

Analisis Cuaca di Szeged Tahun 2006- 2016 Menggunakan Model Naïve Bayes

December 13, 2022



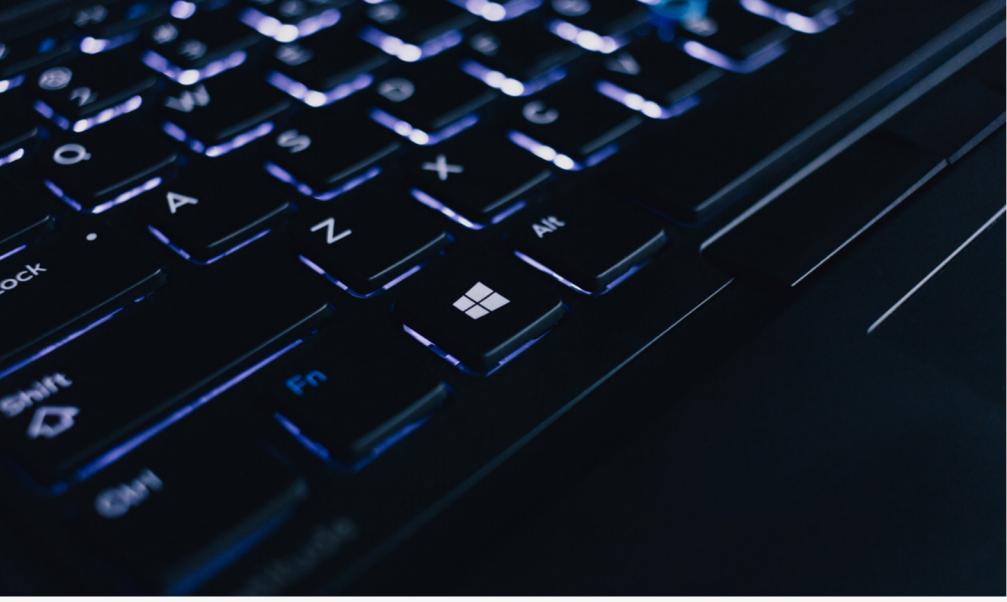
Nama Anggota:

- Ken Dahana - 160420115
- Viqram Ananta Wataf - 160420119



Latar Belakang

Peramalan cuaca adalah proses teknologi dan ilmu pengetahuan yang digunakan untuk memprediksi kondisi atmosfer untuk lokasi yang diberikan. Ilmuwan juga telah berusaha memprediksi cuaca secara formal dan informal. Analisis cuaca di Szeged menggunakan model naive bayes merupakan sebuah studi yang bertujuan untuk mengklasifikasikan cuaca di Szeged berdasarkan data historisnya.



Rumusan Masalah

Prakiraan cuaca telah menjadi yang paling menonjol di antara yang paling masalah yang menyusahkan secara eksperimental dan teknologi dunia pada abad terakhir. Lingkungan Perubahan telah mencari banyak pertimbangan sejak lama karena perubahan mendadak yang terjadi. Ada beberapa batasan dalam pelaksanaan cuaca yang lebih baik meramalkan sehingga akhirnya sulit memprediksi cuaca di sini dan sekarang dengan efektivitas

