

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

PT. X merupakan perusahaan farmasi yang fokus dalam memproduksi dan mengembangkan produk kesehatan seperti obat resep, obat bebas, vitamin, dan suplemen. Penelitian ini dilakukan untuk mengembangkan model rekomendasi yang dapat memberikan rekomendasi 10 produk kepada customer melalui aplikasi B2B e-commerce milik distributor PT. X.

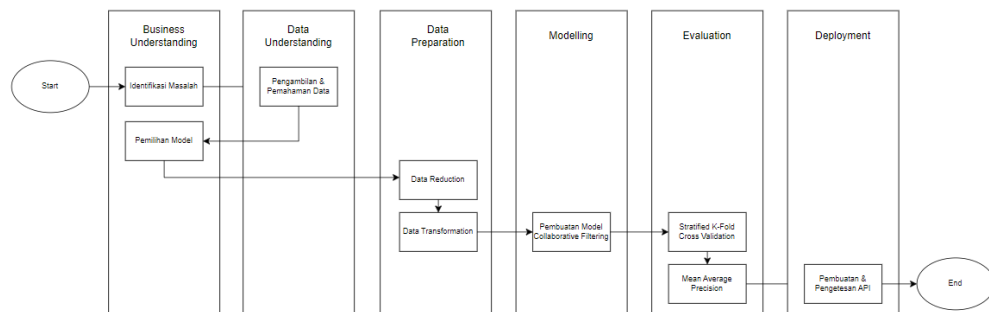
3.2 Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode kualitatif dan kuantitatif. Metode kualitatif merupakan metode yang melakukan pendekatan melalui wawancara, analisis dokumen, dan studi kasus, sedangkan metode kuantitatif merupakan metode yang terdiri dari pengumpulan serta analisis data numerik [43]. Metode kualitatif yang digunakan pada penelitian ini adalah proses wawancara mengenai studi kasus sistem rekomendasi produk bersama dengan tim data science dari PT. X, dan metode kuantitatif yang digunakan pada penelitian ini adalah pembuatan model sistem rekomendasi yang memerlukan perhitungan matematis pada matriks similaritas serta *formula* atau rumus *collaborative filtering* berbasis *user*, *item*, dan *hybrid*. Selain itu, metode kuantitatif juga digunakan pada tahap evaluasi yang diperlukan untuk menghitung nilai presisi menggunakan *formula* atau rumus *mean average precision*.

3.2.1 Alur Penelitian

Alur penelitian yang dilakukan pada penelitian ini dapat dilihat pada gambar 3.1. Alur tersebut dimulai dengan melakukan identifikasi masalah, pengambilan data, pemilihan model sistem rekomendasi, pengelolaan data, hingga pembuatan serta pengetesan API. Seluruh alur penelitian tersebut dibungkus menggunakan framework CRISP-DM dimana alur identifikasi masalah, pengambilan data, dan pemilihan model termasuk ke tahap business

understanding dan data understanding. Proses pengelolaan data yang terdiri dari *data reduction*, *data transformation* merupakan alur yang dilakukan pada tahap *data preparation*, kemudian alur pembuatan model rekomendasi menggunakan *collaborative filtering* merupakan alur yang dilakukan pada tahap *modelling*, dan evaluasi menggunakan teknik *stratified k-fold cross validation* dengan metrik pengukuran *mean average precision* merupakan alur yang dilakukan pada tahap *evaluation*. Penelitian ini ditutup dengan proses pembuatan serta pengetesan API yang alur pada tahap *deployment*.



Gambar 3. 1 Alur Penelitian

Pada tahap mengidentifikasi masalah, ditemukan bahwa model sistem rekomendasi milik PT. X yang menggunakan *package recommenderlab* masih belum memiliki evaluasi secara teknikal sehingga hasil rekomendasi yang dihasilkan tidak dapat diukur tingkat presisinya. Oleh karena itu, tujuan dilakukannya penelitian ini adalah untuk melakukan evaluasi secara teknikal pada model sistem rekomendasi yang ada sekaligus membuat sistem rekomendasi lain yang dapat mengungguli *package recommenderlab* berdasarkan nilai evaluasi tersebut.

