

```
import numpy as np
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import classification_report
```

```
df1 = pd.read_csv("GSE52194.csv")
df2 = pd.read_csv("GSE69240.csv")
df3 = pd.read_csv("GSE71651.csv")
```

```
df = pd.concat([df1,df2,df3], axis=0)
df.head()
```

	ID	class	ENSG00000000003	ENSG00000000005	ENSG000000000419	ENSG0000000000457	ENSG0000000000458
0	SRR1027171	Tumor	8.063609	2.814594	9.012854	7.923587	7
1	SRR1027172	Tumor	8.301166	1.476355	8.498698	7.574031	7
2	SRR1027173	Tumor	8.258444	1.476355	8.770748	8.566797	8
3	SRR1027174	Tumor	8.418663	3.151186	8.514820	7.507616	7
4	SRR1027175	Tumor	9.036324	2.594064	9.145899	8.372162	8

5 rows × 58737 columns

```
print(len(df.columns))
print(len(df))
df4 = df.drop('ID', axis=1)
```

58737
88

```
print(df4["class"].unique())
df4['class'] = (df4["class"] == ' Tumor').apply(int)
```

[' Tumor' ' Normal']

```
print(len(df4["class"] == 1))
display(df4)
```

88

	class	ENSG00000000003	ENSG00000000005	ENSG000000000419	ENSG000000000457	ENSG000000000458
0	1	8.063609	2.814594	9.012854	7.923587	7
1	1	8.301166	1.476355	8.498698	7.574031	7
2	1	8.258444	1.476355	8.770748	8.566797	8
3	1	8.418663	3.151186	8.514820	7.507616	7
4	1	9.036324	2.594064	9.145899	8.372162	8
...
28	1	10.163420	5.795390	8.910083	9.915531	10
29	1	9.977373	5.909457	8.844046	10.065302	10
30	0	9.360885	6.081324	8.784923	9.801867	9
31	0	9.945123	7.288354	8.590931	10.116428	10
32	0	9.929910	7.354764	9.102810	9.898801	9

88 rows × 58736 columns

```
y = df4['class']
df = df4.iloc[:,1:]
print( y.value_counts() )
```

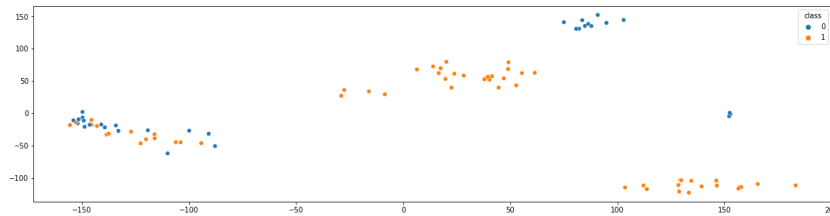
```

from sklearn.decomposition import PCA
pca = PCA(n_components=2)
X_pca = pca.fit_transform(df4)

import matplotlib.pyplot as plt
import seaborn as sns
plt.figure(figsize = (20,5))
sns.scatterplot(X_pca[:,0],X_pca[:,1], hue = y)

1      57
0      31
Name: class, dtype: int64
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f5e6dea7130>

```



▼ Split dataset into Train and Test & SMOTE Over_Sampling

```

from sklearn.model_selection import train_test_split
from imblearn.over_sampling import SMOTE

X = df4.drop(['class'], axis = 1)
print(len(X))
y = df4["class"]
print(len(y))

X = pd.get_dummies(X)
sm = SMOTE(sampling_strategy='minority', random_state=42)
X_res, y_res = sm.fit_resample(X, y)

X_train, X_test, y_train, y_test = train_test_split(X_res, y_res, test_size=0.2, random_state=43)
print(X_test.shape)

88
88
(23, 58735)

```

▼ Train Models

```

names = ["Logistic Regression", "Nearest Neighbors", "Linear SVM", "RBF SVM", "Gaussian Process",
         "Decision Tree", "Random Forest", "Neural Net", "AdaBoost",
         "Naive Bayes", "QDA"]

```

```

from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.gaussian_process import GaussianProcessClassifier
from sklearn.gaussian_process.kernels import RBF
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis

```

```

classifiers = [
    LogisticRegression(max_iter=300),
    KNeighborsClassifier(),
    SVC(kernel="linear", C=0.025),

