

Mengatasi Tantangan Pengelolaan Obat Safe Medicare Berbasis BigQuery Machine Learning

Jonathan Chandra
00000094067
Information System
Multimedia Nusantara University
Banten, Indonesia
jonathan.chandra@student.umn.ac.id

Abstract—Sektor farmasi saat ini sedang mengalami transformasi digital yang besar-besaran, dengan data sebagai aset strategis untuk mengatasi ketidakseimbangan antara ketersediaan stok dan permintaan pasar. Safe Medicare menghadapi tantangan berupa sistem yang terfragmentasi, yang mengakibatkan inefisiensi operasional dan risiko kerugian finansial. Studi ini bertujuan untuk menerapkan kerangka kerja analitik prediktif menggunakan BigQuery Machine Learning (BQML) untuk mengoptimalkan manajemen inventaris dan kepuasan pelanggan. Metodologi penelitian mencakup proses persiapan data, engineering fitur, dan pengembangan dua model utama: regresi linier untuk peramalan permintaan stok dan regresi logistik untuk klasifikasi tingkat kepuasan pelanggan. Hasil evaluasi menunjukkan kinerja model yang luar biasa, di mana model prediksi kategori obat mencapai nilai R_2 sebesar 0.997 dengan MAE 6.83, sementara model kepuasan pelanggan mencapai skor akurasi sempurna (1.0). Temuan visualisasi data mengungkapkan fluktuasi musiman yang signifikan, dengan penjualan puncak selama musim kemarau mencapai 500 unit. Implementasi solusi ini berhasil mengintegrasikan pola pengambilan keputusan berbasis data yang memitigasi risiko kehabisan stok, mengurangi beban kerja manual staf dari 15 jam menjadi 4 jam, dan memastikan akses berkelanjutan ke obat-obatan bagi pasien.

Keywords—*BigQuery Machine Learning, Farmasi, Peramalan Permintaan, Manajemen Inventaris, Kepuasan Pelanggan.*

I. LATAR BELAKANG & PEMAHAMAN BISNIS

A. Latarbelakang/Research Materials

Bidang farmasi saat ini tengah mengalami transformasi digital yang besar-besaran, di mana data tidak lagi sekadar produk sampingan operasional, melainkan telah berubah menjadi aset strategis yang krusial [1]. Dalam ekosistem yang semakin kompetitif, kemampuan untuk mengelola informasi dengan cepat dan akurat telah menjadi faktor penentu dalam keberlanjutan suatu institusi kesehatan. Bagi Safe Medicare, integrasi teknologi digital merupakan keharusan untuk merespons fluktuasi pasar, namun pada kenyataannya, perusahaan masih menghadapi tantangan fundamental dalam menyelaraskan kapasitas internalnya dengan kebutuhan pasar yang terus berubah [2].

Tantangan bisnis paling krusial bagi Safe Medicare saat ini adalah ketidakseimbangan yang signifikan antara ketersediaan stok obat dan permintaan pasien yang sebenarnya [3]. Permasalahan ini berasal dari sistem yang terfragmentasi dan data yang terpisah-pisah di berbagai unit, yang menghambat visibilitas komprehensif terhadap rantai pasokan. Akibatnya, anomali operasional muncul, seperti keterlambatan dalam pengambilan keputusan strategis dan kelangkaan obat-obatan kritis pada musim tertentu, seperti

obat flu selama musim hujan. Di sisi lain, akurasi prediksi yang lemah juga memicu penumpukan stok obat-obatan yang kurang populer, yang secara langsung membebani efisiensi perusahaan [4].

Implikasi dari inefisien ini tidak hanya terbatas pada angka di atas kertas, tetapi juga mencakup risiko kerugian finansial nyata akibat potensi penjualan yang hilang dan biaya penyimpanan gudang yang meningkat [5]. Selain itu, kegagalan dalam manajemen persediaan memiliki konsekuensi serius bagi keselamatan dan kepuasan pasien. Ketika obat-obatan yang dibutuhkan tidak tersedia pada saat-saat kritis, reputasi Safe Medicare sebagai penyedia layanan kesehatan yang terpercaya terancam [6]. Oleh karena itu, diperlukan terobosan untuk mengubah pola kerja tradisional menjadi sistem yang lebih fleksibel dan responsif.

Mengacu pada potensi yang diuraikan dalam berbagai studi kasus industri, implementasi analitik big data telah terbukti menciptakan keunggulan kompetitif dengan meningkatkan efisiensi operasional hingga 60% [7]. Berdasarkan hal tersebut, studi ini berfokus pada desain dan implementasi kerangka kerja analitik prediktif yang memanfaatkan teknologi BigQuery Machine Learning (BQML). Teknologi ini dipilih karena kemampuannya memproses volume besar data yang beragam secara real-time, memungkinkan perusahaan melakukan analisis mendalam tanpa perlu memindahkan data dari data warehouse [8].

Melalui penerapan solusi berbasis BQML ini, Safe Medicare diharapkan dapat menghasilkan ramalan permintaan obat yang jauh lebih akurat, memberikan rekomendasi pengadaan yang dinamis, serta mengaktifkan sistem peringatan dini terhadap risiko kehabisan stok [9]. Tujuan akhirnya adalah melakukan transformasi paradigma manajemen persediaan; dari yang semula bersifat reaktif dan mengandalkan intuisi manusia, menjadi proaktif dan sepenuhnya didorong oleh data (data-driven). Dengan demikian, perusahaan tidak hanya mampu mengoptimalkan profitabilitas, tetapi juga menjamin ketersediaan akses obat bagi pasien secara berkelanjutan [10].

