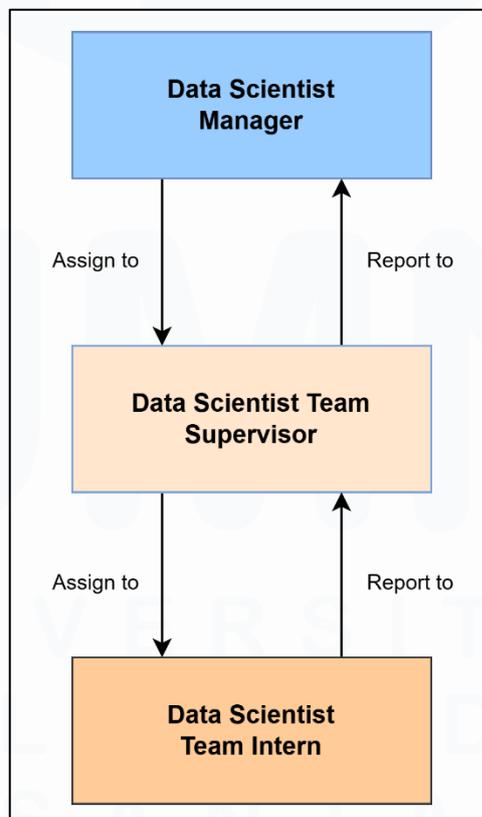


BAB III

PELAKSANAAN KERJA MAGANG

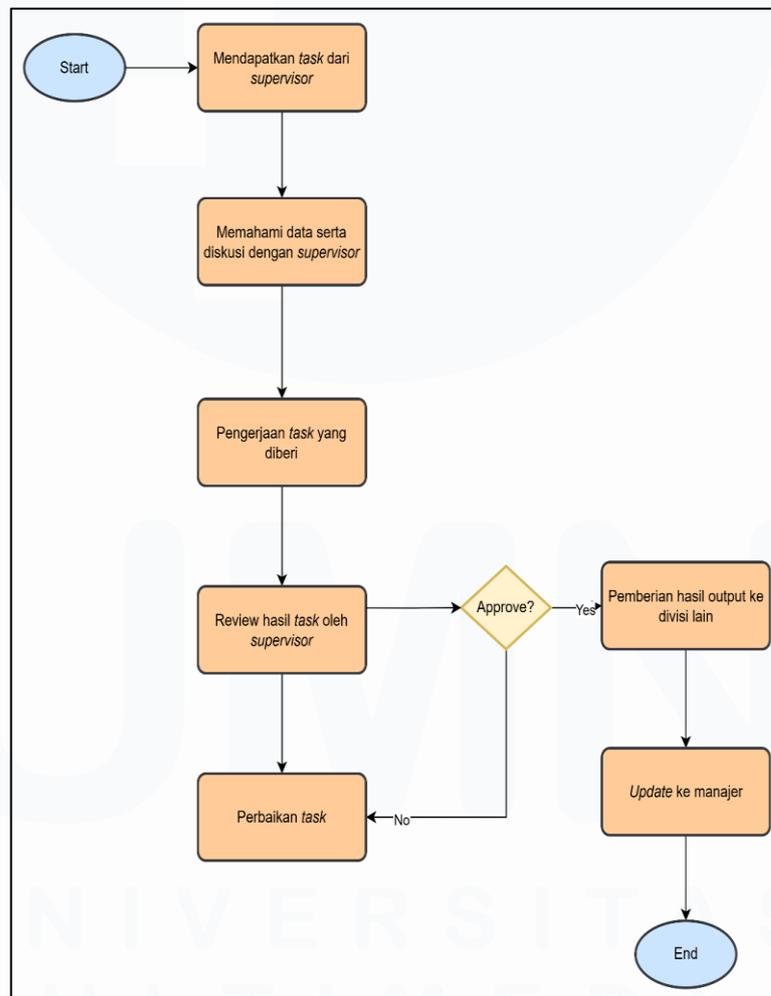
3.1 Kedudukan dan Koordinasi

Posisi yang diduduki selama pelaksanaan kerja magang di PT Global Loyalty Indonesia adalah posisi *intern data scientist*. Pekerjaan utama mencakup mengubah data mentah menjadi data informatif yang dapat berguna bagi perusahaan untuk tumbuh dan berkembang, serta mengartikan data hasil analisa dari berbagai sumber untuk memberi solusi langkah ke depan. Alur koordinasi tim data *scientist* dapat dilihat pada gambar 3.1. Pemberian pekerjaan dimulai dari data *scientist manager* yang saat ini sedang diduduki oleh Bapak Denny Desanleon Yuwono dan diberi kepada *supervisor* yaitu, Bapak Ivan Nur Amanda, yang pada akhirnya diberikan kepada *intern data scientist*.



Gambar 3.1 Alur Koordinasi Tim Data Science

Koordinasi pelaporan pekerjaan pun terjadi sebaliknya, dari *intern data scientist* melaporkan pekerjaan kepada *supervisor*, yang pada akhirnya dilaporkan kepada data *scientist manager* juga. Koordinasi serta komunikasi yang dilakukan juga menggunakan *platform online* seperti Trello. Trello digunakan untuk memberikan *update* pengerjaan *task* yang dikerjakan, yang pada akhirnya akan di *review* bersama saat *weekly meeting* di setiap hari senin. Alur pelaksanaan kerja magang dapat dilihat pada gambar 3.2. Alur tersebut menunjukkan secara detail bagaimana proses pengerjaan sebuah *task* dikerjakan oleh *intern data scientist*.



Gambar 3.2 Alur Pelaksanaan Kerja Magang

Alur pelaksanaan kerja magang dimulai dengan mendapatkan *task* dari *supervisor*. *Task* yang diberi dapat mencakup data-data atau bahkan *query* yang dibutuhkan atau dapat dimanfaatkan untuk pengerjaan *task*. Pelaksanaan kerja pun dilanjut dengan pemahaman data serta diskusi dengan *supervisor* mengenai *task* yang akan dikerjakan. Setelah itu, pengerjaan *task* dimulai, dan setelah itu akan dilaporkan dan di *review* oleh *supervisor*. Jika *task* yang telah dikerjakan di *approve* maka hasil *output* dari pengerjaan *task* pun akan di beri kepada divisi lain yang membutuhkan, setelah itu, *supervisor* akan *update* ke manajer. Jika saat proses *review* tidak di *approve*, maka *intern data scientist* akan melakukan perbaikan *task* yang mana akan diikuti dengan *review* oleh *supervisor* kembali.

3.2 Tugas dan Uraian Kerja Magang

3.2.1 Tugas Kerja Magang

Selama masa magang, fokus utama pekerjaan berada pada pengerjaan program *targeted voucher* yang dijalankan oleh PT Global Loyalty Indonesia. Program ini bertujuan untuk mengirimkan *voucher* secara lebih tepat sasaran kepada *member* berdasarkan perilaku dalam aplikasi Alfagift mereka. Terdapat dua kategori produk utama yang menjadi fokus pengerjaan program ini, yaitu produk ‘tagI’ yang merupakan produk *online* dan hanya dapat dibeli melalui aplikasi Alfagift, yang mencakup kategori ‘GROWING UP MILK SUPER PREMIUM’ dan kategori *toiletries*. Kegiatan yang dilakukan mencakup seluruh proses pendukung program tersebut, mulai dari pemahaman data, pembangunan model prediktif, hingga pengolahan dan integrasi *output* model untuk menghasilkan daftar *member* yang menjadi target *voucher* setiap bulan. Selain itu, dilakukan pula analisis lanjutan terhadap perilaku pelanggan serta evaluasi performa *voucher* yang telah dikirimkan.

3.2.2 Uraian Kerja Magang

Tabel 3.1 berisi rincian pekerjaan serta waktu pelaksanaan kegiatan selama masa magang di PT Global Loyalty Indonesia. Kegiatan diawali dengan pengenalan perusahaan, pemahaman struktur kerja tim

data, serta proses setup *tools* yang digunakan dalam pekerjaan sehari-hari, seperti Jupyter Notebook dan SQL Server. Setelah itu, dilakukan penarikan dan eksplorasi data transaksi pelanggan untuk memahami perilaku *member* dalam aplikasi Alfagift. Pekerjaan kemudian berlanjut pada proses pembangunan model *machine learning* untuk memprediksi pelanggan yang berpotensi menerima *voucher*, di mana hasil klasifikasi model diolah dan digabungkan dengan informasi *member* untuk membentuk daftar target *voucher*. Proses ini dilakukan secara berkala untuk mendukung pengiriman *voucher* bulanan, termasuk *voucher* pada bulan Maret, April, Mei, dan Juni 2025.

Selain pengolahan data dan pemodelan, dilakukan pula analisis visualisasi dan segmentasi perilaku pelanggan, seperti mengidentifikasi loyalitas *member* dan kebiasaan pembelian pada kategori tertentu. Salah satu fokus analisis diarahkan pada kategori produk seperti ‘GROWING UP MILK SUPER PREMIUM’ dan *toiletries*, yang termasuk dalam cakupan program *targeted voucher*. Proses interpretasi model dan analisis lanjutan juga dilakukan untuk memahami lebih dalam hasil prediksi serta efektivitas strategi yang diterapkan. *Monitoring redeemed voucher* menjadi bagian penting dalam evaluasi program, di mana data *redeem voucher* digunakan untuk mengukur tingkat keberhasilan dan perilaku pelanggan.

Tabel 3.1 Tugas dan Uraian Kerja Magang

No.	Pekerjaan	Waktu Pengerjaan
1	Pengenalan perusahaan, setup tools, dan penarikan data sesuai <i>request</i> .	3 Februari – 7 Februari (Minggu 1)
2	Eksplorasi data, pemahaman proyek, dan pembangunan model <i>machine learning</i> untuk program <i>voucher targeted</i> .	10 Februari – 21 Februari (Minggu 2-3)
3	Pengolahan <i>output</i> model dan penggabungan <i>list member</i> untuk <i>voucher targeted</i> Maret 2025.	17 Februari – 21 Februari (Minggu 3)

No.	Pekerjaan	Waktu Pengerjaan
4	Eksplorasi data, visualisasi, dan analisis loyalitas <i>member</i> .	24 Februari – 7 Maret (Minggu 4-5)
5	Data <i>preparation</i> dan modeling untuk <i>voucher targeted</i> April 2025.	10 Maret – 21 Maret (Minggu 6-7)
6	Pengolahan <i>output</i> model dan penggabungan <i>list member</i> untuk <i>voucher targeted</i> April 2025.	17 Maret – 21 Maret (Minggu 7)
7	Interpretasi model <i>targeted</i> produk <i>toiletries</i>	17 Maret – 21 Maret (Minggu 7)
8	Analisa <i>behavior member targeted toiletries</i> April 2025.	24 Maret – 4 April (Minggu 8-9)
9	Data <i>preparation</i> dan modeling untuk <i>voucher targeted</i> Mei 2025	7 April – 18 April (Minggu 10-11)
10	Pengolahan <i>output</i> model dan penggabungan <i>list member</i> untuk <i>voucher targeted</i> Mei 2025.	21 April – 25 April (Minggu 12)
11	Analisis transaksi <i>member targeted</i> dan <i>breakdown</i> kategori <i>toiletries</i>	28 April – 2 Mei (Minggu 13)
12	<i>Monitoring voucher redeemed</i> bulan April 2025	5 Mei – 16 Mei (Minggu 14-15)
13	Data <i>preparation</i> , modeling, dan pengolahan <i>output</i> untuk <i>voucher targeted</i> Juni 2025.	5 Mei – 23 Mei (Minggu 14-16)

3.2.2.1 Pengenalan Perusahaan, Setup Tools, dan Penarikan Data Sesuai Request

Minggu pertama pelaksanaan kerja magang dimulai dengan pengenalan profil perusahaan, sistem kerja, dan lingkungan kerja dengan adanya penjelasan dari tim HC (*Human Capital*) GLI. Pengenalan tersebut pun diikuti dengan perkenalan dengan tim data *scientist* termasuk manajer, *supervisor* atau mentor. Selanjutnya, mentor menjelaskan *tools* atau *software* yang nantinya dapat digunakan dalam pelaksanaan pekerjaan. Instalasi hingga setup *tools* seperti Google Bigquery, server Jupyter Notebook, dan DBeaver juga dilakukan dengan arahan mentor. Proses-proses tersebut dirancang untuk memberikan pemahaman mengenai bagaimana perusahaan beroperasi, bagaimana

alur kerja dalam tim data *scientist* berlangsung, serta bagaimana *tools* dan *software* yang digunakan dapat mendukung pelaksanaan tugas dan proyek selama masa magang. Dengan demikian, adaptasi dan pemahaman terhadap peran serta tanggung jawab dalam setiap tahap pekerjaan yang diberikan dapat dilakukan dengan cepat.

Setelah proses pengenalan serta setup tersebut dilakukan, penjelasan mengenai proyek yang akan di *handle* juga dijelaskan oleh manajer dan juga *supervisor*. Selain pemahaman mengenai proyek yang akan dikerjakan, ada juga *user request* dari divisi lain yang meminta untuk menarik atau melengkapi data. Gambar 3.3 memperlihatkan Microsoft Excel yang diberikan sebagai *request* melengkapi data. Terdapat 5 kolom yaitu, ‘Kode Registrasi’, ‘Tanggal’, ‘Jam’, ‘Branch’, dan juga ‘Nama Branch’. Karena isi dari kolom ‘Branch’ dan ‘Nama Branch’ itu tidak ada (*missing value*), maka tim *data science* diminta untuk mengisinya berdasarkan kode registrasi yang ada.

	A	B	C	D	E
	Kode Registrasi	Tanggal	Jam	Branch	Nama Branch
1					
2		6-Jan-25	4:40:07	N#A	None
3		6-Jan-25	7:59:30	N#A	None
4		6-Jan-25	9:09:32	N#A	None
5		6-Jan-25	9:17:58	N#A	None
6		6-Jan-25	9:23:10	N#A	None
7		6-Jan-25	10:15:33	N#A	None
8		6-Jan-25	10:55:08	N#A	None
9		6-Jan-25	11:10:55	N#A	None
10		6-Jan-25	14:16:00	N#A	None
11		6-Jan-25	15:25:22	N#A	None
12		6-Jan-25	17:00:46	N#A	None
13		6-Jan-25	17:05:38	N#A	None
14		6-Jan-25	17:39:19	N#A	None
15		6-Jan-25	18:27:38	N#A	None
16		6-Jan-25	18:47:34	N#A	None
17		6-Jan-25	19:58:29	N#A	None
18		7-Jan-25	6:40:35	N#A	None
19		7-Jan-25	7:33:44	N#A	None
20		7-Jan-25	7:41:52	N#A	None
21		7-Jan-25	8:35:38	N#A	None
22		7-Jan-25	8:54:13	N#A	None
23		7-Jan-25	10:10:40	N#A	None

Gambar 3.3 Request Melengkapi Data

Gambar 3.4 menunjukkan tangkapan layar salah satu *tools* yang digunakan oleh tim data *scientist* di GLI untuk menarik, melihat, atau membuat *query* yaitu, Google BigQuery. Dengan Google BigQuery tersebut tim data *scientist* dapat mengakses data – data dari aplikasi

Alfagift sendiri. Dalam gambar 3.4 terdapat *query* menarik data kode cabang, dan nama cabang yang ada di dalam kode *voucher* yang tertera.

```

1 select distinct [REDACTED].VOUCHER_CODE,
2 [REDACTED].BRANCH_CODE, (select branch_name from [REDACTED] mb
3 where branch_code = [REDACTED].branch_code) branch_name
4 from [REDACTED]
5 where
6 [REDACTED].DATE >= '2024-10-01'
7 and [REDACTED].VOUCHER_CODE IN (

```

Gambar 3.4 Query Menarik Data

Query tersebut pun menghasilkan *output* tabel yang berisikan kolom kode *voucher* sesuai yang tertera, kode cabang, dan juga nama cabangnya. Setelah itu, tabel tersebut di *import* menjadi *dataframe* pada Jupyter Notebook, yang pada akhirnya dilakukannya *LEFT JOIN* dengan tabel pada Gambar 3.3 agar ‘Branch’ dan ‘Nama Branch’ terisi yang dapat dilihat pada Gambar 3.5.

