

BAB III

METODOLOGI PENELITIAN

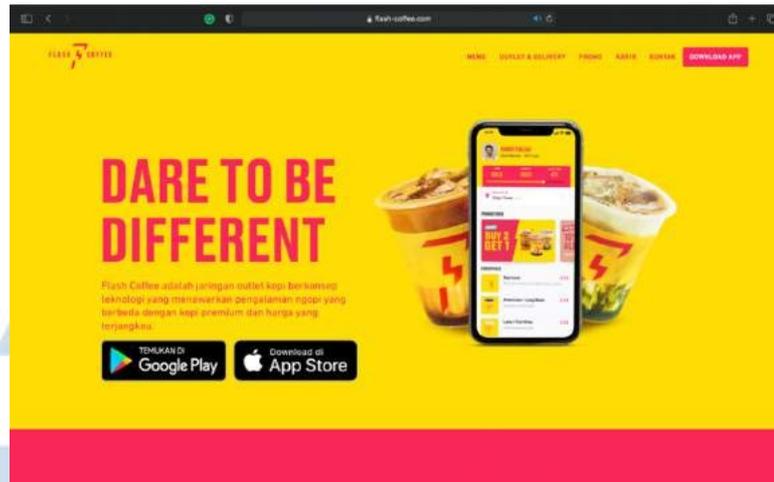
3.1 Gambaran Umum Objek Penelitian

Objek yang diteliti adalah Flash Coffee. Penelitian ini akan meneliti mengenai perilaku konsumen yang melakukan pemesanan melalui aplikasi Flash App yang berwilayah di Singapura. Data yang digunakan adalah data *customer* yang berisi total sales, last transaction, dan promo / campaign apa saja yang digunakan.

3.1.1 Profil Perusahaan

Flash Coffee adalah salah *tech-enabled coffee chains* dengan pertumbuhan tercepat di Asia. Flash Coffee bertujuan untuk mendemokratisasikan kopi berkualitas tinggi agar dapat diakses oleh berbagai negara di Asia dan menyajikan menu kopi premium dengan harga terjangkau. Pelanggan dapat menggunakan aplikasi Flash Coffee untuk memesan dan membayar secara *online*, memilih untuk mengambil pesanan mereka di salah satu etalase kuning merupakan *iconic* Flash Coffee. Flash Coffee beroperasi di Singapura, Indonesia, Thailand, Taiwan, dan Hongkong, serta akan diluncurkan di 6 area lain di seluruh Asia Pasifik pada tahun 2022. Flash Coffee adalah *start up* Seri A yang didukung oleh Rocket Internet, Delivery Hero, White Star Capital, Global Founders Capital dan Conny & Co.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 3. 1 Website Flash Coffee

3.1.2 Produk Perusahaan

Pada Flash Coffee terdapat beberapa kategori produk yaitu *The Brew Crew*, *Essentials*, *Signatures*, *Non-Coffee Signatures*, *Ice Tea*, dan *Snacks*. Produk perusahaan Flash Coffee dapat dilihat pada Tabel 3.1.

Tabel 3. 1 List Produk Flash Coffee

Kategori	Menu
The Brew Crew	Nutella Cold Brew, Melaka Cold Brew, Cold Brew White, Cold Brew Black
Essentials	Piccolo, Flat White, Vanilla Latte, Oatly Cappuccino, Espresso, Mocha, Caramel Latte, Cappuccino, Latte, Americano
Signatures	Macadamia Latte, Milo Espresso Shake, Espresso Tonic, Espresso Soda, Nutella Latte, Melaka Latte, Lychee Espresso Soda, Avo Latte
Non-Coffee Signatures	Milo Dinosaur Shake, Choco Banana Protein Shake, Strawberry Coconut Protein Shake,

