

# **-Dokumen Project Plan**

## **Asah led by Dicoding in association with Accenture**

**ID Tim Capstone Project : A25-CS324**

**ID Use Case : AC-06 Customer Segmentation for Personalized Retail Marketing**

**List Anggota :**

1. M001D5Y1463 - Naufal Akmal Rizqulloh - Machine Learning - **Aktif**
2. M001D5Y0111 - Ahmad Zaidan Al-Anshory - Machine Learning - **Aktif**
3. M298D5Y1979 - Yoga Fatiquurrahman - Machine Learning - **Aktif**

### **A. Ringkasan Eksekutif**

#### **1. Problem Statement :**

Perusahaan ritel (*Where*) saat ini mengalami penurunan efektivitas promosi (*What*) karena strategi pemasaran yang digunakan belum mempertimbangkan perilaku pelanggan secara spesifik (*Why*). Kondisi ini terjadi di era digital dengan data pelanggan yang terus bertambah dari berbagai sumber transaksi (*When*). Akibatnya, perusahaan (*Who*) menghadapi pemborosan anggaran promosi dan kesulitan mempertahankan loyalitas pelanggan (*How*). Oleh karena itu, dibutuhkan cara baru yang dapat membantu perusahaan memahami pola perilaku pelanggan untuk membuat promosi yang lebih tepat sasaran.

#### **2. Research Questions:**

Bagaimana cara mengelompokkan pelanggan berdasarkan perilaku dan nilai transaksi mereka agar perusahaan dapat menyusun strategi promosi yang lebih personal, efisien, dan berdampak pada peningkatan loyalitas pelanggan?

#### **3. Latar Belakang:**

Dunia bisnis ritel kini bergeser ke arah yang semakin kompetitif. Perusahaan berlomba-lomba memanfaatkan data pelanggan untuk memahami kebiasaan belanja dan preferensi mereka. Sayangnya, masih banyak perusahaan yang melakukan segmentasi secara umum, misalnya hanya berdasarkan umur atau lokasi. Padahal, pelanggan memiliki perilaku yang berbeda-beda dan tidak bisa diperlakukan sama. Jika hal ini dibiarkan, perusahaan bisa kehilangan pelanggan potensial dan mengeluarkan biaya

pemasaran yang besar tanpa hasil maksimal. Melalui proyek ini, tim ingin membuat model pengelompokan pelanggan yang didasarkan pada data perilaku dan nilai transaksi, sehingga perusahaan bisa mengenali pelanggan secara lebih mendalam dan membuat promosi yang sesuai kebutuhan mereka.

#### 4. Tujuan:

Tim memilih proyek ini karena topiknya sangat relevan dengan kondisi dunia ritel saat ini dan memberikan peluang besar untuk belajar langsung dari kasus nyata. Kami tertarik untuk memahami bagaimana data dapat digunakan untuk membantu perusahaan mengambil keputusan yang lebih tepat dalam pemasaran. Selain itu, kami ingin menghasilkan solusi yang benar-benar menyelesaikan masalah, bukan hanya menambah fitur. Prinsip yang kami pegang adalah membuat solusi yang menjadi painkiller, bukan sekadar vitamin. Dalam pelaksanaannya, kami menerapkan pendekatan berpikir kreatif seperti design thinking agar solusi yang kami kembangkan tidak hanya akurat secara teknis, tetapi juga bermanfaat dan mudah diterapkan di dunia kerja. Oleh karena itu, kami menawarkan solusi **PROMIND: Promotion Intelligence through RFM-Based Customer Segmentation and Machine Learning for Personalized Retail Marketing**, sebuah sistem cerdas yang membantu perusahaan memahami perilaku pelanggan dan merancang strategi promosi yang lebih personal, efisien, serta berdampak langsung pada peningkatan loyalitas pelanggan.