	A	B	C	D	E
	Kode Registrasi	Tanggal	Jam	Branch	Nama Branch
2		2025-01-06 00:00:00	04:40:07	2PZ1	
3		2025-01-06 00:00:00	07:59:30	2AZ1	
4		2025-01-06 00:00:00	09:09:32		
5		2025-01-06 00:00:00	09:17:58	VZ01	
6		2025-01-06 00:00:00	09:23:10	OZ01	
7		2025-01-06 00:00:00	10:15:33	2AZ1	
8		2025-01-06 00:00:00	10:55:08	2JZ1	
9		2025-01-06 00:00:00	11:10:55	TZ01	
10		2025-01-06 00:00:00	14:16:00		
11		2025-01-06 00:00:00	15:25:22	TZ01	
12		2025-01-06 00:00:00	17:00:46	OZ01	
13		2025-01-06 00:00:00	17:05:38	1YZ1	
14		2025-01-06 00:00:00	17:39:19	PZ01	
15		2025-01-06 00:00:00	18:27:38		
16		2025-01-06 00:00:00	18:47:34	1AZ1	
17		2025-01-06 00:00:00	19:58:29		
18		2025-01-07 00:00:00	06:40:35		
19		2025-01-07 00:00:00	07:33:44		
20		2025-01-07 00:00:00	07:41:52	2JZ1	
21		2025-01-07 00:00:00	08:35:38	2MZ1	
22		2025-01-07 00:00:00	08:54:13		
23		2025-01-07 00:00:00	10:10:40	VZ01	
24		2025-01-07 00:00:00	10:12:20	BZ01	
25		2025-01-07 00:00:00	10:50:54		

Gambar 3.5 Output Data Request

3.2.2.2 Eksplorasi data, pemahaman proyek, dan pembangunan model *machine learning* untuk program *voucher targeted*

Selama masa pelaksanaan kerja magang, proyek data *science* di-handle secara langsung. Proyek yang diberikan adalah salah satu program yang berjalan pada PT GLI sendiri. Program ini disebut *voucher targeted outer Jawa*. Program ini bertujuan untuk membantu salah satu permasalahan yang dialami oleh PT GLI, yaitu adanya persaingan dengan kompetitor, PT GLI mengusahakan untuk tetap menaikkan *sales* di luar pulau Jawa pada kategori tertentu. Program ini menargetkan *member-member* yang berada di luar pulau Jawa untuk diberikan *voucher* pembelian pada aplikasi Alfagift.

Pada minggu kedua, data yang akan digunakan untuk proyek telah diberikan. Data tersebut merupakan data transaksi *member* yang sebelumnya sudah pernah tertarget. Data ini mencakup informasi mengenai penggunaan *voucher*, demografi *member*, detail transaksi, perilaku/event *member* pada aplikasi Alfagift, serta waktu pembelian. Proses penggunaan data dilakukan melalui Jupyter/Python, dimulai dengan mengimpor *library* Python yang diperlukan, seperti ditunjukkan pada Gambar 3.6.

```
In [1]: 1 import warnings
        2 import sys
        3 import csv
        4 import os
        5 import pandas as pd
        6 sys.path.append('')
        7 if not sys.warnoptions:
        8     warnings.simplefilter("ignore")
        9     os.environ["PYTHONWARNINGS"] = "ignore"
        10
        11 import ds_db
        12 import helper_db as helper_db
        13 from helper_db import read_bq
```

Gambar 3.6 Import Library Python

File path khusus ditambahkan ke dalam Python path agar *script* dapat mengakses modul internal perusahaan, yaitu *ds_db* dan *helper_db*, yang berisi fungsi-fungsi bantu untuk koneksi dan pengambilan *data* dari

BigQuery. Fungsi `read_bq` yang diimpor dari `helper_db` digunakan untuk menjalankan *query* SQL ke BigQuery dan mengembalikan hasilnya dalam bentuk *dataframe*, sehingga dapat dianalisis lebih lanjut menggunakan Python.

```

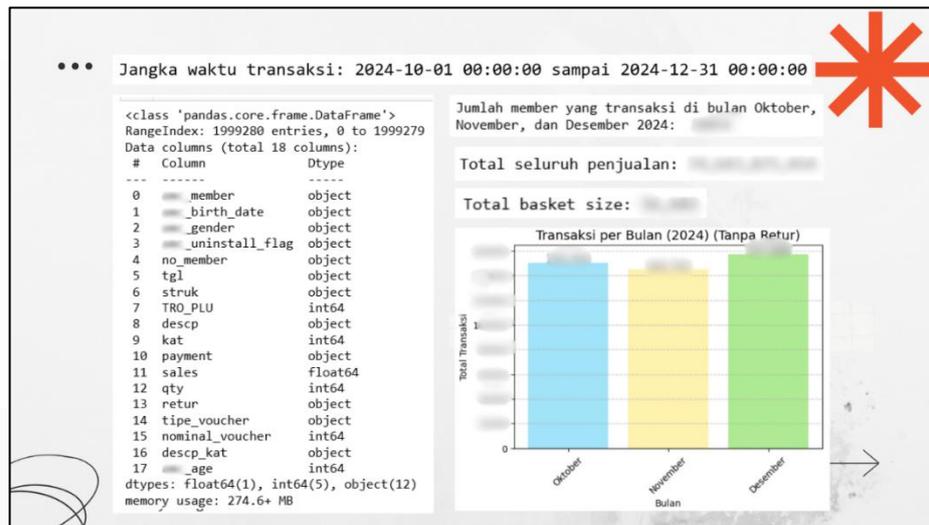
1 q=f'''
2 SELECT A._member, A._birth_date, A._gender,A._uninstall_flag, B.*
3 FROM
4 RIGHT JOIN A
5 ON A._member = B.no_member;
6 result = read_bq(q)
7 result

```

	_member	_birth_date	_gender	_uninstall_flag	no_member	tgl	struk	TRO_PLU	descp	kat	payment
0				N					KIN YOG RICH BERRIES PET 200ML		Aladin
1				N					BRICKS DELIVERY TRUCK		Aladin
2				N					BRICKS ALBI		Aladin
3				N					AQUA AIR MNRL BKL NAS GLN 19L		Aladin
4				N					BRICKS ALFAMART STORE		Aladin
...
1999275				N					BEBELAC 4 VANILLA 1.8KG (OL)		transfer-mandiri
1999276				N					BEBELAC 4 VANILLA 1.8KG (OL)		transfer-mandiri

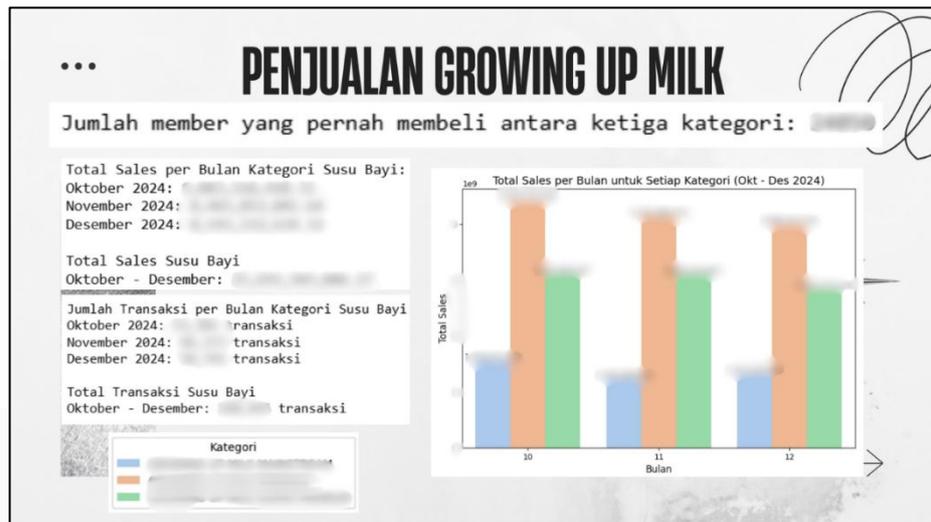
Gambar 3.7 Menarik Data

Gambar 3.7 menampilkan proses penarikan data di Jupyter melalui BigQuery. *Query* SQL yang ditulis menggabungkan dua tabel, yaitu tabel yang diberi alias 'A' dan alias 'B' dengan menggunakan *RIGHT JOIN*. Tujuan dari *query* tersebut adalah untuk mengambil semua data dari tabel 'B' dan menambahkan data tambahan 'birth_date', 'gender', dan juga 'uninstall_flag' berdasarkan nomor-nomor *member* yang terdapat di tabel 'A'.



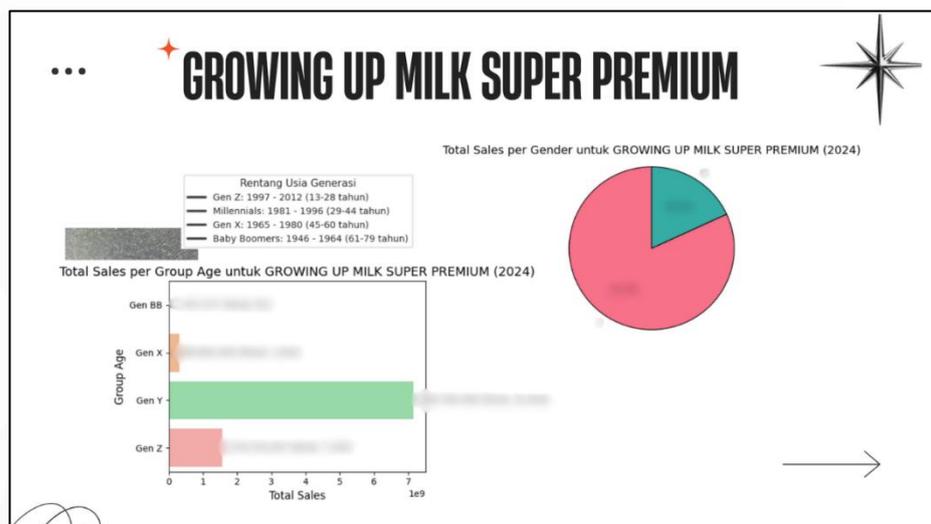
Gambar 3.8 Eksplorasi Data

Setelah data ditarik, langkah awal yang dilakukan yaitu eksplorasi data secara menyeluruh untuk memahami struktur, jumlah data, tipe data, serta konteks dari data tersebut. Analisis dilakukan menggunakan *library* Python seperti Pandas, NumPy, dan Matplotlib/Seaborn untuk menggambarkan distribusi data serta hubungan antar variabel. Gambar 3.8 menampilkan hasil eksplorasi data yang telah dilakukan. Hasil analisis tersebut juga dipresentasikan kepada tim data *science* pada saat *weekly meeting*. Gambar 3.8 atau *slide* pertama menampilkan fitur-fitur yang terdapat pada tabel ‘raw’ beserta tipe data masing-masing kolom, jangka waktu transaksi, dan *insights* umum mengenai data transaksi.



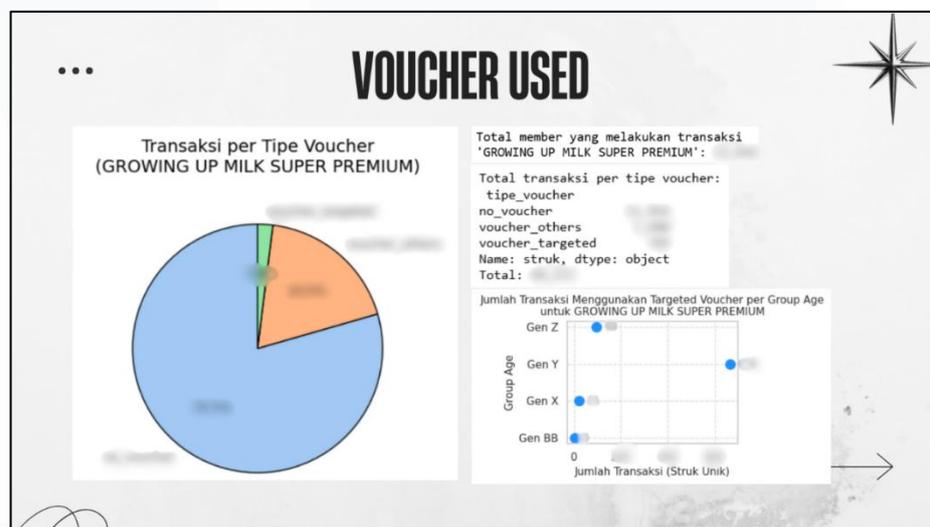
Gambar 3.9 *Insights* Kategori Susu Pertumbuhan

Gambar 3.9 menampilkan *insights* mengenai kategori ‘GROWING UP MILK’ atau kategori susu pertumbuhan, seperti jumlah *member* yang pernah membeli kategori susu pertumbuhan, total *sales*, dan jumlah transaksi selama tiga bulan. Setelah detail penjualan susu pertumbuhan telah diketahui, maka analisa dikerucutkan lagi ke analisa kategori susu pertumbuhan *super premium* atau ‘GROWING UP MILK SUPER PREMIUM’ seperti pada gambar 3.10.



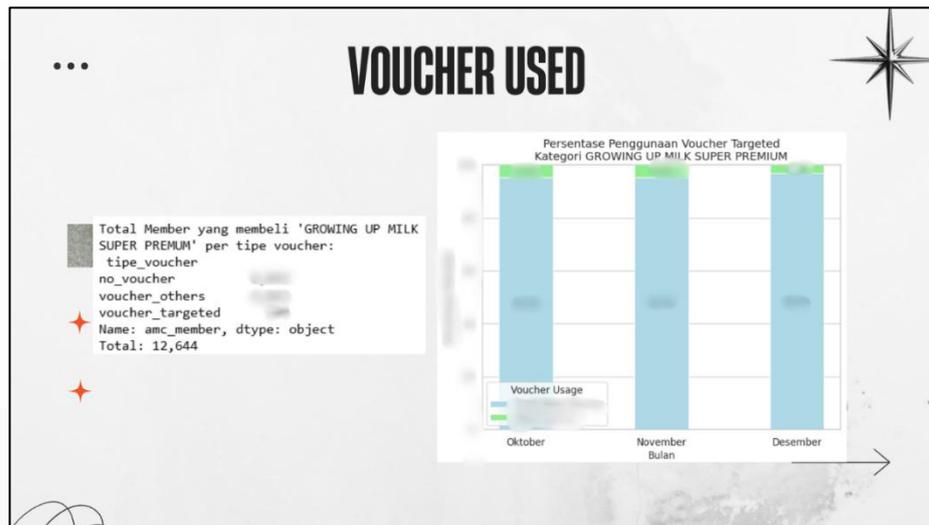
Gambar 3.10 *Insights* Kategori *Growing Up Milk Super Premium*

Terdapat *pie chart* yang menunjukkan *total sales per gender* serta *bar chart total sales per group age* pada Gambar 3.10. Visualisasi ini bertujuan untuk memberikan gambaran umum mengenai profil konsumen berdasarkan jenis kelamin dan kelompok usia pada kategori ‘GROWING UP MILK SUPER PREMIUM’. Dengan penyajian data ini, perusahaan dapat memperoleh *insight* awal mengenai segmentasi konsumen dan potensi target pasar, sebagai dasar untuk analisa lebih lanjut dan pengambilan keputusan strategis di bidang pemasaran produk susu pertumbuhan premium.



Gambar 3.11 *Insights* Penggunaan *Voucher*

Gambar 3.11 menampilkan *insights* mengenai penggunaan *voucher* oleh *member*. Visualisasi ini bertujuan untuk memberikan gambaran umum tentang penggunaan *voucher* yang ditujukan secara spesifik kepada segmen tertentu. Data yang disajikan mencakup informasi seperti jumlah transaksi yang telah menggunakan *voucher targeted*, tidak menggunakan *voucher*, maupun yang menggunakan *voucher* lain. Frekuensi penggunaan *voucher* berdasarkan *group age* juga dapat dilihat pada *chart* yang dibuat merancang program promosi yang lebih tepat sasaran di masa mendatang.



