⌄ Lab4

```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn import svm
from sklearn.ensemble import RandomForestClassifier
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import f1_score
from sklearn.svm import SVC
```

```python
#UPLOAD THE smart_grid_stability_augmented dataset. Drop the 'stab'column, keep axis=1. Save in variable STAB.
```

⌄ **1. Seperate the STAB dataset to train and test, stabf is y, the others are x, make test dataset 30% of the total dataset (10)**

```python
#Write your code here.
#hint: Since y is the column 'stabf', X should be STAB with the dropped stabf (like you did above). Use random state 42.
#ANSWER:
```

⌄ **2. Apply Random Forest to this dataset, predict stabf, show the accuracy. (10)**

```python
#Write your code here.
#hint use RandomForestClassifier and use random state 42.
#ANSWER:
```

⌄ **3. Apply SVM to this dataset, predict stabf, show the accuracy. (10)**

```python
#Write your code here.
#hint use SVC and use random state 42.
```

⌄ **4. Standardize and Normalize the data. Use the first 2000 observations only. Then use PCA to reduce to 2 dimensions. Then plot the reduced data (10)**

```python
#Write your code here.
#hint: first convert STAB into dataframe, use StandardScaler. Look at the practice
#ANSWER:
```

⌄ **5a. Apply K-means on the dataset. Determine optimal number of clusters using silhouette score (10)**

```python
#hint: apply KMeans to each cluster. Predict on the scaled X.
#ANSWER:
silhouette_scores = []
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans
```

⌄

5b. Apply Agglomerative Clustering on the dataset. Determine optimal number of clusters using silhouette score (10)

```
#ANSWER
import matplotlib.pyplot as plt
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import silhouette_score
silhouette_scores = []
```

5c. Compare the two methods. Explain any differences you can see (10)

Write answer here

6. Load a new "Wine" dataset.

```
#write your code here
#ANSWER
```

7. Create a Model using DBSCANS (10). Use variables Malic_Acid and Customer_segment. Compare results for eps=0.5 and eps=0.2 for min_samples=2

```
#hint: look at the practice
#write your code here
#ANSWER
```

8. Explain the different techniques used in this lab. What are their uses and under which conditions are they used?(20)

Write your answer