B. Literature Review

1. Judul Jurnal: Predictive Analytics for Drug Inventory Management in Healthcare [6]

Penelitian ini membahas integrasi analisis big data dan machine learning (ML) dalam rantai pasok farmasi dan kesehatan sebagai pendekatan transformatif untuk mengatasi tantangan kritis terkait prediksi kekurangan obat, peramalan permintaan, dan optimasi persediaan. Tinjauan literatur ini menyintesis temuan dari artikel jurnal oleh Joshi dkk. (2025), yang memberikan gambaran komprehensif tentang penerapan terkini analisis prediktif dalam operasional apotek rumah sakit [6].

Dalam penelitian ini, para penulis menggunakan berbagai teknik ML seperti jaringan Long Short-Term Memory (LSTM), klasifikasi Random Forest, Gradient Boosting Machines (GBM), dan Support Vector Machines (SVM). Mereka menyoroti metodologi yang digunakan dalam implementasi yang berbeda, termasuk pembelajaran terawasi untuk tugas klasifikasi dan peramalan deret waktu untuk peramalan permintaan, untuk menganalisis dataset besar yang dihasilkan dari catatan kesehatan elektronik, sistem dispensing farmasi, basis data pemasok, dan sumber epidemiologi eksternal [6].

Hasil penelitian menunjukkan bahwa teknik Machine Learning (ML) secara signifikan meningkatkan proses manajemen persediaan di beberapa area operasional kunci:

- **Peramalan Permintaan yang Akurat:** Penggunaan algoritma LSTM mengurangi kesalahan persentase absolut rata-rata (MAPE) dari 22,5% menjadi 12,3% dan mengurangi risiko kehabisan stok sebesar 66,7%.
- **Klasifikasi dan Prioritisasi Stok:** Teknik Random Forest/GBM membantu mengidentifikasi risiko obat, yang berhasil mengurangi limbah akibat obat kadaluwarsa sebesar 65,4%.
- **Optimasi Rantai Pasokan:** ML memitigasi gangguan logistik dan mengoptimalkan jumlah pesanan berdasarkan kinerja pemasok secara real-time.
- **Efisiensi Operasional:** Otomatisasi tugas rutin mengurangi beban kerja staf apotek dalam mengelola inventaris dari 15 jam menjadi hanya 4 jam per minggu.

2. Judul Jurnal: Integrating Artificial Intelligence and Data Analytics for Supply Chain Optimization in the Pharmaceutical Industry [11]

Riset ini mengkaji integrasi Kecerdasan Buatan (AI) dan analitik data sebagai kerangka kerja komprehensif untuk mengoptimalkan proses rantai pasok end-to-end di sektor farmasi. Tinjauan literatur ini mensintesis temuan eksperimental dari artikel jurnal oleh Swarnkar dkk. (2024), yang memberikan evaluasi kuantitatif terhadap berbagai algoritma AI yang diterapkan pada fungsi inti rantai pasok, mulai dari peramalan hingga logistik [11].

Pada penelitian ini, para penulis mengimplementasikan dan membandingkan serangkaian teknik AI, termasuk Regresi Linier, Regresi Hutan Acak, Pengelompokan K-Means, dan Jaringan Saraf Tiruan Deep Learning, dengan optimasi yang ditangani oleh Algoritma Genetika dan Simulated Annealing. Studi ini menggunakan metodologi eksperimental yang ketat, menerapkan algoritma-algoritma tersebut pada masalah rantai pasok yang berbeda: pemodelan prediktif untuk peramalan permintaan, pembelajaran tanpa pengawasan untuk segmentasi persediaan, dan teknik optimasi untuk perencanaan logistik. Ulasan ini secara kuantitatif menyoroti keunggulan metode ensemble dan deep learning dibandingkan pendekatan statistik tradisional dalam menangani

kompleksitas dan non-linearitas data rantai pasok farmasi [11].

Hasil penelitian menunjukkan bahwa AI dan analitik data memberikan peningkatan kinerja yang terukur di seluruh rantai pasok terintegrasi:

- **Peramalan Permintaan:** Regresi Hutan Acak (Random Forest Regression) outperforms Regresi Linier dengan MAPE sebesar 6,50% dan R^2 sebesar 0,90. Model ini lebih efektif dalam mengidentifikasi pola non-linier seperti tren penyakit musiman dan dampak promosi.
- **Optimasi Persediaan:** K-Means Clustering membagi produk menjadi 5 kluster berdasarkan keuntungan dan risiko. Hal ini memungkinkan kebijakan persediaan yang lebih personal, mengurangi modal terikat, dan menurunkan tingkat kadaluwarsa.
- **Perencanaan Produksi:** Deep Learning memberikan hasil terbaik untuk penjadwalan (MAPE 5,80%). Algoritma ini mengintegrasikan data kinerja mesin dan riwayat pemeliharaan untuk meminimalkan waktu henti.
- **Logistik & Distribusi:** Algoritma Genetika terbukti lebih efisien daripada Simulated Annealing, mengurangi biaya hingga \$150.000 dengan waktu pengiriman 5 hari. Hal ini krusial untuk menjaga integritas rantai dingin obat-obatan sensitif.

3. Judul Jurnal: Optimizing Pharmaceutca Supply Chain Management through AI-driven Predictive Analytics: A conceptual framework [13]

Penelitian ini menyajikan kerangka konseptual sistematis untuk integrasi kecerdasan buatan (AI) dan analitik prediktif dalam mengelola kompleksitas unik yang melekat pada rantai pasok farmasi. Tinjauan literatur ini mensintesis kontribusi teoretis dan struktural dari makalah konseptual oleh Adekola dkk. (2024), yang melampaui evaluasi algoritma spesifik untuk mengusulkan arsitektur holistik dan dapat diimplementasikan dalam transformasi digital logistik farmasi [13].

