11/16/24, 9:33 AM kmeans - Colab

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import seaborn as sns

1 data=pd.read_csv('/content/Iris.csv')

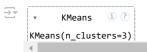
1 data.head()

Id SepalLengthCm SepalWidthCm

0 1 5.1 3.5
```

Id SepalLengthCm SepalWidthCm PetalLengthCm PetalWidthCm Species 1.4 0.2 Iris-setosa 3.5 1 2 4.9 3.0 1.4 0.2 Iris-setosa **2** 3 4.7 3 2 1.3 0.2 Iris-setosa 3 4.6 3.1 1.5 0.2 Iris-setosa 0.2 Iris-setosa 4 5 5.0 1.4 3.6

- 1 from sklearn.preprocessing import LabelEncoder
- 2 le=LabelEncoder()
- 3 data['Species']=le.fit_transform(data['Species'])
- 1 from sklearn.model_selection import train_test_split
- 2 x=data.drop(['Species','Id'],axis=1)
- 3 y=data['Species']
- 4 x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
- 1 from sklearn.cluster import KMeans
- 2 km=KMeans(n_clusters=3)
- 3 km.fit(x_train)



- 1 from sklearn.metrics import accuracy_score
- 2 accuracy_score(y_test,km.predict(x_test))
- 1 sns.pairplot(data,hue='Species')

11/16/24, 9:33 AM kmeans - Colab

