y LAPORAN AKHIR

FINAL PROJECT

PREDICTING PACKAGE STATUS IN E-COMMERCE LOGISTICS

Diajukan untuk memenuhi persyaratan kelulusan

Bootcamp Data Science Rakamin

Disusun oleh :

Mardio Edana Putra

Mochamad Rizqi

Muhammad Gusti Maulana Alfin

Nena Herdiani



**LAST VOYAGE**

**TAHUN 2024**

# ABSTRAKSI

Proyek ini dilaksanakan sebagai bagian dari Final Project Bootcamp Rakamin dengan tujuan membantu Amazon, perusahaan e-commerce global yang berfokus pada penjualan produk retail dan layanan digital, untuk mengelola proses pengembalian barang secara lebih efektif. Amazon Sales Dataset digunakan dalam proyek ini, yang mencakup data terkait transaksi, pengiriman, dan informasi logistik lainnya.

Misi utama proyek ini adalah membangun model prediksi yang dapat mendeteksi risiko pengembalian barang, sehingga memungkinkan Amazon untuk mengambil tindakan preventif dan meningkatkan efisiensi operasional. Dengan demikian, proyek ini diharapkan dapat memberikan dampak positif terhadap pengalaman pelanggan dan pengelolaan logistik.

Pelaksanaan proyek dimulai dengan eksplorasi dataset untuk memahami pola data, dilanjutkan dengan pemrosesan data seperti pembersihan dan analisis eksploratori. Model machine learning yang digunakan adalah XGBoost, yang dipilih berdasarkan kemampuannya dalam menangani dataset kompleks. Model ini berhasil mencapai recall sebesar 83% pada tahap awal dan meningkat menjadi 95% setelah tuning hyperparameter, menunjukkan kemampuan yang tinggi dalam mendeteksi paket yang berisiko untuk dikembalikan.

Fitur-fitur utama yang memengaruhi prediksi meliputi wilayah pengiriman (misalnya, South dan West), kategori produk seperti pakaian etnik dan pakaian bawah, serta waktu pengiriman pada akhir pekan. Hasil dari model ini menunjukkan bahwa implementasi sistem prediksi dapat mengurangi potensi pengembalian barang hingga 15%, sekaligus mendukung pengambilan keputusan strategis di Amazon.

Secara keseluruhan, kegiatan Final Project Bootcamp Rakamin ini memberikan pengalaman praktis dalam penerapan machine learning untuk menyelesaikan permasalahan bisnis nyata, khususnya di industri e-commerce. Keberhasilan proyek ini mencerminkan pentingnya kolaborasi, analisis data yang mendalam, serta pemanfaatan teknologi untuk menghasilkan solusi yang berdampak.

Kata kunci: *Amazon, e-commerce, machine learning, XGBoost, prediksi pengembalian barang, efisiensi logistik.*

# KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga kami dapat menyelesaikan laporan Final Project Bootcamp Rakamin ini dengan baik. Laporan ini disusun sebagai salah satu bentuk tanggung jawab kami dalam mengaplikasikan ilmu dan keterampilan yang telah kami pelajari selama mengikuti program Bootcamp Rakamin.

Kami menyadari bahwa laporan ini tidak akan terselesaikan tanpa dukungan dan kontribusi dari berbagai pihak. Oleh karena itu, dengan rasa hormat dan tulus, kami mengucapkan terima kasih kepada:

1. **Mentor, Dedi Irwanto Samosir**, atas bimbingan, masukan, dan arahan yang sangat berharga selama proses pengerjaan proyek ini.
2. **Rekan-rekan satu tim**:
   * **Mardio Edana Putra** (Project Manager) atas kepemimpinan dan koordinasi yang efektif.
   * **Mochamad Rizqi** (Data Engineer) atas dedikasinya dalam mempersiapkan data yang berkualitas.
   * **Muhammad Gusti Maulana Alfin** (Data Scientist) atas pengembangan model machine learning yang optimal.
   * **Nena Herdiani** (Data/Business Analyst) atas analisis data dan visualisasi yang membantu memahami wawasan dari data.
3. **Instruktur dan staf Bootcamp Rakamin**, atas penyelenggaraan program yang memberikan pengalaman belajar yang bermanfaat dan mendalam.
4. **Keluarga dan teman-teman**, yang selalu memberikan dukungan moral dan semangat kepada kami selama mengikuti program ini.

Kami menyadari bahwa laporan ini masih memiliki kekurangan, dan kami sangat terbuka untuk menerima masukan serta kritik yang membangun demi penyempurnaan di masa mendatang. Semoga laporan ini dapat bermanfaat bagi pihak-pihak yang memerlukan dan menjadi kontribusi positif dalam pengembangan ilmu pengetahuan, khususnya di bidang data dan teknologi.

Akhir kata, kami berharap kerja keras dan hasil yang telah dicapai ini dapat menjadi langkah awal yang baik untuk menghadapi tantangan di masa depan.

**Jakarta, 8 Januari 2025**  
Kelompok 1 Last Voyage

# DAFTAR ISI

[ABSTRAKSI 1](#_2et92p0)

[KATA PENGANTAR 2](#_tyjcwt)

[DAFTAR ISI 3](#_3dy6vkm)

[DAFTAR GAMBAR 4](#_1t3h5sf)

[DAFTAR TABEL 4](#_2s8eyo1)

[DAFTAR LAMPIRAN 4](#_17dp8vu)

[BAB I PENDAHULUAN 5](#_3rdcrjn)

[1.1](#_lnxbz9) Latar belakang 5

[1.2](#_35nkun2) Lingkup 5

[1.3](#_44sinio) Tujuan 6

[BAB II KELOMPOK FINAL PROJECT LAST VOYAGE 9](#_2jxsxqh)

[2.1](#_z337ya) Struktur Organisasi 9

[2.2](#_4i7ojhp) Lingkup Pekerjaan 10

[2.3](#_2xcytpi) Deskripsi Pekerjaan 11

[2.4](#_3whwml4) Jadwal Kerja 11

[BAB III Predicting Package Status in E-Commerce Logistics 15](#_qsh70q)

[3.1](#_1pxezwc) Deskripsi Persoalan 15

[3.2](#_2p2csry) Proses Pelaksanaan 15

[3.3](#_23ckvvd) Pencapaian Hasil 16

[BAB IV PENUTUP 21](#_2grqrue)

[4.1](#_3fwokq0) Kesimpulan 21

[4.2](#_1v1yuxt) Saran 22

[4.3 Referensi 24](#_4f1mdlm)

[BAB V 25](#_2u6wntf)

[Lampiran A PPT Final Project Presentation 25](#_19c6y18)

[Lampiran B Notulensi Mentoring 33](#_3tbugp1)

[Lampiran C Dokumen Teknik 36](#_28h4qwu)

# DAFTAR GAMBAR

[**Gambar II- 1 Struktur Organisasi Kelompok Last Voyage** 9](#_3j2qqm3)

[**Gambar II- 2 Jadwal Kerja** 12](#_2bn6wsx)

[**Gambar III- 1 Alur pemrosesan dataset** 16](#_3o7alnk)

[**Gambar III- 2 Feature Importance XGboost** 19](#_1hmsyys)

[**Gambar III- 3 Penerapan Machine Learning pada Proses Bisnis** 19](#_41mghml)

# DAFTAR TABEL

[**Tabel III- 1 Hasil evaluasi 4 Model** 18](#_ihv636)

[**Tabel III- 2 Hasil evaluasi model dengan tuning hyperparameter** 18](#_32hioqz)

# DAFTAR LAMPIRAN

[**Lampiran A PPT Final Project Presentation 25**](#_19c6y18)

[**Lampiran B Notulensi Mentoring 33**](#_3tbugp1)

[**Lampiran C Dokumen Teknik 36**](#_28h4qwu)

# BAB I PENDAHULUAN

## Latar belakang

Revolusi Industri 4.0 menyediakan perubahan di berbagai macam sektor, salah satunya adalah digitalisasi ekonomi. Salah satu sektor dengan pertumbuhan tercepat dalam digitalisasi ekonomi adalah *E-Commerce.* Data dari **Statista** menyebutkan bahwa penjualan ritel online global diperkirakan mencapai $6,3 triliun pada tahun 2024. Berdasarkan laporan dari Google dengan judul “e-Conomy SEA 2023”, E-Commerce memiliki nilai GMV dan Revenue tertinggi di setiap tahun dan diprediksi sama di tahun 2025 (Google & Temasek, 2023). Oleh sebab itu, pemilihan tema mengenai E-Commerce menarik perhatian kelompok kami dan dapat memberikan insight yang bermanfaat bagi sebuah perusahaan yang bergerak di bidang tersebut.

Bisnis *e-commerce* telah mengalami pertumbuhan pesat dalam beberapa tahun terakhir, didorong oleh kemajuan teknologi dan perubahan perilaku konsumen. Konsumen kini lebih memilih berbelanja online karena kenyamanan, pilihan produk yang lebih banyak, dan kemudahan dalam membandingkan harga. Namun, meskipun memiliki peluang yang luas, bisnis e-commerce juga menghadapi sejumlah tantangan yang dapat memengaruhi keberhasilan dan pertumbuhan mereka.

Tantangan dalam bisnis *E-Commerce* sangat bervariasi dan terus berkembang seiring dengan pertumbuhan industri ini. Beberapa tantangan yang sering terjadi adalah ketatnya persaingan, pengelolaan logistik dan pengiriman barang, keamanan data, pemasaran dan penjangkauan pelanggan, dan lain-lain. Oleh sebab itu, dengan memahami dan mengatasi tantangan ini, perusahaan e-commerce dapat membangun pondasi yang kuat untuk pertumbuhan dan keberhasilan di pasar yang kompetitif.

## Lingkup

Lingkup proyek ini mencakup analisis data penjualan dari platform *e-commerce* Amazon untuk mengidentifikasi pola dan faktor yang memengaruhi status pengiriman barang. Fokus utama adalah memprediksi status pengiriman (***delivered, returned to seller*, dan *rejected***) dengan menggunakan pendekatan *machine learning*. Proyek ini juga mencakup:

* Pemrosesan data untuk memastikan dataset bersih, lengkap, dan siap digunakan dalam analisis serta pemodelan.
* Analisis eksplorasi untuk memahami distribusi data, pola, dan hubungan antara variabel.
* Pengembangan model prediktif dengan tujuan utama meningkatkan metriks evaluasi model *machine learning* yaitu *recall* dalam mendeteksi status "*returned to seller*" dan "*rejected.*"
* Identifikasi fitur-fitur yang memengaruhi status pengiriman barang.
* Memberikan rekomendasi *actionable* kepada perusahaan berdasarkan hasil analisis untuk meningkatkan efisiensi operasional dan pengalaman pelanggan.

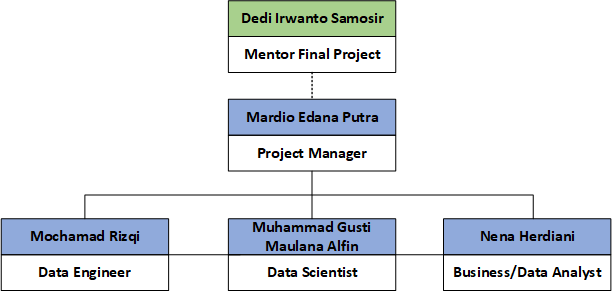
## Tujuan

Tujuan mengikuti Final Project Bootcamp Rakamin adalah untuk:

1. Meningkatkan pemahaman dalam pemrosesan data, analisis, dan pengembangan model *machine learning* yang aplikatif dalam hal tema yang dipilih yaitu di industri *e-commerce*.
2. Mengasah keterampilan dalam menyelesaikan permasalahan bisnis berbasis data, seperti mengidentifikasi pola, tren, dan *business insight* yang relevan.
3. Memahami proses pengambilan keputusan berbasis data, dari identifikasi masalah hingga pembuatan rekomendasi bisnis yang *actionable*.
4. Meningkatkan kemampuan teknis dalam penggunaan *software* seperti Python, SQL, dan visualisasi data menggunakan *tools* seperti Tableau atau Power BI.
5. Membangun pengalaman praktis dalam kerja tim lintas fungsi, yang melibatkan peran seperti Data Scientist, Data Engineer, dan Business Analyst.
6. Menghasilkan proyek akhir yang dapat digunakan sebagai portofolio untuk karier di bidang data science atau analitik bisnis.