Setelah mengetahui permasalahan yang dihadapi, tahap selanjutnya adalah melakukan pengumpulan data dari BigQuery yang sudah disiapkan oleh PT. X. Data tersebut merupakan sumber data utama untuk membuat model sistem rekomendasi baru yang akan dikomparasikan dengan model sistem rekomendasi lama. Model sistem rekomendasi baru akan menggunakan algoritma *collaborative filtering* berbasis *user*, *item*, dan *hybrid* dengan

matriks similaritas *cosine similarity*, dan *manhattan distance*. Penentuan model sistem rekomendasi didasarkan pada referensi penelitian terdahulu serta penyesuaian data yang dapat diperoleh dari PT. X. Kemudian, akan dilakukan pengelolaan data yang terdiri dari tahap *data reduction*, *data preparation*, pembuatan model rekomendasi, dan evaluasi menggunakan *stratified k-fold cross validation* dan metrik pengukuran *mean average precision* untuk mencari model sistem rekomendasi yang dapat memberikan hasil terbaik berdasarkan evaluasi teknikal. Setelah menemukan model yang dapat memberikan hasil rekomendasi terbaik kepada *customer*, akan dibuat sebuah API yang digunakan untuk mengintegrasikan hasil rekomendasi dengan aplikasi B2B *e-commerce*.

3.2.2 Metode *Data Mining*

Penelitian ini memiliki fokus untuk melakukan komparasi pada model sistem rekomendasi dan menghasilkan rekomendasi 10 produk kepada customer menggunakan model terbaik. Keseluruhan hal tersebut akan dilakukan menggunakan metode *data mining*. Terdapat beberapa *framework* pada data mining yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM), dan *Sample, Explore, Modify, Model, dan Assess* (SEMMA). Melalui perbandingan pada tabel 3.1, *framework* data mining yang akan digunakan pada penelitian ini adalah CRISP-DM [29].

Tabel 3. 1 Perbandingan *Framework Data Mining*

Framework	CRISP-DM	SEMMA
Definisi	CRISP-DM merupakan proses dari <i>framework data mining</i> yang bersifat <i>industry-independent</i> .	SEMMA merupakan proses dari <i>framework data mining</i> yang dibuat oleh <i>SAS Institute</i> .
Tahapan	<i>Business understanding, data understanding, data preparation, modeling, evaluation, dan deployment</i>	<i>Sample, explore, modify, model, dan assess</i>
Fleksibilitas	Lebih fleksibel dan cocok untuk berbagai proyek <i>data mining</i> dan <i>machine learning</i>	Lebih terfokus ke penggunaan <i>software SAS</i> dan sering digunakan untuk pendamping metode lain

Terdapat beberapa tahapan yang perlu dilakukan pada penelitian menggunakan *framework* CRISP-DM, yaitu sebagai berikut.

3.2.2.1 Business Understanding

Tahap pertama yang dilakukan adalah pemahaman bisnis. Pada penelitian ini, studi kasus yang diangkat adalah model sistem rekomendasi produk farmasi milik PT. X. Pada model sistem rekomendasi tersebut, setiap *customer* akan diberikan rekomendasi berupa 10 produk yang dapat berbeda antar *customer*. Pada saat ini, perusahaan sudah memiliki model sistem rekomendasi menggunakan *package recommenderlab* dari bahasa pemrograman R. Namun pada saat membangun model sistem rekomendasi tersebut, perusahaan belum melakukan evaluasi secara teknikal sehingga tidak dapat dilakukan evaluasi lebih lanjut. Oleh karena itu, penelitian ini akan melakukan evaluasi secara teknikal menggunakan teknik *stratified k-folds cross validation* dan metrik pengukuran *mean average precision*. Selain itu, penelitian juga akan mengembangkan model sistem rekomendasi baru yang akan dibandingkan dengan model yang sudah ada menggunakan teknik evaluasi yang sama.

3.2.2.2 Data Understanding

Tahap kedua yang dilakukan adalah pemahaman data. Pada penelitian ini, data yang digunakan adalah data histori pemesanan *customer* pada aplikasi B2B *e-commerce* di cabang Bandung, Banjarmasin, dan Samarinda selama 3 bulan. Data tersebut disimpan pada tools BigQuery yang digunakan perusahaan sebagai *data warehouse*. Data terdiri dari *customer_id*, *product_id*, *quantity*, *channel*, dan *branch*. *Channel* merupakan kategori *customer* seperti *hospital* dan *pharmacy*, sedangkan *branch* merupakan pengelompokan cabang berdasarkan area distribusi seperti Bandung, Banjarmasin, dan

Samarinda. Populasi data yang digunakan terdiri dari 877 *customer* dan 409 produk.

