



# **Analisis Deteksi Anomali untuk Peningkatan Integritas Data pada Sistem Inventaris dan Transaksi Pembelian**

Laporan Ujian Tengah Semester

Disusun untuk Memenuhi Tugas Mata Kuliah **IF25-32025 - Penambangan Data**

**Disusun oleh:**

<b>Nama Penyusun</b>		<b>NIM</b>
Fiqri Aldiansyah		122140152
Rustian Afencius Marbun		122140155
A KEVIN SERGIAN		122140125

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
INSTITUT TEKNOLOGI SUMATERA**

**2025**

## Abstrak

*Manajemen data yang akurat merupakan fondasi esensial bagi efisiensi operasional. Penelitian ini bertujuan untuk melakukan analisis deteksi anomali secara komprehensif terhadap dua dataset, 'Stok' dan 'Pembelian', guna mengidentifikasi masalah integritas dan konsistensi data. Hasil analisis mengungkap adanya anomali sistemik yang sangat kritis, yaitu kegagalan sinkronisasi fundamental antara kedua dataset. Hal ini dibuktikan dengan temuan **2.026 "Produk Hantu"** (produk tercatat dalam pembelian namun tidak terdaftar di master stok) dan **1.516 produk di data stok tanpa riwayat pembelian**. Selain itu, ditemukan pula **164 produk** dengan stok ekstrem (outlier), duplikasi kode produk, dan inkonsistensi penulisan satuan. Kesimpulan utama adalah adanya masalah fundamental dalam proses manajemen data master yang memerlukan perbaikan alur kerja bisnis. Penelitian ini merekomendasikan implementasi alur kerja pendaftaran produk baru yang terintegrasi dan standarisasi data master untuk memastikan keandalan sistem inventaris.*

**Kata Kunci:** Deteksi Anomali, Integritas Data, Kualitas Data, Analisis Data Eksploratif, Sistem Inventaris.

# 1 Pendahuluan

## 1.1 Latar Belakang

Di era digital, data menjadi aset strategis yang menentukan daya saing perusahaan. Dalam konteks manajemen inventaris, akurasi data stok dan transaksi pembelian secara langsung mempengaruhi kemampuan perusahaan dalam mengelola rantai pasok, mengoptimalkan modal kerja, dan memberikan layanan pelanggan yang prima. Kesalahan atau inkonsistensi dalam data dapat menyebabkan masalah serius, mulai dari kelebihan atau kekurangan stok (*overstocking/understocking*) hingga pengambilan keputusan pengadaan yang tidak tepat.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Apa saja jenis anomali (kualitas, integritas, statistik) yang terdapat di dalam masing-masing dataset?
2. Apakah terdapat anomali sistemik yang mengindikasikan masalah pada proses bisnis saat kedua dataset dibandingkan?
3. Bagaimana strategi yang dapat digunakan untuk membedakan antara anomali kesalahan input dan transaksi sah yang bernilai ekstrem?

## 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk:

1. Mengidentifikasi, mengklasifikasikan, dan menginterpretasikan berbagai jenis anomali data.
2. Menganalisis akar penyebab potensial dari anomali yang ditemukan.
3. Memberikan rekomendasi strategis berbasis data yang dapat ditindaklanjuti.

## 1.4 Batasan Masalah

Analisis dalam penelitian ini terbatas pada dataset ‘Dataset UTS - Stok.csv’ dan ‘Dataset UTS - Pembelian.csv’ yang disediakan. Analisis tidak mencakup data dari periode waktu lain atau sumber data eksternal lainnya yang tidak disediakan.

---

## 2 Landasan Teori

### 2.1 Deteksi Anomali

Deteksi anomali (atau deteksi outlier) adalah proses mengidentifikasi item, peristiwa, atau observasi yang tidak sesuai dengan pola yang diharapkan atau item lain dalam sebuah dataset [1]. Anomali dapat mengindikasikan berbagai insiden, seperti fraud, kegagalan sistem, atau kesalahan input data.

