

Python Based Multi File Procurement Data Integration Automation at PT. XYZ

Diyan Pebriyanti^{1*}, Nur Fitrianti Fahrudin¹

¹Program Studi Sistem Informasi, Institut Teknologi Nasional Bandung

*Email Korespondensi: pebriyantidiyan@gmail.com

Naskah Masuk 30 November 2025 | Direvisi 30 Januari 2026 | Diterima 15 Februari 2026

ABSTRAK

Pengolahan data pengadaan yang melibatkan banyak file Microsoft Excel masih banyak dilakukan secara manual, terutama pada perusahaan berskala besar. Proses ini memerlukan waktu yang lama dan memiliki risiko kesalahan yang tinggi akibat ketidakkonsistenan format data serta human error. Penelitian ini bertujuan untuk meningkatkan efisiensi dan akurasi pengisian data pengadaan melalui penerapan otomatisasi berbasis bahasa pemrograman Python. Metodologi yang digunakan adalah CRISP-DM (Cross Industry Standard Process for Data Mining) dengan fokus pada tahapan business understanding, data understanding, dan data preparation. Data penelitian berasal dari PT. XYZ berupa beberapa file Excel Nota Kategori Belanja dan Belanja Regional dengan jumlah data mencapai ribuan baris tiap filenya. Proses otomatisasi dilakukan menggunakan Python pada platform Google Colab, meliputi pembersihan data, normalisasi format, agregasi data, serta integrasi dan pencocokan data berdasarkan nama kota atau kabupaten. Hasil penelitian menunjukkan bahwa otomatisasi mampu mempercepat proses pengisian kuantitas secara signifikan serta meningkatkan konsistensi dan akurasi data dibandingkan metode manual. Meskipun masih terdapat keterbatasan akibat ketidaksesuaian data sumber, solusi yang diusulkan terbukti efektif dalam mengurangi beban kerja manual dan meminimalkan kesalahan pengolahan data pengadaan.

Kata kunci: otomatisasi data, Python, CRISP-DM, integrasi data, data pengadaan

ABSTRACT

Procurement data processing involving numerous Microsoft Excel files is still often done manually, especially in large-scale companies. This process is time-consuming and carries a high risk of errors due to inconsistent data formats and human error. This study aims to improve the efficiency and accuracy of procurement data entry through the implementation of automation based on the Python programming language. The methodology used is CRISP-DM (Cross Industry Standard Process for Data Mining) with a focus on the stages of business understanding, data understanding, and data preparation. The research data comes from PT. XYZ in the form of several Excel files of Expenditure Category Notes and Regional Expenditures with data reaching thousands of rows per file. The automation process was carried out using Python on the Google Colab platform, including data cleaning, format normalization, data aggregation, and data integration and matching based on city or district names. The results of the study show that automation can significantly speed up the quantity entry process and improve data consistency and accuracy compared to manual methods. Although there are still limitations due to inconsistencies in source data, the proposed solution has proven effective in reducing manual workload and minimizing procurement data processing errors.

Keywords: data automation, Python, CRISP_DM, data integration, procurement data

1. PENDAHULUAN

Perkembangan teknologi informasi dan digitalisasi data telah menjadi faktor kunci dalam meningkatkan efisiensi dan efektivitas operasional organisasi, baik pada sektor publik maupun swasta. Data tidak lagi hanya berfungsi sebagai arsip, melainkan sebagai aset strategis yang berperan penting dalam mendukung proses

pengambilan keputusan yang cepat, tepat, dan akurat [1][2]. Oleh karena itu, pengelolaan data yang baik, terstruktur, dan terotomatisasi menjadi kebutuhan yang tidak dapat dihindari.

Dalam praktiknya, banyak perusahaan masih menghadapi kendala dalam pengolahan data operasional, khususnya pada proses pengadaan barang dan jasa. Salah satu permasalahan yang sering muncul adalah pengisian dan pencocokan data pengadaan yang dilakukan secara manual menggunakan Spreadsheet maupun Excel [3]. Proses manual ini umumnya membutuhkan waktu yang lama, terutama ketika jumlah data yang diolah sangat besar dan tersebar dalam banyak *file*. Selain itu, metode manual juga memiliki tingkat risiko kesalahan yang tinggi (*human error*), seperti kesalahan *input*, inkonsistensi format data, dan ketidaksesuaian hasil perhitungan.

PT. XYZ sebagai objek penelitian dalam kegiatan ini juga menghadapi permasalahan serupa, khususnya pada pengelolaan banyaknya data pengadaan yang terdiri data Nota Kategori Belanja dan Belanja Regional. Data Belanja Regional berisi informasi pembelanjaan berdasarkan wilayah kota atau kabupaten, sedangkan Nota Kategori Belanja digunakan sebagai dokumen rekapitulasi pengadaan yang memerlukan pengisian kolom kuantitas secara akurat. Dalam kondisi awal, proses pengisian kuantitas pada Nota Kategori Belanja dilakukan dengan mencocokkan data dari beberapa *file* Belanja Regional secara manual. Proses ini menjadi semakin kompleks karena perbedaan format penulisan nama wilayah, adanya nilai kosong atau simbol tertentu, serta jumlah data yang mencapai ribuan baris.

