Projeto Final

Grupo:

- Adriano Felipe Cabral Filho (afcf)
- Felipe Rodrigues de Souza (frs3)
- Lucas Vinícius José da Silva (lvjs)
- Pedro José Carneiro de Souza (pjcs)
- Vitor Sousa Silva (vss2)

In [1]:

```
import time
import optuna
import numpy as np
import pandas as pd
from numpy import std
from numpy import mean
import scikitplot as skplt
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings('ignore')

from random import randint
from tensorflow import keras

from sklearn.neural_network import MLPClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import cross_val_score
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from tensorflow.keras.callbacks import EarlyStopping
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score, roc_auc_score, average_precision_score, plot_confusion_matrix

ds = pd.read_csv('TRNcod.xls', delimiter = "\t")

# Shuffle no dataset
ds = ds.sample(frac=1).reset_index(drop=True)

def compute_performance_metrics(y, y_pred_class, y_pred_scores=None):
    accuracy = accuracy_score(y, y_pred_class)
    recall = recall_score(y, y_pred_class)
    precision = precision_score(y, y_pred_class)
    f1 = f1_score(y, y_pred_class)
    performance_metrics = (accuracy, recall, precision, f1)
    if y_pred_scores is not None:
        skplt.metrics.plot_ks_statistic(y, y_pred_scores)
        plt.show()
        y_pred_scores = y_pred_scores[:, 1]
        auroc = roc_auc_score(y, y_pred_scores)
        aupr = average_precision_score(y, y_pred_scores)
        performance_metrics = performance_metrics + (auroc, aupr)
        plt.suptitle('Acurácia: {:3.3f}\nRecall: {:3.3f}\nPrecision: {:3.3f}\nF1: {:3.3f}\nAUROC: {:3.3f}\nAURP: {:3.3f}'.format(accuracy, recall, precision, f1, aur
    return performance_metrics

def compute_performance_metrics_sem_plot(y, y_pred_class, y_pred_scores=None):
    accuracy = accuracy_score(y, y_pred_class)
    recall = recall_score(y, y_pred_class)
    precision = precision_score(y, y_pred_class)
    f1 = f1_score(y, y_pred_class)
    performance_metrics = (accuracy, recall, precision, f1)
    if y_pred_scores is not None:
        # skplt.metrics.plot_ks_statistic(y, y_pred_scores)
        # plt.show()
        y_pred_scores = y_pred_scores[:, 1]
        auroc = roc_auc_score(y, y_pred_scores)
        aupr = average_precision_score(y, y_pred_scores)
        performance_metrics = performance_metrics + (auroc, aupr)
    return performance_metrics

def compute_performance_metrics_sem_plot2(y, y_pred_class, y_pred_scores, rede_trial):
    accuracy = accuracy_score(y, y_pred_class)
    recall = recall_score(y, y_pred_class)
    precision = precision_score(y, y_pred_class)
    f1 = f1_score(y, y_pred_class)
    performance_metrics = (accuracy, recall, precision, f1)
    if y_pred_scores is not None:
        skplt.metrics.plot_ks_statistic(y, y_pred_scores)
        # plt.show()
        y_pred_scores = y_pred_scores[:, 1]
        auroc = roc_auc_score(y, y_pred_scores)
        aupr = average_precision_score(y, y_pred_scores)
        performance_metrics = performance_metrics + (auroc, aupr)
        plt.title(label=rede_trial, y=0.9)
        plt.suptitle('Acurácia: {:3.3f}\nRecall: {:3.3f}\nPrecision: {:3.3f}\nF1: {:3.3f}\nAUROC: {:3.3f}\nAURP: {:3.3f}'.format(accuracy, recall, precision, f1, aur
        plt.savefig(rede_trial, dpi=100)
        plt.close()
    return performance_metrics
```

In [ ]:

```
# print([d for d in ds.columns])
```

Tratamento do Dataset

In [2]:

```
# inadimplentes = pd.DataFrame(list(filter(lambda x: x == 1, ds['IND_BOM_1_2'])))

# Selecionando quem é inadimplente
inadimplentes = ds[ds['IND_BOM_1_2'] == 1]

# Selecionando quem é adimplente
adimplente = ds[ds['IND_BOM_1_2'] == 0]

global treino_ina
global teste_ina
global valid_ina
global treino_adi
global teste_adi
global valid_adi


# Dividindo datasets
treino_ina = inadimplentes[:int(len(inadimplentes)/2)]
teste_ina  = inadimplentes[int(len(inadimplentes)/2):int((len(inadimplentes)*3)/4)]
valid_ina  = inadimplentes[int((len(inadimplentes)*3)/4):]

treino_adi = adimplente[:int(len(adimplente)/2)]
teste_adi  = adimplente[int(len(adimplente)/2):int((len(adimplente)*3)/4)]
valid_adi  = adimplente[int((len(adimplente)*3)/4):]

# Equalizando tamanho de datasets treino e validação dos inadimplentes ao de adimplentes
treino_ina = treino_ina.loc[treino_ina.index.repeat(2)].drop('INDEX', axis=1)
treino_ina["COPIA"] = treino_ina.duplicated()
treino_ina.sort_values(by="COPIA", inplace=True, ignore_index=True)
treino_ina = treino_ina.iloc[ : ( len(treino_adi) - len(treino_ina) ), :  ]
```

```python
treino_ina.drop(columns=["COPIA"], axis=1, inplace=True)

treino_adi.drop(columns=['INDEX'], axis=1, inplace=True)

valid_ina  = valid_ina.loc[valid_ina.index.repeat(2)].drop('INDEX', axis=1)
valid_ina["COPIA"] = valid_ina.duplicated()
valid_ina.sort_values(by="COPIA", inplace=True, ignore_index=True)
valid_ina = valid_ina.iloc[ : ( len(valid_adi) - len(valid_ina) ), :  ]
valid_ina.drop(columns=["COPIA"], axis=1, inplace=True)
# Fim da equalização

print('Tamanhos de inadimplentes: \nTreino: {}\nTeste: {}\nValidação: {}\n'.format(len(treino_ina.values), len(teste_ina.values), len(valid_ina.values)))
print('Tamanhos de adimplentes: \nTreino: {}\nTeste: {}\nValidação: {}\n'.format(len(treino_adi.values), len(teste_adi.values), len(valid_adi.values)))
print('Colunas: {}'.format(len([d for d in ds.columns])))
ds.drop(labels="INDEX", axis=1, inplace=True)
```

```
Tamanhos de inadimplentes:
Treino: 127549
Teste: 33524
Validação: 63775

Tamanhos de adimplentes:
Treino: 127549
Teste: 63774
Validação: 63775

Colunas: 246
```

In [3]:

```python
# https://stackoverflow.com/questions/29294983/how-to-calculate-correlation-between-all-columns-and-remove-highly-correlated-on
# https://psicometriaonline.com.br/como-testar-a-normalidade-da-amostra-com-kolgomorov-smirnov-e-shapiro-wilk/

# # Create correlation matrix
# corr_matrix = ds.corr().abs()

# # Select upper triangle of correlation matrix
# upper = corr_matrix.where(np.triu(np.ones(corr_matrix.shape), k=1).astype(np.bool))

# # Find features with correlation greater than 0.89
# to_drop = [column for column in upper.columns if any(upper[column] > 0.89)]

# # Drop features
# # ds.drop(to_drop, axis=1, inplace=True)
# ds.columns


# Colunas mais correlacionadas
# l = []
# for c in ds.columns[:-1]:
#     l.append({c: ds[c].corr(ds['IND_BOM_1_1'])})
#     # if (ds[c].corr(ds['IND_BOM_1_1']) > 0.0):
#         # print( c, ds[c].corr(ds['IND_BOM_1_1']) )
# l.sort(key=lambda x: list(x.values())[0])
# print(l)
```

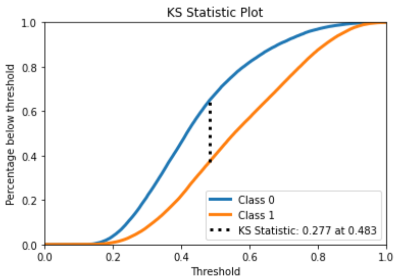
MLP

In [3]:

```python
df_treino = treino_ina.drop(axis=1, labels=['IND_BOM_1_1', 'IND_BOM_1_2']).append(treino_adi.drop(axis=1, labels=['IND_BOM_1_1', 'IND_BOM_1_2']))
df_teste = teste_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(teste_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_teste.drop(columns=['INDEX'], axis=1, inplace=True)
df_validacao = valid_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(valid_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao.drop(columns=['INDEX'], axis=1, inplace=True)

best_mlp = []

def mlp(trial):
    layers              = trial.suggest_int('layers', 1, 2)
    solver              = trial.suggest_categorical('solver', ['lbfgs', 'sgd', 'adam'])
    neurons             = trial.suggest_int('neurons', 2, 32)
    activation          = trial.suggest_categorical('activation', ['identity', 'logistic', 'tanh', 'relu'])
    learning_rate       = trial.suggest_categorical('learning rate', ['constant', 'adaptive'])
    learning_rate_init  = trial.suggest_float('initial learning rate', 0.001, 0.1)
    mlp = MLPClassifier(random_state=1,
                    max_iter=10000,
                    solver=solver,
                    activation=activation,
                    learning_rate=learning_rate,
                    hidden_layer_sizes=(neurons,) if layers==1 else (neurons, neurons),
                    learning_rate_init=learning_rate_init,
                    early_stopping=True).fit(df_treino, [0]*len(treino_ina)+[1]*len(treino_adi))

    best_mlp.append(mlp)

    mlp_pred_class = mlp.predict(df_validacao)
    mlp_pred_scores = mlp.predict_proba(df_validacao)

    accuracy, recall, precision, f1, auroc, aupr = compute_performance_metrics_sem_plot2([0]*len(valid_ina)+[1]*len(valid_adi), mlp_pred_class, mlp_pred_scores, "MLP

    return accuracy

study_0 = optuna.create_study(direction="maximize")
study_0.optimize(mlp, n_trials=64)

compute_performance_metrics([0]*len(teste_ina)+[1]*len(teste_adi), best_mlp[study_0.best_trial.number].predict(df_teste), best_mlp[study_0.best_trial.number].predict
plot_confusion_matrix(best_mlp[study_0.best_trial.number], df_teste, [0]*len(teste_ina)+[1]*len(teste_adi), normalize='true')
```