	P.B.B Protein Shake, Hot Chocolate, Golden Turmeric Latte, Nutella Shake, Nutella Hot Chocolate, Matcha-Damia Shake, Matcha Latte, Lotus Biscuit Shake, Lemon Espresso Soda)
Ice Tea	Ice Tea, Fig & Cinnamon Ice Tea, Milk Tea, Lychee Ice Tea, Lemon Ice Tea
Snacks	Dalgona Biscuit, Lemon Strawberry Tart, Chocolate Tart, Nutella S'mores Waffles, Lotus Biscuit Waffles, Honey Almond Waffles, Ham & Cheese Croissant, Flash x Saladstop! Smoked Salmon with Quinoa Wrap, Waffle, Flash x Saladstop! Baked Salmon with Pesto Salad, Flash x P.Osh Brownie, Pain au Chocolat, Choco Banana Oats, Butter Croissant, Blueberry Almond Oats

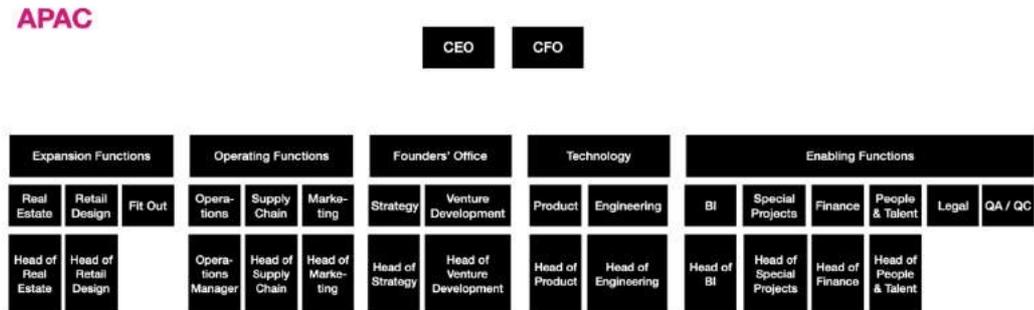
3.1.3 Visi dan Misi Perusahaan

Visi Flash Coffee adalah mendemokratisasikan kopi berkualitas tinggi dan membuatnya dapat diakses oleh kelas menengah Asia yang sedang naik daun. Misi Flash Coffee adalah untuk membangun merek baru di seluruh wilayah, proses operasional kelas atas dan rantai nilai yang terintegrasi penuh [20]. Para pendiri juga berencana untuk mendirikan kantor pusat regionalnya di Singapura, di mana mereka akan mempekerjakan lebih dari 90 anggota staf baru pada akhir tahun 2021.

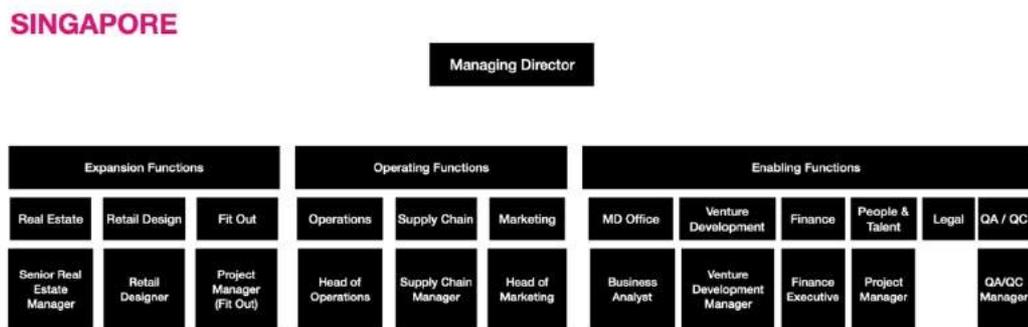
3.1.4 Struktur Organisasi Perusahaan

Flash Coffee merupakan perusahaan yang tersebar di beberapa negara di benua Asia. Setiap negeri memiliki struktur organisasi masing-masing, namun terdapat struktur inti perusahaan

yang bernama “APAC”. Dimana pada bagian “APAC” terdiri dari beberapa departemen yang terdapat pada gambar dibawah ini:



