

NPM : 11122246
Nama : Reymondo Saputra S
Kelas : 3KA01

CRISP – DM

CRISP-DM (Cross Industry Standard Process for Data Mining) adalah sebuah metodologi yang umum digunakan dalam proyek data mining dan analisis data. Metodologi ini membantu dalam memberikan panduan langkah demi langkah yang terstruktur untuk menyelesaikan proyek berbasis data.

CRISP-DM terdiri dari 6 tahapan:

1. Business Understanding (Pemahaman Bisnis)
2. Data Understanding (Pemahaman Data)
3. Data Preparation (Persiapan Data)
4. Modeling (Pemodelan)
5. Evaluation (Evaluasi)
6. Deployment (Penerapan).

Penerapan CRISP – DM:

1. Business Understanding (Pemahaman Bisnis)

Latar Belakang:

RetailMart adalah sebuah toko retail besar yang menawarkan berbagai macam produk mulai dari kebutuhan sehari-hari hingga elektronik. Saat ini, mereka ingin meningkatkan penjualan dan loyalitas pelanggan karena persaingan yang semakin ketat di industri retail.

Tujuan Bisnis:

RetailMart ingin mengembangkan strategi pemasaran yang lebih tepat sasaran dengan fokus untuk meningkatkan pendapatan dan loyalitas pelanggan. Secara khusus, mereka ingin memahami pola pembelian pelanggan mereka untuk mengidentifikasi segmen pelanggan yang berbeda, sehingga kampanye pemasaran dapat disesuaikan dengan kebutuhan setiap segmen tersebut.

Kriteria Kesuksesan:

- Peningkatan penjualan melalui kampanye pemasaran yang lebih efektif.
- Peningkatan retensi dan loyalitas pelanggan berdasarkan hasil dari kampanye yang lebih personal.
- Identifikasi segmen pelanggan yang jelas dan dapat ditindaklanjuti untuk membantu keputusan pemasaran di masa depan.

Ruang Lingkup Proyek: Proyek ini akan fokus pada analisis data penjualan pelanggan selama 12 bulan terakhir untuk memahami pola pembelian dan segmentasi pelanggan. Kampanye pemasaran awal akan diuji pada beberapa segmen dengan evaluasi berdasarkan respons kampanye, seperti tingkat pembelian kembali dan peningkatan jumlah transaksi.

Dengan mendefinisikan pemahaman bisnis seperti ini, tim data di RetailMart dapat memulai langkah-langkah berikutnya dalam CRISP-DM, seperti pemahaman data dan persiapan data, dengan fokus yang jelas pada tujuan yang ingin dicapai.

2. *Data Understanding* (Pemahaman Data)

Tujuan Bisnis:

RetailMart telah mengidentifikasi tujuan bisnis mereka untuk meningkatkan strategi pemasaran dan memahami perilaku pelanggan.

- **Data Transaksi Penjualan:** RetailMart mengakses riwayat transaksi dari sistem point-of-sale (POS) untuk melihat pola pembelian, jenis produk yang paling laku, waktu pembelian, serta penggunaan diskon dan promosi.
- **Data Demografi Pelanggan:** Data ini berasal dari program loyalitas pelanggan dan berisi informasi seperti usia, jenis kelamin, lokasi geografis, serta tingkat pendapatan. RetailMart menggunakannya untuk memahami segmen pelanggan yang berbeda.
- **Data Interaksi Pelanggan:** RetailMart juga memiliki data dari interaksi pelanggan di berbagai saluran, seperti chat online, email, serta media sosial. Data ini mencakup keluhan pelanggan, feedback produk, dan tingkat kepuasan pelanggan.