# BAB II KELOMPOK FINAL PROJECT LAST VOYAGE

## Struktur Organisasi



**Gambar II- 1 Struktur Organisasi Kelompok Last Voyage**

**Penjelasan Posisi:**

1. **Mentor Proyek (Dedi Irwanto Samosir)**: Bertanggung jawab memberikan arahan, bimbingan, dan evaluasi terhadap seluruh tahapan proyek, serta memastikan kelancaran implementasi proyek sesuai dengan tujuan yang ditetapkan.
2. **Project Manager (Mardio Edana Putra)**: Memimpin tim dan mengawasi proyek secara keseluruhan, mulai dari perencanaan hingga pelaksanaan. Project Manager juga bertugas memastikan bahwa proyek berjalan sesuai jadwal serta menjadi penghubung antara tim dan mentor.
3. **Data Engineer (Mochamad Rizqi)**: Bertanggung jawab dalam menyiapkan dan memanipulasi data, memastikan bahwa data yang digunakan oleh tim dalam proyek berkualitas dan siap untuk dianalisis dan dilakukan pemodelan.
4. **Data Scientist (Muhammad Gusti Maulana Alfin)**: Berperan dalam membangun model prediktif dan analisis data yang mendalam untuk menghasilkan wawasan yang berguna bagi keputusan bisnis yang diambil.
5. **Data/Business Analyst (Nena Herdiani)**: Berperan dalam menganalisis data bisnis dan memberikan rekomendasi berdasarkan hasil analisis yang relevan dengan tujuan bisnis proyek.

**Pembagian Unit atau Divisi:**

* **Tim Proyek (Shaded Area)**: Project Manager, Data Engineer, Data Scientist, dan Data/Business Analyst tergabung dalam satu unit proyek, yang memiliki fokus pada penyelesaian proyek dan pencapaian tujuan akhir.
* **Mentor Proyek (Garis Putus-Putus)**: Mentor proyek bertindak sebagai pihak eksternal yang memberikan dukungan dan pengarahan.

## Lingkup Pekerjaan

Berikut adalah lingkup pekerjaan dari setiap anggota yang ada di dalam kelompok Final Project.

1. **Project Manager (Mardio Edana Putra)**

* **Lingkup Pekerjaan:**

Mengawasi seluruh alur proyek, membagikan pengerjaan tugas, memastikan tugas selesai tepat waktu, mengelola komunikasi tim, dan menjaga kualitas hasil proyek.

* **Kaitan dengan Project:**

Berperan dalam perencanaan dan eksekusi proyek untuk memastikan tujuan analitik dan bisnis tercapai sesuai jadwal. Bertanggung jawab memastikan keluaran proyek sesuai kebutuhan pemangku kepentingan.

1. **Data Engineer (Mochamad Rizqi)**

* **Lingkup Pekerjaan:**

Mengelola data mentah, membersihkan, dan mempersiapkan dataset yang diberikan agar siap digunakan untuk analisis dan pengembangan model.

* **Kaitan dengan Project:**

Mendukung pengembangan model prediktif dengan menyediakan data yang berkualitas dan infrastruktur yang efisien. Ini mencakup ekstraksi fitur dan integrasi data dari berbagai sumber yang disediakan.

1. **Data Scientist (Muhammad Gusti Maulana Alfin)**

* **Lingkup Pekerjaan:**

Membuat model *machine learning* untuk memprediksi status pengiriman barang, mengevaluasi performa model, dan memastikan hasil yang optimal.

* **Kaitan dengan Project:**

Fokus pada pengembangan model prediktif dengan tingkat *recall* yang tinggi untuk mendeteksi status pengiriman bermasalah (*returned/rejected*) dan memberikan wawasan strategis berbasis data.

1. **Business and Data Analyst (Nena Herdiani)**

* **Lingkup Pekerjaan:**

Melakukan analisis data untuk menemukan wawasan bisnis yang *actionable* dan membuat visualisasi data yang mudah dipahami pemangku kepentingan.

* **Kaitan dengan Project:**

Berperan dalam menerjemahkan hasil model dan analisis data menjadi rekomendasi strategis untuk meningkatkan efisiensi proses pengiriman dan kepuasan pelanggan.

**Kesimpulan:**Lingkup pekerjaan setiap anggota terintegrasi dengan tujuan proyek untuk memprediksi status pengiriman barang dan memberikan rekomendasi berbasis data yang relevan dengan tantangan operasional *e-commerce.*

***Kolaborasi Tim:***

Meskipun setiap anggota memiliki peran dan tanggung jawab utama masing-masing, kami saling membantu satu sama lain untuk memastikan proyek berjalan lancar. Pendekatan kolaboratif ini memungkinkan kelompok kami untuk mengatasi hambatan bersama, berbagi wawasan, dan mencapai hasil yang maksimal dengan memanfaatkan keahlian dari setiap anggota.

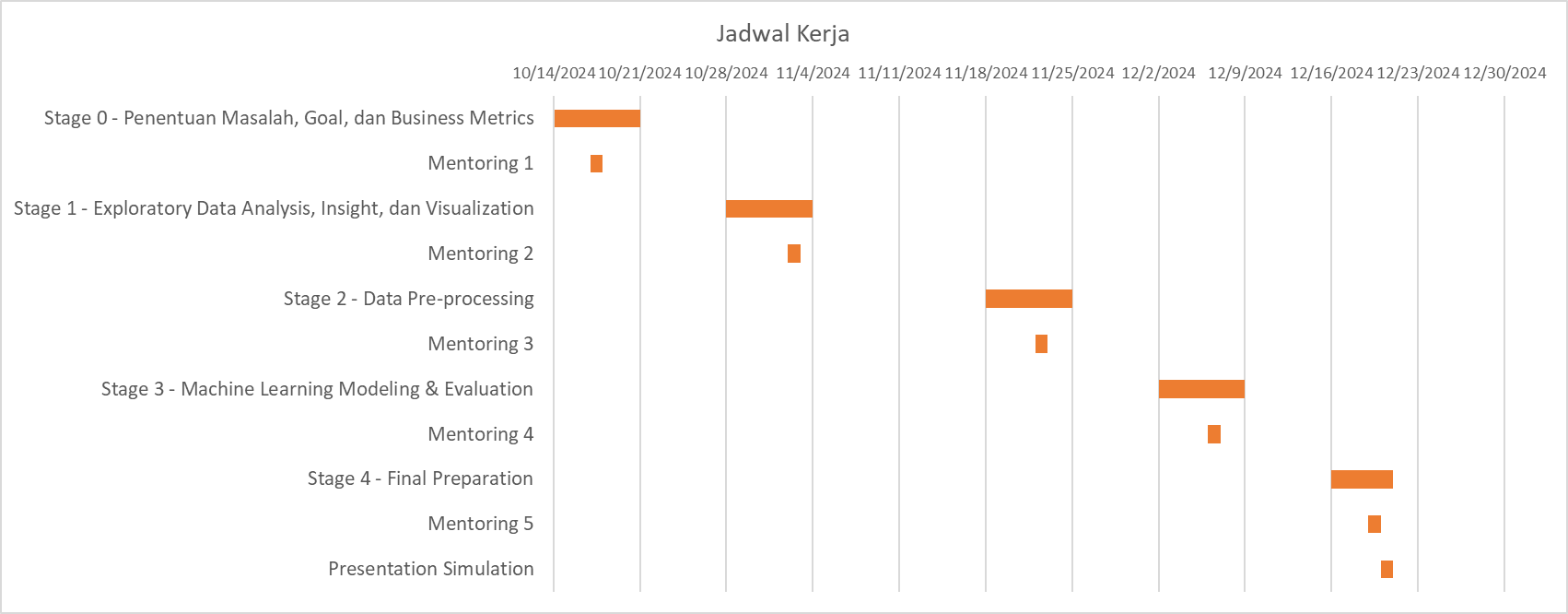
## Deskripsi Pekerjaan

Deskripsi pekerjaan secara lebih rinci adalah sebagai berikut:

1. **Pemrosesan Data**: Membersihkan data, mengatasi nilai yang hilang atau *outlier*, dan melakukan encoding pada fitur kategorikal.
2. **Analisis Eksplorasi**: Mengevaluasi pola penjualan berdasarkan faktor-faktor seperti wilayah, kategori produk, ukuran dan status pengiriman.
3. **Pengembangan Model**: Membangun dan menguji model *machine learning* (misalnya, Random Forest, XGBoost) untuk memprediksi status pengiriman dengan fokus pada peningkatan *recall.*
4. **Evaluasi Model**: Menggunakan metriks evaluasi seperti *recall, precision,* dan F1-score untuk mengukur kinerja model.
5. **Pembuatan Visualisasi**: Menyajikan hasil analisis dan wawasan dalam bentuk dashboard atau laporan visual.
6. **Rekomendasi Bisnis**: Memberikan solusi untuk mengurangi tingkat retur dan penolakan pengiriman barang.

## Jadwal Kerja

Berikut adalah gambaran *timeline* pelaksanaan *final project.*

**

**Gambar II- 2 Jadwal Kerja**

*Final Project Bootcamp* Rakamin dimulai dengan **Stage 0 (14-20 Oktober 2024)**, yaitu penentuan masalah, tujuan proyek, dan metrik bisnis yang relevan. Pada tahapan ini juga terdapat **Mentoring 1**, di mana anggota kelompok berdiskusi dengan mentor terkait hasil penentuan masalah, *goal*, dan metrik bisnis, serta menerima saran untuk perbaikan jika diperlukan.

Selanjutnya, **Stage 1 (28 Oktober - 3 November 2024)** dilakukan eksplorasi data, analisis pola, dan visualisasi untuk mendapatkan wawasan awal. Hasil dari tahap ini didiskusikan dalam **Mentoring 2**, di mana mentor memberikan masukan terkait *insight* yang diperoleh dan visualisasi yang telah dibuat.

Pada **Stage 2 (18-24 November 2024)**, dilakukan *pre-processing* data, seperti pembersihan, pengolahan, dan persiapan data agar siap digunakan untuk pemodelan. Progres tahap ini dibahas dalam **Mentoring 3**, dimana mentor memberikan umpan balik terhadap langkah *pre-processing* yang telah dilakukan.

**Stage 3 (2-8 Desember 2024)** berfokus pada pengembangan model *machine learning*, evaluasi performa model, dan pengoptimalan. Progres model dan evaluasinya didiskusikan dalam **Mentoring 4**, dengan tujuan memastikan model telah sesuai dengan tujuan bisnis dan memperoleh masukan untuk peningkatan.

Pada **Stage 4 (16-20 Desember 2024)**, kami mempersiapkan presentasi proyek. Hasil kerja tahap ini dibahas dalam **Mentoring 5**, yang bertujuan untuk memvalidasi isi dari presentasi, serta memberikan saran untuk penyempurnaan. Akhirnya, pada **21 Desember 2024**, kelompok melaksanakan simulasi presentasi untuk mempersiapkan penyampaian hasil proyek kepada pemangku kepentingan dalam hal ini mentor Rakamin yang berpengalaman di bidangnya, memastikan presentasi dapat dilakukan dengan baik dan sesuai dengan ekspektasi.