Ketidakefisienan proses yang masih manual tersebut berdampak langsung pada waktu penyelesaian pekerjaan dan kualitas data yang dihasilkan. Waktu kerja menjadi lebih lama, sementara potensi terjadinya kesalahan pengisian data semakin besar. Apabila kesalahan ini tidak segera terdeteksi, maka dapat memengaruhi laporan keuangan, proses evaluasi pengadaan, serta pengambilan keputusan manajerial di tingkat selanjutnya. Oleh sebab itu, diperlukan suatu solusi yang mampu mengotomatisasi proses pengolahan data pengadaan agar lebih cepat, konsisten, dan akurat.

Apabila proses pengolahan data pengadaan tetap dilakukan secara manual, risiko kesalahan pengisian data akan semakin meningkat seiring bertambahnya volume dan kompleksitas data. Kesalahan tersebut tidak hanya berdampak pada keterlambatan penyusunan laporan pengadaan, tetapi juga dapat memengaruhi akurasi laporan keuangan dan kualitas pengambilan keputusan manajerial. Selain itu, proses manual yang memakan waktu berpotensi menurunkan efisiensi kerja dan meningkatkan ketergantungan pada sumber daya manusia, sehingga kurang optimal untuk diterapkan pada lingkungan organisasi dengan aktivitas pengadaan yang bersifat rutin dan berkelanjutan.

Pemanfaatan bahasa pemrograman Python dalam pengolahan data menjadi salah satu solusi yang banyak digunakan saat ini [4]. Python memiliki berbagai pustaka (*library*) yang mendukung proses pembersihan data, transformasi data, hingga integrasi antar *file* dalam jumlah besar [4][5]. Selain itu, Python bersifat fleksibel dan mudah diimplementasikan, sehingga cocok digunakan untuk mengotomatisasi pekerjaan berulang yang sebelumnya dilakukan secara manual [5][16]. Dengan dukungan platform seperti Google Colab, proses pengolahan data juga dapat dilakukan secara efisien tanpa memerlukan instalasi lingkungan pemrograman yang kompleks [6].

Beberapa penelitian sebelumnya telah membahas pemanfaatan otomatisasi dan pengolahan data menggunakan Python untuk meningkatkan efisiensi kerja, khususnya pada proses analisis data dan integrasi *dataset* [7][8][9]. Namun, sebagian besar penelitian tersebut masih berfokus pada pembangunan model analitik atau prediktif, serta menggunakan data dengan struktur yang relatif seragam. Di sisi lain, kajian yang secara khusus membahas otomatisasi pengisian dan pencocokan data pengadaan berbasis *multi file* Excel dengan karakteristik data yang tidak konsisten masih terbatas, terutama pada konteks implementasi langsung di lingkungan perusahaan. Selain itu, penerapan metodologi CRISP-DM dalam penelitian terdahulu umumnya mencakup seluruh tahapan hingga pemodelan, sementara pemanfaatannya yang difokuskan pada tahapan persiapan dan integrasi data untuk mendukung proses operasional belum banyak dikaji. Oleh karena itu, penelitian ini hadir untuk mengisi celah tersebut dengan mengusulkan solusi otomatisasi pengolahan data pengadaan berbasis Python yang menitikberatkan pada integrasi data *multi file* dan peningkatan efisiensi proses operasional.

Berdasarkan permasalahan dan celah penelitian yang telah diidentifikasi, penelitian ini memberikan kontribusi dalam bentuk penerapan solusi otomatisasi pengolahan data pengadaan berbasis Python yang difokuskan pada integrasi data *multi file* Excel dengan karakteristik data yang tidak konsisten. Selain itu, penelitian ini menunjukkan pemanfaatan metodologi CRISP-DM sebagai kerangka kerja sistematis dalam mendukung proses pengolahan data operasional tanpa melibatkan tahapan pemodelan. Kontribusi ini diharapkan dapat menjadi referensi praktis bagi organisasi dalam meningkatkan efisiensi pengolahan data pengadaan serta memperkaya kajian terkait otomatisasi data pada bidang sistem informasi.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada penerapan otomatisasi pengolahan data Nota Kategori Belanja dan Belanja Regional menggunakan Python dengan pendekatan metodologi CRISP-DM (*Cross Industry Standard Process for Data Mining*). Metodologi ini dipilih karena menyediakan kerangka kerja sistematis dalam memahami permasalahan bisnis, karakteristik data, serta tahapan pengolahan data yang terstruktur [10]. Diharapkan, melalui penerapan otomatisasi ini, proses pengisian kuantitas pada Nota Belanja Kategori dapat dilakukan dengan lebih efisien dan akurat, serta mampu meminimalkan kesalahan akibat intervensi manual.

Penelitian ini diharapkan dapat memberikan kontribusi praktis bagi perusahaan dalam meningkatkan kualitas pengolahan data pengadaan, serta menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan otomatisasi data, pemanfaatan Python, dan penerapan metodologi CRISP-DM dalam konteks pengolahan data operasional.

