

**ELBOW METHOD**  
**PRAKTIKUM PEMBELAJARAN MESIN**



**Disusun Oleh:**

Nashirudin Baqiy

24060119130045

**Lab A2**

**DEPARTEMEN ILMU KOMPUTER / INFORMATIKA**  
**FAKULTAS SAINS DAN MATEMATIKA**  
**UNIVERSITAS DIPONEGORO**  
**2021**

# **BAB I PENDAHULUAN**

## **1.1. Rumusan Masalah**

- 1.1.1. Implementasikan elbow method!

## **1.2. Tujuan**

- 1.2.1. Mengimplementasikan elbow method untuk mencari jumlah kluster yang tepat dari suatu dataset.

## **1.3. Dasar Teori**

Pembelajaran mesin (ML) adalah studi tentang algoritma komputer yang dapat ditingkatkan secara otomatis melalui pengalaman dan penggunaan data. Itu dilihat sebagai bagian dari kecerdasan buatan. Algoritma pembelajaran mesin membangun model berdasarkan data sampel, yang dikenal sebagai "data pelatihan", untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit untuk melakukannya. Algoritma pembelajaran mesin digunakan dalam berbagai macam aplikasi, seperti dalam kedokteran, penyaringan email, pengenalan suara, dan visi komputer, di mana sulit atau tidak mungkin untuk mengembangkan algoritma konvensional untuk melakukan tugas-tugas yang diperlukan.

Bagian dari pembelajaran mesin terkait erat dengan statistik komputasi, yang berfokus pada pembuatan prediksi menggunakan komputer; tetapi tidak semua pembelajaran mesin adalah pembelajaran statistik. Studi optimasi matematika memberikan metode, teori, dan domain aplikasi ke bidang pembelajaran mesin. Data mining adalah bidang studi terkait, dengan fokus pada analisis data eksplorasi melalui pembelajaran tanpa pengawasan. Beberapa implementasi pembelajaran mesin menggunakan data dan jaringan saraf dengan cara yang meniru kerja otak biologis. Dalam penerapannya di seluruh masalah bisnis, machine learning juga disebut sebagai analitik prediktif.

## BAB II PEMBAHASAN

### 2.1. Deskripsi dataset

Dataset yang digunakan adalah dataset bunga iris. Dataset ini terdiri atas empat fitur yaitu Sepal Length, Sepal Width, Petal Length, dan Petal Width. Serta terdiri atas tiga kelas, yaitu Iris-setosa, Iris-versicolor, dan Iris-virginica. Pada Iris dataset, jumlah keseluruhan data adalah 150 baris dengan masing-masing kelas memiliki 50 baris data. Dalam eksperimen ini kita akan memakai 2 fitur dari data iris, tanpa menggunakan label dari data tersebut untuk melakukan clustering. Data iris terdiri dari data pengukuran fitur (`iris.data`) dan label (`iris.target_names`).

### 2.2. Menentukan nilai untuk cluster

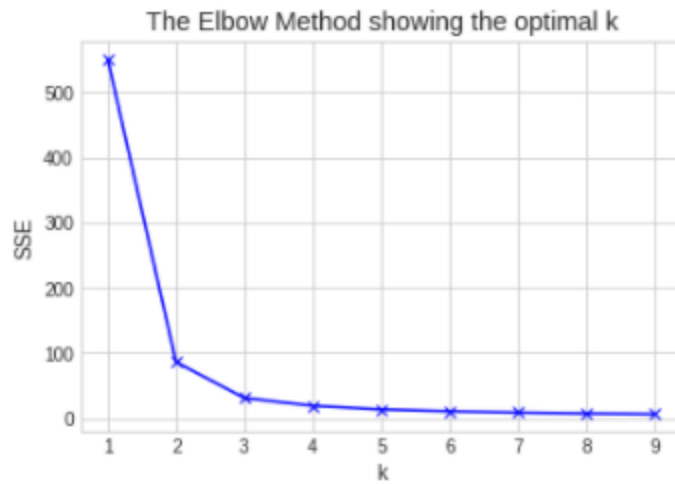
Untuk menentukan nilai yang tepat untuk K, kita dapat menghitung Sum of Squared error dari cluster yang dihasilkan. Dalam python hal tersebut dapat dilakukan dengan menyimpan nilai inertia dalam variable SSE sebagai berikut.

```
X = np.array(list(zip(features[2], features[3]))).reshape(len(features[3]), 2)
SSE = []
K = range(1, 10)
for k in K:
    kmean = KMeans(n_clusters=k).fit(X)
    kmean.fit(X)
    SSE.append(kmean.inertia_)
print("SSE ke-", k, SSE)
```