## B. Cakupan Proyek dan Hasil Kerja

Week	Task
1	<ol style="list-style-type: none"><li>1. Menentukan tujuan proyek, ruang lingkup, serta pembagian tanggung jawab setiap anggota tim.</li><li>2. Mengumpulkan dan memahami dataset pelanggan (misalnya Online Retail Dataset dari Kaggle atau dataset ritel serupa).</li><li>3. Melakukan data cleaning dan preprocessing (menghapus duplikasi, menangani missing values, dan standarisasi format).</li><li>4. Menyusun rancangan workflow proyek, struktur folder, dan dokumentasi awal.</li><li>5. Menyusun serta mengirimkan dokumen Project Plan kepada dosen pembimbing/mentor.</li></ol>

2	<ol style="list-style-type: none"> <li>1. Melakukan exploratory data analysis (EDA) untuk memahami pola perilaku pelanggan dan tren pembelian.</li> <li>2. Menghitung metrik RFM (<i>Recency, Frequency, Monetary</i>) dan menyiapkan fitur tambahan berbasis perilaku pelanggan.</li> <li>3. Melakukan normalisasi/standarisasi fitur untuk mendukung proses clustering.</li> <li>4. Menerapkan metode K-Means clustering untuk segmentasi pelanggan dan menentukan jumlah kluster optimal menggunakan Elbow Method atau Silhouette Score.</li> <li>5. Melakukan visualisasi awal hasil segmentasi.</li> </ol>
3	<ol style="list-style-type: none"> <li>1. Menyempurnakan hasil clustering dengan tuning parameter dan validasi hasil segmen.</li> <li>2. Menafsirkan setiap segmen pelanggan (misalnya Loyal Customers, Potential Loyalists, At Risk, Lost).</li> <li>3. Mengintegrasikan algoritma machine learning (misalnya Decision Tree atau Random Forest) atau deep learning DBSCAN untuk prediksi respons pelanggan terhadap promosi.</li> <li>4. Menyusun dashboard analitik awal untuk memvisualisasikan hasil segmentasi dan performa model.</li> </ol>
4	<ol style="list-style-type: none"> <li>1. Melakukan evaluasi performa model dengan metrik akurasi, precision-recall, dan business relevance.</li> <li>2. Menyusun rekomendasi strategi promosi personal berdasarkan hasil segmentasi dan prediksi.</li> <li>3. Mendesain solusi PROMIND Dashboard untuk membantu tim pemasaran menentukan strategi berbasis data.</li> <li>4. Menyiapkan dokumentasi kode, laporan analisis, dan marketing insight visual.</li> </ol>
5	<ol style="list-style-type: none"> <li>1. Melakukan final review model dan menyiapkan laporan akhir proyek.</li> <li>2. Merancang rencana deployment sistem PROMIND (contoh: dashboard berbasis Streamlit/Power BI dengan integrasi data ritel real-time).</li> <li>3. Menyiapkan bahan presentation deck untuk final showcase dan refleksi pembelajaran.</li> <li>4. Menyusun dokumentasi teknis dan rekomendasi pengembangan lanjutan.</li> </ol>

## C. Jadwal Pengerjaan

No	Task Name	Week 1			Week 2			Week 3			Week 4			Week 5			Week 6			Week 7		
		Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
		1 - 7 Nov			8 - 14 Nov			15- 21 Nov			21- 28 Nov			29 - 5 Des			6 - 12 Des			13-15 Des		
Machine Learning																						
Prepare and Cleaning Data: YOGA																						
1	Prepare Dataset																					
2	Data Loading & Initial Check																					
3	Exploratory Data Analysis (EDA)																					
4	Pre-processing Data																					
Modelling: NAUFAL																						
5	RFM & Data Scaling																					
6	Design Model Architecture																					
7	Train & Optimization Model																					
8	Characteristics Analysis																					
9	Segment Naming																					
Output Visualization and Recommendation: AHMAD ZAIDAN																						
10	Segment & Persona Visualization																					
11	Marketing Recommendations																					
12	Finalization & Documentation																					

## D. Uraian Rencana Penugasan/Job Desk Setiap Learning Path

Dalam use case ini, seluruh anggota tim berada pada learning path Machine Learning, sehingga fokus utama tugas adalah membangun pipeline analisis berbasis data untuk menghasilkan segmentasi pelanggan yang bermakna dan actionable. Tim Machine Learning bertanggung jawab mulai dari memahami

dan memproses data transaksi pelanggan, mengembangkan model segmentasi menggunakan pendekatan RFM (*Recency, Frequency, Monetary*) yang dikombinasikan dengan algoritma clustering seperti K-Means atau DBSCAN, hingga menghasilkan insight bisnis dan rekomendasi strategi pemasaran personalisasi. Proses ini mencakup pembersihan dan transformasi data, perhitungan metrik RFM, pemodelan dan evaluasi kluster dengan metrik seperti Silhouette Score, serta visualisasi hasil untuk mendukung pengambilan keputusan strategis. Dengan demikian, tim ML tidak hanya fokus pada aspek teknis pemodelan, tetapi juga mengonversi hasil analisis menjadi strategi pemasaran berbasis data yang relevan dan dapat diterapkan oleh perusahaan ritel online.