2. METODOLOGI

Metodologi penelitian yang digunakan dalam penelitian ini bertujuan untuk memberikan kerangka kerja yang sistematis dalam menyelesaikan permasalahan pengisian dan pencocokan data Nota Kategori Belanja dan Belanja Regional. Pendekatan yang digunakan adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*), yang merupakan metodologi standar lintas industri dalam pengolahan dan analisis data [17]. Metodologi ini dipilih karena mampu mengakomodasi proses pemahaman bisnis, pemahaman data, hingga persiapan data secara terstruktur dan berulang.

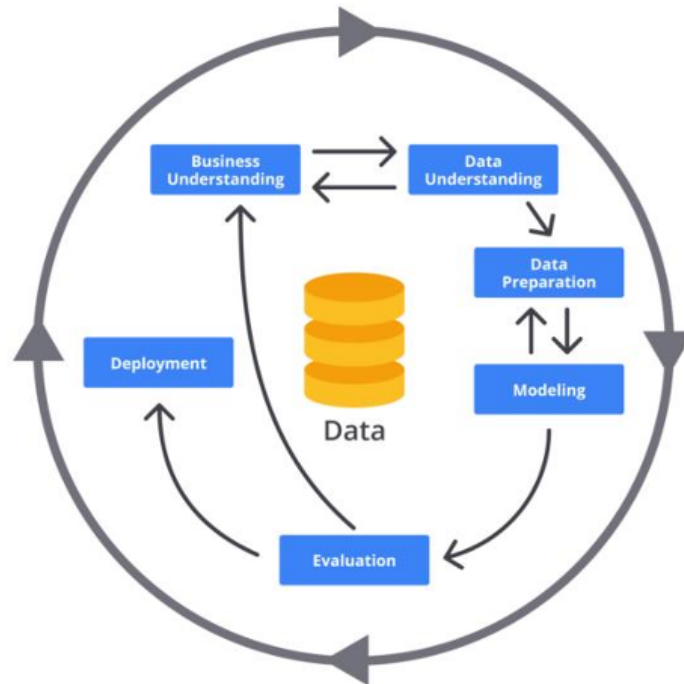
Secara umum, penelitian ini juga menggunakan pendekatan deskriptif dengan studi kasus pada PT. XYZ, di mana fokus penelitian diarahkan pada penerapan otomatisasi pengolahan data pengadaan berbasis Python. Data yang digunakan bersumber dari *file* internal perusahaan berupa beberapa *file* Microsoft Excel dengan volume data yang besar dan struktur yang beragam.

Pemilihan bahasa pemrograman Python dalam penelitian ini didasarkan pada kemampuannya dalam menangani pengolahan data berskala besar secara efisien serta dukungan pustaka yang lengkap, seperti *pandas* dan *numpy*, yang memudahkan proses pembersihan, transformasi, agregasi, dan integrasi data dari berbagai sumber [11][18]. Python juga bersifat fleksibel dan mudah dikembangkan, sehingga sesuai untuk mengotomatisasi proses pengolahan data yang bersifat repetitif dan sebelumnya dilakukan secara manual [12].

Selain itu, penggunaan platform Google Colab dipilih karena menyediakan lingkungan pemrograman berbasis *cloud* yang tidak memerlukan instalasi perangkat lunak tambahan, mendukung eksekusi kode Python secara interaktif, serta memudahkan pengelolaan dan pemrosesan *file* Excel dalam jumlah besar [13][19]. Google Colab juga memungkinkan proses pengolahan data dilakukan secara lebih cepat dan efisien, serta mendukung kolaborasi dan replikasi penelitian, sehingga solusi yang diusulkan lebih mudah diterapkan dan dikembangkan di lingkungan organisasi[19].

2.1 Kerangka Metodologi Penelitian

Kerangka metodologi penelitian ini mengacu pada tahapan CRISP-DM yang terdiri dari enam fase utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* seperti yang terdapat pada Gambar 1 [14]. Namun, dalam penelitian ini tahapan yang difokuskan meliputi *business understanding*, *data understanding*, dan *data preparation*, karena tujuan utama penelitian adalah menghasilkan data yang telah terintegrasi dan siap digunakan, bukan membangun model prediktif.



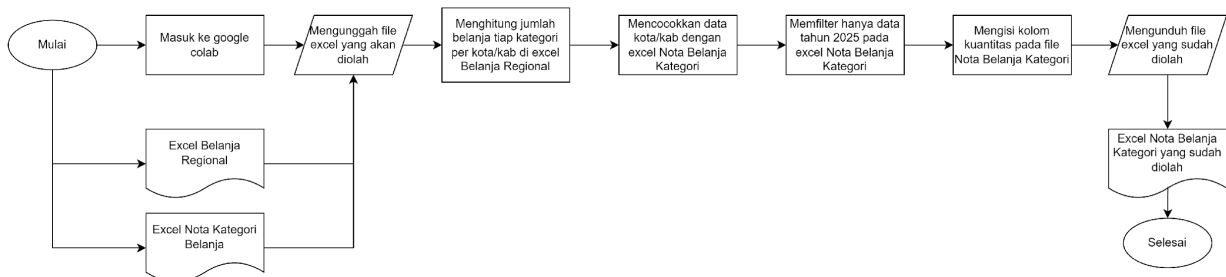
Gambar 1. Tahapan CRISP-DM

Alur metodologi dimulai dari identifikasi permasalahan bisnis yang dihadapi perusahaan, dilanjutkan dengan pemahaman karakteristik data yang digunakan, serta diakhiri dengan proses persiapan dan pengolahan data secara otomatis menggunakan bahasa pemrograman Python. Setiap tahapan dilakukan secara bertahap dan saling berkaitan untuk memastikan hasil yang diperoleh sesuai dengan kebutuhan bisnis.