Gambar 3.12 *Insights Penggunaan Voucher Targeted*

Gambar 3.12 menampilkan informasi lebih spesifik mengenai penggunaan *voucher targeted* oleh *member* yang membeli produk kategori ‘GROWING UP MILK SUPER PREMIUM’. Visualisasi ini menunjukkan total *member* yang melakukan pembelian berdasarkan tipe *voucher* yang digunakan, sehingga dapat membantu dalam mengidentifikasi efektivitas masing-masing jenis *voucher* dalam mendorong pembelian pada segmen produk *premium*.

```

df1['_PLU'] = df1['_PLU'].astype(int)
df1['kat'] = df1['kat'].astype(int)
df1['sales'] = df1['sales'].astype(float)
df1['qty'] = df1['qty'].astype(int)
df1['nominal_voucher'] = df1['nominal_voucher'].astype(int)

df['_member'] = df['_member'].astype(str)
df['no_member'] = df['no_member'].astype(str)
df['tgl'] = pd.to_datetime(df['tgl'])
df['_birth_date'] = pd.to_datetime(df['_birth_date'])

df['_gender'] = df['_gender'].astype(str)
df['_uninstall_flag'] = df['_uninstall_flag'].astype(str)
df['struk'] = df['struk'].astype(str)
df['descp'] = df['descp'].astype(str)
df['payment'] = df['payment'].astype(str)
df['retur'] = df['retur'].astype(str)
df['tipe_voucher'] = df['tipe_voucher'].astype(str)
df['descp_kat'] = df['descp_kat'].astype(str)

```

Gambar 3.13 *Convert Data Type*

Setelah eksplorasi awal dilakukan, tahap data *preparation* juga dilaksanakan, termasuk penanganan *missing values*, penghapusan duplikasi, serta konversi tipe data yang tidak sesuai. Konversi tipe data dilakukan untuk memastikan penggabungan data dapat berjalan dengan baik. Gambar 3.13 menampilkan potongan kode Python yang digunakan untuk mengganti tipe data pada dua tabel, yaitu tabel 'df' dan 'df1'. Tabel 'df' berisi data transaksi *member targeted* pada periode Oktober 2024 hingga Desember 2024, sedangkan tabel 'df1' merupakan data transaksi *member targeted* pada bulan Januari 2025.

```
df = pd.concat([df, df1], ignore_index=True)
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2663336 entries, 0 to 2663335
Data columns (total 17 columns):
#   Column                Dtype
---  -
0   _member                object
1   _birth_date            datetime64[ns]
2   _gender                object
3   _uninstall_flag       object
4   no_member              object
5   tgl                    datetime64[ns]
6   struk                  object
7   ._PLU                  int64
8   descp                  object
9   kat                    int64
10  payment                object
11  sales                  float64
12  qty                    int64
13  retur                  object
14  tipe_voucher           object
15  nominal_voucher        int64
16  descp_kat              object
dtypes: datetime64[ns](2), float64(1), int64(4), object(10)
memory usage: 345.4+ MB
```

Gambar 3.14 *Concat* tabel

Setelah kedua tabel dipastikan memiliki tipe data yang sama, maka proses *concat* pun dilakukan seperti yang ditampilkan pada gambar 3.14. Fitur-fitur yang tidak relevan juga dihapus demi meningkatkan performa model nantinya. Tujuan utama model adalah untuk mengklasifikasi apakah seorang *member* akan menukarkan (*redeem*) *voucher* yang akan diberikan. Untuk itu, *feature engineering* juga

dilakukan untuk membuat variabel-variabel baru yang berpotensi meningkatkan akurasi klasifikasi.

```
11]: df['redeemed_voucher'] = df.apply(
    lambda row: 1 if row['tipe_voucher'] == 'voucher_targeted' and
        row['descp_kat'] == 'GROWING UP MILK SUPER PREMIUM'
        else 0,
    axis=1
)

12]: #Total transaksi member
df['jumlah_transaksi'] = df.groupby('_member')['struk'].transform('count')

13]: #Total pembelian member
df['total_sales'] = df.groupby('_member')['sales'].transform('sum')

14]: #Basket size semua transaksi
df['basket_size'] = df['total_sales'] / df['jumlah_transaksi']

15]: #Pernah membeli produk growing up milk sebelumnya atau tidak
produk_kategori = ['GROWING UP MILK MAINSTREAM', 'GROWING UP MILK PREMIUM', 'GROWING UP MILK SUPER PREMIUM']
df['purchased_gum'] = df['descp_kat'].isin(produk_kategori)
df['purchased_gum'] = df.groupby('_member')['purchased_gum'].transform('max')

16]: #Jumlah voucher yang pernah digunakan, targeted ataupun voucher others
df['total_voucher_per_member'] = df.groupby('_member')['tipe_voucher'].transform(lambda x: x.isin(
    ['voucher_others', 'voucher_targeted']).sum())

17]: #jumlah retur yang telah dilakukan member
df['jumlah_retur'] = df.groupby('_member')['retur'].transform(lambda x: (x == 'Retur').sum())
```

Gambar 3.15 *Feature Engineering*

Gambar 3.15 menampilkan beberapa potongan kode Python yang digunakan dalam proses *feature engineering* untuk menyiapkan data sebelum dilakukan pemodelan. Pembuatan kolom-kolom baru seperti 'redeemed_voucher' sebagai target klasifikasi, serta fitur tambahan seperti 'jumlah_transaksi', 'total_sales', 'basket_size', 'purchased_gum', 'total_voucher_per_member', 'jumlah_retur', dan 'qty_purchased' per *member* bertujuan untuk memberikan informasi atau data yang lebih lengkap kepada model. Dengan adanya fitur-fitur tersebut, diharapkan algoritma *machine learning* dapat mengenali pola dan hubungan dalam data lebih baik sehingga model dapat lebih akurat dalam mengklasifikasi kemungkinan seorang *member* akan *redeem voucher*.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

```

22]: df_per_member = df.groupby('member').agg({
    'birth_date': 'first',
    'gender': 'first',
    'uninstall_flag': 'first',
    'tgl': 'min',
    'age': 'first',
    'Group_age': 'first',
    'redeemed_voucher': 'max',
    'jumlah_transaksi': 'first',
    'total_sales': 'first',
    'basket_size': 'first',
    'purchased_gum': 'first',
    'total_voucher_per_member': 'first',
    'jumlah_retur': 'first',
    'qty_purchased': 'first'
}).reset_index()

df_per_member

```

Gambar 3.16 Agregasi Data

Gambar 3.16 menampilkan proses agregasi data menggunakan metode *groupby* pada Python. Dalam proses ini, data dikelompokkan berdasarkan kolom nomor *member* agar setiap baris merepresentasikan satu *member* unik. Kemudian dilakukan agregasi dengan fungsi yang disesuaikan untuk setiap kolom menggunakan metode *.agg()*. Misalnya, kolom 'birth_date', 'gender', dan 'age' menggunakan fungsi 'first' karena nilainya tidak berubah per transaksi dan cukup diambil satu kali sebagai identitas *member*. Kolom 'tgl' menggunakan fungsi 'min' untuk mendapatkan tanggal transaksi pertama. Sementara kolom 'redeemed_voucher' menggunakan fungsi 'max' untuk mengetahui apakah *member* tersebut pernah melakukan penukaran *voucher* (nilai 1) atau tidak sama sekali (nilai 0). Kolom fitur yang telah ditambahkan sebelumnya juga menggunakan fungsi 'first' karena sebelumnya sudah dibuat sesuai per *member*. Proses agregasi ini dilakukan untuk merubah data yang awalnya per transaksi menjadi per *member*, sehingga dapat digunakan dalam pemodelan klasifikasi.

```
pernah_liat = pd.read_csv('Angelin/pernah_liat.csv')
pernah_liat.head(3)
```

	EVENT_NAME	EVENT_UUID	member_id	plu	descp	cart_id	cart_source	date_event	next_event_time	prev_event_time
0	view_product				S-26 ULTIMA PROCAL CAN 850G(OL)	NaN	NaN	2025-01-01 15:19:32	2025-01-01 15:20:21	2025-01-01 15:18:50
1	view_product				S-26 ULTIMA PROCAL CAN 850G(OL)	NaN	NaN	2025-01-01 15:20:21	2025-01-01 18:15:51	2025-01-01 15:19:32
2	view_product				S-26 ULTIMA PROCAL CAN 850G(OL)	NaN	NaN	2025-01-01 18:15:51	NaN	2025-01-01 15:20:21

Gambar 3.17 Tabel 'pernah_liat'

Setelah data transaksi dan profil berhasil diagregasi menjadi per *member*, fitur tambahan ditambahkan seperti yang ditunjukkan pada Gambar 3.17, yaitu tabel `pernah_liat` yang memuat informasi mengenai aktivitas *member* pada aplikasi Alfagift yang melakukan melihat produk atau `view_product`. Tabel ini memiliki beberapa kolom penting, antara lain 'EVENT_NAME' yang menunjukkan jenis *event* (dalam hal ini `view_product`), 'EVENT_UUID' sebagai identifikasi unik untuk setiap *event*, 'member_id' untuk mengidentifikasi pengguna, serta 'plu' dan 'descp' yang menggambarkan kode dan deskripsi produk yang dilihat (misalnya S-26 ULTIMA PROCAL CAN 850G(OL)). Selain itu, terdapat kolom 'cart_id' dan 'cart_source' yang berkaitan dengan keranjang belanja dalam aplikasi, meskipun dalam contoh data ini nilainya masih NaN atau kosong. Kolom 'date_event' mencatat waktu terjadinya *event*, sementara 'next_event_time' dan 'prev_event_time' memberikan konteks waktu terhadap aktivitas yang terjadi sebelum dan sesudahnya.

Tabel `pernah_liat` ini digunakan untuk menambahkan fitur *event* baru yaitu 'view_product' pada data transaksi dan profil. Apabila nomor *member* ada dalam tabel `pernah_liat`, maka nilai pada kolom fitur akan diisi dengan 1. Sebaliknya, jika nomor *member* tidak terdapat dalam tabel tersebut, maka nilai yang diberikan adalah 0. Fitur ini membantu dalam mengidentifikasi apakah seorang *member* pernah melihat produk susu premium, yang dapat membantu model dalam analisis pola dalam data.

```
target_dan_affinity = pd.read_csv('Angelin/buy_affinity_and_target_together.csv')
target_dan_affinity.head(3)
```

	ponta_id	struk	date	item1	descp_kategori1	item2	descp_kategori2
0			2024-10-25	BEBELAC GOLD 3 VANILA BOX 700G	GROWING UP MILK PREMIUM	MY BABY SHP BLACK&SHINE 100ML	BABY SHAMPOO
1			2024-10-25	CHIL-KID 3DHA VNL BOX 190G	GROWING UP MILK PREMIUM	CUSSONS B.COLG SPARKL JOY 100ML	BABY COLOGNE
2			2024-10-25	SGM EKSPLOK SOYA VANILA BOX700G	GROWING UP MILK MAINSTREAM	ZWITSAL B.SHP ALOE VERA 100ML	BABY SHAMPOO

Gambar 3.18 Tabel 'target_dan_affinity'

Gambar 3.18 menampilkan tabel 'target_dan_affinity' yang berisi data aktivitas *member* yang pernah melakukan pembelian produk susu pertumbuhan serta produk *affinity*. Produk *affinity* merupakan produk yang berkaitan dengan kategori tertentu, dalam hal ini adalah yang berkaitan dengan kategori susu pertumbuhan. Contoh produk *affinity* susu pertumbuhan meliputi popok, sampo bayi, parfum bayi, dan produk perawatan bayi lainnya. Sama seperti pada tabel 'pernah_liat', apabila nomor *member* ditemukan dalam tabel ini, maka nilai pada fitur akan diisi dengan 1. Sebaliknya, jika tidak ditemukan, maka nilainya adalah 0.

```
buy_affinity = pd.read_csv('/buy_affinity.csv')
buy_affinity.head(3)
```

	date	struk_affinity	ponta_id	plu	descp_kat	descp
0	2024-11-19			GROWING UP MILK MAINSTREAM	CHILGO PWD3+ VNL BOX 950G (OL)	
1	2024-11-19				BABY CEREAL	SUN BRS MERAH EKO SAC 120G
2	2024-11-19				PANTS DIAPERS	MIU BABY PANTS XL-24

Gambar 3.19 Tabel 'buy_affinity'

Gambar 3.19 menunjukkan tabel *buy_affinity* yang memuat data aktivitas pembelian produk *affinity* yang dilakukan secara terpisah, tanpa disertai pembelian produk target ('GROWING UP MILK SUPER PREMIUM'). Fitur ini ditambahkan ke dalam data transaksi dan profil *member* dengan pendekatan yang sama seperti pada tabel aktivitas sebelumnya, yaitu dengan memberikan nilai 1 jika *member* tercatat dalam tabel *buy_affinity*, dan 0 jika tidak.

```
repurchased = pd.read_csv('Angelin/ beli_lg_pas_obis.csv')
repurchased.head(3)
```

	ponta_id	descp	QTY	date	next_event_time	prev_event_time	perkiraan_habis_hari	perkiraan_habis_hari_multiplied	selisih_waktu_next
0		CHIL-KID 3DHA VNL BOX 780G	1	2024-11-27	2024-12-03	2024-11-07	7.222222	7.222222	6
1		ENFAGROW ESNTL 3 VANILA BOX800G	1	2024-12-30	2025-01-04	2024-12-23	7.407407	7.407407	5
2		ENFAGROW A+ 4 VANILLA 3KG (DL)	1	2024-12-05	2024-12-05	NaN	27.777778	27.777778	0

Gambar 3.20 Tabel 'repurchased'

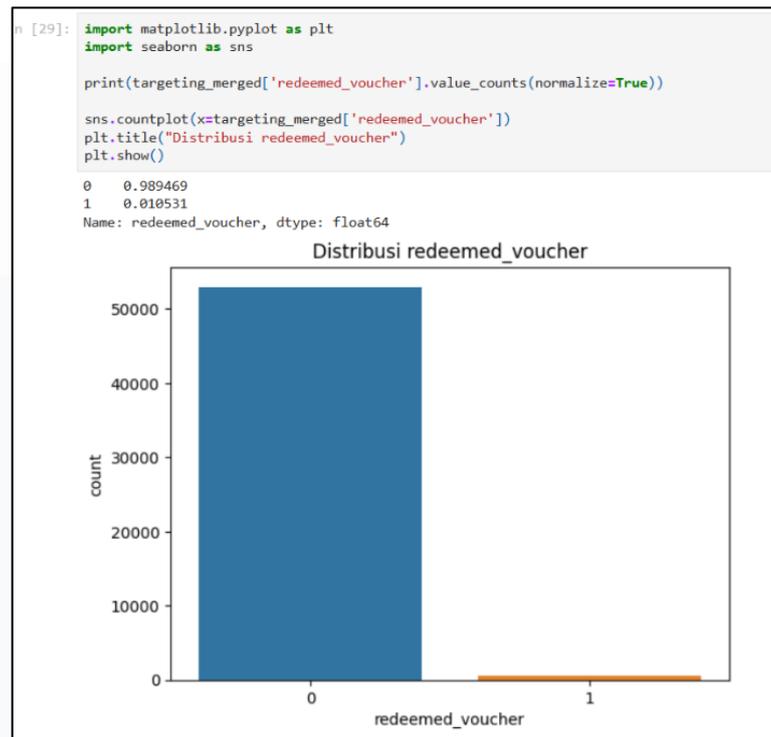
Gambar 3.20 memperlihatkan tabel 'repurchased' yang memuat data aktivitas *member* yang telah melakukan pembelian ulang terhadap produk susu pertumbuhan super premium. Tabel ini mencatat informasi penting seperti tanggal pembelian pertama dan tanggal pembelian berikutnya, serta menghitung jumlah hari antara kedua transaksi tersebut. Data ini digunakan untuk mengidentifikasi pola loyalitas dan kebiasaan *repeat order* dari *member* terhadap produk target, yang sangat berguna dalam analisis pola retensi pelanggan.

```
add_to_cart = pd.read_csv('Angelin/atc_tok.csv')
add_to_cart.head(3)
```

	EVENT_NAME	EVENT_UIID	member_id	plu	descp	cart_id	cart_source	date_event	struk_online	next_event_time	prev_event_time
0	add_to_cart				MORIGRO MADU BOX 200G (DL)		DETAIL_PRODUCT	2024-09-29 17:26:50	NaN	2024-09-29 17:27:10	NaN
1	add_to_cart				MORIGRO MADU BOX 200G (DL)		DETAIL_PRODUCT	2024-09-30 15:54:49	NaN	2024-09-30 15:54:52	2024-09-29 17:27:10
2	add_to_cart				MORIGRO MADU BOX 200G (DL)		DETAIL_PRODUCT	2024-09-30 15:54:52	NaN	2024-10-03 11:11:16	2024-09-30 15:54:49

Gambar 3.21 Tabel 'add_to_cart'

Gambar 3.21 menampilkan tabel 'add_to_cart' yang mencatat aktivitas pengguna yang menambahkan produk target ke dalam keranjang belanja. Meskipun *cart* belum tentu dilanjutkan ke proses pembelian, data ini tetap penting karena menunjukkan adanya minat atau ketertarikan terhadap produk tersebut. Fitur ini ditambahkan ke dalam data transaksi dan profil untuk memperkaya informasi mengenai interaksi awal pengguna dengan produk.



Gambar 3.22 Distribusi 'redeemed_voucher'

Setelah penambahan fitur-fitur yang relevan ke dalam data transaksi dan profil pengguna, tahap selanjutnya adalah melihat distribusi dari label target yang akan digunakan dalam proses pelatihan model (*data train*). Pada Gambar 3.22 ditampilkan distribusi dari variabel 'redeemed_voucher', yaitu label biner yang menunjukkan apakah seorang *member* telah menukarkan *voucher* (1) atau tidak (0).

Berdasarkan hasil visualisasi dan perhitungan proporsi yang ditampilkan, terlihat bahwa data sangat tidak seimbang (*imbalanced*). Mayoritas *member*, tidak menukarkan *voucher targeted* (kelas 0), sementara sedikit yang menukarkan *voucher targeted* (kelas 1). Ketidakseimbangan ini penting untuk diperhatikan dalam tahap pemodelan karena dapat memengaruhi kinerja model prediksi, terutama dalam mengklasifikasi kelas minoritas.

```

from sklearn.preprocessing import LabelEncoder, StandardScaler
encoder = LabelEncoder()
targeting_merged['Group_age'] = encoder.fit_transform(targeting_merged['Group_age'])
targeting_merged['_gender'] = encoder.fit_transform(targeting_merged['_gender'])

targeting_merged = targeting_merged[['_member',
                                     'jumlah_transaksi',
                                     'total_sales',
                                     'basket_size',
                                     'total_voucher_per_member',
                                     'purchased_gum',
                                     'Group_age',
                                     '_gender',
                                     'jumlah_retur',
                                     'qty_purchased',
                                     'buy_target_and_affinity',
                                     'buy_affinity',
                                     'add_to_cart',
                                     'repurchased',
                                     'redeemed_voucher'
                                     ]]
targeting_merged.set_index('_member', inplace=True)
targeting_merged

```

Gambar 3.23 Encode Data

Gambar 3.23 menunjukkan proses *encoding* variabel kategorikal, yaitu ‘Group_age’ dan ‘amc_gender’, ke dalam format numerik menggunakan LabelEncoder dari *library* sklearn.preprocessing. Hal ini penting karena sebagian besar algoritma *machine learning* hanya dapat bekerja dengan data numerik. Selanjutnya, dilakukan pemilihan fitur-fitur yang akan digunakan dalam pemodelan. Fitur-fitur tersebut mencakup informasi profil *member*, aktivitas transaksi, serta interaksi dengan produk dan *voucher* yang sudah disiapkan sebelumnya. Variabel target atau label dalam proses ini adalah *redeemed_voucher*. Terakhir, kolom *member* dijadikan sebagai indeks dari DataFrame agar memudahkan identifikasi setiap baris data berdasarkan ID member-nya.