# BAB III Predicting Package Status in E-Commerce Logistics

## Deskripsi Persoalan

Dalam industri e-commerce, pengelolaan logistik merupakan salah satu elemen yang sangat krusial. Tingkat pengembalian produk selama periode tertentu (April-Juni 2022) mencapai 6,38% dengan total kerugian penjualan sebesar INR 1.275.310. Meskipun angka ini berada dalam standar industri, dampaknya terhadap revenue dan efisiensi operasional perlu dievaluasi lebih dalam. Proyek ini difokuskan pada perusahaan Amazon**,** salah satu pemain utama di industri e-commerce global. Dataset yang digunakan adalah **Amazon Sales Dataset**, yang terdiri dari beberapa kolom seperti *date****,*** *id****,*** *qty****,*** *amount****,*** *category****,*** *size*, dan lainnya. Dataset ini memberikan informasi transaksi penjualan yang relevan untuk analisis prediktif dan pengembangan model machine learning.

Salah satu tantangan utama dalam proyek ini adalah keterbatasan dataset yang tersedia. Dataset yang digunakan sebagian besar terdiri dari fitur kategorikal, seperti status pesanan, lokasi pengiriman, dan jenis layanan, sedangkan fitur numerik yang biasanya memberikan informasi lebih mendalam untuk analisis prediktif (misalnya waktu pengiriman atau berat paket) jumlahnya sangat terbatas. Keterbatasan ini menyulitkan pembangunan model machine learning yang dapat diandalkan untuk memprediksi status pengiriman dengan akurasi tinggi.

Tujuan utama proyek ini adalah membangun model machine learning yang fokus pada prediksi status pengiriman paket. Selain itu, proyek ini bertujuan untuk mengidentifikasi fitur-fitur utama yang memengaruhi pengembalian produk sehingga perusahaan dapat mengembangkan strategi proaktif untuk menangani pelanggan yang berisiko melakukan pengembalian.

## Proses Pelaksanaan

Proses pelaksanaan proyek ini terdiri dari beberapa tahapan utama:

1. Pengumpulan dan Eksplorasi Data

Dataset yang digunakan diperoleh dari sistem logistik perusahaan e-commerce. Data tersebut kemudian dieksplorasi untuk memahami pola dan distribusi nilai dari setiap fitur.

1. Preprocessing Data

* **Penanganan Missing Values:** Mengisi nilai kosong berdasarkan pola atau rata-rata.
* **Encoding Fitur Kategorikal:** Menggunakan teknik encoding seperti One-Hot Encoding dan Label Encoding untuk fitur kategorikal.
* **Normalisasi Data:** Melakukan normalisasi untuk memastikan fitur numerik memiliki skala yang seragam.

1. Feature Engineering

* Membuat fitur baru seperti weekdays dan weekend agar mengetahui pengiriman yang efektif terjadi saat weekdays atau weekend.

1. Modeling

* Algoritma yang digunakan meliputi Logistic Regression, kNN, Random Forest, dan XGBoost.

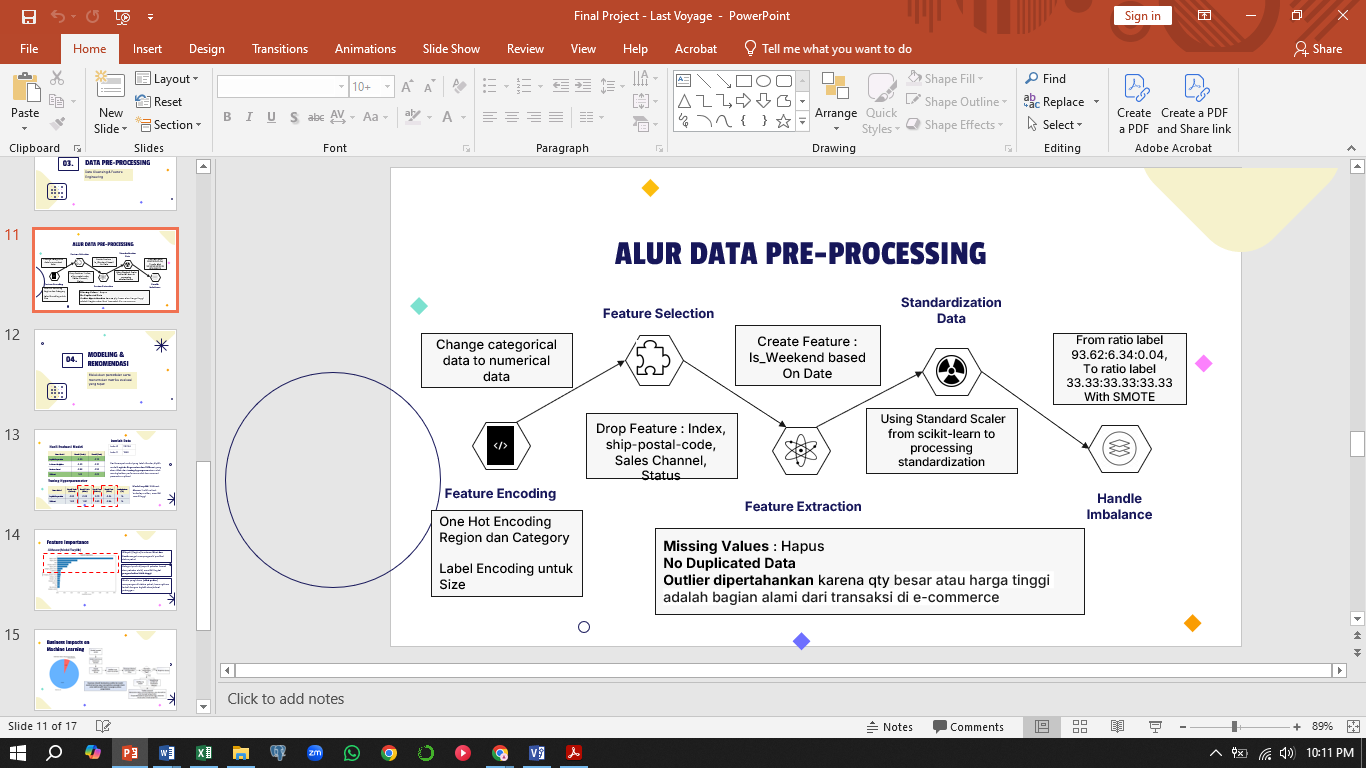
1. Kolaborasi Tim dan Tools

* Proyek ini dilakukan dalam tim menggunakan alat seperti Google Colab dan Jupyter Notebook untuk coding.
* Pengetahuan yang relevan dari bootcamp, seperti penggunaan library Python (Pandas, Scikit-learn), sangat mendukung keberhasilan proyek.

1. Kendala yang Dihadapi

* Komunikasi Tim: Jadwal yang padat antar anggota tim menjadi salah satu tantangan utama, terutama karena beberapa anggota memiliki kewajiban lain di luar proyek. Solusi: Menentukan Jadwal rapat dan komunikasi melalui grup.
* **Keterbatasan Data Numerik:** Menyiasatinya dengan feature engineering untuk menambahkan informasi baru.

Secara keseluruhan, rangkuman alur pemrosesan data dijelaskan pada gambar di bawah ini.



**Gambar III- 1 Alur pemrosesan dataset**

## Pencapaian Hasil

Proyek ini menghasilkan sejumlah pencapaian utama yang mendukung pengembangan sistem prediksi status pengiriman pada logistik *e-commerce*:

* + - 1. Kinerja Model
* Model terbaik yang dihasilkan adalah **XGBoost**, dengan nilai recall sebesar 85%.
* Model mampu memprediksi status pengiriman dengan tingkat akurasi yang memadai untuk implementasi awal.
  + - 1. Identifikasi Fitur Penting
* Analisis menunjukkan bahwa **wilayah pengiriman (Region)**, terutama wilayah **South** dan **West**, memiliki pengaruh signifikan terhadap prediksi status paket.
* Fitur lain seperti **kategori produk (pakaian etnik dan bawah)**, serta **waktu pengiriman (akhir pekan)** juga berdampak signifikan terhadap risiko pengembalian.
  + - 1. Dampak Bisnis
* Implementasi model ini memungkinkan perusahaan untuk lebih proaktif dalam menangani potensi pengembalian barang dengan notifikasi awal kepada penjual.
* Prediksi status pengiriman berbasis model ini berpotensi mengurangi **lost sales hingga 15%** dengan memitigasi risiko pengembalian barang secara efektif.
  + - 1. Rekomendasi

1. **Fokus Wilayah:** Tingkatkan analisis logistik di wilayah dengan tingkat pengembalian tinggi, seperti South dan West, serta optimalkan jalur pengiriman.
2. **Kategori Produk:** Perbaiki deskripsi produk, panduan ukuran, dan visualisasi untuk kategori dengan tingkat pengembalian tinggi.
3. **Jadwal Pengiriman:** Hindari pengiriman di akhir pekan untuk mengurangi potensi pengembalian barang.
4. **Edukasi Pelanggan:** Jalankan kampanye edukasi kepada pelanggan untuk meningkatkan pemahaman mengenai proses pengiriman dan pengembalian barang.

* Hasil Evaluasi Model

**Tabel III- 1 Hasil evaluasi 4 Model**

| Nama Model | Recall (Train) | Recall (Test) |
| --- | --- | --- |
| Logistic Regression | 0.80 | 0.73 |
| k-Nearest Neighbor | 0.83 | 0.27 |
| Random Forest | 0.96 | 0.20 |
| XGBoost | 1.00 | 0.83 |

* Tuning Hyperparameter

**Tabel III- 2 Hasil evaluasi model dengan tuning hyperparameter**

| Nama Model | Recall Train (Before) | Recall Test  (After) | Recall Test (Before) | Recall Test (After) | Peningkatan  (%) |
| --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.80 | 0.88 | 0.73 | 0.84 | 15 |
| XGBoost | 1.00 | 1.00 | 0.83 | 0.95 | 14 |

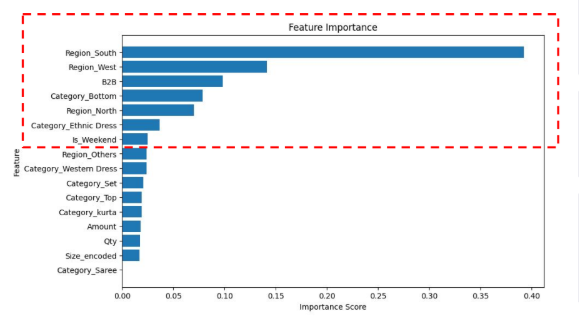
Model terpilih: **XGBoost**

XGBoost dipilih sebagai model yang dengan nilai *Recall test* yang tinggi (0.83), dan setelah tuning hyperparameter menjadi (0.95). Hal ini memastikan bahwa model sangat baik dalam mendeteksi status yang harus ditolak dan menghindari kesalahan kritis dalam memprediksi kelas positif (status yang seharusnya ditolak).

Dalam bisnis *e-commerce*, mengelola pengembalian barang adalah proses yang sering mempengaruhi kepuasan pelanggan dan juga beban logistik perusahaan. Dengan recall yang tinggi, kita memastikan bahwa hampir semua paket yang perlu dikembalikan akan terdeteksi.

Walaupun *recall* tinggi dapat menyebabkan beberapa kesalahan dalam prediksi (misalnya, *false positives*, di mana paket yang seharusnya diterima malah dikembalikan), hal ini lebih mudah untuk diperbaiki. Dalam banyak kasus, kesalahan ini bisa diselesaikan dengan verifikasi manual atau proses pengembalian otomatis, sehingga konsumen yang menerima paket yang salah bisa segera diberi solusi.