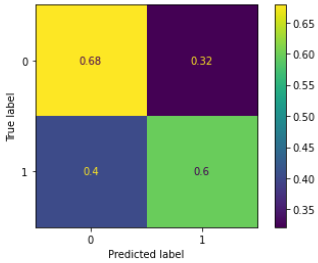
```
[I 2021-12-12 20:40:00,301] A new study created in memory with name: no-name-5eef8c69-9c6f-47d8-a5c4-ecc8d67965ca
[I 2021-12-12 20:40:18,250] Trial 0 finished with value: 0.6264288514308114 and parameters: {'layers': 1, 'solver': 'adam', 'neurons': 18, 'activation': 'logistic',
 'learning rate': 'constant', 'initial learning rate': 0.09109204532809982}. Best is trial 0 with value: 0.6264288514308114.
[I 2021-12-12 20:40:40,852] Trial 1 finished with value: 0.6360721285770287 and parameters: {'layers': 2, 'solver': 'adam', 'neurons': 15, 'activation': 'logistic',
 'learning rate': 'constant', 'initial learning rate': 0.02671932906926546}. Best is trial 1 with value: 0.6360721285770287.
[I 2021-12-12 20:41:46,359] Trial 2 finished with value: 0.6304978439827519 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 12, 'activation': 'identity',
'learning rate': 'adaptive', 'initial learning rate': 0.03621666853560101}. Best is trial 1 with value: 0.6360721285770287.
[I 2021-12-12 20:42:10,535] Trial 3 finished with value: 0.6264602116816934 and parameters: {'layers': 1, 'solver': 'adam', 'neurons': 21, 'activation': 'logistic',
 'learning rate': 'constant', 'initial learning rate': 0.08231492707611938}. Best is trial 1 with value: 0.6360721285770287.
[I 2021-12-12 20:51:22,641] Trial 4 finished with value: 0.6238102704821639 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 13, 'activation': 'logistic',
'learning rate': 'constant', 'initial learning rate': 0.05759457395850182}. Best is trial 1 with value: 0.6360721285770287.
[I 2021-12-12 20:52:55,717] Trial 5 finished with value: 0.6397255978047824 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 14, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.05073428158010421}. Best is trial 5 with value: 0.6397255978047824.
[I 2021-12-12 20:53:13,893] Trial 6 finished with value: 0.6279733437867503 and parameters: {'layers': 2, 'solver': 'adam', 'neurons': 25, 'activation': 'identity',
 'learning rate': 'constant', 'initial learning rate': 0.09725322173179218}. Best is trial 5 with value: 0.6397255978047824.
[I 2021-12-12 21:21:45,168] Trial 7 finished with value: 0.6137201097608781 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 24, 'activation': 'logistic',
'learning rate': 'constant', 'initial learning rate': 0.07493011350038593}. Best is trial 5 with value: 0.6397255978047824.
[I 2021-12-12 21:22:10,530] Trial 8 finished with value: 0.6260446883575068 and parameters: {'layers': 2, 'solver': 'adam', 'neurons': 26, 'activation': 'identity',
 'learning rate': 'adaptive', 'initial learning rate': 0.08143607994406217}. Best is trial 5 with value: 0.6397255978047824.
[I 2021-12-12 21:22:31,764] Trial 9 finished with value: 0.6298784790278322 and parameters: {'layers': 1, 'solver': 'adam', 'neurons': 31, 'activation': 'identity',
 'learning rate': 'adaptive', 'initial learning rate': 0.0029219374686136946}. Best is trial 5 with value: 0.6397255978047824.
[I 2021-12-12 21:23:17,957] Trial 10 finished with value: 0.6386044688357507 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 2, 'activation': 'tanh', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.05601229348856481}. Best is trial 5 with value: 0.6397255978047824.
[I 2021-12-12 21:24:17,334] Trial 11 finished with value: 0.6403684829478636 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 3, 'activation': 'tanh', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.058684739035340376}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:25:00,901] Trial 12 finished with value: 0.6304272834182674 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 2, 'activation': 'relu', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.040036252787325935}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:26:14,714] Trial 13 finished with value: 0.6381732653861231 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 7, 'activation': 'tanh', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.06396536383373262}. Best is trial 11 with value: 0.6403684829478636.
```

```
[I 2021-12-12 21:27:27,430] Trial 14 finished with value: 0.6386907095256762 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 8, 'activation': 'tanh', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.023461433443856032}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:28:23,501] Trial 15 finished with value: 0.6306781654253234 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 6, 'activation': 'relu', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.046411131254829784}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:29:37,028] Trial 16 finished with value: 0.6389651117208938 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 10, 'activation': 'tanh', 'lea
rning rate': 'adaptive', 'initial learning rate': 0.06814065309124875}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:30:30,872] Trial 17 finished with value: 0.639952959623677 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 4, 'activation': 'tanh', 'learn
ing rate': 'adaptive', 'initial learning rate': 0.01783685450605295}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:31:32,469] Trial 18 finished with value: 0.6382203057624461 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 4, 'activation': 'tanh', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.0019502841572633378}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:32:23,578] Trial 19 finished with value: 0.6390591924735398 and parameters: {'layers': 2, 'solver': 'sgd', 'neurons': 2, 'activation': 'tanh', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.014819792680873717}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:33:10,804] Trial 20 finished with value: 0.6373814190513524 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 5, 'activation': 'tanh', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.010432296668493837}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:34:45,892] Trial 21 finished with value: 0.6339161113288906 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 18, 'activation': 'tanh', 'lea
rning rate': 'adaptive', 'initial learning rate': 0.044787335905702504}. Best is trial 11 with value: 0.6403684829478636.
[I 2021-12-12 21:35:52,029] Trial 22 finished with value: 0.6405331242649941 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 10, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.03244903765068676}. Best is trial 22 with value: 0.6405331242649941.
[I 2021-12-12 21:37:02,740] Trial 23 finished with value: 0.6391611132889063 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 10, 'activation': 'tanh', 'lea
rning rate': 'adaptive', 'initial learning rate': 0.028660478797339396}. Best is trial 22 with value: 0.6405331242649941.
[I 2021-12-12 21:38:10,534] Trial 24 finished with value: 0.6386358290866327 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 9, 'activation': 'relu', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.01907698752783516}. Best is trial 22 with value: 0.6405331242649941.
[I 2021-12-12 21:38:59,282] Trial 25 finished with value: 0.639890239121913 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 4, 'activation': 'tanh', 'learn
ing rate': 'adaptive', 'initial learning rate': 0.03232268086860157}. Best is trial 22 with value: 0.6405331242649941.
[I 2021-12-12 21:40:04,229] Trial 26 finished with value: 0.6405644845158761 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 11, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.038643117434869326}. Best is trial 26 with value: 0.6405644845158761.
[I 2021-12-12 21:43:35,466] Trial 27 finished with value: 0.6347785182281458 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 12, 'activation': 'logisti
c', 'learning rate': 'adaptive', 'initial learning rate': 0.039904849524831816}. Best is trial 26 with value: 0.6405644845158761.
[I 2021-12-12 21:44:52,344] Trial 28 finished with value: 0.6391062328498628 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 16, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.05366628490571898}. Best is trial 26 with value: 0.6405644845158761.
[I 2021-12-12 21:46:45,291] Trial 29 finished with value: 0.6375695805566445 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 20, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.06339905039066138}. Best is trial 26 with value: 0.6405644845158761.
[I 2021-12-12 21:47:55,821] Trial 30 finished with value: 0.6395923167385339 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 10, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.04409202754598887}. Best is trial 26 with value: 0.6405644845158761.
[I 2021-12-12 21:49:18,923] Trial 31 finished with value: 0.6398667189337515 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 7, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.021911482787623882}. Best is trial 26 with value: 0.6405644845158761.
[I 2021-12-12 21:50:24,758] Trial 32 finished with value: 0.6409721677773422 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 4, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.009951325676263768}. Best is trial 32 with value: 0.6409721677773422.
[I 2021-12-12 21:52:01,987] Trial 33 finished with value: 0.6404782438259506 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 11, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.008408671743724755}. Best is trial 32 with value: 0.6409721677773422.
[I 2021-12-12 21:54:02,284] Trial 34 finished with value: 0.6400078400627205 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 15, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.007362889987935016}. Best is trial 32 with value: 0.6409721677773422.
[I 2021-12-12 21:57:20,599] Trial 35 finished with value: 0.642077616620933 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 11, 'activation': 'logistic',
'learning rate': 'constant', 'initial learning rate': 0.033070586613238195}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-12 22:01:12,800] Trial 36 finished with value: 0.6339239513916112 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 13, 'activation': 'logisti
c', 'learning rate': 'constant', 'initial learning rate': 0.0319077297544169}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-12 22:39:57,125] Trial 37 finished with value: 0.6151626813014505 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 17, 'activation': 'logisti
c', 'learning rate': 'constant', 'initial learning rate': 0.03450138547533955}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-12 22:43:42,885] Trial 38 finished with value: 0.6339239513916112 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 13, 'activation': 'logisti
c', 'learning rate': 'constant', 'initial learning rate': 0.02704411432849982}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-12 23:54:18,071] Trial 39 finished with value: 0.6213563308506468 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 15, 'activation': 'logisti
c', 'learning rate': 'constant', 'initial learning rate': 0.0399957916779562}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:09:31,791] Trial 40 finished with value: 0.6316268130145041 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 8, 'activation': 'logistic',
'learning rate': 'constant', 'initial learning rate': 0.049273331692431914}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:09:51,746] Trial 41 finished with value: 0.6339474715797726 and parameters: {'layers': 1, 'solver': 'adam', 'neurons': 11, 'activation': 'logistic',
'learning rate': 'constant', 'initial learning rate': 0.010854735220074064}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:13:11,356] Trial 42 finished with value: 0.6347785182281458 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 12, 'activation': 'logisti
c', 'learning rate': 'constant', 'initial learning rate': 0.00519559261136187}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:14:14,175] Trial 43 finished with value: 0.6411995295962367 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 11, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.027330145261761467}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:14:40,638] Trial 44 finished with value: 0.6302391219129753 and parameters: {'layers': 2, 'solver': 'adam', 'neurons': 14, 'activation': 'logistic',
'learning rate': 'constant', 'initial learning rate': 0.03687419172854359}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:15:47,128] Trial 45 finished with value: 0.6289298314386516 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 20, 'activation': 'identity',
 'learning rate': 'adaptive', 'initial learning rate': 0.024662516355434302}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:17:40,359] Trial 46 finished with value: 0.6418502548020384 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 6, 'activation': 'logistic',
'learning rate': 'adaptive', 'initial learning rate': 0.028779074513895504}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:18:34,729] Trial 47 finished with value: 0.6389572716581733 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 5, 'activation': 'logistic',
'learning rate': 'adaptive', 'initial learning rate': 0.014650655662845234}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:20:42,078] Trial 48 finished with value: 0.6377577420619365 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 8, 'activation': 'logistic',
'learning rate': 'adaptive', 'initial learning rate': 0.0275306395503197}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:22:33,489] Trial 49 finished with value: 0.6384476675813406 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 7, 'activation': 'logistic',
'learning rate': 'constant', 'initial learning rate': 0.03687077552966993}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:23:10,877] Trial 50 finished with value: 0.6304586436691494 and parameters: {'layers': 2, 'solver': 'lbfgs', 'neurons': 6, 'activation': 'identity',
'learning rate': 'adaptive', 'initial learning rate': 0.020495614484324305}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:24:20,289] Trial 51 finished with value: 0.6401019208153665 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 9, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.03190980364358179}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:25:30,590] Trial 52 finished with value: 0.6389337514700117 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 11, 'activation': 'logisti
c', 'learning rate': 'adaptive', 'initial learning rate': 0.0400359190111914}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:26:42,131] Trial 53 finished with value: 0.640219521756174 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 9, 'activation': 'logistic', 'l
earning rate': 'adaptive', 'initial learning rate': 0.028037408497633733}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:26:50,444] Trial 54 finished with value: 0.5 and parameters: {'layers': 1, 'solver': 'adam', 'neurons': 6, 'activation': 'relu', 'learning rate': 'a
daptive', 'initial learning rate': 0.015832565019236026}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:28:10,010] Trial 55 finished with value: 0.6410035280282242 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 32, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.048175673918137527}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:29:45,287] Trial 56 finished with value: 0.6388318306546452 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 31, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.04841491164720835}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:31:08,128] Trial 57 finished with value: 0.6387612700901607 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 29, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.05194928908258793}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:42:43,797] Trial 58 finished with value: 0.6227440219521756 and parameters: {'layers': 1, 'solver': 'lbfgs', 'neurons': 23, 'activation': 'logisti
c', 'learning rate': 'adaptive', 'initial learning rate': 0.04466075817322117}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:43:39,819] Trial 59 finished with value: 0.6408702469619757 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 3, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.024194789209265138}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:44:32,639] Trial 60 finished with value: 0.6301293610348883 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 3, 'activation': 'relu', 'lear
ning rate': 'adaptive', 'initial learning rate': 0.023153056684111124}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:45:21,609] Trial 61 finished with value: 0.6391924735397884 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 2, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.04250165681913789}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:46:13,405] Trial 62 finished with value: 0.6406036848294786 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 3, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.03565707925542141}. Best is trial 35 with value: 0.642077616620933.
[I 2021-12-13 00:47:08,798] Trial 63 finished with value: 0.6411995295962367 and parameters: {'layers': 1, 'solver': 'sgd', 'neurons': 3, 'activation': 'logistic',
 'learning rate': 'adaptive', 'initial learning rate': 0.03015863343420548}. Best is trial 35 with value: 0.642077616620933.
```