Gambar 3.24 Data Train

Gambar 3.24 menampilkan data yang telah melalui proses pemilihan fitur (*feature selection*) dan *encoding*. Dari *dataset* awal yang

berisi berbagai informasi transaksi dan atribut pelanggan, hanya fitur-fitur yang dianggap relevan dengan proses klasifikasi 'redeemed_voucher' yang disertakan dalam data pelatihan ini. Beberapa fitur numerik seperti 'jumlah_transaksi', 'total_sales', 'basket_size', 'total_voucher_per_member', dan 'qty_purchased' digunakan tanpa transformasi kategorikal, sedangkan fitur kategorikal seperti 'purchased_gum', 'Group_age', dan gender telah dikodekan ke dalam format numerik (misalnya True/False menjadi 1/0 atau M/F menjadi 1/0), sehingga dapat digunakan dalam model *machine learning*.

```
from imblearn.over_sampling import SMOTE

X = targeting_merged.drop(columns=['redeemed_voucher'])
y = targeting_merged['redeemed_voucher']

smote = SMOTE(sampling_strategy=0.2, random_state=42)
X_resampled, y_resampled = smote.fit_resample(X, y)

print(y_resampled.value_counts(normalize=True))
```

0	0.833336
1	0.166664

Name: redeemed_voucher, dtype: float64

Gambar 3.25 Pembagian Data

Karena ditemukan ketidakseimbangan pada variabel target (redeemed_voucher), dilakukan teknik *oversampling* menggunakan SMOTE (*Synthetic Minority Oversampling Technique*) untuk menyeimbangkan distribusi kelas seperti yang ditampilkan pada gambar 3.25. Teknik SMOTE digunakan dengan parameter `sampling_strategy=0.2` yang berarti kelas minoritas akan ditingkatkan hingga mencapai 20% dari total data setelah *resampling*. Proses ini dilakukan sebelum data dibagi (*split*) ke dalam data latih dan data uji, dengan tujuan agar proses pembelajaran model memiliki data minoritas yang cukup untuk dilatih secara lebih representatif. Hasil *resampling* menghasilkan *dataset* baru `X_resampled` dan `y_resampled` yang telah lebih seimbang dan siap untuk dibagi dan digunakan dalam proses modeling.

```

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X_resampled, y_resampled, test_size=0.2, random_state=42, stratify=y_resampled
)

from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report

clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_train)

y_pred = clf.predict(X_test)

print(classification_report(y_test, y_pred))

```

	precision	recall	f1-score	support
0	0.98	0.98	0.98	10598
1	0.92	0.91	0.91	2120
accuracy			0.97	12718

Gambar 3.26 Data Split dan Modeling

Setelah data siap, pembagian data menjadi data latih dan data uji juga dilakukan yang kemudian diterapkan beberapa algoritma *machine learning* seperti *Decision Tree*, *Random Forest*, dan *Gradient Boosting*. Evaluasi model dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Dari beberapa model yang diuji, model *Random Forest* menunjukkan performa terbaik dalam mengklasifikasi *redeemed voucher* seperti yang terlihat pada gambar 2.36.

3.2.2.3 Pengolahan output model dan penggabungan *list member* untuk *voucher targeted* Maret 2025

Setelah model klasifikasi *redeemed voucher* selesai dibangun dan dievaluasi, langkah berikutnya adalah mengolah hasil klasifikasi tersebut agar dapat dimanfaatkan oleh tim yang akan menyebarkan *voucher* tersebut. *Output* dari model hanya berupa *ID member* yang diprediksi memiliki kemungkinan tinggi untuk menukarkan *voucher*. *ID member* hasil prediksi ini diambil dari model dengan mempertimbangkan *threshold* probabilitas tertentu, yang menunjukkan seberapa besar kemungkinan *member* akan melakukan *redeem voucher*. Dalam kasus ini, hanya *member* dengan probabilitas di atas *threshold* yang dipilih

untuk dimasukkan ke dalam daftar target. Proses tersebut dapat dilihat pada gambar 3.27 yang menunjukkan *code python*.

```
nyoba = X.copy()
nyoba['pred'] = clf.predict(X)
nyoba['real'] = y

nyoba
nyoba[nyoba['pred'] == 1]
len(tuple(nyoba[nyoba['pred'] == 1].index))
tuple(nyoba[nyoba['pred'] == 1].index)

import pandas as pd
index_pred = tuple(nyoba[nyoba['pred'] == 1].index)
df_member_id = pd.DataFrame(index_pred, columns=['Member_ID'])
df_member_id.to_excel("Angelin/Targeted Member Model/member_id_list.xlsx", index=False)
```

Gambar 3.27 Penggabungan *List Member ID*

Langkah selanjutnya adalah menggabungkan *list ID member* hasil model tersebut dengan *list ID member* yang dihasilkan dari *query* historis yang biasa digunakan oleh perusahaan. *Query* historis ini disusun berdasarkan data histori pembelian *member* terhadap produk ‘GROWING UP MILK SUPER PREMIUM’ tanpa menggunakan pendekatan *AI/machine learning*. Biasanya, *query* tersebut mengambil data *member* yang sebelumnya pernah membeli produk dalam kategori tersebut dalam periode tertentu, seperti 3 atau 6 bulan terakhir. Penggabungan dilakukan untuk menyatukan dua pendekatan berbeda yaitu, pendekatan berbasis histori dan pendekatan prediktif berbasis *machine learning* dengan tujuan memperluas dan mempertahankan atau *retain sales* kategori produk susu pertumbuhan *premium*. Proses penggabungan dilakukan dengan menggunakan Python, di mana kedua daftar *ID member* dicat.

Member id	
index	
20259	████████████████████
20260	████████████████████
20261	████████████████████
20262	████████████████████
20263	████████████████████
...	...
30899	████████████████████
30900	████████████████████
30901	████████████████████
30902	████████████████████
30903	████████████████████

Gambar 3.28 Output List Member

Hasil akhir dari penggabungan ini adalah daftar final ID *member* luar pulau Jawa yang menjadi target penerima *voucher* untuk produk ‘GROWING UP MILK SUPER PREMIUM’ pada periode Maret 2025 seperti yang dapat dilihat pada gambar 3.28. Daftar ini kemudian disiapkan dalam format *file* yang diminta, yaitu *csv*, dan diserahkan kepada *supervisor* yang pada akhirnya akan diberi kepada tim lain untuk proses distribusi *voucher* melalui aplikasi Alfagift.

3.2.2.4 Eksplorasi data, visualisasi, dan analisis loyalitas *member*

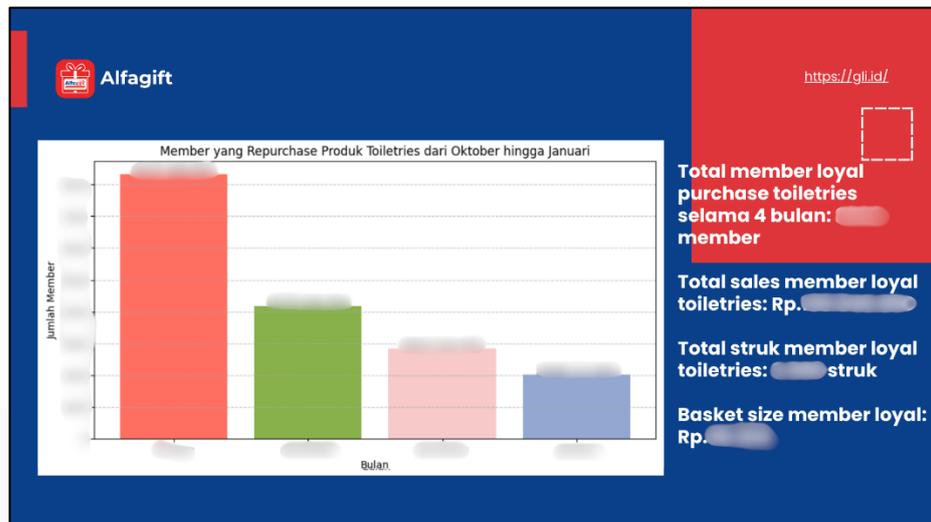
Setelah proses pemodelan untuk kategori ‘GROWING UP MILK SUPER PREMIUM’ selesai dan daftar target *member* sudah diberikan kepada *supervisor*, langkah berikutnya adalah melakukan eksplorasi data dan visualisasi untuk memahami perilaku *member* secara lebih mendalam, khususnya terhadap kategori produk yang menjadi fokus yaitu ‘GROWING UP MILK SUPER PREMIUM’ dan *toiletries*. Eksplorasi data dilakukan untuk melihat pola-pola yang mungkin ada dalam data historis pembelian *member targeted* susu pertumbuhan super premium terhadap kecenderungan loyalitas *member* kategori *toiletries*. Produk kategori *toiletries* yang dimaksud dan digunakan adalah produk

perlengkapan mandi bayi, perlengkapan mandi anak, dan popok.



Gambar 3.29 *Powerpoint* Eksplorasi Data

Gambar 3.29 menampilkan *slide* awal dari hasil eksplorasi data yang dilakukan menggunakan Python, kemudian dirangkum dalam bentuk presentasi PowerPoint. *Slide* ini menjelaskan konteks eksplorasi yang dilakukan. Penjelasan sumber data yang digunakan dan temuan umum dari eksplorasi awal, seperti jumlah data, distribusi member yang pernah menukar voucher, serta insight awal lainnya ditampilkan. Tampilan ini berguna untuk memberikan gambaran ringkas namun informatif kepada tim *data science* saat *weekly meeting*.



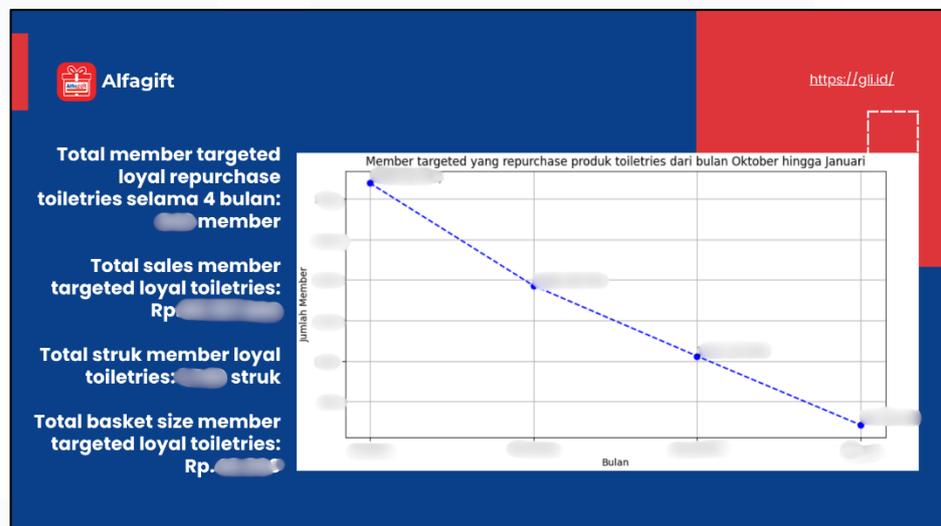
Gambar 3.30 Powerpoint Visualisasi Loyalitas Member

Gambar 3.30 menunjukkan visualisasi mengenai perilaku *repurchase* (pembelian ulang) produk *toiletries* oleh *member* Alfagift selama periode empat bulan, dari Oktober hingga Januari. Grafik batang di sisi kiri menampilkan tren jumlah *member* yang melakukan *repurchase* setiap bulannya. Sementara itu, di sisi kanan *slide* terdapat rangkuman metrik penting seperti total *member* loyal yang melakukan pembelian selama empat bulan, total *sales* yang dihasilkan, total jumlah struk transaksi, serta nilai rata-rata pembelian (*basket size*).



Gambar 3.31 Powerpoint Data Member Targeted

Gambar 3.31 menampilkan data hasil segmentasi *member* Alfagift yang ditargetkan untuk menerima *voucher* produk *toiletries* pada bulan Maret. Tampilan tabel memperlihatkan informasi penting seperti *member ID*, jumlah *struk*, *latency* (jarak waktu antar transaksi), *recent* (transaksi terakhir), serta informasi cabang dan *whitelist* produk. Selain itu, pada ditampilkan juga jumlah *member* yang masuk dalam daftar *targeted* serta total *member* yang tercatat pernah membeli produk *toiletries* selama empat bulan terakhir.



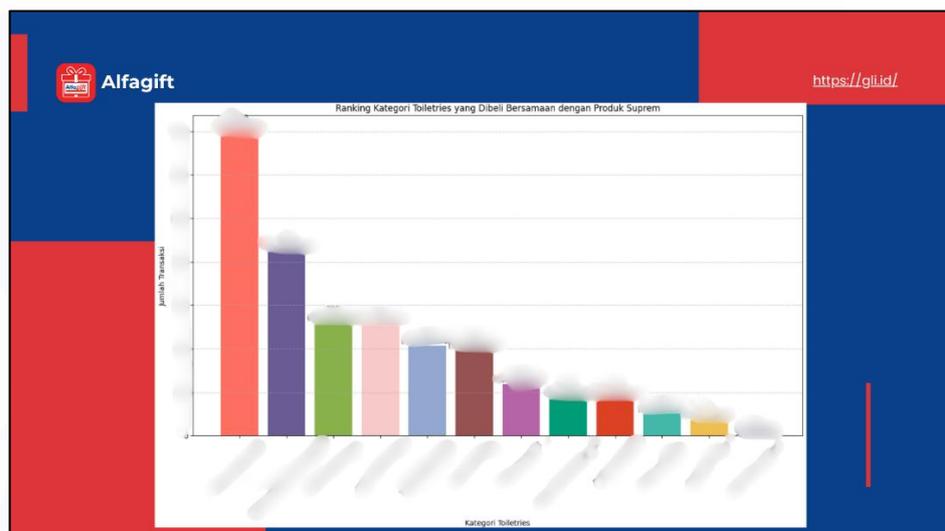
Gambar 3.32 *Powerpoint* Analisis Loyalitas *Member*

Gambar 3.32 menampilkan data analisis loyalitas *member* Alfagift yang secara konsisten melakukan pembelian ulang (*repurchase*) produk *toiletries* dalam periode empat bulan terakhir. Grafik di sisi kanan menunjukkan tren jumlah *member* yang melakukan *repurchase* setiap bulannya. Di sisi kiri, ditampilkan metrik utama sebagai hasil dari analisis tersebut, yaitu total jumlah *member* loyal, total nilai transaksi (*sales*), total jumlah *struk* transaksi, serta total *basket size* atau rata-rata nilai pembelian per transaksi dari *member* loyal tersebut.



Gambar 3.33 *Powerpoint* Analisis Terhadap Susu Pertumbuhan Super Premium

Gambar 3.33 menampilkan data dari segmentasi *member* Alfacift yang tidak hanya menunjukkan loyalitas dalam pembelian ulang produk *toiletries* selama empat bulan, tetapi juga membeli produk susu pertumbuhan super premium. Informasi yang disajikan mencakup jumlah *member* yang memenuhi dua kriteria tersebut, total penjualan dari kelompok *member* ini, total struk transaksi yang dihasilkan, serta nilai rata-rata pembelian per transaksi (*basket size*).



Gambar 3.34 *Powerpoint* Visualisasi Rank

Gambar 3.34 menunjukkan peringkat kategori produk *toiletries* yang paling sering dibeli bersamaan dengan produk ‘GROWING UP MILK SUPER PREMIUM’ oleh *member* Alfagift. Visualisasi berbentuk diagram batang ini menampilkan jumlah transaksi untuk setiap kategori *toiletries*. Kategori dengan jumlah transaksi tertinggi berada di peringkat pertama, diikuti oleh kategori lainnya secara menurun. Secara keseluruhan, hasil eksplorasi dan visualisasi ini memberikan gambaran yang lebih menyeluruh mengenai perilaku dan loyalitas *member* terhadap produk-produk utama yang dianalisis. Temuan-temuan dari analisis ini dapat digunakan untuk memperkuat strategi *marketing* berikutnya, tidak hanya dalam hal distribusi *voucher*, tetapi juga dalam penyusunan program loyalitas, penawaran *personalisasi*, serta kampanye promosi yang berbasis data.