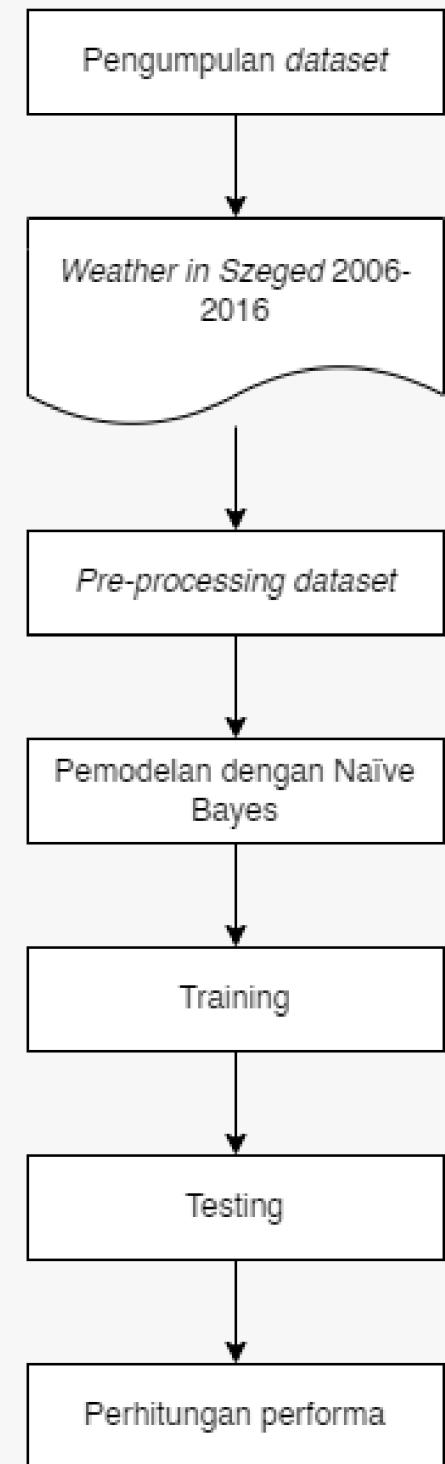


METODE PENELITIAN

Terdapat lima tahapan yang dilakukan dalam penelitian menggunakan model naive bayes.

- Pengumpulan Dataset
- Pre-processing Dataset
- Pembentukan Model
- Training dan Testing
- Perhitungan Performa

Gambar disamping, disajikan diagram blok alur proses penelitian secara lebih detail.



Gambar 1. Diagram Blok Alur Proses Penelitian
Naive Bayes



METODE PENELITIAN

A. Pengumpulan *Dataset*

Tahap awal dalam penelitian ini adalah Pengumpulan dataset. Pada tabel 1 disajikan jumlah data dan fitur yang ada dalam dataset.

Jumlah Data	Fitur/Atribut
96453	12

Tabel 1. Fitur-fitur yang tersedia di *dataset*

Sedangkan pada tabel 2 disajikan visualisasi detail data pada dataset yang digunakan untuk penelitian ini.

No	Nama	Class
1	Summary	Di slide berikutnya
2	Precip Type	Rain, Snow, Hail

Tabel 2. Fitur-Fitur Pada Dataset

ISI CLASS DARI SUMMARY

Partly Cloudy, Mostly Cloudy, Overcast, Foggy, Breezy and Mostly Cloudy, Clear, Breezy and Partly Cloudy, Breezy and Overcast, Humid and Mostly Cloudy, Humid and Partly Cloudy, Windy and Foggy, Windy and Overcast, Breezy and Foggy, Windy and Partly Cloudy , Breezy, Dry and Partly Cloudy, Windy and Mostly Cloudy' Dangerously Windy and Partly Cloudy, Dry, Windy, Humid and Overcast, Light Rain, Drizzle, Windy and Dry, Dry and Mostly Cloudy, Breezy and Dry, dan Rain.

METODE PENELITIAN

B. Pembuatan Model dan Training-Testing

Dengan menggunakan data ini dengan data testing, dapat memperkirakan kondisi cuaca. Model training set pelatihan dan pengetahuan itu digunakan sebagai data testing untuk memprediksi nilai. Jumlah data yang digunakan untuk data training adalah 70% sedangkan sisanya digunakan untuk data testing.

Berikut detail konfigurasi environment perangkat keras yang dipakai selama uji coba.

Parameter	Spesifikasi
CPU	Intel(R) Core(TM) i7-10750H CPU @2.60GHz 2.59 GHz
RAM	16 GB
Sistem Operasi	Windows 10
GPU	Intel(R) UHD Graphics

Tabel 3. Spesifikasi Perangkat Keras

METODE PENELITIAN

C. Perhitungan Performa

Pada tahapan ini, dilakukan proses perhitungan metriks performa Naïve Bayes yang didasarkan pada Teorema Bayesian. Persamaan (1) menunjukkan perhitungan atau rumus dari teorema bayesian.

