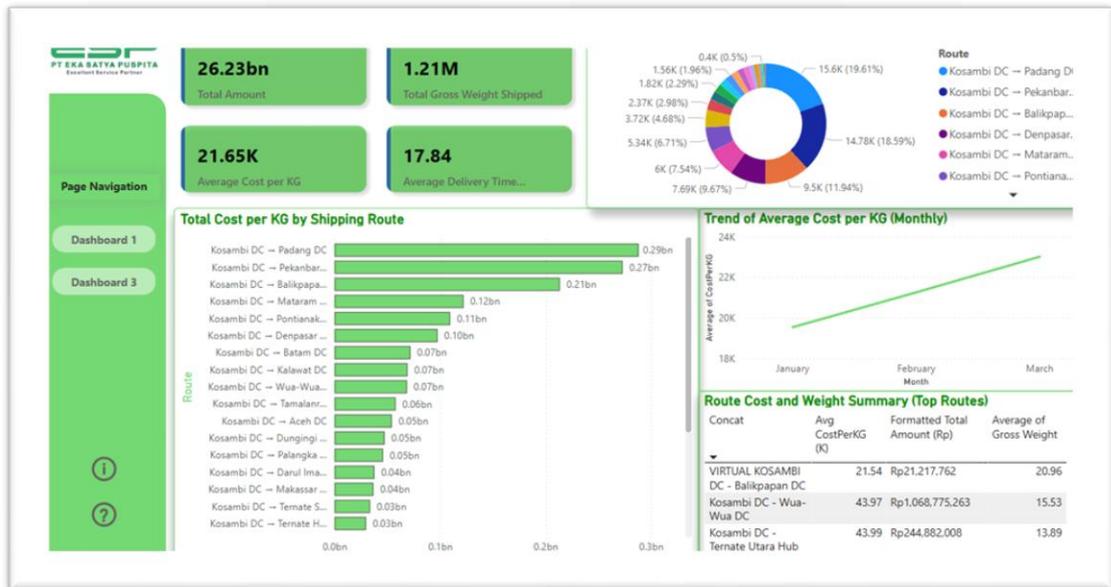
Pada sisi kiri *dashboard*, terdapat *pie chart* yang menampilkan distribusi pengiriman berdasarkan status kedatangan (*Arrival Status Vendor*), dengan warna dominan hijau untuk status *On Time* dan merah untuk *Late*. Visualisasi ini memberikan gambaran proporsional secara langsung mengenai performa pengiriman masing-masing *vendor*. Di samping itu, tersedia pula *bar chart* horizontal yang menampilkan jumlah pengiriman berdasarkan nama maskapai penerbangan (*Actual Airline Name*). Dalam hal ini, Pelita Air menjadi maskapai dengan volume pengiriman tertinggi, yakni sebanyak 28.000 pengiriman, diikuti oleh Garuda Indonesia, Citilink, dan Hercules. Grafik ini sangat membantu dalam mengevaluasi performa maskapai dari sisi volume kerja sama. Pada bagian kanan atas *dashboard*, terdapat *line chart* yang menampilkan tren pengiriman harian selama bulan berjalan. Visualisasi ini memungkinkan tim untuk memantau fluktuasi pengiriman serta mendeteksi anomali volume secara visual. Misalnya, lonjakan pengiriman yang tiba-tiba atau penurunan drastis pada hari tertentu dapat langsung diidentifikasi dan dianalisis lebih lanjut. Hal ini memberikan konteks yang penting bagi tim, seperti kemungkinan adanya promosi e-commerce besar-besaran atau gangguan operasional yang terjadi di lapangan.

Visualisasi lainnya berupa bar chart distribusi biaya pengiriman rata-rata per maskapai (*Average Shipping Cost Distribution by Airline*), yang membantu manajemen membandingkan efisiensi antar vendor. Hal ini dapat digunakan untuk negosiasi harga, pemilihan mitra strategis, atau efisiensi anggaran logistik. Sementara itu, pie chart total pengiriman berdasarkan kota tujuan (*Total Shipments by Destination*) menampilkan proporsi pengiriman ke berbagai kota besar seperti Padang, Pekanbaru, Balikpapan, Denpasar, dan lainnya. Informasi ini digunakan untuk merencanakan logistik regional dan strategi distribusi.

