

W5-QUI-11422010

by Turnitin

Submission date: 26-Feb-2025 05:13AM (UTC+0400)

Submission ID: 2598305272

File name: BY2r596glnuwTjU9FM3F.docx (280.7K)

Word count: 771

Character count: 5117

Laporan Praktikum
4143204 Data Science

DATA CLASS



Disusun Oleh
David Kristian Silalahi
11422010

4
Sarjana Terapan Teknologi Rekayasa Perangkat Lunak

Fakultas Vokasi
Institut Teknologi Del

Data Class

1. Pendahuluan

Machine Learning (ML) adalah cabang dari *Artificial Intelligence* (AI) yang berfokus pada pengembangan sistem yang dapat belajar dari data serta membuat keputusan atau prediksi tanpa harus diprogram secara eksplisit. Dalam penerapan ML, data memiliki peran yang sangat penting karena menjadi dasar bagi model untuk mengenali pola dan hubungan antar variabel.

Salah satu konsep fundamental dalam ML adalah *data class*, yaitu pengelompokan data berdasarkan karakteristik tertentu. Konsep ini membantu model dalam memahami dan memproses informasi dengan lebih sistematis, terutama dalam tugas-tugas yang melibatkan klasifikasi dan pengelompokan data.

Konsep *data class* dalam ML dapat dikaji dari dua perspektif utama:

1. **Sebagai Representasi Data** – *Data class* digunakan untuk menyusun dan mengelola data dalam format yang lebih terstruktur, sehingga mempermudah proses analisis dan pemrosesan lebih lanjut.
2. **Sebagai Label dalam Klasifikasi** – Dalam *supervised learning*, *data class* berfungsi sebagai label atau kategori yang digunakan oleh model untuk belajar dan membuat prediksi terhadap data baru.

2. Konsep Data Class dalam Machine Learning

2.1 Data Class sebagai Representasi Data

Dalam pemrograman Python, konsep *data class* dapat diimplementasikan menggunakan modul *dataclasses*. Modul ini memungkinkan pengelolaan data dalam bentuk yang lebih terstruktur, mempermudah akses, serta meningkatkan keterbacaan kode. Penggunaan *data class* sangat berguna dalam pemrosesan data untuk keperluan *machine learning*, terutama dalam menyusun dataset yang rapi dan mudah dimanipulasi.

Secara umum, *data class* mengacu pada kategori atau label yang digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu. Dalam konteks *machine learning*, istilah ini sering dikaitkan dengan tugas klasifikasi, di mana model bertujuan untuk mengidentifikasi dan menetapkan label tertentu pada data masukan. Contoh umum dari penggunaan *data class* adalah dalam klasifikasi email, di mana sistem harus menentukan apakah sebuah email termasuk dalam kategori "spam" atau "bukan spam". Dalam kasus ini, "spam" dan "bukan spam" merupakan dua *data class* yang berbeda yang membantu model dalam membedakan dan mengenali pola dari fitur-fitur yang diberikan.

Selain dalam klasifikasi, konsep *data class* juga dapat diterapkan dalam berbagai tugas *machine learning* lainnya seperti regresi dan *clustering*. Dalam regresi, meskipun tidak secara eksplisit menggunakan *data class* sebagai label diskrit, data sering dikelompokkan berdasarkan atribut tertentu untuk mempermudah analisis. Sedangkan dalam *clustering*, data dikelompokkan ke dalam beberapa kelas tanpa label awal, memungkinkan model menemukan pola tersembunyi dalam dataset.

Dengan memanfaatkan *data class*, model *machine learning* dapat lebih mudah memahami hubungan antara variabel input (*features*) dan output (*labels*). Hal ini memungkinkan pengambilan keputusan yang lebih akurat dalam berbagai tugas prediktif, baik itu untuk keperluan pengklasifikasian, pengelompokan, maupun analisis data yang lebih kompleks.

Contoh Implementasi Data Class dalam Python

```
[1]: from dataclasses import dataclass

@dataclass
class DataPoint:
    feature1: float
    feature2: float
    label: int

# Contoh penggunaan
data = DataPoint(5.3, 2.1, 1)
print(data)

DataPoint(feature1=5.3, feature2=2.1, label=1)
```

Penjelasan:

- DataPoint merupakan sebuah kelas yang merepresentasikan satu titik data dengan dua atribut fitur (feature1, feature2) serta label (label).
- Dengan menggunakan dataclass, kita mendapatkan metode `__init__`, `__repr__`, dan `__eq__` secara otomatis tanpa harus menuliskannya secara eksplisit.

2.2 Data Class sebagai Label dalam Klasifikasi

1 Dalam supervised learning, model machine learning dilatih menggunakan dataset yang sudah dikategorikan (berlabel). Data class dalam konteks ini merujuk pada label yang digunakan untuk membedakan kategori data yang berbeda.