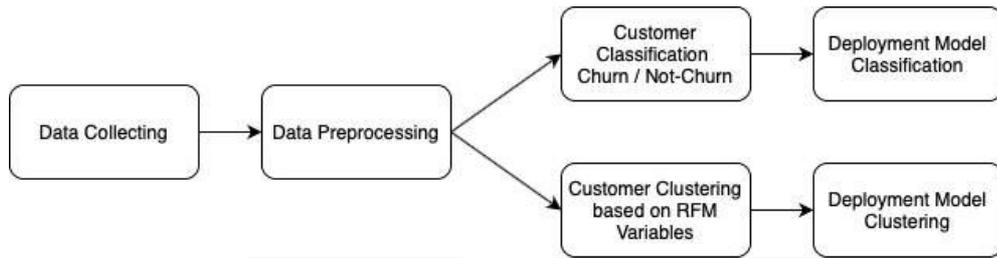
Gambar 3. 2 Struktur Organisasi Flash Coffee APAC



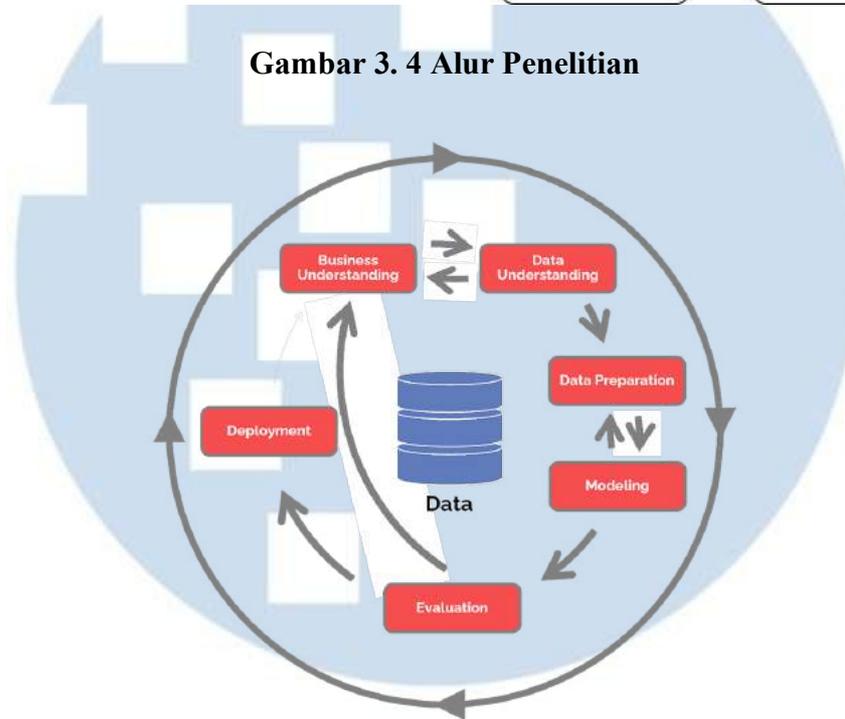
Gambar 3. 3 Struktur Organisasi Flash Coffee Singapore

3.2 Metode Penelitian

Metode yang digunakan pada penelitian ini adalah metode kuantitatif dengan menghitung tingkat akurasi pada model klasifikasi. Beberapa proses dalam penelitian ini terdapat pada Gambar 3.5 dengan menggunakan *framework* CRISP-DM. Selain itu terdapat alur penelitian pada Gambar 3.4 yang telah disesuaikan pada penelitian ini dikarenakan terdapat 2 *study* yang terkandung dalam penelitian ini yaitu *customer classification* dan *customer clustering* pada *customer* Flash Coffee. Pada bagian *customer clustering* bukan kelanjutan dari *customer classification* yang telah dibuat.