3.2.2.5 Data preparation dan modeling untuk *voucher targeted April 2025*

Pada bulan sebelumnya, Februari 2025, proses *preparation* serta *modeling* telah dilakukan terhadap kategori ‘GROWING UP MILK SUPER PREMIUM’ untuk bulan Maret 2025. Namun pada tahap ini, selain untuk produk susu pertumbuhan, kategori *toiletries* juga ingin ditambahkan. Kategori *toiletries* yang termasuk untuk program *voucher targeted* luar pulau jawa ini adalah produk perlengkapan mandi bayi, anak, serta popok. Fokus utamanya masih sama yaitu, membangun model klasifikasi yang dapat mengklasifikasikan apakah seorang *member* akan menukarkan *voucher* untuk kategori yang ditarget atau tidak. Sama seperti yang dilakukan pada kategori *growing up milk super premium*, proses dimulai dengan persiapan data, yaitu tahap di mana data mentah dibersihkan dan disusun agar siap digunakan untuk pemodelan.

dilakukan karena sama seperti program *voucher targeted* ‘GROWING UP MILK SUPER PREMIUM’ bulan Maret 2025, program *voucher targeted* produk *toiletries* juga ditujukan untuk para *member* yang ada di pulau luar Jawa.

```
toiletries_luar_jawa = toiletries_luar_jawa.fillna(0)

toiletries_luar_jawa['purchased_gum'] = (toiletries_luar_jawa['qty_gum'] > 0).astype(int)
toiletries_luar_jawa['purchased_affinity'] = (toiletries_luar_jawa['qty_affinity'] > 0).astype(int)
toiletries_luar_jawa['purchased_toiletries'] = (toiletries_luar_jawa['qty_toileters'] > 0).astype(int)
toiletries_luar_jawa['atc_toiletries'] = (toiletries_luar_jawa['atc_target'] > 0).astype(int)
toiletries_luar_jawa['atc_affinity'] = (toiletries_luar_jawa['atc_affinity'] > 0).astype(int)
toiletries_luar_jawa['view_toiletries'] = (toiletries_luar_jawa['vp_target'] > 0).astype(int)
toiletries_luar_jawa['view_affinity'] = (toiletries_luar_jawa['vp_affinity'] > 0).astype(int)

toiletries_luar_jawa['tbtd_delivery_city'] = toiletries_luar_jawa['tbtd_delivery_city'].astype(str)
toiletries_luar_jawa['tbtd_delivery_kecamatan'] = toiletries_luar_jawa['tbtd_delivery_kecamatan'].astype(str)
toiletries_luar_jawa['tbtd_delivery_provinsi'] = toiletries_luar_jawa['tbtd_delivery_provinsi'].astype(str)
```

Gambar 3.37 Penanganan Nilai Hilang dan *Feature Engineer*

Gambar 3.37 menampilkan proses yang dilakukan setelah penggabungan data (*merge*) selesai dilakukan, proses yang dilakukan yaitu menangani nilai yang hilang (*missing values*), yang dalam kasus ini diisi dengan nilai 0. Selanjutnya, dilakukan proses *feature engineering* untuk menciptakan fitur-fitur baru yang dapat membantu model dalam memahami pola perilaku pengguna. Contohnya adalah fitur yang menunjukkan apakah seorang *member* melakukan pembelian produk *toiletries* dalam 3 bulan terakhir, atau total aktivitas (*event*) yang berkaitan dengan produk *toiletries*. Fitur-fitur tersebut dikonversi ke dalam format biner (0/1) untuk memudahkan algoritma dalam melakukan proses klasifikasi.

```

encoder = LabelEncoder()
toiletries_luar_jawa['group_age'] = encoder.fit_transform(toiletries_luar_jawa['group_age'])
toiletries_luar_jawa['_gender'] = encoder.fit_transform(toiletries_luar_jawa['_gender'])
toiletries_luar_jawa['tbtd_delivery_city'] = encoder.fit_transform(toiletries_luar_jawa['tbtd_delivery_city'])

toiletries_luar_jawa['tbtd_delivery_store'] = encoder.fit_transform(toiletries_luar_jawa['tbtd_delivery_store'])
toiletries_luar_jawa['tbtd_delivery_kecamatan'] = encoder.fit_transform(toiletries_luar_jawa['tbtd_delivery_kecamatan'])
toiletries_luar_jawa['tbtd_delivery_provinsi'] = encoder.fit_transform(toiletries_luar_jawa['tbtd_delivery_provinsi'])

model = toiletries_luar_jawa[['_member',
                              '_gender',
                              'group_age',
                              'struk',
                              'sales',
                              'basket',
                              'voucher_others_used',
                              'qty',
                              'purchased_gum',
                              'basket_gum',
                              'purchased_affinity',
                              'basket_affinity',
                              'purchased_toiletries',
                              'basket_toileters',
                              'atc_toiletries',
                              'atc_affinity',
                              'view_toiletries',
                              'view_affinity',
                              'tbtd_delivery_city',
                              'redeemed_targeted']]

model.set_index('_member', inplace=True)
model

```

Gambar 3.38 Data *Encoding* dan Pemilihan Fitur

Setelah proses pembersihan data selesai, dilakukan tahap *encoding* untuk mengubah nilai kategorikal menjadi numerik menggunakan metode *Label Encoding* seperti yang dapat dilihat pada gambar 3.38. Kolom-kolom seperti 'group_age', 'gender', serta lokasi pengiriman seperti 'tbtd_delivery_city', 'tbtd_delivery_store', 'tbtd_delivery_kecamatan', dan 'tbtd_delivery_provinsi' dikonversi agar dapat diproses oleh algoritma *machine learning*. Selanjutnya, dipilih sejumlah fitur yang dianggap relevan untuk proses pemodelan, yang mencakup informasi demografis, perilaku pembelian, penggunaan *voucher*, serta interaksi dengan produk seperti *toiletries*, *GUM (Growing Up Milk)*, dan produk-produk *affinity*. Fitur-fitur ini mencerminkan berbagai aktivitas konsumen, seperti pembelian, penambahan ke keranjang (*add to cart*), dan interaksi tampilan (*view product*). Terakhir, kolom identitas *member* dijadikan indeks untuk menjaga keunikan data per individu dan memastikan struktur data yang lebih rapi untuk proses analisis selanjutnya.

```

X = model.drop(columns=['redeemed_targeted'])
y = model['redeemed_targeted']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
X, y, test_size=0.2, random_state=42, stratify=y)

```

Gambar 3.39 Data *Splitting*

Setelah data siap digunakan, langkah selanjutnya adalah memisahkan fitur (X) dan target (y) seperti yang ditunjukkan pada gambar 3.39. Kolom 'redeemed_targeted' dipilih sebagai variabel target karena merepresentasikan apakah konsumen menukarkan *voucher* yang ditargetkan atau tidak. Sementara itu, seluruh kolom lain digunakan sebagai fitur prediktor. Setelah itu, data dibagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) menggunakan fungsi `train_test_split` dari *library* scikit-learn.

```

: from xgboost import XGBClassifier
  from sklearn.metrics import classification_report

xgb_model = XGBClassifier(
    n_estimators=200,
    learning_rate=0.1,
    max_depth=6,
    random_state=42,
    scale_pos_weight=len(y_train[y_train == 0]) / len(y_train[y_train == 1]),
    use_label_encoder=False,
    eval_metric='logloss'
)

xgb_model.fit(X_train, y_train)

y_pred = xgb_model.predict(X_test)

print(classification_report(y_test, y_pred))

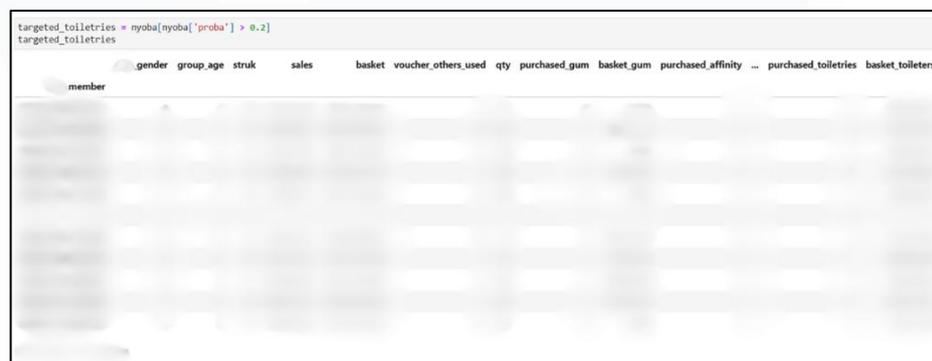
```

	precision	recall	f1-score	support
0	0.99	0.91	0.94	10691
1	0.16	0.57	0.25	338
accuracy			0.90	11029

Gambar 3.40 Pemodelan XGBoost

Gambar 3.40 menampilkan model XGBoost dilatih menggunakan data yang telah dibagi sebelumnya. Setelah pelatihan, model diuji dan dievaluasi menggunakan *classification report*. Hasil menunjukkan bahwa meskipun akurasi keseluruhan tinggi (96%), namun

Gambar 3.41 menunjukkan proses penggabungan hasil prediksi model klasifikasi XGBoost dengan data transaksi pelanggan untuk keperluan analisis lebih lanjut. Data asli disalin ke dalam variabel 'nyoba', lalu dilakukan *reset index* dan konversi kolom ID *member* ke dalam format *string* agar dapat diproses dengan baik. Model kemudian digunakan untuk menghasilkan dua jenis prediksi, label kelas 'pred' yang menunjukkan apakah seorang *member* diperkirakan akan menukarkan *voucher* atau tidak, dan probabilitas 'proba' yang merepresentasikan tingkat keyakinan model terhadap prediksi tersebut. Label aktual 'real' dari *dataset* juga ditambahkan untuk keperluan evaluasi model.



Gambar 3.42 Penentuan *Threshold* Untuk *Voucher Targeted* April 2025

Gambar 3.42 menunjukkan proses penyaringan data berdasarkan nilai probabilitas 'proba' hasil prediksi model klasifikasi. Pada tahap ini, diterapkan *threshold* sebesar 0.2, yang berarti hanya *member* dengan probabilitas lebih dari 0.2 yang akan dimasukkan ke dalam kelompok *targeted_toiletries*. Dengan kata lain, hanya *member* yang diprediksi memiliki kemungkinan minimal 20% untuk menukarkan *voucher toiletries* yang akan ditargetkan dalam kampanye promosi. *Threshold* ini digunakan sebagai batas kelayakan untuk memastikan bahwa hanya *member* dengan kecenderungan yang cukup tinggi terhadap penukaran *voucher* yang dipilih, sehingga distribusi *voucher* menjadi lebih tepat sasaran.

```

index_pred = tuple(targeted_toiletries.index)

df_member_id = pd.DataFrame(index_pred, columns=['Member_ID'])

len(targeted_toiletries.index)

df_member_id.to_excel("/.../toiletries_april.xlsx", index=False)

print("File member_id_list.xlsx berhasil disimpan!")
File member_id_list.xlsx berhasil disimpan!

```

Gambar 3.43 Penyimpanan Data Untuk *Voucher Targeted* April 2025

Gambar 3.43 menunjukkan proses penyimpanan hasil prediksi *member* yang ditargetkan untuk menerima *voucher toiletries*. *Member* dengan probabilitas di atas *threshold* 0.2 disaring, lalu ID-nya dikonversi ke dalam *DataFrame* dan disimpan ke *file* Excel bernama *toiletries_april.xlsx*. Langkah ini dilakukan agar daftar *member* tersebut dapat diberikan dengan mudah untuk keperluan tim lain selanjutnya.

Untuk proses pengolahan output list member ID kategori susu pertumbuhan super premium juga sama dilakukan seperti kategori *toiletries*. Setelah diperoleh daftar ID member dari kedua model, baik untuk kategori *toiletries* maupun susu pertumbuhan super premium, langkah selanjutnya adalah menggabungkan hasil prediksi ini dengan daftar ID member yang dihasilkan dari query historis pembelian. Proses penggabungan ini dilakukan menggunakan Python, di mana seluruh ID *member* dari hasil model *toiletries* dan susu pertumbuhan *super premium* serta hasil *query* historis masing-masing kategori akan digabungkan menjadi satu daftar. Pendekatan ini dilakukan agar strategi distribusi *voucher* mencakup mempertahankan pelanggan setia yang sudah pernah membeli produk (retensi), sekaligus mencoba menjangkau pelanggan baru yang berpotensi tertarik berdasarkan hasil prediksi dari model *machine learning*. Hasil akhir dari proses ini adalah sebuah *file* daftar ID *member* target dalam format *.csv*. *File* ini disiapkan untuk diserahkan kepada *supervisor*, yang kemudian akan meneruskannya ke tim lainnya

untuk proses distribusi *voucher* pada bulan April 2025 secara digital melalui aplikasi Alfagift.

3.2.2.7 Interpretasi model *targeted* produk *toiletries*

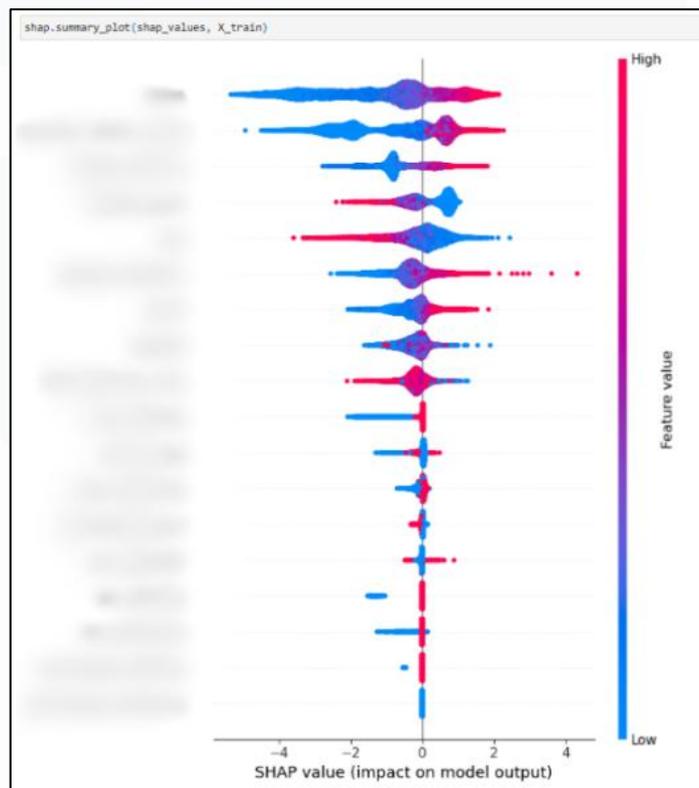
Tahap terakhir adalah interpretasi model, yaitu menjelaskan hasil yang diperoleh dari model dan memahami faktor-faktor apa saja yang paling berpengaruh dalam prediksi penukaran *voucher*. Misalnya, model dapat menunjukkan bahwa frekuensi kunjungan ke aplikasi, total belanja produk *toiletries*, atau kebiasaan menukarkan *voucher* sebelumnya merupakan variabel yang sangat penting. Interpretasi ini dilakukan agar cara kerja atau hasil dari model dapat dimengerti dan juga memberikan wawasan yang dapat digunakan selanjutnya.



Gambar 3.44 Interpretasi Model Menggunakan LIME

Gambar 3.44 menunjukkan hasil interpretasi model menggunakan LIME (*Local Interpretable Model-agnostic Explanations*), sebuah metode yang digunakan untuk menjelaskan klasifikasi model *machine learning* secara lokal pada satu data uji. Pada gambar ini, model menjelaskan alasan di balik prediksi untuk data ke-10, dengan memperlihatkan probabilitas prediksi yaitu 52% untuk kelas *Negative* (tidak menukarkan *voucher*) dan 48% untuk kelas *Positive* (menukarkan *voucher*). Visualisasi batang di bagian bawah menampilkan fitur-fitur yang paling berkontribusi terhadap hasil prediksi tersebut, di mana fitur yang berwarna oranye memberikan kontribusi terhadap prediksi *Positive*. Panjang batang menunjukkan seberapa besar pengaruh

fitur terhadap prediksi akhir. Kolom 'Feature' dan 'Value' menampilkan nama fitur dan nilai sebenarnya pada data tersebut, sehingga pengguna dapat memahami faktor-faktor spesifik apa yang memengaruhi keputusan model. Interpretasi ini penting agar hasil prediksi model dapat dimengerti secara intuitif dan dapat dipercaya dalam pengambilan keputusan bisnis.



Gambar 3.45 Interpretasi Model Menggunakan SHAP

Gambar 3.45 menampilkan interpretasi model menggunakan SHAP (*SHapley Additive exPlanations*), yang memberikan pemahaman mendalam terhadap kontribusi masing-masing fitur terhadap hasil prediksi. Grafik ini merupakan *summary* plot yang memperlihatkan fitur-fitur paling berpengaruh berdasarkan nilai SHAP. Setiap titik mewakili satu observasi (baris data), dengan posisi horizontal menunjukkan seberapa besar pengaruh fitur tersebut terhadap *output* model (semakin ke kanan berarti semakin mendorong ke prediksi positif, dan sebaliknya).

Warna titik menunjukkan nilai fitur, di mana merah menandakan nilai tinggi dan biru menunjukkan nilai rendah. Fitur yang berada di urutan atas adalah yang paling berkontribusi dalam keputusan model secara keseluruhan. Visualisasi ini berguna untuk mengetahui fitur apa yang paling penting dalam memengaruhi keputusan model dan bagaimana nilai suatu fitur tertentu dapat meningkatkan atau menurunkan kemungkinan prediksi, sehingga memudahkan pemahaman serta pengambilan keputusan yang lebih tepat.

3.2.2.8 Analisa behavior member targeted toiletries April 2025

Setelah daftar *member* yang menjadi target penerima *voucher* untuk kategori produk *toiletries* dan susu pertumbuhan *super premium* di bulan April 2025 ditentukan, langkah berikutnya adalah melakukan analisis perilaku atau *behavior* dari para *member* tersebut. Tujuan dari analisis ini adalah untuk lebih memahami karakteristik pelanggan yang terpilih oleh model, serta mengevaluasi apakah mereka memang merupakan segmen yang potensial untuk ditargetkan dalam promosi. Informasi ini penting untuk memastikan bahwa strategi pemasaran yang disusun sudah tepat sasaran dan dapat meningkatkan efektivitas distribusi *voucher*.