```

```

SVC(gamma=2, C=1),
GaussianProcessClassifier(1.0 * RBF(1.0)),
DecisionTreeClassifier(max_depth=5, random_state=43),
RandomForestClassifier(max_depth=5, random_state=43),
MLPClassifier(alpha=1, max_iter=1000),
AdaBoostClassifier(),
GaussianNB(),
QuadraticDiscriminantAnalysis()]

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.fit_transform(X_test)

for name, clf in zip(names, classifiers):
    clf.fit(X_train, y_train)
    score = clf.score(X_test, y_test)
    print(f"Accuracy of {name} Classifier is:{score}")

Accuracy of Logistic Regression Classifier is:0.9130434782608695
Accuracy of Nearest Neighbors Classifier is:0.8695652173913043
Accuracy of Linear SVM Classifier is:0.9130434782608695
Accuracy of RBF SVM Classifier is:0.391304347826087
Accuracy of Gaussian Process Classifier is:0.9130434782608695
Accuracy of Decision Tree Classifier is:0.782608695652174
Accuracy of Random Forest Classifier is:0.9130434782608695
Accuracy of Neural Net Classifier is:0.6521739130434783
Accuracy of AdaBoost Classifier is:0.8695652173913043
Accuracy of Naive Bayes Classifier is:0.8695652173913043
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:926: UserWarning: Variables are collinear
  warnings.warn("Variables are collinear")
Accuracy of QDA Classifier is:0.9565217391304348

```

▼ Logistic Regression

```

from sklearn.linear_model import LogisticRegression

lr_model = LogisticRegression(random_state=42, max_iter=300,)
lr_model.fit(X_train, y_train)
lr_accuracy = lr_model.score(X_test, y_test)
print(f"Accuracy of Logistic Regression Classifier is:{lr_accuracy}")

Accuracy of Logistic Regression Classifier is:0.9130434782608695

```

▼ K Nearest Neighbor Algorithm

```

from sklearn.neighbors import KNeighborsClassifier

knn_model = KNeighborsClassifier()
knn_model.fit(X_train, y_train)
knn_accuracy = knn_model.score(X_test_scaled, y_test)
print(f"Accuracy of kNN Classifier is:{knn_accuracy}")

Accuracy of kNN Classifier is:0.8695652173913043
/usr/local/lib/python3.8/dist-packages/sklearn/base.py:450: UserWarning: X does not have valid feature names, but KNeighborsClassifier v
  warnings.warn(

```

Linear Support Vector Machines (SVM)

```

from sklearn import svm

svm_model = svm.SVC(random_state=42, kernel='linear')
svm_model.fit(X_train, y_train)

svm_accuracy = svm_model.score(X_test, y_test)
print(f"Accuracy of Linear SVM Classifier is:{svm_accuracy}")

Accuracy of Linear SVM Classifier is:0.9130434782608695

```

▼ RBF Support Vector Machines (SVM)

```
from sklearn import svm
from sklearn.model_selection import StratifiedShuffleSplit
from sklearn.model_selection import GridSearchCV

C_range = np.logspace(-2, 10, 13)
gamma_range = np.logspace(-9, 3, 13)
param_grid = dict(gamma=gamma_range, C=C_range)
cv = StratifiedShuffleSplit(n_splits=5, test_size=0.2, random_state=43)
grid = GridSearchCV(SVC(), param_grid=param_grid, cv=cv)
grid.fit(X_train_scaled, y_train)

print("The best parameters are %s with a score of %0.2f"
      % (grid.best_params_, grid.best_score_))

rbf_svm_model = svm.SVC(gamma=0.01, C=25, class_weight={1: 2})
rbf_svm_model.fit(X_train_scaled, y_train)

rbf_svm_accuracy = rbf_svm_model.score(X_test_scaled, y_test)
print(f"Accuracy of RBF SVM Classifier is:{rbf_svm_accuracy}")

Accuracy of RBF SVM Classifier is:0.391304347826087
```

▼ Gaussian Process Classifier

```
from sklearn.gaussian_process import GaussianProcessClassifier
from sklearn.gaussian_process.kernels import RBF

kernel = 1.0 * RBF(1.0)
gpc_model = GaussianProcessClassifier(kernel, random_state=43, max_iter_predict=10000, n_jobs=-1)
gpc_model.fit(X_train, y_train)

gpc_accuracy = gpc_model.score(X_test, y_test)
print(f"Accuracy of Gaussian Process Classifier is:{gpc_accuracy}")

Accuracy of Gaussian Process Classifier is:0.9130434782608695
```