Setelah mengumpulkan data ini, RetailMart akan melakukan eksplorasi awal untuk memahami struktur dan kualitas data tersebut. Langkah-langkah yang dilakukan meliputi:

- **Mengidentifikasi Kualitas Data:** RetailMart menganalisis apakah terdapat missing values, duplikasi, atau inkonsistensi dalam data. Misalnya, apakah semua transaksi tercatat dengan benar, atau apakah ada pelanggan yang terdaftar dua kali dengan identitas yang berbeda.
- **Mengeksplorasi Distribusi Data:** RetailMart melihat distribusi jumlah transaksi per bulan, segmentasi umur pelanggan, atau frekuensi interaksi pelanggan di media sosial. Distribusi ini membantu dalam memahami pola dan tren yang ada.
- **Menentukan Keterkaitan Antar Variabel:** RetailMart mengevaluasi keterkaitan antar variabel, seperti apakah ada hubungan antara lokasi geografis pelanggan dengan jenis produk yang dibeli, atau apakah umur pelanggan memengaruhi preferensi produk tertentu.

Melalui eksplorasi ini, RetailMart dapat memperoleh wawasan awal tentang pola dalam data yang nantinya akan digunakan untuk persiapan data dan pemodelan di fase berikutnya.

3. *Data Preparation* (Persiapan Data)

Tujuan Bisnis:

Setelah menyelesaikan tahap Data Understanding dan mengidentifikasi berbagai pola serta kualitas data, RetailMart melanjutkan ke tahap **Data Preparation**. Dalam tahap ini, fokus RetailMart adalah membersihkan dan mempersiapkan data agar siap digunakan dalam analisis lebih lanjut.

Langkah-langkah yang dilakukan RetailMart dalam tahap ini meliputi:

- a) **Menangani Missing Values (Nilai yang Hilang):** Tim data RetailMart menemukan bahwa beberapa entri data, seperti informasi demografi pelanggan dan data transaksi, memiliki nilai yang hilang. RetailMart memilih beberapa strategi penanganan:
 - Menghapus entri yang tidak lengkap jika nilai yang hilang sangat signifikan.
 - Menggunakan imputasi, seperti mengganti nilai yang hilang dengan rata-rata, median, atau metode interpolasi lainnya untuk variabel-variabel yang penting.
- b) **Menghapus Duplikasi:** Dalam eksplorasi sebelumnya, ditemukan bahwa beberapa pelanggan terdaftar lebih dari sekali atau ada transaksi yang tercatat dua kali. RetailMart melakukan pembersihan dengan menghapus entri yang duplikat, memastikan setiap pelanggan dan transaksi hanya tercatat satu kali.
- c) **Menangani Outliers (Pencilan):** RetailMart juga menemukan beberapa transaksi dengan jumlah yang jauh lebih besar dari rata-rata, yang bisa jadi merupakan kesalahan input. Outlier ini dianalisis lebih lanjut, dan jika dipastikan tidak relevan, mereka dihapus atau disesuaikan agar tidak mendistorsi analisis berikutnya.
- d) **Integrasi Data dari Berbagai Sumber:** Data dari transaksi penjualan, data demografi pelanggan, dan interaksi di berbagai saluran kini diintegrasikan menjadi satu dataset

yang kohesif. Proses integrasi ini dilakukan dengan menggabungkan data berdasarkan kunci identifikasi yang unik, seperti ID pelanggan, sehingga data dari berbagai sumber dapat disinkronkan dengan benar.

- e) **Transformasi Data:** Data yang ada dalam format yang berbeda (misalnya, format tanggal yang tidak seragam) dinormalisasi agar konsisten. RetailMart juga mengubah beberapa kategori menjadi format numerik atau biner untuk memudahkan analisis lebih lanjut.
- f) **Pembuatan Fitur-Fitur Baru (Feature Engineering):** RetailMart mulai membuat fitur-fitur baru yang dapat memberikan nilai lebih dalam analisis, seperti:
 - **Frekuensi Pembelian:** Menghitung berapa kali pelanggan melakukan pembelian dalam periode tertentu.
 - **Rata-rata Pengeluaran Per Kunjungan:** Mengkalkulasi rata-rata jumlah uang yang dihabiskan setiap kali pelanggan datang ke toko.
 - **Loyalitas Pelanggan:** Menambahkan fitur loyalitas yang diukur berdasarkan seberapa sering pelanggan berbelanja atau berinteraksi dengan brand di berbagai saluran.