Out[3]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f7b8deded30>
```

```
<Figure size 432x288 with 0 Axes>
```



Plot da matriz de confusão do melhor classificador dentre os testados acima

Resultados do MLP

In [4]:
```
optuna.visualization.plot_optimization_history(study_0)
```

In [5]:
```
optuna.visualization.plot_slice(study_0)
```

In [6]:
```
optuna.visualization.plot_param_importances(study_0)
```

Comentários acerca dos resultados do MLP

Random Forest

In [ ]:
```
# Cuidado: usar esses parâmetros muito elevados, ou o default (100 estimadores e profundidade ilimitada) vai travar seu computador

# Parâmetros default:
# n_estimators=100, *,
# criterion="gini",
# max_depth=None,
# min_samples_split=2,
# min_samples_leaf=1,
# min_weight_fraction_leaf=0.,
# max_features="auto",
# max_leaf_nodes=None,
# min_impurity_decrease=0.,
# min_impurity_split=None,
# bootstrap=True,
# oob_score=False,
# n_jobs=None,
# random_state=None,
# verbose=0,
# warm_start=False,
# class_weight=None,
# ccp_alpha=0.0,
# max_samples=None
```

Teste de Kolmogorov-Smirnov (KS) e matriz de confusão da Random Forest

Acima, podemos ver os resultados de ambos. O teste de Kolmogorov-Smirnov assemelha-se bastante a uma distribuição normal. A matriz foi bem sucedida para identificar resultados falsos, para casos verdadeiro seu resultado foi pouco acima de 50%.

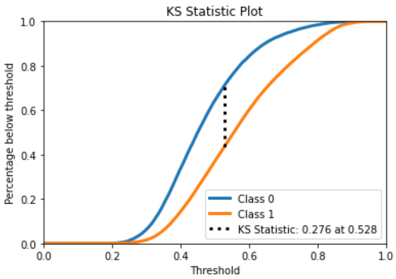
In [3]:
```
ds_dropado = ds.drop(labels=['IND_BOM_1_2', 'IND_BOM_1_1'], axis=1)
df_treino = treino_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(treino_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao = valid_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(valid_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao.drop(columns=['INDEX'], axis=1, inplace=True)

best_rf = []

def rf(trial):
    max_depth           = trial.suggest_int("max_depth", 16, 64)
    criterion           = trial.suggest_categorical("criterion", ["gini", "entropy"])
    n_estimators        = trial.suggest_int("n_estimators", 8, 32)
    min_samples_leaf    = trial.suggest_int("min_samples_leaf", 2, 8)
    min_samples_split   = trial.suggest_int("min_samples_split", 2, 8)
    random_forest = RandomForestClassifier(n_estimators=n_estimators,
                                    max_depth=max_depth,
                                    criterion=criterion,
                                    bootstrap=True,
                                    min_samples_leaf=min_samples_leaf,
                                    min_samples_split=min_samples_split).fit(df_treino, [0]*len(treino_ina)+[1]*len(treino_adi))

    # Código opcional para logar e cronometrar tempo
    # print('Random Forest com {} estimadores e profundidade máxima {}, critério {}, min amostras de folhas {} e min divisão de amostras {}\n'.format(
    #       n_estimators, max_depth, criterion, min_samples_leaf, min_samples_split
    # ))
    # start = time.time()
    # n_scores = cross_val_score(random_forest, ds_dropado, ds['IND_BOM_1_2'], scoring='accuracy', n_jobs=1, error_score='raise')
    # print('Acurácia e desvio padrão: %.3f (%.3f)' % (mean(n_scores), std(n_scores)))
    # print("Tempo: ", format(time.time() - start, '.3f'), 's\n', sep='')

    best_rf.append(random_forest)

    rf_pred_class  = random_forest.predict(df_validacao)
    rf_pred_scores = random_forest.predict_proba(df_validacao)

    accuracy, recall, precision, f1, auroc, aupr = compute_performance_metrics_sem_plot2([0]*len(valid_ina)+[1]*len(valid_adi), rf_pred_class, rf_pred_scores, "RF/RF

    return accuracy

study_1 = optuna.create_study(direction="maximize")
study_1.optimize(rf, n_trials=64)
```