2.2 Business Understanding

Tahap *business understanding* merupakan langkah awal dalam metodologi CRISP-DM yang bertujuan untuk memahami konteks bisnis, tujuan, dan permasalahan yang ingin diselesaikan [14]. Pada tahap ini dilakukan identifikasi proses pengolahan data pengadaan yang berjalan di PT. XYZ, khususnya pada proses pengisian dan pencocokan data Nota Kategori Belanja dengan data Belanja Regional.

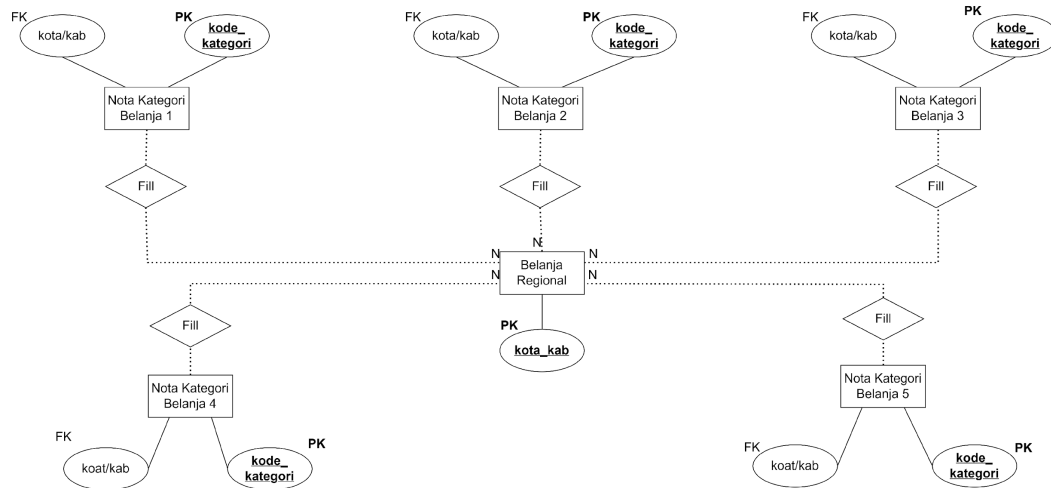
Permasalahan utama yang diidentifikasi adalah proses pengisian kuantitas pada Nota Belanja Kategori yang masih dilakukan secara manual dengan mencocokkan data dari beberapa *file* Belanja Regional. Proses tersebut membutuhkan waktu yang lama, bersifat repetitif, serta memiliki risiko kesalahan yang tinggi. Oleh karena itu, tujuan bisnis dari penelitian ini adalah mengembangkan solusi otomatisasi yang mampu mempercepat proses pengisian data, meningkatkan akurasi hasil, serta mengurangi ketergantungan pada proses manual. Adapun alur untuk *automasi* pengisian data dijelaskan pada Gambar 2.



Gambar 2. Alur kerja *automasi* pengisian data

2.3 Data Understanding

Tahap *data understanding* dilakukan untuk memahami struktur, isi, dan karakteristik data yang digunakan dalam penelitian [14][15]. Data yang digunakan terdiri dari dua jenis utama *file* Microsoft Excel yang memiliki struktur serupa namun jumlah baris data yang berbeda-beda, yaitu data lima *file* Excel Belanja Regional dan lima *file* Excel Nota Kategori Belanja, yang masing-masing *file* memiliki ribuan baris data. Penggambaran struktur tabel dan hubungan antar *file* terdapat pada Gambar 3.



Gambar 3. Diagram hubungan antar *file*

Dari diagram hubungan antar *file* yang tertera pada Gambar 3 menunjukkan bahwa Nota Kategori Belanja dan Belanja Regional memiliki relasi *one optional to many optional* yang terhubung melalui atribut kota/kabupaten, dengan hubungan sebagai berikut:

- Tiap satu kota/kabupaten pada tabel Nota Kategori Belanja bisa saja banyak/beberapa kali ditemukan di tabel Belanja Regional.
- Tidak semua kota/kabupaten yang ada di tabel Nota Kategori Belanja ditemukan di tabel Belanja Regional maupun sebaliknya.