Persamaan (2) menunjukkan dari masalah yang akan di klasifikasikan.

Persamaan (3) menunjukkan rumus dari perhitungan berdasarkan asumsi naïve independen.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$$P(y = c_k|x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_n|c_k)P(y=c_k)}{P(x_1, x_2, \dots, x_n)} \quad (2)$$

$$P(y = c_k|x_1, x_2, \dots, x_n) = \frac{P(y=c_k) \prod_{i=1}^n P(x_i | y=c_k)}{P(x_1, x_2, \dots, x_n)} \quad (3)$$



METODE PENELITIAN

C. Perhitungan Performa

Persamaan (4) menunjukkan rumus dari perkiraan probabilitas menggunakan MLE (maximum likelihood estimator).

Persamaan (5) menunjukkan perhitungan dari naïve bayes training untuk fitur kategori.

Persamaan (6) menunjukkan perhitungan dari Gaussian Naïve Bayes.

$$P(x_i = v | y = c_k) = \frac{\text{Count}(x_i=v \wedge y=c_k)}{\text{Count}(y=c_k)} \quad (4)$$

$$P(x_i = v | y = c_k) = \frac{\text{Count}(x_i=v \wedge y=c_k)+1}{\text{Count}(y=c_k)+n_i} \quad (5)$$

$$P(x_i | y = c_k) = \frac{1}{\sqrt{2\pi}\sigma_{c_k}} \exp\left(-\frac{(x_i - \mu_{c_k})^2}{2\sigma_{c_k}^2}\right) \quad (6)$$

PADA VARIABEL PERCIPITATION TYPE

HASIL DAN PEMBAHASAN

Pada pengategorian variabel Percipitation Type, kita akan mengodekan setiap nilai string yang ada menjadi integer sehingga bisa dikenali dan diproses oleh library Gaussian Naive Bayes

```
from sklearn.model_selection import train_test_split
import pandas as pd
weathernp = filtered_weather.to_numpy()
percipnp = weather['Precip Type'].to_numpy()
print(percipnp)
unique, codedpercipnp = np.unique(percipnp, return_inverse=True)
print(codedpercipnp)

[979] ✓ 0.9s
...
... ['rain' 'rain' 'rain' ... 'rain' 'rain' 'rain']
[1 1 1 ... 1 1 1]
```

Gambar 2. Mengodifikasi Percipitation Type Sehingga Berbentuk Angka

HASIL DAN PEMBAHASAN

Setelah mengodifikasi, kita harus membagi bagian dataset mana yang akan menjadi training dataset dan bagian mana yang akan menjadi testing dataset. Pada penelitian kali ini, kami memilih perbandingan 60%-40%. Bagian 60% akan dipakai untuk training model Gaussian Naive Bayes dan 40% akan dipakai untuk testing model.

```
xtrain, xtest, ytrain, ytest = train_test_split(weathernp, codedpercipnp, test_size=0.4, random_state=1)
8]    ✓ 0.8s
```