3.2.2.3 Data Preparation

Tahap ketiga yang dilakukan adalah preparasi data. Awalnya, data yang disimpan dari BigQuery akan dibaca oleh Python di Jupyter Notebook agar dapat diolah lebih lanjut. Setelah data terbaca, akan dilakukan proses pembersihan data atau *data reduction* yang hanya mengambil data pemesanan dari *customer* dengan minimal pemesanan 5 produk. Kemudian akan dilakukan transformasi nilai *quantity* atau kuantitas dari produk yang dipesan menjadi 1 dan 0. Di saat yang bersamaan, akan dilakukan *cross validation* menggunakan *stratified k-folds* yang akan membagi data untuk pelatihan dan pengujian sebanyak *k* kali. Setelah itu, data pelatihan akan diubah menjadi bentuk *user item preference* agar dapat dilakukan pemodelan, dan data pengujian akan disimpan ke bentuk *dictionary*.

3.2.2.4 Modelling

Tahap keempat yang dilakukan adalah tahap pembuatan sistem rekomendasi untuk menghasilkan rekomendasi 10 produk. Pada tahap ini, akan dibuatkan sistem rekomendasi baru menggunakan algoritma *collaborative filtering* berbasis *user*, *item*, dan *hybrid* yang menggunakan matriks similaritas *cosine similarity*, dan *manhattan distance*. Secara keseluruhan, terdapat 6 model yang akan dibuat yaitu *collaborative filtering* berbasis *user*, *item*, *hybrid* menggunakan *cosine similarity*, dan *collaborative filtering* berbasis *user*, *item*, *hybrid* menggunakan *manhattan distance*.

3.2.2.5 Evaluation

Tahap kelima yang dilakukan adalah tahap evaluasi dari setiap model yang telah dibuat. Selain itu, tahap ini juga akan melakukan komparasi dengan melakukan evaluasi pada sistem rekomendasi *package*

recommenderlab yang sedang dijalankan oleh PT. X. Evaluasi dilakukan menggunakan metrik pengukuran *mean average precision*. Model sistem rekomendasi yang terbaik adalah model sistem rekomendasi yang dapat memberikan nilai *mean average precision* paling tinggi.

3.2.2.6 Deployment

Tahap terakhir yang dilakukan adalah tahap penyebaran model sistem rekomendasi yang telah dibuat. Pada tahap ini, model sistem rekomendasi yang akan digunakan untuk menghasilkan rekomendasi 10 produk dan disebarakan adalah model yang telah dipilih berdasarkan nilai *mean average precision* tertinggi. Model sistem rekomendasi tersebut akan dibuat di file API untuk menghasilkan 10 rekomendasi produk kepada customer menggunakan FastAPI. Rekomendasi kemudian akan dipanggil setiap kali ada *customer* yang membuka aplikasi B2B *e-commerce* melalui *request* yang dikirimkan ke API.

3.3 Teknik Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data histori pendistribusian produk pada aplikasi *B2B e-commerce*. Data tersebut terdiri dari *customer_id*, *product_id*, dan *quantity* yang sudah dilakukan *masking* menjadi 1 dan 0. Selain itu, juga terdapat data terkait informasi *channel* dan *branch* dari setiap *customer*. Pengumpulan data histori pendistribusian produk diperoleh dari tabel-tabel BigQuery yang sudah disiapkan oleh perusahaan. Data *customer_id* dan *product_id* yang diberikan adalah data yang sudah dilakukan *masking* agar kerahasiaan data tetap terjaga.

3.3.1 Populasi dan Sampel

Populasi data yang diperoleh dari perusahaan adalah data distribusi produk di cabang farmasi Bandung, farmasi Banjarmasin, dan farmasi Samarinda. Farmasi Bandung mewakili cabang besar dengan 451 *customer*, farmasi Banjarmasin mewakili cabang sedang dengan 241 *customer*, dan farmasi Samarinda mewakili cabang kecil dengan 185 *customer*. Ketiga cabang

