



PENGGUNAAN BAHASA PYTHON UNTUK ANALISIS DAN VISUALISASI DATA PENDUDUK DI DESA SUMBERJO, NGANJUK

Angelina Kencana Putri^{1*}, Dhani Ichsanuddin Nur²

^{1,2} Universitas Pembangunan Nasional “Veteran” Jawa Timur

*E-mail: 21012010025@student.upnjatim.ac.id¹, dhaniin.ma@upnjatim.ac.id²

ABSTRAK

Di era perkembangan teknologi yang sangat pesat ini melakukan suatu pekerjaan lebih banyak dengan penggunaan alat digital, yang mana ini memudahkan manusia untuk melakukan pekerjaan tanpa harus dengan cara manual. Salah satunya adalah melakukan analisis dan visualisasi data dengan menggunakan python. Python ini merupakan bahasa pemrograman yang memiliki banyak manfaat. Tiap tahun data akan semakin bertambah banyak dan semakin banyak data yang ada itu jika tidak menggunakan python akan memungkinkan terjadinya kesalahan penginputan atau analisis data. Maka dari itu penggunaan python ini berguna untuk meminimalisir kesalahan data penduduk dan memudahkan dalam membaca suatu data. Pada saat ini pengolahan data penduduk di desa sumberjo masih menggunakan cara manual dan dengan menggunakan Microsoft excel yang kurang efisien jika dilakukan pada saat ini. Banyaknya data yang ada jika terus menggunakan cara seperti itu akan kurang efisien jika masih terus dilakukan. Maka dari itu diadakan pelatihan dasar dan pengimplementasian penggunaan python untuk memudahkan petugas desa dalam melakukan analisis dan visualisasi data. Pelatihan ini dilakukan dari python yang paling dasar yaitu memproses data menggunakan google colab. Ini berguna untuk efisiensi waktu, tenaga dan meminimalisir kesalahan dalam menganalisis maupun visualisasi data penduduk.

Kata kunci: Python, Analisis data, Visualisasi data

USE OF PYTHON LANGUAGE FOR ANALYSIS AND VISUALIZATION OF POPULATION DATA IN SUMBERJO VILLAGE, NGANJUK

ABSTRACT

In this era of very rapid technological development, doing more work uses digital tools, which makes it easier for people to do work without having to do it manually. One of them is analyzing and visualizing data using Python. Python is a programming language that has many benefits. Every year the data will increase and the more data there is, if you don't use Python, it will be possible for data input or analysis errors to occur. Therefore, using Python is useful for minimizing population data errors and making it easier to read data. Currently processing population data in Sumberjo village still uses manual methods and uses Microsoft Excel which is less efficient when done at this time. The large amount of data that exists if you continue to use this method will be less efficient if it continues. Therefore, basic training was held and implementation of the use of Python to make it easier for village officials to carry out data analysis and visualization. This training is carried out using the most basic Python, namely processing data using Anaconda google colab. This is useful for efficiency of time, energy and minimizing errors in analyzing and visualizing population data.

Keywords: Python, Data analysis, Data visualization

PENDAHULUAN

Data kependudukan sangatlah penting. Kumpulan angka, informasi, fakta, atau observasi yang dikumpulkan, direkam, atau disimpan disebut data. Dalam konteks yang lebih luas, data digunakan sebagai dasar untuk analisis, pengambilan keputusan, dan pemahaman lebih lanjut tentang fenomena atau peristiwa. Data digunakan untuk menganalisis pola, tren, atau hubungan antara variabel, dan membantu dalam pengambilan keputusan. Data digunakan untuk melacak suatu proses atau entitas. Sebuah desa dipimpin oleh seorang kepala desa dan dibantu oleh beberapa staf, terutama sekretariat. Seorang sekretaris di desa bertanggung jawab untuk membantu kepala desa dalam menganalisis data, yaitu mencari, mengumpulkan, menghimpun, dan mengolah data, serta informasi yang berkaitan dengan bidang teknis administrasi desa. Mereka juga harus menyiapkan bahan untuk menyusun kebijakan, pedoman, dan petunjuk teknis yang berkaitan dengan perangkat desa, menyusun dan



menyajikan data. Namun, data saat ini yang ada di desa Sumberjo disajikan dalam bentuk tabel sederhana yang sulit dipahami dengan cepat. Data kependudukan belum ditampilkan dalam visualisasi data dalam bentuk grafik sehingga kesulitan untuk mengidentifikasi data. Penggunaan software yang memudahkan pendataan, penganalisisan dan visualisasi data untuk pendataan penduduk di desa sangatlah penting.

Permasalahan yang muncul pada di desa sumberjo ini adalah dalam pengelolaan dan penyajian data kependudukan yang ada. Pegawai desa yang ada di desa sumberjo ini belum mengenal serta menggunakan python dan hanya menggunakan microsoft excel dan juga pendataan secara manual yang dirasa untuk digunakan saat ini kurang efisien, karena jika menggunakan python akan lebih mudah untuk melakukan analisis data kependudukan. Permasalahan ini juga disebabkan oleh kurangnya sarana informasi yang jelas tentang penyajian data. Dengan menggunakan bahasa pemrograman, dapat mengotomatisasi proses penyajian data. Dapat membuat skrip atau program yang memungkinkan pengumpulan, pemrosesan, dan penampilan data secara otomatis, yang menghemat waktu dan usaha. Dengan menggunakan bahasa pemrograman, dapat membuat antarmuka interaktif yang memungkinkan pengguna berinteraksi dengan data. Antarmuka ini dapat terdiri dari elemen seperti grafik interaktif, formulir input, atau aplikasi berbasis web yang memungkinkan pengguna mempelajari lebih lanjut tentang data.

