

PROGRAM STUDI SARJANA SISTEM INFORMASI
PROPOSAL PROYEK PENAMBANGAN DATA



Customer Clustering for CRM in XYZ Store

DISUSUN OLEH:

12S17001	Krista Lumbantoruan
12S17042	Aulia SL Pakpahan
12S17049	Paddy T Silitonga

INSTITUT TEKNOLOGI DEL
NOVEMBER 2020

Daftar Isi

Bab 1. Business Understanding	1
1.1 <i>Business Objectives</i>	1
1.2 <i>Situation Assesment</i>	2
1.3 <i>Data Mining Goal</i>	2
1.4 <i>Produce Project Plan</i>	2
Bab 2. Data Understanding	4
2.1 <i>Collect Initial Data</i>	4
2.2. <i>Describe Data</i>	4
2.3. <i>Explore Data</i>	4
2.4. <i>Verify Data Quality</i>	5
Referensi	5

Bab 1. Business Understanding

1.1 *Business Objectives*

Perkembangan teknologi yang semakin pesat dimanfaatkan oleh para pelaku bisnis dengan memanfaatkan teknologi digital dan internet. Salah satu contoh pelaku bisnis yaitu Toko XYZ yang merupakan toko bahan makanan yang memiliki banyak pelanggan dengan karakteristik yang berbeda-beda. Toko XYZ mengalami masalah dalam meningkatkan profit bisnis dan menentukan parameter yang tepat dalam membuat strategi pemasaran. Dalam hal ini, Toko XYZ ingin membuat keputusan dalam memperoleh strategi yang dapat meningkatkan kualitas pemasaran. Pada proyek ini, penulis akan mencoba menganalisis perilaku dari pelanggan dengan memberikan solusi untuk permasalahan Toko XYZ.

Kualitas pemasaran yang baik dapat memicu profit yang lebih besar. Salah satu cara untuk memperoleh keuntungan dalam bidang pemasaran adalah memahami pelanggan. *Customer Relationship Management* (CRM) merupakan sebuah pendekatan strategis yang berkonsentrasi untuk meningkatkan nilai terhadap perusahaan melalui penanganan hubungan yang tepat dengan pelanggan utama dan juga dengan kelompok pelanggan lainnya. Karena itu, pengelompokan pelanggan atau *customer clustering* merupakan langkah yang baik dalam pemasaran untuk menggambarkan perilaku pelanggan terhadap perusahaan.

Pada proyek ini, *customer clustering* dilakukan dengan analisis RFM (*Recency, Frequency, and Monetary*), yaitu melakukan pengelompokan berdasarkan kapan terakhir pelanggan melakukan transaksi, sering tidaknya melakukan transaksi, dan transaksi yang dilakukan pada jumlah sedikit atau besar. Dengan demikian, tujuan proyek ini adalah untuk memahami karakteristik pelanggan dengan melakukan pengelompokan berdasarkan data transaksi belanja pelanggan yang dianalisis menggunakan model RFM. Pengelompokan tersebut dilakukan dengan melihat reaksi pasar pelanggan atau transaksi pemasaran yang dilakukan agar strategi pemasaran dan layanan yang diberikan dengan tepat.

1.2 Situation Assesment

Toko XYZ merupakan toko yang menjual berbagai jenis bahan makanan. Toko ini ingin meningkat kualitas pemasaran dengan membuat strategi pemasaran yang tepat untuk memicu profit yang lebih besar. Rekaman data pemasaran pelanggan Toko XYZ dapat digunakan untuk meninjau perilaku pelanggan. Rekaman data tersebut antara lain waktu transaksi, id pelanggan, id produk, dan lain-lain. Atribut terkait transaksi akan menjadi fokus dalam pengelompokan pelanggan di Toko XYZ berdasarkan analisis model RFM. Batasan dari proyek ini adalah rekaman data yang digunakan pada Toko XYZ adalah pada 1 November 2000 sampai dengan 28 Februari 2001. Penentuan keputusan dalam pembuatan strategi adalah berfokus pada data transaksi pelanggan dengan menggunakan analisis RFM.

1.3 Data Mining Goal

Tujuan dari penambahan data ini adalah agar didapatkan suatu pengetahuan mengenai perilaku pelanggan sehingga dapat menjadi salah satu parameter untuk membuat keputusan atau strategi yang tepat dalam meningkatkan kualitas pemasaran yang memicu peningkatan profit.

1.4 Produce Project Plan

Proyek ini akan menggunakan algoritma *K-Means Clustering* pada *python*. Data pelanggan akan diekstraksi berdasarkan tiga variabel RFM setiap pelanggan pada Toko XYZ. Setiap pelanggan diakumulasikan masing-masing nilai RFM, yaitu *Recency*, *Frequency*, dan *Monetary*. *Recency* yaitu rentan terakhir kali transaksi dilakukan. Semakin kecil rentangnya, maka nilai R akan semakin besar. *Frequency* yaitu jumlah transaksi dalam satu periode. Semakin banyak *frequency*, maka nilai F akan semakin besar. *Monetary*, yaitu nilai pelanggan berupa uang yang dikeluarkan pelanggan pada periode tersebut maka nilai M semakin besar. Setelah dilakukan ekstraksi data pada tahap *data preparation*, maka dilakukan penentuan jumlah *cluster* menggunakan *Elbow method*. Pemodelan data akan digunakan untuk mengoptimalkan hasil pengelompokan. *Modeling* merupakan tahap untuk melakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal.