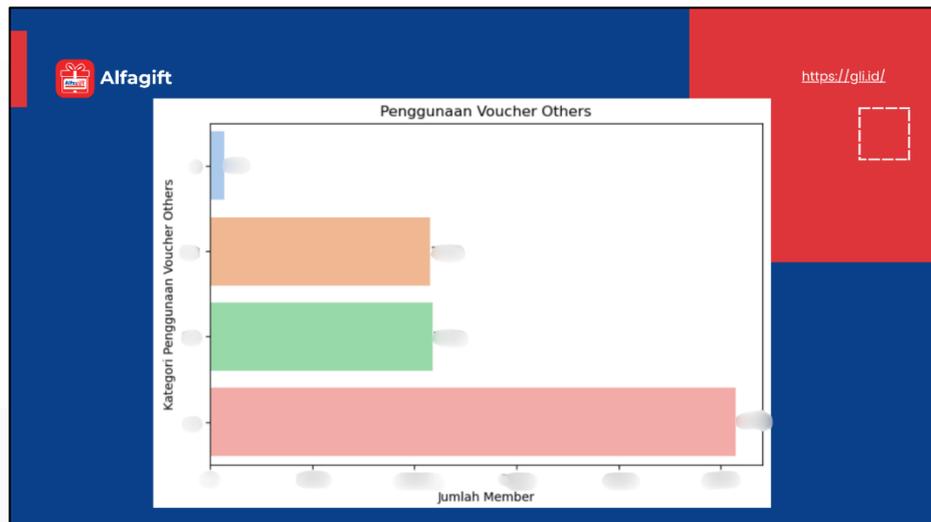
Gambar 3.46 Slide Update Penggunaan Voucher Targeted Bulan Maret 2025

Hasil dari analisis yang telah dilakukan juga dibuat dalam PowerPoint *slides* untuk digunakan sebagai *update* kepada seluruh tim data *science* pada saat *weekly meeting* dilakukan. Pada saat *weekly meeting* pun *update* terkait dengan performa *model* pada program *voucher targeted* dilakukan. Gambar 3.46 menampilkan *update* kategori susu pertumbuhan *super premium* bulan Maret 2025. Setelah *update* dilakukan, *slide-slide* hasil analisis *behavior member* pun ditampilkan.



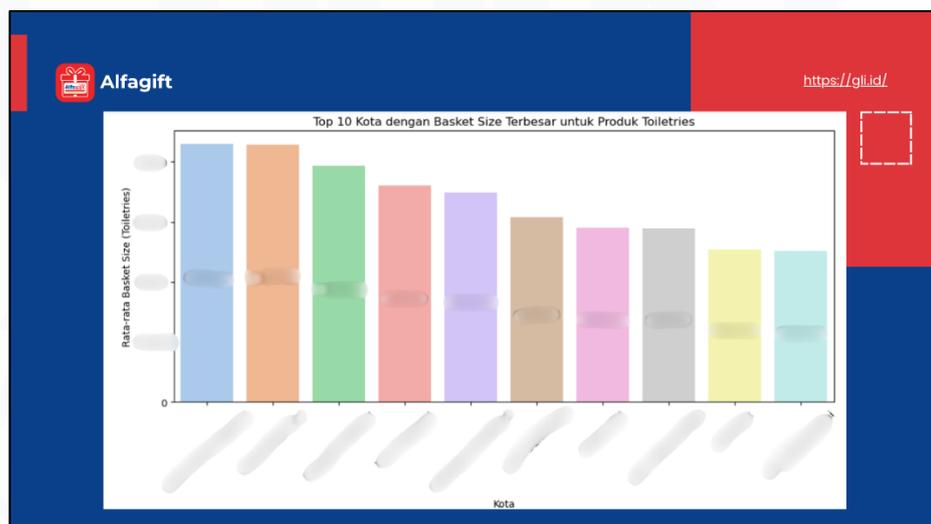
Gambar 3.47 Slide Analisa Konversi Aktivitas Hingga Pembelian

Gambar 3.47 menampilkan grafik batang yang menunjukkan jumlah *member* Alfagift pada setiap tahapan *funnel* aktivitas terhadap produk *toiletries*, yaitu melihat produk (*view product*), menambahkan ke keranjang (*add to cart*), dan melakukan pembelian (*buy*). Di samping grafik, terdapat penjelasan ringkas yang menyebutkan total *member* dan jumlah *member* pada masing-masing tahap. Visualisasi ini memberikan gambaran alur perilaku atau *behavior* pengguna terhadap produk *toiletries* dan dapat digunakan untuk menggali *insight* mengenai konversi pengguna dari tahap awal hingga akhir dalam proses pembelian.



Gambar 3.48 Slide Penggunaan *Voucher* Lain Produk *Toiletries*

Gambar 3.48 menampilkan diagram batang horizontal yang menunjukkan distribusi jumlah *member* Alfagift berdasarkan kategori penggunaan *voucher* lainnya atau selain *voucher targeted* untuk produk *toiletries*. Sumbu Y merepresentasikan kategori penggunaan *voucher*, sementara sumbu X menunjukkan jumlah *member*. Visualisasi ini memberikan gambaran tentang seberapa banyak *member* yang menggunakan jenis *voucher* lain, sehingga dapat dimanfaatkan untuk memahami preferensi penggunaan promo dan efektivitas kategori *voucher* dalam menarik pelanggan.



Gambar 3.49 Slide Rank Basket Size Produk Toiletries Terbesar Berdasarkan Kota

Gambar 3.49 menampilkan visualisasi perbandingan rata-rata *basket size* (ukuran keranjang belanja) untuk produk *toiletries* di sepuluh kota teratas di Indonesia. Sumbu vertikal menunjukkan nilai rata-rata *basket size*, sedangkan sumbu horizontal menunjukkan nama-nama kota. Grafik ini memberikan gambaran tentang kota mana saja yang memiliki kecenderungan belanja produk *toiletries* yang tinggi.



Gambar 3.50 Slide Member Targeted GUM Super premium Yang Ditemukan Pada Targeted Toiletries

Gambar 3.50 menampilkan diagram lingkaran yang menunjukkan proporsi *member targeted GUM super premium* yang juga menjadi target pada kategori produk *toiletries*. Diagram menunjukkan jumlah *member* yang hanya termasuk dalam target GUM super premium, serta jumlah *member* yang termasuk dalam keduanya. Visualisasi ini memberikan gambaran tentang potensi *overlap* antara kedua target dan dapat dimanfaatkan untuk menyusun strategi promosi lintas produk yang lebih efektif.



Gambar 3.51 Slide Insights Analisa Behavior Member Targeted Toiletries April 2025

Gambar 3.51 merupakan slide yang berisi *insights* terkait analisis perilaku *member targeted* untuk produk *toiletries* pada bulan April 2025. Visualisasi ini menunjukkan temuan bahwa sebagian besar *member* yang menukarkan (*redeem*) *voucher toiletries* juga memanfaatkan *voucher* lain, yang menunjukkan kecenderungan mereka untuk aktif dalam berbagai promosi. *Insight* ini juga dapat membuka peluang untuk strategi *cross-selling*, di mana *member* yang sudah aktif dalam satu jenis promo kemungkinan besar tertarik dengan promo lainnya. Selain itu, *insight* ini juga menyebutkan bahwa faktor lokasi geografis turut memengaruhi tingkat penggunaan *voucher targeted*.

Hasil dari analisis perilaku ini kemudian digunakan sebagai dasar untuk menyusun strategi penawaran yang lebih personal. Dengan memahami pola dan preferensi *member*, perusahaan dapat membuat program *voucher* yang lebih menarik dan relevan. Dengan pendekatan berbasis data seperti ini, diharapkan konversi penukaran *voucher* bisa meningkat, sekaligus memperkuat loyalitas *member* terhadap produk-produk *toiletries* yang ditawarkan oleh Alfagift.

3.2.2.9 Data *preparation* dan modeling untuk *voucher targeted* Mei 2025

Setelah data terbaru sudah tersedia dan dapat digunakan, proses pemodelan pun berlanjut. Sama seperti proses-proses yang telah dilakukan sebelumnya, dimulai dari data *preparation* (pembersihan data, *feature engineering*, agregasi data, pemilihan fitur), diikuti dengan penggunaan teknik *oversampling* untuk menangani ketidakseimbangan data (kategori susu pertumbuhan *super premium*), lalu diikuti dengan pembagian data menjadi data uji dan data latih, hingga pemodelan dan evaluasi. Pada bulan April proses pemodelan data mulai dilakukan dengan kategori susu pertumbuhan *super premium*.



```
1 trans = pd.read_parquet('...')
2 trans
```

no_member	tgl	struk	..._PLU	descp	kat	payment	sales	qty	retur	tipe_voucher	nominal_voucher	descp_kat
...
...
...
...
...
...
...
...
...

Gambar 3.52 Data Transaksi Member *Targeted* GUM *Suprem* Februari - April 2025

Gambar 3.52 menunjukkan data transaksi *member-member* yang sebelumnya pernah di target program *voucher targeted outer* Jawa kategori ‘GROWING UP MILK SUPER PREMIUM’ atau GUM *Suprem* pada bulan Februari hingga April 2025. Data ini berisi informasi pembelian secara rinci, termasuk tanggal transaksi, produk yang dibeli, kuantitas, serta informasi terkait penggunaan *voucher*. Seluruh data ini menjadi data awal untuk proses analisis, karena mencerminkan perilaku *member* yang telah terpapar program promosi. Langkah selanjutnya yang dilakukan dari data ini adalah melakukan proses *feature engineering* dan agregasi, yaitu mengubah data transaksi menjadi bentuk yang lebih

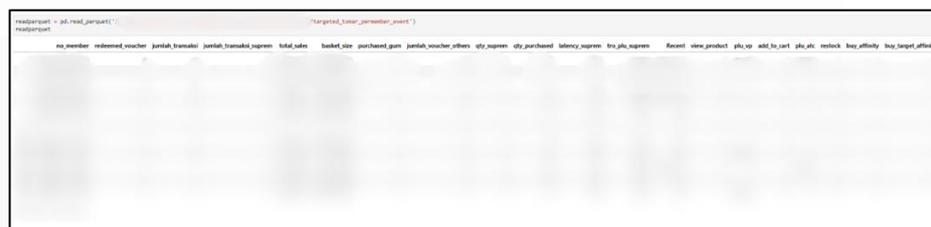
ringkas dan terstruktur, yaitu satu baris per *member* untuk memudahkan pemodelan prediktif yang ingin dibuat.



no_member	redeemed_voucher	jumlah_transaksi	jumlah_transaksi_suprem	total_sales	basket_size	purchased_gum	jumlah_voucher_others	qty_suprem	qty_purchased	latency_suprem	plu_suprem	Recent

Gambar 3.53 Data Per Member Targeted GUM Suprem Februari – April 2025

Gambar 3.53 menunjukkan data yang telah diproses menjadi format satu baris per *member*. Data ini merupakan hasil dari proses *feature engineering* terhadap data transaksi mentah sebelumnya, dengan tujuan menambahkan variabel-variabel yang dapat merepresentasikan perilaku dan aktivitas setiap *member* secara lebih baik. Fitur-fitur yang dihasilkan adalah ‘redeemed_voucher’, ‘jumlah_transaksi’, ‘jumlah_transaksi_suprem’, ‘total_sales’, ‘basket_size’, ‘purchased_gum’, ‘jumlah_voucher_others’, ‘qty_suprem’, ‘qty_purchased’, ‘latency_suprem’, ‘plu_suprem’, dan ‘Recent’.



no_member	redeemed_voucher	jumlah_transaksi	jumlah_transaksi_suprem	total_sales	basket_size	purchased_gum	jumlah_voucher_others	qty_suprem	qty_purchased	latency_suprem	plu_suprem	Recent	view_product	plu_up	add_to_cart	plu_at	redlock	buy_affinity	buy_target_affinity	

Gambar 3.54 Data Per Member Targeted Voucher GUM Suprem Februari – April 2025 Dengan Event

Gambar 3.54 menampilkan data hasil penggabungan antara data transaksi yang telah diolah menjadi satu baris per *member* dengan data aktivitas pengguna (*event*). Data *event* ini mencerminkan interaksi pengguna di aplikasi, seperti kunjungan halaman produk (*view product*), penambahan produk ke keranjang (*add to cart*), atau aktivitas lainnya yang menunjukkan minat terhadap produk tertentu. Karena sumber data

transaksi dan *event* berasal dari tabel yang berbeda, proses penarikan data dilakukan secara terpisah. Penggabungan keduanya baru dilakukan pada tahap akhir dalam proses persiapan data.

```
5 model = targeted[['no_member',
6                   'jumlah_transaksi',
7                   'jumlah_transaksi_suprem',
8                   'redeemed_voucher',
9                   'total_sales',
10                  'basket_size',
11                  'purchased_gum',
12                  'jumlah_voucher_others',
13                  'qty_suprem',
14                  'qty_purchased',
15                  'latency_suprem',
16                  'view_product',
17                  'add_to_cart',
18                  'restock',
19                  'buy_affinity',
20                  'buy_target_affinity'
21                  ]]
22 model.set_index('no_member', inplace=True)
23 model
```

Gambar 3.55 Pemilihan Data Train Untuk *Targeted Voucher GUM Suprem Mei 2025*

Gambar 3.55 menunjukkan proses pemilihan data *train* untuk program *voucher targeted GUM Suprem* selama periode Februari hingga April 2025. Sebelum memasuki tahapan *oversampling*, pembagian data, dan pembangunan model prediktif, terlebih dahulu dilakukan proses seleksi terhadap fitur-fitur yang dianggap relevan untuk digunakan dalam pelatihan model. Dengan kolom 'no_member' digunakan sebagai indeks utama untuk mengidentifikasi masing-masing *member*. Sebelum pemilihan fitur-fitur ini juga tidak lupa dilakukan penanganan nilai yang hilang.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

```

1 from imblearn.over_sampling import SMOTE
2 from sklearn.model_selection import train_test_split
3 from sklearn.ensemble import RandomForestClassifier
4 from sklearn.metrics import classification_report
5
6 X = model.drop(columns=['redeemed_voucher'])
7 y = model['redeemed_voucher']
8
9 smote = SMOTE(sampling_strategy=0.3, random_state=42)
10 X_resampled, y_resampled = smote.fit_resample(X, y)
11
12 print(y_resampled.value_counts(normalize=True))

```

redeemed_voucher	proportion
0	0.769248
1	0.230752

Name: proportion, dtype: float64

Gambar 3.56 Penerapan Teknik *Oversampling* Pada Data Train *Targeted Voucher* GUM *Suprem* Februari - April 2025

Gambar 3.56 memperlihatkan penerapan teknik *oversampling* yang dilakukan pada data *train*. Setelah proses pemilihan fitur selesai dilakukan, langkah selanjutnya adalah menangani ketidakseimbangan kelas dalam data, yaitu kondisi di mana jumlah data *member* yang menukarkan *voucher* jauh lebih sedikit dibandingkan yang tidak. Untuk mengatasi hal ini, digunakan teknik *oversampling* dengan metode SMOTE (*Synthetic Minority Over-sampling Technique*).

```

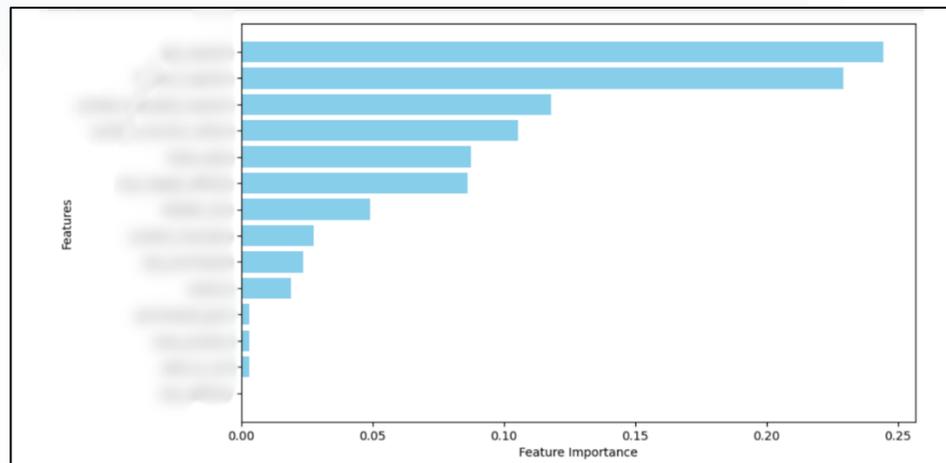
: 1 clf = RandomForestClassifier(
2     n_estimators=50,
3     max_depth=10,
4     random_state=42
5 )
6
7 clf.fit(X_train, y_train)
8
9 y_pred = clf.predict(X_test)
10
11 print(classification_report(y_test, y_pred))

```

	precision	recall	f1-score	support
0	1.00	0.96	0.98	9356
1	0.88	0.99	0.93	2806
accuracy			0.97	12162

Gambar 3.57 Data Modeling Untuk *Targeted Voucher* GUM *Suprem* Mei 2025

Gambar 3.57 menggambarkan proses data modeling. Setelah data *train* diolah dan diseimbangkan menggunakan teknik *oversampling*, langkah berikutnya adalah membagi data menjadi data latih dan data uji. Kemudian, dilakukan proses pemodelan menggunakan beberapa algoritma *machine learning*, seperti Logistic Regression, Random Forest, dan XGBoost. Masing-masing model dievaluasi menggunakan metrik *classification report* untuk mengukur kinerjanya dalam memprediksi penukaran *voucher*. Dari hasil evaluasi tersebut, model Random Forest dipilih karena memberikan performa terbaik dalam hal akurasi dan keseimbangan prediksi antar kelas.