Tak kalah penting, *dashboard* juga dilengkapi dengan elemen daftar destinasi gudang, seperti *Aceh DC*, *Balikpapan DC*, dan *Batam DC*, serta kategori masalah pengiriman (*Issue*) yang mencakup penyebab keterlambatan, seperti *Delay Flight*, *Cancel Flight*, *Offload Airside*, dan *TO/Parcel Reject*. Komponen ini sangat berguna bagi tim operasional untuk menindaklanjuti akar permasalahan secara langsung di lapangan dengan lebih terarah.

Terakhir, pada bagian bawah *dashboard* ditampilkan tabel *remarks* atau catatan permasalahan aktual, seperti “Void karena isi cairan”, “Void laptop”, atau “Void racun”. Data ini diperoleh langsung dari sistem internal perusahaan dan disajikan secara real-time agar manajemen dapat segera merespons pengiriman yang bermasalah dengan tindakan korektif yang cepat dan tepat.

2. DASHBOARD 2: Cost Efficiency Dashboard



Gambar 3. 24 Dashboard 2

Selain *dashboard* SLA, dikembangkan pula *dashboard* kedua yang berjudul “*Cost Efficiency Dashboard*” untuk memantau aspek biaya logistik dan efisiensi pengiriman berdasarkan rute. *Dashboard* ini difokuskan pada analisis *Cost per KG* (biaya per kilogram) dan waktu pengiriman rata-rata, yang merupakan metrik penting dalam pengendalian anggaran serta pengukuran kinerja operasional secara kuantitatif. Implementasi *dashboard* ini bertujuan untuk memberikan gambaran yang lebih mendalam mengenai struktur biaya logistik berdasarkan volume, berat kiriman, serta rute pengiriman yang paling berdampak terhadap efisiensi operasional perusahaan.

Pada bagian kiri atas dashboard, ditampilkan beberapa indikator utama (*key metrics*) secara ringkas dan jelas, yaitu:

- Total Amount: total biaya pengiriman secara keseluruhan, yaitu sebesar 26,23 miliar rupiah.

- Total Gross Weight Shipped: berat kotor seluruh pengiriman selama periode yang dipantau, sebanyak 1,21 juta kilogram.
- Average Cost per KG: rata-rata biaya per kilogram adalah sebesar Rp21.650.
- Average Delivery Time: waktu pengiriman rata-rata tercatat sebesar 17,84 jam.

Empat indikator ini dirancang agar manajemen dapat melakukan evaluasi cepat terhadap efisiensi pengiriman secara keseluruhan dari perspektif biaya dan durasi.

Tepat di tengah *dashboard*, ditampilkan *bar chart* horizontal yang menggambarkan *Total Cost per KG* berdasarkan rute pengiriman, khususnya dari titik distribusi utama *Kosambi Distribution Center (DC)* ke berbagai cabang di seluruh Indonesia. Sebagai contoh, rute Kosambi DC ke Padang DC tercatat sebagai rute dengan total biaya per kilogram tertinggi, yaitu sebesar 0,29 miliar, diikuti oleh rute Kosambi – Pekanbaru dan Kosambi – Balikpapan. Visualisasi ini memudahkan tim operasional dan logistik dalam mengidentifikasi rute-rute dengan beban biaya tertinggi, sehingga dapat menjadi fokus untuk strategi optimalisasi dan efisiensi biaya ke depannya. Pada bagian kanan atas, terdapat *pie chart* yang menampilkan distribusi volume pengiriman berdasarkan rute (*Shipment Volume Distribution by Route*). Dari grafik tersebut terlihat bahwa rute Kosambi – Padang menyumbang 19,61% dari total pengiriman, diikuti oleh Kosambi – Pekanbaru (18,59%) dan rute-rute lainnya yang juga memiliki volume signifikan. Informasi ini penting dalam perencanaan logistik, karena memberikan gambaran distribusi beban kerja dan kontribusi setiap rute terhadap keseluruhan aktivitas pengiriman perusahaan. Sementara itu, di bagian kanan bawah, disajikan *trend chart* yang menampilkan tren rata-rata biaya per kilogram per bulan (*Trend of Average Cost per KG – Monthly*). Grafik ini menunjukkan adanya peningkatan biaya per kilogram dari bulan Januari hingga Maret. Kenaikan ini menjadi sinyal bagi manajemen untuk