```
[I 2021-12-13 09:53:46,701] A new study created in memory with name: no-name-645ee9d7-04f2-49f6-b887-c3745391ff0c
[I 2021-12-13 09:54:27,698] Trial 0 finished with value: 0.6122540180321443 and parameters: {'max_depth': 62, 'criterion': 'gini', 'n_estimators': 20, 'min_samples_l
eaf': 4, 'min_samples_split': 8}. Best is trial 0 with value: 0.6122540180321443.
[I 2021-12-13 09:54:47,460] Trial 1 finished with value: 0.6049392395139161 and parameters: {'max_depth': 37, 'criterion': 'gini', 'n_estimators': 10, 'min_samples_l
eaf': 5, 'min_samples_split': 6}. Best is trial 0 with value: 0.6122540180321443.
[I 2021-12-13 09:55:08,387] Trial 2 finished with value: 0.5943237945903567 and parameters: {'max_depth': 55, 'criterion': 'entropy', 'n_estimators': 9, 'min_samples
_leaf': 3, 'min_samples_split': 2}. Best is trial 0 with value: 0.6122540180321443.
[I 2021-12-13 09:55:55,276] Trial 3 finished with value: 0.6175460603684829 and parameters: {'max_depth': 57, 'criterion': 'gini', 'n_estimators': 24, 'min_samples_l
eaf': 5, 'min_samples_split': 2}. Best is trial 3 with value: 0.6175460603684829.
[I 2021-12-13 09:56:17,979] Trial 4 finished with value: 0.6098706389651117 and parameters: {'max_depth': 57, 'criterion': 'gini', 'n_estimators': 11, 'min_samples_l
eaf': 7, 'min_samples_split': 7}. Best is trial 3 with value: 0.6175460603684829.
[I 2021-12-13 09:56:42,197] Trial 5 finished with value: 0.6196550372402979 and parameters: {'max_depth': 25, 'criterion': 'gini', 'n_estimators': 13, 'min_samples_l
eaf': 8, 'min_samples_split': 3}. Best is trial 5 with value: 0.6196550372402979.
[I 2021-12-13 09:57:13,298] Trial 6 finished with value: 0.621858094864759 and parameters: {'max_depth': 22, 'criterion': 'entropy', 'n_estimators': 15, 'min_samples
_leaf': 3, 'min_samples_split': 4}. Best is trial 6 with value: 0.621858094864759.
```

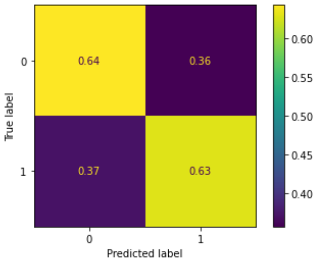
```
[I 2021-12-13 09:57:32,841] Trial 7 finished with value: 0.6154292434339474 and parameters: {'max_depth': 24, 'criterion': 'gini', 'n_estimators': 11, 'min_samples_l
eaf': 6, 'min_samples_split': 7}. Best is trial 6 with value: 0.621858094864759.
[I 2021-12-13 09:58:26,212] Trial 8 finished with value: 0.6281615052920423 and parameters: {'max_depth': 25, 'criterion': 'entropy', 'n_estimators': 27, 'min_sample
s_leaf': 8, 'min_samples_split': 3}. Best is trial 8 with value: 0.6281615052920423.
[I 2021-12-13 09:58:55,945] Trial 9 finished with value: 0.6015209721677773 and parameters: {'max_depth': 61, 'criterion': 'entropy', 'n_estimators': 13, 'min_sample
s_leaf': 3, 'min_samples_split': 7}. Best is trial 8 with value: 0.6281615052920423.
[I 2021-12-13 10:00:00,991] Trial 10 finished with value: 0.6258408467267739 and parameters: {'max_depth': 34, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 8, 'min_samples_split': 5}. Best is trial 8 with value: 0.6281615052920423.
[I 2021-12-13 10:01:05,384] Trial 11 finished with value: 0.6252685221481772 and parameters: {'max_depth': 35, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 8, 'min_samples_split': 5}. Best is trial 8 with value: 0.6281615052920423.
[I 2021-12-13 10:02:12,021] Trial 12 finished with value: 0.62203057624461 and parameters: {'max_depth': 46, 'criterion': 'entropy', 'n_estimators': 32, 'min_samples
_leaf': 7, 'min_samples_split': 4}. Best is trial 8 with value: 0.6281615052920423.
[I 2021-12-13 10:02:54,762] Trial 13 finished with value: 0.6353978831830654 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 27, 'min_sampl
es_leaf': 8, 'min_samples_split': 5}. Best is trial 13 with value: 0.6353978831830654.
[I 2021-12-13 10:03:40,725] Trial 14 finished with value: 0.6359702077616621 and parameters: {'max_depth': 17, 'criterion': 'entropy', 'n_estimators': 27, 'min_sampl
es_leaf': 7, 'min_samples_split': 3}. Best is trial 14 with value: 0.6359702077616621.
[I 2021-12-13 10:04:20,609] Trial 15 finished with value: 0.6345041160329282 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 26, 'min_sampl
es_leaf': 7, 'min_samples_split': 4}. Best is trial 14 with value: 0.6359702077616621.
[I 2021-12-13 10:04:55,441] Trial 16 finished with value: 0.6341826734613877 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 22, 'min_sampl
es_leaf': 6, 'min_samples_split': 3}. Best is trial 14 with value: 0.6359702077616621.
[I 2021-12-13 10:05:42,165] Trial 17 finished with value: 0.6375068600548804 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:06:39,220] Trial 18 finished with value: 0.6207840062720502 and parameters: {'max_depth': 31, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:07:14,116] Trial 19 finished with value: 0.6116346530772246 and parameters: {'max_depth': 43, 'criterion': 'entropy', 'n_estimators': 18, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:08:05,138] Trial 20 finished with value: 0.6302548020384163 and parameters: {'max_depth': 20, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 4, 'min_samples_split': 8}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:08:48,371] Trial 21 finished with value: 0.635280282242258 and parameters: {'max_depth': 17, 'criterion': 'entropy', 'n_estimators': 27, 'min_sample
s_leaf': 7, 'min_samples_split': 5}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:09:33,867] Trial 22 finished with value: 0.622046256370051 and parameters: {'max_depth': 29, 'criterion': 'entropy', 'n_estimators': 24, 'min_sample
s_leaf': 7, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:10:30,295] Trial 23 finished with value: 0.6310544884359075 and parameters: {'max_depth': 20, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 6, 'min_samples_split': 4}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:11:23,776] Trial 24 finished with value: 0.6251195609564877 and parameters: {'max_depth': 29, 'criterion': 'entropy', 'n_estimators': 25, 'min_sampl
es_leaf': 8, 'min_samples_split': 5}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:12:12,854] Trial 25 finished with value: 0.6365817326538612 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 28, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:12:55,559] Trial 26 finished with value: 0.6323010584084673 and parameters: {'max_depth': 19, 'criterion': 'entropy', 'n_estimators': 22, 'min_sampl
es_leaf': 5, 'min_samples_split': 7}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:13:54,519] Trial 27 finished with value: 0.6251430811446491 and parameters: {'max_depth': 27, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:14:36,783] Trial 28 finished with value: 0.6289925519404155 and parameters: {'max_depth': 21, 'criterion': 'entropy', 'n_estimators': 23, 'min_sampl
es_leaf': 6, 'min_samples_split': 7}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:15:06,389] Trial 29 finished with value: 0.6340415523324187 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 19, 'min_sampl
es_leaf': 4, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:15:48,125] Trial 30 finished with value: 0.6263347706781655 and parameters: {'max_depth': 23, 'criterion': 'entropy', 'n_estimators': 21, 'min_sampl
es_leaf': 7, 'min_samples_split': 8}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:16:31,880] Trial 31 finished with value: 0.6349274794198354 and parameters: {'max_depth': 17, 'criterion': 'entropy', 'n_estimators': 27, 'min_sampl
es_leaf': 8, 'min_samples_split': 5}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:17:20,841] Trial 32 finished with value: 0.6356017248137985 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 28, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:18:19,378] Trial 33 finished with value: 0.6347706781654253 and parameters: {'max_depth': 18, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 5, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:19:23,979] Trial 34 finished with value: 0.6195923167385339 and parameters: {'max_depth': 48, 'criterion': 'entropy', 'n_estimators': 28, 'min_sampl
es_leaf': 6, 'min_samples_split': 7}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:20:09,688] Trial 35 finished with value: 0.6296432771462172 and parameters: {'max_depth': 20, 'criterion': 'gini', 'n_estimators': 25, 'min_samples_
leaf': 7, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:21:15,639] Trial 36 finished with value: 0.6259427675421403 and parameters: {'max_depth': 26, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 7, 'min_samples_split': 2}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:22:01,328] Trial 37 finished with value: 0.6262877303018424 and parameters: {'max_depth': 22, 'criterion': 'gini', 'n_estimators': 30, 'min_samples_
leaf': 5, 'min_samples_split': 7}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:22:37,810] Trial 38 finished with value: 0.6375068600548804 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 25, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:23:18,640] Trial 39 finished with value: 0.6210348882791062 and parameters: {'max_depth': 38, 'criterion': 'gini', 'n_estimators': 25, 'min_samples_
leaf': 6, 'min_samples_split': 3}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:24:04,972] Trial 40 finished with value: 0.6144021952175618 and parameters: {'max_depth': 52, 'criterion': 'entropy', 'n_estimators': 24, 'min_sampl
es_leaf': 5, 'min_samples_split': 8}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:24:48,979] Trial 41 finished with value: 0.6371854174833399 and parameters: {'max_depth': 18, 'criterion': 'entropy', 'n_estimators': 28, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:25:35,580] Trial 42 finished with value: 0.6276127009016073 and parameters: {'max_depth': 23, 'criterion': 'entropy', 'n_estimators': 26, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:26:21,413] Trial 43 finished with value: 0.6293923951391611 and parameters: {'max_depth': 19, 'criterion': 'entropy', 'n_estimators': 28, 'min_sampl
es_leaf': 2, 'min_samples_split': 7}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:27:07,457] Trial 44 finished with value: 0.6343551548412387 and parameters: {'max_depth': 18, 'criterion': 'entropy', 'n_estimators': 26, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:28:13,325] Trial 45 finished with value: 0.6287416699333594 and parameters: {'max_depth': 24, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 5, 'min_samples_split': 5}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:29:13,042] Trial 46 finished with value: 0.619921599372795 and parameters: {'max_depth': 63, 'criterion': 'entropy', 'n_estimators': 28, 'min_sample
s_leaf': 6, 'min_samples_split': 7}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:30:18,779] Trial 47 finished with value: 0.6325519404155233 and parameters: {'max_depth': 22, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 7, 'min_samples_split': 2}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:31:26,828] Trial 48 finished with value: 0.6248451587612701 and parameters: {'max_depth': 28, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:32:00,787] Trial 49 finished with value: 0.6139004312034496 and parameters: {'max_depth': 31, 'criterion': 'gini', 'n_estimators': 17, 'min_samples_
leaf': 5, 'min_samples_split': 4}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:32:45,580] Trial 50 finished with value: 0.6224852998823991 and parameters: {'max_depth': 25, 'criterion': 'entropy', 'n_estimators': 23, 'min_sampl
es_leaf': 6, 'min_samples_split': 5}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:33:26,897] Trial 51 finished with value: 0.6372951783614269 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 28, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 17 with value: 0.6375068600548804.
[I 2021-12-13 10:34:11,745] Trial 52 finished with value: 0.6378518228145825 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 52 with value: 0.6378518228145825.
[I 2021-12-13 10:34:57,130] Trial 53 finished with value: 0.6372559780478244 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 52 with value: 0.6378518228145825.
[I 2021-12-13 10:35:47,459] Trial 54 finished with value: 0.6356958055664446 and parameters: {'max_depth': 19, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 8, 'min_samples_split': 6}. Best is trial 52 with value: 0.6378518228145825.
[I 2021-12-13 10:36:43,169] Trial 55 finished with value: 0.6318149745197962 and parameters: {'max_depth': 21, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 7, 'min_samples_split': 7}. Best is trial 52 with value: 0.6378518228145825.
[I 2021-12-13 10:37:35,757] Trial 56 finished with value: 0.6356566052528421 and parameters: {'max_depth': 18, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 8, 'min_samples_split': 5}. Best is trial 52 with value: 0.6378518228145825.
[I 2021-12-13 10:38:28,012] Trial 57 finished with value: 0.6381967855742846 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 6, 'min_samples_split': 6}. Best is trial 57 with value: 0.6381967855742846.
[I 2021-12-13 10:39:22,348] Trial 58 finished with value: 0.6395531164249314 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 58 with value: 0.6395531164249314.
[I 2021-12-13 10:40:27,236] Trial 59 finished with value: 0.6305292042336339 and parameters: {'max_depth': 21, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 6, 'min_samples_split': 5}. Best is trial 58 with value: 0.6395531164249314.
[I 2021-12-13 10:40:41,075] Trial 60 finished with value: 0.6258173265386123 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 8, 'min_sample
s_leaf': 7, 'min_samples_split': 7}. Best is trial 58 with value: 0.6395531164249314.
[I 2021-12-13 10:41:29,399] Trial 61 finished with value: 0.6369188553508428 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 7, 'min_samples_split': 6}. Best is trial 58 with value: 0.6395531164249314.
[I 2021-12-13 10:42:20,026] Trial 62 finished with value: 0.6380086240689925 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 8, 'min_samples_split': 6}. Best is trial 58 with value: 0.6395531164249314.
[I 2021-12-13 10:43:17,386] Trial 63 finished with value: 0.6343551548412387 and parameters: {'max_depth': 19, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 8, 'min_samples_split': 6}. Best is trial 58 with value: 0.6395531164249314.

---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-3-23bc67534477> in <module>
     39 study_1 = optuna.create_study(direction="maximize")
     40 study_1.optimize(rf, n_trials=64)
---> 41 compute_performance_metrics([0]*len(teste_ina)+[1]*len(teste_adi), best_rf[study_1.best_trial.number].predict(df_teste), best_rf[study_1.best_trial.number].p
redict_proba(df_teste))
     42 plot_confusion_matrix(best_rf[study_1.best_trial.number], df_teste, [0]*len(teste_ina)+[1]*len(teste_adi), normalize='true')

NameError: name 'df_teste' is not defined
```

Plot da matriz de confusão do melhor classificador dentre os testados acima

In [4]:
```
df_teste = teste_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(teste_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_teste.drop(columns=['INDEX'], axis=1, inplace=True)
compute_performance_metrics([0]*len(teste_ina)+[1]*len(teste_adi), best_rf[study_1.best_trial.number].predict(df_teste), best_rf[study_1.best_trial.number].predict_p
plot_confusion_matrix(best_rf[study_1.best_trial.number], df_teste, [0]*len(teste_ina)+[1]*len(teste_adi), normalize='true')
```



Out[4]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f0a35b8b1c0>

<Figure size 432x288 with 0 Axes>
```



Resultados da Random Forest

In [5]:
```
optuna.visualization.plot_optimization_history(study_1)
```

In [6]:
```
optuna.visualization.plot_slice(study_1)
```

In [7]:
```
optuna.visualization.plot_param_importances(study_1)
```

Comentários acerca dos resultados da Random Forest

A profundidade máxima é, sem dúvidas, o maior diferencial entre todos os hiperparâmetros, e menores profundidades são melhores; é possível perceber através do gráfico que quanto maior a profundidade melhor vai ser o desempenho, até atingir-se um limite. A quantidade de estimadores tem um impacto bastante menor, no gráfico, é possível ver que, excluindo-se os outliers, o número de estimadores realmente não interfere bastante no resultado final. A quantidade mínima de samples nas folhas tem impacto ainda menor, pode-se dizer que é inútil alterar este parâmetro.