### 2.2 Pendekatan Deteksi Anomali

Metode deteksi anomali dapat diklasifikasikan ke dalam beberapa pendekatan, di antaranya:

- **Pendekatan Statistik:** Metode ini mengasumsikan bahwa data normal mengikuti distribusi statistik tertentu. Data yang deviasinya jauh dari distribusi tersebut dianggap anomali. Salah satu teknik yang populer adalah **Interquartile Range (IQR)**, yang efektif untuk data yang tidak terdistribusi normal.
- **Pendekatan Berbasis Aturan (Rule-Based):** Metode ini menggunakan aturan logika bisnis yang telah didefinisikan sebelumnya untuk mengidentifikasi anomali. Contohnya, "sebuah KODE produk tidak boleh merujuk pada lebih dari satu NAMA PRODUK".

### 2.3 Integritas dan Kualitas Data

Kualitas data mengacu pada kondisi data yang mencerminkan akurasi, kelengkapan, konsistensi, dan keandalannya. Integritas data adalah pemeliharaan dan jaminan akurasi dan konsistensi data selama siklus hidupnya. Anomali seringkali merupakan gejala dari buruknya kualitas dan integritas data.

## 3 Metodologi Penelitian

### 3.1 Alur Kerja Analisis

Penelitian ini mengikuti alur kerja analisis data yang sistematis: Pemuatan Data, Pembersihan Data, Analisis Data Eksploratif (EDA), Deteksi Anomali Internal, Deteksi Anomali Gabungan, dan Kesimpulan.

### 3.2 Teknik Pemuatan dan Pembersihan Data

Tantangan utama pada tahap awal adalah format data non-standar. Dataset ‘Stok’ berformat *fixed-width*, sedangkan ‘Pembelian’ berformat laporan teks kompleks. Digunakan teknik parsing custom dengan ‘pandas’ yang meliputi ‘pd.read fwf’ dengan spesifikasi kolom manual dan parsing baris per baris menggunakan ‘str.slice()’ dan ‘regex’. Proses pembersihan mencakup koreksi tipe data (terutama string numerik format Indonesia ‘1.234,56’), penanganan nilai yang hilang, dan penghapusan duplikat.

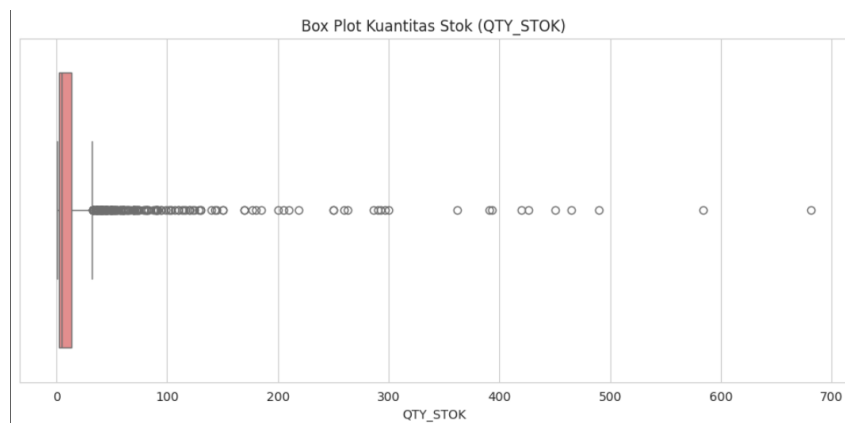
## 4 Hasil dan Pembahasan

Analisis data mengungkap tiga jenis temuan utama: anomali kualitas data internal, anomali sistemik antar dataset, dan validasi hipotesis mengenai transaksi bernilai besar.

### 4.1 Anomali Kualitas Data Internal

Analisis pada masing-masing dataset mengungkap masalah pada kualitas pencatatan data.

- **Pada Dataset Stok:** Ditemukan adanya duplikasi kode produk, inkonsistensi penulisan satuan, dan **164 produk** dengan kuantitas stok ekstrem (outlier).
- **Pada Dataset Pembelian:** Ditemukan transaksi dengan kuantitas pembelian ekstrem. Transaksi pembelian tunggal terbesar yang terdeteksi bernilai **Rp 510.000,00**.