▼ Decision Trees

```
from sklearn import tree

max_depth_range = np.linspace(1, 10, 10)
param_grid = dict(max_depth=max_depth_range)
cv = StratifiedShuffleSplit(n_splits=5, test_size=0.2, random_state=43)
grid = GridSearchCV(DecisionTreeClassifier(class_weight={1: 0}), param_grid=param_grid, cv=cv)
grid.fit(X_train, y_train)

print("The best parameters are %s with a score of %0.2f"
      % (grid.best_params_, grid.best_score_))

The best parameters are {'max_depth': 1.0} with a score of 0.53

tree_model = tree.DecisionTreeClassifier( max_depth=1, random_state=53)
tree_model.fit(X_train, y_train)

tree_accuracy = tree_model.score(X_test, y_test)
print(f"Accuracy of Decision Tree Classifier is:{tree_accuracy}")

Accuracy of Decision Tree Classifier is:0.8260869565217391
```

▼ Random Forest

```
from sklearn.ensemble import RandomForestClassifier

max_depth_range = np.linspace(1, 27, 27)
```

```

max_features_range = np.arange(1, 27, 1)
param_grid = dict(max_depth=max_depth_range, max_features=max_features_range)
cv = StratifiedShuffleSplit(n_splits=5, test_size=0.2, random_state=43)
grid = GridSearchCV(RandomForestClassifier(class_weight={1: 0}, random_state=43), param_grid=param_grid, cv=cv)
grid.fit(X_train, y_train)

print("The best parameters are %s with a score of %.2f"
      % (grid.best_params_, grid.best_score_))

from sklearn.ensemble import RandomForestClassifier

rf_model = RandomForestClassifier(max_depth=25, n_estimators=400, max_features=150, random_state=43, n_jobs=-1)
rf_model.fit(X_train, y_train)

rf_accuracy = rf_model.score(X_test, y_test)
print(f"Accuracy of Random Forest Classifier is:{rf_accuracy}")

Accuracy of Random Forest Classifier is:0.9130434782608695

```

▼ Neural Network

```

from sklearn.neural_network import MLPClassifier

nn_model = MLPClassifier(solver='lbfgs', alpha=1e-5, hidden_layer_sizes=(100,), random_state=43, max_iter=10000, learning_rate='adaptive')
nn_model.fit(X_train, y_train)
nn_accuracy = nn_model.score(X_test, y_test)
print(f"Accuracy of MLP Classifier is:{nn_accuracy}")

Accuracy of MLP Classifier is:0.9130434782608695

```

▼ AdaBoost

```

from sklearn.ensemble import AdaBoostClassifier

ada_model = AdaBoostClassifier(random_state=43, n_estimators=100)
ada_model.fit(X_train, y_train)
ada_accuracy = ada_model.score(X_test, y_test)
print(f"Accuracy of Ada Boost Classifier is:{ada_accuracy}")

Accuracy of Ada Boost Classifier is:0.8695652173913043

```

▼ Gaussian Naive Bayes

```

from sklearn.naive_bayes import GaussianNB

gnb_model = GaussianNB()
gnb_model.fit(X_train, y_train)
gnb_accuracy = gnb_model.score(X_test, y_test)
print(f"Accuracy of Gaussian Naive Bayes Classifier is:{gnb_accuracy}")

Accuracy of Gaussian Naive Bayes Classifier is:0.8695652173913043

```

▼ Quadratic Discriminant Analysis

```

from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis

qda_model = QuadraticDiscriminantAnalysis()
qda_model.fit(X_train, y_train)
qda_accuracy = qda_model.score(X_test, y_test)
print(f"Accuracy of Quadratic Discriminant Analysis Classifier is:{qda_accuracy}")

/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
  warnings.warn("Variables are collinear")
Accuracy of Quadratic Discriminant Analysis Classifier is:0.956217391304348

```

▼ kNN

```
knn_model = KNeighborsClassifier(n_neighbors=5)
knn_model.fit(X_train, y_train)
```

```
▼ KNeighborsClassifier
KNeighborsClassifier()
```

```
knn_model_pred = knn_model.predict(X_test)
```

```
print(classification_report(y_test, knn_model_pred))
```

	precision	recall	f1-score	support
0	0.75	1.00	0.86	9
1	1.00	0.79	0.88	14
accuracy			0.87	23
macro avg	0.88	0.89	0.87	23
weighted avg	0.90	0.87	0.87	23