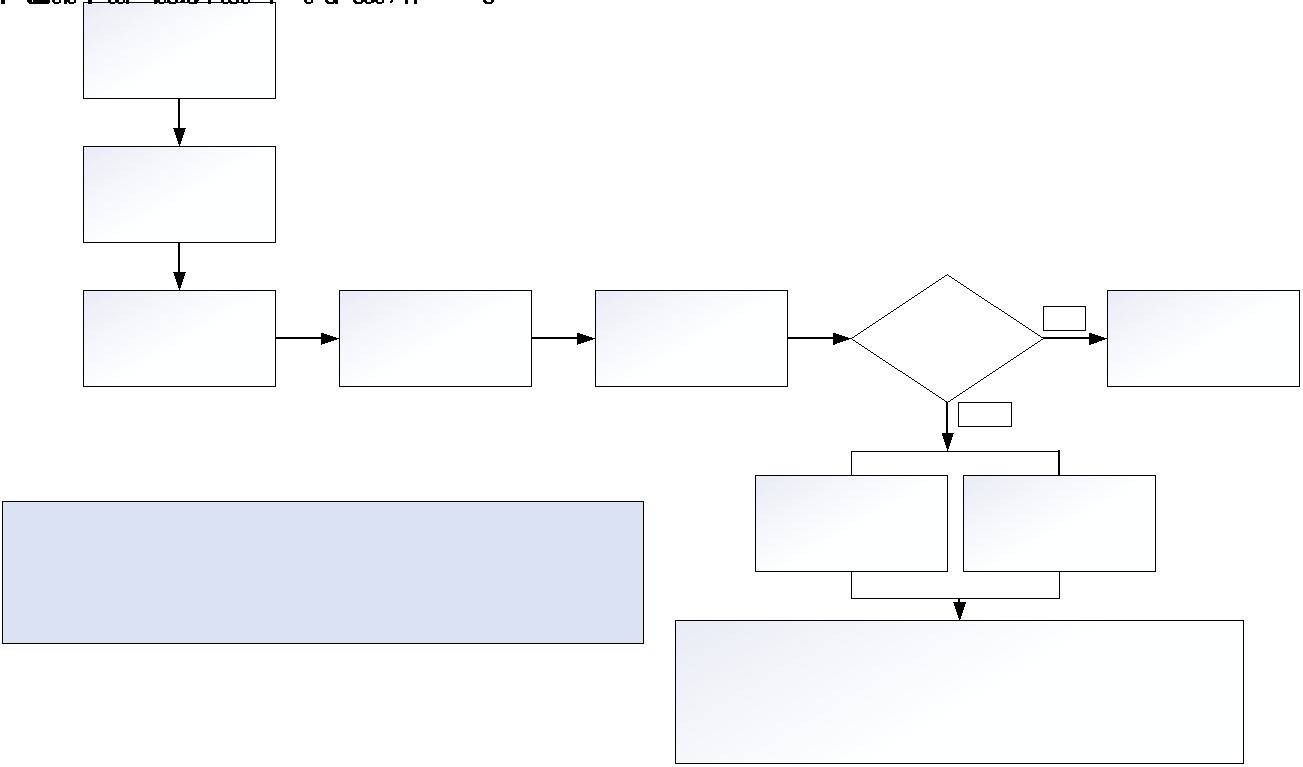
* *Feature Importance* (XGBoost)



**Gambar III- 2 Feature Importance XGboost**

* Wilayah (Region) terutama **West dan South** sangat mempengaruhi prediksi status paket.
* Kategori produk (seperti pakaian bawah atau pakaian etnik) memiliki tingkat **pengembalian lebih tinggi.**
* Waktu pengiriman (**akhir pekan**) mempengaruhi status paket, kemungkinan terkait dengan logistik atau jadwal pelanggan.

**Implementasi Penerapan Machine Learning pada Proses Bisnis**



**Gambar III- 3 Penerapan Machine Learning pada Proses Bisnis**

Proses penerapan machine learning dalam e-commerce dimulai saat pembeli melakukan transaksi, diikuti dengan sistem yang memproses pembayaran dan mencatat pesanan. Setelah penjual menerima pemberitahuan untuk mengirimkan barang, sistem e-commerce memberi notifikasi kepada pembeli. Model machine learning digunakan untuk memprediksi status pengiriman dan risiko pengembalian produk berdasarkan berbagai faktor. Jika risiko pengembalian tinggi, sistem akan mengirimkan notifikasi kepada penjual dan menghubungi pembeli untuk mengambil langkah proaktif, seperti memberikan klarifikasi, mengganti metode pengiriman, atau menawarkan diskon. Tindakan preventif seperti diskon tambahan, asuransi, atau metode pengiriman yang lebih tepat dapat diambil untuk mengurangi risiko pengembalian sebelum pengiriman barang dilanjutkan.

# BAB IV PENUTUP

## Kesimpulan

* + - 1. **Pencapaian Utama**
         1. Kinerja Model: Model terbaik yang dihasilkan adalah XGBoost, dengan nilai recall sebesar 83% setelah tuning hyperparameter. Model ini mampu memprediksi status pengiriman dengan akurasi yang memadai untuk implementasi awal.
  1. Identifikasi Fitur Penting:
* Wilayah pengiriman (terutama South dan West) memberikan pengaruh signifikan terhadap risiko pengembalian.
* Fitur kategori produk, seperti pakaian etnik dan pakaian bawah, serta waktu pengiriman (akhir pekan), juga memiliki dampak besar.
  1. Dampak Bisnis: Implementasi model ini dapat mengurangi potensi pengembalian barang hingga 15%, meningkatkan efisiensi operasional, serta membantu pengambilan keputusan logistik yang lebih strategis.

1. **Rekomendasi Strategis**
   * + - 1. Optimalisasi Wilayah: Fokuskan analisis logistik dan promosi di wilayah dengan tingkat pengembalian tinggi, seperti South dan West, untuk mengurangi risiko pengembalian.
         2. Kategori Produk: Tingkatkan deskripsi produk, panduan ukuran, dan visualisasi pada kategori dengan tingkat pengembalian tinggi. Promosikan kategori yang diminati, seperti Set dan Kurta, melalui kampanye pemasaran khusus.
         3. Jadwal Pengiriman: Hindari pengiriman di akhir pekan untuk meminimalkan risiko pengembalian barang.
         4. Edukasi Pelanggan: Jalankan kampanye untuk meningkatkan pemahaman pelanggan tentang kebijakan pengembalian dan proses pengiriman.
2. **Kendala dan Solusi**

Komunikasi Tim: Kendala jadwal anggota tim diselesaikan dengan menentukan jadwal rapat yang teratur dan komunikasi intensif melalui grup diskusi.

Keterbatasan Data Numerik: Dilakukan *feature engineering* untuk menambahkan informasi baru, seperti hari kerja vs akhir pekan, agar analisis lebih relevan.

1. Business Insights
   * + - 1. Performa Produk: Kategori Set dan Kurta menunjukkan performa pengiriman dan penerimaan yang baik, mencerminkan preferensi pelanggan.
         2. Distribusi Wilayah: Negara bagian seperti Maharashtra dan Karnataka mendominasi jumlah pesanan, menunjukkan potensi untuk fokus promosi di wilayah tersebut.
         3. Efektivitas Metode Pengiriman: Metode Expedited Shipping paling efektif, terutama untuk kategori populer.

Model prediksi berbasis XGBoost berhasil memberikan wawasan yang actionable bagi perusahaan untuk memitigasi risiko pengembalian barang, meningkatkan efisiensi logistik, serta memberikan rekomendasi strategis untuk peningkatan kepuasan pelanggan. Proyek ini menunjukkan bahwa pendekatan berbasis data dapat mendukung pengambilan keputusan bisnis yang lebih baik dan meningkatkan daya saing dalam industri e-commerce.

## Saran

Optimalisasi Infrastruktur Logistik

* Distribusi Wilayah: Fokuskan pada wilayah dengan tingkat pengembalian barang tinggi, seperti South dan West, dengan mengoptimalkan jalur pengiriman, menambahkan mitra logistik, atau mendirikan pusat distribusi lebih dekat ke pelanggan.
* Kemitraan Strategis: Jalin kerja sama dengan penyedia logistik yang menawarkan layanan pengiriman cepat, terutama untuk wilayah dengan permintaan tinggi seperti Maharashtra dan Karnataka.

Pemanfaatan Data Real-Time

* Tracking Dinamis: Integrasikan model prediksi dengan sistem manajemen logistik real-time untuk memberikan notifikasi awal mengenai potensi masalah pengiriman.
* Feedback Loop: Gunakan data terbaru dari proses pengiriman untuk terus memperbarui model, sehingga akurasinya tetap relevan terhadap tren logistik yang berubah.

Peningkatan Customer Experience

* Personalized Notifications: Kirimkan pemberitahuan otomatis kepada pelanggan tentang status pengiriman barang dengan informasi yang jelas, termasuk estimasi waktu pengiriman yang lebih akurat.
* Proaktif dalam Risiko Pengembalian: Identifikasi pelanggan dengan risiko tinggi untuk melakukan pengembalian dan berikan penawaran personal, seperti diskon untuk pembelian berikutnya, guna meningkatkan kepuasan mereka.

Inovasi pada Kategori Produk

* Fokus pada Produk Populer: Investasikan lebih banyak pada kategori seperti Set dan Kurta yang memiliki tingkat permintaan tinggi. Tingkatkan variasi desain, stok, dan promosi untuk mempertahankan minat pelanggan.
* Pengelolaan Produk Bermasalah: Tinjau ulang produk dengan tingkat pengembalian tinggi dan evaluasi deskripsi, kualitas, serta ukurannya. Pertimbangkan untuk menawarkan panduan ukuran interaktif atau fitur coba secara virtual.
  + - 1. Penguatan Analisis Pasar
* Preferensi Regional: Analisis lebih dalam kebutuhan dan preferensi pelanggan di wilayah dengan pesanan rendah untuk menciptakan strategi penetrasi pasar yang lebih efektif.
* Segmentasi Pelanggan: Gunakan data pelanggan untuk mengidentifikasi segmen yang paling menguntungkan dan arahkan promosi sesuai dengan preferensi mereka.
  + - 1. Penyelarasan Tim dan Proses
* Pelatihan Internal: Berikan pelatihan bagi tim terkait penggunaan hasil model prediksi untuk pengambilan keputusan logistik dan strategi bisnis.
* Kolaborasi Antardepartemen: Pastikan tim logistik, pemasaran, dan customer service memiliki komunikasi yang terintegrasi untuk memanfaatkan rekomendasi model secara maksimal.

1. Monitoring dan Evaluasi Berkelanjutan

* Key Performance Indicators (KPIs): Tetapkan metrik seperti tingkat pengembalian barang, akurasi pengiriman, dan tingkat kepuasan pelanggan sebagai tolok ukur keberhasilan implementasi.
* Review Berkala: Lakukan evaluasi triwulanan terhadap kinerja model dan dampaknya pada bisnis untuk menentukan area yang memerlukan perbaikan atau pengembangan lebih lanjut.

1. Strategi Pemasaran Berbasis Wawasan

* Promosi yang Tepat Sasaran: Gunakan wawasan dari model untuk membuat kampanye promosi yang sesuai, misalnya penawaran diskon khusus di wilayah tertentu atau bundling produk dengan tingkat pengembalian rendah.
* Edukasi Pelanggan: Jalankan kampanye edukasi untuk membantu pelanggan memahami fitur produk, panduan ukuran, dan kebijakan pengembalian untuk mengurangi ketidakpuasan.