Misalnya, dalam klasifikasi kondisi daun mangga berdasarkan jenis penyakit, kita memiliki tiga kelas utama:

- **Kelas 0:** Daun sehat
- **Kelas 1:** Daun terkena jamur
- **Kelas 2:** Daun terkena bakteri

Struktur dataset dapat direpresentasikan sebagai berikut:

```
X = [[5.3, 2.1], [3.2, 1.7], [4.1, 2.3]] # Fitur (misalnya panjang & lebar daun)
y = [0, 1, 2] # Label kategori
```

Model machine learning kemudian akan belajar mengenali pola dari data ini dan menggunakannya untuk melakukan prediksi terhadap input yang baru.

3. Implementasi Data Class dalam Machine Learning

3.1 Data Class sebagai Struktur Penyimpanan Data Latih

```
[3]: from dataclasses import dataclass
    from typing import List

    @dataclass
    class Dataset:
        features: List[List[float]]
        labels: List[int]

    # Contoh dataset
    data = Dataset(
        features=[[5.3, 2.1], [3.2, 1.7], [4.1, 2.3]],
        labels=[0, 1, 2]
    )
    print(data)

    Dataset(features=[[5.3, 2.1], [3.2, 1.7], [4.1, 2.3]], labels=[0, 1, 2])
```

Pendekatan ini memudahkan penyimpanan dataset dalam format yang lebih terstruktur serta lebih mudah dikelola dalam pengolahan lebih lanjut.

3.2 Penggunaan Data Class dalam PyTorch

Framework deep learning seperti PyTorch sering kali menggunakan konsep dataset yang disusun menggunakan data class.

Contoh Custom Dataset dalam PyTorch

```
[9]: import torch
    from torch.utils.data import Dataset

    class LeafDataset(Dataset):
        def __init__(self, features, labels):
            self.features = torch.tensor(features, dtype=torch.float32)
            self.labels = torch.tensor(labels, dtype=torch.long)

        def __len__(self):
            return len(self.features)

        def __getitem__(self, idx):
            return self.features[idx], self.labels[idx]

    # Contoh penggunaan
    data_features = [[5.3, 2.1], [3.2, 1.7], [4.1, 2.3]]
    data_labels = [0, 1, 2]
    dataset = LeafDataset(data_features, data_labels)

    # Akses satu data
    print(dataset[0])

    C:\Users\DAVID\anaconda3\lib\site-packages\torch\utils\_pytree.py:185: FutureWarning: optree is installed but the version is too old to support PyTorch
    Dynamo in C++ pytree. C++ pytree support is disabled. Please consider upgrading optree using 'python3 -m pip install --upgrade 'optree>=0.11.0''.
    warnings.warn(
    (tensor([5.3000, 2.1000]), tensor(0))
```

Penjelasan:

- LeafDataset merupakan sebuah class yang memungkinkan dataset disusun dalam format yang kompatibel dengan DataLoader dalam PyTorch.
- Data dikonversi ke dalam bentuk tensor untuk mendukung efisiensi perhitungan selama training model deep learning.

4. Kesimpulan

Data class dalam machine learning memiliki dua kegunaan utama:

1. **Sebagai Representasi Data** – Memanfaatkan dataclass dalam Python untuk menyusun data dalam format yang lebih rapi.
2. **Sebagai Label Klasifikasi** – Digunakan dalam supervised learning untuk mengklasifikasikan data ke dalam kategori tertentu.

Dengan mengimplementasikan data class, pengelolaan dataset menjadi lebih terstruktur dan efisien, baik untuk algoritma machine learning tradisional maupun deep learning seperti yang digunakan dalam PyTorch.


5. Referensi

- ³ Géron, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, 2019.
- Dokumentasi Python: <https://docs.python.org/>
- Dokumentasi PyTorch: <https://pytorch.org/docs/>

Bukti Upload :


https://github.com/davidkristia/Kuis-Data-Science_11422010


Bukti Turnitin


Dupli Checker


Date: 25-02-2025

Plagiarism Scan Report

**3%**
Plagiarism

**0%**
Exact Match

**3%**
Partial Match

**97%**
Unique

Words	779
Characters	5812
Sentences	55
Paragraphs	54
Read Time	4 minute(s)
Speak Time	6 minute(s)

Content Checked For Plagiarism

Laporan Praktikum
4143204 Data Science

DATA CLASS

Disusun Oleh
David Kristian Silalahi
11422010
Sarjana Terapan Teknologi Rekayasa Perangkat Lunak


Matched Source

Similarity 6%

Title:[ageron/handson-ml3 - GitHub](#)

This project aims at teaching you the fundamentals of Machine Learning in python. It contains the example code and solutions to the exercises in the third edition of my O'Reilly book Hands-on Machine Learning with Scikit-Learn:

<https://github.com/ageron/handson-ml3>

Check By: Dupli Checker

ORIGINALITY REPORT

11 %	11 %	4 %	5 %
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.jktgadget.com Internet Source	4 %
2	artikelpendidikan.id Internet Source	3 %
3	www.coursehero.com Internet Source	2 %
4	jtiik.ub.ac.id Internet Source	1 %
5	geograf.id Internet Source	1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off