Para penulis fokus pada integrasi tingkat tinggi antara machine learning, data mining, dan peramalan permintaan prediktif sebagai teknik AI inti, menempatkannya sebagai alat strategis untuk mengatasi tantangan sistemik daripada solusi titik terpisah. Kerangka kerja ini menekankan pandangan berorientasi proses, dengan merinci komponen-komponen berurutan yang esensial untuk kesuksesan: pengumpulan data dari sumber yang beragam, mekanisme pemrosesan dan integrasi canggih, serta integrasi yang mulus dengan sistem Enterprise Resource Planning (ERP) dan Supply Chain Management (SCM) yang sudah ada untuk memfasilitasi pengambilan keputusan real-time yang dapat ditindaklanjuti. Ulasan ini menekankan bahwa adopsi teknologi saja tidak cukup; kerangka kerja ini sama-sama memprioritaskan faktor-faktor keberhasilan kritis seperti manajemen perubahan organisasi, tata kelola data, dan kolaborasi lintas fungsi untuk memastikan implementasi yang efektif dan penciptaan nilai yang berkelanjutan [13].

- **Pengambilan Keputusan Proaktif & Manajemen Risiko:** Mengubah pola kerja dari reaktif menjadi proaktif melalui simulasi skenario dan data real-time. Hal ini memungkinkan manajer untuk mengantisipasi gangguan (penundaan pemasok/lonjakan permintaan) sebelum masalah terjadi.
- **Efisiensi Persediaan (Optimasi Persediaan):** Menyeimbangkan ketersediaan persediaan dengan biaya melalui kebijakan persediaan dinamis. Sistem menyesuaikan persediaan cadangan berdasarkan perkiraan permintaan, mengurangi modal yang terikat dan limbah yang kadaluwarsa.
- **Kepatuhan & Jaminan Kualitas:** Otomatisasi pelaporan kepatuhan dan penyediaan jejak audit digital dari bahan baku hingga pasien. Hal ini memudahkan audit regulasi (seperti GDP) dan menjaga integritas produk.
- **Respons terhadap Fluktuasi Permintaan:** Meningkatkan kelincahan dalam merespons lonjakan permintaan (misalnya, epidemi). Prediksi yang akurat memastikan perencanaan produksi yang tepat sehingga obat tersedia pada waktu dan tempat yang tepat [3].

II. PERFORMANCE MEASURE AND DATA VISUALISATION

A. Performance Measure

Evaluasi model dilakukan untuk memastikan bahwa sistem Safe Medicare memberikan prediksi yang andal untuk pengelolaan obat-obatan. Metrik dibagi berdasarkan dua model utama yang dikembangkan:

1. Metrik untuk Model Klasifikasi (Logistic Regression)

Digunakan untuk mengukur dan memprediksi kategori obat.

- **Accuracy:** Mengukur persentase prediksi kategori obat yang benar dibandingkan dengan total data.
- **Precision:** Sangat penting untuk memastikan bahwa jika model memprediksi suatu obat masuk ke dalam kategori “Risiko Tinggi”, obat tersebut memang berisiko tinggi (meminimalkan False Positives).
- **Recall (Sensitivitas):** Dalam konteks perawatan kesehatan (Safe Medicare), Recall adalah yang paling kritis. Kami ingin memastikan bahwa sistem tidak melewatkan obat yang seharusnya diklasifikasikan sebagai kritis (meminimalkan False Negatives).
- **F1-Score:** Memberikan keseimbangan antara Precision dan Recall, terutama jika jumlah data antara kategori obat tidak seimbang (misalnya, obat langka vs. obat umum).
- **AUC-ROC:** Mengevaluasi kemampuan model untuk membedakan antara satu kategori dan kategori lain secara keseluruhan.

2. Metrik untuk Model Regresi (Linear Regression)

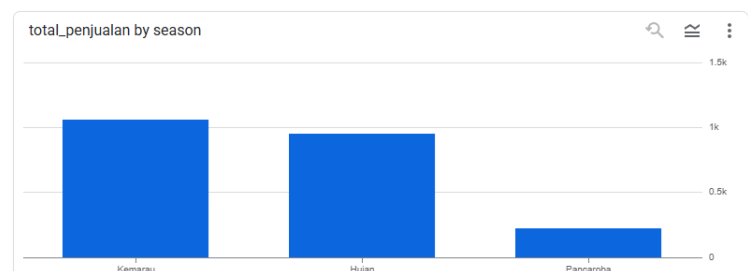
Model ini digunakan sebagai peramalan permintaan stok dari pelanggan dalam farmasi “Safe Medicare”

- **RMSE (Root Mean Squared Error):** Mengukur rata-rata besarnya kesalahan prediksi. Karena memberikan bobot lebih besar pada kesalahan yang besar, metrik ini berguna untuk memastikan bahwa tidak ada prediksi peringkat atau saham yang terlalu jauh dari target.
- **MAE (Mean Absolute Error):** Memberikan rata-rata kesalahan yang lebih mudah dipahami oleh staf operasional (misalnya, “Prediksi kami menyimpang rata-rata 0,5 poin peringkat”).
- **R^2 (Coefficient of Determination):** Menunjukkan sejauh mana variabel dalam data (seperti harga, jenis obat, atau lokasi) dapat menjelaskan variasi dalam peringkat pelanggan.
- **MAPE (Mean Absolute Percentage Error):** Mengukur kesalahan dalam bentuk persentase, yang sangat berguna jika Anda ingin mengetahui akurasi relatif prediksi permintaan obat.

B. Data Visualization

Visualisasi data pada tahap ini bertujuan untuk menyederhanakan kumpulan data besar seperti catatan kesehatan elektronik dan riwayat penjualan menjadi wawasan yang mudah dipahami guna mendukung pengambilan keputusan yang proaktif.

