

W5QUI11422021-
1740575066487

by Turnitin [™]

Submission date: 26-Feb-2025 02:06PM (UTC+0100)

Submission ID: 2599256767

File name: W5QUI11422021-1740575066487.docx (64.08K)

Word count: 673

Character count: 4681

LAPORAN PRAKTIKUM

Data Science



DISUSUN OLEH :

Lamria Sitorus – 11422021

STr TEKNOLOGI REKAYASA PERANGKAT LUNAK

IT DEL

SITOLUAMA

Dataset dalam Data Science

Pengertian Dataset Dataset adalah kumpulan data yang terorganisir dalam suatu format tertentu dan digunakan dalam berbagai analisis di bidang Data Science. Dataset dapat berbentuk tabel, gambar, teks, audio, atau jenis data lainnya yang dikumpulkan dan diolah untuk menghasilkan wawasan yang berguna. Dalam Data Science, dataset digunakan untuk eksplorasi data, penggalian pola, analisis statistik, visualisasi data, serta pengembangan model prediktif dan deskriptif.

Tujuan utama dari dataset dalam Data Science adalah untuk mendukung proses analisis data yang dapat menghasilkan informasi yang relevan bagi pengambilan keputusan. Dataset membantu dalam mengidentifikasi tren, menemukan anomali, melakukan segmentasi data, serta memahami hubungan antara berbagai variabel. Selain itu, dataset yang berkualitas dapat meningkatkan akurasi hasil analisis serta membantu menghindari bias dalam pengolahan data.

Jenis-Jenis Dataset Dataset dalam Data Science dapat diklasifikasikan ke dalam beberapa kategori berdasarkan karakteristiknya:

1. Numerical Dataset

Numerical dataset berisi data dalam bentuk ² angka yang dapat dihitung atau diukur. Data ini bersifat kuantitatif dan sering digunakan dalam analisis statistik serta pemodelan prediktif. Contoh: data suhu, berat badan, dan pendapatan individu.

2. Bivariate Dataset

Bivariate dataset terdiri dari dua variabel yang dianalisis untuk melihat hubungan atau korelasi di antara keduanya. Misalnya, hubungan antara jumlah jam belajar dan nilai ujian mahasiswa.

3. Multivariate Dataset

Multivariate dataset mencakup lebih dari dua variabel dalam satu set data. Analisis multivariat digunakan untuk memahami interaksi antar variabel. Contoh: dataset yang berisi informasi usia, jenis kelamin, pendapatan, dan tingkat pendidikan seseorang.

4. Categorical Dataset

Categorical dataset mengandung data dalam bentuk kategori atau kelompok. Data ini tidak bersifat numerik dan tidak dapat dihitung secara langsung. Contoh: jenis kelamin (laki-laki/perempuan), status pernikahan, dan warna mobil.

5. Correlation Dataset

Correlation dataset berisi informasi mengenai hubungan antar variabel dalam bentuk matriks korelasi. Nilai korelasi membantu dalam menentukan apakah dua variabel memiliki hubungan linier positif, negatif, atau tidak berkorelasi sama sekali.

Proses Pengelolaan Dataset dalam Data Science Dataset dalam Data Science harus dikelola dengan baik agar dapat digunakan secara efektif dalam analisis. Berikut adalah tahapan utama dalam pengelolaan dataset:

1. Pengumpulan Data

Data dikumpulkan dari berbagai sumber seperti database, API, sensor, survei, atau web scraping. Sumber data yang beragam memungkinkan analisis yang lebih komprehensif.

2. Pembersihan Data (Data Cleaning)

Data sering kali mengandung kesalahan, duplikasi, atau nilai yang hilang. Proses pembersihan mencakup penghapusan data yang tidak relevan, penanganan data yang hilang, dan normalisasi format data agar lebih seragam.

3. Eksplorasi dan Transformasi Data

Eksplorasi data dilakukan untuk memahami distribusi data, menemukan pola, dan mengidentifikasi outlier. Transformasi data melibatkan pengolahan data menjadi format yang lebih sesuai untuk analisis, seperti mengonversi teks menjadi angka atau menyamakan skala nilai.

4. Pembagian Dataset

Jika dataset digunakan untuk pemodelan, maka dapat dibagi menjadi beberapa bagian, seperti data pelatihan (training data), data validasi (validation data), dan data pengujian (testing data). Ini membantu dalam evaluasi model yang lebih akurat.

5. Visualisasi Data

Visualisasi digunakan untuk menampilkan pola dalam dataset dengan cara yang lebih mudah dipahami. Contoh teknik visualisasi meliputi grafik batang, diagram pencar (scatter plot), heatmap, dan diagram lingkaran.

6. Penyimpanan dan Pengelolaan Dataset

Dataset yang telah diproses disimpan dalam format yang sesuai seperti CSV, JSON, atau database SQL/NoSQL agar mudah diakses untuk analisis lebih lanjut. Untuk skala besar, penyimpanan dapat menggunakan Cloud Storage seperti Google Drive, AWS S3, atau layanan big data lainnya.

Secara keseluruhan, pengelolaan dataset dalam Data Science mencakup pengumpulan, pembersihan, eksplorasi, analisis, visualisasi, serta penyimpanan data. Dengan pengelolaan yang baik, dataset dapat digunakan untuk menghasilkan wawasan yang berharga dan membantu pengambilan keputusan berbasis data.

Contoh dataset

MNIST Dataset adalah kumpulan data yang berisi gambar angka tulisan tangan dari 0 hingga 9. Dataset ini terdiri dari 60.000 gambar untuk pelatihan dan 10.000 gambar untuk pengujian, dengan setiap gambar memiliki resolusi 28x28 piksel dalam format grayscale. MNIST sering digunakan sebagai benchmark dalam pengembangan model machine learning, terutama dalam tugas pengenalan pola dan computer vision. Karena kesederhanaannya, dataset ini menjadi salah satu yang paling populer untuk mempelajari dan menguji algoritma deep learning, seperti jaringan saraf tiruan dan convolutional neural networks (CNN).

Referensi

<https://books.google.co.id/books?hl=id&lr=&id=svXUDQAAQBAJ&oi=fnd&pg=PP1&dq=>

[pengertian+dataset+dalam+data+science&ots=0Ww2Hve_Wa&sig=ZUr0XNguQigl2p7_lik6Jm78h4g&redir_esc=y#v=onepage&q=pengertian%20dataset&f=false](#)

ORIGINALITY REPORT

7%

SIMILARITY INDEX

6%

INTERNET SOURCES

3%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to Universitas Sumatera Utara

Student Paper

2%

2

prezi.com

Internet Source

1%

3

pt.scribd.com

Internet Source

1%

4

www.dumax.org

Internet Source

1%

5

Doddi Yuniardi, Rani Puspita, Ridwan Ridwan.
"MENINGKATKAN EFISIENSI PENGENDALIAN
SUHU BOILER PADA PLTSa BURANGKENG
MELALUI ANALISIS KOMPREHENSIF BERBASIS
MACHINE LEARNING", Jurnal Teknik dan
Science, 2023

Publication

1%

6

www.biologie.uni-hamburg.de

Internet Source

1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On