Gambar 3.58 *Feature Importances* Untuk *Targeted Voucher GUM Suprem Mei 2025*

Gambar 3.58 menampilkan *feature importances* dari model Random Forest yang digunakan dalam pemodelan program *voucher targeted GUM Suprem* untuk periode Mei 2025. Visualisasi ini menunjukkan sejauh mana masing-masing fitur berkontribusi dalam proses prediksi model terhadap perilaku penukaran *voucher* oleh *member*. Semakin tinggi nilai *importance* suatu fitur, semakin besar pengaruhnya dalam membantu model membedakan antara *member* yang kemungkinan besar akan menukarkan *voucher* dan yang tidak.

Setelah proses pemodelan untuk kategori ‘GROWING UP MILK SUPER PREMIUM’ pada bulan Mei 2025 selesai dilakukan, tahapan

selanjutnya adalah memulai proses pemodelan untuk kategori *toiletries*. Proses ini mengikuti alur yang serupa, dimulai dari persiapan data, pembuatan fitur (*feature engineering*), hingga pemisahan data menjadi data latih dan data uji. Teknik *oversampling* kembali digunakan untuk menangani ketidakseimbangan kelas, dan berbagai algoritma *machine learning* diterapkan untuk membangun model prediktif.

```

Data columns (total 44 columns):
#  Column                               Non-Null Count  Dtype
---  -
0   punta_id                               non-null        object
1   _gender                                 non-null        object
2   umur                                    non-null        float64
3   struk                                   non-null        int64
4   sales                                  non-null        float64
5   qty                                    non-null        int64
6   basket                                 non-null        float64
7   sales_toileters                        non-null        float64
8   qty_toileters                          non-null        float64
9   basket_toileters                       non-null        float64
10  sales_gum                              non-null        float64
11  qty_gum                                non-null        float64
12  struk_gum                              non-null        int64
13  basket_gum                             non-null        float64
14  sales_affinity                         non-null        float64
15  qty_affinity                           non-null        float64
16  struk_affinity                         non-null        int64
17  basket_affinity                        non-null        float64
18  voucher_used                           non-null        object
19  vp_all                                  non-null        float64
20  atc_all                                 non-null        float64
21  atc_and_buy_all                        non-null        float64
22  vp_target                              non-null        float64
23  atc_target                             non-null        float64
24  atc_and_buy_target                    non-null        float64
25  vp_affinity                            non-null        float64
26  atc_affinity                           non-null        int64
27  atc_and_buy_affinity                  non-null        float64
28  tbtd_delivery_type                    non-null        float64
29  tbtd_delivery_store                   non-null        object
30  tbtd_delivery_city                    non-null        object
31  tbtd_delivery_kecamatan               non-null        object
32  tbtd_delivery_address                 non-null        object
33  tbtd_delivery_provinsi                non-null        object
34  tbtd_delivery_longitude                non-null        float64
35  tbtd_delivery_latitude                non-null        float64
36  redeemed_targeted                     non-null        int64
37  voucher_others_used                   non-null        int64
38  purchased_gum                          non-null        int64
39  purchased_affinity                    non-null        int64
40  purchased_toiletries                  non-null        int64
41  atc_toiletries                        non-null        int64
42  view_toiletries                       non-null        int64
43  view_affinity                         non-null        int64
dtypes: float64(23), int64(13), object(8)

```

Gambar 3.59 Data Untuk *Targeted Voucher Toiletries* Mei 2025

Gambar 3.59 menunjukkan data yang telah dipersiapkan untuk keperluan pemodelan *targeted voucher* kategori *toiletries* pada bulan Mei 2025. Data ini berasal dari berbagai tabel transaksi, aktivitas pengguna,

serta informasi demografis *member-member* Alfagift yang berada di luar pulau Jawa, yang kemudian digabungkan (*merge*) dan diolah menjadi format satu baris per *member*. Setiap baris mewakili satu ID *member* dengan total 44 kolom fitur yang mencakup informasi seperti jumlah transaksi, nilai belanja, jumlah produk yang dibeli, aktivitas pengguna (seperti *view product* dan *add to cart*), serta riwayat penggunaan *voucher*. Selain itu, data juga mencakup atribut lokasi pengiriman seperti kota, provinsi, dan koordinat geografis.

```

toiletries['gender'] = encoder.fit_transform(toiletries['gender'])
toiletries['tbtd_delivery_provinsi'] = encoder.fit_transform(toiletries['tbtd_delivery_provinsi'])

model = toiletries[['ponta_id',
                    'gender',
                    'umur',
                    'struk',
                    'sales',
                    'basket',
                    'voucher_others_used',
                    'qty',
                    'purchased_gum',
                    'basket_gum',
                    'basket_affinity',
                    'basket_toileters',
                    'atc_toiletries',
                    'sales_toileters', 'qty_toileters', 'struk_toileters',
                    'sales_gum', 'qty_gum', 'struk_gum', 'basket_gum',
                    'sales_affinity', 'qty_affinity', 'struk_affinity', 'basket_affinity',
                    'atc_affinity',
                    'view_toiletries',
                    'view_affinity',
                    'tbtd_delivery_provinsi',
                    'redeemed_targeted']]
model.set_index('ponta_id', inplace=True)
model

```

Gambar 3.60 Data *Encoding* Serta Pemilihan Fitur Untuk Model *Targeted Voucher Toiletries* Mei 2025

Gambar 3.60 menggambarkan tahap pengolahan data lanjutan untuk pemodelan *targeted voucher* produk *toiletries* bulan Mei 2025, yaitu proses *encoding* dan pemilihan fitur. Pada tahap ini, kolom kategorikal seperti 'gender' dan 'tbtd_delivery_provinsi' terlebih dahulu dikonversi ke bentuk numerik menggunakan teknik label *encoding* agar dapat diproses oleh algoritma *machine learning*. Setelah itu, dilakukan pemilihan fitur-fitur yang dianggap relevan dan berkontribusi dalam prediksi penukaran *voucher*. Fitur-fitur tersebut mencakup informasi demografis, aktivitas belanja (seperti jumlah transaksi, total penjualan, dan ukuran keranjang belanja), interaksi dengan kategori produk tertentu

(seperti GUM, *affinity*, dan *toiletries*), serta aktivitas pengguna terhadap *voucher* dan produk di aplikasi. Data akhir kemudian disusun dalam format indeks berdasarkan ID *member* ('*ponta_id*'), sehingga siap digunakan pada tahap pemodelan selanjutnya.

```

1 from imblearn.over_sampling import SMOTE
2 from sklearn.model_selection import train_test_split
3 from sklearn.ensemble import RandomForestClassifier
4 from sklearn.metrics import classification_report
5
6 X = model.drop(columns=['redeemed_targeted'])
7 y = model['redeemed_targeted']
8
9 smote = SMOTE(sampling_strategy=0.3, random_state=42)
10 X_resampled, y_resampled = smote.fit_resample(X, y)
11
12 print(y_resampled.value_counts(normalize=True))

```

redeemed_targeted	proportion
0	0.769231
1	0.230769

Name: proportion, dtype: float64

Gambar 3.61 Penerapan Teknik *Oversampling* Untuk Model *Targeted Voucher Toiletries Mei 2025*

Gambar 3.61 menunjukkan penerapan teknik *oversampling* pada data modeling. Sama seperti pada kategori *GUM Suprem*, data pada kategori ini juga menunjukkan ketidakseimbangan yang signifikan antara *member* yang menukarkan *voucher* (*redeemed*) dan yang tidak. Untuk mengatasi masalah ini, teknik *oversampling* pun juga digunakan, dalam kasus ini SMOTE agar distribusi kelas menjadi sedikit lebih seimbang.

```

1 X_train, X_test, y_train, y_test = train_test_split(
2     X_resampled, y_resampled, test_size=0.3, random_state=42, stratify=y_resampled
3 )

```

```

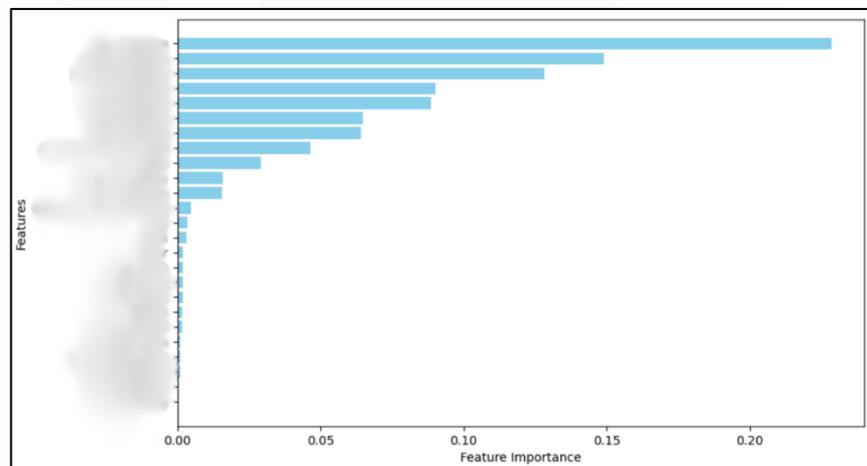
1 clf = RandomForestClassifier(
2     n_estimators=50,
3     max_depth=10,
4     random_state=42
5 )
6
7 clf.fit(X_train, y_train)
8
9 y_pred = clf.predict(X_test)
10
11 print(classification_report(y_test, y_pred))

```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	217893
1	0.96	0.97	0.96	65368
accuracy			0.98	283261

Gambar 3.62 Pembagian Data Uji Dan Latih Serta Pemodelan Untuk Model *Targeted Voucher Toiletries Mei 2025*

Gambar 3.62 memperlihatkan proses pembagian data menjadi data latih dan data uji serta *tahap* awal pemodelan. Tahapan ini dilakukan dengan metode yang sama seperti pada pemodelan sebelum-sebelumnya. Setelah data diseimbangkan melalui teknik *oversampling*, data kemudian dibagi menjadi dua bagian, data latih yang digunakan untuk melatih model *machine learning*, dan data uji yang digunakan untuk mengukur performa model terhadap data yang belum pernah dilihat sebelumnya. Pada tahapan ini juga mulai dilakukan pelatihan beberapa algoritma seperti Random Forest, Logistic Regression, atau XGBoost untuk mencari model dengan performa terbaik berdasarkan metrik evaluasi seperti *precision*, *recall*, dan F1-score. Pendekatan yang konsisten ini bertujuan agar hasil pemodelan antar program dapat dibandingkan secara adil dan memperoleh hasil yang optimal untuk masing-masing kategori produk.



Gambar 3.63 *Feature Importances* Untuk Model *Targeted Voucher Toiletries* Mei 2025

Gambar 3.63 menampilkan hasil *feature importance* dari model *machine learning* yang digunakan untuk memprediksi keberhasilan penukaran *voucher* pada program *targeted voucher* kategori *toiletries* bulan Mei 2025. Visualisasi ini menunjukkan sejauh mana masing-masing fitur atau variabel dalam data berkontribusi terhadap prediksi yang dihasilkan oleh model. Dengan menggunakan algoritma seperti

Random Forest yang dapat mengukur pentingnya setiap fitur, dapat diketahui fitur-fitur mana yang paling mempengaruhi keputusan model.

3.2.2.10 Pengolahan *output* model dan penggabungan *list member* untuk *voucher targeted* Mei 2025

Setelah proses pemodelan selesai dan diperoleh hasil prediksi dari model *machine learning*, langkah selanjutnya adalah mengolah *output* model tersebut menjadi daftar *member* yang akan ditargetkan, sesuai dengan permintaan *supervisor*. Pengelolaan ini mencakup transformasi hasil prediksi menjadi format yang dapat digunakan secara operasional, seperti *file* daftar ID *member*, serta penyaringan berdasarkan *threshold* atau prediksi probabilitas tertentu. Proses ini tentunya dilakukan untuk kedua kategori produk yang menjadi fokus dalam program *targeted voucher* bulan Mei 2025.

```
target = X.copy()
target['pred'] = clf.predict(X)
target['real'] = y
target['proba'] = clf.predict_proba(X)[:, 1]
```

target	jumlah_transaksi	total_sales	basket_size	purchased_gum	jumlah_voucher_others	qty_suprem	qty_purchased	latency_suprem	view_product	add_to_cart	restock	buy_affinity	buy_target_affinity	pred	real	proba
no_member																0.252946
																0.335747
																0.000000
																0.011461
																0.000000
																...
																0.000000
																0.000000
																0.084534
																0.000000
																0.000000

Gambar 3.64 Pengolahan *Output Model Targeted Voucher GUM Suprem* Mei 2025

Gambar 3.64 menampilkan kode yang digunakan untuk menghasilkan *output* klasifikasi. Data X yang berisi fitur-fitur model *dicopy* terlebih dahulu untuk memastikan tidak mengubah data asli, lalu ditambahkan tiga kolom baru yaitu, 'pred' yang merupakan hasil prediksi dari model (`clf.predict(X)`), 'real' yang menunjukkan label aktual (y), serta 'proba' yang berisi probabilitas bahwa masing-masing data termasuk dalam kelas 1 atau positif. Bagian `[:, 1]` digunakan untuk mengambil probabilitas dari kelas target positif (*redeem voucher*), karena fungsi `predict_proba` mengembalikan probabilitas untuk masing-masing kelas. Untuk kategori *GUM Suprem*, seleksi *member* dilakukan

Gambar 3.66 menunjukkan proses pengolahan *output* model untuk kategori *targeted voucher toiletries* pada bulan Mei 2025. Sama seperti pada kategori susu pertumbuhan *super premium*, hasil klasifikasi *model* pada kategori *toiletries* ini juga difilter berdasarkan nilai probabilitas. Angka ambang batas (*threshold*) yang digunakan adalah 0,29, sehingga hanya *member* dengan probabilitas lebih dari 0,29 yang akan ditargetkan.

targeted		
Member ID	Struk Alfagift	Branch
1	1	101
2	2	102
3	3	103
4	4	104
5	5	105
6	6	106
7	7	107
8	8	108
9	9	109
10	10	110
11	11	111
12	12	112
13	13	113
14	14	114
15	15	115
16	16	116
17	17	117
18	18	118
19	19	119
20	20	120

Gambar 3.67 List Member Targeted Voucher Toiletries Mei 2025

Gambar 3.67 menampilkan daftar akhir *member* yang telah terpilih untuk menerima *voucher* kategori *toiletries* berdasarkan hasil prediksi model. Data ini telah disesuaikan dengan *request supervisor*, dengan menambahkan kolom seperti jumlah struk pembelian melalui Alfagift serta kode cabang toko ('branch') terkait. Informasi ini nantinya akan digunakan oleh tim terkait dalam tahap pengiriman *voucher* kepada *member-member targeted*.

3.2.2.11 Analisis transaksi member targeted dan breakdown kategori toiletries

Setelah proses pengolahan *output* model *targeted* untuk bulan Mei 2025 selesai dilakukan, dilakukan analisis sesuai permintaan

supervisor. Analisis dilakukan terhadap data transaksi *member* yang telah menjadi *target* kategori *toiletries* selama periode Januari hingga April 2025. Analisis ini bertujuan untuk memahami perilaku belanja konsumen pada departemen ‘BABY & KID NEEDS’, khususnya pada sub kategori ‘BABY DIAPERS’, ‘BABY TOILETRIES’, dan ‘KIDS TOILETRIES’. Data yang digunakan disajikan dalam bentuk pivot *table* seperti pada gambar 3.68. Pivot *table* tersebut memuat informasi mengenai total penjualan, jumlah *gross margin*, serta jumlah transaksi pada masing-masing sub kategori.