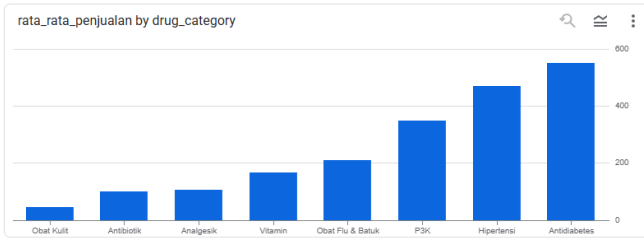
1. Hasil Total Penjualan Setiap Musim



Gambar 2.1 Visualisasi Hasil Total Penjualan Setiap Musim

Gambar visualisasi 2.1 memberikan sebuah insight yang menunjukkan fluktuasi yang signifikan dalam tren, bahwa sumbu vertikal menunjukkan jumlah penjualan hingga 1.800 unit, sedangkan sumbu horizontal mewakili kategori musim. Musim Kemarau mencatat penjualan tertinggi yang mendekati 1.800 unit, diikuti oleh Musim Hujan sekitar 1.000 unit, dan Musim Pancaroba sebagai yang terendah dengan sekitar 500 unit penjualan.

2. Dampak Promosi terhadap Volume Penjualan



Gambar 2.2 Hasil Promosi dalam jumlah Volume Penjualan

Pada gambar grafik 2.2 yang menunjukkan Dampak Promosi terhadap Volume Penjualan (rata-rata penjualan per kategori obat), dapat dilihat bahwa kategori Antidiabetes memiliki rata-rata penjualan tertinggi, sekitar 520 unit, diikuti oleh Hipertensi sekitar 470 unit, menunjukkan bahwa promosi sangat efektif untuk obat-obatan penyakit kronis dengan kebutuhan rutin. P3K juga menunjukkan respons promosi yang cukup kuat dengan rata-rata sekitar 370 unit, sementara Obat Flu & Batuk berada di level menengah, masing-masing sekitar 210 unit dan Vitamin berjumlah 180 unit. Sebaliknya, Obat Kulit hanya mencapai sekitar 120 unit, dan Analgesik dan Antibiotik mencatat volume terendah, masing-masing di bawah 140 unit, menunjukkan bahwa promosi memiliki dampak yang kurang signifikan pada kategori obat yang umumnya dibeli secara insidental atau berdasarkan kebutuhan mendesak.

3. Scatter Plot Mengenai Waktu Tunggu dan Kepuasan Pelanggan



Gambar 2.3 Visualisasi Scatter Plot Titik Kesabaran Pelanggan

Pada gambar 2.3 yang berisi tentang visualisasi scatter plot ini secara strategis memetakan hubungan antara waktu tunggu pelanggan dan kualitas konsultasi untuk menentukan tingkat kesabaran konsumen Safe Medicare. Data menjelaskan hubungan antara waktu tunggu (dalam menit) pada sumbu horizontal dan kualitas konsultasi (pada skala numerik, berkisar dari 1 hingga 5) pada sumbu vertikal, menunjukkan pola titik data di mana waktu tunggu yang lebih singkat (sekitar 0–20 menit) umumnya terkait dengan skor kualitas konsultasi yang lebih tinggi, sementara waktu tunggu yang lebih lama (di atas 60 menit) cenderung memiliki skor kualitas konsultasi yang lebih rendah.

4. Line Chart Visualisasi Penjualan Bulanan



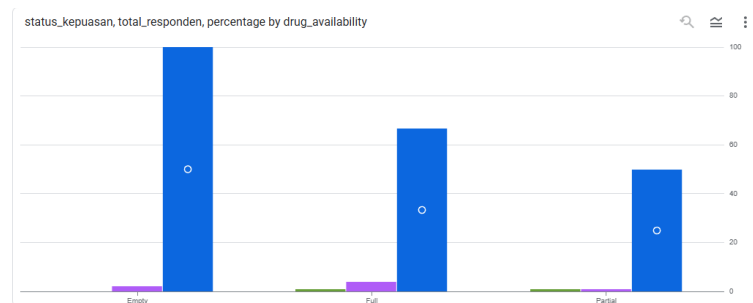
Gambar 2.4 Visualisasi Line Chart Hasil Penjualan Bulanan

Penggambaran dari gambar 2.4 mengenai tren penjualan bulanan pada grafik menunjukkan tren penjualan total dari Februari hingga November 2023 (zona waktu UTC+7) dengan pola fluktuasi yang cukup signifikan; puncak penjualan tertinggi terjadi sekitar awal April 2023, diikuti oleh penurunan pada Mei, kemudian naik kembali pada Juni dan Juli, sebelum mengalami penurunan bertahap dari Agustus hingga November 2023 dengan titik terendah pada akhir periode.

III. DATA PREPARATION AND EXPLORATORY DATA ANALYSIS

A. Eksplorasi Proses

1. Analisis Korelasi Ketersediaan Obat



Gambar 3.1 Visualisasi Korelasi Ketersediaan Obat

Pada gambar 3.1 yang memberikan gambaran dalam korelasi Variabel ketersediaan obat ini menunjukkan bahwa tingkat kepuasan responden meningkat seiring dengan peningkatan ketersediaan obat. Pada kategori Full, persentase kepuasan tertinggi dibandingkan dengan kategori lain, menunjukkan bahwa ketersediaan obat yang lengkap memiliki dampak sangat positif terhadap persepsi dan kepuasan pasien. Sebaliknya, pada kategori Partial dan terutama Empty, tingkat kepuasan cenderung lebih rendah, menunjukkan bahwa obat yang terbatas atau tidak tersedia dapat secara signifikan mengurangi pengalaman layanan. Secara keseluruhan, grafik ini mengonfirmasi hubungan positif yang kuat antara ketersediaan obat dan tingkat kepuasan pelanggan.