Gambar 3. 4 Alur Penelitian



Gambar 3. 5 CRISP-DM Process

3.2.1 Business Understanding

Pada tahap *business understanding*, memiliki tujuan bisnis dan tujuan dari data mining yang akan dilakukan. Pada penelitian ini tujuan bisnis yang dimaksud adalah untuk mengurangi adanya churn pada pelanggan Flash Coffee, serta tujuan data mining pada penelitian ini adalah untuk mengklasifikasikan adanya *customer* yang sudah *churn* dan membagi *customer* ke beberapa *cluster* berdasarkan nilai *recency*, *frequency* dan *monetary* melalui proses *data mining*.

3.2.2 Data Understanding

Data yang akan digunakan pada penelitian ini adalah data yang berasal database Flash Coffee terkait Flash App yang berisi *data order customer* melalui aplikasi di wilayah Singapura. Dalam data tersebut terdapat beberapa data seperti *id order*, *transaction time*, *pick up time*, *payment status*, *store id*, *customer ref id*, *total pay*, *product name* serta *last update customer* dalam menggunakan aplikasi. Periode pengumpulan data yaitu Februari 2021 hingga Februari 2022.

3.2.3 Data Preparation

Data yang telah dikumpulkan dalam bentuk dataset akan dilakukan proses *preprocessing* yaitu proses dimana data dilakukan pengelolaan / perubahan dengan tujuan membersihkan *missing value*, menghapus *noise* dan data yang tidak konsisten. *Noise Removal* sangat penting dalam data preprocessing karena dalam dataset memiliki kemungkinan adanya data “null” dan ketidakseimbangan atribut dalam *dataset* [17]. Dalam data preprocessing juga berkaitan dengan proses *feature selection*, pemilihan fitur adalah langkah dimana terjadinya proses pengurangan jumlah *variable* input ketika mengembangkan model prediktif. *Feature selection berfungsi* untuk mengurangi biaya komputasi pemodelan dan dapat meningkatkan kinerja model. *Feature selection* menggunakan metode pemilihan fitur berbasis *statistic* yang melibatkan evaluasi korelasi antara setiap *variable* input dengan *variable* target dan memilih *variable* input yang memiliki tingkat korelasi terkuat. Kinerja klasifikasi dapat meningkat jika dataset berisi variabel yang sangat prediktif. Oleh karena itu, fokus pada pemilihan fitur penting dan mengurangi jumlah atribut yang tidak relevan dapat meningkatkan kinerja klasifikasi.

3.2.4 Modelling

3.2.4.1 Customer Classification and Prediction

Dalam penelitian ini terdapat dua jenis pelanggan yaitu pelanggan yang tetap setia kepada perusahaan atau *non-churn* dan pelanggan yang tidak setia atau *churn*. Model *machine learning* yang dibuat akan mengklasifikasikan data pelanggan menggunakan kumpulan data berlabel. Algoritma klasifikasi yang akan digunakan dalam penelitian ini adalah algoritma yang lebih unggul diantara *random forest* dan *logistic regression*. Kedua algoritma tersebut berdasarkan jurnal [17] dan [18] memiliki tingkat akurasi yang tinggi dalam melakukan klasifikasi adanya *churn* pada pelanggan. Setelah mengklasifikasikan pelanggan berdasarkan model tersebut, maka langkah selanjutnya adalah mengidentifikasi faktor-faktor yang menyebabkan *churn*.

3.2.4.2 Customer Profiling through Clustering

Cluster pelanggan dibuat untuk membagi beberapa data pelanggan ke dalam beberapa partisi berdasarkan hubungan antara variable 1 dan variable lainnya. Clustering dilakukan dengan cara menemukan beberapa pola serupa dalam data seperti bentuk, ukuran, warna, atau perilaku dan membaginya sesuai pola tersebut. Pada bagian ini, data yang telah siap diolah akan dilakukan proses *clustering* berdasarkan keunikannya masing-masing. *Clustering* membagi data menjadi beberapa cluster, contohnya cluster customer yang sangat royal - royal - tidak royal. Berdasarkan *cluster* tersebut dapat ditelusuri kebiasaan membeli para pelanggan sehingga dapat menentukan strategi yang tepat bagi setiap *cluster* yang ada. *Cluster* yang akan dibuat adalah menggunakan algoritma k-means. Berdasarkan jurnal [17] dan [19], k-means memiliki performa yang baik dalam membentuk *cluster*.