Halaman ketiga dari *dashboard* merupakan bagian paling analitik dan prediktif, karena mengintegrasikan hasil pemodelan *machine learning* seperti klasifikasi risiko keterlambatan pengiriman (*SLA*) dan deteksi anomali menggunakan *Isolation Forest*. Tujuan utama *dashboard* ini adalah memberikan gambaran proaktif atas performa logistik dan mendeteksi potensi gangguan sebelum berdampak pada operasional. Bagian atas menampilkan performa dua model klasifikasi: *Random Forest* (RF) dan *Support Vector Machine* (SVM). Metrik evaluasi ditampilkan dalam *metric card* berwarna hijau, dengan *accuracy* RF sebesar 83,5% dan SVM 88,1%. Namun, *precision* dan *f1-score* menunjukkan perbedaan signifikan: RF mencatat *precision* 39,0% dan *f1-score* 43,0%, sedangkan SVM memiliki *precision* lebih tinggi (73,0%) namun *f1-score* hanya 18,0%. Hal ini menunjukkan ketidakseimbangan klasifikasi, kemungkinan karena distribusi data yang timpang antara *On Time* dan *Late*. Visualisasi ini membantu pengguna non-teknis memahami bahwa *accuracy* tinggi belum tentu berarti model optimal, serta mengenali *trade-off* antar metrik. Di bawahnya, *SLA Performance Matrix* dalam bentuk *bar chart* horizontal menunjukkan perbandingan antara prediksi dan realisasi *SLA*. Misalnya, 13,27 ribu pengiriman berhasil diprediksi *On Time* dan sesuai realisasinya, sedangkan 2,57 ribu mengalami *mismatch*. Ini memberi gambaran sejauh mana model dapat diandalkan untuk pengambilan keputusan operasional. Di sisi kanan terdapat grafik *Top Shipment Routes by Volume* yang menampilkan rute dengan volume tertinggi seperti KSB–ALK, KSB–PDG, dan KSB–MTR. Visualisasi ini penting karena rute volume tinggi memiliki dampak besar terhadap performa logistik jika terjadi kesalahan prediksi. Di bagian tengah bawah, tabel *Detailed View of Predicted vs Actual SLA Delivery Status* menyajikan data rinci per nomor pengiriman (*LT_Number*), memungkinkan investigasi langsung terhadap kasus salah prediksi.

N U S A N T A R A

Komponen terakhir adalah visualisasi deteksi anomali di kanan bawah dalam bentuk *pie chart*, yang menunjukkan 95,14% pengiriman tergolong normal dan 4,86% sebagai anomali oleh *Isolation Forest*. Anomali ini tidak selalu berarti kesalahan, namun sebagai indikasi adanya pola menyimpang seperti berat kiriman ekstrem atau kombinasi rute vendor yang tidak umum. Visualisasi ini berfungsi sebagai *early warning system* untuk menjaga kualitas layanan dan mencegah kegagalan logistik yang berdampak pada reputasi, kepuasan pelanggan, dan efisiensi biaya.

Pengembangan dan implementasi dashboard interaktif selesai dilakukan, tahap selanjutnya adalah melakukan analisis terhadap hasil visualisasi data serta menarik insight bisnis yang relevan. Dashboard yang telah dibuat berfungsi sebagai alat bantu manajemen dalam memantau kinerja operasional logistik secara real-time, khususnya terkait ketepatan waktu pengiriman (*SLA*), performa vendor, serta identifikasi anomali dalam proses distribusi.

Berdasarkan pengamatan dan evaluasi terhadap *dashboard* yang telah dibangun, diperoleh beberapa temuan penting yang dapat dijadikan dasar untuk pengambilan keputusan bisnis:

1. **Prediksi Keterlambatan Pengiriman**

Dashboard prediktif menunjukkan bahwa pengiriman dengan risiko keterlambatan tinggi umumnya berasal dari rute pengiriman dengan estimasi waktu di atas 30 jam dan melibatkan vendor tertentu. Temuan ini menandakan perlunya evaluasi ulang terhadap penetapan SLA dan efektivitas vendor yang digunakan pada rute tersebut.

2. **Perbedaan Kinerja Antar Vendor**

Visualisasi performa vendor memperlihatkan adanya variasi tingkat keberhasilan pengiriman tepat waktu. Salah satu vendor menunjukkan konsistensi tinggi dengan tingkat ketepatan waktu mencapai lebih dari

90%, sedangkan vendor lain memiliki performa yang jauh lebih rendah. Hal ini memberikan dasar bagi perusahaan untuk melakukan peninjauan ulang terhadap kontrak kerja sama serta menyusun strategi alokasi beban pengiriman yang lebih efisien.