Teste com profundidade menor

In [8]:
```
def rf(trial):
    max_depth           = trial.suggest_int("max_depth", 2, 16)
    criterion           = trial.suggest_categorical("criterion", ["gini", "entropy"])
    n_estimators        = trial.suggest_int("n_estimators", 8, 32)
    min_samples_leaf    = trial.suggest_int("min_samples_leaf", 2, 8)
    min_samples_split   = trial.suggest_int("min_samples_split", 2, 8)
    random_forest = RandomForestClassifier(n_estimators=n_estimators,
                                    max_depth=max_depth,
                                    criterion=criterion,
                                    bootstrap=True,
                                    min_samples_leaf=min_samples_leaf,
                                    min_samples_split=min_samples_split).fit(df_treino, [0]*len(treino_ina)+[1]*len(treino_adi))

    rf_pred_class  = random_forest.predict(df_validacao)
    rf_pred_scores = random_forest.predict_proba(df_validacao)

    accuracy, recall, precision, f1, auroc, aupr = compute_performance_metrics_sem_plot([0]*len(valid_ina)+[1]*len(valid_adi), rf_pred_class, rf_pred_scores)

    return accuracy

study_1 = optuna.create_study(direction="maximize")
study_1.optimize(rf, n_trials=64)
```

```
[I 2021-12-13 10:51:57,381] A new study created in memory with name: no-name-f01eceaf-879c-435b-b6ec-274904bcbc97
[I 2021-12-13 10:52:39,920] Trial 0 finished with value: 0.6385966287730301 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 26, 'min_sample
s_leaf': 7, 'min_samples_split': 3}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:52:55,472] Trial 1 finished with value: 0.624076832614661 and parameters: {'max_depth': 6, 'criterion': 'entropy', 'n_estimators': 20, 'min_samples_
leaf': 8, 'min_samples_split': 6}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:53:07,963] Trial 2 finished with value: 0.626765974127793 and parameters: {'max_depth': 7, 'criterion': 'entropy', 'n_estimators': 14, 'min_samples_
leaf': 4, 'min_samples_split': 7}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:53:30,821] Trial 3 finished with value: 0.6291807134457076 and parameters: {'max_depth': 7, 'criterion': 'entropy', 'n_estimators': 28, 'min_samples
_leaf': 6, 'min_samples_split': 8}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:53:33,727] Trial 4 finished with value: 0.5870717365738926 and parameters: {'max_depth': 2, 'criterion': 'entropy', 'n_estimators': 9, 'min_samples_
leaf': 5, 'min_samples_split': 4}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:53:43,658] Trial 5 finished with value: 0.619890239121913 and parameters: {'max_depth': 5, 'criterion': 'gini', 'n_estimators': 19, 'min_samples_lea
f': 3, 'min_samples_split': 3}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:54:03,966] Trial 6 finished with value: 0.6362524500196002 and parameters: {'max_depth': 11, 'criterion': 'gini', 'n_estimators': 19, 'min_samples_l
eaf': 4, 'min_samples_split': 6}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:54:15,594] Trial 7 finished with value: 0.6327165817326539 and parameters: {'max_depth': 11, 'criterion': 'entropy', 'n_estimators': 10, 'min_sample
s_leaf': 5, 'min_samples_split': 7}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:54:20,507] Trial 8 finished with value: 0.5952881223049784 and parameters: {'max_depth': 2, 'criterion': 'gini', 'n_estimators': 21, 'min_samples_le
af': 8, 'min_samples_split': 6}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:54:44,117] Trial 9 finished with value: 0.6287259898079185 and parameters: {'max_depth': 16, 'criterion': 'gini', 'n_estimators': 16, 'min_samples_l
eaf': 2, 'min_samples_split': 4}. Best is trial 0 with value: 0.6385966287730301.
[I 2021-12-13 10:55:32,696] Trial 10 finished with value: 0.6388318306546452 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 7, 'min_samples_split': 2}. Best is trial 10 with value: 0.6388318306546452.
[I 2021-12-13 10:56:24,752] Trial 11 finished with value: 0.6388631909055272 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 7, 'min_samples_split': 2}. Best is trial 11 with value: 0.6388631909055272.
[I 2021-12-13 10:57:14,551] Trial 12 finished with value: 0.6370364562916503 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 7, 'min_samples_split': 2}. Best is trial 11 with value: 0.6388631909055272.
[I 2021-12-13 10:57:59,903] Trial 13 finished with value: 0.6392551940415523 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 7, 'min_samples_split': 2}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 10:58:33,113] Trial 14 finished with value: 0.6382987063896511 and parameters: {'max_depth': 13, 'criterion': 'entropy', 'n_estimators': 25, 'min_sampl
es_leaf': 6, 'min_samples_split': 2}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 10:59:19,846] Trial 15 finished with value: 0.6379772638181106 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 8, 'min_samples_split': 3}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 10:59:49,959] Trial 16 finished with value: 0.6384790278322227 and parameters: {'max_depth': 12, 'criterion': 'entropy', 'n_estimators': 24, 'min_sampl
es_leaf': 6, 'min_samples_split': 4}. Best is trial 13 with value: 0.6392551940415523.
```

```
[I 2021-12-13 11:00:17,885] Trial 17 finished with value: 0.6363857310858487 and parameters: {'max_depth': 9, 'criterion': 'entropy', 'n_estimators': 29, 'min_sample
s_leaf': 7, 'min_samples_split': 2}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:01:07,539] Trial 18 finished with value: 0.6390827126617012 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 6, 'min_samples_split': 3}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:01:47,663] Trial 19 finished with value: 0.6385182281458252 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 27, 'min_sampl
es_leaf': 6, 'min_samples_split': 5}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:02:17,313] Trial 20 finished with value: 0.6374206193649549 and parameters: {'max_depth': 10, 'criterion': 'gini', 'n_estimators': 30, 'min_samples_
leaf': 4, 'min_samples_split': 3}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:03:09,793] Trial 21 finished with value: 0.6387142297138377 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 7, 'min_samples_split': 2}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:04:02,376] Trial 22 finished with value: 0.6373814190513524 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 6, 'min_samples_split': 3}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:04:34,656] Trial 23 finished with value: 0.6388867110936888 and parameters: {'max_depth': 13, 'criterion': 'entropy', 'n_estimators': 23, 'min_sampl
es_leaf': 8, 'min_samples_split': 2}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:05:06,942] Trial 24 finished with value: 0.6366601332810663 and parameters: {'max_depth': 13, 'criterion': 'entropy', 'n_estimators': 24, 'min_sampl
es_leaf': 8, 'min_samples_split': 4}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:05:34,747] Trial 25 finished with value: 0.637765582124657 and parameters: {'max_depth': 12, 'criterion': 'entropy', 'n_estimators': 22, 'min_sample
s_leaf': 8, 'min_samples_split': 3}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:05:57,087] Trial 26 finished with value: 0.6376636613092904 and parameters: {'max_depth': 13, 'criterion': 'entropy', 'n_estimators': 16, 'min_sampl
es_leaf': 5, 'min_samples_split': 2}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:06:17,250] Trial 27 finished with value: 0.6322304978439828 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 12, 'min_sampl
es_leaf': 6, 'min_samples_split': 5}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:07:02,586] Trial 28 finished with value: 0.6391297530380243 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 7, 'min_samples_split': 3}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:07:43,808] Trial 29 finished with value: 0.6383535868286946 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 27, 'min_sampl
es_leaf': 7, 'min_samples_split': 3}. Best is trial 13 with value: 0.6392551940415523.
[I 2021-12-13 11:08:27,730] Trial 30 finished with value: 0.6394825558604469 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 7, 'min_samples_split': 4}. Best is trial 30 with value: 0.6394825558604469.
[I 2021-12-13 11:09:10,202] Trial 31 finished with value: 0.6385966287730301 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 7, 'min_samples_split': 4}. Best is trial 30 with value: 0.6394825558604469.
[I 2021-12-13 11:09:52,904] Trial 32 finished with value: 0.6395060760486084 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 7, 'min_samples_split': 3}. Best is trial 32 with value: 0.6395060760486084.
[I 2021-12-13 11:10:24,566] Trial 33 finished with value: 0.6377577420619365 and parameters: {'max_depth': 12, 'criterion': 'entropy', 'n_estimators': 26, 'min_sampl
es_leaf': 7, 'min_samples_split': 4}. Best is trial 32 with value: 0.6395060760486084.
[I 2021-12-13 11:11:06,749] Trial 34 finished with value: 0.6369266954135633 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 8, 'min_samples_split': 5}. Best is trial 32 with value: 0.6395060760486084.
[I 2021-12-13 11:11:47,956] Trial 35 finished with value: 0.6369972559780478 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 28, 'min_sampl
es_leaf': 7, 'min_samples_split': 4}. Best is trial 32 with value: 0.6395060760486084.
[I 2021-12-13 11:12:22,140] Trial 36 finished with value: 0.6362054096432771 and parameters: {'max_depth': 11, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 8, 'min_samples_split': 3}. Best is trial 32 with value: 0.6395060760486084.
[I 2021-12-13 11:12:37,271] Trial 37 finished with value: 0.6194590356722853 and parameters: {'max_depth': 5, 'criterion': 'entropy', 'n_estimators': 28, 'min_sample
s_leaf': 5, 'min_samples_split': 3}. Best is trial 32 with value: 0.6395060760486084.
[I 2021-12-13 11:12:59,036] Trial 38 finished with value: 0.6326617012936103 and parameters: {'max_depth': 8, 'criterion': 'entropy', 'n_estimators': 26, 'min_sample
s_leaf': 7, 'min_samples_split': 4}. Best is trial 32 with value: 0.6395060760486084.
[I 2021-12-13 11:13:39,947] Trial 39 finished with value: 0.639937279498236 and parameters: {'max_depth': 14, 'criterion': 'gini', 'n_estimators': 31, 'min_samples_l
eaf': 6, 'min_samples_split': 5}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:14:11,577] Trial 40 finished with value: 0.6372246177969424 and parameters: {'max_depth': 10, 'criterion': 'gini', 'n_estimators': 31, 'min_samples_
leaf': 6, 'min_samples_split': 5}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:14:50,899] Trial 41 finished with value: 0.6387377499019992 and parameters: {'max_depth': 14, 'criterion': 'gini', 'n_estimators': 29, 'min_samples_
leaf': 7, 'min_samples_split': 5}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:15:32,659] Trial 42 finished with value: 0.6363622108976872 and parameters: {'max_depth': 15, 'criterion': 'gini', 'n_estimators': 31, 'min_samples_
leaf': 6, 'min_samples_split': 6}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:16:10,575] Trial 43 finished with value: 0.638000784006272 and parameters: {'max_depth': 13, 'criterion': 'gini', 'n_estimators': 31, 'min_samples_l
eaf': 7, 'min_samples_split': 6}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:16:52,128] Trial 44 finished with value: 0.6355546844374755 and parameters: {'max_depth': 16, 'criterion': 'gini', 'n_estimators': 27, 'min_samples_
leaf': 6, 'min_samples_split': 4}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:17:27,710] Trial 45 finished with value: 0.6395374362994904 and parameters: {'max_depth': 12, 'criterion': 'gini', 'n_estimators': 29, 'min_samples_
leaf': 5, 'min_samples_split': 4}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:18:01,366] Trial 46 finished with value: 0.6373030184241474 and parameters: {'max_depth': 12, 'criterion': 'gini', 'n_estimators': 28, 'min_samples_
leaf': 4, 'min_samples_split': 7}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:18:28,643] Trial 47 finished with value: 0.6382987063896511 and parameters: {'max_depth': 11, 'criterion': 'gini', 'n_estimators': 25, 'min_samples_
leaf': 5, 'min_samples_split': 5}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:19:09,123] Trial 48 finished with value: 0.6384555076440611 and parameters: {'max_depth': 14, 'criterion': 'gini', 'n_estimators': 31, 'min_samples_
leaf': 5, 'min_samples_split': 4}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:19:43,378] Trial 49 finished with value: 0.637577420619365 and parameters: {'max_depth': 13, 'criterion': 'gini', 'n_estimators': 29, 'min_samples_l
eaf': 4, 'min_samples_split': 4}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:20:05,046] Trial 50 finished with value: 0.6363151705213642 and parameters: {'max_depth': 12, 'criterion': 'gini', 'n_estimators': 19, 'min_samples_
leaf': 6, 'min_samples_split': 6}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:20:49,425] Trial 51 finished with value: 0.6386907095256762 and parameters: {'max_depth': 15, 'criterion': 'gini', 'n_estimators': 31, 'min_samples_
leaf': 7, 'min_samples_split': 3}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:21:37,440] Trial 52 finished with value: 0.6382595060760486 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 6, 'min_samples_split': 3}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:22:19,715] Trial 53 finished with value: 0.63546844374755 and parameters: {'max_depth': 16, 'criterion': 'gini', 'n_estimators': 28, 'min_samples_le
af': 3, 'min_samples_split': 4}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:23:04,333] Trial 54 finished with value: 0.6387063896511173 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 29, 'min_sampl
es_leaf': 7, 'min_samples_split': 5}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:23:55,136] Trial 55 finished with value: 0.6391532732261858 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 5, 'min_samples_split': 2}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:24:47,569] Trial 56 finished with value: 0.6376793414347315 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 5, 'min_samples_split': 2}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:25:27,509] Trial 57 finished with value: 0.6390670325362603 and parameters: {'max_depth': 13, 'criterion': 'gini', 'n_estimators': 32, 'min_samples_
leaf': 5, 'min_samples_split': 2}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:25:37,163] Trial 58 finished with value: 0.6043041944335554 and parameters: {'max_depth': 3, 'criterion': 'entropy', 'n_estimators': 27, 'min_sample
s_leaf': 4, 'min_samples_split': 2}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:25:46,486] Trial 59 finished with value: 0.6308192865542924 and parameters: {'max_depth': 10, 'criterion': 'entropy', 'n_estimators': 8, 'min_sample
s_leaf': 5, 'min_samples_split': 2}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:26:35,694] Trial 60 finished with value: 0.6386201489611917 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 31, 'min_sampl
es_leaf': 6, 'min_samples_split': 4}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:27:24,511] Trial 61 finished with value: 0.6386907095256762 and parameters: {'max_depth': 15, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 7, 'min_samples_split': 3}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:28:14,807] Trial 62 finished with value: 0.6389651117208938 and parameters: {'max_depth': 16, 'criterion': 'entropy', 'n_estimators': 30, 'min_sampl
es_leaf': 8, 'min_samples_split': 3}. Best is trial 39 with value: 0.639937279498236.
[I 2021-12-13 11:29:04,177] Trial 63 finished with value: 0.6376479811838495 and parameters: {'max_depth': 14, 'criterion': 'entropy', 'n_estimators': 32, 'min_sampl
es_leaf': 6, 'min_samples_split': 2}. Best is trial 39 with value: 0.639937279498236.
```

