

Assignment 01

Python/AI Refresher



FAKULTAS
ILMU
KOMPUTER

Review Python

Pada pengerjaan assignment/tugas kali ini (dan seterusnya), bahasa Python akan digunakan secara intensif. Bagi yang ingin melakukan review/refresh pengetahuan tentang Python, silakan mengakses playlist video YouTube Pemrograman Python (gratis) di:



Tautan: <https://s.id/pymooc-ui>

Selain itu, terdapat pula OpenCourseWare (OCW) UI untuk Dasar-Dasar Pemrograman 1, yang berisi slides, video, lab, tugas pemrograman, dan contoh soal UTS/UAS DDP 1. Selamat menikmati:



Tautan: <https://ocw.ui.ac.id/course/view.php?id=142>

Review kNN (k-Nearest Neighbor)

Sesuai dengan namanya (nearest neighbors), algoritma ini mengklasifikasikan suatu data baru dengan melihat kedekatan data tersebut terhadap data lama yang sudah memiliki label. Contoh, kamu mempunyai dataset yang berisi informasi rumah dengan label rumah mewah dan rumah tidak mewah. kNN akan menyimpan informasi mengenai ukuran yang serupa, seperti kemiripan luas rumah dan juga harga rumah. Ketika data rumah baru masuk, kNN akan mengecek kesamaan informasi data rumah tersebut terhadap data lama yang telah dipelajari kNN. Hasil pengecekan akan dijadikan sebagai kesimpulan: rumah baru tersebut masuk ke dalam kelas yang mana.

Apa saja yang perlu kita perhatikan sebelum mengimplementasikan kNN?

- Jumlah k: Jumlah k sangat menentukan performa kNN. k merepresentasikan jumlah data point terdekat, yang akan digunakan untuk menentukan suatu data point masuk ke dalam kategori apa. k yang kecil rentan dengan noise yang dapat membuat model overfitting, sedangkan k yang besar dapat merusak aspek lokalitas yang berujung ke penurunan performa model klasifikasi (underfitting).
- Rumus jarak yang digunakan: Rumus jarak yang digunakan untuk menghitung kedekatan data point terhadap data yang lainnya juga sangat menentukan performa dari algoritma kNN. Jika perhitungan jarak yang digunakan kurang sesuai, kalkulasi similarity akan tidak tepat dan berujung membuat klasifikasi menjadi tidak sesuai.

Ada beberapa rumus jarak yang biasa digunakan dalam kNN, di antaranya adalah:

- Manhattan distance:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

- Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}$$

Sebagai tambahan, berikut materi kNN yang pernah diajarkan sebelumnya agar Anda dapat mempelajari kNN secara lebih dalam. Silakan akses materi [di sini](#).

Berikut contoh penggunaan kNN untuk klasifikasi menggunakan library scikit-learn:

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import accuracy_score

# Load dataset iris
```

```
iris = load_iris()
X = iris.data
y = iris.target

# Split dataset menjadi training dan testing dengan proporsi 80%
training dan 20% testing
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Buat model kNN untuk klasifikasi
knn_classifier = KNeighborsClassifier(n_neighbors=3)

# Fit model dengan data training
knn_classifier.fit(X_train, y_train)

# Prediksi data testing
y_pred = knn_classifier.predict(X_test)

# Hitung akurasi model
accuracy = accuracy_score(y_test, y_pred)
print("Accuracy:", accuracy)
```

Tugas: Klasifikasi dan regresi dengan kNN from scratch



Tautan:

<https://www.cnnindonesia.com/gaya-hidup/20190822091453-262-423633/4-alasan-kerang-a-balon-dikenal-dengan-harga-selangit>

Pada tugas ini, Anda akan menggunakan dataset Abalone (dataset.csv) yang sudah disesuaikan dengan keperluan pada tugas kali ini. Silakan download dataset [di sini](#).

Datasetnya terdiri dari:

- 8 atribut
 - Length – Panjang dari cangkang kerang abalone

- Diameter – Diameter dari kerang abalone
- Height – Tinggi dari abalone
- Whole weight – Berat keseluruhan dari kerang abalone
- Shucked weight – Berat dari daging kerang abalone
- Viscera weight – Berat dari organ dalam kerang abalone
- Shell weight – Berat kerang abalone yang sudah dikeringkan
- Class – Kelas dari kerang abalone
- 4177 baris data

Untuk task classification, Anda dibebaskan memilih atribut yang akan digunakan untuk menentukan class dari kerang abalone. Untuk task regression, Anda diharuskan untuk menggunakan atribut “Length” dan “Height” dalam menentukan “Whole weight” dari kerang abalone.

Anda akan diberikan kode class kNN yang harus Anda lengkapi sendiri. Anda diperbolehkan menambah fungsi atau parameter lain yang dibutuhkan. Class ini WAJIB digunakan untuk mengimplementasikan kNN, baik untuk classification maupun regression. Apabila tidak menggunakan class ini, jawaban Anda TIDAK AKAN DINILAI. Silakan download template [di sini](#).

Anda diperbolehkan untuk menggunakan library pendukung seperti pandas dan numpy untuk membantu Anda dalam mengerjakan tugas ini. Harap konsultasikan dengan asdos terkait penggunaan library pendukung.

Pada bagian akhir, terdapat pertanyaan yang harus dijawab. Apabila Anda mencari referensi dari internet, Anda WAJIB menyertakan sumbernya. Silakan jawab pertanyaan dengan jelas dan lengkap secara langsung di file tersebut.

Dilarang keras menyontek. Plagiarisme tidak ditoleransi dan akan dikenai penalti atau nilai akhir E.

Pengurangan nilai akibat keterlambatan pengumpulan tugas akan ditentukan berdasarkan jumlah menit keterlambatan Anda dalam mengumpulkan. Misalnya, apabila terlambat 1 menit, nilai akhir akan dikurangi 1 poin, apabila terlambat 10 menit, nilai akhir akan dikurangi 10 poin, dan seterusnya.

Bobot penilaian:

- Implementasi class kNN (25%)
- Hasil percobaan dan analisis (50%)
- Menjawab pertanyaan (25%)

Pengumpulan tugas:

Kumpulkan berkas dengan format penamaan seperti berikut:

Assignment1_[NPM]_[NamaLengkap].ipynb

Contoh:

Assignment1_2006596535_FransiscoWilliamSudianto.ipynb