Row Labels	Sum of sales	Sum of gm	Jumlah Transaksi
BABY DIAPERS			
PANTS DIAPERS			
REGULAR DIAPERS			
BABY TOILETRIES			
BABY BAR SOAP			
BABY COLOGNE			
BABY LIQUID DET & SOFTENER			
BABY LIQUID SOAP			
BABY LOTION, OIL & CREAM			
BABY SHAMPOO			
BABY TALCUM POWDER			
KIDS TOILETRIES			
KIDS COLOGNE			
KIDS HAIR VITAMIN			
KIDS LIQUID SOAP			
KIDS SHAMPOO			
TOOTH BRUSH KIDS			
TOOTH PASTE KIDS			

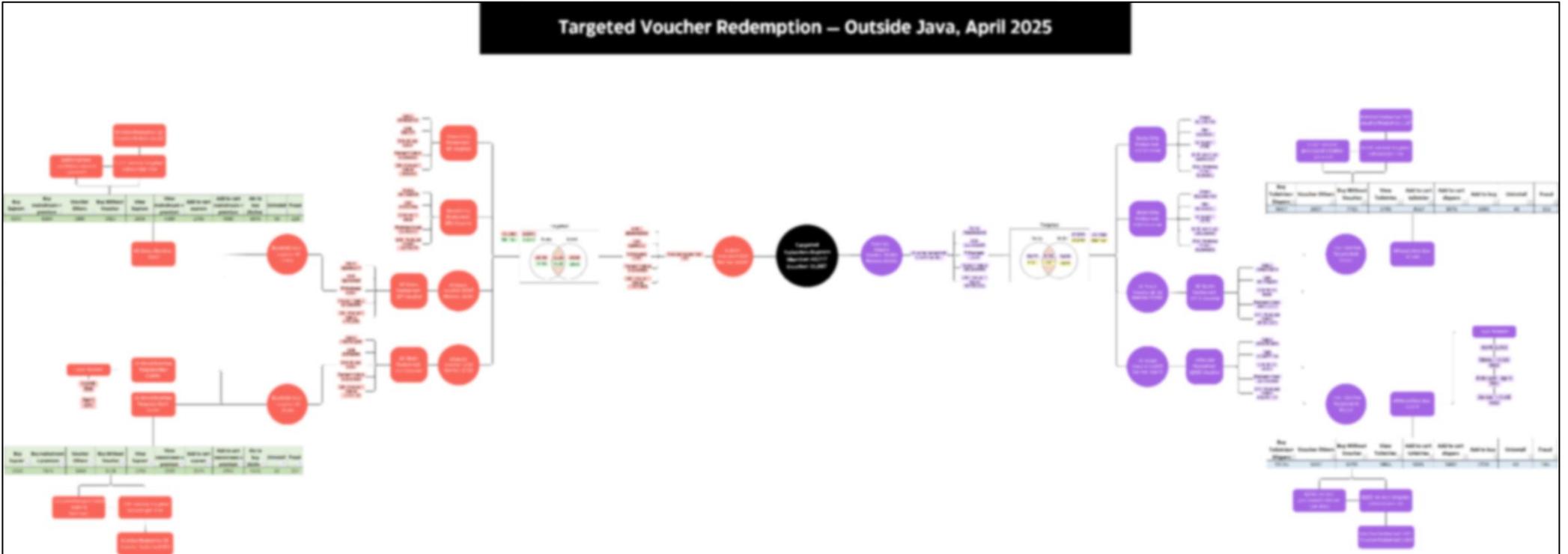
Gambar 3.68 Breakdown Transaksi *Toiletries* Januari-April 2025

3.2.2.12 Monitoring *voucher redeemed* bulan April 2025

Pada tahap ini dilakukan *monitoring* terhadap data *voucher redeemed* April 2025. Hal ini dilakukan untuk mengevaluasi efektivitas dari program *voucher* yang telah dijalankan. Proses *monitoring* dilakukan dengan visualisasi dalam bentuk *chart*, yang menampilkan mulai dari distribusi *voucher*, *redemption*, hingga aktivitas *member* yang tertarget. *Chart* dapat dilihat pada gambar 3.69.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

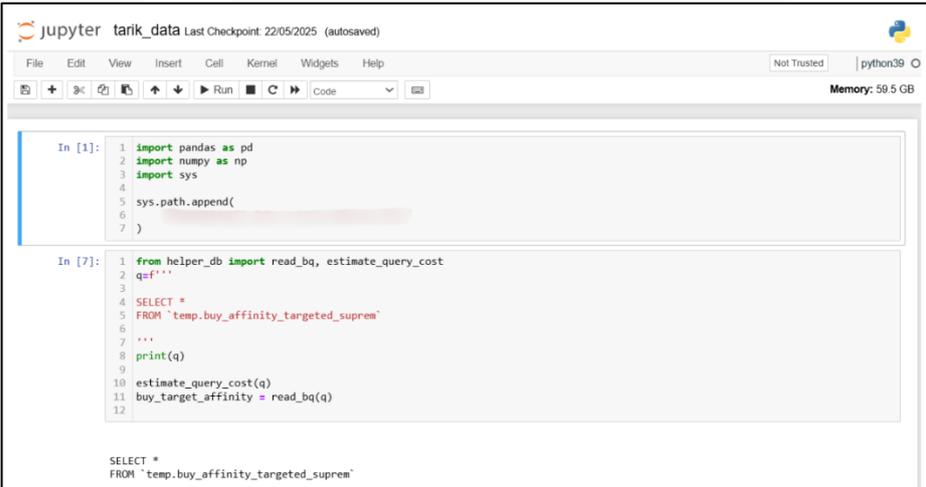
Targeted Voucher Redemption – Outside Java, April 2025



Gambar 3.69 Monitoring Chart Voucher Redeemed Bulan April 2025

3.2.2.13 Data preparation, modeling, dan pengolahan *output* untuk *voucher targeted Juni 2025*

Pada bulan Juni 2025, program pemberian *voucher* secara bertarget untuk *member* di wilayah luar Pulau Jawa masih tetap dilanjutkan sebagai bagian dari strategi pemasaran yang berkelanjutan. Untuk mendukung keberhasilan program ini, pada bulan Mei dilakukan proses pemodelan ulang dengan menggunakan data transaksi dan perilaku *member* tambahan. Tujuan dari pemodelan ini adalah untuk mengidentifikasi kembali nomor-nomor *member* yang memenuhi kriteria tertentu sehingga dapat dijadikan sasaran utama dalam distribusi *voucher*.



```
Jupyter tarik_data Last Checkpoint: 22/05/2025 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Not Trusted python39 Memory: 59.5 GB

In [1]: 1 import pandas as pd
        2 import numpy as np
        3 import sys
        4
        5 sys.path.append(
        6
        7 )

In [7]: 1 from helper_db import read_bq, estimate_query_cost
        2 q="""
        3
        4 SELECT *
        5 FROM `temp.buy_affinity_targeted_suprem`
        6
        7 """
        8 print(q)
        9
        10 estimate_query_cost(q)
        11 buy_target_affinity = read_bq(q)
        12

SELECT *
FROM `temp.buy_affinity_targeted_suprem`
```

Gambar 3.70 Menarik Data Untuk *Targeted Voucher GUM Suprem Juni 2025*

Gambar 3.70 memperlihatkan tahapan proses penarikan data yang digunakan untuk modeling kategori produk ‘GROWING UP MILK SUPER PREMIUM’ untuk bulan Juni 2025. Data yang diambil mencakup data-data seperti histori pembelian, frekuensi transaksi, serta aktivitas *member* pada aplikasi Alfagift. Data-data yang diambil bersumber dari beberapa tabel yang ada dalam Google BigQuery *database* PT GLI.

```

Data columns (total 34 columns):
#   Column                               Dtype
---  -
0   no_member                             object
1   redeemed_voucher                       int64
2   jumlah_transaksi                       int64
3   jumlah_transaksi_suprem                 int64
4   total_sales                            float64
5   basket_size                            float64
6   qty_purchased                          int64
7   purchased_gum                          int64
8   qty_suprem                             int64
9   jumlah_voucher_others                  int64
10  latency_suprem                         float64
11  ^~o_plu_suprem                          object
12  recent                                  datetime64[ns]
13  recency_all                             int64
14  recency_gum_pm                          float64
15  frekuensi_hari_belanja                  int64
16  persen_trans_suprem                     float64
17  avg_qty_per_trans_suprem                float64
18  ratio_qty_suprem                        float64
19  jumlah_kategori_unik                    int64
20  voucher_targeted_non_suprem              int64
21  jumlah_voucher_all                       int64
22  median_hari_antar_trans                  float64
23  recency_suprem                          float64
24  view_product                            int64
25  plu_vp                                  int64
26  add_to_cart                             int64
27  plu_atc                                  int64
28  restock                                  int64
29  mean_perkiraan_habis_hari_multiplied    float64
30  median_selisih_waktu_next                float64
31  count_restock                           float64
32  buy_affinity                             int64
33  buy_target_affinity                      int64
dtypes: Float64(2), Int64(3), datetime64[ns](1), float64(10), int64(16), object(2)
memory usage: 7.6+ GB

```

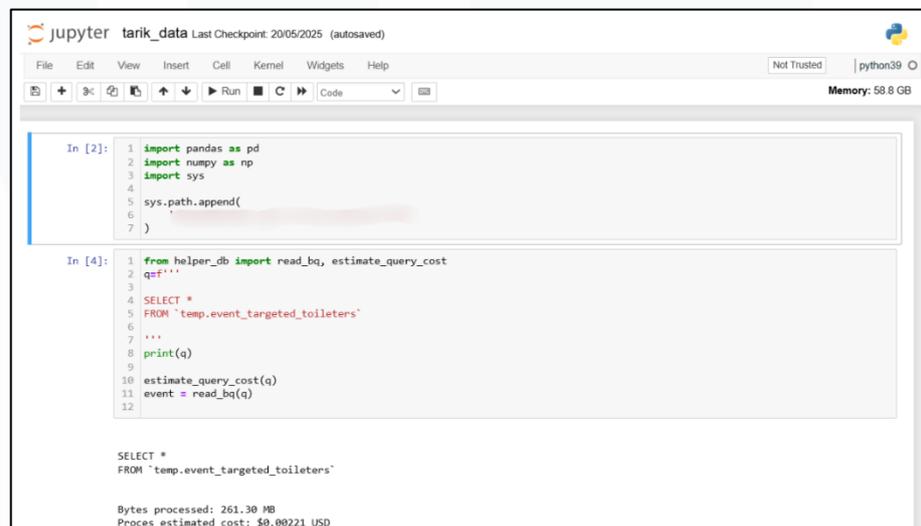
Gambar 3.71 Fitur-fitur yang digunakan untuk modeling *Targeted Voucher GUM Suprem Juni 2025*

Setelah penarikan data dilakukan, persiapan data, seperti pembersihan data, *feature engineering*, penggabungan data, dan agregasi juga dilakukan. Gambar 3.71 menampilkan fitur-fitur yang sudah siap untuk digunakan dalam proses pemodelan. *Dataset* ini terdiri dari 34 kolom yang merepresentasikan berbagai aspek perilaku *member*, mulai dari jumlah dan frekuensi transaksi, penggunaan *voucher* sebelumnya, jenis produk yang dibeli, hingga waktu pembelian terakhir.

Training Accuracy: 0.9972267210968873				
Testing Accuracy : 0.9967903918827975				
Classification Report (Test Data):				
	n	estimate		
		precision	recall	f1-score
	support			
0	44578	1.00	1.00	1.00
1	13373	0.99	0.99	0.99
accuracy			1.00	57951

Gambar 3.72 Evaluasi model *Targeted Voucher GUM Suprem Juni 2025*

Gambar 3.72 menunjukkan hasil evaluasi dari model *machine learning* yang digunakan untuk pemodelan *targeted voucher* GUM *Suprem* bulan Juni 2025. Model yang digunakan adalah *Random Forest*, dan hasilnya menunjukkan performa yang sangat baik dengan akurasi pelatihan sebesar 99,72% dan akurasi pengujian sebesar 99,68%. Selain itu, metrik *precision*, *recall*, dan *f1-score* pada data pengujian masing-masing mencapai nilai hampir sempurna, baik untuk kelas 0 maupun kelas 1.



```
jupyter tarik_data Last Checkpoint: 20/05/2025 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Not Trusted python39 Memory: 58.8 GB

In [2]:
1 import pandas as pd
2 import numpy as np
3 import sys
4
5 sys.path.append(
6
7 )

In [4]:
1 from helper_db import read_bq, estimate_query_cost
2 qf'''
3
4 SELECT *
5 FROM `temp.event_targeted_toiletries`
6
7 '''
8 print(q)
9
10 estimate_query_cost(q)
11 event = read_bq(q)
12

SELECT *
FROM `temp.event_targeted_toiletries`

Bytes processed: 261.30 MB
Proces estimated cost: $0.00221 USD
```

Gambar 3.73 Menarik Data Untuk *Targeted Voucher Toiletries* Juni 2025

Gambar 3.73 memperlihatkan tahapan proses penarikan data yang digunakan untuk modeling kategori produk *Toiletries* untuk bulan Juni 2025. Proses penarikan data ini dilakukan dengan cara yang sama seperti pada kategori GUM *Suprem*, yaitu mencakup data histori pembelian, frekuensi transaksi, serta aktivitas *member* pada aplikasi Alfacift. Data-data tersebut diambil dari berbagai tabel yang tersedia dalam Google BigQuery *database* milik PT GLI, dan menjadi dasar utama dalam proses pemodelan *targeted voucher*.

```

1 model = toiletries[['_ponta_id',
2                   'sales', 'qty', 'basket',
3                   'sales_toileters', 'qty_toileters', 'basket_toileters',
4                   'sales_gum', 'qty_gum', 'basket_gum',
5                   'sales_affinity', 'qty_affinity', 'basket_affinity',
6                   'struk', 'struk_toileters', 'struk_gum', 'struk_affinity',
7                   |
8                   'redeemed_targeted', 'voucher_others_used',
9                   'vp_all', 'atc_all', 'atc_and_buy_all',
10                  'vp_target', 'atc_target', 'atc_and_buy_target',
11                  'vp_affinity', 'atc_affinity', 'atc_and_buy_affinity',
12                  'tbtd_delivery_city',
13                  |
14                  'purchased_gum', 'purchased_affinity', 'purchased_toiletries',
15                  'atc_toiletries', 'atc_affinity',
16                  'view_affinity', 'view_toiletries']]
17 model.set_index('_ponta_id', inplace=True)
18 model
19

```

Gambar 3.74 Fitur-fitur Modeling Untuk *Targeted Voucher Toiletries* Juni 2025

Gambar 3.74 menampilkan daftar fitur yang digunakan dalam proses pemodelan untuk *Targeted Voucher* kategori produk *Toiletries* pada bulan Juni 2025. Fitur-fitur ini dikembangkan dari hasil pengolahan data transaksi dan aktivitas *member*, mencakup berbagai indikator seperti volume pembelian, frekuensi kunjungan, pemanfaatan *voucher* sebelumnya, serta interaksi dengan berbagai kategori produk lainnya. Secara keseluruhan, terdapat 34 variabel yang merepresentasikan pola perilaku konsumen yang berpotensi membantu dalam mengidentifikasi *member* yang tepat untuk diberikan *voucher*.

Training Accuracy: 0.9837602388844753				
Testing Accuracy : 0.9833836182099436				
Classification Report (Test Data):				
	precision	recall	f1-score	support
0	0.99	0.98	0.99	240680
1	0.95	0.98	0.96	72204
accuracy			0.98	312884

Gambar 3.75 Hasil Evaluasi Model Untuk *Targeted Voucher Toiletries* Juni 2025

Gambar 3.75 memperlihatkan hasil evaluasi model yang digunakan untuk prediksi *targeted voucher* kategori *Toiletries* bulan Juni 2025. Dalam proses ini, algoritma *Random Forest* digunakan karena mampu menangani data dengan jumlah fitur yang banyak serta memiliki

waktu untuk mempelajari dan memahami makna dari istilah tersebut.

2. Struktur *database* perusahaan yang kompleks juga menjadi tantangan utama, dikarenakan volume data yang sangat besar serta sistem penyimpanan yang terdiri dari banyak tabel dengan relasi yang tidak selalu langsung terlihat. Kesulitan muncul dalam menentukan tabel mana yang harus diambil untuk memperoleh data yang dibutuhkan, terutama saat data spesifik diperlukan. Data yang dicari tersebar di berbagai tabel dan memerlukan proses *join* yang cukup rumit agar dapat digabungkan menjadi satu *dataset* yang utuh.
3. Ketika melakukan klasifikasi terhadap *dataset* yang memiliki distribusi kelas yang tidak seimbang (*imbalanced data*). Dalam kondisi seperti ini, model cenderung “berpihak” pada kelas mayoritas, yang membuat klasifikasi terhadap kelas minoritas menjadi kurang akurat. Masalah ini menjadi cukup signifikan karena hasil klasifikasi menjadi kurang representatif terhadap kondisi aktual.

3.4 Solusi atas Kendala yang Ditemukan

Bagian ini berisi solusi atas kendala yang ditemukan selama proses kerja magang

1. Untuk mengatasi ketidaktahuan terhadap struktur data dan konteks industri perusahaan, langkah awal yang dilakukan adalah aktif mencari referensi internal seperti dokumentasi teknis, atau presentasi yang tersedia di lingkungan kerja. Selain itu, diskusi langsung dengan *supervisor* atau rekan kerja juga sangat membantu dalam menjelaskan istilah-istilah teknis dan bisnis yang belum *familiar*. Dengan bertanya secara aktif serta mencatat setiap istilah penting yang muncul, pemahaman pun meningkat seiring waktu.
2. Dalam menghadapi struktur *database* yang kompleks, solusi yang diterapkan adalah dengan memetakan struktur *database* secara bertahap. Hal ini dilakukan dengan memahami diagram relasi antar

tabel berdasarkan informasi dari *metadata* dan eksplorasi skema *database* menggunakan *query* sederhana. Proses ini membantu dalam mengidentifikasi tabel-tabel utama dan relasinya. Selain itu, berdiskusi langsung dengan rekan kerja atau tim terkait juga memberikan kejelasan mengenai tabel mana yang sebaiknya digunakan dan cara optimal untuk melakukan *join* antar tabel.

3. Untuk menangani permasalahan data yang tidak seimbang (*imbalanced data*) dalam proses klasifikasi, diterapkan teknik *oversampling*, khususnya dengan metode SMOTE (*Synthetic Minority Over-sampling Technique*). Teknik ini membantu menyeimbangkan distribusi antara kelas mayoritas dan minoritas, sehingga model *machine learning* dapat belajar secara lebih baik terhadap kedua kelas. Dengan melakukan *oversampling* sebelum proses pelatihan model, performa klasifikasi terhadap kelas minoritas menjadi lebih akurat dan representatif terhadap kondisi sebenarnya.