**Tabel Job Desk Anggota Tim**

Team	Job Desk	Output
Anggota 1: Data Understanding & Preprocessing	<ul style="list-style-type: none"> <li>- Mengambil dan memahami dataset <i>Online Retail</i> (UCI) dari Kaggle.</li> <li>- Melakukan data cleaning (menghapus duplikasi, transaksi negatif, missing values).</li> <li>- Menghitung metrik RFM (Recency, Frequency, Monetary) untuk tiap pelanggan</li> <li>- Melakukan feature scaling (MinMaxScaler atau StandardScaler).</li> </ul>	Dataset bersih dan siap model, disertai tabel RFM yang terstruktur.
Anggota 2: Modeling & Clustering	<ul style="list-style-type: none"> <li>- Membangun model segmentasi pelanggan berbasis RFM + algoritma clustering (K-Means, DBSCAN).</li> <li>- Menentukan jumlah cluster optimal menggunakan Elbow Method dan Silhouette Score.</li> <li>- Menginterpretasikan hasil tiap cluster (misal: "<i>Champions</i>", "<i>Loyal Customers</i>", "<i>At Risk</i>", "<i>Big Spenders</i>").</li> </ul>	Notebook model segmentasi RFM dengan evaluasi dan visualisasi kluster.
Anggota 3:	<ul style="list-style-type: none"> <li>- Membuat visualisasi interaktif</li> </ul>	Dashboard atau laporan

Insight  
Generation &  
Marketing  
Strategy

menggunakan  
Matplotlib/Seaborn/Plotly  
untuk menampilkan hasil  
segmentasi.

- Menghubungkan hasil model dengan marketing persona.
- Merancang rekomendasi strategi pemasaran personalisasi berbasis tiap segmen pelanggan (misalnya loyalty program, reactivation campaign, discount targeting).

insight pemasaran  
berbasis hasil  
segmentasi.

## E. Sumber Daya Proyek

Untuk mendukung kelancaran pelaksanaan proyek capstone ini, diperlukan berbagai sumber daya yang mencakup perangkat lunak, lingkungan pengembangan, dataset, dan referensi akademik yang relevan. Seluruh sumber daya dipilih berdasarkan kesesuaian dengan kebutuhan analisis data dan pemodelan machine learning pada use case segmentasi pelanggan berbasis RFM (Sari & Cengiz, 2023). Bahasa pemrograman Python digunakan karena fleksibilitas dan kelengkapan ekosistem datanya, sedangkan library seperti Scikit-learn, Pandas, dan Seaborn berperan penting dalam tahapan preprocessing, pemodelan, serta visualisasi hasil. Dataset diperoleh dari platform terpercaya seperti Kaggle dan UCI Repository, sementara referensi jurnal dan repositori GitHub digunakan untuk memperkuat landasan ilmiah serta implementasi teknis. Adapun rincian sumber daya dan fungsinya dijabarkan pada tabel berikut.

**Tabel Sumber Daya Proyek**

Resources	Tools	Objectives
Bahasa Pemrograman	Python	Bahasa utama untuk analisis data, mendukung library seperti pandas, numpy, scikit-learn. Umum digunakan dalam riset RFM dan clustering (Pranata & Saepudin, 2021).
Framework & Library Machine Learning	Scikit-learn, Pandas, Numpy, Matplotlib, Seaborn, Plotly	Scikit-learn untuk implementasi algoritma K-Means, DBSCAN, MinMaxScaler, dan evaluasi