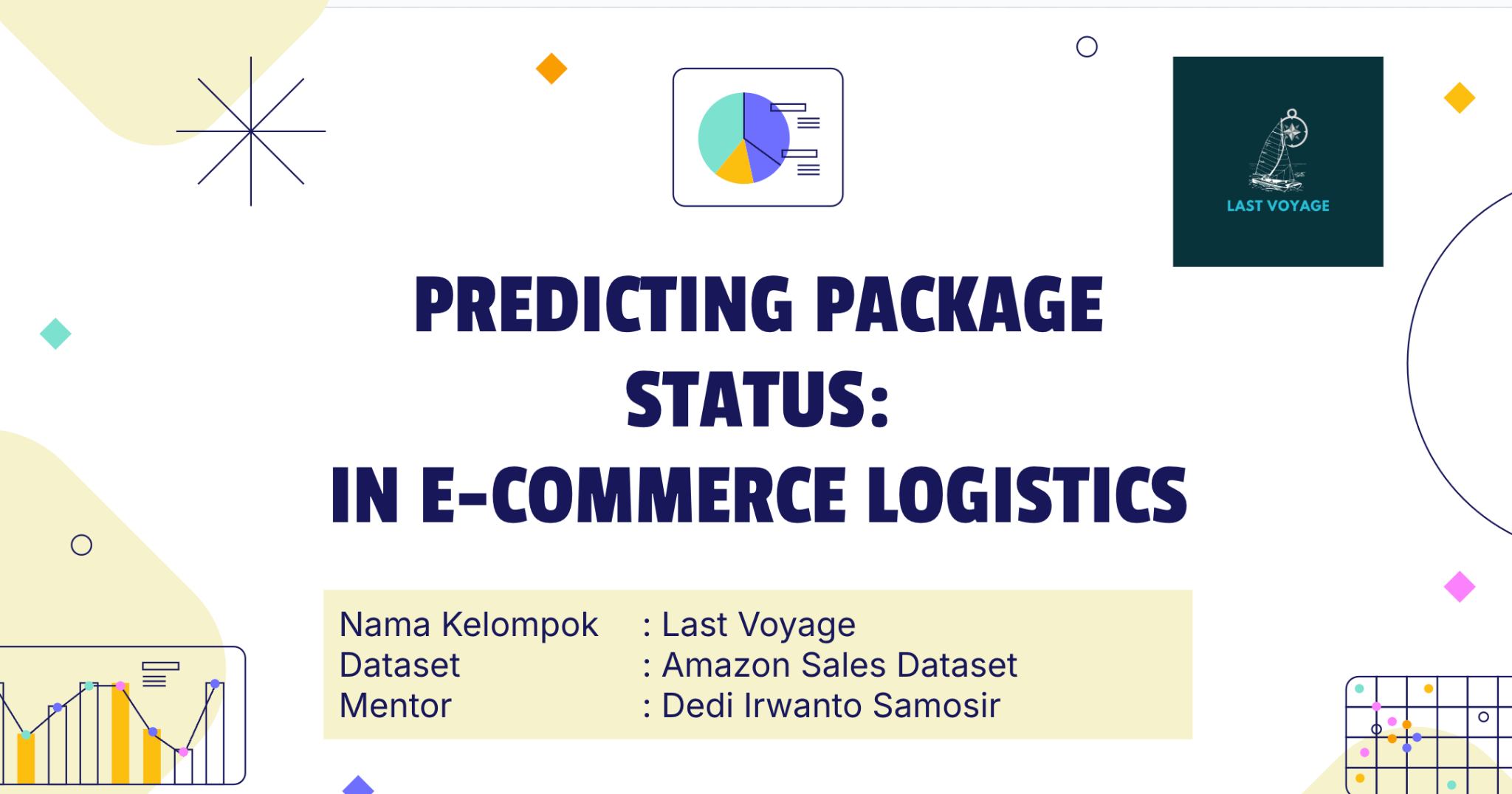
Implementasi proyek ini memiliki potensi besar untuk meningkatkan efisiensi operasional, meminimalkan kerugian, dan meningkatkan pengalaman pelanggan. Keberhasilan jangka panjangnya bergantung pada pemanfaatan data yang berkelanjutan, kolaborasi internal, dan respons cepat terhadap perubahan dalam kebutuhan pelanggan dan logistik.

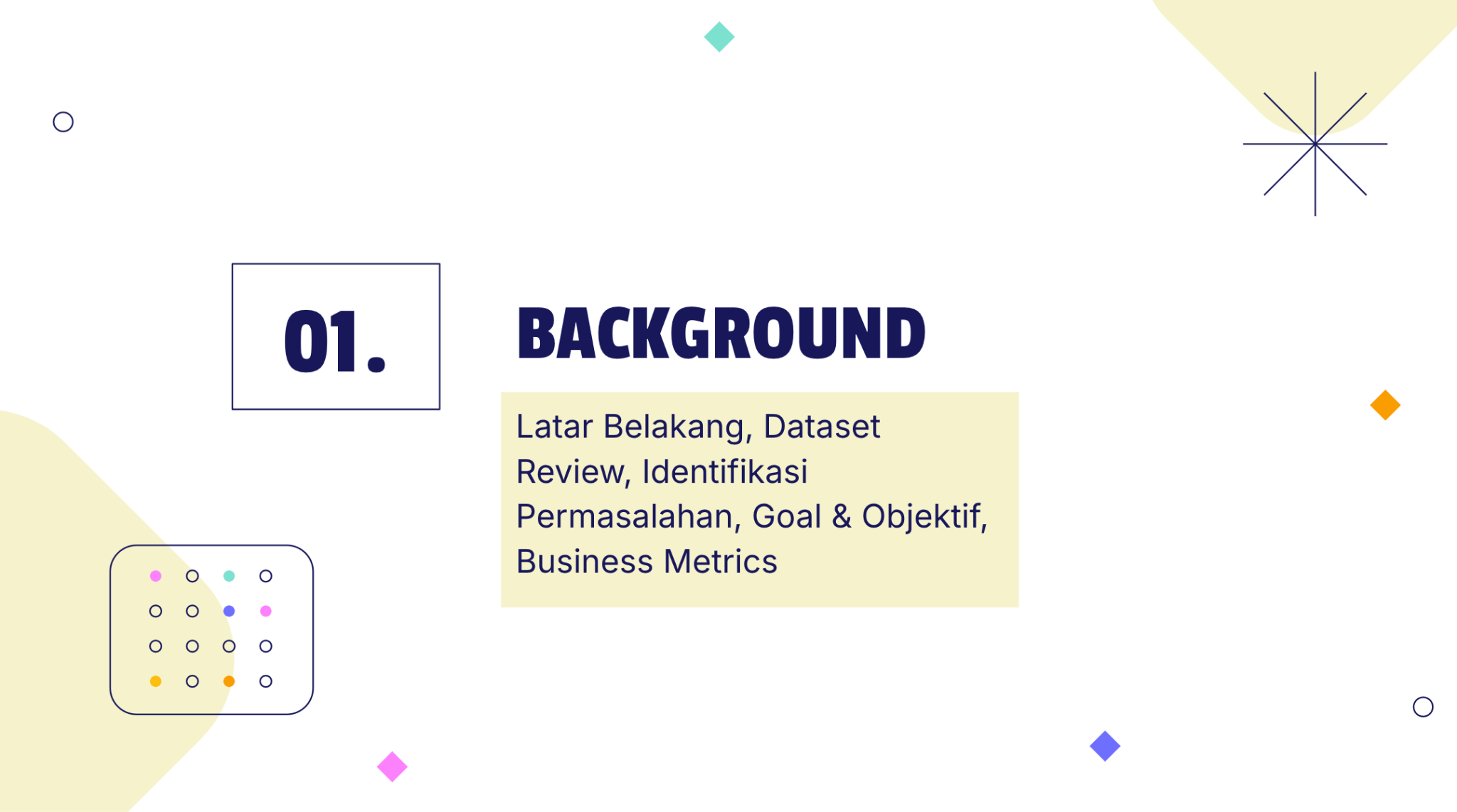
## 4.3 Referensi

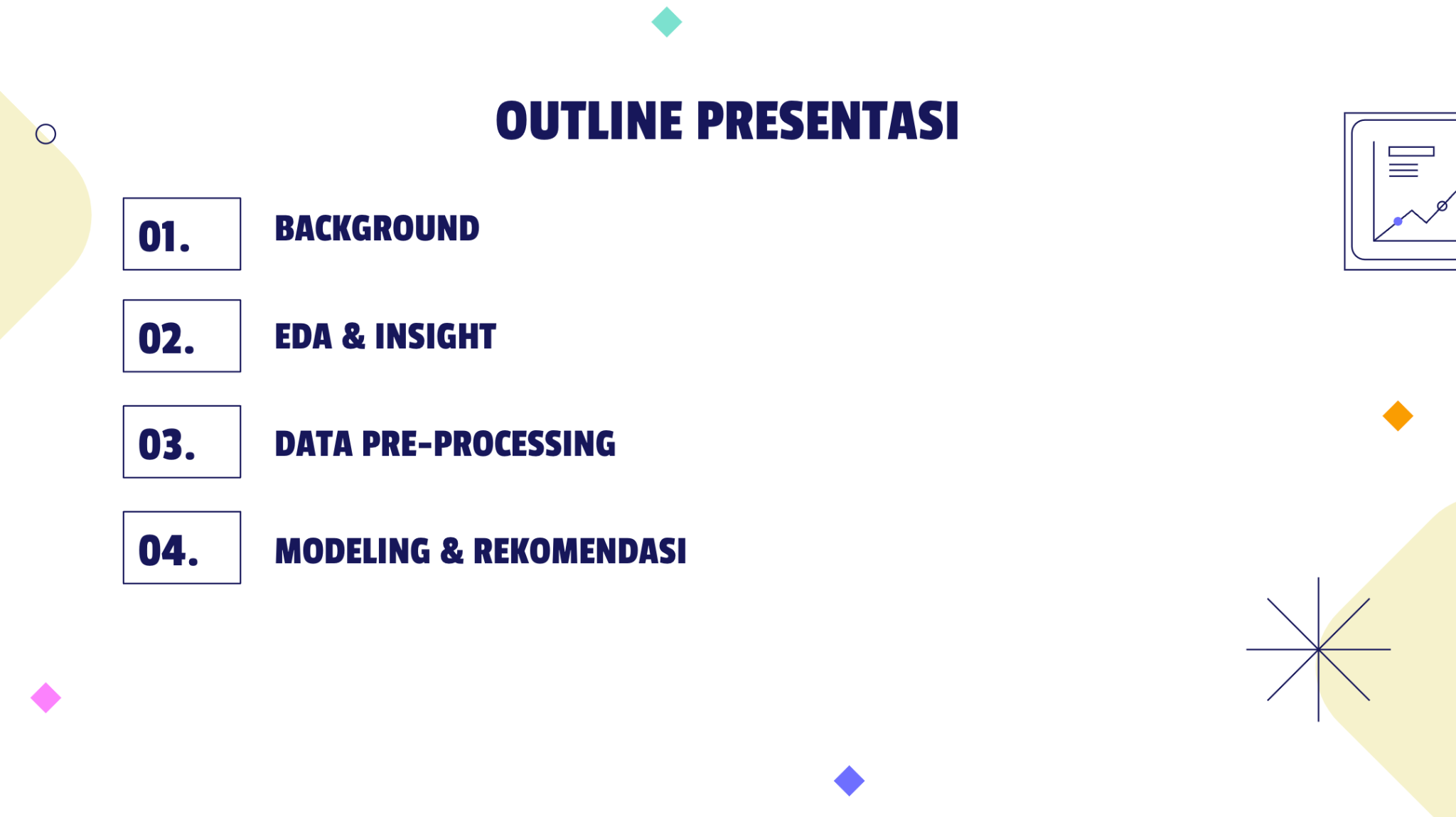
* + - 1. Statista. (n.d.). *Global retail e-commerce sales from 2014 to 2024 (in trillion U.S. dollars)*. Retrieved from <https://www.statista.com>
      2. Google & Temasek. (2023). *e-Conomy SEA 2023*. Retrieved from https://economysea.withgoogle.com