Tabel 1. Kamus data Belanja Regional

Nama	Type Data	Keterangan	Isi Data
Kota/kab	<i>String</i> (<i>Not Null</i>) <i>Primary Key</i>	Nama kota/kabupaten	KOTA MEDAN, KOTA WONOGIRI, KAB. BADUNG
Nomor	<i>Integer</i> (<i>Not Null</i>)	Nomorurut data	1, 2, 3, ..., 469
Kelompok_wilayah	<i>String</i> (<i>Not Null</i>)	Wilayah kota/kabupaten	ACEH DAN SUMATRA UTARA, YOGYAKARTA DAN JAWA TENGAH BAGIAN TIMUR, BALI DAN NUSA TENGGARA
Wilayah_spesifik	<i>String</i> (<i>Not Null</i>)	Wilayah spesifik kota/kabupaten	MEDAN, WONOGIRI, BALI
Kode_supplier	<i>Integer</i> (<i>Not Null</i>)	Kode supplier	437259, 111123, 739200
Area	<i>String</i> (<i>Not Null</i>)	Area kota/kabupaten	ACEH & MEDAN; YOGYAKARTA & REMBANG; BALI & NUSA TENGGARA TIMUR
Objek_belanja	<i>String</i> (<i>Not Null</i>)	Daerah area belanja	SUPPLIER MEDAN, SUPPLIER UTAMA WONOGIRI, SUPPLIER BADUNG

Nama	Tipe Data	Keterangan	Isi Data
Lokasi_supplier	<i>String</i> (<i>Not Null</i>)	Lokasi tempat belanja	SUPPLIER 8 MEDAN, SUPPLIER UTAMA/GD1 WONOGIRI, SUPPLIER BADUNG
Alamat	<i>String</i> (<i>Not Null</i>)	Alamat spesifik	Jl. Lintas Sumatra No.37; Jl. Wonosari No.26-28; Jl. Raya Dalung RT/RW 002/003
Provinsi	<i>String</i> (<i>Not Null</i>)	Provinsi dari kota/kabupaten	SUMATRA UTARA, JAWA TENGAH, BALI
Belanja_kategori1	<i>Integer</i> (<i>Null</i>)	Kuantitas belanja kategori 1	5, 1, 2
Belanja_kategori2	<i>Integer</i> (<i>Null</i>)	Kuantitas belanja kategori 2	1, -, 1
Belanja_kategori3	<i>Integer</i> (<i>Null</i>)	Kuantitas belanja kategori 3	-, -, 1
Belanja_kategori4	<i>Integer</i> (<i>Null</i>)	Kuantitas belanja kategori 4	2, 1, 1
Belanja_kategori5	<i>Integer</i> (<i>Null</i>)	Kuantitas belanja kategori 5	-, 2, -
Jumlah_belanja	<i>Integer</i> (<i>Null</i>)	Jumlah kuantitas keseluruhan kategori belanja	8, 4, 4
Harga_satuan_kategori1	<i>Integer</i> (<i>Not Null</i>)	Harga satuan belanja kategori 1	1.002.380; 2.478.000; 1.388.000
Harga_satuan_kategori2	<i>Integer</i> (<i>Not Null</i>)	Harga satuan belanja kategori 2	2.000.009; 3.000.003; 2.999.008
Harga_satuan_kategori3	<i>Integer</i> (<i>Not Null</i>)	Harga satuan belanja kategori 3	3.584.555; 4.888.000; 3.273.333
Harga_satuan_kategori4	<i>Integer</i> (<i>Not Null</i>)	Harga satuan belanja kategori 4	2.488.330; 3.478.333; 2.379.489
Harga_satuan_kategori5	<i>Integer</i> (<i>Not Null</i>)	Harga satuan belanja kategori 5	2.347.449; 5.000.589; 1.444.930
Subtotal_harga_kategori1	<i>Integer</i> (<i>Not Null</i>)	Subtotal harga satuan dikali kuantitas kategori 1	5.422.714; 8.329.490; 7.450.498
Subtotal_harga_kategori2	<i>Integer</i> (<i>Not Null</i>)	Subtotal harga satuan dikali kuantitas kategori 2	7.844.925; 7.483.599; 7.600.000
Subtotal_harga_kategori3	<i>Integer</i> (<i>Not Null</i>)	Subtotal harga satuan dikali kuantitas kategori 3	9.484.760; 5.300.488; 8.588.555
Subtotal_harga_kategori4	<i>Integer</i> (<i>Not Null</i>)	Subtotal harga satuan dikali kuantitas kategori 4	7.490.333; 4.389.222; 8.399.000
Subtotal_harga_kategori5	<i>Integer</i> (<i>Not Null</i>)	Subtotal harga satuan dikali kuantitas kategori 5	7.555.599; 8.599.005; 5.499.990
Total_harga	<i>Integer</i> (<i>Not Null</i>)	Total harga keseluruhan kategori	43.300.500; 36.399.000; 23.499.599
Total_harga_3bulan	<i>Integer</i> (<i>Not Null</i>)	Total harga keseluruhan kategori per 3 bulan	97.394.000; 117.266.990; 78.600.566

Tabel 2. Kamus data Kategori Belanja

Nama	Tipe Data	Keterangan	Isi Data
Kode_kategori	<i>String</i> (<i>Not Null</i>) <i>Primary Key</i>	Kode data dengan identitas tahun	KK_2024_001, KK_2024_002, ..., KK_2026_515
Kota/kab	<i>String</i> (<i>Not Null</i>) <i>Foreign Key</i>	Foreign key dari tabel Data Belanja Regional, berisi nama kota/kabupaten beserta identitas kelas	CAT1 KOTA BANDUNG, CAT1 KAB. BANDUNG TIMUR, CAT1 KAB. BANDUNG BARAT
Satuan	<i>String</i> (<i>Not Null</i>)	Satuan belanja	paket
Harga_satuan	<i>Integer</i> (<i>Not Null</i>)	Harga satuan layanan	325.400; 320.143; 103.740
Kuantitas	<i>Integer</i> (<i>Null</i>)	Kuantitas pemesanan dari tiap kota/kabupaten	3, 2, 7