		cluster (Elbow, Silhouette) serta (sisanya) membuat visualisasi RFM score, distribusi pelanggan, dan hasil clustering agar interpretatif bagi tim marketing.
Environment	Jupyter notebook: google colab, notebook kaggle	Menyediakan lingkungan komputasi kolaboratif berbasis cloud dengan GPU support.
Dataset	<a href="#">Dataset1</a> , <a href="#">Dataset2</a> , <a href="#">Dataset3</a> , UCI Machine Learning Repository (2015)	Dataset transaksi pelanggan retail dari Kaggle yang umum digunakan untuk analisis RFM dan segmentasi pelanggan.
Referensi Jurnal, GitHub, Dokumentasi model	Chen & Guestrin (2016), Han et al. (2012), Hosseini et al. (2010), Kohonen (2013), Nguyen et al. (2020), Sohn & Kim (2008), Xu & Wunsch (2005)	
	Github ref: <a href="#">wahyudesu</a> , <a href="#">miraytopal</a> Kaggle ref: <a href="#">animalina</a> Hugging face ref: <a href="#">qwen</a>	
Cloud backend	Google drive, GitHub & Kaggle	Penyimpanan hasil proyek, notebook, dan laporan tim agar mudah diakses lintas anggota.

## F. Rencana Manajemen Risiko dan Isu

ID	Nama Resiko	Kategori Risiko	Deskripsi Risiko ( Identifikasi )	Strategi Mitigasi ( Respons )
R1	Kualitas Data Buruk	Teknis	Dataset mengandung data duplikat, nilai kosong, atau ketidaksesuaian	Melakukan data cleaning menyeluruh, validasi data sebelum

			format yang dapat menghambat analisis.	pemodelan, dan membuat log data yang diperbarui secara berkala.
R2	Ketidakseimbangan Waktu Antar Anggota Tim	Manajemen Tim	Anggota memiliki jadwal berbeda sehingga pengerjaan proyek bisa tertunda.	Menetapkan timeline mingguan yang jelas, melakukan check-in meeting rutin, dan pembagian tugas sesuai kapasitas anggota.
R3	Kesalahan Pemilihan Jumlah Cluster	Teknis / Metodologi	Kesalahan dalam menentukan jumlah cluster optimal dapat membuat hasil segmentasi tidak akurat.	Menggunakan metode evaluasi seperti Elbow Method dan Silhouette Score untuk menentukan jumlah cluster terbaik.
R4	Keterbatasan Komputasi	Teknis	Proses pemodelan dan visualisasi pada dataset besar dapat memakan waktu lama di Colab.	Mengoptimalkan penggunaan penggunaan runtime Colab, memanfaatkan subset data saat eksperimen, dan menyimpan hasil antar proses
R5	Ketidaksesuaian Interpretasi Hasil	Analisis / Komunikasi	Interpretasi segmen pelanggan bisa berbeda antar anggota tim.	Menetapkan kriteria segmentasi yang disepakati bersama dan memverifikasi hasil dengan cross-check antar anggota sebelum finalisasi laporan.



## Referensi

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
- Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
- Hosseini, S. B., Maleki, A., & Gholamian, M. R. (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Systems with Applications, 37(7), 5259–5264. <https://doi.org/10.1016/j.eswa.2009.12.070>
- Kohonen, T. (2013). Essentials of the self-organizing map. Neural Networks, 37, 52–65. <https://doi.org/10.1016/j.neunet.2012.09.018>
- Nguyen, B., Simkin, L., & Canhoto, A. (2020). The dark side of digital personalization: An agenda for research and practice. Journal of Business Research, 116, 209–221. <https://doi.org/10.1016/j.jbusres.2020.05.046>
- Pranata, A., & Saepudin, E. (2021). Customer segmentation using RFM model and K-Means clustering: A case study on online retail dataset. International Journal of Computer Applications, 183(43), 18–23. <https://doi.org/10.5120/ijca2021921804>
- Sari, T., & Cengiz, H. (2023). Customer segmentation using Recency, Frequency, and Monetary analysis with machine learning techniques. Procedia Computer Science, 219, 108–117. <https://doi.org/10.1016/j.procs.2023.02.012>
- Sohn, S. Y., & Kim, H. S. (2008). Searching customer patterns of mobile service using clustering and quantitative association rule. Expert Systems with Applications, 34(2), 1070–1077. <https://doi.org/10.1016/j.eswa.2006.12.033>
- UCI Machine Learning Repository. (2015). Online Retail Data Set. University of California, Irvine. <https://archive.ics.uci.edu/ml/datasets/online+retail>
- Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. IEEE Transactions on Neural Networks, 16(3), 645–678. <https://doi.org/10.1109/TNN.2005.845141>