Gambar 3. Pembagian dataset yang akan dipakai sebagai training dan testing

HASIL DAN PEMBAHASAN

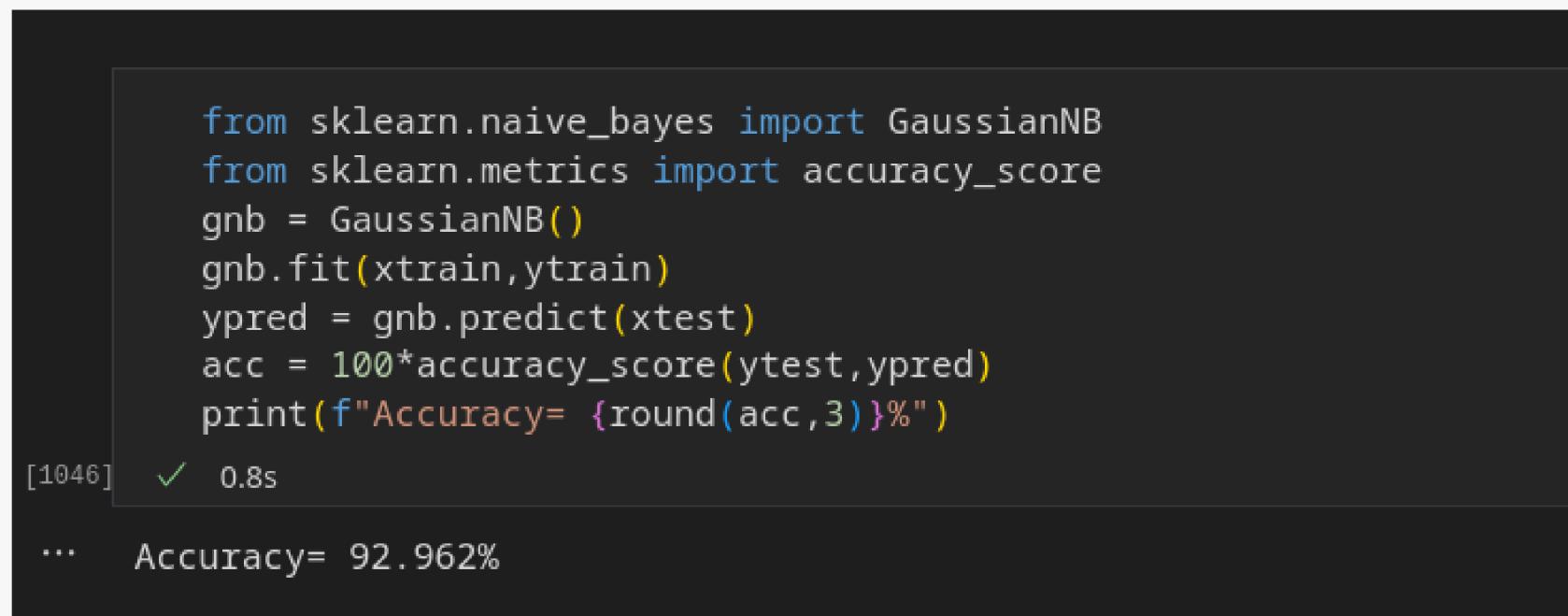
Kemudian, cari hubungan nilai antara variabel-variabel yang ada dengan Percipitation Type untuk mengetahui seberapa berat suatu variabel mempengaruhi hasil. Karena tidak ada nilai-nilai yang diabsolutkan disini kurang dari 1 maka seluruh fitur dinyatakan sangat berpengaruh terhadap hasil

```
for i in filtered_weather.columns:  
    print(f'Correlation between {i} and Precip Type class: {np.corrcoef(filtered_weather[i], codedpercipnp)[0,1]}')  
[1047] ✓ 0.5s  
... Correlation between Temperature (C) and Precip Type class: -0.5422453613181669  
Correlation between Apparent Temperature (C) and Precip Type class: -0.5452639564269441  
Correlation between Humidity and Precip Type class: 0.22389594188257944  
Correlation between Wind Speed (km/h) and Precip Type class: -0.06831941192287515  
Correlation between Wind Bearing (degrees) and Precip Type class: -0.040390018975976556  
Correlation between Visibility (km) and Precip Type class: -0.29199925464572113  
Correlation between Pressure (millibars) and Precip Type class: 0.006472496786503377
```

Gambar 4. Menghitung Korelasi Antara Semua Variable Dengan Percipitation Type

HASIL DAN PEMBAHASAN

Setelah melakukan pengkodean dan membagi data mana saja yang menjadi test dan train, maka akan diproses perhitungannya menggunakan gnb.fit dan gnb.predict. Kemudian akan dihitung keakurasiannya menggunakan gnb.score. Pada penelitian kali ini, hasil yang didapatkan cukup akurat dalam memprediksi jenis hujan apa yang cocok jika data-data diberikan