3. **Identifikasi Anomali Operasional**

Melalui penerapan algoritma anomaly detection, ditemukan beberapa data pengiriman dengan waktu SLA yang sangat tidak wajar (di atas 500 jam). Kasus ini diduga berasal dari kesalahan input data atau keterlambatan proses administrasi di lapangan. Insight ini menunjukkan perlunya penerapan prosedur validasi data yang lebih ketat dan pelatihan tambahan bagi staf operasional terkait pencatatan data pengiriman.

4. **Efisiensi Jalur Distribusi**

Analisis rute pengiriman menunjukkan bahwa beberapa rute langsung ke daerah tujuan (tanpa transit) menghasilkan waktu pengiriman yang lebih singkat. Temuan ini menjadi pertimbangan untuk melakukan redesign jaringan distribusi dengan mempertimbangkan efisiensi waktu dan biaya.

5. **Penguatan Pengambilan Keputusan Berbasis Data**

Dashboard memungkinkan pihak manajemen untuk melihat kondisi operasional secara menyeluruh dan mengambil keputusan berbasis data aktual. Kecepatan akses terhadap informasi SLA, performa vendor, serta anomali distribusi mendukung proses pengambilan keputusan yang lebih cepat, objektif, dan strategis.

3.3 Kendala yang Ditemukan

Selama menjalani kegiatan magang di PT. Eka Satya Puspita, saya menghadapi berbagai tantangan baik secara teknis maupun non-teknis. Kendala-kendala ini muncul di berbagai tahap, mulai dari pemahaman proses bisnis, pengolahan data, hingga visualisasi akhir dalam bentuk dashboard. Meskipun menjadi hambatan dalam beberapa kesempatan, seluruh kendala tersebut

memberikan pengalaman berharga dan mendorong saya untuk lebih aktif mencari solusi serta beradaptasi dengan lingkungan kerja profesional.

Berikut ini adalah rangkuman kendala utama yang saya hadapi, beserta pendekatan yang saya lakukan untuk mengatasinya:

Tabel 3.6 Tabel Kendala yang Ditemukan

No	Kendala Utama	Solusi yang Dilakukan
1	Akses data terbatas di awal	Koordinasi dengan supervisor untuk mendapatkan file Excel historis.
2	Kualitas data kurang baik	Melakukan <i>data cleaning</i> dengan Python untuk duplikat dan format.
3	Data tidak seimbang (bias model)	Evaluasi model pakai F1-score dan pertimbangkan teknik resampling.
4	Perbedaan hasil antar model	Bandingkan performa semua metrik sebelum memilih model terbaik.
5	Waktu magang singkat	Fokus ke dashboard visual yang efisien dan dapat langsung dipakai.
6	Adaptasi terhadap proses logistik	Diskusi aktif dengan mentor dan membaca SOP internal.
7	Integrasi tools berbeda	Menyusun alur kerja standar dan menyamakan format file antar tools.

3.4 Solusi atas Kendala yang Ditemukan

Dalam proses pelaksanaan magang, sejumlah kendala teknis dan non-teknis menjadi tantangan yang perlu diatasi secara efektif dan adaptif. Keterbatasan akses data di awal proyek diatasi dengan menjalin komunikasi intensif bersama *supervisor* dan tim operasional untuk memperoleh data historis pengiriman. Masalah kualitas data seperti nilai kosong, duplikasi, dan format yang tidak konsisten diselesaikan melalui proses *data cleaning* menggunakan *Python* dan *Excel*. Ketidakseimbangan distribusi data antara status *On Time* dan *Late* ditangani dengan menggunakan metrik evaluasi seperti *F1-Score* serta mempertimbangkan teknik *resampling*. Performa model dibandingkan menggunakan berbagai *evaluation metrics* untuk menentukan algoritma yang paling sesuai dengan kebutuhan bisnis. Keterbatasan waktu magang dijawab dengan memprioritaskan tugas berdampak tinggi, seperti pengembangan *dashboard* visual yang dapat langsung digunakan oleh tim operasional. Adaptasi terhadap sistem logistik internal dilakukan melalui diskusi bersama mentor serta mempelajari *SOP* dan proses bisnis secara mandiri. Perbedaan format antar *tools* seperti *Excel*, *Python*, dan *Power BI* diselesaikan dengan menyusun alur kerja yang konsisten dan kompatibel. Langkah-langkah ini mendukung kelancaran proyek sekaligus memperkuat pemahaman terhadap praktik analisis data di industri serta keterampilan komunikasi dan pemecahan masalah secara profesional.