correlation_target _label_30 9010 split .02

January 2, 2023

[ ]:
```
# Importing the packages
import sys
import numpy as np
np.set_printoptions(threshold=sys.maxsize)
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import sklearn
import random
from sklearn.metrics import␣
↪confusion_matrix,accuracy_score,classification_report,RocCurveDisplay,ConfusionMatrixDispla
```

[ ]:
```
pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
pd.set_option('display.width', None)
pd.set_option('display.max_colwidth', None)
```

[ ]:
```
# Importing the dataset
df = pd.read_csv('dataset_30.csv')
df.drop(['index'], axis=1, inplace=True)
#df.head()
```

[ ]:
```
# if your dataset contains missing value, check which column has missing values
#df.isnull().sum()
```

[ ]:
```
#df.dropna(inplace=True)
```

[ ]:
```
from sklearn import preprocessing


col = df.columns[:]

lab_en= preprocessing.LabelEncoder()

for c in col:
    df[c]= lab_en.fit_transform(df[c])

#df.head(50)
```

[ ]:
```
##print(df.corr()['Result'].sort_values())
## correlation values of features with target label
corr_col = abs(df.corr()['Result']).sort_values(ascending=False)
corr_col = corr_col.rename_axis('Col').reset_index(name='Correlation')
corr_col
```

[ ]:

	Col	Correlation
0	Result	1.000000
1	SSLfinal_State	0.714741
2	URL_of_Anchor	0.692935
3	Prefix_Suffix	0.348606
4	web_traffic	0.346103
5	having_Sub_Domain	0.298323
6	Request_URL	0.253372
7	Links_in_tags	0.248229
8	Domain_registeration_length	0.225789
9	SFH	0.221419
10	Google_Index	0.128950
11	age_of_domain	0.121496
12	Page_Rank	0.104645
13	having_IPhaving_IP_Address	0.094160
14	Statistical_report	0.079857
15	DNSRecord	0.075718
16	Shortining_Service	0.067966
17	Abnormal_URL	0.060488
18	URLURL_Length	0.057430
19	having_At_Symbol	0.052948
20	on_mouseover	0.041838
21	HTTPS_token	0.039854
22	double_slash_redirecting	0.038608
23	port	0.036419
24	Links_pointing_to_page	0.032574
25	Redirect	0.020113
26	Submitting_to_email	0.018249
27	RightClick	0.012653
28	Iframe	0.003394
29	Favicon	0.000280
30	popUpWidnow	0.000086

[ ]:
```
def correlation (corr_col, threshold):
        corr_feature = set()
        for index, row in corr_col.iterrows():
                if row['Correlation'] < threshold or np.
 ↪isnan(row['Correlation']):
                        corr_feature.add(row['Col'])
        return corr_feature
```

[ ]:
```
corr_feature = correlation(corr_col,.02)
len(set(corr_feature))
```

[ ]: 5

[ ]:
```
corr_feature
```

[ ]: {'Favicon', 'Iframe', 'RightClick', 'Submitting_to_email', 'popUpWidnow'}

[ ]:
```
df.drop(corr_feature, axis=1, inplace=True)
```

[ ]:
```
# # Remove features having correlation coeff. between +/- 0.03
# df.drop(['Favicon','Iframe','Redirect',
#                   'popUpWidnow','RightClick','Submitting_to_email'], axis=1,␣
↪inplace=True)
```

[ ]:
```
len(df.columns)
```

[ ]: 26

[ ]:
```
#df.head()
```

[ ]:
```
a=len(df[df.Result==0])
b=len(df[df.Result==1])
```

[ ]:
```
print("Count of Legitimate Websites = ", a)
print("Count of Phishy Websites = ", b)
```

```
Count of Legitimate Websites =  4898
Count of Phishy Websites =  6157
```

[ ]:
```
# df.corr()
```

[ ]:
```
# #Using Pearson Correlation
# plt.figure(figsize=(30,30))
# corr = df.corr()
# sns.heatmap(corr, annot=True, cmap=plt.cm.CMRmap_r)
# plt.show()
```

[ ]:
```
# # with the following function we can select highly correlated features
# # it will remove the first feature that is correlated with anything other␣
↪feature

# def correlation(dataset, threshold):
#     col_corr = set()  # Set of all the names of correlated columns
#     corr_matrix = dataset.corr()
#     for i in range(len(corr_matrix.columns)):
#         for j in range(i):
```

```
#             if abs(corr_matrix.iloc[i, j]) > threshold: # we are interested␣
↪in absolute coeff value
#                 colname = corr_matrix.columns[i]  # getting the name of column
#                 col_corr.add(colname)
#     return col_corr
```

[ ]:
```
# corr_features = correlation(df, 0.8)
# len(set(corr_features))
```

[ ]:
```
# corr_features
```

[ ]:
```
#df.head()
```

[ ]:
```
#from sklearn import preprocessing


# col =df[df.columns[:]]

# lab_en= preprocessing.LabelEncoder()

# for c in col:
#     df[c]= lab_en.fit_transform(df[c])

# df.head()
```

[ ]:
```
X = df.drop(['Result'], axis=1, inplace=False)
#X.head()
#same work
##inplace true modifies the og data & does not return anything
##inplace false does not modify og data but returns something whoch we store in␣
↪a var
# X= df.drop(columns='Result')
# X.head()
```

[ ]:
```
#df.head()
```

[ ]:
```
y = df['Result']
y = pd.DataFrame(y)
y.head()
```

[ ]:

	Result
0	0
1	0
2	0
3	0
4	1

[ ]:
```
# separate dataset into train and test
from cProfile import label
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
    X,
    y,
    test_size=0.1,
    random_state=10)

X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

[ ]:
```
((9949, 25), (1106, 25), (9949, 1), (1106, 1))
```

[ ]:
```
#X_test.head()
```

[ ]:
```
print("Training set has {} samples.".format(X_train.shape[0]))
print("Testing set has {} samples.".format(X_test.shape[0]))
```

```
Training set has 9949 samples.
Testing set has 1106 samples.
```

[ ]:
```
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LogisticRegression

# defining parameter range
param_grid = {'penalty' : ['l2'],
             'C' : [0.1, 1, 10, 20, 30],
             'solver' : ['lbfgs','newton-cg','liblinear','sag','saga'],
             'max_iter' : [2500, 5000]}

grid_logr = GridSearchCV(LogisticRegression(), param_grid, refit = True, cv =
 ↪10, verbose = 3, n_jobs = -1)

# fitting the model for grid search
grid_logr.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_logr.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_logr.best_estimator_)
print(grid_logr.best_score_)
```

```
Fitting 10 folds for each of 50 candidates, totalling 500 fits
{'C': 1, 'max_iter': 2500, 'penalty': 'l2', 'solver': 'liblinear'}
LogisticRegression(C=1, max_iter=2500, solver='liblinear')
0.9286367450936777
```

[ ]:
```python
logr_model = grid_logr.best_estimator_

# Performing training
#logr_model = logr.fit(X_train, y_train.values.ravel())
```

[ ]:
```python
logr_predict  =  logr_model.predict(X_test)
```

[ ]:
```python
# from sklearn.metrics import confusion_matrix,accuracy_score
# cm = confusion_matrix(y_test, dct_pred)
# ac = accuracy_score(y_test, dct_pred)
```

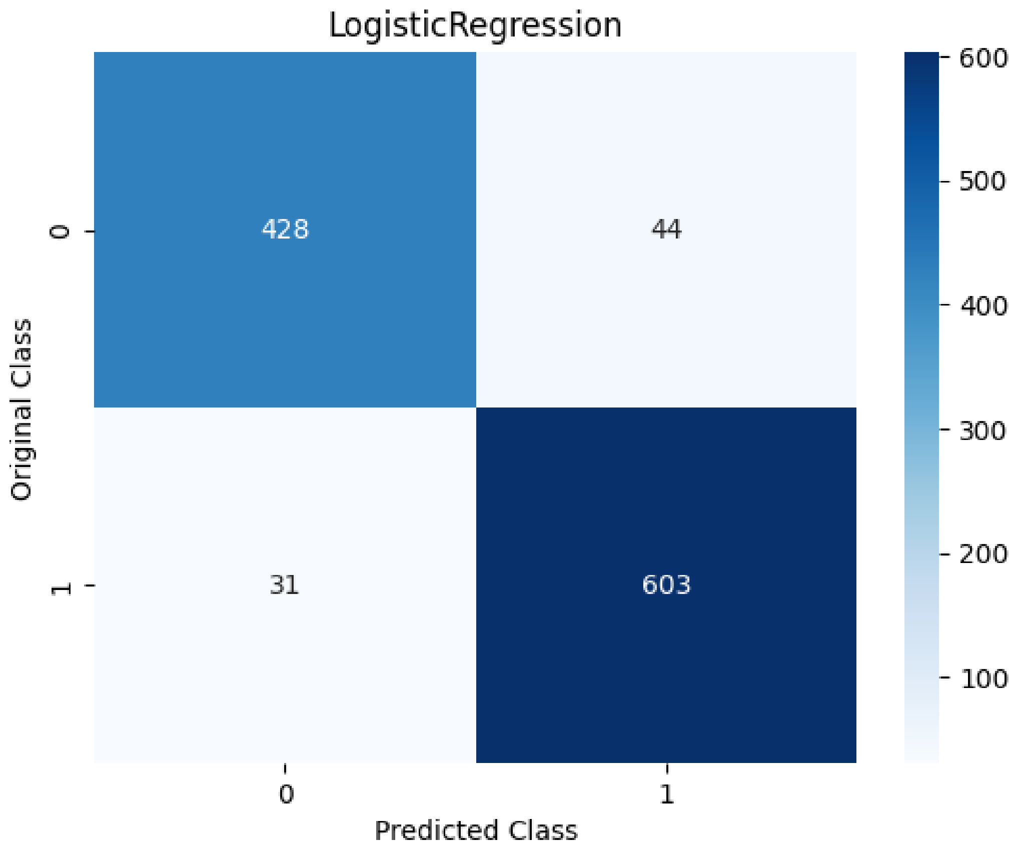
[ ]:
```python
print ("Accuracy of logr classifier : ", accuracy_score(y_test,␣
↪logr_predict)*100)
```

```
Accuracy of logr classifier :  93.21880650994575
```

[ ]:
```python
print(classification_report(y_test, logr_predict))
```