3.2.5 Evaluation

Dalam penelitian ini, model klasifikasi *churn* yang diusulkan dievaluasi menggunakan akurasi, presisi, *recall*, dan *f-measure*. Akurasi akan menghitung metrik yang mengidentifikasi sejumlah contoh yang diklasifikasikan dengan benar [17]. Dalam hal ini, akurasi adalah jumlah berapa banyak klasifikasi yang benar dilakukan oleh model klasifikasi terhadap data *churn* dan tidak *churn*. Akurasi adalah metrik paling sederhana dan paling banyak digunakan untuk mengukur kinerja sebuah *classifier* [25]. Model *accuracy* didefinisikan sebagai jumlah klasifikasi yang diprediksi dengan benar oleh model dibagi dengan jumlah total prediksi yang dibuat. Hal tersebut merupakan salah satu cara untuk menilai kinerja model. Berikut rumus tingkat akurasi:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Rumus 3. 1 Rumus Accuracy

Keterangan :

- TP / True Positive adalah data churn yang diprediksi churn.
- TN / True Negative adalah data tidak churn yang diprediksi tidak churn.
- FP / False Positive adalah data tidak churn yang diprediksi churn.
- FN / False Negative adalah data churn yang diprediksi tidak churn.

Tingkat TP juga dikenal sebagai sensitivitas yang memberitahu bagian mana dari data yang diklasifikasikan dengan benar sebagai positif. *True Positive* adalah jumlah dari banyaknya data *churn* yang benar diprediksi *churn*. Perhitungan tingkat TP dilakukan dengan rumus berikut:

$$TP Rate = \frac{True Positive}{Actual Positive}$$

Rumus 3. 2 Rumus TP Rate

Tingkat FP memberitahu bagian mana dari data yang salah diklasifikasikan sebagai positif. False Positive adalah banyaknya data yang tidak *churn* namun diprediksi *churn*. Hasil dari tingkat FP harus rendah untuk semua pengklasifikasi. Rumus tingkat FP adalah:

$$FP\ Rate = \frac{False\ Positives}{Actual\ Negatives}$$

Rumus 3. 3 Rumus FP Rate

Akurasi, juga dikenal sebagai *Positive Predictive Value* (PPV), menunjukkan bagian mana dari data prediksi yang positif. PPV juga biasa dikenal sebagai *precision*. *Precision* menggambarkan tingkat keakuratan antara data yang diminta dengan hasil prediksi yang diberikan oleh model. Pada penelitian ini, *Precision* merupakan perbandingan antara *True Positive* dengan banyaknya data yang diprediksi positif *churn*.

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)}$$

Rumus 3. 4 Rumus Precision

Recall atau *Sensitivity* (*True Positive Rate*) menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi. *Recall* merupakan perbandingan antara True Positive dengan banyaknya data yang *actual* positif *churn*.

$$Recall = \frac{(True\ Positive)}{(True\ Positive + False\ Negative)}$$

Rumus 3. 5 Rumus Recall

Nilai *F-Measure* adalah pertukaran antara pengklasifikasian yang benar pada semua titik data dan memastikan bahwa setiap kelas berisi poin hanya satu kelas. *F-Measure* merupakan *harmonic mean* dari nilai *precision* dan *recall*. Nilai terbaik F-Measure adalah 1.0 dan nilai terburuknya adalah 0. Jika F-Measure punya skor yang baik maka mengindikasikan bahwa model memiliki *precision* dan *recall* yang baik.

$$F - Measure = 2 * C \frac{Precision * Recall}{Precision + Recall} D$$

Rumus 3. 6 Rumus F-Measure

Selanjutnya adalah overfitting yang merupakan suatu kondisi dimana nilai akurasi pada data *train* jauh lebih tinggi dibandingkan nilai akurasi pada data *test*. Berdasarkan penelitian terdahulu, model yang memiliki overfitting adalah model yang kurang baik dikarenakan model tersebut hanya "menghafal" pola yang diamati dari data sebelumnya namun tidak benar-benar "mempelajari" pola yang relevan [24].