Dengan semua data yang telah dibersihkan, disatukan, dan ditransformasi, RetailMart kini memiliki dataset yang lebih siap untuk proses modeling di tahap selanjutnya. Fitur-fitur baru yang relevan juga disiapkan untuk meningkatkan efektivitas analisis dan pemodelan.

4. *Modeling* (Pemodelan)

Tujuan Bisnis:

Setelah melalui tahap Data Preparation dan memiliki dataset yang bersih serta fitur-fitur yang relevan, RetailMart kini memasuki tahap **Modeling**. Dalam fase ini, tujuan RetailMart adalah membangun dan menguji model yang dapat memberikan wawasan lebih dalam tentang perilaku pelanggan dan membantu dalam pengambilan keputusan bisnis, seperti segmentasi pelanggan.

Langkah-langkah yang dilakukan RetailMart dalam tahap Modeling meliputi:

- a) **Memilih Teknik Pemodelan yang Tepat:**
 - Berdasarkan tujuan bisnis RetailMart, salah satu teknik yang dipilih adalah **clustering** untuk melakukan segmentasi pelanggan. Clustering memungkinkan RetailMart untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu, seperti frekuensi pembelian, preferensi produk, atau total pengeluaran.
 - Selain itu, RetailMart juga mempertimbangkan teknik lain, seperti **regresi** untuk memprediksi total pengeluaran pelanggan di masa depan dan **classification** untuk mengidentifikasi pelanggan yang paling mungkin beralih ke kompetitor.
- b) **Membangun Beberapa Model:**
 - RetailMart membangun beberapa model clustering menggunakan teknik **K-Means Clustering** dan **Hierarchical Clustering** untuk menemukan kelompok pelanggan yang berbeda.
 - Untuk prediksi, RetailMart juga membangun beberapa model regresi, termasuk **Linear Regression** dan **Random Forest Regression**, untuk melihat model mana yang memberikan hasil terbaik dalam memprediksi total pengeluaran pelanggan.
 - Model **Classification** seperti **Decision Trees** dan **Logistic Regression** juga dicoba untuk mengklasifikasikan pelanggan berdasarkan loyalitas mereka.
- c) **Pengujian Model:**

- Setiap model diuji dengan menggunakan data yang disiapkan sebelumnya. RetailMart membagi dataset menjadi **training data** dan **testing data** untuk mengevaluasi kinerja masing-masing model.
 - **Cross-validation** juga digunakan untuk memastikan hasil dari model tidak hanya bagus di satu subset data, tetapi juga dapat diaplikasikan dengan baik pada data yang lebih umum.
- d) **Validasi untuk Menghindari Overfitting:**
- RetailMart menerapkan berbagai teknik validasi untuk mencegah **overfitting**, di mana model bekerja sangat baik pada data training tetapi tidak dapat menggeneralisasi pada data baru. Teknik yang digunakan termasuk **k-fold cross-validation** dan **regularization**.
 - Setelah validasi dilakukan, model dengan performa terbaik, baik dari segi akurasi maupun kemampuan generalisasi, dipilih untuk digunakan lebih lanjut.
- e) **Evaluasi Kinerja Model:**
- Setiap model dievaluasi menggunakan metrik kinerja yang relevan, seperti **Silhouette Score** untuk model clustering atau **Mean Squared Error (MSE)** untuk model regresi. RetailMart membandingkan hasil dari berbagai model untuk menentukan model yang paling sesuai dengan tujuan bisnis.

Setelah tahap modeling ini, RetailMart siap untuk melakukan evaluasi lebih lanjut terhadap model yang dipilih, memastikan model tersebut memenuhi kebutuhan bisnis, dan dapat diintegrasikan dalam operasional perusahaan.