Menurut Tsipitsis dan Chorianopoulos [1], analisis RFM digunakan untuk memahami karakteristik *customer*. Ada 5 label pelanggan yang akan digunakan pada penelitian ini yaitu, *superstars customers*, *golden customer*, *occasional customers*, *everyday shoppers*, dan *dormant customers*. Label ini dipisahkan berdasarkan transaksi yang dilakukan pelanggan. Pelanggan yang sering melakukan transaksi dan dengan nilai yang besar akan masuk ke dalam label *superstars customers*. *Golden customer* yaitu pelanggan sering melakukan transaksi dengan nilai transaksi kedua tertinggi. *Occasional customer* yaitu label untuk pelanggan yang tidak terlalu sering melakukan transaksi namun memiliki nilai transaksi yang tinggi. *Everyday shopper* untuk pelanggan yang sering bertransaksi dengan nilai transaksi yang rendah. Dan untuk pelanggan yang memiliki nilai transaksi rendah dan sudah lama tidak bertransaksi masuk ke dalam label *dormant customers*.

Hasil dari proses *K-Means Clustering* akan berupa informasi yang menunjukkan jumlah pelanggan setiap cluster, titik pusat atau centroid, dan nilai performa cluster. Hasil cluster akan dilakukan pengujian dengan menggunakan SSE atau *Sum of Square Error*. Selanjutnya, hasil *clustering* akan dianalisis dengan memberikan keputusan apakah teknik pemodelan yang dipergunakan dapat dijadikan standar dalam menentukan tujuan proyek. Setelah itu, akan dilakukan *deployment plan* dengan menjelaskan gambaran mengenai rencana terhadap pembuatan laporan, lalu melakukan *produce final report* memberikan visualisasi dari laporan yang telah dibuat berdasarkan pada *deployment plan*.

Bab 2. Data Understanding

2.1 Collect Initial Data

Pengumpulan data dilakukan dengan eksplorasi di internet dan melakukan peninjauan kecocokan dengan penelitian yang dilakukan. Pada akhirnya, dataset yang akan digunakan diperoleh dari <https://www.kaggle.com/chiranjivdas09/ta-feng-grocery-dataset>. Data ini sebelumnya sudah pernah digunakan dalam penelitian *clustering customer segmentation* berdasarkan LRFM dengan *tools* R. Bedanya pada penelitian yang akan penulis lakukan adalah penulis akan melakukan pengelompokan terhadap *customer* berdasarkan RFM menggunakan python. *Code program* untuk penelitian ini juga belum ada ditemukan di github dengan menggunakan *dataset* yang sama dan *tools* yang sama, sehingga dapat dipastikan penelitian ini tidak melakukan *plagiarism*.

2.2 Describe Data

Pada proyek ini, *dataset* yang digunakan adalah **Ta-Feng Grocery** yang merupakan kumpulan data belanja bahan makanan pada sebuah supermarket. Dataset tersebut merupakan hasil rekaman data pemasaran pelanggan dari 1 November 2000 sampai dengan 28 Februari 2001.

2.3 Explore Data

Dataset Ta-Feng Grocery terdiri dari 817.741 transaksi milik 32.266 pengguna dengan 23.812 item. Dataset terdiri dari 10 atribut, yaitu *transaction_dt*, *customer_id*, *age_group*, *pin_code*, *product_subclass*, *product_id*, *amount*, *asset*, dan *sales_price*.

Table 1 Atribut dan Tipe Data Mentah

Atribut	Tipe data pada Excel	Keterangan
TRANSACTION_DT	Number	Tanggal transaksi
CUSTOMER_ID	Number	Id pelanggan yang melakukan transaksi
AGE_GROUP	Number	Grup usia pelanggan yang melakukan transaksi

PIN_CODE	Number	Kode pos pelanggan yang melakukan transaksi
PRODUCT_SUBCLASS	Number	Subkelas produk yang dibeli pelanggan
PRODUCT_ID	Number	Id produk yang dibeli pelanggan
AMOUNT	Number	Jumlah produk yang dibeli
ASSET	Number	Aset yang dimiliki oleh produk
SALES_PRICE	Number	Total harga produk

2.4. Verify Data Quality

Kualitas data pada **Ta-Feng Grocery** belum baik, sehingga masih dibutuhkan *data preprocessing* untuk menghasilkan kualitas data yang baik. Pada *dataset*, terdapat data yang kotor dan bernilai kosong. Untuk mengatasi masalah ini, maka tahap selanjutnya akan dilakukan *preprocessing data* untuk mendukung proses penambangan data dan menghasilkan pengelompokan yang tepat.

Referensi

- [1] E. Haddi, X. Liu, and Y. Shi, "SENTIMENT ANALYSIS OF HOTEL REVIEW USING NAÏVE BAYES ALGORITHM AND INTEGRATION OF INFORMATION GAIN AND GENETIC ALGORITHM AS FEATURE SELECTION METHODS," *Int. Semin. Sci. Issues Trends Bekasi*, 2014.
- [2] Subburathinam, K. (2016). *CUSTOMER SEGMENTATION FRAMEWORK USING REDEFINED*. May 2015.