Data Belanja Regional yang terdapat pada Tabel 1 berisi informasi mengenai wilayah kota atau kabupaten, kategori belanja, serta jumlah paket belanja untuk masing-masing kategori. Sementara itu, data Nota Kategori Belanja seperti pada Tabel 2 berisi informasi pengadaan yang memerlukan pengisian kolom kuantitas berdasarkan data dari Belanja Regional. Pada tahap ini juga dilakukan identifikasi permasalahan data diantaranya adalah sebagai berikut:

- Perbedaan format penulisan nama wilayah
- Adanya ketidakkonsistenan data yang bernilai “0”, tanda strip (“-”), atau *null*,
- Perbedaan tipe data pada beberapa kolom.

Data pada *file* Belanja Regional digunakan untuk mengisi kolom kuantitas pada *file* Nota Kategori Belanja berdasarkan kecocokan nama kota/kabupaten.

2.4 Data Preparation

Tahap *data preparation* bertujuan untuk menyiapkan data agar siap digunakan dalam proses otomatisasi pengisian Nota Belanja Kategori [20]. Pada tahap ini dilakukan pembersihan data dengan menangani nilai kosong dan simbol tertentu, penyesuaian tipe data numerik, serta penyeragaman penulisan nama kota atau kabupaten untuk memastikan konsistensi antar *file* [15][20]. Selanjutnya, data Belanja Regional diagregasi untuk memperoleh jumlah paket belanja per wilayah yang digunakan sebagai dasar pengisian kolom kuantitas. Seluruh proses pengolahan dan integrasi data dilakukan menggunakan bahasa pemrograman Python sehingga pengisian kuantitas dapat dilakukan secara otomatis, konsisten, dan lebih efisien dibandingkan metode manual. Untuk lebih jelasnya terkait apa saja yang dilakukan pada tahap *data preparation* dapat dilihat pada Tabel 3.

Tabel 3. Tahapan proses *data preparation*

Tahap	Proses
<i>Data cleaning</i>	• Penggabungan kolom untuk dijadikan <i>header</i> kolom • Hapus spasi berlebih • Mengganti nilai strip (“-”) menjadi nol (“0”) • Mengganti nilai <i>null</i> pada kuantitas menjadi nol (“0”)
<i>Data transformation</i>	• Mengganti tipe data pada kolom numerik menjadi integer • Agregasi jumlah data pengadaan per kota/kabupaten • Normalisasi nama kota/kabupaten menjadi huruf kecil • <i>Merge dataset</i> berdasarkan nama kota • Memperbarui nilai kuantitas di Nota Kategori Belanja dengan nilai dari Excel Belanja Regional berdasarkan nama kota/kabupaten

3. HASIL DAN PEMBAHASAN

Implementasi dilakukan menggunakan bahasa pemrograman Python untuk mengolah beberapa *file* Microsoft Excel dengan volume data yang besar dan struktur yang seragam.

3.1 Hasil Implementasi Otomatisasi

Hasil implementasi menunjukkan bahwa proses otomatisasi berhasil mengisi kolom kuantitas pada Nota Kategori Belanja secara otomatis berdasarkan hasil agregasi data Belanja Regional. Dari total ± 5.000 baris data Nota Kategori Belanja dan data Belanja Regional yang diproses, sebanyak lebih dari 3.600 baris data ($\pm 72\%$) berhasil terisi secara otomatis, sementara sisanya tidak terisi karena tidak ditemukannya data pengadaan pada wilayah terkait atau nama kota/kabupaten tidak ditemukan pada data sumber.

Proses otomatisasi menggantikan metode manual yang sebelumnya dilakukan dengan pencocokan data satu per satu antar *file* Excel. Dengan penerapan pembersihan dan normalisasi data pada tahap *data preparation*, tingkat kecocokan data antar *file* meningkat secara signifikan dibandingkan kondisi awal sebelum dilakukan standarisasi format data.

3.2 Analisis Efisiensi Waktu

Dari sisi efisiensi waktu, otomatisasi pengolahan data menunjukkan peningkatan yang signifikan. Berdasarkan pengamatan proses kerja, pengisian kuantitas secara manual untuk satu periode data pengadaan dengan ribuan baris membutuhkan waktu rata-rata 10 sampai dengan 16 jam kerja. Setelah diterapkan otomatisasi menggunakan Python, seluruh proses pengolahan data, mulai dari pembersihan hingga pengisian kuantitas, dapat diselesaikan dalam waktu 30 sampai 45 menit.