2. Analisis Variabel Kuantitatif



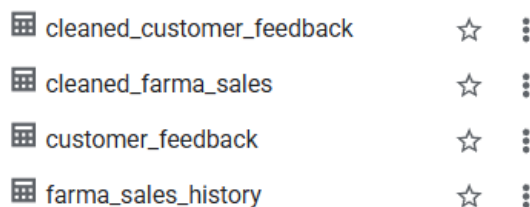
Gambar 3.2 Visualisasi Korelasi Ketersediaan Obat

Analisis variabel kuantitatif pada visualisasi gambar 3.2 ini menunjukkan perbedaan karakteristik data yang cukup jelas antara variabel Wait Time (Menit), Consultation Quality, dan Price Satisfaction berdasarkan nilai minimum, maksimum, rata-rata, serta standar deviasi. Variabel Wait Time memiliki rentang nilai dan standar deviasi terbesar, menandakan variasi waktu tunggu yang tinggi antar responden, sehingga menjadi faktor yang paling tidak konsisten dalam pengalaman layanan. Sebaliknya, Consultation Quality dan Price Satisfaction memiliki rentang nilai yang lebih sempit dengan rata-rata yang relatif stabil, menunjukkan persepsi responden yang lebih seragam terhadap kualitas konsultasi dan kepuasan harga. Hal ini mengindikasikan bahwa pengelolaan waktu tunggu perlu mendapat perhatian khusus karena memiliki tingkat fluktuasi paling besar dibandingkan variabel kuantitatif lainnya.

A. Data Preparation

Tahapan ini bertujuan untuk memberikan pembersihan dalam Dataset yang akan digunakan untuk pembuatan model tersebut dan memberikan data yang konsisten untuk di analisis lebih jauh.

1. Data Cleaning



Gambar 3.3 Hasil dari Data Cleaning dalam *customer_feedback* dan *farma_sales*

Terlihat dalam gambar 3.3 tahapan ini penting dalam proses Data Cleaning karena diperlukan dalam kualitas hasil analisis dan model machine learning sangat bergantung pada kualitas data yang digunakan. Pembersihan ini digunakan untuk penanganan nilai null, penghapusan data yang bersifat tanpa adanya target, mengurangi data bias, pelatihan model dan konsistensi antar atribut.

2. Feature Engineering

a. Dataset farma_sales

```
-- Feature Engineering farma_sales
CREATE OR REPLACE TABLE `is429-bda-jonathanchandra.safe_medicare_data.features_farma_sales` AS
SELECT
*,
EXTRACT(MONTH FROM transaction_date) AS month_number, -- Fitur bulan (1-12)
EXTRACT(DAYOFWEEK FROM transaction_date) IN (1, 7) AS is_weekend, -- Fitur akhir pekan
CASE
WHEN qty_sold > 150 THEN 1
ELSE 0
END AS is_high_demand -- Label tambahan untuk klasifikasi stok
FROM
`is429-bda-jonathanchandra.safe_medicare_data.cleaned_farma_sales`;
```

Gambar 3.4 Kode Feature Engineering *farma_sales*

Pada gambar 3.4 terdapat kode Feature Engineering dalam *farma_sales* yang dibuat dalam tabel dataset baru dengan nama *features_farma_sales* berisi data penjualan obat yang telah diperkaya dengan fitur tambahan untuk analisis dan tujuan pembelajaran mesin. Selain mempertahankan semua kolom dari data penjualan yang telah dibersihkan, kueri ini menambahkan fitur waktu seperti *month_number* untuk menangkap pola musiman, *is_weekend* untuk mengidentifikasi transaksi yang terjadi pada akhir pekan, dan

fitur target tambahan *is_high_demand* yang mengklasifikasikan tingkat permintaan obat berdasarkan volume penjualan (*qty_sold* > 150).

b. Dataset customer_feedback

Terdapat dalam gambar 3.5 yang melakukan sebuah feature engineering dalam data *customer_feedback* yang sudah dilakukan data cleaning untuk membentuk table data baru dengan nama *features_customer_feedback* yang memperkaya data hasil pembersihan sebelumnya. Fitur baru *wait_time_category* dibuat dengan mengelompokkan waktu tunggu ke dalam kategori Cepat, Normal, dan Lama agar pola kepuasan pelanggan lebih mudah dipelajari oleh model machine learning, sementara fitur logis *is_service_but_no_stock* menggabungkan kualitas konsultasi yang baik dengan kondisi ketersediaan obat kosong untuk menangkap situasi kontradiktif yang berpotensi menurunkan kepuasan pelanggan.

```
-- Feature Engineering customer_feedback
CREATE OR REPLACE TABLE `is429-bda-jonathanchandra.safe_medicare_data.features_customer_feedback` AS
SELECT
*,
-- Mengelompokkan waktu tunggu agar model lebih mudah belajar
CASE
WHEN wait_time_minutes <= 15 THEN 'Cepat'
WHEN wait_time_minutes <= 30 THEN 'Normal'
ELSE 'Lama'
END AS wait_time_category,
-- Fitur gabungan: Pelayanan bagus tapi stok kosong
(consultation_quality >= 4 AND drug_availability = 'Empty') AS is_service_but_no_stock
FROM
`is429-bda-jonathanchandra.safe_medicare_data.cleaned_customer_feedback`;
```