▼ Linear SVM

```
from sklearn.svm import LinearSVC
```

```
svm_model = LinearSVC(max_iter = 1000)
svm_model = svm_model.fit(X_train,y_train)
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/svm/_base.py:1206: ConvergenceWarning: Liblinear failed to converge, increase the number
warnings.warn(
```

```
y_pred = svm_model.predict(X_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.82	1.00	0.90	9
1	1.00	0.86	0.92	14
accuracy			0.91	23
macro avg	0.91	0.93	0.91	23
weighted avg	0.93	0.91	0.91	23

▼ RBF SVM

```
from sklearn import svm
```

```
rbf_svm_model = svm.SVC()
rbf_svm_model = rbf_svm_model.fit(X_train, y_train)
```

```
rbf_svm_pred = rbf_svm_model.predict(X_test)
print(classification_report(y_test, rbf_svm_pred))
```

	precision	recall	f1-score	support
0	0.75	1.00	0.86	9
1	1.00	0.79	0.88	14
accuracy			0.87	23
macro avg	0.88	0.89	0.87	23
weighted avg	0.90	0.87	0.87	23

▼ Decision Tree

```
from sklearn import tree
```

```
tree_model = tree.DecisionTreeClassifier()
tree_model = tree_model.fit(X_train,y_train)
```

```
tree_pred = tree_model.predict(X_test)
print(classification_report(y_test, tree_pred))
```

	precision	recall	f1-score	support
0	0.73	0.89	0.80	9
1	0.92	0.79	0.85	14
accuracy			0.83	23
macro avg	0.82	0.84	0.82	23
weighted avg	0.84	0.83	0.83	23

▼ Random Forest

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.tree import DecisionTreeClassifier
```

```
random_model = DecisionTreeClassifier(max_depth=None, min_samples_split=2,random_state=0)
random_model = RandomForestClassifier()
random_model = random_model.fit(X_train,y_train)
```

```
random_model_pred = random_model.predict(X_test)
print(classification_report(y_test, random_model_pred))
```

	precision	recall	f1-score	support
0	0.82	1.00	0.90	9
1	1.00	0.86	0.92	14
accuracy			0.91	23
macro avg	0.91	0.93	0.91	23
weighted avg	0.93	0.91	0.91	23

▼ ADA boost

```
from sklearn.ensemble import AdaBoostClassifier
```

```
ada_model = AdaBoostClassifier(random_state=43, n_estimators=100)
ada_model = ada_model.fit(X_train, y_train)
```

```
ada_pred = ada_model.predict(X_test)
print(classification_report(y_test, ada_pred))
```

	precision	recall	f1-score	support
0	0.80	0.89	0.84	9
1	0.92	0.86	0.89	14
accuracy			0.87	23
macro avg	0.86	0.87	0.87	23
weighted avg	0.87	0.87	0.87	23

▼ Naive Bayes

```
from sklearn.naive_bayes import GaussianNB
```

```
gnb_model = GaussianNB()
gnb_model = gnb_model.fit(X_train,y_train)
```

```
y_pred = gnb_model.predict(X_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.75	1.00	0.86	9
1	1.00	0.79	0.88	14
accuracy			0.87	23
macro avg	0.88	0.89	0.87	23
weighted avg	0.90	0.87	0.87	23

▼ Neural Network

```
from sklearn.neural_network import MLPClassifier
```

```
nn_model = MLPClassifier(solver='lbfgs', alpha=1e-5, hidden_layer_sizes=(100,), random_state=43, max_iter=10000, learning_rate='adaptive')
nn_model = nn_model.fit(X_train,y_train)
```

```
y_pred = nn_model.predict(X_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	0.82	1.00	0.90	9
1	1.00	0.86	0.92	14
accuracy			0.91	23
macro avg	0.91	0.93	0.91	23
weighted avg	0.93	0.91	0.91	23

▼ QuadraticDiscriminantAnalysis

```
from sklearn.discriminant_analysis import QuadraticDiscriminantAnalysis
from sklearn.metrics import accuracy_score
```

```
qda_model = QuadraticDiscriminantAnalysis()
qda_model = qda_model.fit(X_train, y_train)
```

```
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
```

```
y_pred = qda_model.predict(X_test)
print(classification_report(y_test, y_pred))
```

	precision	recall	f1-score	support
0	1.00	0.89	0.94	9
1	0.93	1.00	0.97	14
accuracy			0.96	23
macro avg	0.97	0.94	0.95	23
weighted avg	0.96	0.96	0.96	23