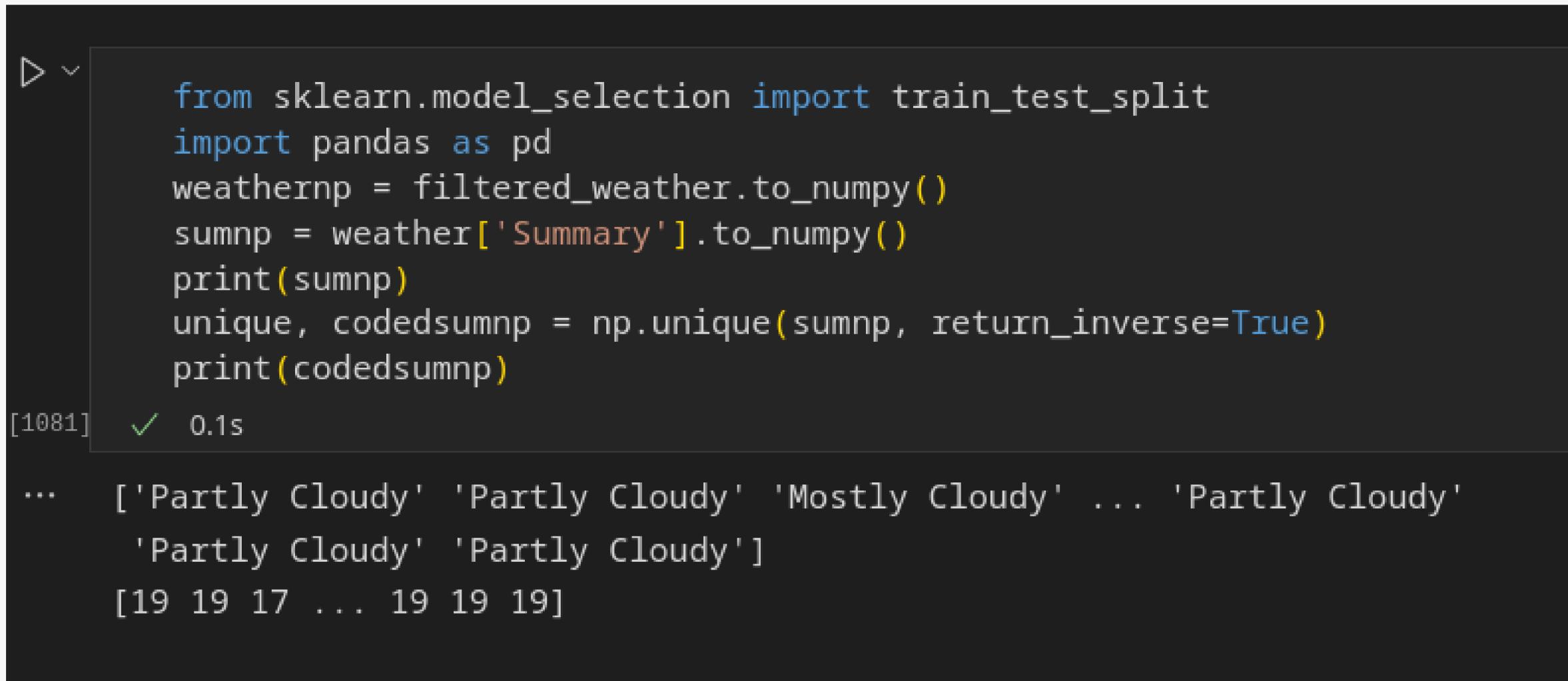
```
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
gnb = GaussianNB()
gnb.fit(xtrain,ytrain)
ypred = gnb.predict(xtest)
acc = 100*accuracy_score(ytest,ypred)
print(f"Accuracy= {round(acc,3)}%")
[1046] ✓ 0.8s
... Accuracy= 92.962%
```

Gambar 5. Menghitung Gaussian Naive Bayes dan Mengevaluasi Keakurasianya

PADA VARIABEL SUMMARY

Prosesnya mirip dengan Percipitation Type

HASIL DAN PEMBAHASAN



```
from sklearn.model_selection import train_test_split
import pandas as pd
weathernp = filtered_weather.to_numpy()
sumnp = weather['Summary'].to_numpy()
print(sumnp)
unique, codedsumnp = np.unique(sumnp, return_inverse=True)
print(codedsumnp)

[1081] ✓ 0.1s
...
['Partly Cloudy' 'Partly Cloudy' 'Mostly Cloudy' ... 'Partly Cloudy'
 'Partly Cloudy' 'Partly Cloudy']
[19 19 17 ... 19 19 19]
```