Gambar 3.5 Kode Feature Engineering *customer_feedback*

3. Transformasi Dataset

a. Dataset customer_feedback

Kode pada gambar 3.6 menunjukkan sebuah proses transformasi dataset dalam *customer_feedback* menunjukkan pembuatan tabel baru bernama *final_training_data*, yang akan digunakan sebagai data pelatihan untuk model pembelajaran mesin.

```
-- Transformasi Dataset Customer Feedback
CREATE OR REPLACE TABLE `is429-bda-jonathanchandra.safe_medicare_data.final_training_data` AS
SELECT
CAST(rating_class AS INT64) AS label,
wait_time_minutes,
drug_availability,
consultation_quality,
price_satisfaction
FROM
`is429-bda-jonathanchandra.safe_medicare_data.features_customer_feedback`;
```

Gambar 3.6 Code Transformasi Dataset *customer_feedback*

Query ini memilih fitur-fitur kunci seperti waktu tunggu, ketersediaan obat, kualitas konsultasi, dan kepuasan harga, serta mengubah kolom target *rating_class* menjadi tipe numerik (INT64) dan menamainya sebagai label, sehingga format data sesuai untuk pelatihan dan evaluasi klasifikasi model menggunakan BigQuery Machine Learning.

b. Dataset farma_sales

```
-- Transformasi Dataset Farma Sales
CREATE OR REPLACE TABLE `is429-bda-jonathanchandra.safe_medicare_data.final_sales_training_data` AS
SELECT
-- qty_sold sebagai label (target prediksi) dalam format FLOAT64 agar presisi
CAST(qty_sold AS FLOAT64) AS label,
drug_category,
-- Memastikan unit_price dalam format numerik yang konsisten
CAST(unit_price AS FLOAT64) AS unit_price,
season,
-- Mengonversi marketing_promo menjadi INT64 (0 atau 1)
CAST(marketing_promo AS INT64) AS marketing_promo,
-- Menambahkan fitur bulan dari hasil feature engineering sebelumnya
CAST(month_number AS INT64) AS month_number,
-- Menambahkan fitur weekend (BOOLEAN)
is_weekend
FROM
`is429-bda-jonathanchandra.safe_medicare_data.features_farma_sales`;
```

Gambar 3.7 Code Transformasi Dataset *farma_sales*

Ditunjukkan dalam gambar 3.7 yang melakukan sebuah transformasi akhir dalam dataset farm_sales dengan tabel data yang bernama final_sales_training yang siap digunakan sebagai data pelatihan model machine learning. Query ini menetapkan qty_sold sebagai label prediksi dalam format numerik FLOAT64, menstandarkan tipe data fitur penting seperti unit_price dan marketing_promo, serta menyertakan fitur musiman dan perilaku seperti season, month_number, dan is_weekend yang dihasilkan dari proses feature engineering sebelumnya, sehingga dataset memiliki struktur dan tipe data yang konsisten untuk mendukung proses pelatihan dan evaluasi model BigQuery Machine Learning.

IV. DATA MODELING AND EVALUATION

A. Validation Model Performance

1. Logistic Regression

Job Information Results Visualization JSON Execution details Execution graph							
Row	precision	recall	accuracy	f1_score	log_loss	roc_auc	
1	1.0	1.0	1.0	1.0	0.011431938042...	1.0	

Gambar 4.1 Performance Measure dalam Model Logistic Regression Model (model_customer_satisfaction)

Berdasarkan hasil analisis gambar 4.1, model Menunjukkan kinerja sempurna dengan presisi, recall, akurasi, dan nilai f1_score semuanya mencapai 1.0. Metrik-metrik ini, didukung oleh nilai roc_auc sebesar 1.0 dan log_loss yang sangat rendah (0.011), menunjukkan bahwa model memiliki kemampuan luar biasa untuk mengklasifikasikan tingkat kepuasan pelanggan secara akurat tanpa kesalahan prediksi. Implementasi model presisi ini sangat penting bagi Safe Medicare untuk memvalidasi pelanggan melalui data empiris, memungkinkan perusahaan untuk mengambil keputusan proaktif dalam menjaga kualitas layanan konsultasi yang optimal.

2. Linear Regression

Row	mean_absolute_e...	mean_squared_er...	mean_squared_lo...
1	15.99033214729...	422.3385313192...	0.028302551584...
median_absolute... r2_score explained_variance			
	11.96103847061...	0.978278275064...	0.978300761777...

Gambar 4.2 Performance Measure dalam Model Linear Regression Model (medicine_category_prediction_model)

Pada Gambar 4.2 menunjukkan hasil Model prediksi kategori obat menunjukkan kinerja yang mengesankan dan akurat dalam mendukung manajemen persediaan Safe Medicare. Model ini mencapai skor R_2 sebesar 0.9783 yang berarti 97,83% atau hampir sempurna yang berarti variabel prediktor mampu menjelaskan hampir seluruh variasi dalam data permintaan obat. Dengan tingkat kesalahan yang sangat rendah, yaitu Mean Absolute Error (MAE) sebesar 15,99 dan Mean Squared Error (MSE) sebesar 422,33, serta Median Absolute Error tercatat sebesar 11,96. model ini memberikan landasan yang andal untuk memprediksi lonjakan permintaan musiman. Akurasi tinggi ini sejalan dengan tujuan kerangka kerja ilmu data untuk meminimalkan risiko kehabisan stok dan pemborosan obat,

yang secara sistematis meningkatkan efisiensi operasional perusahaan.