In [9]:

```
optuna.visualization.plot_optimization_history(study_1)
```

In [10]:

```
optuna.visualization.plot_slice(study_1)
```

In [11]:

```
optuna.visualization.plot_param_importances(study_1)
```

Comentários acerca dos resultados da segunda execução

Os resultados foram melhores, é possível perceber uma tendência de crescimento em torno do 16, entretanto, a melhora foi de menos de 1% (a melhora foi de 0.04%). Tendo em vista que esta execução foi mais rápida (11 minutos a menos) que a primeira (que demorou um total de 50 minutos), então, o teste foi válido.

Gradient Boosting

In [ ]:

```
# Parâmetros default:
# *,loss='deviance',
# learning_rate=0.1,
# n_estimators=100,
# subsample=1.0,
# criterion='friedman_mse',
# min_samples_split=2,
# min_samples_leaf=1,
# min_weight_fraction_leaf=0.,
# max_depth=3,
```

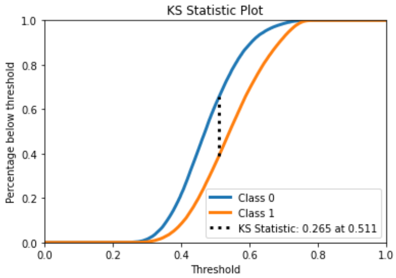
```python
    # min_impurity_decrease=0.,
    # min_impurity_split=None,
    # random_state=None,
    # max_features=None, verbose=0,
    # max_leaf_nodes=None,
    # warm_start=False,
    # validation_fraction=0.1,
    # n_iter_no_change=None,
    # tol=1e-4,
    # ccp_alpha=0.0
```

In [3]:

```python
ds_dropado = ds.drop(labels=['IND_BOM_1_2', 'IND_BOM_1_1'], axis=1)
df_treino = treino_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(treino_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao = valid_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(valid_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao.drop(columns=['INDEX'], axis=1, inplace=True)
df_teste = teste_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(teste_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_teste.drop(columns=['INDEX'], axis=1, inplace=True)

best_gb = []

def gb(trial):
    loss                 = trial.suggest_categorical("loss", ["deviance", "exponential"])
    max_depth            = trial.suggest_int("max_depth", 16, 64)
    subsample            = trial.suggest_float("subsample", 0.01, 0.7)
    n_estimators         = trial.suggest_int("n_estimators", 8, 32)
    learning_rate        = trial.suggest_float("learning_rate", 0.01, 0.7)
    min_samples_leaf     = trial.suggest_int("min_samples_leaf", 2, 8)

    # Código opcional para logar execuções
    # print('Gradient Boosting com {} estimadores e profundidade máxima {}, critério {}, min amostras de folhas {}, subsample de {} e taxa de aprendizagem {}\n'.form
    #     n_estimators, max_depth, loss, min_samples_leaf, subsample, learning_rate
    # ))


    gradient_boost = GradientBoostingClassifier(n_estimators = n_estimators,
                                       max_depth = max_depth,
                                       min_samples_leaf = min_samples_leaf,
                                       learning_rate = learning_rate,
                                       loss = loss,
                                       subsample = subsample).fit(df_treino, [0]*len(treino_ina)+[1]*len(treino_adi))
    best_gb.append(gradient_boost)

    gb_pred_class = gradient_boost.predict(df_validacao)
    gb_pred_scores = gradient_boost.predict_proba(df_validacao)

    accuracy, recall, precision, f1, auroc, aupr = compute_performance_metrics_sem_plot2([0]*len(valid_ina)+[1]*len(valid_adi), gb_pred_class, gb_pred_scores, "GB/GB

    return accuracy

study_2 = optuna.create_study(direction="maximize")
study_2.optimize(gb, n_trials=64)

compute_performance_metrics([0]*len(teste_ina)+[1]*len(teste_adi), best_gb[study_2.best_trial.number].predict(df_teste), best_gb[study_2.best_trial.number].predict_p
plot_confusion_matrix(best_gb[study_2.best_trial.number], df_teste, [0]*len(teste_ina)+[1]*len(teste_adi), normalize='true')
```