Hal ini menunjukkan adanya pengurangan waktu kerja lebih dari 96% dibandingkan metode manual. Peningkatan efisiensi tersebut berdampak langsung pada produktivitas kerja, karena proses pengolahan data pengadaan yang bersifat rutin dan berulang dapat diselesaikan dengan lebih cepat, sehingga sumber daya manusia dapat dialokasikan pada aktivitas analisis dan pengambilan keputusan yang lebih strategis.

3.3 Analisis Akurasi Data

Dari segi akurasi, otomatisasi berbasis Python mampu mengurangi potensi kesalahan akibat *human error*. Pada proses manual, kesalahan pengisian kuantitas dan ketidakkonsistenan perhitungan sering terjadi, terutama pada data dengan format yang tidak seragam. Setelah diterapkan otomatisasi, hasil pengisian kuantitas menjadi lebih konsisten karena seluruh perhitungan dilakukan secara sistematis berdasarkan data yang telah dibersihkan dan distandarisasi.

Namun demikian, masih terdapat beberapa data Nota Kategori Belanja yang kolom kuantitasnya tidak terisi. Hal ini disebabkan oleh pada kota/kabupaten tersebut tidak melakukan pengadaan ataupun tidak ditemukannya kecocokan nama kota/kabupaten pada data Belanja Regional. Kondisi tersebut menunjukkan bahwa kualitas dan konsistensi data sumber sangat memengaruhi keberhasilan proses otomatisasi.

3.4 Pembahasan

Berdasarkan hasil yang diperoleh, dapat disimpulkan bahwa penerapan otomatisasi pengolahan data menggunakan Python dengan pendekatan CRISP-DM efektif dalam meningkatkan efisiensi dan akurasi pengisian data pengadaan. Pendekatan ini sangat relevan untuk diterapkan pada organisasi yang memiliki volume data besar dan proses pengolahan data yang bersifat repetitif.

Meskipun demikian, keberhasilan otomatisasi sangat bergantung pada kualitas data awal. Oleh karena itu, diperlukan standarisasi penulisan data sejak tahap *input* agar proses integrasi data dapat berjalan lebih optimal. Pengembangan lanjutan juga dapat dilakukan dengan menambahkan mekanisme validasi otomatis atau pemetaan nama wilayah untuk meningkatkan tingkat kecocokan data.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat disimpulkan beberapa poin utama sebagai berikut:

1. Penerapan otomatisasi pengolahan data pengadaan berbasis Python terbukti mampu meningkatkan efisiensi waktu secara signifikan. Proses pengisian kuantitas pada Nota Kategori Belanja yang sebelumnya membutuhkan waktu berjam-jam dengan metode manual dapat diselesaikan dalam hitungan menit melalui proses otomatisasi, sehingga mengurangi beban kerja yang bersifat repetitif.
2. Otomatisasi yang diimplementasikan juga meningkatkan konsistensi dan akurasi data pengadaan. Proses pembersihan, normalisasi, dan integrasi data yang dilakukan secara sistematis mampu meminimalkan kesalahan pengisian data akibat *human error* serta menghasilkan data yang lebih seragam dan dapat diandalkan.
3. Metodologi CRISP-DM efektif digunakan sebagai kerangka kerja dalam pengolahan dan integrasi data operasional. Fokus pada tahapan *business understanding*, *data understanding*, dan *data preparation* memungkinkan penelitian ini menghasilkan data yang siap digunakan tanpa memerlukan pembangunan model analitik atau prediktif.
4. Penelitian ini memberikan kontribusi praktis bagi organisasi dalam meningkatkan kualitas pengolahan data pengadaan. Solusi yang diusulkan dapat dijadikan referensi untuk penerapan otomatisasi serupa pada proses pengolahan data lainnya. Penelitian selanjutnya disarankan untuk mengembangkan mekanisme validasi dan pemetaan data yang lebih lanjut serta memperluas cakupan otomatisasi pada jenis data pengadaan yang berbeda.

Untuk penelitian selanjutnya, disarankan untuk dilakukan pengembangan pada aspek standarisasi dan validasi data, serta perluasan cakupan otomatisasi agar dapat diterapkan pada jenis data pengadaan lainnya. Dengan demikian, pemanfaatan otomatisasi data diharapkan dapat memberikan kontribusi yang lebih optimal dalam mendukung pengambilan keputusan berbasis data.