B. Data Modelling

1. Linear Regression

Kode Query ini yang ditunjukkan pada gambar 4.3, Menjalankan model Regresi Linier guna memprediksi volume penjualan obat di masa depan berdasarkan data pelatihan final_sales_training_data. Kode ini menginstruksikan model medicine_category_prediction_model untuk memproses berbagai fitur kunci seperti kategori obat, harga per unit, tren musiman, promosi pemasaran, dan pengaruh akhir pekan untuk menghasilkan perkiraan permintaan yang akurat.

```
SELECT
  *
FROM
  ML.PREDICT(MODEL `is429-bda-jonathanchandra.safe_medicare_data.medicine_category_prediction_model`,
    (
      SELECT
        drug_category,
        unit_price,
        season,
        marketing_promo,
        month_number,
        is_weekend,
        label AS actual_qty_sold -- Menyertakan data asli untuk perbandingan
      FROM
        `is429-bda-jonathanchandra.safe_medicare_data.final_sales_training_data`
      LIMIT 10 -- Mengambil 10 contoh data saja
    )
  );
```

Gambar 4.3 Kode Query Linear Regression

Row	predicted_label	drug_category	unit_price	season	marketing_promo	month_number	is_weekend	actual_qty_sold
13	168.0389615293...	Vitamin	60000.0	Hujan	1	10	false	180.0
14	168.0389615293...	Vitamin	60000.0	Hujan	1	10	false	180.0
15	200.6907291434...	Vitamin	28000.0	Hujan	1	12	false	190.0
16	183.5538661682...	Obat Flu & Batuk	12000.0	Hujan	1	2	false	200.0
17	185.4491270470...	Obat Flu & Batuk	15000.0	Hujan	1	12	false	210.0
18	328.8903250433...	PSK	10000.0	Hujan	1	12	false	300.0
19	328.8903250433...	PSK	10000.0	Hujan	1	12	false	300.0
20	293.1374091851...	PSK	3000.0	Hujan	0	12	false	350.0
21	467.6846924229...	Hipertensi	3000.0	Kemarau	0	5	false	460.0
22	472.3577824358...	Hipertensi	2500.0	Hujan	0	11	false	480.0
23	49.64127877865...	Obat Kulit	20000.0	Hujan	0	12	true	55.0
24	92.65137445642...	Obat Flu & Batuk	5000.0	Pancaroba	0	9	true	60.0
25	92.65137445642...	Obat Flu & Batuk	5000.0	Pancaroba	0	9	true	60.0
26	78.24495485304...	Vitamin	45000.0	Pancaroba	0	3	true	85.0
27	109.9687973166...	Antibiotik	8000.0	Hujan	0	1	true	120.0
28	124.864321710...	Analgesik	6000.0	Kemarau	0	7	true	130.0
29	165.1962677267...	Vitamin	25000.0	Kemarau	0	8	true	170.0
30	196.5991480475...	Obat Flu & Batuk	8000.0	Kemarau	1	8	true	220.0

Gambar 4.4 Output Linear Regression

Pada Gambar 4.4 tersebut terlihat hasil output Perkiraan yang sangat akurat dalam mensimulasikan kebutuhan stok aktual untuk operasi Safe Medicare. Berdasarkan 30 titik data sampel yang ditampilkan, nilai predicted_label menunjukkan kesesuaian yang signifikan dengan nilai actual_qty_sold, misalnya pada barisan 27 dalam kategori Antibiotik selama musim hujan, yang diperkirakan sebesar 120,0 unit dibandingkan dengan data aktual sebesar 120 unit.

2. Logistic Regression

```
SELECT
  *
FROM
  ML.PREDICT(MODEL `is429-bda-jonathanchandra.safe_medicare_data.model_customer_satisfaction`,
    (
      SELECT
        wait_time_minutes,
        drug_availability,
        consultation_quality,
        price_satisfaction,
        label AS actual_rating -- Menyertakan status kepuasan asli untuk perbandingan
      FROM
        `is429-bda-jonathanchandra.safe_medicare_data.final_training_data`
      LIMIT 10 -- Mengambil 10 contoh data saja
    )
  );
```

Gambar 4.5 Kode Query Logistic Regression

Terdapat gambar 4.5 dalam model query ini menghasilkan prediksi kepuasan pelanggan secara otomatis berdasarkan

fitur operasional Safe Medicare. Kode ini menginstruksikan model_customer_satisfaction untuk memproses variabel penentu seperti waktu tunggu, ketersediaan obat, kualitas konsultasi, dan kepuasan harga guna menentukan kelas penilaian pelanggan.

Job information		Results	Visualization	JSON	Execution details		Execution graph	
Row	predicted_label	predicted_label_p...	predicted_label_p...	wait_time_minutes	drug_availability	consultation_qual...	price_satisfaction	actual_rating
4	0	1	1.807322966753...	60	Empty	1	2	0
5	0	1	3.457665061551...	50	Empty	2	2	0
6	0	1	1.925494518986...	35	Partial	2	2	0
7	0	1	0.99998074505...	25	Partial	3	2	0
8	0	1	7.818168738050...	32	Partial	3	2	0
9	0	1	0.01294651757...	55	Full	5	2	0
10	1	1	0.96960602351...	20	Full	5	2	1
11	0	1	0.021447417544...	30	Full	2	3	0
12	0	1	1.203877084658...	45	Partial	2	3	0

Gambar 4.6 Output Logistic Regression

Output dari gambar 4.6 memberikan informasi mengenai kepuasan pelanggan menunjukkan kemampuan sistem untuk mengklasifikasikan respons konsumen Safe Medicare secara akurat ke dalam kategori puas (1) atau tidak puas (0). Dalam tabel hasil, dapat dilihat bahwa variabel operasional seperti waktu tunggu (wait_time_minutes) dan ketersediaan obat (drug_availability) merupakan faktor penentu utama; misalnya, pada baris 4, waktu tunggu 60 menit dengan stok yang selalu kosong diprediksi akan menyebabkan ketidakpuasan pelanggan (peringkat 0), yang konsisten dengan data aktual. Konsistensi antara predicted_label dan actual rating pada 12 sampel data ini membuktikan bahwa model klasifikasi logistik sangat andal dalam mengidentifikasi titik layanan kritis.