As per accuracy

- QuadraticDiscriminantAnalysis – 96%
- Neural Networ = 91%
- Random Forest = 91%
- Linear SVM = 91 %
- Logistic Regression = 91%

▼ QDA k-fold Validation

```
from sklearn.model_selection import cross_val_score

#manipulated the cv to produce the highest level of accuracy
scores = cross_val_score(qda_model, X_train, y_train, cv=10)

print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std() * 2))

/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
/usr/local/lib/python3.8/dist-packages/sklearn/discriminant_analysis.py:878: UserWarning: Variables are collinear
warnings.warn("Variables are collinear")
Accuracy: 0.88 (+/- 0.20)
```

▼ Area under the curve

```
from sklearn.metrics import roc_auc_score

roc_auc = roc_auc_score(y_test, y_pred)

print("ROC AUC:", roc_auc)

ROC AUC: 0.9444444444444444
```

▼ Determining the number of Fetures using GridSearchCV

```
from sklearn.model_selection import GridSearchCV

param_grid = {'reg_param': [0, 0.1, 1, 10]}

grid_search = GridSearchCV(estimator=qda_model, param_grid=param_grid, cv=10)

grid_search.fit(X_res, y_res)

best_hyperparameters = grid_search.best_params_
best_accuracy = grid_search.best_score_

print("Best hyperparameters: ", best_hyperparameters)
print("Best score: ", best_accuracy)

print("List of available parameters: ", qda_model.get_params().keys())
```

▼ Applying UFS with PCA Technique

```
#extracted tumor class genes
t_x = df4['class'].isin([1])
tumor_x = df4[t_x]
print(tumor_x.shape)
```

```
(57, 58736)
```

```
t_y = y.isin([1])
tumor_y = y[t_y]
print(tumor_y.shape)
```

```
(57,)
```

```
from sklearn.decomposition import PCA
pca = PCA()
```

```
X_pca = abs(pca.fit_transform(tumor_x))
```

```
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import f_classif
```

```
header = df4.columns
```

```
print(len(header))
```

```
k = 10
selector = SelectKBest(score_func=f_classif, k=k)
```

```
X_new = selector.fit_transform(X_pca, tumor_y)
print(y.value_counts())
```

```
mask = selector.get_support()
print(len(mask))
```

```
selected_header = [header[i] for i, x in enumerate(mask) if x]
```

```
extracted_features = pd.DataFrame(data=X_new, columns=selected_header)
```

```
header_df = pd.DataFrame(data=header, columns=['header'])
```

```
result = pd.concat([header_df, extracted_features], axis=1)
```

```
58736
1      57
0      31
Name: class, dtype: int64
57
/usr/local/lib/python3.8/dist-packages/sklearn/feature_selection/_univariate_selection.py:108: RuntimeWarning: invalid value encountered
  msb = ssbn / float(dfbn)
```

```
print(X_new.shape)
print(result.head(2))
```

```
(57, 10)
      header  ENSG00000004139  ENSG00000004142  ENSG00000004399  \
0      class      7.532712      4.214933      0.878545
1  ENSG00000000003      0.818142      1.394575      0.292131
      ENSG00000004455  ENSG00000004468  ENSG00000004478  ENSG00000004487  \
0      3.864689      1.054819      2.685564      0.07222
1      0.717469      0.407136      1.758345      0.19824
      ENSG00000004534  ENSG00000004660  ENSG00000004700
```