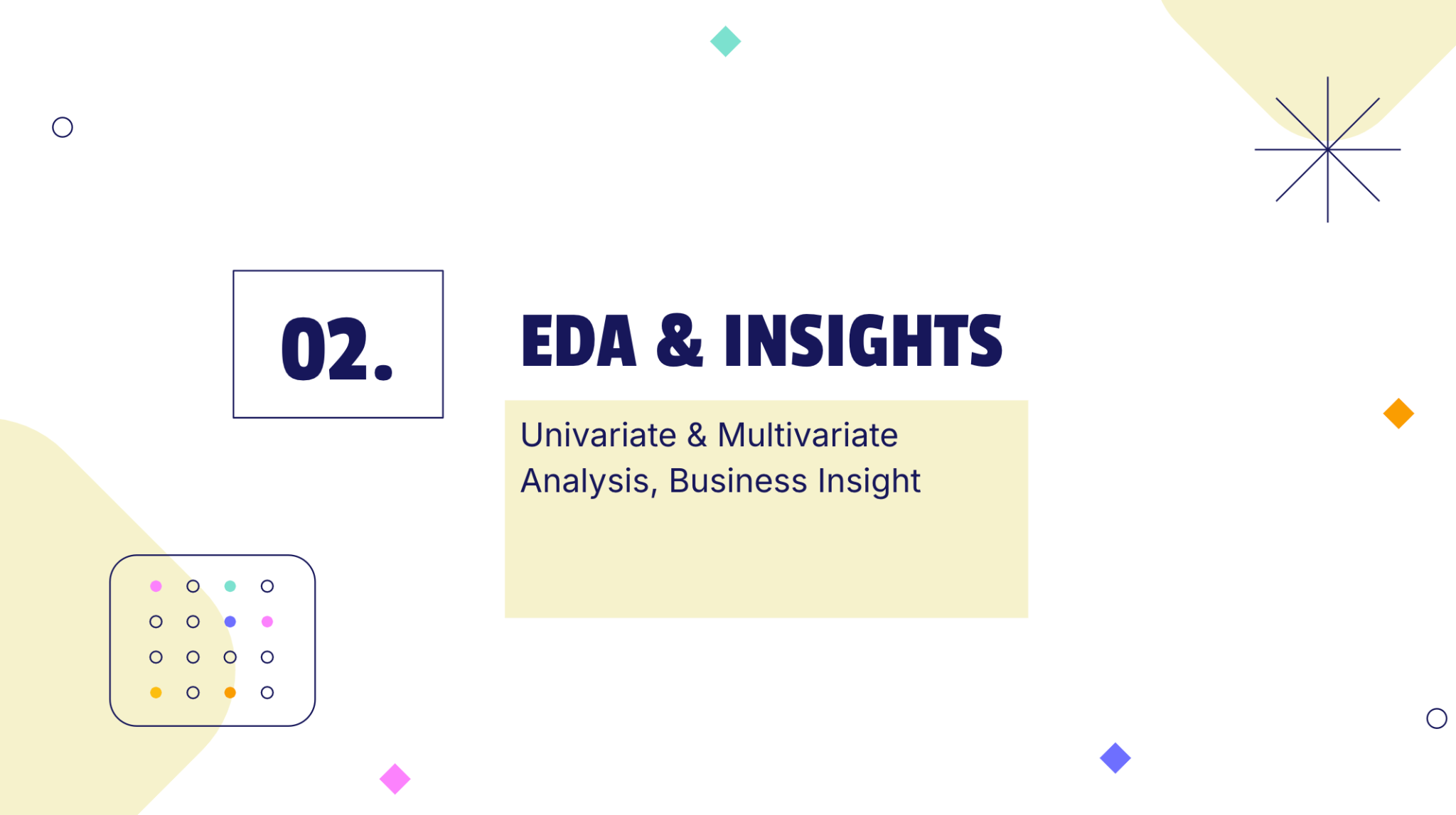
# BAB V

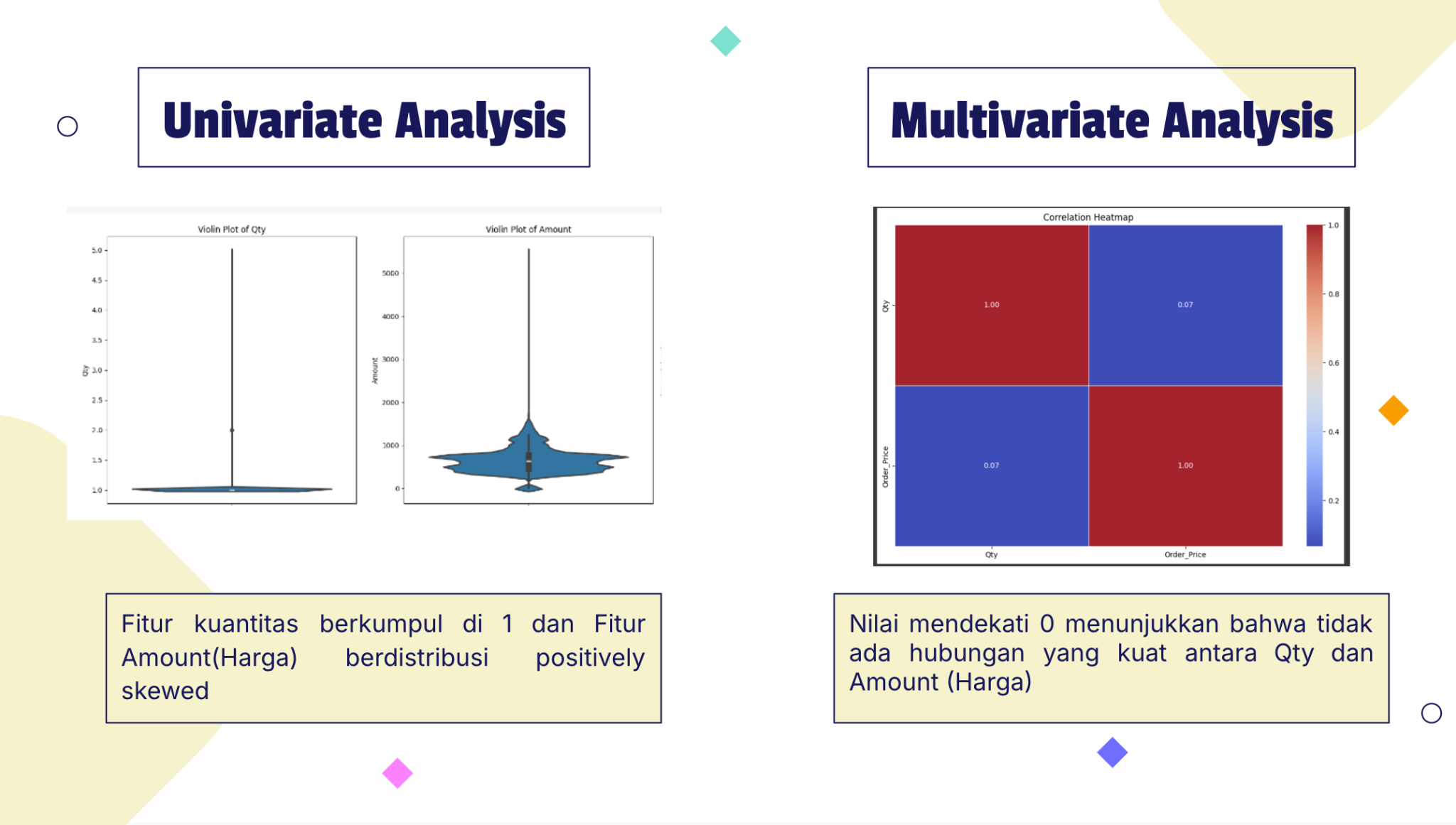
## Lampiran A PPT Final Project Presentation