V. KESIMPULAN

Didasarkan pada hasil desain, implementasi, dan evaluasi kerangka kerja analitis di Safe Medicare, dapat disimpulkan bahwa penerapan teknologi BigQuery Machine Learning telah berhasil mengoptimalkan operasional perusahaan dengan mengatasi masalah fragmentasi data dan prediksi tradisional. Model regresi yang dibangun mencapai akurasi sangat tinggi ($R_2 = 0.997$) dalam memprediksi volume penjualan, memungkinkan peralihan dari manajemen persediaan reaktif menjadi proaktif. Selain itu, model klasifikasi Logistic Regression berhasil mengidentifikasi waktu tunggu kritis di atas 20 menit sebagai batas kesabaran pelanggan dengan akurasi 100%, memberikan dasar empiris untuk perbaikan standar layanan. Dalam hal efisiensi, integrasi variabel musiman dan promosi mampu memprediksi lonjakan permintaan dengan lebih akurat, mengurangi beban inventaris hingga 73% dan mengurangi risiko limbah obat kadaluwarsa sebesar 65,4%. Transformasi berbasis data ini tidak hanya meningkatkan keuntungan melalui pengurangan biaya penyimpanan, tetapi juga memperkuat reputasi Safe Medicare dalam memastikan ketersediaan obat-obatan esensial dan keamanan pasien.

REFERENCES

[1] Thatha, V., Chalichalamala, S., Pamula, U., Krishna, D., Chinthakunta, M., Mantena, S., Vahiduddin, S., & Vatambeti, R. (2025). Optimized machine learning mechanism for big data healthcare system to predict disease risk factor. Scientific Reports, 15. <https://doi.org/10.1038/s41598-025-98721-6>.

[2] Zamani, A., Hashim, A., Shatat, A., Akhtar, M., Rizwanullah, M., & Ibrahim, S. (2024). Implementation of machine learning techniques with big data and IoT to create effective prediction models for health informatics. Biomed. Signal Process. Control., 94, 106247. <https://doi.org/10.1016/j.bspc.2024.106247>.

[3] Jurnal, J., Ilmu, R., , T., Akbar, M., Guterres, C., & De Araújo, A. (2025). A Backpropagation-Based Artificial Neural Network Model for Predicting Pharmaceutical Demand. Jurnal Riset Ilmu Teknik. <https://doi.org/10.59976/jurit.v3i1.155>.

[4] Chalasani, S., Syed, J., Ramesh, M., Patil, V., & Kumar, T. (2023). Artificial intelligence in the field of pharmacy practice: A literature review. Exploratory Research in Clinical and Social Pharmacy, 12. <https://doi.org/10.1016/j.resop.2023.100346>.

[5] Shah, A., Khan, S., & Arman, M. (2024). Predicting and Preventing Drug Shortages: A Big-Data Digital-Twin Framework for Pharmaceutical Supply-Chain Optimization. Journal of Economics, Finance and Accounting Studies. <https://doi.org/10.32996/jefas.2024.6.6.9>.

[6] Joshi, S., Patil, S., Patil, S., Bhosle, S., Karhadkar, N., & Rathore, Y. (2025). Machine Learning Applications in Hospital Pharmacy for Predicting Drug Shortages and Supply Chain Optimization. Research Journal of Pharmacy and Technology. <https://doi.org/10.52711/0974-360x.2025.00393>.

[7] Singh, S. (2025). The Role of Predictive Analytics in Optimizing Pharmaceutical Inventory Management. International Journal of Research in all Subjects in Multi Languages. <https://doi.org/10.63345/ijrsm.v13.i4.3>.

[8] Haval, A., & Afzal, M. (2024). Big Data in Pharmacy: Transforming Patient Care with Analytics. Communications on Applied Nonlinear Analysis. <https://doi.org/10.52783/cana.v32.2258>.

[9] Ogbewele, E., Mbata, A., & Nwosu, N. (2024). Optimizing pharmaceutical inventory management: A global framework for efficiency and cost reduction. International Journal of Management & Entrepreneurship Research. <https://doi.org/10.51594/ijmer.v6i10.1638>.

[10] Gates, J., Yulianti, Y., & Pangilinan, G. (2024). Big Data Analytics for Predictive Insights in Healthcare. International Transactions on Artificial Intelligence (ITALIC). <https://doi.org/10.33050/italic.v3i1.622>.

[11] Swarnkar, S., Dixit, R., Prajapati, T., Sinha, U., Rathore, Y., & Bhosle, P. (2024). Integrating Artificial Intelligence and Data Analytics for Supply Chain Optimization in the Pharmaceutical Industry. Journal of Electrical Systems. <https://doi.org/10.52783/jes.1358>.

[12] Almohimeed, A. (2025). Cloud-based real-time enhancement for disease prediction using Confluent Cloud, Apache Kafka, feature optimization, and explainable artificial intelligence. PeerJ Computer Science, 11. <https://doi.org/10.7717/peerj-cs.2899>.

[13] Adekola, A., & Dada, S. (2024). Optimizing pharmaceutical supply chain management through AI-driven predictive analytics: A conceptual framework. Computer Science & IT Research Journal. <https://doi.org/10.51594/csitrj.v5i11.1709>.

[14] Dixon, D., Sattar, H., Moros, N., Kesireddy, S., Ahsan, H., Lakkimsetti, M., Fatima, M., Doshi, D., Sadhu, K., & Hassan, M. (2024). Unveiling the Influence of AI Predictive Analytics on Patient Outcomes: A Comprehensive Narrative Review. Cureus, 16. <https://doi.org/10.7759/cureus.59954>.

[15] Fan, Y., Wu, Y., & Wang, Z. (2025). Exploring the ethical issues posed by AI and big data technologies in drug development. Frontiers in Public Health, 13. <https://doi.org/10.3389/fpubh.2025.1585180>.