

Pemilihan Sampel Berdasarkan Jenis Spesies Bunga pada Dataset Iris

¹Muhammad Bayu Syuhada, ²Anisa Fitriyani, ³Ukasyah Muntaha, ⁴Johannes Krisjon Silitonga,
⁵Smertniki Javid Ahmedthian

Jurusan Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Lampung Selatan, Indonesia

Email: muhammad.122450007@student.itera.ac.id, anisa.122450019@student.itera.ac.id,
ukasyah.122450028@student.itera.ac.id, johannes.122450043@student.itera.ac.id,
smertniki.122450115@student.itera.ac.id

PENDAHULUAN

Pengambilan sampel merupakan langkah penting dalam proses analisis data dan berperan penting dalam mempengaruhi validitas dan reliabilitas hasil penelitian. Artikel ini menjelaskan proses pemilihan sampel yang sesuai untuk analisis dataset iris, terutama pada spesies bunga. Dataset ini mencakup tiga spesies bunga Iris yang masing-masing spesies memiliki karakteristik morfologi yang unik, yaitu Iris setosa, Iris virginica, dan Iris versicolor.

Dalam artikel ini, kami akan memperkenalkan konsep pemilihan sampel berdasarkan jenis spesies bunga pada dataset Iris. Strategi pengambilan sampel secara acak dan strategi penyortiran dipertimbangkan dalam pemilihan sampel. Selain itu, pemilihan sampel yang tepat membutuhkan pemahaman yang lebih mendalam mengenai perbedaan morfologis antara ketiga spesies bunga tersebut serta strategi yang sesuai dalam merepresentasikan keakuratan dari masing-masing jenis spesies bunga tersebut.

Teknik yang digunakan untuk pengambilan sampel ini menggunakan random sampling. Pengacakan sampel dilakukan dengan menggunakan bahasa pemrograman Python dengan memanfaatkan package seperti random untuk pengacakan yang efisien. Dengan package random akan mengilustrasikan bagaimana memilih sampel dengan pengacakan yang efektif dari dataset Iris yang mencakup tiga spesies bunga yang berbeda. Metode yang digunakan untuk pengacakan sampel dataset Iris adalah metode closure, dimana teknik pengacakan ini akan memastikan bahwa setiap elemen dalam populasi memiliki probabilitas yang sama untuk dipilih dalam sampel.

METODE

1. Module

Module adalah satu fungsi dalam bahasa pemrograman python yang berguna untuk memudahkan pengguna dalam membuat satu program, module ini merupakan salah satu bagian perangkat lunak. Banyak cara yang bisa dilakukan oleh perangkat lunak ini dalam membantu kita pengguna untuk membuat sebuah program sebagai contoh kita bisa import objek modules, costume import name, dan sebagainya. Dalam module kita bisa menjelajahi fungsi tersebut dengan menggunakan dir dan help, fungsi dir untuk menampilkan semua fungsi yang diimplementasikan di setiap module yang kita gunakan.

2. Closure

Dalam istilah umum dalam programing untuk deklarasi fungsi yang berada di dalam nested function (Prayogo, N.A 2024). Closure juga bisa disimpan kedalam variabel. Penerapan

closure juga bisa kita lakukan pada nilai argument sebagai parameter fungsi, tetapi parameter ini harus dengan sesuai kebutuhan.

Closure dalam python digunakan untuk panggilan balik untuk mengakses outer function variabel bahkan setelah outer function tersebut sudah tertutup menjadi variabel lokal. Penggunaan closure sangat berguna jika ingin memanggil variabel dan menghindari variable global yang menyembunyikan data serta dapat menjadi salah satu solusi yang sangat baik untuk beberapa contoh kasus mudah.

3. Metode Simple Random Sampling

Random Sampling atau biasa disingkat Random Sampling merupakan suatu cara pengambilan sampel dimana setiap anggota populasi diberikan opportunity (kesempatan) yang sama untuk terpilih menjadi sampel (Permadina Kanah Arieska, 2018). Dalam Random Sampling, setiap individu dalam populasi memiliki peluang yang sama untuk dipilih sebagai sampel, yang diharapkan akan menghasilkan hasil yang tidak bias dari total populasi. Metode Random Sampling adalah jenis pengumpulan sampel probabilitas yang digunakan dalam penelitian. Setiap individu dalam populasi memiliki peluang yang sama untuk dipilih sebagai sampel, yang diharapkan akan menghasilkan hasil yang tidak bias dari total populasi.

Teknik pengambilan Random Sampling dapat dilakukan dengan cara acak dari segmen kecil individu atau anggota dari keseluruhan populasi. Metode ini juga membutuhkan penamaan atau penomoran populasi target dan kemudian menggunakan beberapa jenis metode undian. Terdapat empat tipe teknik Random Sampling, yaitu pemilihan acak dari segmen kecil individu atau anggota dari keseluruhan populasi.