3.2.6 Deployment

Tahap selanjutnya adalah *deployment* dengan menerapkan model yang telah dibuat ke data baru dengan jumlah data sebanyak 18.367 baris dengan rentang waktu dari bulan Maret 2022 hingga Mei 2022. Model klasifikasi diimplementasikan ke data baru untuk melakukan *labelling* berdasarkan hasil pembelajaran pada *training* model sebelumnya pada proses *modelling*. Sedangkan, model *clustering* diterapkan untuk membagi data baru menjadi beberapa *cluster* untuk tujuan *customer segmentation*.

3.3 Variabel Penelitian

Pada penelitian ini menggunakan metode *classification* dan *clustering* dalam mengolah data *customer*. *Customer* akan diklasifikasi *churn* atau *non-churn* dan proses selanjutnya adalah membagi *customer* menjadi beberapa *cluster* untuk memperjelas kelompok-kelompok yang ada sehingga dapat menjadi target *marketing* yang tepat bagi perusahaan untuk mengurangi adanya *churn*.

3.3.1 Variabel Dependen

Variabel Dependen atau variabel terikat adalah variabel yang dipengaruhi oleh variabel lain. Pada penelitian ini terdapat 1 variabel dependen yaitu *Churn Customer* yang dianggap *churn* adalah *customer* yang tidak melakukan transaksi lebih dari 30 hari, dimana hal tersebut

berdasarkan dari ketentuan perusahaan Flash Coffee sendiri yang mendefinisikan *indicator churn*. Dalam bisnisnya, Flash Coffee setiap bulan melakukan perhitungan *retention rate* dan *revenue* perusahaan.

Tabel 3. 2 Variabel Dependen

Variabel Penelitian	Deskripsi Variabel
<i>Churn</i>	Pelanggan yang tidak melakukan transaksi selama lebih dari 30 hari akan dianggap <i>churn</i> .

3.3.2 Variabel Independen

Variabel Independen atau variabel bebas adalah variabel yang tidak dipengaruhi oleh variabel lain. Pada penelitian ini terbagi menjadi 2 bagian yaitu variabel independen berdasarkan algoritma Random Forest dan algoritma Logistic Regression. Tabel 3.3 adalah penjelasan terkait kumpulan variabel independen pada model dengan algoritma Random Forest yang terdiri dari *storeName*, *productName*, *sumFlashPoint*, *totalPayments*, *pointAfter*, *totalPay*, *getFlashPoint*, *totalItems*, *promoName1*, *pointBefore*, *userPlatform*, *cnt_trx*, *cnt_promo*, *paymentMethod*, *items_qty*, *totalFood*, *food_qty*.

Tabel 3. 3 Variabel Independen Random Forest

Variabel Penelitian	Deskripsi Variabel
<i>storeName</i>	Nama <i>store</i> tempat <i>customer</i> melakukan pembelian produk Flash Coffee.
<i>productName</i>	Nama produk yang dibeli oleh <i>customer</i> .
<i>sumFlashPoint</i>	Jumlah <i>flashpoint</i> yang didapatkan dari keseluruhan transaksi tersebut.
<i>totalPayments</i>	Total uang yang dibayarkan oleh <i>customer</i> dari seluruh transaksi.