Analisis data penduduk adalah langkah penting dalam perencanaan pembangunan desa yang efektif. Dengan menggunakan alat analisis yang tepat, desa dapat memahami kebutuhan penduduk, dan merencanakan program yang lebih baik. Python adalah salah satu bahasa pemrograman yang kuat yang dapat digunakan untuk mempermudah analisis dan visualisasi data penduduk. Di era teknologi, pemrosesan tampilan data sekarang mudah karena menggunakan tools (Prasetya et al., 2019). Bahasa pemrograman Python adalah salah satu alat yang paling populer saat ini. Baik untuk bisnis perusahaan, penginputan data penduduk, analisis data, dan visualisasi data. Membuat keputusan data terkait berapa penduduk di setiap RT dari berbagai dusun yang ada di desa tersebut merupakan pengelompokan penduduk dari analisis data. Sekretaris desa dapat menemukan tren dan pola dengan memeriksa data historis saat ini. Oleh karena itu, analisis data menjadi penting bagi pengolahan data. Pengolahan data satu per satu dengan secara manual adalah tugas yang sulit. Namun, tugas-tugas sulit seperti pemrosesan data dapat disederhanakan dengan komputer dan kecerdasan buatan (Siregar et al., 2023). Penumpukan data akan terjadi karena data semakin banyak. Pengolahan data yang baik dapat meminimalisir kesalahan data dan juga dapat mengidentifikasi kebutuhan masa depan yang berperan penting dalam pengembangan desa. Analisis data dapat dilakukan dalam berbagai cara yaitu menggunakan ilmu data atau metode lainnya. Menggambarkan data secara nyata baik yang berbentuk tabel, grafik bar, grafik pie, grafik garis, peta, dan diagram disebut visualisasi data (Husna et al., 2023). Python adalah bahasa pemrograman yang mudah digunakan untuk melakukan pengolahan dan penyajian data di era teknologi modern. Python ini merupakan bahasa pemrograman tingkat tinggi yang dibuat oleh Guido van Rossum yang mana berorientasi pada objek. Python adalah bahasa multifungsi, kaya akan fitur, dan dapat dijelaskan secara maksimal serta tersusun rapi. Seringkali, kumpulan data yang besar dan kompleks dapat disederhanakan dengan menggunakan bahasa pemrograman berorientasi objek ini. Karena komunitas Python yang besar dan aktif, banyak sumber daya, tutorial, dan dukungan yang tersedia (Radhi et al., 2021). Jika menghadapi masalah atau kesulitan dengan analisis data, dapat dengan mudah menemukan bantuan. Python mudah diintegrasikan dengan banyak platform dan perangkat lunak. Data analisis dapat bekerja dengan data dari berbagai sumber dan mengintegrasikannya ke dalam alur kerja mereka berkat hal ini. Karena popularitasnya yang tinggi dalam bidang pembelajaran mesin dan kecerdasan buatan, data analisis dapat dengan mudah membuat model prediktif dan algoritma cerdas.

Python adalah bahasa pemrograman interpretatif yang dapat digunakan untuk berbagai tujuan, dan fokusnya pada keterbacaan kode membuat sintaks lebih mudah dipahami. Python juga sangat mudah dipelajari bahkan bagi pemula maupun orang yang sudah mahir menggunakan bahasa pemrograman lainnya, karena itu Python banyak diminati oleh pemula dan profesional karena bahasa pemrogramannya yang mudah dipelajari (Junaidi et al., 2023). Python dipilih sebagai bahasa pemrograman software untuk menganalisis data dan visualisasi data juga karena memiliki perpustakaan yang luas yang dapat digunakan untuk berbagai aplikasi, mulai dari pemrograman statistik hingga pembelajaran mendalam. Python adalah bahasa pemrograman terbaik yang membuat



sintak lebih mudah dipahami. Python ini juga menawarkan banyak paket visualisasi, ini dikarenakan mempunyai visual yang membuat data lebih gampang diakses dan dipakai dengan membuat berbagai bagan dan grafik interaktif. Ini menjadikan python sebagai salah satu alat yang harus dimiliki dan ini tidak hanya untuk analisis data tetapi juga untuk segala ilmu data. Dengan demikian, Python membuat penggunaannya lebih mudah memahami data (Junaidi et al., 2023). Python adalah bahasa pemrograman open-source, yang berarti mendapatkan akses gratis ke kode sumbernya. Ini membuatnya murah bagi organisasi dan individu yang menangani data. Data science adalah bidang interdisipliner yang berfokus pada pengumpulan, analisis, interpretasi, dan penggunaan data untuk pengambilan keputusan dan pengungkapan wawasan. Ini melibatkan penggunaan teknik statistik, pemodelan matematika, pemrograman komputer, dan pengetahuan domain yang mendalam untuk menggali dan menganalisis data. Bidang ini terus berkembang karena munculnya teknologi dan alat baru untuk mengelola dan menganalisis data yang lebih besar.

Untuk meningkatkan visualisasi data, google colab mempunyai beberapa keuntungan. Antara lain sebagai berikut: akses cloud computing gratis pengguna dapat memanfaatkan sumber daya komputasi termasuk GPU dan TPU tanpa biaya, kemudahan kolaborasi memungkinkan tim bekerja sama dalam pengembangan visualisasi data bersamaan, integrasi sumber daya dimana google drive dapat digunakan untuk impor dan ekspor data dengan mudah dan pustaka python yang kuat. Menggunakan google colab untuk visualisasi data, notebook interaktif memungkinkan untuk mengeksplorasi data interaktif, menjalankan kode secara iteratif, melihat hasil visualisasi, dan menyesuaikan visualisasi sesuai kebutuhan. membuat visualisasi yang menarik: dari grafik garis sederhana hingga visualisasi data yang kompleks dengan palet warna dan desain yang menarik (Husna et al., 2023).

METODE

Kegiatan sosialisasi pemahaman terhadap python pada staf kantor desa ini dilaksanakan di kantor Desa Sumberjo, Kabupaten Nganjuk. Pelaksanaan kegiatan ini diberikan kepada staf sekretariat desa yang ada di kantor Desa Sumberjo, khususnya untuk bagian operator atau sekretaris desa yang bertugas untuk membantu kepala desa menghimpun data kependudukan sebagai bahan untuk analisis dan visualisasi data. Kegiatan ini dilengkapi dengan modul dan video tutorial yang menjelaskan cara mengolah dan menampilkan data penduduk secara praktis menggunakan Python, software anaconda yang bekerja menggunakan google colab. Harapannya setelah dilakukan bimbingan, arahan, dan pelatihan teknis untuk staf sekretariat desa akan memudahkan sekretaris desa mengolah, menganalisis dan memvisualisasikan data dengan baik dan efisiensi waktu kerja meningkat.

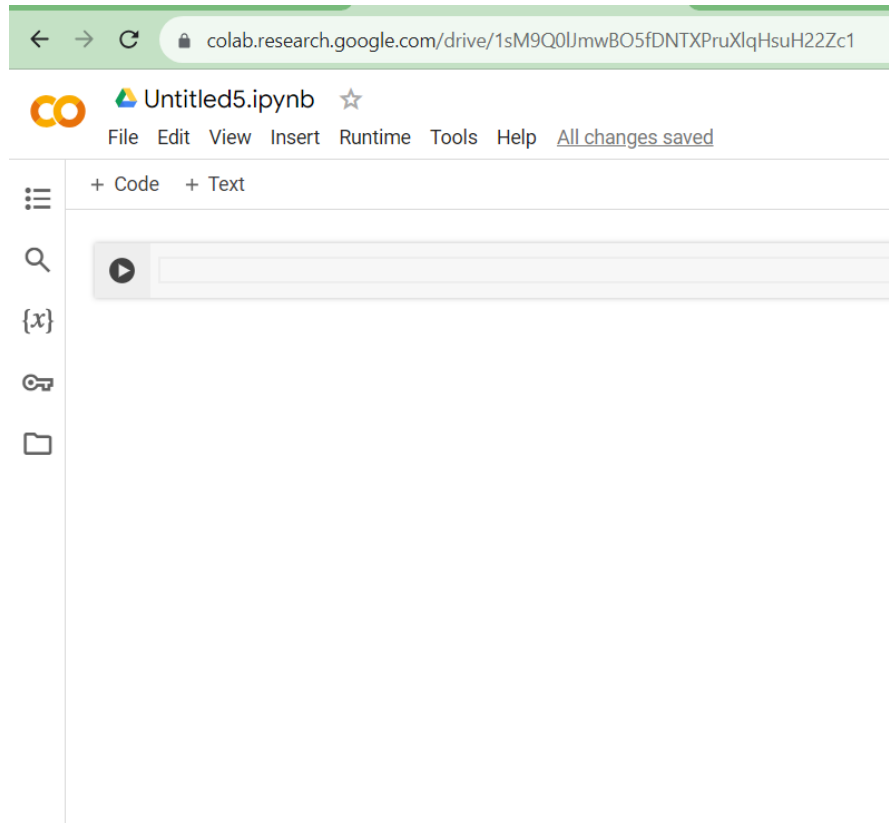
Adapun metode pengabdian masyarakat ini diawali dengan tahap persiapan yaitu koordinasi dan meminta persetujuan dengan Kepala Desa Sumberjo terkait rencana kegiatan yang akan dilaksanakan dan menjelaskan secara detail maksud dan tujuan melakukan kegiatan ini. Kedua yaitu tahap pelaksanaan yang dilaksanakan pada hari jumat tanggal 27 Oktober 2023 pada pukul 09.00 WIB hingga selesai dan dilakukan pertemuan di kantor Desa Sumberjo, Kabupaten Nganjuk bersama dengan kepala desa dan sekretaris desa Sumberjo.

HASIL DAN PEMBAHASAN

Para programmer harus mahir dalam bahasa pemrograman. Bahasa program, juga dikenal sebagai pemrograman sendiri, adalah suatu pedoman yang diberikan kepada komputer untuk melakukan tugas-tugas tertentu untuk menyelesaikan masalah. Python adalah salah satu dari banyak bahasa pemrograman yang semakin populer di seluruh dunia dalam beberapa tahun terakhir ini. Python adalah bahasa pemrograman komputer yang umum digunakan untuk membuat situs web, software, dan aplikasi, serta mengotomatiskan tugas dan analisis data. Bahasa pemrograman juga dapat membuat berbagai program, bukan hanya untuk menyelesaikan masalah tertentu saja. Python ini menjadi bahasa pemrograman yang paling banyak digunakan karena mudah digunakan dan serbaguna.

Program ini memberikan instruksi tentang cara memakai python, pengenalan ikon dan alat yang tersedia untuk Python, mengiput data, dan menyajikan grafik dengan menggunakan python yang bekerja di google colab. Untuk mendapatkan informasi dari data yang digunakan untuk pendataan data penduduk, eksplorasi data diperlukan. Google colab, juga dikenal sebagai google colaboratory, adalah sebuah platform berbasis cloud yang dibuat oleh google yang memanfaatkan bahasa pemrograman

Python untuk pengembangan dan analisis data. Google colab sangat disukai oleh pengembang, peneliti, dan praktisi ilmu data karena memberikan infrastruktur komputasi yang kuat tanpa biaya tambahan, memungkinkan pengguna untuk membuat dan menjalankan notebook interaktif di lingkungan berbasis web dengan akses gratis ke sumber daya komputasi berkekuatan tinggi seperti unit pemrosesan grafis (GPU) dan unit pemrosesan tensor (TPU). Dalam hal ini, kita akan mencoba menggunakan google colab dengan melakukan pencarian pada google. Setelah sudah melakukan pencarian di google bisa membuka google colab yang tampil pada menu google yang paling atas seperti pada gambar diatas. Lalu akan tampil menu seperti di bawah ini dan pilih menu new notebook jika baru menggunakan google colab dan open colab untuk membuka notebook yang sudah dibuat.



Gambar 1. Tampilan pada *new notebook google colab*

(Sumber: colab.research.google.com)

Gambar 1 diatas merupakan tampilan kerja pada google colab yang memiliki banyak fitur yang tersedia untuk mengoperasikan google colab ini. Berikut ini merupakan tabel petunjuk fitur yang ada di python google colab beserta fungsinya:

Tabel 1. Fitur Google Colab

FITUR	FUNGSI
File	fitur "File" Google Colab digunakan untuk membuka notebook yang sudah ada di Google Drive atau mengimpor notebook dari GitHub atau URL, yang memudahkan akses dan manajemen notebook. Simpan dalam Google Drive, menggunakan opsi ini untuk menyimpan notebook yang sedang dikerjakan ke Google Drive. Ini adalah cara yang baik untuk mengamankan pekerjaan dan membuat salinan cadangan. Simpan sebagai, menggunakan opsi ini untuk menyimpan salinan notebook yang sedang dikerjakan ke Google Drive Pustaka Data memungkinkan pengelolaan dataset dan file lainnya yang digunakan dalam sesi Colab. Mengunggah file ke sesi Colab dari komputer lokal atau mengunduh file dari sesi Colab ke komputer lokal. Google Colab otomatis menyimpan riwayat revisi dari notebook. Selain itu, Anda akan menemukan catatan teknis di bagian bawah menu "File", yang memberikan informasi tentang penyimpanan sementara, penggunaan CPU, GPU, dan versi Python yang digunakan oleh sesi Colab



FITUR	FUNGSI
Edit	Jika ingin mengembalikan perubahan yang tidak diinginkan atau membatalkan tindakan yang telah dibatalkan sebelumnya. Menghapus konten yang dipilih dari sel aktif dan menyalinnya ke clipboard. Menyalin konten yang dipilih dari sel aktif ke clipboard tanpa menghapusnya dari sel awal. Memasukkan konten yang ada di clipboard ke dalam sel aktif. Setelah menggunakan fungsi "Cut" atau "Copy" untuk mengubah jenis sel aktif antara "Code" (kode Python) dan "Text" (teks Markdown), ini berguna. Dengan bantuan ini, dapat menulis kode atau teks penjelasan dalam satu notebook.
Insert	Insert pada Google colab untuk menambah kode sel, teks sel, section header sel, sel kode awal, cuplikan kode, dan tambah bidang formulir
View	Google Colab's menu "View" menawarkan berbagai pilihan untuk mengatur tampilan dan penampilan notebook.
Runtime	Menjalankan semua sel di notebook secara berurutan, dari atas ke bawah. Jalankan Sebelum, jalankan sel yang dipilih dan semua sel di atasnya. Jalankan Setelah, jalankan sel yang dipilih dan semua sel di bawahnya. Jalankan Teks yang Dipilih, menjalankan kode yang dipilih dalam sel teks.
Tools	Menggunakan shortcut keyboard untuk mencari dan menjalankan perintah tertentu dalam colab, mengakses berbagai pengaturan colab, seperti tema, editor kode, font, dan sebagainya. Keyboard shortcuts, yang menampilkan daftar pintasan keyboard yang dapat digunakan dalam colab. Menyusun daftar isi berdasarkan markup teks dalam notebook Form, membuat formulir interaktif yang memungkinkan pengguna mengisinya.
Help	Menampilkan daftar keyboard shortcuts yang dapat digunakan dalam Colab. Memberikan rekomendasi tentang cara menggunakan Colab dengan lebih baik, memberikan tutorial singkat tentang penggunaan Google Colab. Colab FAQ, berisi jawaban atas pertanyaan yang sering diajukan tentang Google Colab. Mengarahkan ke forum komunitas Google Colab untuk mendapatkan bantuan atau berbicara dengan pengguna lain. Melaporkan masalah atau kesalahan yang mungkin ditemui saat menggunakan Colab.

Untuk mencari, mengganti, menambah, dan menghapus data, pertama-tama, harus memiliki data populasi penduduk yang akan dianalisis. Data ini dapat diambil dalam bentuk file CSV, Excel, atau bahkan diambil secara online jika tersedia. Untuk mengimpor data penduduk ke dalam notebook colab ini dapat menggunakan pustaka seperti Pandas untuk membaca file data dari penyimpanan lokal atau mengambilnya dari sumber online. Gunakan Pandas untuk melakukan eksplorasi awal terhadap data penduduk, ini berguna untuk dapat melihat struktur data, melihat statistik deskriptif, dan menemukan pola atau tren yang menarik. Dengan menggunakan google colab ini juga dapat dengan mudah mengirimkan tautan notebook colab kepada orang lain atau mengundang mereka untuk bekerja sama mengedit data secara online. Google colab notebook juga dapat memudahkan untuk mencari, megedit data sesuai dengan yang diinginkan, ini memudahkan penggunaanya dalam mencari dan mengedit data yang mempunyai banyak kolom dan baris.

Kode-kode yang sering digunakan untuk mencari dan mengedit data diantaranya sebagai berikut:

- Cara melihat semua data pada tabel: `df_namatabelnya`
- Cara melihat kolom apa saja yang ada pada data: `df_namatabelnya.columns`
- Cara mencari satu kolom: `df_namatabelnya['kolomya']`
- Cara mencari multiple kolom: `df_household[['kolomnya', 'kolomya', 'kolomya']]`
- Cara mencari menggunakan `.iloc` merujuk pada lokasi index baris atau kolomnya sehingga harus integer: `df.iloc[baris, kolom]`
- Cara mencari menggunakan `.loc` merujuk pada lokasi nama baris atau kolomnya: `df.loc[baris, kolom]`
- Cara mencari sesuai kondisi yang dibutuhkan: `df[df['nama_kolom'] <comparison_operator> <value>].` Contoh nyata: `df_nama_tabel[(df_nama_tabel['nama_kolom'] == 'isi_kolom_yang_diinginkan') | (df_nama_tabel['nama_kolom'] >= angka_ingin_ditampilkan)]`
Contoh `comparison_operator` adalah seperti `==, !=, >, >=, <, <=`.



- Cara mengetahui banyaknya baris atau kolom `.shape[0]` = banyaknya baris, `.shape[1]` = banyaknya kolom : `df_nama_tabelnya.shape`
- Cara mengetahui tipe data pada kolom: `df_household.info()`
- Cara mengubah tipe data: `df_namatabel['kolomnya'] = df_namatabel['kolomnya'].astype('tipedata')`
- Cara menambahkan kolom baru: `df_namatabel['kolombaru'] = isi yang diinginkan`
- Cara menambahkan kolom baru dengan isian dari kolom yang sudah ada: `df_namatabel['kolombaru'] = df_namatabel['nama kolom yang isinya ingin dipakai']+df_namatabel['nama kolom yang isinya ingin dipakai ']`
- Cara menghapus kolom: `df_namatabel.drop('kolomnya', axis=1, inplace=True)`
- Cara menghapus baris: `df_namatabel([4,5,6], axis=0, inplace=True)`

Sebelum melakukan visualisasi data sebaiknya melakukan EDA dan CRISP-DM. Exploratory Data Analysis (EDA) adalah suatu pendekatan dalam ilmu data yang bertujuan untuk secara sistematis menyelidiki dan memahami sifat-sifat dasar dalam sebuah dataset. Sebelum memasuki tahap analisis yang lebih mendalam atau pembangunan model, EDA memberikan fondasi yang kokoh untuk memahami konteks dan karakteristik data yang dimiliki. Metode analisis ini menyediakan berbagai alat untuk meringkas dan memperoleh pengetahuan tentang sekumpulan data dengan cepat melalui penggunaan grafik sebagai visualisasi data (Wibowo et al., 2022). Langkah pertama dalam menganalisis data adalah mempelajari karakteristik data. Analisis data eksploratif sangat penting untuk membantu menelaah dan menemukan karakteristik data yang dapat membantu dalam pemilihan model yang tepat. Dua alat penting dalam EDA untuk menunjukkan karakteristik data adalah ringkasan numerik dan diagram kotak, atau boxplot. Ringkasan numerik menunjukkan dimana data memusat dan bagaimana sebaran datanya, sedangkan boxplot menunjukkan distribusi data, yang sangat penting untuk menentukan metode analisis statistika yang tepat untuk data.

Langkah-Langkah dalam EDA antara lain sebagai berikut ini:

- a) Ringkasan Statistik: Mengevaluasi statistik deskriptif seperti mean, median, kuartil, dan deviasi standar untuk memahami sentralitas dan variabilitas data.
- b) Visualisasi Data: Menggunakan berbagai grafik dan plot untuk menggambarkan distribusi data, relasi antar variabel, dan tren.
- c) Handling Missing Data: Mengidentifikasi dan menangani nilai-nilai yang hilang atau kosong dalam dataset.
- d) Deteksi Anomali: Meneliti outlier atau anomali yang dapat mempengaruhi hasil analisis.
- e) Analisis Korelasi: Mengukur korelasi antar variabel untuk memahami hubungan di antara mereka.
- f) Analisis Dimensi dan Transformasi: Mengeksplorasi dimensi data dan melakukan transformasi jika diperlukan.
- g) Segmentasi dan Kelompok: Menganalisis data berdasarkan kelompok tertentu untuk memahami variasi di dalamnya.

Manfaat dari EDA antara lain seperti dibawah ini:

- Meminimalkan Kesalahan Interpretasi: EDA membantu mengurangi risiko kesalahan interpretasi dengan memberikan pandangan yang komprehensif tentang data.
- Pemahaman Konteks: Memahami konteks di balik data membantu para analis membuat keputusan yang lebih informasional dan kontekstual.
- Identifikasi Potensi Masalah: EDA membantu mengidentifikasi potensi masalah atau pertanyaan-pertanyaan baru yang perlu diteliti.

Tujuan EDA yaitu sebagai berikut:

- Mendapatkan Pemahaman Awal: EDA membantu mengungkap informasi awal tentang data, termasuk distribusi, pola, dan karakteristik utama.
- Identifikasi Anomali: EDA membantu mendeteksi nilai-nilai yang tidak biasa atau anomali dalam dataset, yang dapat memengaruhi hasil analisis.
- Perencanaan Analisis Selanjutnya: EDA membimbing para analis data untuk merumuskan pertanyaan analisis yang lebih mendalam dan merencanakan pendekatan yang sesuai.



Exploratory Data Analysis merupakan langkah pertama yang krusial dalam penelitian data, membuka pintu menuju penemuan wawasan yang mendalam dan pemahaman yang lebih baik terhadap fenomena yang direpresentasikan oleh data.

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi yang banyak digunakan dalam penambangan data dan penelitian pembelajaran mesin. Ini adalah model proses yang berfungsi sebagai dasar untuk proses ilmu data. Metodologi ini digunakan untuk mengeksplorasi dan menemukan ide-ide baru atau hipotesis yang dapat diuji lebih lanjut. Model proses CRISP-DM terdiri dari enam fase berurutan: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penyebaran. Metodologi CRISP-DM telah digunakan dalam berbagai penelitian, seperti pengelompokan perguruan tinggi swasta, menentukan jalur penjualan, dan menerapkan data mining (Suhanda et al., 2020). Metodologi ini banyak digunakan karena menyediakan pendekatan terstruktur untuk penambangan data dan penelitian pembelajaran mesin. Ini membantu peneliti untuk mengidentifikasi masalah bisnis, memahami data, menyiapkan data, membangun model, mengevaluasi model, dan menyebarkan model. Sangatlah penting untuk memahami metodologi CRISP-DM dan bagaimana hal itu dapat digunakan untuk memandu penambangan data dan penelitian pembelajaran mesin. Dengan mengikuti enam fase berurutan dari model proses CRISP-DM, peneliti dapat memastikan bahwa penelitian mereka terstruktur, sistematis, dan efektif. Ini dapat membantu memastikan bahwa penelitian menghasilkan hasil berkualitas tinggi yang dapat digunakan untuk membuat keputusan berdasarkan data (Christian & Qi, 2022). Berikut adalah cara langkah-langkah CRISP-DM dapat dihubungkan dengan visualisasi data:

1. Business Understanding (Pemahaman Bisnis):

- Identifikasi kebutuhan visualisasi untuk memahami tren, pola, atau insight yang relevan dengan tujuan bisnis.
- Buat representasi visual konsep atau proses bisnis yang ingin dimengerti.

2. Data Understanding (Pemahaman Data):

- Visualisasikan distribusi variabel untuk memahami karakteristik data.
- Gunakan grafik untuk mengeksplorasi hubungan antar variabel.
- Identifikasi anomali atau pola menarik melalui visualisasi.

3. Data Preparation (Persiapan Data):

- Visualisasikan hasil pembersihan data dan transformasi untuk memastikan pemahaman yang baik tentang data yang telah dipersiapkan.
- Gunakan grafik atau visualisasi untuk mendeteksi outlier atau kesalahan data.

4. Modeling (Modeling):

- Visualisasikan hasil dari model pembelajaran mesin, seperti grafik karakteristik ROC atau confusion matrix.
- Gunakan visualisasi untuk membandingkan performa model yang berbeda.

5. Evaluation (Evaluasi):

- Gunakan visualisasi untuk membandingkan hasil prediksi model dengan nilai sebenarnya.
- Buat grafik atau diagram untuk menggambarkan kinerja model secara keseluruhan.

6. Deployment (Penerapan):

- Visualisasikan tren atau insight yang dihasilkan oleh model dan terapkan dalam konteks bisnis yang sesuai.

7. Documentation (Dokumentasi):

- Sertakan visualisasi dalam laporan dan presentasi untuk menjelaskan temuan dan keputusan yang diambil.
- Dokumentasikan jenis visualisasi yang paling efektif dalam menyampaikan informasi.
- Dokumentasi yang Mudah Dipahami: Dokumentasi yang dihasilkan membantu penggunaan kembali dan pemahaman proyek oleh orang lain dalam organisasi

Keuntungan Menggunakan CRISP-DM:

- Struktur yang Jelas: CRISP-DM menyediakan struktur yang jelas dan langkah-langkah yang dapat diikuti, membantu mengelola kompleksitas proyek data mining.
- Pentingnya Pemahaman Bisnis: Penekanan pada pemahaman bisnis membantu menjaga fokus proyek pada tujuan bisnis dan menghindari pengumpulan informasi yang tidak relevan.



- **Fleksibilitas dan Adaptabilitas:** Metodologi ini bersifat fleksibel dan dapat diadaptasi sesuai dengan kebutuhan proyek dan perubahan kebutuhan bisnis.
- **Mengatasi Masalah Kualitas Data:** CRISP-DM membantu mengatasi masalah kualitas data melalui tahap pemahaman dan persiapan data yang menyeluruh.
- **Peningkatan Keterlibatan Pemangku Kepentingan:** Dengan memahami bisnis dan menyajikan hasil analisis dengan jelas, CRISP-DM membantu meningkatkan keterlibatan pemangku kepentingan.
- **Pengambilan Keputusan yang Terinformasi:** Dengan fokus pada evaluasi model dan interpretasi hasil, CRISP-DM membantu organisasi membuat keputusan yang lebih terinformasi berdasarkan data.

Dengan menyematkan visualisasi data pada setiap tahap CRISP-DM, tim dapat memahami data dengan lebih baik, mengidentifikasi pola atau masalah, dan menyampaikan hasil dengan lebih efektif kepada pemangku kepentingan.

EDA (Exploratory Data Analysis) dan CRISP-DM (Cross-Industry Standard Process for Data Mining) dapat digunakan secara bersamaan dalam siklus hidup analisis data. Meskipun keduanya memiliki fokus dan tujuan yang berbeda, seringkali mereka saling melengkapi dalam proyek analisis data yang komprehensif. Berikut adalah cara di mana keduanya dapat diintegrasikan:

1. Memulai dengan EDA

- **Pemahaman Awal Data:** EDA dapat memberikan pemahaman awal yang mendalam tentang karakteristik dasar data, pola, dan anomali.
- **Formulasi Pertanyaan Analisis:** Hasil EDA dapat membantu merumuskan pertanyaan analisis yang lebih spesifik atau menemukan area kepentingan yang perlu lebih mendalam diteliti.

2. Integrasi dengan CRISP-DM:

- **Tahap Bisnis dan Pemahaman Data CRISP-DM:** Hasil EDA menjadi input berharga dalam tahap bisnis dan pemahaman data CRISP-DM. Pemahaman awal dari EDA membentuk dasar untuk memahami bagaimana data dapat mendukung tujuan bisnis.
- **Penentuan Variabel dan Pertimbangan Model:** Analisis EDA dapat membantu dalam penentuan variabel yang akan digunakan dalam model dan memberikan wawasan tentang hubungan antar variabel.
- **Evaluasi dan Peningkatan Model:** Hasil dari tahap pemodelan CRISP-DM dapat kembali dievaluasi menggunakan analisis EDA untuk memastikan bahwa model memahami dan merepresentasikan data dengan baik.

3. Siklus Iteratif:

- **Iterasi antara EDA dan CRISP-DM:** Proses analisis data seringkali bersifat iteratif. Hasil dari tahap CRISP-DM dapat memunculkan pertanyaan baru yang memerlukan eksplorasi lebih lanjut melalui EDA.
- **Penyempurnaan Model:** Jika hasil pemodelan menunjukkan kekurangan atau tidak sesuai dengan temuan dari EDA, model dapat disempurnakan atau dimodifikasi.

4. Akhir Proyek:

- **Presentasi dan Komunikasi Hasil:** Analisis EDA dapat digunakan untuk mendukung presentasi hasil proyek secara lebih intuitif dan dapat dimengerti oleh pemangku kepentingan yang mungkin tidak terlibat dalam aspek teknis CRISP-DM.

Dengan mengintegrasikan EDA dan CRISP-DM, tim analisis data dapat memanfaatkan kekuatan keduanya untuk mendapatkan pemahaman yang lebih holistik dan mendalam tentang data dan menghasilkan hasil analisis yang lebih efektif.

Visualisasi data adalah proses menampilkan data atau informasi dalam format visual seperti grafik, diagram, peta, atau bentuk visual lainnya. Tujuan utamanya adalah untuk membuat data lebih mudah dipahami dan dianalisis, memungkinkan untuk menemukan pola, tren, dan hubungan di antara dataset yang kompleks (Sitio et al., 2022). Ini dicapai melalui visualisasi data yang mudah dipahami, yang memudahkan pengambilan keputusan, dan memberi tahu orang lain tentang hasil analisis. Sesuai dengan jenis data dan informasi yang ingin disampaikan, metode visualisasi data dapat menggunakan berbagai jenis grafik, seperti bar chart, line chart, pie chart, heatmap, scatter plot, dan jenis grafik lainnya. Pemahaman, analisis, dan komunikasi data dataset bergantung pada visualisasi data (Harmaja et al., 2022). Sebagai platform komputasi cloud berbasis notebook, Google Colab telah memberi

pengguna kemampuan yang luar biasa untuk melihat data secara interaktif dan efisien. Dengan menggunakan visualisasi data, data dapat divisualisasikan secara visual. Ini membantu dalam mengungkap pola: memudahkan pemahaman pola, tren, dan hubungan yang mungkin tidak terlihat dalam data mentah, komunikasi yang efektif: membantu dalam menyampaikan hasil analisis dengan cara yang mudah dipahami dan menarik (Sitio et al., 2022). Dalam konteks visualisasi data, Google Colab dapat digunakan untuk menganalisis dan memvisualisasikan data, seperti data populasi. Platform ini menyediakan fitur pembelajaran mesin bawaan yang memudahkan pengguna untuk menganalisis dan memvisualisasikan data. Selain itu, Google Colab dapat digunakan untuk mengembangkan aplikasi web dan software yang berhubungan dengan data, sehingga memudahkan akses dan penggunaan informasi secara lebih efektif. Untuk melakukan pengolahan data perlu memasukan kode pengolahan dan nama data yang dimasukkan untuk memperoleh visualisasi data. Jika sudah memiliki data dalam bentuk file CSV atau Excel bisa menggunakan kode seperti dibawah ini untuk membuat diagram batang:

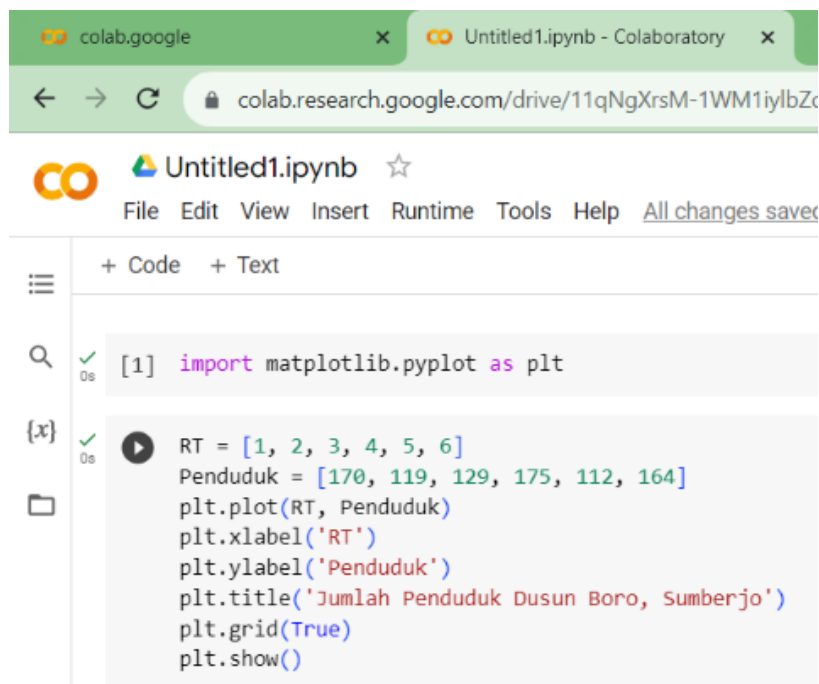
```
for col in ['namakolom','namakolom', 'namakolom','namakolom']:
```

```
    plt.figure(figsize=(8,4))
```

```
    sns.countplot(df[col])
```

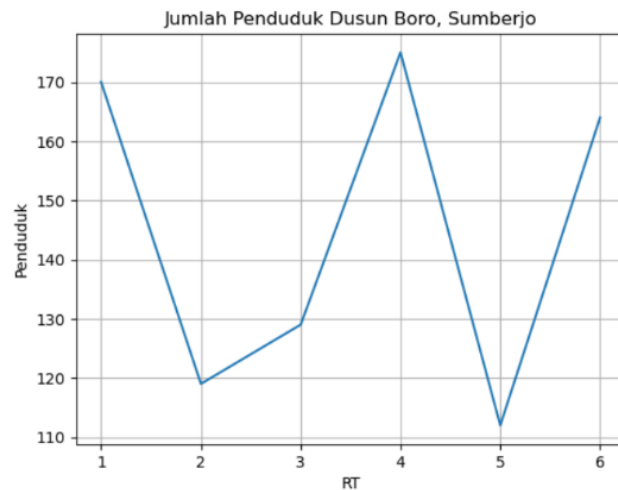
```
    plt.show()
```

Jika tidak memasukkan dokumen, bisa juga diinput secara langsung dengan cara seperti gambar di bawah ini untuk membuat garis line:



```
colab.google x Untitled1.ipynb - Colaboratory x
colab.research.google.com/drive/11qNgXrsM-1WM1iyIbZc
Untitled1.ipynb ☆
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[1] import matplotlib.pyplot as plt
RT = [1, 2, 3, 4, 5, 6]
Penduduk = [170, 119, 129, 175, 112, 164]
plt.plot(RT, Penduduk)
plt.xlabel('RT')
plt.ylabel('Penduduk')
plt.title('Jumlah Penduduk Dusun Boro, Sumberjo')
plt.grid(True)
plt.show()
```