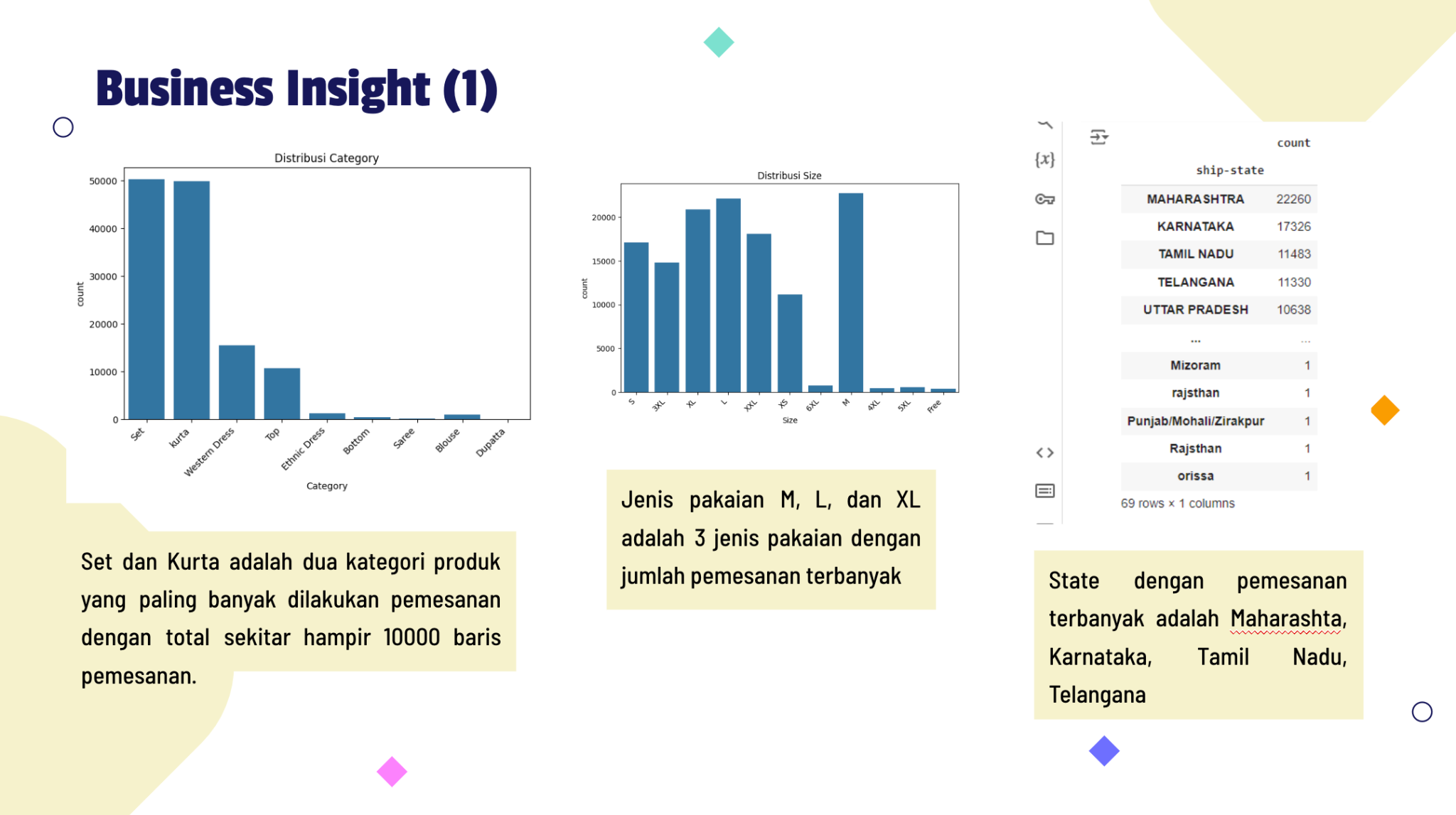
**

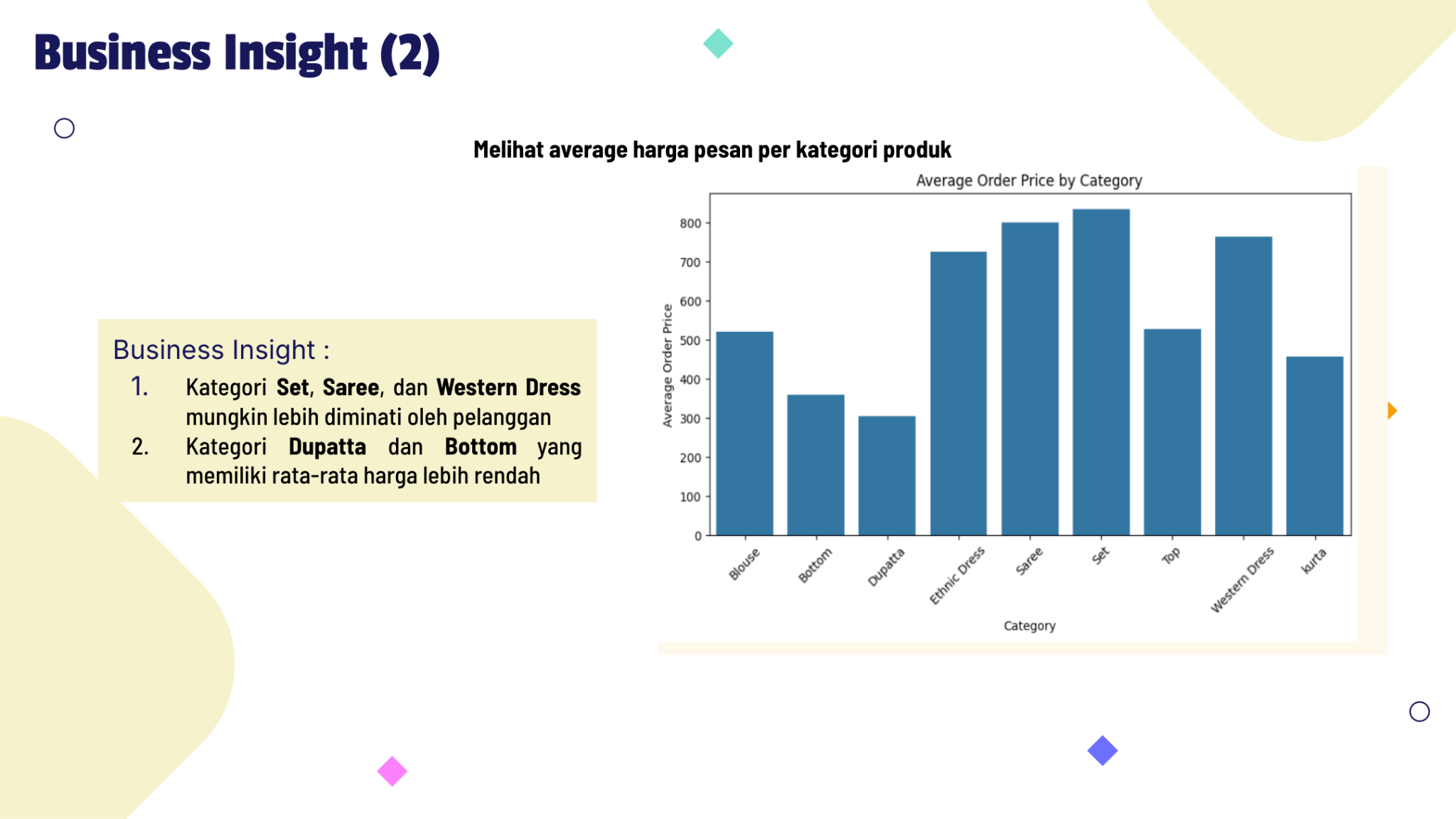




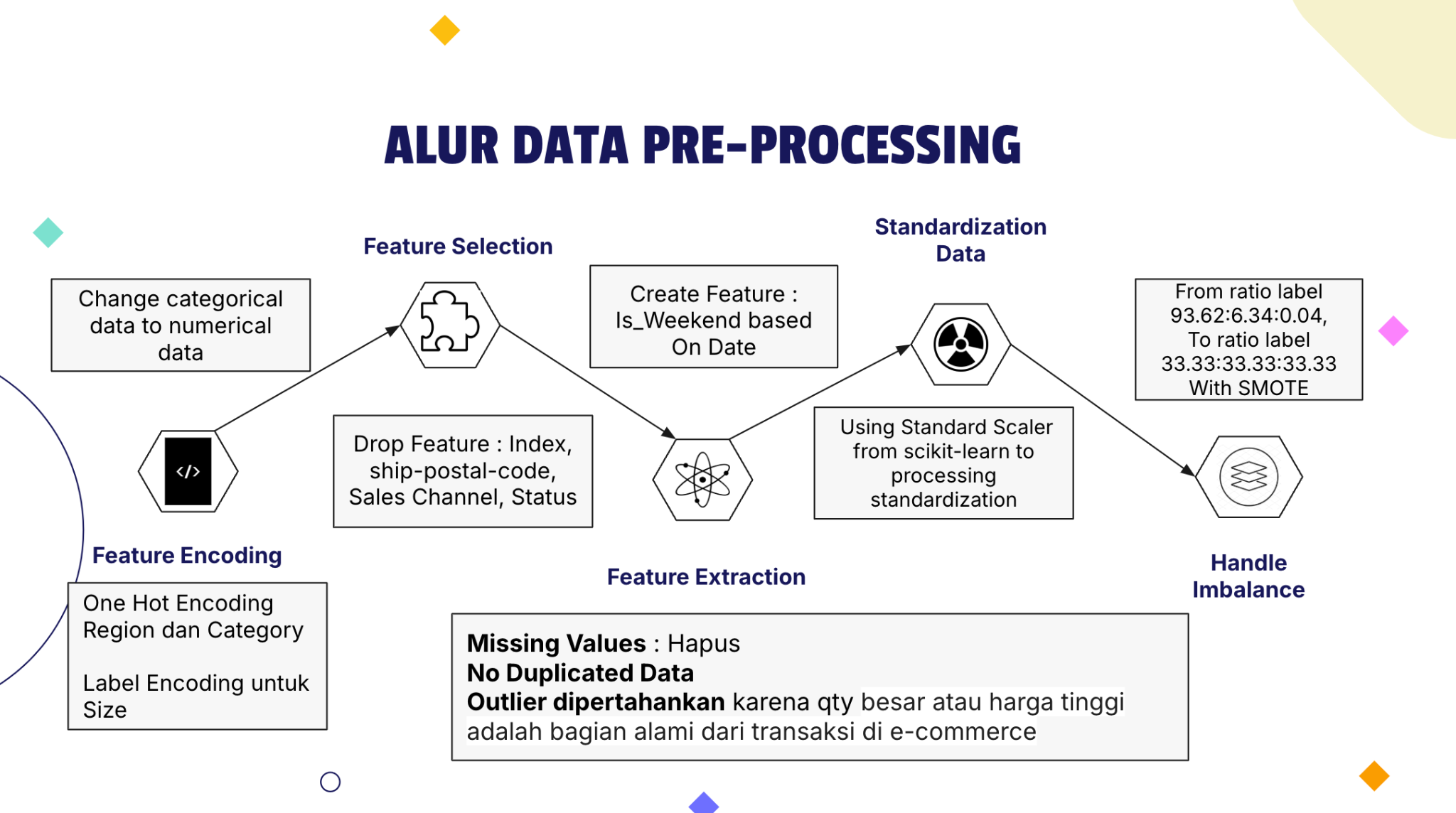




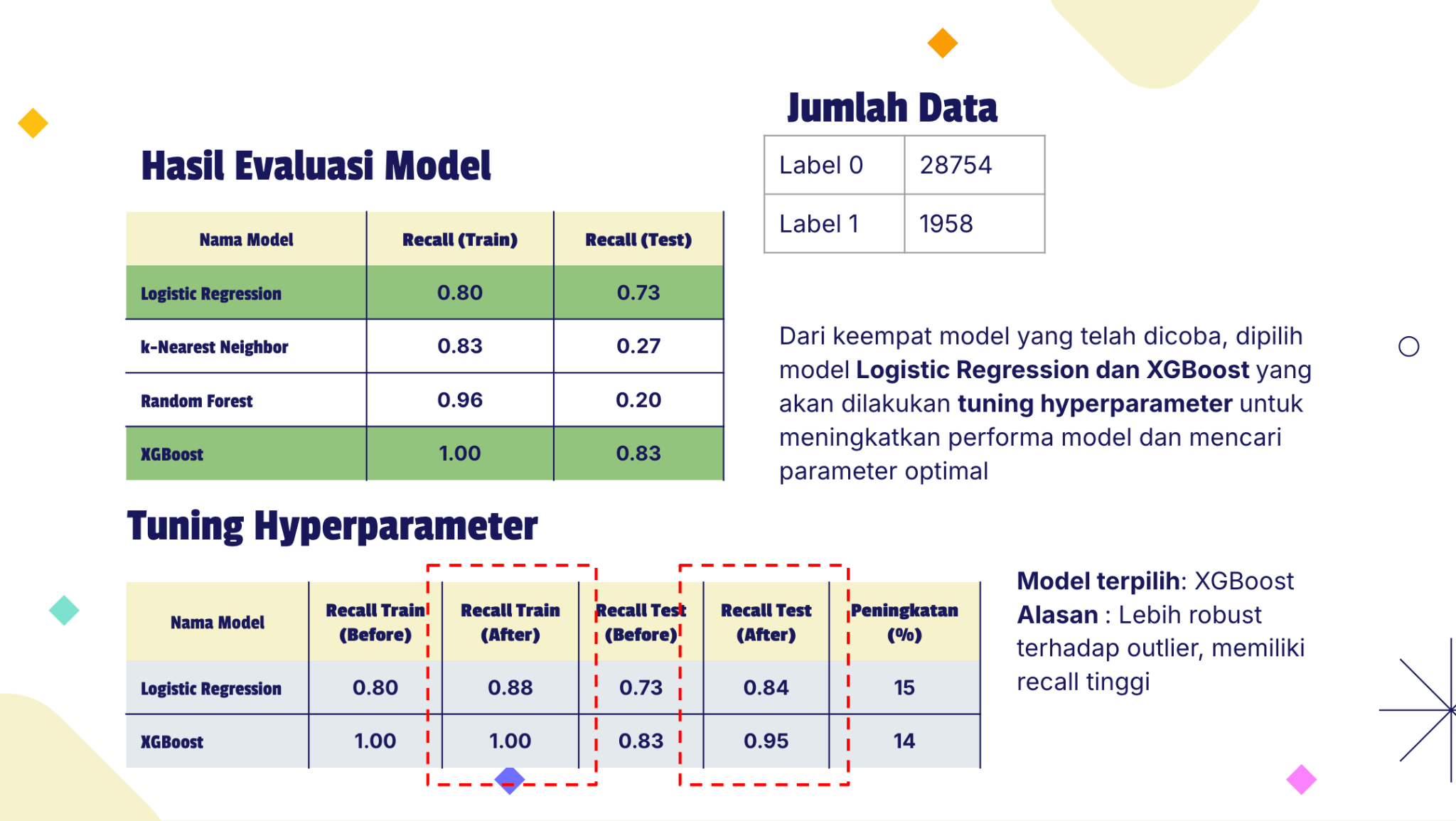


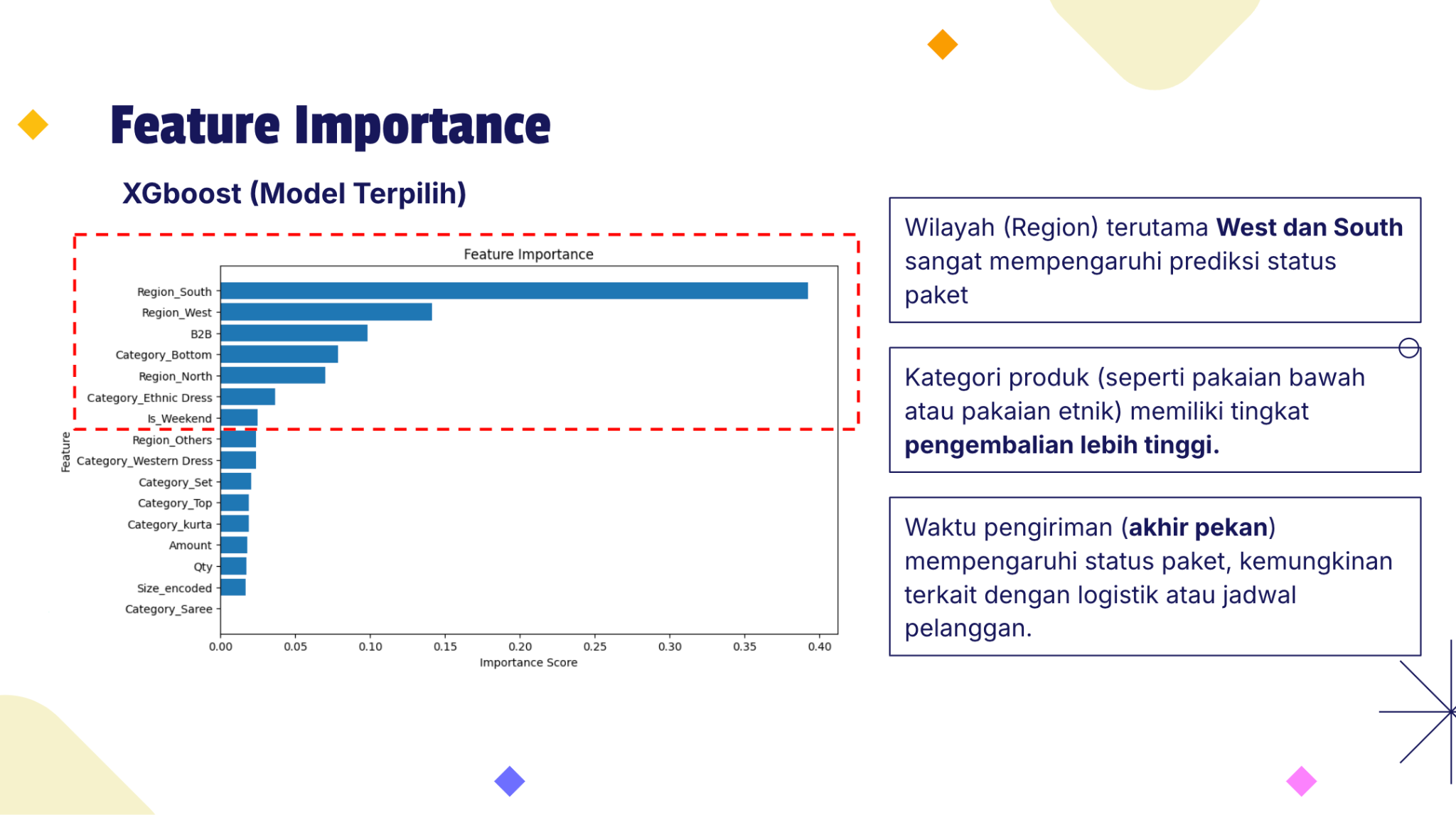


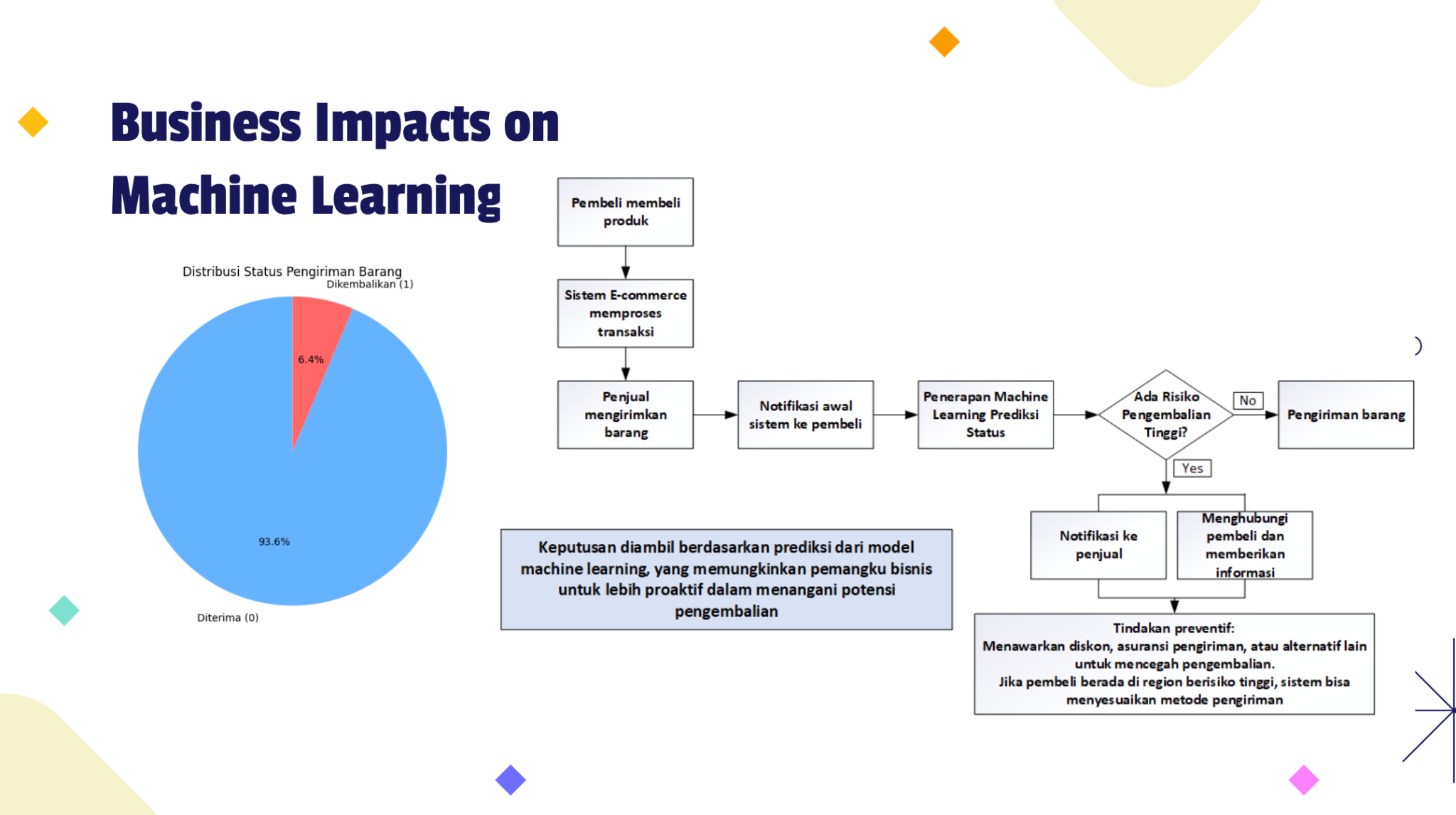














## Lampiran B Notulensi Mentoring

| Minggu/Tgl | Kegiatan | Hasil |
| --- | --- | --- |
| 18/Oktober 2024 | 1. Presentasi dataset yang sudah diolah oleh kelompok 1 mengenai e-commerce  2. Business Metrics  3. Goals, identifikasi masalah, | 1. Dataset yang tersedia belum bisa menunjukkan masalah yang perlu diselesaikan, Rating dan jumlah pembeli tidak bisa dijadikan database untuk rekomendasi produk dan menjadi sebuah  machine learning, dataset berupa analisis hasil penjualan dan evaluasi.  2. Goals ditambahkan berupa angka yang menjadi tujuan (% kenaikan/revenue),  3. Businnes metrics ditambahkan perhitungan total revenue, presentase pengunjung yang melakukan pembelian |

| 3/ November 2024 | 1. Penentuan kolom yang berpengaruh untuk target machine learning  2. Penentuan kolom target dan isinya yang berpengaruh untuk machine learning  3. Membahas isi dataset dan analisis statistic, univariate, multivariate, dan insight.  4. Membahas data pengiriman produk | 1. Kolom target difokuskan pada status delivered to buyer, rejected by customers dan returned to buyers (melihat dari sisi konsumen)  2. Kolom ship-state difokuskan dibandingkan city untuk menyederhanakan jumlah unik baris  3. Mapping untuk jenis level pengiriman, kategori, dan ukuran produk  4. Dataset ukuran, jenis produk dan status pengiriman untuk mengetahui produk mana yang banyak di cancel, yang banyak diterima, dan yang banyak di return  5. data Returned dan Damage bisa dipertimbangkan disatukan |
| --- | --- | --- |