0	2.490519	0.837601	9.719311e-13
1	0.042381	0.297540	9.719311e-13

GRAPHS

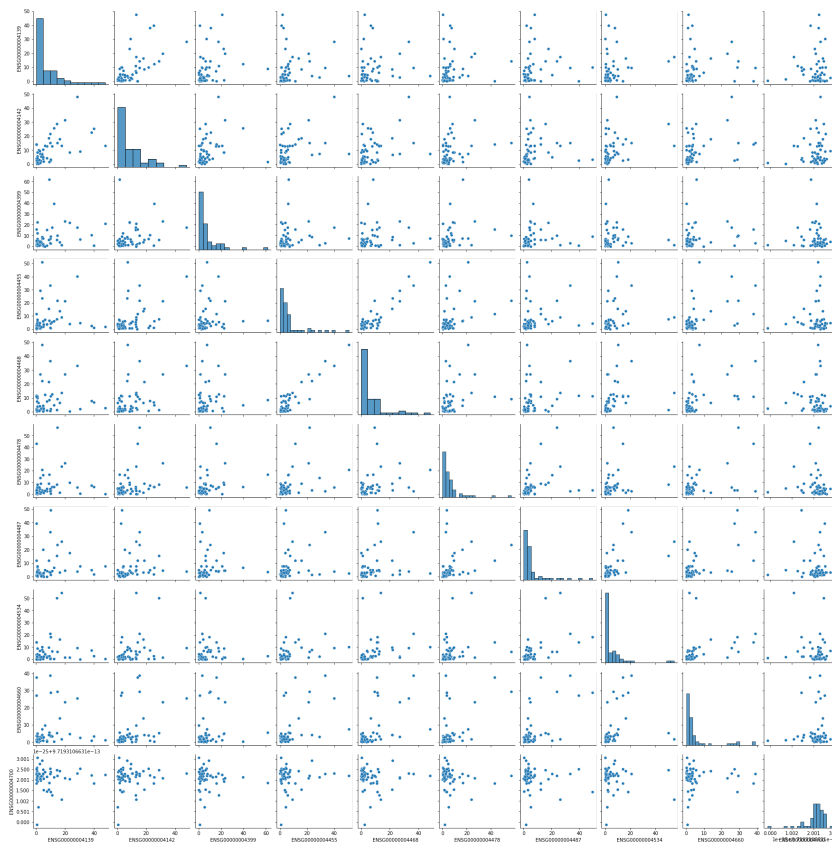
Pairplot

```
import seaborn as pairplot
```

```
ax = sns.pairplot(extracted_features)
```

```
plt.savefig("top_10_Pairplot.png")
```

```
plt.show()
```



Heatmap

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

```
corr = extracted_features.corr()
```

```
ax = sns.heatmap(corr, cmap='coolwarm', annot = True)
```

```

name = "Top Genes"
genes_name = ["SARM", "POLDIP2", "PLXND1", "AK2", "CD68", "FKBP4",
              "KDM1A", "RBM6", "CAMKK1", "RECQL"]

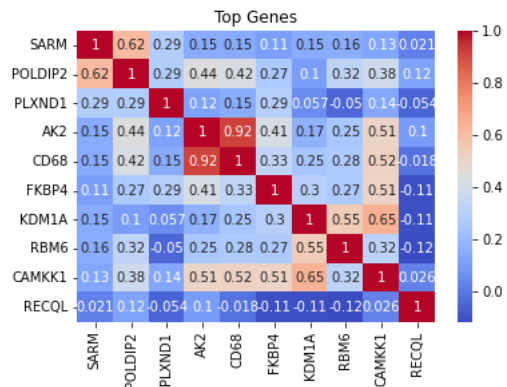
ax.set_xticklabels(genes_name)
ax.set_yticklabels(genes_name)

plt.title(name)

plt.savefig("top_10_heatmap.png")

plt.show()

```



Bar plot

```

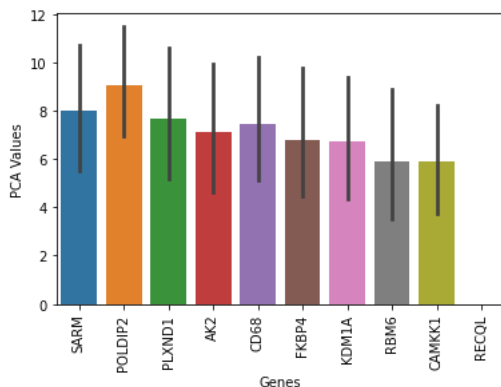
ax = sns.barplot(data = extracted_features,)

genes_name = ["SARM", "POLDIP2", "PLXND1", "AK2", "CD68", "FKBP4",
              "KDM1A", "RBM6", "CAMKK1", "RECQL"]

ax.set_xticklabels(genes_name)
plt.xticks(rotation=90)

plt.xlabel('Genes')
plt.ylabel('PCA Values')
plt.show()

```



Saving my model and downloading

```

import pickle

# Save the model
with open("QDA_model.pkl", "wb") as f:
    pickle.dump(qda_model, f)

```

```
# Load the model
with open("QDA_model.pkl", "rb") as f:
    best_model = pickle.load(f)
```