```
              precision    recall  f1-score   support

           0       0.93      0.91      0.92       472
           1       0.93      0.95      0.94       634

    accuracy                           0.93      1106
   macro avg       0.93      0.93      0.93      1106
weighted avg       0.93      0.93      0.93      1106
```

[ ]:
```python
sns.heatmap(confusion_matrix(y_test, logr_predict), annot=True, fmt='g',␣
↪cmap='Blues')
plt.title("LogisticRegression")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```




```
[ ]: # from sklearn.neighbors import KNeighborsClassifier

     # #training_accuracy=[]
     # test_accuracy=[]

     # neighbors=range(1,10)
     # ##values.ravel() converts vector y to flattened array
     # for i in neighbors:
     #     knn=KNeighborsClassifier(n_neighbors=i)
     #     knn_model = knn.fit(X_train,y_train.values.ravel())
     #     #training_accuracy.append(knn.score(X_train,y_train.values.ravel()))
     #     test_accuracy.append(knn_model.score(X_test,y_test.values.ravel()))
```

```
[ ]: # plt.plot(neighbors,test_accuracy,label="test accuracy")
     # plt.ylabel("Accuracy")
     # plt.xlabel("number of neighbors")
     # plt.legend()
     # plt.show()
```

[ ]:
```python
from sklearn.neighbors import KNeighborsClassifier

# defining parameter range
param_grid = {'n_neighbors': [1,2,3,4,5,6,7,8,9,10]}

grid_knn = GridSearchCV(KNeighborsClassifier(), param_grid, refit = True, cv =
 ↪10, verbose = 3, n_jobs = -1)

# fitting the model for grid search
grid_knn.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_knn.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_knn.best_estimator_)
print(grid_knn.best_score_)
```

```
Fitting 10 folds for each of 10 candidates, totalling 100 fits
{'n_neighbors': 1}
KNeighborsClassifier(n_neighbors=1)
0.9614035974641819
```

[ ]:
```python
knn_model = grid_knn.best_estimator_
#knn_model = knn.fit(X_train,y_train.values.ravel())
```

[ ]:
```python
#print ("Accuracy of knn classifier: ", max(test_accuracy)*100)
knn_predict = knn_model.predict(X_test)
```

[ ]:
```python
print('The accuracy of knn Classifier is: ', 100.0 * accuracy_score(y_test,
 ↪knn_predict))
```

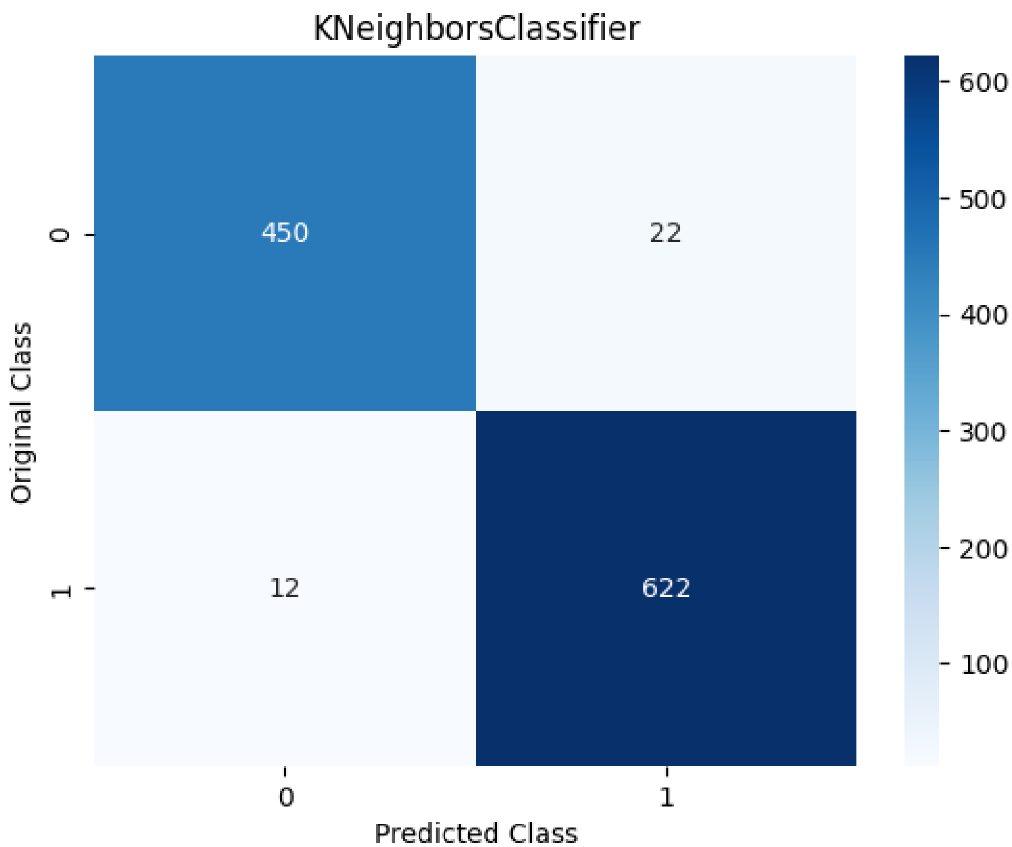
```
The accuracy of knn Classifier is:  96.9258589511754
```

[ ]:
```python
print(classification_report(y_test, knn_predict))
```

```
              precision    recall  f1-score   support

           0       0.97      0.95      0.96       472
           1       0.97      0.98      0.97       634

    accuracy                           0.97      1106
   macro avg       0.97      0.97      0.97      1106
weighted avg       0.97      0.97      0.97      1106
```

[ ]:
```python
sns.heatmap(confusion_matrix(y_test, knn_predict), annot=True, fmt='g',
 ↪cmap='Blues')
plt.title("KNeighborsClassifier")
```

```
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```



[ ]:
```
# # here is the change
# knn_y_pred_proba = knn.predict_proba(X_test)
# knn_y_pred_proba_positive = knn_y_pred_proba[:, 1]

# RocCurveDisplay.from_predictions(y_test,knn_y_pred_proba_positive)

# fig, ax = plt.subplots()
# RocCurveDisplay.from_estimator(
#     logreg, X_test, y_test, ax = ax)

# logreg_y_decision = logreg.decision_function(X_test)
# metrics.RocCurveDisplay.
↪from_predictions(y_test,logreg_y_decision,ax=ax,name="logreg predictions")
```

[ ]:
```python
from sklearn.svm import SVC

# defining parameter range
param_grid = {'C': [0.1, 1, 10],
              'gamma': [1, 0.1, 0.01],
              'kernel': ['linear','poly', 'rbf', 'sigmoid']}

grid_svc = GridSearchCV(SVC(), param_grid, refit = True, cv = 10, verbose = 3,␣
 ↪n_jobs = -1)

# fitting the model for grid search
grid_svc.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_svc.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_svc.best_estimator_)
print(grid_svc.best_score_)
```

```
Fitting 10 folds for each of 36 candidates, totalling 360 fits
{'C': 10, 'gamma': 1, 'kernel': 'rbf'}
SVC(C=10, gamma=1)
0.9668303287059038
```

[ ]:
```python
svc_model = grid_svc.best_estimator_
#svc_model = svc.fit(X_train,y_train.values.ravel())
```

[ ]:
```python
svc_predict = svc_model.predict(X_test)
```

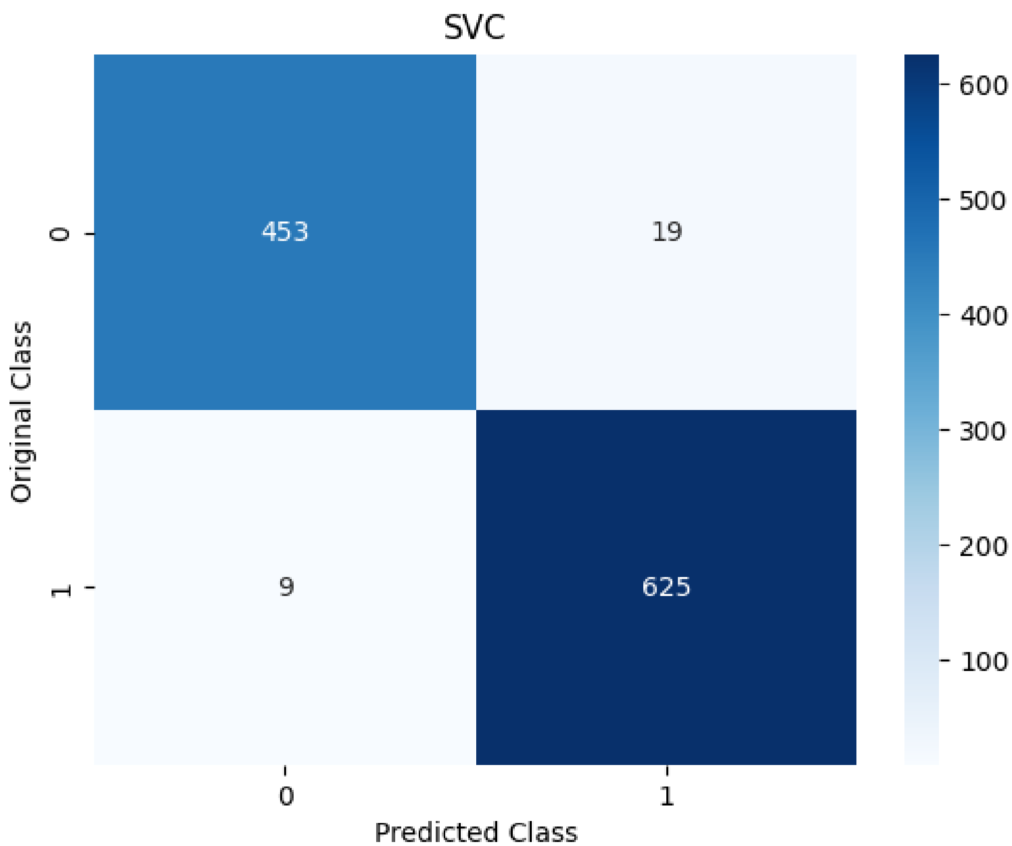
[ ]:
```python
print('The accuracy of svc Classifier is: ', 100.0 * accuracy_score(y_test,␣
 ↪svc_predict))
```

```
The accuracy of svc Classifier is:  97.46835443037975
```

[ ]:
```python
print(classification_report(y_test, svc_predict))
```

```
              precision    recall  f1-score   support

           0       0.98      0.96      0.97       472
           1       0.97      0.99      0.98       634

    accuracy                           0.97      1106
   macro avg       0.98      0.97      0.97      1106
weighted avg       0.97      0.97      0.97      1106
```

```
[ ]: sns.heatmap(confusion_matrix(y_test, svc_predict), annot=True, fmt='g',␣
     ↪cmap='Blues')
     plt.title("SVC")
     plt.xlabel('Predicted Class')
     plt.ylabel('Original Class')
     plt.show()
```




```
[ ]: from sklearn.svm import NuSVC

     # defining parameter range
     param_grid = {'nu': [0.1, 0.5],
                   'gamma': [1, 0.1, 0.01],
                   'kernel': ['linear','poly', 'rbf', 'sigmoid']}

     grid_nusvc = GridSearchCV(NuSVC(), param_grid, refit = True, verbose = 3, cv =␣
     ↪10, n_jobs = -1)

     # fitting the model for grid search
     grid_nusvc.fit(X_train, y_train.values.ravel())
```

```
# print best parameter after tuning
print(grid_nusvc.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_nusvc.best_estimator_)
print(grid_nusvc.best_score_)
```

```
Fitting 10 folds for each of 24 candidates, totalling 240 fits
{'gamma': 1, 'kernel': 'rbf', 'nu': 0.1}
NuSVC(gamma=1, nu=0.1)
0.9669308312184665
```

[ ]:
```
nusvc_model = grid_nusvc.best_estimator_
#nusvc_model = nusvc.fit(X_train, y_train.values.ravel())
```

[ ]:
```
nusvc_predict = nusvc_model.predict(X_test)
```

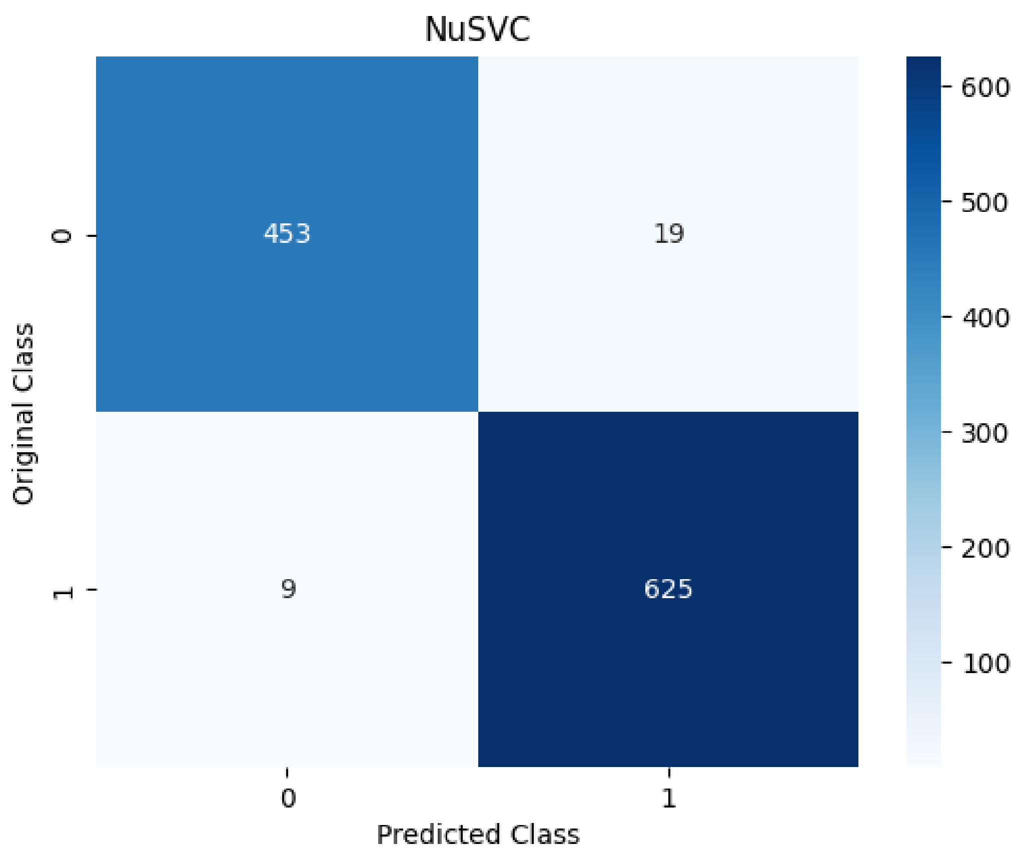
[ ]:
```
print('The accuracy of nusvc Classifier is: ', 100.0 * accuracy_score(y_test,␣
↪nusvc_predict))
```

```
The accuracy of nusvc Classifier is:  97.46835443037975
```

[ ]:
```
print(classification_report(y_test, nusvc_predict))
```

```
              precision    recall  f1-score   support

           0       0.98      0.96      0.97       472
           1       0.97      0.99      0.98       634

    accuracy                           0.97      1106
   macro avg       0.98      0.97      0.97      1106
weighted avg       0.97      0.97      0.97      1106
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test, nusvc_predict), annot=True, fmt='g',␣
↪cmap='Blues')
plt.title("NuSVC")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```



[ ]:
```
from sklearn.svm import LinearSVC

# defining parameter range
param_grid = {'C': [0.1, 1, 10, 20, 30],
                        'penalty': ['l1','l2'],
                        'loss': ['squared_hinge'],
                        'dual': [False],
                        'tol': [.1,.01,.001]}

grid_lsvc = GridSearchCV(LinearSVC(), param_grid, refit = True, verbose = 3, cv␣
↪= 10, n_jobs = -1)

# fitting the model for grid search
grid_lsvc.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_lsvc.best_params_)

# print how our model looks after hyper-parameter tuning
```

```
print(grid_lsvc.best_estimator_)
print(grid_lsvc.best_score_)
```

```
Fitting 10 folds for each of 30 candidates, totalling 300 fits
{'C': 1, 'dual': False, 'loss': 'squared_hinge', 'penalty': 'l2', 'tol': 0.01}
LinearSVC(C=1, dual=False, tol=0.01)
0.9289383537405337
```

[ ]:
```
lsvc_model = grid_lsvc.best_estimator_
#lsvc_model = lsvc.fit(X_train, y_train.values.ravel())
```

[ ]:
```
lsvc_predict = lsvc_model.predict(X_test)
```

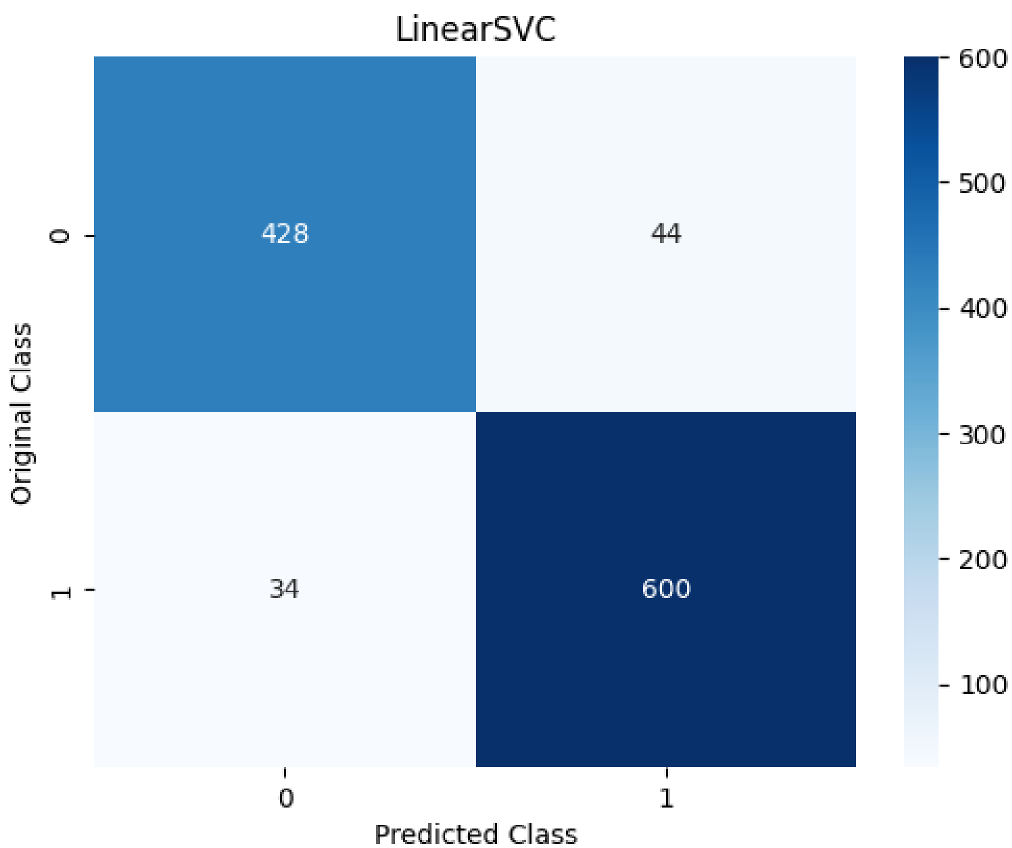
[ ]:
```
print('The accuracy of lsvc Classifier is: ', 100.0 * accuracy_score(y_test,␣
 ↪lsvc_predict))
```

```
The accuracy of lsvc Classifier is:  92.94755877034359
```

[ ]:
```
print(classification_report(y_test, lsvc_predict))
```

```
              precision    recall  f1-score   support

           0       0.93      0.91      0.92       472
           1       0.93      0.95      0.94       634

    accuracy                           0.93      1106
   macro avg       0.93      0.93      0.93      1106
weighted avg       0.93      0.93      0.93      1106
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test, lsvc_predict), annot=True, fmt='g',␣
 ↪cmap='Blues')
plt.title("LinearSVC")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```




```
[ ]: from sklearn.ensemble import AdaBoostClassifier

     # defining parameter range
     param_grid = {'n_estimators': [40,50,100,200,300]}

     grid_ada = GridSearchCV(AdaBoostClassifier(), param_grid, refit = True, verbose␣
      ↪= 3, cv = 10, n_jobs = -1)

     # fitting the model for grid search
     grid_ada.fit(X_train, y_train.values.ravel())

     # print best parameter after tuning
     print(grid_ada.best_params_)

     # print how our model looks after hyper-parameter tuning
     print(grid_ada.best_estimator_)
     print(grid_ada.best_score_)
```

```
Fitting 10 folds for each of 5 candidates, totalling 50 fits
{'n_estimators': 100}
```

```
AdaBoostClassifier(n_estimators=100)
0.9371798630981871
```

[ ]:
```
ada_model = grid_ada.best_estimator_
#ada_model = ada.fit(X_train,y_train.values.ravel())
```

[ ]:
```
ada_predict = ada_model.predict(X_test)
```

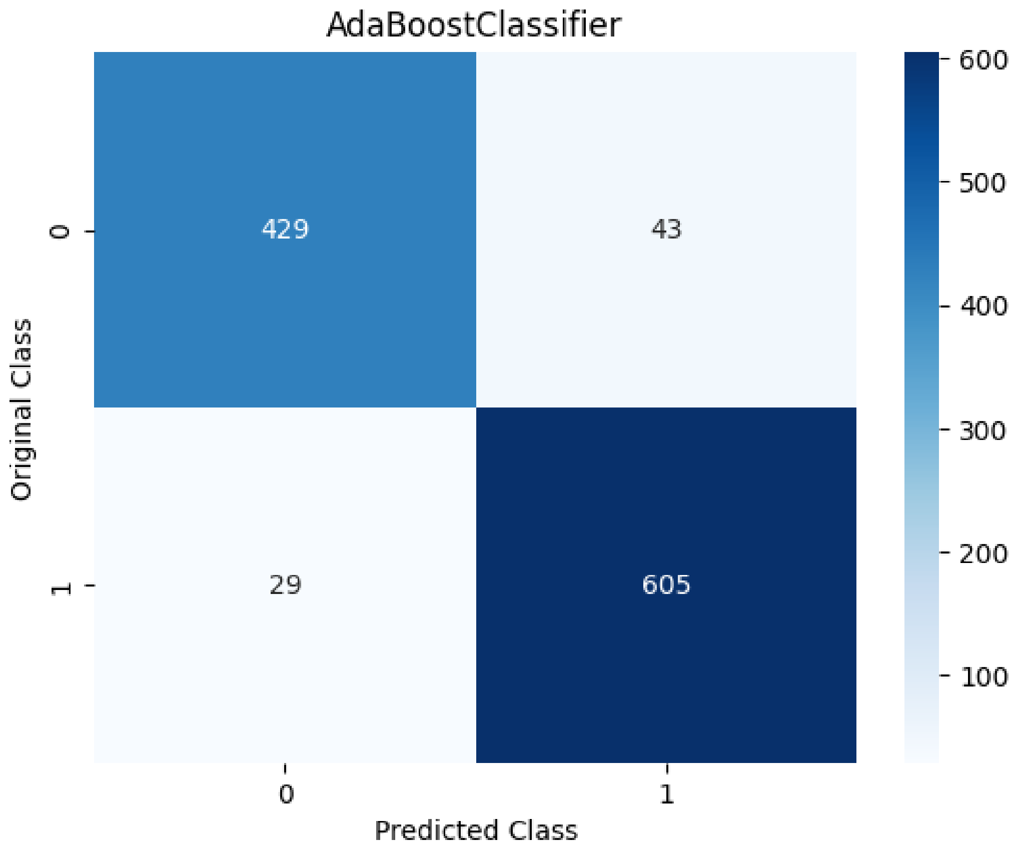
[ ]:
```
print('The accuracy of Ada Boost Classifier is: ', 100.0 *␣
↪accuracy_score(ada_predict,y_test))
```

```
The accuracy of Ada Boost Classifier is:  93.49005424954791
```

[ ]:
```
print(classification_report(y_test, ada_predict))
```

```
              precision    recall  f1-score   support

           0       0.94      0.91      0.92       472
           1       0.93      0.95      0.94       634

    accuracy                           0.93      1106
   macro avg       0.94      0.93      0.93      1106
weighted avg       0.93      0.93      0.93      1106
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test, ada_predict), annot=True, fmt='g',␣
↪cmap='Blues')
plt.title("AdaBoostClassifier")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```



[ ]:
```python
from xgboost import XGBClassifier


# defining parameter range
param_grid = {
    "gamma": [.01, .1, .5],
    "n_estimators": [50,100,150,200,250]
}

grid_xgb = GridSearchCV(XGBClassifier(), param_grid, refit = True, verbose = 3,␣
 ↪cv = 10, n_jobs = -1)

# fitting the model for grid search
grid_xgb.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_xgb.best_params_)

# print how our model looks after hyper-parameter tuning
```

```python
print(grid_xgb.best_estimator_)
print(grid_xgb.best_score_)
```

```
Fitting 10 folds for each of 15 candidates, totalling 150 fits
{'gamma': 0.01, 'n_estimators': 250}
XGBClassifier(base_score=0.5, booster='gbtree', callbacks=None,
              colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1,
              early_stopping_rounds=None, enable_categorical=False,
              eval_metric=None, gamma=0.01, gpu_id=-1, grow_policy='depthwise',
              importance_type=None, interaction_constraints='',
              learning_rate=0.300000012, max_bin=256, max_cat_to_onehot=4,
              max_delta_step=0, max_depth=6, max_leaves=0, min_child_weight=1,
              missing=nan, monotone_constraints='()', n_estimators=250,
              n_jobs=0, num_parallel_tree=1, predictor='auto', random_state=0,
              reg_alpha=0, reg_lambda=1, …)
0.9722577677117984
```

[ ]:
```python
xgb_model = grid_xgb.best_estimator_
#xgb_model = xgb.fit(X_train,y_train)
```

[ ]:
```python
xgb_predict=xgb_model.predict(X_test)
```

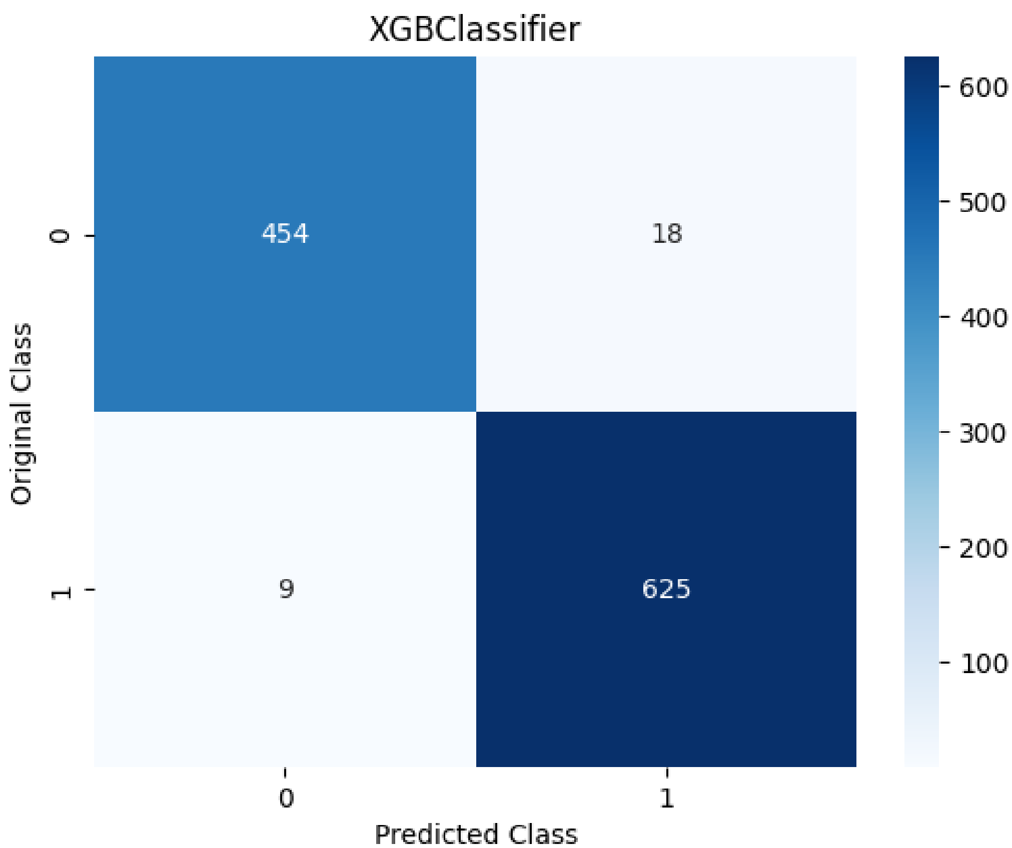
[ ]:
```python
print('The accuracy of XGBoost Classifier is: ' , 100.0 *␣
↪accuracy_score(xgb_predict,y_test))
```

```
The accuracy of XGBoost Classifier is:  97.55877034358048
 97.55877034358048
```

[ ]:
```python
print(classification_report(y_test, xgb_predict))
```

```
              precision    recall  f1-score   support

           0       0.98      0.96      0.97       472
           1       0.97      0.99      0.98       634

    accuracy                           0.98      1106
   macro avg       0.98      0.97      0.97      1106
weighted avg       0.98      0.98      0.98      1106
```

[ ]:
```python
sns.heatmap(confusion_matrix(y_test, xgb_predict), annot=True, fmt='g',␣
↪cmap='Blues')
plt.title("XGBClassifier")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```



[ ]:
```
from sklearn.ensemble import GradientBoostingClassifier

# defining parameter range
param_grid = {
    "learning_rate": [.1,.5,1],
    "n_estimators": [50,100,150,200,250]
}

grid_gbc = GridSearchCV(GradientBoostingClassifier(), param_grid, refit = True,␣
 ↪verbose = 3, cv = 10, n_jobs = -1)

# fitting the model for grid search
grid_gbc.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_gbc.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_gbc.best_estimator_)
```

```
print(grid_gbc.best_score_)
```

```
Fitting 10 folds for each of 15 candidates, totalling 150 fits
{'learning_rate': 1, 'n_estimators': 250}
GradientBoostingClassifier(learning_rate=1, n_estimators=250)
0.9696443990576624
```

[ ]:
```
gbc_model = grid_gbc.best_estimator_
#gbc_model = gbc.fit(X_train,y_train.values.ravel())

#clf = GradientBoostingClassifier(n_estimators=100, learning_rate=1.0,
#    max_depth=1, random_state=0).fit(X_train, y_train)
#clf.score(X_test, y_test)
```

[ ]:
```
gbc_predict = gbc_model.predict(X_test)
```

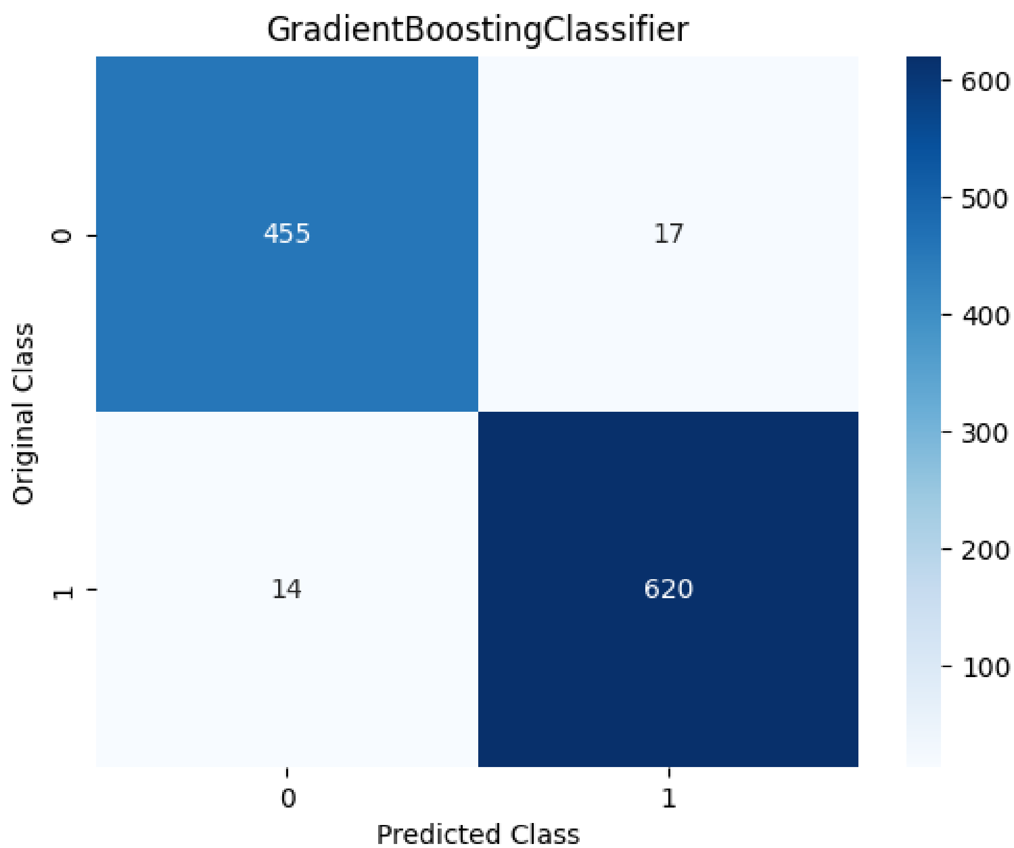
[ ]:
```
print('The accuracy of GradientBoost Classifier is: ' , 100.0 *␣
↪accuracy_score(gbc_predict,y_test))
```

```
The accuracy of GradientBoost Classifier is:  97.19710669077757
```

[ ]:
```
print(classification_report(y_test, gbc_predict))
```

```
              precision    recall  f1-score   support

           0       0.97      0.96      0.97       472
           1       0.97      0.98      0.98       634

    accuracy                           0.97      1106
   macro avg       0.97      0.97      0.97      1106
weighted avg       0.97      0.97      0.97      1106
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test, gbc_predict), annot=True, fmt='g',␣
↪cmap='Blues')
plt.title("GradientBoostingClassifier")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```



[ ]:
```
# gbc_model.get_params().keys()
```

[ ]:
```
# import inspect
# import sklearn
# import xgboost

# models = [xgboost.XGBClassifier]
# for m in models:
#     hyperparams = inspect.signature(m.__init__)
#     print(hyperparams)
# #or
# xgb_model.get_params().keys()
```

[ ]:
```
from sklearn.ensemble import BaggingClassifier
from sklearn.tree import DecisionTreeClassifier

# defining parameter range
param_grid = {
    "base_estimator": [DecisionTreeClassifier()],
    "n_estimators": [50,100,150,200,250]
```

```
}

grid_bag = GridSearchCV(BaggingClassifier(), param_grid, refit = True, verbose␣
↪= 3, cv = 10, n_jobs = -1)

# fitting the model for grid search
grid_bag.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_bag.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_bag.best_estimator_)
print(grid_bag.best_score_)
```

```
Fitting 10 folds for each of 5 candidates, totalling 50 fits
{'base_estimator': DecisionTreeClassifier(), 'n_estimators': 200}
BaggingClassifier(base_estimator=DecisionTreeClassifier(), n_estimators=200)
0.9689408814697229
```

[ ]:
```
bag_model = grid_bag.best_estimator_
#bag_model = bag.fit(X_train, y_train.values.ravel())
```

[ ]:
```
bag_predict = bag_model.predict(X_test)
```

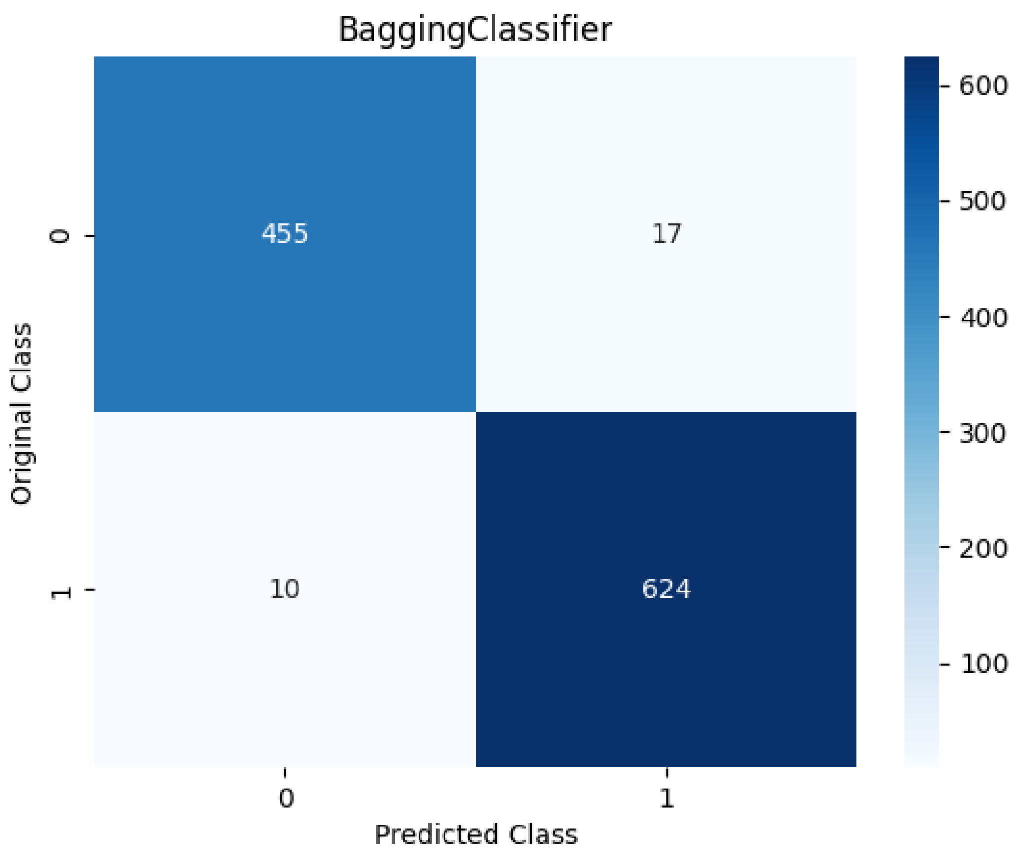
[ ]:
```
print('The accuracy of Bagging Classifier is: ' , 100.0 *␣
↪accuracy_score(y_test, bag_predict))
```

```
The accuracy of Bagging Classifier is:  97.55877034358048
```

[ ]:
```
print(classification_report(y_test, bag_predict))
```

```
              precision    recall  f1-score   support

           0       0.98      0.96      0.97       472
           1       0.97      0.98      0.98       634

    accuracy                           0.98      1106
   macro avg       0.98      0.97      0.98      1106
weighted avg       0.98      0.98      0.98      1106
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test, bag_predict), annot=True, fmt='g',␣
↪cmap='Blues')
plt.title("BaggingClassifier")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```



[ ]:
```
from sklearn.ensemble import RandomForestClassifier

# defining parameter range
param_grid = {
    "n_estimators": [50,100,150,200,250]
}

grid_rfc = GridSearchCV(RandomForestClassifier(), param_grid, refit = True,␣
 ↪verbose = 3, cv = 10, n_jobs = -1)

# fitting the model for grid search
grid_rfc.fit(X_train, y_train.values.ravel())

# print best parameter after tuning
print(grid_rfc.best_params_)

# print how our model looks after hyper-parameter tuning
print(grid_rfc.best_estimator_)
print(grid_rfc.best_score_)
```

```
Fitting 10 folds for each of 5 candidates, totalling 50 fits
{'n_estimators': 200}
RandomForestClassifier(n_estimators=200)
0.970649828619961
```

[ ]:
```
rfc_model = grid_rfc.best_estimator_
#rfc_model = rfc.fit(X_train,y_train.values.ravel())
```

[ ]:
```
rfc_predict = rfc_model.predict(X_test)
```

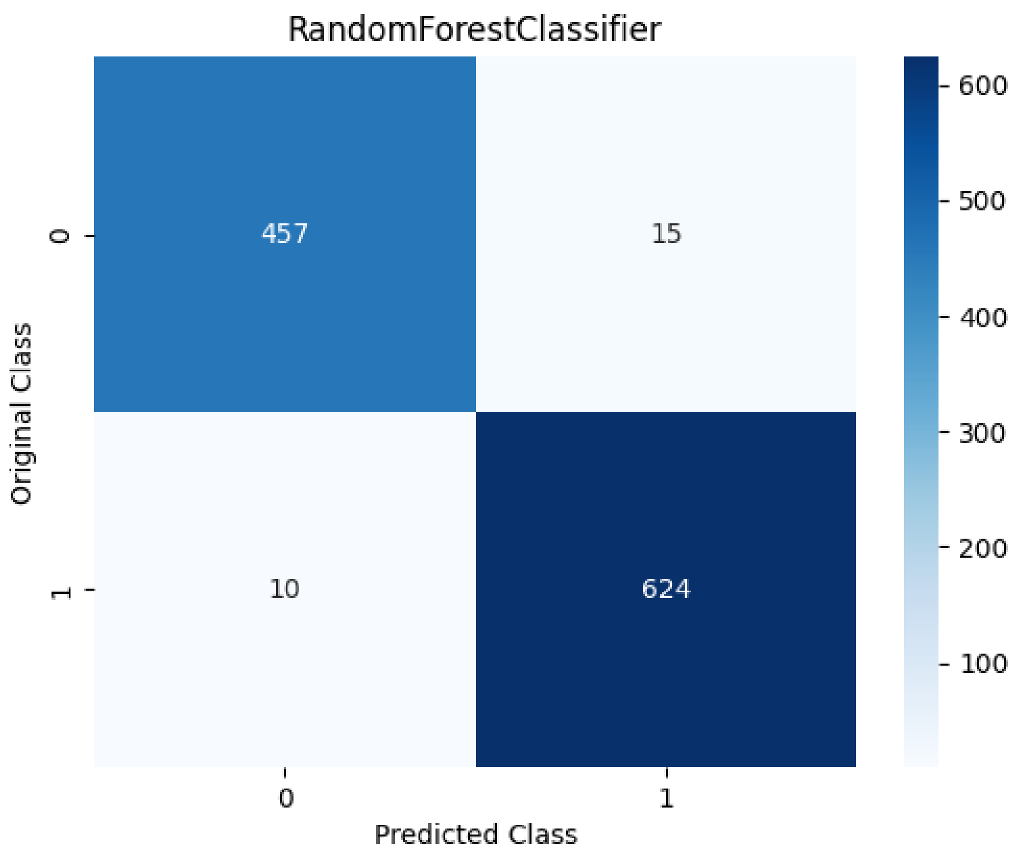
[ ]:
```
print('The accuracy of RandomForest Classifier is: ' , 100.0 *␣
↪accuracy_score(rfc_predict,y_test))
```

```
The accuracy of RandomForest Classifier is:  97.73960216998192
```

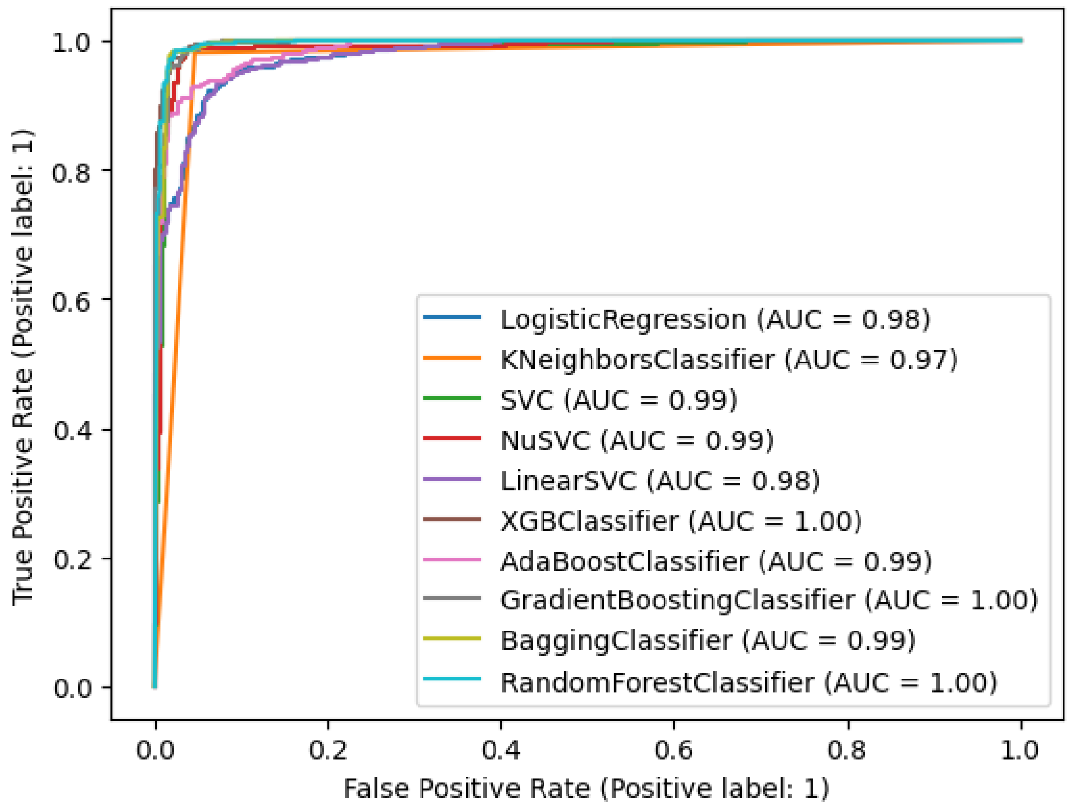
[ ]:
```
print(classification_report(y_test, rfc_predict))
```

```
              precision    recall  f1-score   support

           0       0.98      0.97      0.97       472
           1       0.98      0.98      0.98       634

    accuracy                           0.98      1106
   macro avg       0.98      0.98      0.98      1106
weighted avg       0.98      0.98      0.98      1106
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test, rfc_predict), annot=True, fmt='g',␣
↪cmap='Blues')
plt.title("RandomForestClassifier")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```




```
[ ]: estimators =␣
     ↪[logr_model,knn_model,svc_model,nusvc_model,lsvc_model,xgb_model,ada_model,gbc_model,bag_mo

     for estimator in estimators:
         RocCurveDisplay.from_estimator(estimator,X_test,y_test,ax=plt.gca())
```



[ ]:
```
import tensorflow as tf
#from tensorflow.keras.datasets import imdb
from keras.layers import Embedding, Dense, LSTM, BatchNormalization
from keras.losses import BinaryCrossentropy
from keras.models import Sequential
from keras.optimizers import Adam
#from tensorflow.keras.preprocessing.sequence import pad_sequences

# Model configuration
additional_metrics = ['accuracy']
batch_size = 32
#embedding_output_dims = (X_train.shape[1])
loss_function = BinaryCrossentropy()
#max_sequence_length = (X_train.shape[1])
#num_distinct_words = (X_train.shape[1])
number_of_epochs = 100
optimizer = Adam()
validation_split = 0.20
verbosity_mode = 1

# reshape from [samples, features] into [samples, timesteps, features]
```

```
timesteps = 1
X_train_reshape = X_train.values.ravel().reshape(X_train.shape[0],timesteps,␣
↪X_train.shape[1])
X_test_reshape = X_test.values.ravel().reshape(X_test.shape[0],timesteps,␣
↪X_test.shape[1])

# Disable eager execution
#tf.compat.v1.disable_eager_execution()

# Load dataset
# (x_train, y_train), (x_test, y_test) = imdb.
↪load_data(num_words=num_distinct_words)
# print(x_train.shape)
# print(x_test.shape)

# Pad all sequences
# padded_inputs = pad_sequences(X_train, maxlen=max_sequence_length, value = 0.
↪0) # 0.0 because it corresponds with <PAD>
# padded_inputs_test = pad_sequences(X_test, maxlen=max_sequence_length, value␣
↪= 0.0) # 0.0 because it corresponds with <PAD>

# Define the Keras model
def build_model_lstm():
    model = Sequential()
    #model.add(Embedding(num_distinct_words, embedding_output_dims,␣
↪input_length=max_sequence_length))
    model.add(LSTM(100, input_shape = (timesteps,X_train_reshape.shape[2])))
    model.add(BatchNormalization())
    model.add(Dense(50, activation='relu'))
    model.add(Dense(25, activation='relu'))
    model.add(Dense(10, activation='relu'))
    model.add(Dense(1, activation='sigmoid'))

    # Compile the model
    model.compile(optimizer=optimizer, loss=loss_function,␣
↪metrics=additional_metrics)
    return model

#from keras.wrappers.scikit_learn import KerasClassifier
lstm_model = build_model_lstm()
# Give a summary
lstm_model.summary()

# Train the model
```

```python
history = lstm_model.fit(X_train_reshape, y_train.values.ravel(),␣
    ↪batch_size=batch_size, epochs=number_of_epochs, verbose=verbosity_mode,␣
    ↪validation_split=validation_split)

# Test the model after training
#lstm_predict = lstm_model.predict(X_test_reshape)
test_results = lstm_model.evaluate(X_test_reshape, y_test.values.ravel(),␣
    ↪verbose=False)
print(f'Test results - Loss: {test_results[0]} - Accuracy:␣
    ↪{100*test_results[1]}%')
```

```
Model: "sequential_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_3 (LSTM)               (None, 100)               50400

 batch_normalization_3 (Batc  (None, 100)              400
 hNormalization)

 dense_12 (Dense)            (None, 50)                5050
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_3 (LSTM)               (None, 100)               50400

 batch_normalization_3 (Batc  (None, 100)              400
 hNormalization)

 dense_12 (Dense)            (None, 50)                5050

 dense_13 (Dense)            (None, 25)                1275

 dense_14 (Dense)            (None, 10)                260

 dense_15 (Dense)            (None, 1)                 11

=================================================================
Total params: 57,396
Trainable params: 57,196
Non-trainable params: 200
_________________________________________________________________
Epoch 1/100
249/249 [==============================] - 4s 7ms/step - loss: 0.2264 -
accuracy: 0.9153 - val_loss: 0.3250 - val_accuracy: 0.9312
Epoch 2/100
249/249 [==============================] - 1s 5ms/step - loss: 0.1627 -
```

```
accuracy: 0.9306 - val_loss: 0.1677 - val_accuracy: 0.9347
Epoch 3/100
249/249 [==============================] - 1s 5ms/step - loss: 0.1472 -
accuracy: 0.9372 - val_loss: 0.1360 - val_accuracy: 0.9422
Epoch 4/100
249/249 [==============================] - 1s 4ms/step - loss: 0.1371 -
accuracy: 0.9420 - val_loss: 0.1214 - val_accuracy: 0.9447
Epoch 5/100
249/249 [==============================] - 1s 4ms/step - loss: 0.1275 -
accuracy: 0.9472 - val_loss: 0.1264 - val_accuracy: 0.9487
Epoch 6/100
249/249 [==============================] - 1s 5ms/step - loss: 0.1179 -
accuracy: 0.9520 - val_loss: 0.1253 - val_accuracy: 0.9533
Epoch 7/100
249/249 [==============================] - 1s 4ms/step - loss: 0.1106 -
accuracy: 0.9520 - val_loss: 0.1001 - val_accuracy: 0.9578
Epoch 8/100
249/249 [==============================] - 1s 4ms/step - loss: 0.1019 -
accuracy: 0.9583 - val_loss: 0.0936 - val_accuracy: 0.9628
Epoch 9/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0933 -
accuracy: 0.9626 - val_loss: 0.1031 - val_accuracy: 0.9578
Epoch 10/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0948 -
accuracy: 0.9597 - val_loss: 0.0955 - val_accuracy: 0.9573
Epoch 11/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0870 -
accuracy: 0.9637 - val_loss: 0.1074 - val_accuracy: 0.9543
Epoch 12/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0847 -
accuracy: 0.9639 - val_loss: 0.1048 - val_accuracy: 0.9548
Epoch 13/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0805 -
accuracy: 0.9662 - val_loss: 0.0997 - val_accuracy: 0.9583
Epoch 14/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0785 -
accuracy: 0.9676 - val_loss: 0.0955 - val_accuracy: 0.9643
Epoch 15/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0796 -
accuracy: 0.9651 - val_loss: 0.0999 - val_accuracy: 0.9633
Epoch 16/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0783 -
accuracy: 0.9688 - val_loss: 0.0960 - val_accuracy: 0.9618
Epoch 17/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0712 -
accuracy: 0.9705 - val_loss: 0.0909 - val_accuracy: 0.9643
Epoch 18/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0688 -
```

```
accuracy: 0.9716 - val_loss: 0.1003 - val_accuracy: 0.9593
Epoch 19/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0691 -
accuracy: 0.9711 - val_loss: 0.0968 - val_accuracy: 0.9688
Epoch 20/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0675 -
accuracy: 0.9703 - val_loss: 0.0954 - val_accuracy: 0.9618
Epoch 21/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0647 -
accuracy: 0.9710 - val_loss: 0.0996 - val_accuracy: 0.9623
Epoch 22/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0612 -
accuracy: 0.9737 - val_loss: 0.1011 - val_accuracy: 0.9623
Epoch 23/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0634 -
accuracy: 0.9722 - val_loss: 0.0909 - val_accuracy: 0.9638
Epoch 24/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0642 -
accuracy: 0.9721 - val_loss: 0.1035 - val_accuracy: 0.9663
Epoch 25/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0628 -
accuracy: 0.9755 - val_loss: 0.1050 - val_accuracy: 0.9623
Epoch 26/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0573 -
accuracy: 0.9759 - val_loss: 0.0959 - val_accuracy: 0.9623
Epoch 27/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0630 -
accuracy: 0.9737 - val_loss: 0.1088 - val_accuracy: 0.9603
Epoch 28/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0594 -
accuracy: 0.9750 - val_loss: 0.0900 - val_accuracy: 0.9678
Epoch 29/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0562 -
accuracy: 0.9769 - val_loss: 0.0994 - val_accuracy: 0.9643
Epoch 30/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0540 -
accuracy: 0.9778 - val_loss: 0.1008 - val_accuracy: 0.9673
Epoch 31/100
249/249 [==============================] - 1s 6ms/step - loss: 0.0524 -
accuracy: 0.9779 - val_loss: 0.0982 - val_accuracy: 0.9638
Epoch 32/100
249/249 [==============================] - 2s 6ms/step - loss: 0.0544 -
accuracy: 0.9761 - val_loss: 0.1061 - val_accuracy: 0.9648
Epoch 33/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0517 -
accuracy: 0.9795 - val_loss: 0.0979 - val_accuracy: 0.9668
Epoch 34/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0487 -
```

```
accuracy: 0.9784 - val_loss: 0.0958 - val_accuracy: 0.9678
Epoch 35/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0508 -
accuracy: 0.9784 - val_loss: 0.0951 - val_accuracy: 0.9663
Epoch 36/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0506 -
accuracy: 0.9788 - val_loss: 0.0981 - val_accuracy: 0.9678
Epoch 37/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0516 -
accuracy: 0.9789 - val_loss: 0.0963 - val_accuracy: 0.9673
Epoch 38/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0486 -
accuracy: 0.9774 - val_loss: 0.1457 - val_accuracy: 0.9633
Epoch 39/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0500 -
accuracy: 0.9789 - val_loss: 0.0920 - val_accuracy: 0.9663
Epoch 40/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0481 -
accuracy: 0.9795 - val_loss: 0.1039 - val_accuracy: 0.9618
Epoch 41/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0447 -
accuracy: 0.9809 - val_loss: 0.1169 - val_accuracy: 0.9608
Epoch 42/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0472 -
accuracy: 0.9810 - val_loss: 0.1011 - val_accuracy: 0.9638
Epoch 43/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0468 -
accuracy: 0.9812 - val_loss: 0.1042 - val_accuracy: 0.9658
Epoch 44/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0464 -
accuracy: 0.9785 - val_loss: 0.0977 - val_accuracy: 0.9633
Epoch 45/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0418 -
accuracy: 0.9827 - val_loss: 0.1198 - val_accuracy: 0.9658
Epoch 46/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0527 -
accuracy: 0.9798 - val_loss: 0.1248 - val_accuracy: 0.9608
Epoch 47/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0444 -
accuracy: 0.9805 - val_loss: 0.1167 - val_accuracy: 0.9583
Epoch 48/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0411 -
accuracy: 0.9823 - val_loss: 0.1031 - val_accuracy: 0.9628
Epoch 49/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0444 -
accuracy: 0.9807 - val_loss: 0.1138 - val_accuracy: 0.9603
Epoch 50/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0452 -
```

```
accuracy: 0.9809 - val_loss: 0.1382 - val_accuracy: 0.9598
Epoch 51/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0449 -
accuracy: 0.9807 - val_loss: 0.1041 - val_accuracy: 0.9623
Epoch 52/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0455 -
accuracy: 0.9805 - val_loss: 0.1156 - val_accuracy: 0.9633
Epoch 53/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0446 -
accuracy: 0.9818 - val_loss: 0.1049 - val_accuracy: 0.9633
Epoch 54/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0427 -
accuracy: 0.9812 - val_loss: 0.1170 - val_accuracy: 0.9608
Epoch 55/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0447 -
accuracy: 0.9830 - val_loss: 0.1029 - val_accuracy: 0.9613
Epoch 56/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0452 -
accuracy: 0.9819 - val_loss: 0.1102 - val_accuracy: 0.9663
Epoch 57/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0381 -
accuracy: 0.9849 - val_loss: 0.1041 - val_accuracy: 0.9653
Epoch 58/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0385 -
accuracy: 0.9820 - val_loss: 0.1284 - val_accuracy: 0.9643
Epoch 59/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0383 -
accuracy: 0.9822 - val_loss: 0.1108 - val_accuracy: 0.9673
Epoch 60/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0418 -
accuracy: 0.9833 - val_loss: 0.1369 - val_accuracy: 0.9638
Epoch 61/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0417 -
accuracy: 0.9813 - val_loss: 0.1205 - val_accuracy: 0.9653
Epoch 62/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0396 -
accuracy: 0.9825 - val_loss: 0.1080 - val_accuracy: 0.9638
Epoch 63/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0427 -
accuracy: 0.9823 - val_loss: 0.1277 - val_accuracy: 0.9628
Epoch 64/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0388 -
accuracy: 0.9814 - val_loss: 0.1270 - val_accuracy: 0.9663
Epoch 65/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0395 -
accuracy: 0.9825 - val_loss: 0.1203 - val_accuracy: 0.9668
Epoch 66/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0387 -
```

```
accuracy: 0.9832 - val_loss: 0.1196 - val_accuracy: 0.9693
Epoch 67/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0364 -
accuracy: 0.9847 - val_loss: 0.1024 - val_accuracy: 0.9693
Epoch 68/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0407 -
accuracy: 0.9827 - val_loss: 0.1169 - val_accuracy: 0.9628
Epoch 69/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0431 -
accuracy: 0.9808 - val_loss: 0.1167 - val_accuracy: 0.9638
Epoch 70/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0358 -
accuracy: 0.9847 - val_loss: 0.1103 - val_accuracy: 0.9653
Epoch 71/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0347 -
accuracy: 0.9835 - val_loss: 0.1126 - val_accuracy: 0.9709
Epoch 72/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0340 -
accuracy: 0.9850 - val_loss: 0.1201 - val_accuracy: 0.9633
Epoch 73/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0378 -
accuracy: 0.9832 - val_loss: 0.1060 - val_accuracy: 0.9648
Epoch 74/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0370 -
accuracy: 0.9844 - val_loss: 0.1152 - val_accuracy: 0.9653
Epoch 75/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0394 -
accuracy: 0.9842 - val_loss: 0.1176 - val_accuracy: 0.9653
Epoch 76/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0393 -
accuracy: 0.9817 - val_loss: 0.1027 - val_accuracy: 0.9673
Epoch 77/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0357 -
accuracy: 0.9844 - val_loss: 0.1096 - val_accuracy: 0.9693
Epoch 78/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0377 -
accuracy: 0.9844 - val_loss: 0.1150 - val_accuracy: 0.9668
Epoch 79/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0372 -
accuracy: 0.9833 - val_loss: 0.1130 - val_accuracy: 0.9643
Epoch 80/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0348 -
accuracy: 0.9844 - val_loss: 0.1146 - val_accuracy: 0.9643
Epoch 81/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0362 -
accuracy: 0.9838 - val_loss: 0.1305 - val_accuracy: 0.9653
Epoch 82/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0357 -
```

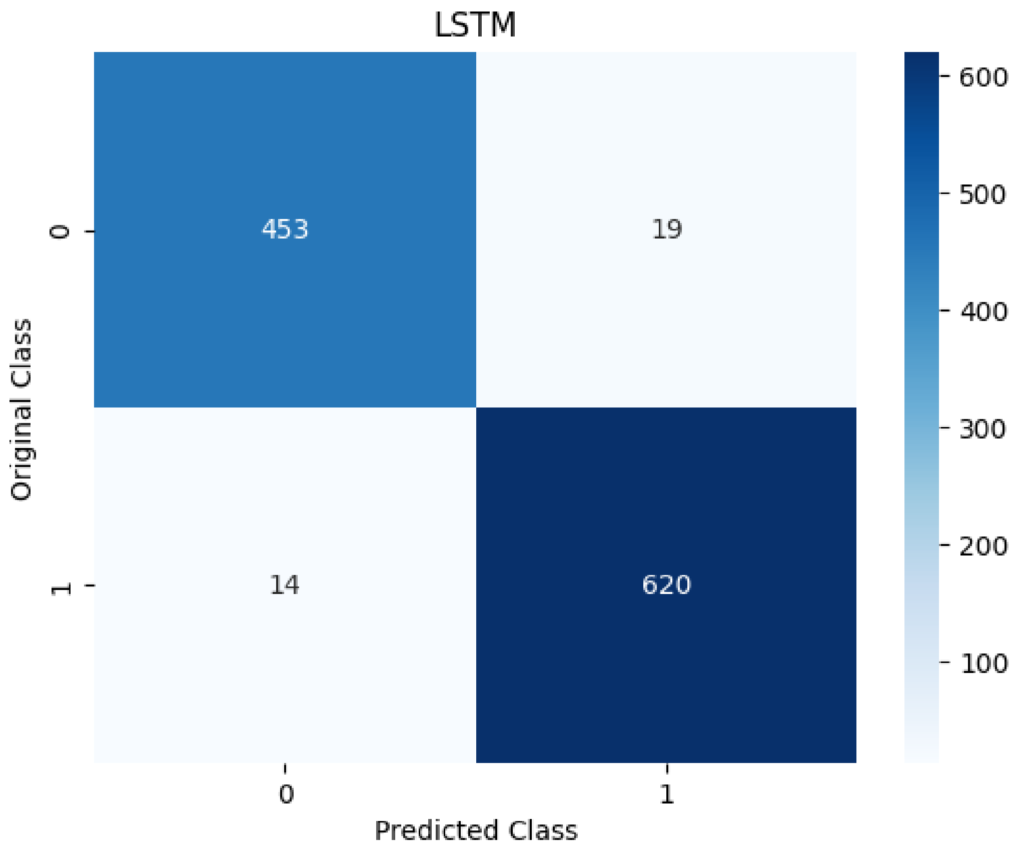
```
accuracy: 0.9845 - val_loss: 0.1094 - val_accuracy: 0.9668
Epoch 83/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0347 -
accuracy: 0.9843 - val_loss: 0.1118 - val_accuracy: 0.9658
Epoch 84/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0357 -
accuracy: 0.9833 - val_loss: 0.1170 - val_accuracy: 0.9653
Epoch 85/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0340 -
accuracy: 0.9840 - val_loss: 0.1167 - val_accuracy: 0.9693
Epoch 86/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0397 -
accuracy: 0.9823 - val_loss: 0.1252 - val_accuracy: 0.9618
Epoch 87/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0377 -
accuracy: 0.9834 - val_loss: 0.1098 - val_accuracy: 0.9719
Epoch 88/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0363 -
accuracy: 0.9854 - val_loss: 0.1176 - val_accuracy: 0.9688
Epoch 89/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0337 -
accuracy: 0.9856 - val_loss: 0.1015 - val_accuracy: 0.9688
Epoch 90/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0374 -
accuracy: 0.9825 - val_loss: 0.1082 - val_accuracy: 0.9658
Epoch 91/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0341 -
accuracy: 0.9840 - val_loss: 0.1126 - val_accuracy: 0.9648
Epoch 92/100
249/249 [==============================] - 1s 5ms/step - loss: 0.0321 -
accuracy: 0.9843 - val_loss: 0.1303 - val_accuracy: 0.9648
Epoch 93/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0323 -
accuracy: 0.9853 - val_loss: 0.1139 - val_accuracy: 0.9683
Epoch 94/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0368 -
accuracy: 0.9838 - val_loss: 0.1612 - val_accuracy: 0.9628
Epoch 95/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0366 -
accuracy: 0.9835 - val_loss: 0.1161 - val_accuracy: 0.9688
Epoch 96/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0353 -
accuracy: 0.9833 - val_loss: 0.1105 - val_accuracy: 0.9683
Epoch 97/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0346 -
accuracy: 0.9840 - val_loss: 0.1376 - val_accuracy: 0.9683
Epoch 98/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0330 -
```

```
accuracy: 0.9843 - val_loss: 0.1121 - val_accuracy: 0.9709
Epoch 99/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0318 -
accuracy: 0.9858 - val_loss: 0.1189 - val_accuracy: 0.9698
Epoch 100/100
249/249 [==============================] - 1s 4ms/step - loss: 0.0320 -
accuracy: 0.9866 - val_loss: 0.1296 - val_accuracy: 0.9678
Test results - Loss: 0.11641994118690491 - Accuracy: 97.01627492904663%
```

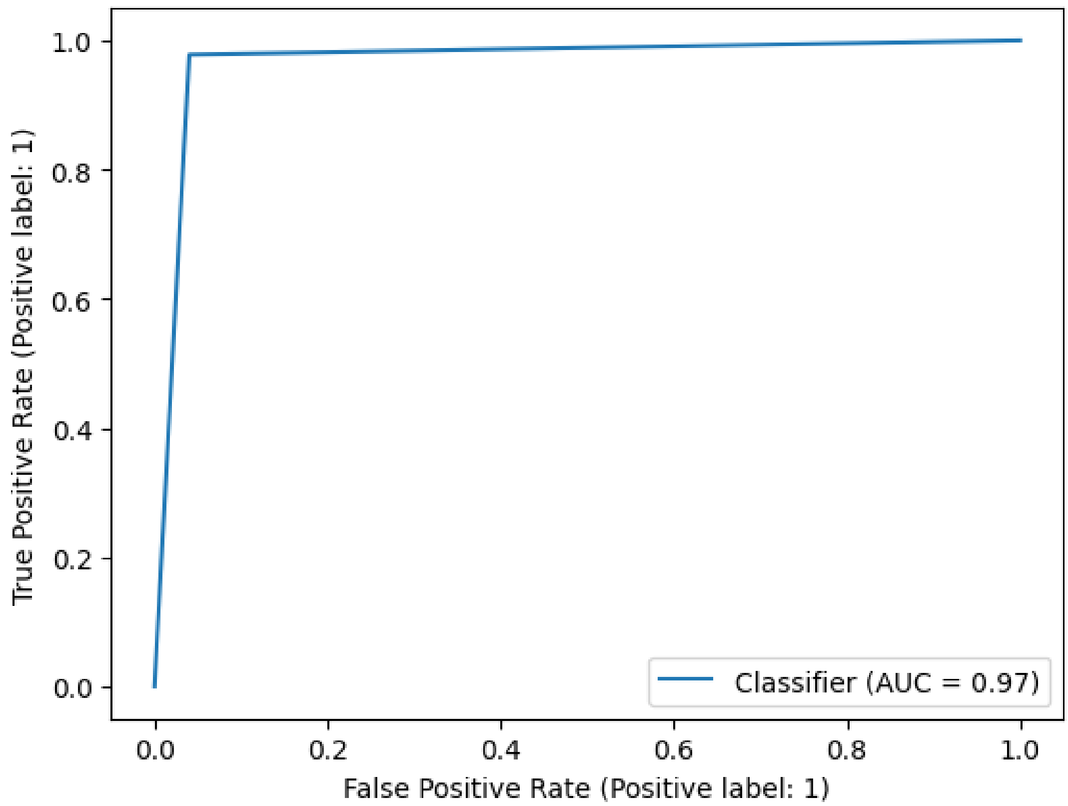
[ ]:
```
lstm_predict_proba = lstm_model.predict(X_test_reshape, batch_size=32)
lstm_predict_class = (lstm_predict_proba > 0.5).astype("int32")
print(classification_report(y_test, lstm_predict_class))
```

```
35/35 [==============================] - 1s 2ms/step
              precision    recall  f1-score   support

           0       0.97      0.96      0.96       472
           1       0.97      0.98      0.97       634

    accuracy                           0.97      1106
   macro avg       0.97      0.97      0.97      1106
weighted avg       0.97      0.97      0.97      1106
```

[ ]:
```
sns.heatmap(confusion_matrix(y_test, lstm_predict_class), annot=True, fmt='g',
  cmap='Blues')
plt.title("LSTM")
plt.xlabel('Predicted Class')
plt.ylabel('Original Class')
plt.show()
```



[ ]:
```
RocCurveDisplay.from_predictions(y_test,lstm_predict_class)
plt.show()
```



[ ]:
```
# print("Trade off between true positive rate and false positive rate")
# from sklearn.metrics import roc_curve
# fpr, tpr, _ = roc_curve(y_test, lstm_predict_class)
# plt.plot(fpr, tpr)
# plt.title('ROC curve')
# plt.xlabel('false positive rate')
# plt.ylabel('true positive rate')
# plt.xlim(0,)
# plt.ylim(0,)
# plt.show()
```

[ ]:
```
# from sklearn.metrics import roc_curve
# fpr, tpr, thresh = roc_curve(y_test, lstm_predict_class)
```

[ ]:
```
# # plot roc curves
# plt.plot(fpr, tpr, linestyle='--',color='orange', label='LSTM')

# # title
# plt.title('ROC curve')
# # x label
# plt.xlabel('False Positive Rate')
```

```
# # y label
# plt.ylabel('True Positive rate')

# plt.legend(loc='best')
# plt.savefig('ROC',dpi=300)
# plt.show()
```

[ ]:
```
# from keras.layers import Flatten
# model = Sequential([
#     Flatten(input_shape=(len(X_test.columns),)),
#     Dense(16, activation=tf.nn.relu),
#         Dense(16, activation=tf.nn.relu),
#     Dense(1, activation=tf.nn.sigmoid),
# ])

# model.compile(optimizer='adam',
#               loss='binary_crossentropy',
#               metrics=['accuracy'])

# model.fit(X_train, y_train, epochs=50, batch_size=1)

# test_loss, test_acc = model.evaluate(X_test, y_test)
# print('Test accuracy:', test_acc)
```

[ ]:
```
# model_pred = model.predict(X_test, batch_size=64)
# model_pred = (model_pred > 0.5).astype(int).reshape(-1,)
# print(classification_report(y_test, model_pred))
```

[ ]:
```
# sns.heatmap(confusion_matrix(y_test, model_pred), annot=True, fmt='g',␣
 ↪cmap='Blues')
# plt.title("Nural network")
# plt.xlabel('Predicted Class')
# plt.ylabel('Original Class')
# plt.show()
```

[ ]:
```
# tensorflow\python\keras\engine\sequential.py:455: UserWarning: model.
 ↪predict_classes() is deprecated and will be removed after 2021-01-01. Please␣
 ↪use instead:* np.argmax(model.predict(x), axis=-1), if your model does␣
 ↪multi-class classification (e.g. if it uses a softmax last-layer activation).
 ↪* (model.predict(x) > 0.5).astype("int32"), if your model does binary␣
 ↪classification (e.g. if it uses a sigmoid last-layer activation).
```