Gambar 1: Box Plot Kuantitas Stok, Menunjukkan Adanya Outlier Signifikan

### 4.2 Anomali Sistemik: Kegagalan Sinkronisasi Antar Dataset

Ini adalah temuan paling kritis. Analisis gabungan menunjukkan bahwa kedua sistem, Stok dan Pembelian, beroperasi secara terpisah.

- **"Produk Hantu":** Ditemukan **2.026** 'KODE' produk yang ada dalam catatan transaksi pembelian tetapi tidak terdaftar sama sekali di data master stok.
- **"Stok Ajaib":** Ditemukan **1.516** 'KODE' produk yang tercatat memiliki stok tetapi tidak memiliki riwayat transaksi pembelian.

Temuan ini secara konklusif membuktikan adanya kegagalan fundamental dalam proses bisnis pencatatan inventaris.



Gambar 2: Ilustrasi Anomali Sistemik - Sistem Stok dan Pembelian yang Tidak Sinkron

### 4.3 Validasi Hipotesis Anomali Bernilai Besar

Soal tugas mengarahkan untuk menyelidiki kasus transaksi dengan total pembayaran di atas Rp 10 juta dan strategi untuk memvalidasinya.

#### 4.3.1 Hasil Investigasi Transaksi lebih besar dari Rp 10 Juta

Analisis berbasis aturan (*rule-based*) diterapkan pada kolom 'NILAI MSK' pada dataset 'Pembelian' yang telah dibersihkan. Hasilnya, **tidak ditemukan satupun transaksi** yang nilainya melebihi Rp 10.000.000. Temuan ini mematahkan hipotesis awal yang diberikan dalam kasus, dan membuktikan bahwa anomali nilai ekstrem pada dataset ini tidak mencapai skala tersebut. Ini adalah sebuah temuan penting yang menunjukkan bahwa analisis data dapat digunakan untuk memvalidasi atau menyangkal asumsi bisnis.

#### 4.3.2 Strategi Membedakan Anomali Kesalahan Input vs. Transaksi Sah

Untuk menjawab bagaimana cara membedakan anomali akibat kesalahan input dari transaksi sah yang memang bernilai besar, pendekatan hibrid yang menggabungkan analisis kuantitatif dan kualitatif direkomendasikan:

1. **Analisis Kontekstual (Kuantitatif):** Sebelum melakukan validasi eksternal, analisis dapat memeriksa data anomali secara kontekstual. Salah satu cara paling efektif adalah dengan menghitung harga per unit ('Harga Satuan = NILAI MSK / QTY MSK'). Jika sebuah transaksi memiliki nilai total yang tinggi namun harga satuannya melonjak secara tidak wajar (misalnya, 100x lipat dari transaksi lain untuk produk yang sama), ini adalah indikator kuat adanya kesalahan input pada kolom 'NILAI MSK' atau 'QTY MSK'.
2. **Validasi Eksternal (Kualitatif):** Ini adalah langkah konfirmasi yang paling menentukan. Data transaksi anomali yang ditemukan harus divalidasi dengan sumber kebenaran eksternal. Caranya adalah dengan menunjukkan temuan tersebut kepada staf yang bertanggung jawab (misalnya, bagian pengadaan atau keuangan) dan membandingkannya dengan 'dokumen fisik' seperti faktur pembelian atau *Purchase Order (PO)*. Jika terdapat diskrepansi antara data digital dan dokumen fisik, maka anomali tersebut terbukti sebagai kesalahan input.

## 5 Penutup

### 5.1 Kesimpulan

Penelitian ini menyimpulkan bahwa masalah anomali utama pada dataset ini bukanlah data outlier acak, melainkan **anomali sistemik yang berakar pada proses bisnis yang tidak terintegrasi**. Kegagalan untuk menyinkronkan data transaksi pembelian dengan data master stok telah menyebabkan integritas data inventaris menjadi sangat rendah dan tidak dapat diandalkan untuk pengambilan keputusan.