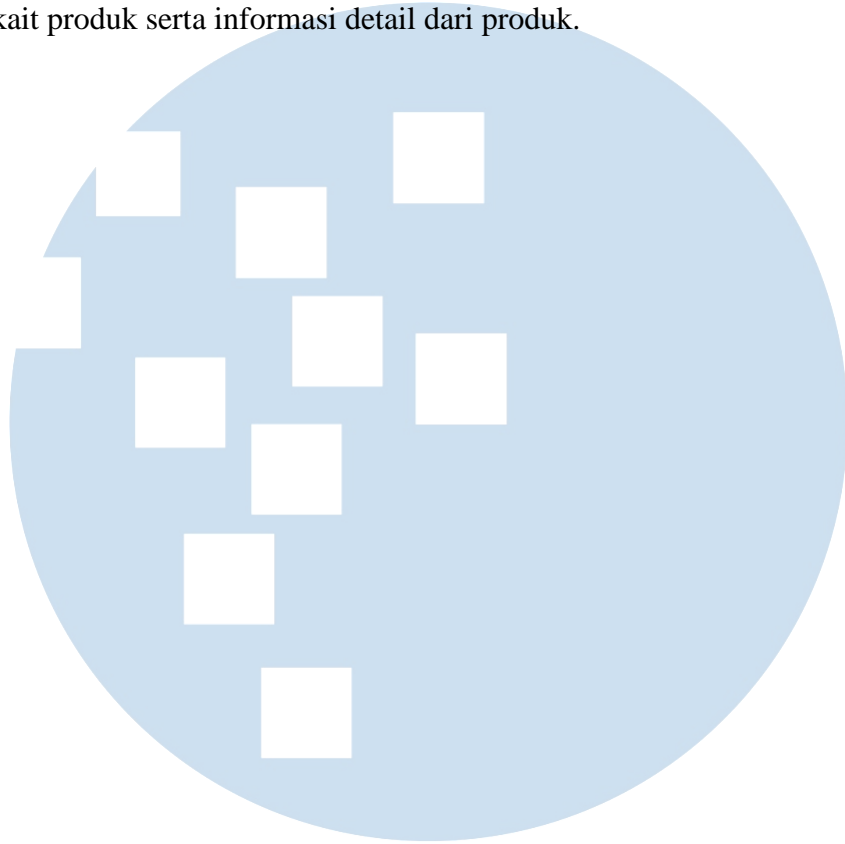
tersebut memiliki total 877 *customer* dan 409 produk. Proses pengambilan sampel pada penelitian ini dilakukan dengan teknik *purposive sampling*. *Purposive sampling* merupakan teknik pengambilan data dari keseluruhan populasi berdasarkan kriteria yang sudah ditetapkan oleh peneliti [44].

Sampel data yang dipilih adalah data distribusi produk kepada customer yang melakukan pemesanan minimal 5 produk selama 3 bulan terakhir. Pemilihan tersebut berdasarkan hasil penelitian yang pernah dilakukan dimana heavy customer atau customer yang memiliki purchasing quantity yang tinggi dapat memperoleh akurasi yang lebih tinggi [45]. Selain itu customer dengan pemesanan kurang dari 5 produk selama 3 bulan terakhir dianggap belum memiliki nilai yang cukup untuk diikuti sertakan pada model collaborative filtering yang dibuat. Sampel dari data yang akan diuji pada model rekomendasi adalah 397 produk serta 526 customer dimana 259 customer berasal dari farmasi Bandung, 183 customer berasal dari farmasi Banjarmasin, dan 84 customer berasal dari farmasi Samarinda.

3.3.2 Periode Pengambilan Data

Proses pengambilan data histori distribusi produk dilakukan pada tools BigQuery. Data histori distribusi produk yang digunakan pada penelitian ini adalah data distribusi di cabang farmasi Bandung, farmasi Banjarmasin, dan farmasi Samarinda selama 3 bulan yang dimulai dari tanggal 1 Oktober 2022 hingga 31 Desember 2022. Data yang diambil hanya berasal dari ketiga cabang tersebut dikarenakan setiap cabang tersebut sudah bisa mempresentasikan cabang besar yaitu farmasi Bandung, cabang sedang yaitu farmasi Banjarmasin, dan cabang kecil yaitu farmasi Samarinda. Selain itu, data yang digunakan hanya 3 bulan dikarenakan data 3 bulan sudah memiliki jumlah transaksi yang cukup untuk dikomparasikan yaitu sebanyak 8645 transaksi. Selain itu, penggunaan data lebih dari 3 bulan akan memerlukan waktu komputasi yang cukup lama mengingat fokus utama pada penelitian ada melakukan komparasi pada 7 model sistem rekomendasi. Tabel data histori distribusi produk yang diambil dari BigQuery tidak mengikut sertakan

informasi tanggal sebuah produk dipesan oleh customer serta informasi detail terkait produk serta informasi detail dari produk.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA