**Prasyarat Kemampuan**

Sebelum mengikuti kelas ini, idealnya Anda sudah lulus kelas [Belajar Dasar Visualisasi Data](https://www.dicoding.com/academies/177) dan [Memulai Pemrograman Dengan Python](https://www.dicoding.com/academies/86). Namun, jika belum menyelesaikan kelas tersebut, Anda harus memiliki kemampuan seperti berikut.

* Memahami dasar-dasar statistik.
* Memahami dasar-dasar visualisasi data.
* Mengetahui berbagai tipe diagram pada visualisasi data beserta kegunaannya.
* Mengetahui berbagai teknik dalam membuat visualisasi data yang efektif.
* Memahami konsep dasar pemrograman menggunakan Python.
* Familier berbagai sintak dasar dalam bahasa pemrograman Python.
* Mampu menggunakan Google Colab, Jupyter Notebook, atau Python IDE lainnya untuk menulis dan menjalankan kode Python.

Nah, bagaimana menurut Anda? Sudah cukupkah bekal Anda untuk mengikuti kelas ini? Jika belum, kami sangat menyarankan Anda untuk mengikuti kelas [Belajar Dasar Visualisasi Data](https://www.dicoding.com/academies/177) dan [Memulai Pemrograman Dengan Python](https://www.dicoding.com/academies/86) terlebih dahulu sebelum lanjut ke materi berikutnya.

**Prasyarat Tools**

Selain kemampuan prasyarat yang sudah Anda ketahui di materi sebelumnya, kelas ini juga memiliki prasyarat tools yang perlu dipenuhi terlebih dahulu. Jika Anda sudah mengikuti seluruh latihan pada kelas Belajar Dasar Visualisasi Data dan Memulai Pemrograman dengan Python, seharusnya sudah familier dengan tools dan kebutuhan ini.

**Lingkungan Pengembangan**

Selama mengikuti kelas ini, tentu Anda akan banyak menulis kode untuk mengembangkan model machine learning. Berikut beberapa daftar lingkungan pengembangan yang dapat Anda pilih untuk digunakan di kelas ini.

* [Jupyter Notebook](https://jupyter.org/)
* [Google Colaboratory](https://colab.research.google.com/notebooks/intro.ipynb#recent=true)
* Integrated Development Environment (IDE), contohnya PyCharm atau Visual Studio
* Anaconda

Glossarium

Berikut adalah glosarium dengan istilah umum yang digunakan pada kelas ini. Anda dapat membaca sekilas materi berikut untuk mengenali istilah-istilah umum yang ada di modul kelas ini. Selain itu, Anda juga dapat mengunjungi kembali halaman ini setiap kali menemukan istilah yang belum dimengerti. Carilah istilah tersebut pada halaman glosarium ini untuk mengidentifikasi makna atau definisinya. Jika masih terdapat kosakata yang tidak Anda pahami dan belum masuk di daftar ini, Anda dapat memberikan saran melalui fitur Laporan Materi.

A

Analisis data

Analisis data merupakan sebuah proses mengumpulkan, mentransformasi, dan menata data untuk menarik kesimpulan, membuat prediksi, serta memberi pertimbangan yang tepat dalam mengambil keputusan.

Analytical skills

Analytical skills merupakan keterampilan yang berhubungan dengan kualitas dan karakteristik yang berhubungan dengan penyelesaian masalah menggunakan fakta.

Analytical thinking

Analytical thinking merupakan keterampilan yang melibatkan pengidentifikasian dan pendefinisian sebuah masalah, kemudian menyelesaikannya berdasarkan data dengan cara yang terorganisasi dan berurutan

Assessing data

Proses menilai kualitas dan struktur dari sebuah data untuk mengidentifikasi berbagai masalah yang terdapat dalam data, seperti missing value, unstandard value, dll.

B

Bias

Sebuah prasangka dan kecenderungan yang mendukung atau menentang suatu hal, individu, atau kelompok lain dengan cara yang kurang adil.

C

Cleaning data

Proses pembersihan data dari berbagai masalah yang dijumpai dalam tahap assessing data.

Confirmation bias

Bias ini muncul karena adanya kecenderungan kita dalam mencari atau menafsirkan informasi untuk mengonfirmasi keyakinan yang sudah ada sebelumnya.

Correlation

Parameter ini digunakan untuk mengidentifikasi korelasi atau hubungan dari dua feature numerik dalam sebuah data.

Covariance

Parameter ini digunakan untuk mengidentifikasi hubungan antar dua feature dalam sebuah dataset.

D

Data

Data merupakan sekumpulan fakta yang dapat direpresentasikan dalam berbagai bentuk seperti angka, gambar, video, teks, hasil pengukuran, dll.

Data analyst

Data analyst merupakan orang yang melakukan proses analisis data.

Data diskret

Data numerik yang hanya bisa direpresentasikan dalam bilangan bulat dan tidak dapat dibagi ke dalam unit yang lebih kecil.

Data ethics

Kumpulan acuan standar dalam menilai benar dan salah sebuah proses pengolahan data.

Data ink ratio

Perbandingan antara tinta (bisa diartikan sebagai elemen visual) yang digunakan untuk mendeskripsikan data dan total tinta yang digunakan dalam satu visualisasi data.

Data kategoris

Tipe data yang menggunakan kelas atau label untuk merepresentasikan kelompok dari suatu informasi.

Data kontinu

Data kuantitatif yang nilainya bisa dibagi atau diubah ke dalam unit yang lebih kecil.

Data kuantitatif

Tipe yang direpresentasikan dalam skala numerik sehingga memungkinkan kita untuk menjalankan operasi matematis.

Data nominal

Tipe data kategoris yang tidak memiliki urutan atau peringkat.

Data ordinal

Tipe data kategoris yang dapat diurutkan berdasarkan peringkat.

Data relationship

Parameter statistik untuk mengidentifikasi hubungan dua atau lebih feature/column/variable dalam sebuah data.

Data security

Upaya dalam menjaga dan melindungi informasi digital (data) dari berbagai pihak yang tidak bertanggung jawab dalam seluruh proses pengolahannya.

Data visualization

Data visualization merupakan tahapan yang harus kita lakukan sebelum membuat kesimpulan dan mengomunikasikan (draw conclusion & communicate) hasil dari proses analisis yang telah dilakukan.

Data wrangling

Data wrangling merupakan sebuah proses atau kumpulan kegiatan yang meliputi pengumpulan data (Gathering data), penilaian data (Assessing data), serta pembersihan data (Cleaning data) sebelum data digunakan dalam proses analisis data.

Data-driven decision making

Data-driven decision making merupakan sebuah kultur, best practice, serta proses dalam penggunaan data dan hasil analisis untuk memandu seluruh pengambilan keputusan bisnis yang tentunya sejalan dengan objektif perusahaan.

Dataset

Kumpulan data yang dikhususkan untuk menyelesaikan tugas tertentu.

Descriptive statistics

Descriptive statistics ialah kumpulan konsep statistik yang umum digunakan untuk mendeskripsikan sebuah data.

Distribusi data

Konsep statistik yang digunakan untuk menunjukkan frekuensi suatu nilai muncul dalam sebuah data.

Duplicate data

Masalah yang terjadi ketika terdapat sebuah observasi (semua nilai dalam satu unit baris) yang memiliki nilai yang sama persis pada setiap kolomnya.

E

Echo chamber

Keadaan yang membuat seseorang hanya menerima informasi dan opini yang sesuai dengan yang mereka percayai

Encryption

Teknik yang memanfaatkan algoritma tertentu untuk mengubah data menjadi bentuk unusable bagi individu atau aplikasi yang tidak memiliki akses terhadap algoritma tersebut.

Explanatory analysis

Proses analisis data yang bertujuan untuk membagikan beberapa insight yang menarik dari sebuah data

Exploratory Data Analysis

Exploratory data analysis merupakan tahap eksplorasi data yang telah dibersihkan guna memperoleh insight dan menjawab pertanyaan analisis.

F

Foreign key

Kolom yang berisi primary key dari tabel lain.

G

Gathering data

Proses mengumpulkan semua data yang dibutuhkan untuk menjawab semua pertanyaan atau masalah bisnis yang ingin kita hadapi.

I

Inaccurate value

Masalah yang muncul ketika nilai dalam sebuah data tidak sesuai dengan hasil observasi.

Inconsistent value

Masalah yang muncul ketika sebuah data memiliki nilai yang tidak konsisten baik dari segi satuan maupun ketentuan penilaian.

Inner join

Proses join yang hanya mengambil nilai yang bersesuaian di kedua tabel.

Interpretation bias

Bias yang terjadi karena adanya kecenderungan kita dalam menginterpretasikan situasi ambigu hanya ke dalam dua keadaan (hitam dan putih).

Interquartile range

Parameter statistik yang menggambarkan selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1).

Invalid value

Masalah ini muncul ketika terdapat beberapa nilai yang tidak masuk akal, tidak sesuai dengan ketentuan, dan background knowledge dari data tersebut.

K

Kredibilitas data

Bagian dari data quality assessment yang digunakan sebagai sebuah tolok ukur untuk menilai tingkat kepercayaan terhadap suatu data.

L

Left join

Proses join yang akan mengambil semua nilai dari tabel kiri beserta nilai yang bersesuaian dari tabel kanan.

Left-skewed distribution

distribusi data yang terjadi ketika sebagian besar populasi data berada pada bagian kanan.

Lie factor

Rasio perbandingan ukuran yang ditampilkan pada grafik dan ukuran yang sebenarnya ada dalam data.

M

Matplotlib

Matplotlib merupakan sebuah library Python yang reliable dan komprehensif untuk mendukung kebutuhan pembuatan visualisasi data.

Mean

Nilai yang diperoleh dari menghitung jumlah keseluruhan data dan dibagi dengan banyaknya data yang dimiliki.

Measuring asymmetric

Parameter statistik untuk menilai ketidaksimetrisan dalam sebuah distribusi data.

Measuring central tendency

Measuring central tendency adalah parameter statistik untuk mengidentifikasi nilai khas atau sentral dalam sebuah data.

Measuring dispersion

Parameter statistik untuk menilai variabilitas atau sebaran nilai dalam sebuah data.

Median

Parameter yang merepresentasikan nilai tengah atau persentil ke-50 dari keseluruhan observasi atau data.

Merge

Merge atau join merupakan teknik untuk menggabungkan dua tabel data menggunakan primary key dan foreign key.

Missing value

Masalah ini muncul karena adanya nilai yang hilang dari sebuah data dan biasanya direpresentasikan sebagai nilai NaN dalam library pandas.

Mode

Parameter ini akan memberikan kita gambaran atau informasi terkait nilai yang paling sering muncul dalam suatu data.

N

NumPy

NumPy (singkatan dari numerical Python) merupakan salah satu library Python yang sangat powerful untuk membuat dan mengolah multi-dimensional arrays (sering juga disebut sebagai matriks atau tensor).

O

Observer bias

Bias ini muncul karena adanya kecenderungan yang berbeda dari setiap individu ketika melakukan observasi.

Outer join

Sering juga disebut full outer join merupakan proses join yang akan mengambil semua nilai dari kedua tabel

Outlier

Titik data yang berada sangat jauh dari titik data yang lain dalam sebuah dataset.

P

Pandas

Pandas merupakan library Python yang spesifik digunakan untuk memanipulasi dan menganalisis data.

Primary key

Sebuah kolom dengan nilai unik yang merepresentasikan suatu data dalam sebuah tabel.

R

Range

Parameter yang digunakan untuk melihat perbedaan antara nilai maksimum dan minimum dari suatu data.

Right join

Proses join yang akan mengambil semua nilai dari tabel kanan beserta nilai yang bersesuaian dari tabel kiri.

Right-skewed distribution

Distribusi data yang memiliki sebagian besar populasi data yang terkonsentrasi pada bagian kiri

S

Sampling bias

Terjadi ketika sampel tidak mewakili populasi secara keseluruhan.

SciPy

SciPy merupakan library Python yang khusus digunakan untuk kebutuhan komputasi saintifik.

Seaborn

Seaborn merupakan library Python yang spesifik digunakan untuk membuat visualisasi data yang atraktif dan informatif.

Skewness

Parameter statistik yang digunakan untuk mengukur kesimetrisan sebuah distribusi data.

Standard deviation

Nilai akar kuadrat dari variance.

Statistik

Statistik merupakan ilmu yang menerapkan berbagai metode saintifik dalam proses pengumpulan, analisis, interpretasi, serta penyajian data.

Symmetric distribution

Distribusi data yang memungkinkan kita untuk membagi data menjadi dua bagian secara simetris serta memiliki nilai mean, median, dan mode yang sama.

T

Tokenization

Metode untuk mengubah elemen tertentu dalam sebuah data menjadi sebuah data random yang berperan sebagai token.

V

Variance

Parameter yang digunakan untuk menggambarkan besar simpangan suatu titik data dari nilai mean.

**Daftar Referensi**

[1] Domo, Inc. “Data Never Sleeps 9.0”. Diakses pada: 30 Oktober 2022. [Online]. Tersedia: [tautan](https://www.domo.com/learn/infographic/data-never-sleeps-9" \t "_blank)

[2] Google Career Certificates. (2021). Foundations: Data, Data, Everywhere. [Online]. Tersedia: [tautan](https://www.coursera.org/learn/foundations-data?specialization=google-data-analytics" \t "_blank)

[3] Tim Stobierski. “The Advantages of Data-Driven Decision-Making”. online.hbs.edu. Tersedia: [tautan](https://online.hbs.edu/blog/post/data-driven-decision-making" \t "_blank) (diakses pada 10 Desember 2022)

[4] Udacity. (2020). Nanodegree Program Become a Data Analyst. [Online]. Tersedia: [tautan](https://www.udacity.com/course/data-analyst-nanodegree--nd002" \t "_blank)

[5]  Navlani, Avinash et al.  “Python Data Analysis - Third Edition”. 2021. [Online]. Tersedia: [tautan](https://learning.oreilly.com/library/view/python-data-analysis/9781789955248/" \t "_blank). Diakses pada 12 Desember 2022

[6] CrowdFlower. “2016 Data Science Report”. CrowdFlower. 2016. [Online]. Tersedia: [tautan](https://www2.cs.uh.edu/~ceick/UDM/CFDS16.pdf" \t "_blank) (diakses pada 15 Desember 2022)

[7] Google Career Certificates. (2021). Prepare Data for Exploration. [Online]. Tersedia: [tautan](https://www.coursera.org/learn/data-preparation" \t "_blank)

[8] IBM. “What is data security?”. IBM. Tersedia: [tautan](https://www.ibm.com/id-en/topics/data-security" \t "_blank) (diakses pada 20 Desember 2022)

[9] McKinney, Wes. “Python for Data Analysis, 3rd Edition”. 2022. [Online]. Tersedia: [tautan](https://learning.oreilly.com/library/view/python-for-data/9781098104023/" \t "_blank) (diakses pada 20 Desember 2022)

[10]  Knaflic, Cole Nussbaumer. “Storytelling with Data: A Data Visualization Guide for Business Professionals”. 2015. [Online]. Tersedia: [tautan](https://learning.oreilly.com/library/view/storytelling-with-data/9781119002253/" \t "_blank) (diakses pada 10 January 2022)

[11] Google Career Certificates. (2021). Ask Questions to Make Data-Driven Decisions. [Online]. Tersedia: [tautan](https://www.coursera.org/learn/ask-questions-make-decisions" \t "_blank)

[12] Tufte, Edward Rolf. “The Visual Display of Quantitative Information”. 2007. [Online]. Tersedia: [tautan](http://faculty.salisbury.edu/~jtanderson/teaching/cosc311/fa21/files/tufte.pdf" \t "_blank) (diakses pada 2 Februari 2023)

[13] Google Career Certificates. (2021). Share Data Through the Art of Visualization. [Online]. Tersedia: [tautan](https://www.coursera.org/learn/visualize-data?specialization=google-data-analytics" \t "_blank)  
[14] Priddy, Sam. “Best practices for building effective dashboards”. Tableau. Tersedia: [tautan](https://www.tableau.com/blog/best-practices-for-building-effective-dashboards" \t "_blank) (diakses pada 25 Februari 2023)