DAFTAR PUSTAKA

- [1] A. Budi Trisnawan, "Pemanfaatan Big Data dalam Sistem Informasi untuk Pengambilan Keputusan Strategis," *Journal of Information System and Education Development*, vol. 3, no. 3, pp. 39–43, Sep. 2025, doi: 10.62386/jised.v3i3.163.
- [2] Alvina Dwi Suwandita, Vania Pijasari, Adinda Eka Diani Prasetyowati, and Mochammad Isa Anshori, "Analisis Data Human Resources Untuk Pengambilan Keputusan: Penggunaan Analisis Data Dan Artificial Intelligence (AI) Dalam Meramalkan Tren Sumber Daya Manusia, Pengelolaan Talenta, Dan Rentensi Karyawan," *Manajemen Kreatif Jurnal*, vol. 1, no. 4, pp. 97–111, Oct. 2023, doi: 10.55606/makreju.v1i4.2161.
- [3] G. Gumilar and R. Wulan, "Sistem Informasi Peminjaman Aset Barang Kantor Di Lapas Kelas II A Cibinong," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 5, no. 5, pp. 819–825, Oct. 2022, doi: 10.32672/jnkti.v5i5.4821.
- [4] Angelina M. T. I. Sambu Ua *et al.*, "Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru," *Jurnal Publikasi Teknik Informatika*, vol. 2, no. 2, pp. 88–99, Jul. 2023, doi: 10.55606/jupti.v2i2.1742.
- [5] Regina Lo *et al.*, "Penggunaan Bahasa Pemrograman Python dalam Menganalisis Hubungan Kualitas Kopi dengan Lokasi Pertanian Kopi," *Jurnal Publikasi Teknik Informatika*, vol. 2, no. 2, pp. 100–109, May 2023, doi: 10.55606/jupti.v2i2.1752.
- [6] Zidanul Akbar, Asrul Suwondo, Rizky Ramadhan, and Abdul Halim Hasugian, "Deteksi Warna Dasar Menggunakan Metode Thresholding HSV dengan OpenCV," *Neptunus: Jurnal Ilmu Komputer Dan Teknologi Informasi*, vol. 3, no. 3, pp. 256–264, Aug. 2025, doi: 10.61132/neptunus.v3i3.1020.
- [7] D. A. Manalu and G. Gunadi, "IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI," *Infotech: Journal of Technology Information*, vol. 8, no. 1, pp. 43–54, Jun. 2022, doi: 10.37365/jti.v8i1.131.
- [8] Andy Hermawan *et al.*, "Optimalisasi Strategi Pemasaran Melalui Analisis RFM pada Dataset Transaksi Ritel Menggunakan Python," *Jurnal Manajemen Riset Inovasi*, vol. 2, no. 4, pp. 254–267, Oct. 2024, doi: 10.55606/mri.v2i4.3246.

- [9] S. S. Mangun and K. Kusrini, "Prediksi Kebakaran Hutan Ibu Kota Nusantara Menggunakan Produk MODIS dengan Algoritma Regresi Linear, Gradient Boosting dan Decision Tree," *Jambura Journal of Informatics*, vol. 1, no. 1, pp. 01–10, Apr. 2025, doi: 10.37905/jji.v1i1.30926.
- [10] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, Oct. 2021, doi: 10.30871/jaic.v5i2.3200.
- [11] H. Z. Yahya and Y. Ramdhani, "Klasifikasi Bumbu Dapur Pasar Menggunakan Metode Deep Neural Network Berbasis Android," *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 6, no. 1, pp. 19–33, Feb. 2023, doi: 10.32672/jnkti.v6i1.5569.
- [12] A. Az Zahra, N. Istiana, and A. Wibowo, "Implementasi Interpolasi Polinomial Bentuk Baku dan Metode Selisih Terbagi Newton Menggunakan Excel dan Google Colab," *GAUSS: Jurnal Pendidikan Matematika*, vol. 8, no. 1, pp. 1–13, Jun. 2025, doi: 10.30656/gauss.v8i1.10589.
- [13] Febby Wilyani, Qonaah Nuryan Arif, and Fitri Aslimar, "Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory," *Jurnal Pelayanan dan Pengabdian Masyarakat Indonesia*, vol. 3, no. 1, pp. 08–14, Mar. 2024, doi: 10.55606/jppmi.v3i1.1087.
- [14] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput Sci*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
- [15] V. Plotnikova, M. Dumas, and F. P. Milani, "Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements," *Data Knowl Eng*, vol. 139, p. 102013, May 2022, doi: 10.1016/j.datak.2022.102013.
- [16] A. A. A. Fernandes, M. Koehler, N. Konstantinou, P. Pankin, N. W. Paton, and R. Sakellariou, "Data Preparation: A Technological Perspective and Review," *SN Comput Sci*, vol. 4, no. 4, p. 425, Jun. 2023, doi: 10.1007/s42979-023-01828-8.
- [17] Dr. S. R. Sukhdeve and S. S. Sukhdeve, "Google Colaboratory," in *Google Cloud Platform for Data Science*, Berkeley, CA: Apress, 2023, pp. 11–34. doi: 10.1007/978-1-4842-9688-2_2.
- [18] R. Nursyanti, G. Prakarsa, S. N. Rahmawati, and A. Andriana, "Implementasi K-Nearest Neighbor untuk Prediksi Harga Bibit Kelapa Sawit dengan Metode CRISP-DM," *SisInfo*, vol. 7, no. 1, pp. 124–134, Feb. 2025, doi: 10.37278/sisinfo.v7i1.1122.
- [19] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, "The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks," *SBC Reviews on Computer Science*, vol. 4, no. 1, pp. 28–43, Oct. 2024, doi: 10.5753/reviews.2024.3757.
- [20] A. Az Zahra, N. Istiana, and A. Wibowo, "Implementasi Interpolasi Polinomial Bentuk Baku dan Metode Selisih Terbagi Newton Menggunakan Excel dan Google Colab," *GAUSS: Jurnal Pendidikan Matematika*, vol. 8, no. 1, pp. 1–13, Jun. 2025, doi: 10.30656/gauss.v8i1.10589.