| 23/ November 2024 | 1. Outlier pada dataset, transformasi fitur  2. Pengaruh kolom kategorikal ketika filtering status  3. Feature encoding (pengelompokkan state dan category)  4. Penambahan feature yang diperlukan | Log transformasi pada fitur yang memiliki outlier tinggi tidak perlu dilakukan asalkan menggunakan model robust terhadap outlier nantinya, pengelompokkan region yang banyak efisien  jika dibagi berdasarkan posisi mata angin, kolom kategorikal yang terpengaruh akibat filtering status tidak apa apa tidak digunakan, status bisa dilakukan label encoding (1,2,3), usulan  penambahan fitur weekdays/weekend berdasarkan kolom date, penambahan feature tidak perlu dilakukan apabila memang dirasa tidak perlu (biasanya hanya untuk business insight |
| --- | --- | --- |

| 7/Desember/2024 | 1. Label status  2. Pemilihan metriks evaluasi yang tepat untuk kasus klasifikasi status paket  3. Penanganan class imbalanced  4. Interpretasi Feature Importance | -Label status untuk klasifikasi untuk angka 1 biasanya menunjukkan status yang akan dijadikan fokus perhatian, dan 0 yang normal  -Metriks evaluasi recall biasanya lebih banyak diterapkan dalam kasus nyata  -Penanganan class imbalanced dengan SMOTE dan RESAMPLE  -Class imbalanced untuk status masih normal saja dalam kasus nyata juga biasanya akan jomplang  -Ambil 5 -6 feature importance teratas saja untuk dijelaskan insight-nya |
| --- | --- | --- |

| 20/Desember/2024 | 1. Identifikasi masalah  2. EDA dan Business Insight  3. Data Pre-Processing  4. Pemodelan dan Rekomendasi | -Untuk identifikasi masalah tampilkan angka detail yang menunjukkan tingkat kembalinya barang dan kerugian yang timbul  -Untuk Business Insight tampilkan grafik penting serta insight-nya juga jangan hanya membaca data  -Data Pre-processing dibuat dalam bentuk diagram alur saja  -Pemodelan cantumkan model yang dipilih, bagaimana machine learning itu diimplementasi, dan rekomendasi berdasarkan hasil pemilihan model |
| --- | --- | --- |

## Lampiran C Dokumen Teknik

Link Github : <https://github.com/Rizqii1/Finpro_Kelompok1>

Link Final Project : <https://drive.google.com/drive/folders/1FI8c9zoe1SG9W-weiLcNBFFuFIjeuE_K?usp=sharing>