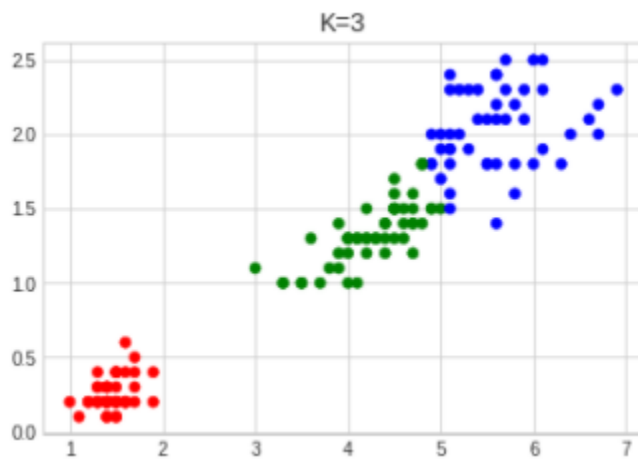
```
ke- 1 [550.8953333333334]
ke- 2 [550.8953333333334, 86.39021984551397]
ke- 3 [550.8953333333334, 86.39021984551397, 31.371358974358973]
ke- 4 [550.8953333333334, 86.39021984551397, 31.371358974358973, 19.465989010989013]
ke- 5 [550.8953333333334, 86.39021984551397, 31.371358974358973, 19.465989010989013, 13.916908757908757]
ke- 6 [550.8953333333334, 86.39021984551397, 31.371358974358973, 19.465989010989013, 13.916908757908757, 11.040239971910452]
ke- 7 [550.8953333333334, 86.39021984551397, 31.371358974358973, 19.465989010989013, 13.916908757908757, 11.040239971910452, 9.203314009661835]
ke- 8 [550.8953333333334, 86.39021984551397, 31.371358974358973, 19.465989010989013, 13.916908757908757, 11.040239971910452, 9.203314009661835, 7.615402244955953]
ke- 9 [550.8953333333334, 86.39021984551397, 31.371358974358973, 19.465989010989013, 13.916908757908757, 11.040239971910452, 9.203314009661835, 7.615402244955953, 6.576538396386223]
```

### 2.3. Plot nilai K

Langkah selanjutnya adalah melakukan plotting nilai K untuk mencari nilai K mana yang sekiranya cocok untuk dijadikan jumlah clusters.

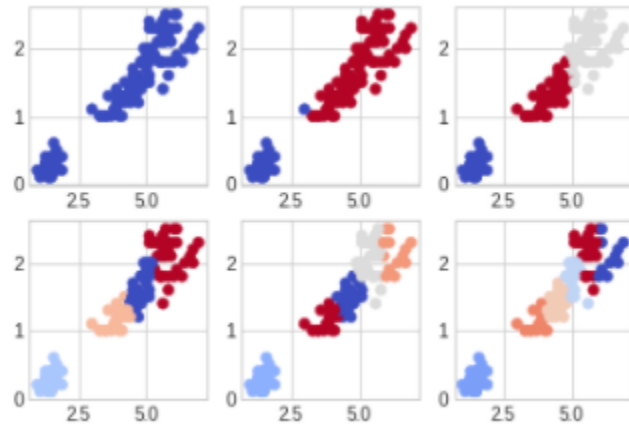


Jika melihat dari grafik hasil plotting diatas maka kita akan memilih nilai dari grafik tersebut yang mulai tidak curam. Nilai K yang mulai tidak curam ada pada K ketiga. Maka dari itu kita akan menggunakan K dengan nilai 3 atau 3 clusters.