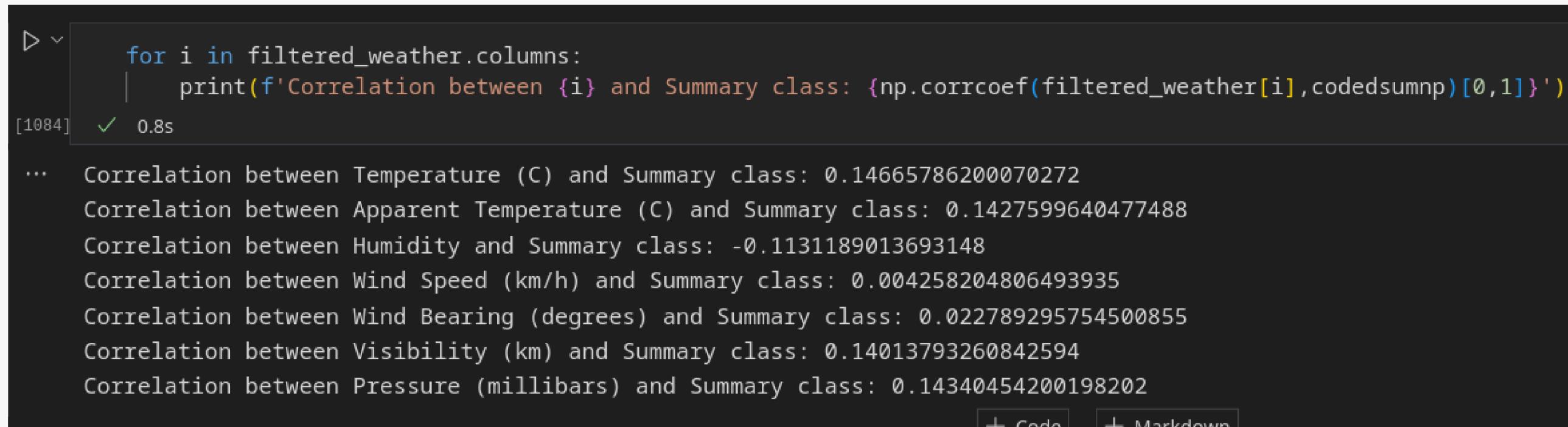
Gambar 6. Mengodifikasi Summary Sehingga Berbentuk Angka

HASIL DAN PEMBAHASAN

```
xtrain, xtest, ytrain, ytest = train_test_split(weathernp, codedsumnp,test_size=0.4, random_state=1)
3] ✓ 0.6s
```