pointAfter	Jumlah <i>flashpoint</i> setelah melakukan transaksi.
totalPay	Jumlah uang yang dibayarkan dalam suatu transaksi.
getFlashPoint	Jumlah <i>flashpoint</i> yang didapatkan dari suatu transaksi.
totalItems	Total <i>items</i> yang dibeli dari keseluruhan transaksi.
promoName1	Nama promo yang digunakan oleh <i>customer</i> .
pointBefore	Jumlah <i>flashpoint</i> sebelum melakukan transaksi.
userPlatform	Sistem operasi (iOS / Android) yang digunakan <i>customer</i> dalam melakukan transaksi.
cnt_trx	Jumlah keseluruhan transaksi yang dilakukan <i>customer</i> .
cnt_promo	Jumlah keseluruhan berapa banyak <i>customer</i> menggunakan promo dalam melakukan transaksi.
paymentMethod	Metode pembayaran yang digunakan oleh <i>customer</i> .
items_qty	Jumlah <i>items</i> yang dibeli dalam suatu transaksi.
totalFood	Jumlah makanan yang dibeli dari keseluruhan transaksi.
food_qty	Jumlah makanan yang dibeli pada suatu transaksi.

Tabel 3. 4 Variabel Independen Logistic Regression

Variabel Penelitian	Deskripsi Variabel
paymentMethod	Metode pembayaran yang digunakan oleh <i>customer</i> .

userPlatform	Sistem operasi (iOS / Android) yang digunakan <i>customer</i> dalam melakukan transaksi.
items_qty	Jumlah <i>items</i> yang dibeli dalam suatu transaksi.
orderType	Tipe <i>order</i> yang dilakukan.
customerTier	Status <i>membership customer</i> pada Flash Coffee App.
cnt_trx	Jumlah keseluruhan transaksi yang dilakukan <i>customer</i> .
totalItems	Total <i>items</i> yang dibeli dari keseluruhan transaksi.
cnt_promo	Jumlah keseluruhan berapa banyak <i>customer</i> menggunakan promo dalam melakukan transaksi.
productName	Nama produk yang dibeli oleh <i>customer</i> .
promoName1	Nama promo yang digunakan oleh <i>customer</i> .
storeName	Nama <i>store</i> tempat <i>customer</i> melakukan pembelian produk Flash Coffee.

Tabel 3.4 merupakan penjelasan terkait variabel independen pada model dengan menggunakan algoritma Logistic Regression yang terdiri dari *paymentMethod*, *userPlatform*, *items_qty*, *orderType*, *customerTier*, *cnt_trx*, *totalItems*, *cnt_promo*, *productName*, *promoName1*, *storeName*.

3.4 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang didapat langsung dari perusahaan Flash Coffee. Data tersebut menjelaskan mengenai data *order* pada *customer* yang melakukan pemesanan melalui Flash App di Singapura. Periode pengambilan data untuk tahap *modelling* adalah data bulan Februari 2021 hingga Februari 2022 dengan jumlah keseluruhan data adalah sebanyak 174,936

records, sedangkan untuk data pada tahap deployment adalah menggunakan data dengan periode Maret 2022 hingga Mei 2022 dengan jumlah data sebanyak 18,367 *records*.

3.5 Teknik Pengambilan Sampel

Teknik pengambilan sampel pada penelitian ini adalah dengan melakukan random sampling. Dimana pada sampel tersebut akan dibagi menjadi 2 bagian yaitu data *training* dan data *testing* untuk dilakukan pembuatan model klasifikasi. Data training yang akan digunakan adalah data dengan persentase 70% dari total data keseluruhan. Sedangkan, data testing menggunakan data dengan persentase 30%.

3.6 Teknik Analisis Data

Pada penelitian ini dilakukan analisis data dengan mengumpulkan data *customer order* yang melakukan pemesanan melalui aplikasi dan berwilayah di Singapura. Setelah mendapatkan data tersebut akan dilakukan *cleaning* dan *preprocessing* data dengan menggunakan bahasa pemrograman Python. Data yang telah siap digunakan akan dibagi menjadi 2 yaitu data *training* dan data *testing*, lalu membuat model dengan menggunakan algoritma *classification*. Model tersebut yang akan mengklasifikasikan *churn customer* Flash Coffee (*churn* atau *non-churn*). Langkah selanjutnya adalah melakukan perhitungan tingkat akurasi pada model untuk menilai seberapa akurat dan tepat model yang telah dibuat untuk melakukan klasifikasi.