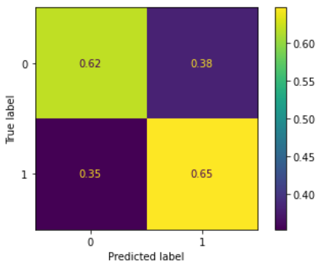
```
[I 2021-12-14 12:35:04,754] A new study created in memory with name: no-name-b8adeb50-a23d-463b-94fc-944a708e9a49
[I 2021-12-14 12:38:01,910] Trial 0 finished with value: 0.5818424147393179 and parameters: {'loss': 'deviance', 'max_depth': 51, 'subsample': 0.41731688066640243,
 'n_estimators': 9, 'learning_rate': 0.42063373317223757, 'min_samples_leaf': 7}. Best is trial 0 with value: 0.5818424147393179.
[I 2021-12-14 12:39:17,529] Trial 1 finished with value: 0.6161662093296746 and parameters: {'loss': 'exponential', 'max_depth': 31, 'subsample': 0.1314123597374676
2, 'n_estimators': 14, 'learning_rate': 0.06671473937284195, 'min_samples_leaf': 5}. Best is trial 1 with value: 0.6161662093296746.
[I 2021-12-14 12:43:08,316] Trial 2 finished with value: 0.5892512740101921 and parameters: {'loss': 'deviance', 'max_depth': 22, 'subsample': 0.5739603892662606, 'n
_estimators': 10, 'learning_rate': 0.4509134670537174, 'min_samples_leaf': 5}. Best is trial 1 with value: 0.6161662093296746.
[I 2021-12-14 12:51:59,330] Trial 3 finished with value: 0.5873069384555076 and parameters: {'loss': 'exponential', 'max_depth': 49, 'subsample': 0.4893683836994049
4, 'n_estimators': 18, 'learning_rate': 0.4457168114947387, 'min_samples_leaf': 6}. Best is trial 1 with value: 0.6161662093296746.
[I 2021-12-14 13:06:53,904] Trial 4 finished with value: 0.5679184633477068 and parameters: {'loss': 'deviance', 'max_depth': 59, 'subsample': 0.4111795716780603, 'n
_estimators': 26, 'learning_rate': 0.668913103794306, 'min_samples_leaf': 6}. Best is trial 1 with value: 0.6161662093296746.
[I 2021-12-14 13:08:28,060] Trial 5 finished with value: 0.5932575460603685 and parameters: {'loss': 'exponential', 'max_depth': 29, 'subsample': 0.1491510181091912
6, 'n_estimators': 15, 'learning_rate': 0.2699796443430259, 'min_samples_leaf': 7}. Best is trial 1 with value: 0.6161662093296746.
[I 2021-12-14 13:13:08,496] Trial 6 finished with value: 0.6314386515092121 and parameters: {'loss': 'exponential', 'max_depth': 50, 'subsample': 0.2579725016543647
5, 'n_estimators': 28, 'learning_rate': 0.04752114921679849, 'min_samples_leaf': 8}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 13:14:41,642] Trial 7 finished with value: 0.5917287338298707 and parameters: {'loss': 'exponential', 'max_depth': 33, 'subsample': 0.1424631140549395
8, 'n_estimators': 15, 'learning_rate': 0.27233566868925524, 'min_samples_leaf': 7}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 13:27:52,328] Trial 8 finished with value: 0.5986828694629557 and parameters: {'loss': 'deviance', 'max_depth': 36, 'subsample': 0.4913296374188595, 'n
_estimators': 32, 'learning_rate': 0.24636709272396276, 'min_samples_leaf': 6}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 13:30:16,812] Trial 9 finished with value: 0.5820070560564484 and parameters: {'loss': 'exponential', 'max_depth': 40, 'subsample': 0.2140168898881506,
'n_estimators': 12, 'learning_rate': 0.41168077342549714, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 13:36:31,477] Trial 10 finished with value: 0.6226421011368091 and parameters: {'loss': 'exponential', 'max_depth': 60, 'subsample': 0.265772968485450
6, 'n_estimators': 26, 'learning_rate': 0.041826506711636735, 'min_samples_leaf': 2}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 13:43:25,624] Trial 11 finished with value: 0.6241787534300275 and parameters: {'loss': 'exponential', 'max_depth': 64, 'subsample': 0.295669885379740
5, 'n_estimators': 26, 'learning_rate': 0.011678327716671377, 'min_samples_leaf': 2}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 13:50:35,465] Trial 12 finished with value: 0.6238494707957664 and parameters: {'loss': 'exponential', 'max_depth': 50, 'subsample': 0.307215464392541
3, 'n_estimators': 26, 'learning_rate': 0.01420892885396989, 'min_samples_leaf': 2}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 13:51:53,525] Trial 13 finished with value: 0.6002195217561741 and parameters: {'loss': 'exponential', 'max_depth': 64, 'subsample': 0.0509978056679876
7, 'n_estimators': 32, 'learning_rate': 0.14529728390732555, 'min_samples_leaf': 3}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:03:13,494] Trial 14 finished with value: 0.6132497059976479 and parameters: {'loss': 'exponential', 'max_depth': 44, 'subsample': 0.693723036182264
3, 'n_estimators': 23, 'learning_rate': 0.13897398485854534, 'min_samples_leaf': 8}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:18:05,511] Trial 15 finished with value: 0.5748490787926304 and parameters: {'loss': 'exponential', 'max_depth': 56, 'subsample': 0.34523407420553,
 'n_estimators': 29, 'learning_rate': 0.6067669229563805, 'min_samples_leaf': 3}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:18:20,596] Trial 16 finished with value: 0.5962289298314386 and parameters: {'loss': 'exponential', 'max_depth': 64, 'subsample': 0.0212087414769754
62, 'n_estimators': 21, 'learning_rate': 0.14153073915135736, 'min_samples_leaf': 8}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:23:37,398] Trial 17 finished with value: 0.624264994119953 and parameters: {'loss': 'exponential', 'max_depth': 55, 'subsample': 0.2321987669565929
2, 'n_estimators': 29, 'learning_rate': 0.07499051604044929, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:29:33,620] Trial 18 finished with value: 0.6029949039592317 and parameters: {'loss': 'exponential', 'max_depth': 54, 'subsample': 0.2237224070115203
2, 'n_estimators': 30, 'learning_rate': 0.19934408569947926, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:31:26,238] Trial 19 finished with value: 0.6150921207369658 and parameters: {'loss': 'exponential', 'max_depth': 45, 'subsample': 0.0879625361394077
4, 'n_estimators': 29, 'learning_rate': 0.08207508291383014, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:36:10,382] Trial 20 finished with value: 0.5887416699333594 and parameters: {'loss': 'exponential', 'max_depth': 45, 'subsample': 0.2032758195502212
8, 'n_estimators': 22, 'learning_rate': 0.3377122823801209, 'min_samples_leaf': 3}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:42:31,861] Trial 21 finished with value: 0.621889455115641 and parameters: {'loss': 'exponential', 'max_depth': 59, 'subsample': 0.2968096705626031
6, 'n_estimators': 24, 'learning_rate': 0.014074752280582241, 'min_samples_leaf': 2}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 14:52:15,721] Trial 22 finished with value: 0.6214190513524108 and parameters: {'loss': 'exponential', 'max_depth': 55, 'subsample': 0.3980963671588362
3, 'n_estimators': 28, 'learning_rate': 0.09287860687893003, 'min_samples_leaf': 3}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 15:00:49,117] Trial 23 finished with value: 0.6276048608388867 and parameters: {'loss': 'exponential', 'max_depth': 64, 'subsample': 0.351878266203777,
'n_estimators': 32, 'learning_rate': 0.016322221599021667, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 15:10:53,484] Trial 24 finished with value: 0.6064445315562524 and parameters: {'loss': 'exponential', 'max_depth': 61, 'subsample': 0.3484974780333164
3, 'n_estimators': 32, 'learning_rate': 0.17788653455169182, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 15:16:36,402] Trial 25 finished with value: 0.6231360250882007 and parameters: {'loss': 'exponential', 'max_depth': 51, 'subsample': 0.241969416133156
9, 'n_estimators': 31, 'learning_rate': 0.09041653157548707, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 15:20:54,991] Trial 26 finished with value: 0.6006507252058016 and parameters: {'loss': 'exponential', 'max_depth': 55, 'subsample': 0.1776160851406041
2, 'n_estimators': 28, 'learning_rate': 0.2151699690282359, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 15:27:34,582] Trial 27 finished with value: 0.6180086240689926 and parameters: {'loss': 'deviance', 'max_depth': 41, 'subsample': 0.4834293819226928,
 'n_estimators': 19, 'learning_rate': 0.016231473948409902, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 15:35:24,070] Trial 28 finished with value: 0.6152410819286555 and parameters: {'loss': 'exponential', 'max_depth': 47, 'subsample': 0.371542521267086
3, 'n_estimators': 24, 'learning_rate': 0.11625962686694558, 'min_samples_leaf': 3}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 15:45:35,957] Trial 29 finished with value: 0.6192316738533908 and parameters: {'loss': 'deviance', 'max_depth': 53, 'subsample': 0.45870540357989786,
 'n_estimators': 31, 'learning_rate': 0.05080678509819957, 'min_samples_leaf': 6}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 16:07:28,987] Trial 30 finished with value: 0.5765660525284202 and parameters: {'loss': 'exponential', 'max_depth': 58, 'subsample': 0.560492431696823
7, 'n_estimators': 28, 'learning_rate': 0.5392764834510377, 'min_samples_leaf': 4}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 16:14:15,682] Trial 31 finished with value: 0.6221952175617405 and parameters: {'loss': 'exponential', 'max_depth': 64, 'subsample': 0.2934739198709125
4, 'n_estimators': 26, 'learning_rate': 0.014186818779906685, 'min_samples_leaf': 2}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 16:21:01,778] Trial 32 finished with value: 0.6291023128185026 and parameters: {'loss': 'exponential', 'max_depth': 62, 'subsample': 0.3207589617172249
6, 'n_estimators': 30, 'learning_rate': 0.05549490976336203, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 16:28:14,515] Trial 33 finished with value: 0.6275813406507252 and parameters: {'loss': 'exponential', 'max_depth': 61, 'subsample': 0.338290657854987
9, 'n_estimators': 30, 'learning_rate': 0.05902615527448696, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 16:35:38,121] Trial 34 finished with value: 0.6281301450411604 and parameters: {'loss': 'exponential', 'max_depth': 63, 'subsample': 0.336877442121633
9, 'n_estimators': 31, 'learning_rate': 0.05253744795324095, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
```

```
[I 2021-12-14 16:46:38,765] Trial 35 finished with value: 0.6180243041944335 and parameters: {'loss': 'exponential', 'max_depth': 63, 'subsample': 0.443655568786719
2, 'n_estimators': 32, 'learning_rate': 0.11773442007845447, 'min_samples_leaf': 6}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 16:53:45,584] Trial 36 finished with value: 0.6145276362210897 and parameters: {'loss': 'deviance', 'max_depth': 18, 'subsample': 0.4022687317313694,
 'n_estimators': 30, 'learning_rate': 0.16454165830233736, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 16:55:17,303] Trial 37 finished with value: 0.5953508428067424 and parameters: {'loss': 'exponential', 'max_depth': 58, 'subsample': 0.263405274305026
5, 'n_estimators': 8, 'learning_rate': 0.3224330090785183, 'min_samples_leaf': 7}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:07:14,748] Trial 38 finished with value: 0.6034731477851822 and parameters: {'loss': 'exponential', 'max_depth': 62, 'subsample': 0.534852684175856
9, 'n_estimators': 27, 'learning_rate': 0.2263816606934624, 'min_samples_leaf': 6}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:11:15,767] Trial 39 finished with value: 0.6266797334378675 and parameters: {'loss': 'exponential', 'max_depth': 48, 'subsample': 0.3735056680838108
6, 'n_estimators': 17, 'learning_rate': 0.04713891892149584, 'min_samples_leaf': 8}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:15:09,123] Trial 40 finished with value: 0.5769345354762838 and parameters: {'loss': 'deviance', 'max_depth': 26, 'subsample': 0.1855605964526241,
 'n_estimators': 31, 'learning_rate': 0.2945668454007403, 'min_samples_leaf': 7}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:22:08,767] Trial 41 finished with value: 0.6298157585260682 and parameters: {'loss': 'exponential', 'max_depth': 60, 'subsample': 0.330867354086940
1, 'n_estimators': 30, 'learning_rate': 0.05007781004029695, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:30:04,222] Trial 42 finished with value: 0.617828302626421 and parameters: {'loss': 'exponential', 'max_depth': 58, 'subsample': 0.3247065661907484,
'n_estimators': 32, 'learning_rate': 0.10762421496513438, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:35:18,802] Trial 43 finished with value: 0.6308584868678949 and parameters: {'loss': 'exponential', 'max_depth': 52, 'subsample': 0.2645281488586697
6, 'n_estimators': 30, 'learning_rate': 0.039914203903854986, 'min_samples_leaf': 6}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:39:56,302] Trial 44 finished with value: 0.6279733437867503 and parameters: {'loss': 'exponential', 'max_depth': 36, 'subsample': 0.259423963197687
3, 'n_estimators': 27, 'learning_rate': 0.04734402416427355, 'min_samples_leaf': 6}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:42:26,112] Trial 45 finished with value: 0.6015288122304978 and parameters: {'loss': 'exponential', 'max_depth': 53, 'subsample': 0.113134608819972,
'n_estimators': 30, 'learning_rate': 0.17140575203818215, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:47:04,633] Trial 46 finished with value: 0.62421795374363 and parameters: {'loss': 'exponential', 'max_depth': 51, 'subsample': 0.27944072640278145,
'n_estimators': 25, 'learning_rate': 0.07120912744176179, 'min_samples_leaf': 7}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 17:57:17,538] Trial 47 finished with value: 0.582391219129753 and parameters: {'loss': 'exponential', 'max_depth': 57, 'subsample': 0.3168330896002284,
'n_estimators': 29, 'learning_rate': 0.39111194079477785, 'min_samples_leaf': 5}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 18:05:17,888] Trial 48 finished with value: 0.617985103880831 and parameters: {'loss': 'exponential', 'max_depth': 60, 'subsample': 0.3849978028074596
5, 'n_estimators': 27, 'learning_rate': 0.12483920986814033, 'min_samples_leaf': 6}. Best is trial 6 with value: 0.6314386515092121.
[I 2021-12-14 18:08:42,549] Trial 49 finished with value: 0.6334300274402195 and parameters: {'loss': 'exponential', 'max_depth': 50, 'subsample': 0.165472995188781
3, 'n_estimators': 31, 'learning_rate': 0.021201155244168258, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:11:36,807] Trial 50 finished with value: 0.6314778518228146 and parameters: {'loss': 'exponential', 'max_depth': 47, 'subsample': 0.1580201575314062
2, 'n_estimators': 28, 'learning_rate': 0.010519353142034338, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:14:46,002] Trial 51 finished with value: 0.6323951391611133 and parameters: {'loss': 'exponential', 'max_depth': 42, 'subsample': 0.1628030833651353
8, 'n_estimators': 29, 'learning_rate': 0.028596553591318043, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:17:28,031] Trial 52 finished with value: 0.6320188161505292 and parameters: {'loss': 'exponential', 'max_depth': 42, 'subsample': 0.1473651568537200
4, 'n_estimators': 28, 'learning_rate': 0.016641431323441203, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:20:03,361] Trial 53 finished with value: 0.629266954135633 and parameters: {'loss': 'exponential', 'max_depth': 43, 'subsample': 0.1560142563418976
4, 'n_estimators': 25, 'learning_rate': 0.02340944562385857, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:21:17,434] Trial 54 finished with value: 0.6297765582124657 and parameters: {'loss': 'exponential', 'max_depth': 36, 'subsample': 0.0743877150032593
6, 'n_estimators': 28, 'learning_rate': 0.010136829107792527, 'min_samples_leaf': 7}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:23:20,802] Trial 55 finished with value: 0.6184319874558997 and parameters: {'loss': 'exponential', 'max_depth': 39, 'subsample': 0.1134749487961623
6, 'n_estimators': 27, 'learning_rate': 0.09003986716630792, 'min_samples_leaf': 7}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:26:29,590] Trial 56 finished with value: 0.6313288906311251 and parameters: {'loss': 'exponential', 'max_depth': 46, 'subsample': 0.1607146315212672
5, 'n_estimators': 29, 'learning_rate': 0.031091349667572658, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:27:16,783] Trial 57 finished with value: 0.559827518620149 and parameters: {'loss': 'exponential', 'max_depth': 47, 'subsample': 0.0350368993762300
3, 'n_estimators': 25, 'learning_rate': 0.49428655220336254, 'min_samples_leaf': 7}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:30:13,941] Trial 58 finished with value: 0.6327636221089769 and parameters: {'loss': 'exponential', 'max_depth': 42, 'subsample': 0.16181694168567,
 'n_estimators': 29, 'learning_rate': 0.030016547923153756, 'min_samples_leaf': 8}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:32:15,654] Trial 59 finished with value: 0.6208545668365347 and parameters: {'loss': 'exponential', 'max_depth': 42, 'subsample': 0.1368467883146354
3, 'n_estimators': 23, 'learning_rate': 0.07924144571123334, 'min_samples_leaf': 8}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:36:17,709] Trial 60 finished with value: 0.6115954527636221 and parameters: {'loss': 'exponential', 'max_depth': 39, 'subsample': 0.2016043567158102
6, 'n_estimators': 28, 'learning_rate': 0.14811803315461564, 'min_samples_leaf': 8}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:39:04,310] Trial 61 finished with value: 0.6326303410427283 and parameters: {'loss': 'exponential', 'max_depth': 45, 'subsample': 0.159784243551669
6, 'n_estimators': 26, 'learning_rate': 0.031211724432891844, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:40:56,445] Trial 62 finished with value: 0.6306938455507644 and parameters: {'loss': 'exponential', 'max_depth': 49, 'subsample': 0.1078547522293435
1, 'n_estimators': 27, 'learning_rate': 0.02965155218920697, 'min_samples_leaf': 6}. Best is trial 49 with value: 0.6334300274402195.
[I 2021-12-14 18:42:14,617] Trial 63 finished with value: 0.6146452371618973 and parameters: {'loss': 'exponential', 'max_depth': 43, 'subsample': 0.0704425676058879
8, 'n_estimators': 29, 'learning_rate': 0.09844070708162139, 'min_samples_leaf': 7}. Best is trial 49 with value: 0.6334300274402195.
```