### 5.2 Saran

Berdasarkan temuan tersebut, berikut adalah saran yang direkomendasikan:

1. **Perbaiki Proses Bisnis (Prioritas Utama):** Mendesain ulang alur kerja sehingga setiap produk baru yang dibeli **wajib didaftarkan** di sistem master stok sebelum transaksi pembeliannya dapat dicatat.
2. **Audit dan Pembersihan Data Master:** Melakukan audit menyeluruh terhadap 1.516 "stok ajaib" untuk melacak asal-usulnya dan membersihkan data master 'Stok' dari duplikasi kode serta inkonsistensi lainnya.
3. **Implementasi Aturan Validasi Sistem:** Menerapkan *business rules* pada sistem input untuk secara otomatis mencegah pencatatan transaksi bagi produk yang tidak terdaftar.



## Pustaka

- [1] Aggarwal, C. C. (2017). *Outlier Analysis*. Springer.
- [2] McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media.

## A Lampiran

### A.1 Lampiran A: Tautan Repositori Proyek

- Tautan GitHub: <https://github.com/fiqrialdnsyh/KELOMPOK-15>
- Tautan Video Penjelasan: <https://bit.ly/VideoTutorialDataMining>

### A.2 Lampiran B: Contoh Detail Anomali yang Ditemukan

Tabel 1: Contoh Anomali Duplikasi Kode pada Data Stok

KODE	NAMA_PRODUK
P0000010	PARATUSIN TAB
P00000101	POLIDENT CEANSER
V0000002	VEGETA JERUK
V00000029	VOLTAREN 50MG

### A.3 Lampiran C: Kode Kunci (Parsing Data Pembelian)

Listing 1: Kode C.1: Skrip Python fungsional untuk mem-parsing file 'Pembelian.csv'

```

1 import pandas as pd
2 import numpy as np
3 import re
4
5 file_path_pembelian = 'path/to/your/file.csv'
6
7 # List untuk menampung data
8 parsed_data = []
9 current_kode, current_nama_produk, current_unit = None, None, None
10
11 with open(file_path_pembelian, 'r', encoding='utf-8', errors='ignore') as f:
12     for line in f:
13         cleaned_line = line.strip().strip('"')
14         if not cleaned_line or '----' in cleaned_line:
15             continue
16
17         # Cek jika ini adalah baris master produk
18         if re.match(r'^[A-Z]\d+\s', cleaned_line):
19             parts = cleaned_line.split('\t') if '\t' in cleaned_line else
20                 cleaned_line.split()
21             if len(parts) >= 2:
22                 current_kode, current_nama_produk, current_unit = parts[0], ' '.
23                     join(parts[1:-1]), parts[-1]
24                 continue
25
26         # Cek jika ini adalah baris transaksi
27         if re.search(r'\d{2}-\d{2}-\d{2}', cleaned_line) and current_kode:
28             parts = cleaned_line.split()

```

```
27         if len(parts) < 4: continue
28
29         tanggal, no_transaksi = parts[0], parts[1]
30
31         if no_transaksi.startswith('1.'): # Transaksi MASUK
32             qty_msk, nilai_msk, qty_klr, nilai_klr = parts[2], parts[3], '0'
33             , '0'
34         else: # Transaksi KELUAR
35             qty_msk, nilai_msk, qty_klr, nilai_klr = '0', '0', parts[2],
36             parts[3]
37
38         parsed_data.append({
39             'KODE': current_kode, 'NAMA_PRODUK': current_nama_produk,
40             'UNIT': current_unit, 'TANGGAL': tanggal, 'NO_TRANSAKSI':
41             no_transaksi,
42             'QTY_MSK': qty_msk, 'NILAI_MSK': nilai_msk,
43             'QTY_KLR': qty_klr, 'NILAI_KLR': nilai_klr
44         })
45
46 df_pembelian = pd.DataFrame(parsed_data)
47
48 # ... (lanjutan proses pembersihan)
```