Gambar 7. Pembagian dataset yang akan dipakai sebagai training dan testing

HASIL DAN PEMBAHASAN

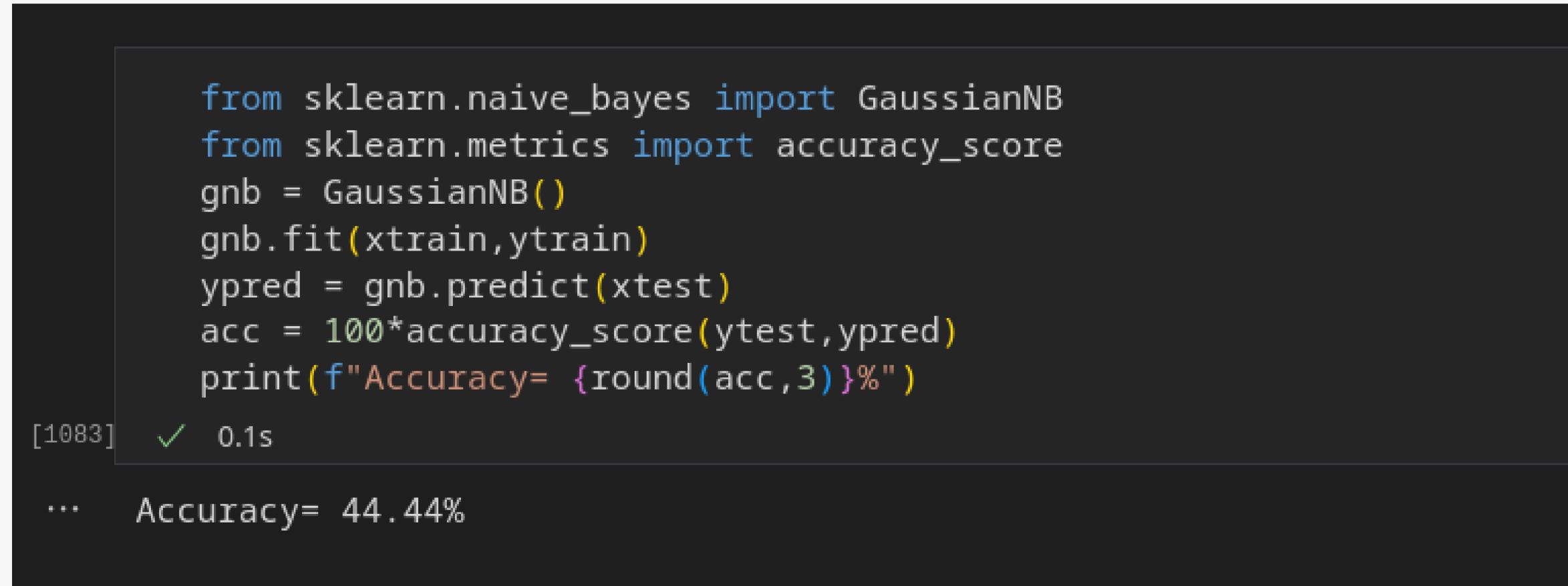


```
for i in filtered_weather.columns:  
    print(f'Correlation between {i} and Summary class: {np.corrcoef(filtered_weather[i], codedsumnp)[0,1]}')  
[1084] 0.8s  
... Correlation between Temperature (C) and Summary class: 0.14665786200070272  
Correlation between Apparent Temperature (C) and Summary class: 0.1427599640477488  
Correlation between Humidity and Summary class: -0.1131189013693148  
Correlation between Wind Speed (km/h) and Summary class: 0.004258204806493935  
Correlation between Wind Bearing (degrees) and Summary class: 0.022789295754500855  
Correlation between Visibility (km) and Summary class: 0.14013793260842594  
Correlation between Pressure (millibars) and Summary class: 0.14340454200198202
```

Gambar 8. Menghitung Korelasi Antara Semua Variable Dengan Summary

HASIL DAN PEMBAHASAN

Saat kami menghitung hasil keakurasi model ini, kami dapatkan nilainya cukup rendah sehingga model tidak dapat diandalkan untuk memprediksi summary apa yang cocok jika data-data diberikan.



```
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import accuracy_score
gnb = GaussianNB()
gnb.fit(xtrain,ytrain)
ypred = gnb.predict(xtest)
acc = 100*accuracy_score(ytest,ypred)
print(f"Accuracy= {round(acc,3)}%")
[1083] ✓ 0.1s
... Accuracy= 44.44%
```

Gambar 9. Menghitung Gaussian Naive Bayes dan Mengevaluasi Keakurasiannya



Kesimpulan

Hasil analisis menunjukkan bahwa ada hubungan yang signifikan antara variabel cuaca dengan variabel independen lainnya, sehingga dapat digunakan untuk memprediksi perubahan cuaca di Szeged. Prediksi cuaca dapat dilakukan menggunakan model Gaussian Naive Bayes. Pada variabel percipitation type, dapat diprediksi secara akurat akan memiliki jenis hujan yang seperti apa. Namun, pada variabel Summary, tidak dapat diprediksi secara akurat dikarenakan banyaknya value yang ada untuk model yang terbatas. Dengan demikian model gaussian naive bayes menggunakan variabel percipitation type cukup akurat memprediksi perubahan cuaca di Szeged.

Terima Kasih.

Apakah ada pertanyaan?

