

Data Science and Business Analytics

Task 2: Prediction using unsupervised machine learning

submitted by: Laxman Velip

```
In [ ]: #changing the current working directory
import os
os.chdir('C:/Users/laxman/Documents/My Bluetooth/data_csv')
```

```
In [ ]: #importing necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [ ]: #data analysis
dataset=pd.read_csv('iris.csv') #importing dataset
print(dataset.head(5)) #first five rows of dataset
print(dataset.isna().sum()) #checking for null values
print(dataset.columns) #printing all columns of dataset
print(dataset['Species'].unique()) #printing unique elements of 'Species' column
X=dataset.iloc[:, [1,2,3,4]].values
```

```
In [ ]: #finding the optimum number of clusters
from sklearn.cluster import KMeans
wcss=[]
for i in range(1,11):
    kmeans_clu=KMeans(n_clusters = i, random_state=56)
    kmeans_clu.fit(X)
    wcss.append(kmeans_clu.inertia_)
```

```
In [ ]: #elbow method visualization
plt.figure(figsize=(10,5))
plt.plot(range(1,11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('wcss')
plt.show()
```

In the above plot elbow occur at 3, hence number of clusters k=3

```
In [ ]: #model building
kmeans=KMeans(n_clusters=3,random_state=56)
y_kmeans=kmeans.fit_predict(X)
```

```
In [ ]: plt.scatter(X[y_kmeans==0,0], X[y_kmeans==0,1], s=100,c='red', label='Iris-versicolour')
plt.scatter(X[y_kmeans==1,0], X[y_kmeans==1,1], s=100,c='blue', label='Iris-setosa')
plt.scatter(X[y_kmeans==2,0], X[y_kmeans==2,1], s=100,c='green', label='Iris-verginica')

plt.scatter(kmeans.cluster_centers_[0, 0], kmeans.cluster_centers_[0, 1], s=300, c='black', label='Centroids')
plt.title('clusters of Species')
plt.xlabel('sepal length')
plt.ylabel('sepal width')
plt.legend()
plt.show()
```

```
In [ ]:
```