Gambar 2. Kode membuat garis line
(Sumber: colab.research.google.com)



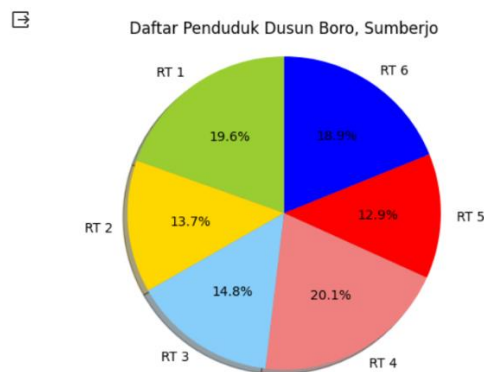
Gambar 3. Garis line jumlah penduduk per RT
(Sumber: Kantor Desa Sumberjo, Nganjuk)

```
labels = ['RT 1', 'RT 2', 'RT 3', 'RT 4', 'RT 5', 'RT 6']
quantity = [170, 119, 129, 175, 112, 164]
colors = ['yellowgreen', 'gold', 'lightskyblue', 'lightcoral', 'red', 'blue', 'yellow', 'olive']

plt.title('Daftar Penduduk Dusun Boro, Sumberjo')
plt.pie(quantity, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=90)

plt.axis('equal')
plt.show()
```

Gambar 4. Kode membuat diagram lingkaran
(Sumber: colab.research.google.com)



Gambar 5. Data diagram lingkaran jumlah penduduk per RT
(Sumber: Kantor Desa Sumberjo, Nganjuk)



Gambar 6. Kegiatan sosialisasi di Desa Sumberjo, Nganjuk
(Sumber: Dokumentasi kegiatan)

Setelah mengikuti pelatihan penggunaan python dan memahami setiap langkahnya, staf di kantor desa diharapkan untuk menerapkannya pada data-data yang ada di desa. Dan hasil dari kegiatan ini akan membantu staf yang ada di kantor desa Sumberjo dalam menjalankan pekerjaan mereka sebagai pegawai desa dengan mengolah dan memvisualisasikan data untuk mendapatkan informasi yang membantu mereka membuat keputusan.

SIMPULAN

Dari sosialisasi yang dilakukan di kantor Desa Sumberjo, Kabupaten Nganjuk ini meningkatkan pemahaman dan pengetahuan pegawai desa Sumberjo. Pengetahuan tentang kegunaan atau manfaat yang diperoleh dari python atau bahasa pemrograman yang saat ini sedang banyak dipakai untuk analisis dan visualisasi data pada google colab. Google colab ini tidak perlu mendownload dan bisa digunakan kapanpun dan jika sudah mempunyai data berupa excel bisa dimasukkan dan dianalisis di google colab juga. Sehingga ini bisa meringankan tugas pegawai desa untuk mengerjakan pekerjaan analisis dan visualisasi data penduduk desa. Maka dari itu diharapkan sosialisasi ini dapat diterapkan oleh pegawai desa untuk menghemat atau efisiensi waktu serta memperoleh visualisasi dan pengolahan data yang akurat.

DAFTAR PUSTAKA

- Christian, Y., & Qi, K. O. Y. R. (2022). Penerapan K-Means Pada Segmentasi Pasar Untuk Riset Pemasaran Pada Startup Early Stage Dengan Menggunakan Crisp-Dm. *Jurikom (Jurnal Riset Komputer)*, 9(4), 966. <https://doi.org/10.30865/Jurikom.V9i4.4486>
- Harmaja, O. J., Purba, W., Siregar, M. P., Manurung, H. T., & Sirait, F. A. (2022). Analisis Tingkat Kepuasan Pelanggan Pada Rudang Hotel Berastagi Menggunakan Metode Customer Satisfaction Index (Csi). *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 5(1), 147. <https://doi.org/10.37600/Tekinkom.V5i1.511>



- Husna, L., Eko, P., & Utomo, P. (2023). Analisis Dan Visualisasi Data Body Performance Menggunakan Tiga Tools Visualisasi Analysis And Visualization Of Body Performance Data Using Three Visualization Tools. *Jurnal Ilmiah Intech : Information Technology Journal Of Umus*, 5(01), 32–40.
- Junaidi, S., Devegi, M., & Kurniawan, H. (2023). *Pelatihan Pengolahan Dan Visualisasi Data Penduduk Menggunakan Python*. 4(1), 151–162. <https://doi.org/10.30812/adma.v4i1.2963>
- Prasetya, B. D., Pamungkas, F. S., & Kharisudin, I. (2019). Pemodelan Dan Peramalan Data Saham Dengan Analisis Time Series Menggunakan Python. *Prisma, Prosiding Seminar Nasional Matematika*, 3, 714–718. <https://journal.unnes.ac.id/sju/index.php/prisma/>
- Radhi, M., Ryan Hamonangan Sitompul, D., Hamonangan Sinurat, S., & Indra, E. (2021). Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima*, 4(2).
- Siregar, B., Evania, C. Della, & Yosia. (2023). Data Analysis Training Using Python At Jne. *Jurnal Pengabdian Masyarakat Bestari*, 2(2), 161–170. <https://doi.org/10.55927/jpmb.v2i2.3056>
- Sitio, A., Sindar, A., Marbun, M., Tiara, D., & Aswin, A. (2022). Pengenalan Data Scientist Pada Peserta Pkbn Al Habib Melalui Belajar Dasar Coding Python. *Jurnal Pengabdian Pada Masyarakat*, 7(1), 194–200. <https://doi.org/10.30653/002.202271.44>
- Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-Dm Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, 6(2), 12–20. <https://doi.org/10.37012/jtik.v6i2.299>
- Wibowo, A. H., Faisah, K., & Devianto, Y. (2022). Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada Pt. Telkominfra. *Jurnal Teknik Informatika Dan Sistem Informasi*, 9(3). <http://jurnal.mdp.ac.id>