PEMBAHASAN

Program kali ini menggunakan tiga modul Python yakni pandas, random dan sklearn.datasets.

```
import pandas as pd
from sklearn.datasets import load_iris
import random
```

Gambar 1. Import Modul

Modul pandas(pd) digunakan untuk mengolah data. Pandas pada program kali ini menggunakan pd. Dataframe untuk membuat data frame dari data jenis spesies dataset Iris. Modul random memiliki fungsi untuk menghasilkan bilangan acak. Modul sklearn.datasets berfungsi untuk mengimpor dataset Iris.

```
# Muat dataset Iris
iris = load_iris()
iris_df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
iris_df['species'] = iris.target_names[iris.target]
```

Gambar 2. Memuat dataset Iris

Dataset Iris dimuat dengan menggunakan `load_iris()`. Data dari dataset Iris dimasukkan ke dalam data frame menggunakan `pd.DataFrame()`. Parameter adalah data fitur (atribut) dari dataset Iris, dan parameter `columns` diisi dengan nama-nama fitur yang diambil dari `iris.feature_names`. Kolom spesies ditambahkan ke data frame dengan menggunakan `iris_df['species'] = iris.target_names[iris.target]`. Nilai dari kolom ini adalah nama spesies yang sesuai dengan target dari dataset Iris. Target ini merupakan indeks dari nama spesies dalam array `iris.target_names`.

```
[9] #Fungsi Random Sampling
def data_sampler(df, replacement):
    def sample(sample_size):
        if replacement:
            return [random.choice(df.index) for _ in range(sample_size)]
        else:
            return random.sample(list(df.index), sample_size)

    return sample
```

Gambar 3. Fungsi Random Sampling

Fungsi `data_sampler` adalah fungsi yang digunakan untuk melakukan sampling acak dari indeks data frame. Fungsi ini mengembalikan fungsi `sample`. Fungsi `sample` akan menggunakan `random.choice()` jika `replacement` adalah `true`, yang berarti sampel diambil dengan pengulangan (sampel bisa duplikat) dan menggunakan `random.sample()` jika `replacement` adalah `false`, yang berarti sampel diambil tanpa pengulangan (sampel harus unik).

```
# Contoh penggunaan pada spesies Setosa dengan sampel acak sebanyak 5
sampel_iris = data_sampler(iris_df[iris_df['species'] == 'setosa'], replacement=True)
sampled_final = sampel_iris(5)
print("Sampel Acak:", sampled_final)
print("Dataframe Sampel Acak:")
print(iris_df.loc[sampled_final])
```

Gambar 4. Penggunaan Random Sampling pada spesies Setosa

Fungsi `data_sampler` digunakan di sini untuk membuat sampel acak dari spesies Setosa dalam DataFrame `iris_df`. Setelah itu mengambil sampel acak sebanyak lima pengulangan.

```
# Contoh penggunaan pada spesies Versicolor dengan sampel acak sebanyak 5
sampel_iris = data_sampler(iris_df[iris_df['species'] == 'versicolor'], replacement=False)
sampled_final = sampel_iris(5)
print("Sampel Acak:", sampled_final)
print("Dataframe Sampel Acak:")
print(iris_df.loc[sampled_final])
```

Gambar 5. Penggunaan Random Sampling pada spesies Versicolor

Sama seperti sebelumnya, hal yang membedakan disini adalah membuat sampel acak dari spesies Versicolor sebanyak lima kali tanpa diulang.

```
# Contoh penggunaan pada spesies Virginica dengan sampel acak sebanyak 5
sampel_iris = data_sampler(iris_df[iris_df['species'] == 'virginica'], replacement=False)
sampled_final = sampel_iris(5)
print("Sampel Acak:", sampled_final)
print("Dataframe Sampel Acak:")
print(iris_df.loc[sampled_final])
```

Gambar 6. Penggunaan Random Sampling pada spesies Virginica

Kode diatas menjelaskan mengenai sampel spesies Virginica diacak sebanyak lima kali tanpa adanya pengulangan.

KESIMPULAN

Berdasarkan percobaan yang sudah dilakukan, dapat disimpulkan bahwa pengacakan sampel pada dataset Iris bisa dilakukan menggunakan Closure pada Python. Penerapan python closure membantu membagi data lebih mudah dan efisien. Pengacakan sampel pada dataset Iris ini dilakukan untuk membagi dataset yang ada menjadi tiga berdasarkan spesies bunga yaitu setosa, versicolor dan virginica. Random sampling membantu memilih data mana yang akan digunakan karena dengan pengacakan sampel setiap data memiliki probabilitas yang sama untuk dipilih dan ini akan mengurangi terjadinya bias data.

DAFTAR PUSTAKA

Nosuke(2013.Modules Python.Bagaimana fungsi dan cara penggunaanya ?.TUTORIAL PYTHON 13.CODEKEY

Prayogo, N. A. (n.d.). Python Closure | Dasar Pemrograman Python. Dasar Pemrograman Python.

Permadina Kanah Arieska, ,. N. (2018). PEMILIHAN TEKNIK SAMPLING BERDASARKAN PERHITUNGAN EFISIENSI RELATIF. *Statistika*, 2.