5. Evaluation (Evaluasi)

Tujuan bisnis:

Setelah menyelesaikan tahap pemodelan dan membangun beberapa model yang dapat memprediksi serta menganalisis perilaku pelanggan, RetailMart memasuki tahap **Evaluation**. Pada tahap ini, RetailMart menilai apakah model yang telah dibuat memberikan hasil yang sesuai dengan tujuan bisnis yang telah ditetapkan sebelumnya, yaitu meningkatkan pemahaman tentang perilaku pelanggan dan mengoptimalkan strategi pemasaran.

Langkah-langkah yang dilakukan RetailMart dalam tahap Evaluation meliputi:

- **Mengevaluasi Kesesuaian Model dengan Tujuan Bisnis:**
 - RetailMart membandingkan hasil model dengan tujuan utama bisnis yang telah ditetapkan pada fase Business Understanding, seperti apakah model clustering membantu mengidentifikasi segmen pelanggan yang relevan untuk kampanye pemasaran yang lebih efektif.
 - Tim data RetailMart juga menilai apakah model prediksi pengeluaran pelanggan dapat memberikan hasil yang akurat dalam memprediksi total pengeluaran dan potensi loyalitas pelanggan di masa mendatang.
- **Analisis Wawasan yang Dihasilkan:**
 - RetailMart menganalisis apakah hasil model benar-benar memberikan **wawasan yang bermanfaat**. Misalnya, apakah model clustering menunjukkan kelompok pelanggan yang bisa diberi penawaran khusus atau apakah model klasifikasi berhasil mengidentifikasi pelanggan yang rentan meninggalkan program loyalitas.

- RetailMart juga melihat apakah fitur-fitur yang dihasilkan dari model, seperti frekuensi pembelian atau rata-rata pengeluaran, dapat digunakan untuk membuat keputusan bisnis yang lebih baik.
- **Memeriksa Kinerja Model dengan Metrik yang Tepat:**
 - RetailMart menggunakan metrik evaluasi yang sudah ditentukan untuk menilai kinerja model. Misalnya, untuk model clustering, **Silhouette Score** digunakan untuk memastikan bahwa segmen pelanggan yang dihasilkan adalah kohesif dan berbeda satu sama lain. Untuk model prediksi, **Mean Squared Error (MSE)** dan **Accuracy** dievaluasi.
 - Jika hasil model tidak memuaskan dari segi performa atau metrik yang diinginkan, RetailMart siap untuk melakukan penyesuaian, seperti menambah data atau melakukan transformasi tambahan pada fitur.
 - **Iterasi Ulang Jika Diperlukan:**
 - Jika hasil evaluasi menunjukkan bahwa model tidak mencapai tujuan bisnis atau hasil yang diinginkan tidak memadai, RetailMart akan kembali ke fase sebelumnya, seperti **Modeling** atau bahkan **Data Preparation**, untuk memperbaiki atau menyempurnakan model.
 - RetailMart dapat mencoba pendekatan pemodelan yang berbeda atau melakukan iterasi pada proses feature engineering, seperti menciptakan fitur baru atau menghilangkan fitur yang tidak relevan.
 - **Mempersiapkan Penerapan Hasil:**
 - Jika model telah terbukti memadai dan relevan dengan tujuan bisnis, RetailMart mulai mempersiapkan penerapan hasil model ini ke dalam proses bisnis nyata. Misalnya, hasil dari model clustering dapat digunakan untuk menciptakan kampanye pemasaran yang lebih tersegmentasi dan tertarget, atau model prediksi dapat digunakan untuk program retensi pelanggan.

Dengan evaluasi ini, RetailMart memastikan bahwa model yang dibangun tidak hanya baik secara teknis, tetapi juga relevan dengan kebutuhan bisnis. Jika hasil model belum memuaskan, tim RetailMart siap untuk melakukan iterasi ulang hingga mendapatkan hasil yang sesuai.