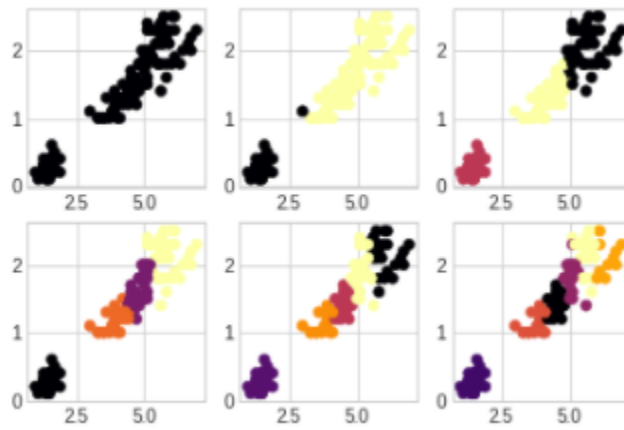


## 2.4 Hasil Clusters

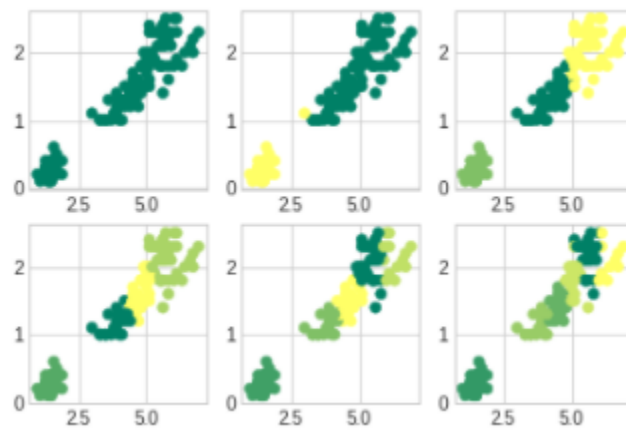
Coolwarm



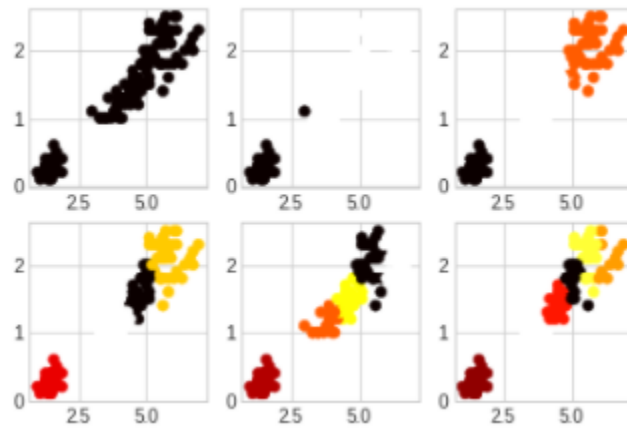
Inferno



Summer



Hot



## **BAB III PENUTUP**

### **4.1. Kesimpulan**

Pada praktikum keempat di mata kuliah pembelajaran mesin kali ini, mempelajari mengenai cara menggunakan elbow method untuk menentukan jumlah cluster dari dataset tanpa label. Dataset yang digunakan pada praktikum ini adalah dataset bunga iris. Cluster yang digunakan dari dataset bunga iris ini adalah petal length dan petal width. Dari hasil praktikum, cluster yang didapat sebanyak 3 cluster.

## DAFTAR PUSTAKA

Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications*, 129, 273-285.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.

Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; Arvin, F., "Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning" *IEEE Transactions on Vehicular Technology*, 2020.

"What is Machine Learning?". *www.ibm.com*. Retrieved 2021-08-15.

Zhou, Victor (2019-12-20). "Machine Learning for Beginners: An Introduction to Neural Networks". *Medium*. Retrieved 2021-08-15.