Out[3]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f6168865be0>
```

```
<Figure size 432x288 with 0 Axes>
```



Teste de Kolmogorov-Smirnov (KS) e matriz de confusão da Gradient Boosting

Acima, podemos ver os resultados de ambos. O teste de Kolmogorov-Smirnov assemelha-se, assim como a Random Forest, bastante a uma distribuição normal. Mais uma vez (semelhante a Random Forest), a matriz foi bem sucedida para identificar resultados verdadeiros e falsos. Entretanto, o Gradient Boosting peca pela sua lenta execução, a qual apresenta resultados inferiores aos do Random Forest.

Resultados do Gradient Boosting

In [4]:
```
optuna.visualization.plot_optimization_history(study_2)
```

In [5]:
```
optuna.visualization.plot_slice(study_2)
```

In [6]:
```
optuna.visualization.plot_param_importances(study_2)
```

Comentários acerca dos resultados do Gredient Boosting

O learning rate é certamente o mais importante dos hiperparâmetros, vemos que quanto mais próximo a 0 melhor o seu desempenho. Um número maior de estimadores foi o que obteve melhores resultados. Para subsamples, o intervalo ao redor de 0.2 obteve os melhores resultados. O min samples leaf obteve resultados consistentes no valor de 6. Por fim, a profundidade máxima se saiu bem com valores entre 40 e 50, o que foi uma surpresa, pois intuitivamente, imaginávamos que maiores profundidades iriam obter melhores resultados.

Regressão Logística

In [ ]:

```
# Parâmetros default
# penalty='l2',
# *,
# dual=False,
# tol=1e-4,
# C=1.0,
# fit_intercept=True,
# intercept_scaling=1,
# class_weight=None,
# random_state=None,
# solver='lbfgs',
# max_iter=100,
# multi_class='auto',
# verbose=0,
# warm_start=False,
# n_jobs=None,
# l1_ratio=None
```

Teste de Kolmogorov-Smirnov (KS) e matriz de confusão da Regressão Logística

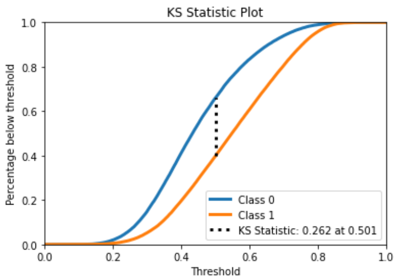
Acima, podemos ver os resultados de ambos. O teste de Kolmogorov-Smirnov assemelha-se, assim como a Random Forest e Gradient Boosting, bastante a uma distribuição normal. Semelhante aos já supracitados, a matriz foi melhor sucedida para identificar resultados falsos, porém, os resultados caíram na identificação de resultados falsos mas foram os melhores até então para identificar resultados verdadeiros.

In [3]:

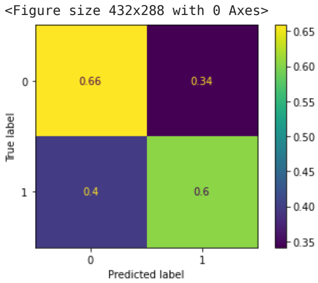
```
ds_dropado = ds.drop(labels=['IND_BOM_1_2', 'IND_BOM_1_1'], axis=1)
df_treino = treino_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(treino_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao = valid_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(valid_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao.drop(columns=['INDEX'], axis=1, inplace=True)
df_teste = teste_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(teste_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_teste.drop(columns=['INDEX'], axis=1, inplace=True)

best_rl = []

def rl(trial):
    # penalty = trial.suggest_categorical("penalty", ["none", "l2", "l1", "elasticnet"])
    solver = trial.suggest_categorical("solver", ["newton-cg", "lbfgs", "liblinear", "sag", "saga"])
    C = trial.suggest_float("C", 0.01, 1.0)
    multi_class = "ovr"

    # Código opcional para log
    # print('Regressão Logística com solver {} e C {}\n'.format(
    #     solver, C
    # ))

    regressao_log = LogisticRegression(max_iter=1000,
                                random_state=1,
                                solver=solver,
                                C=C,
                                multi_class=multi_class).fit(df_treino, [0]*len(treino_ina)+[1]*len(treino_adi))

    best_rl.append(regressao_log)

    rl_pred_class = regressao_log.predict(df_validacao)
    rl_pred_scores = regressao_log.predict_proba(df_validacao)

    accuracy, recall, precision, f1, auroc, aupr = compute_performance_metrics_sem_plot2([0]*len(valid_ina)+[1]*len(valid_adi), rl_pred_class, rl_pred_scores,  "LR/L

    return accuracy

study_3 = optuna.create_study(direction="maximize")
study_3.optimize(rl, n_trials=64)

compute_performance_metrics([0]*len(teste_ina)+[1]*len(teste_adi), best_rl[study_3.best_trial.number].predict(df_teste), best_rl[study_3.best_trial.number].predict_p
plot_confusion_matrix(best_rl[study_3.best_trial.number], df_teste, [0]*len(teste_ina)+[1]*len(teste_adi), normalize='true')
```

```
[I 2021-12-13 17:28:29,982] A new study created in memory with name: no-name-ea7c6a75-d17e-45ea-8b7f-6053360cb5ec
[I 2021-12-13 17:28:45,474] Trial 0 finished with value: 0.631344570756566 and parameters: {'solver': 'saga', 'C': 0.5849769857357946}. Best is trial 0 with value:
 0.631344570756566.
[I 2021-12-13 17:29:00,180] Trial 1 finished with value: 0.631485691885535 and parameters: {'solver': 'saga', 'C': 0.2842694673818212}. Best is trial 1 with value:
 0.631485691885535.
[I 2021-12-13 17:29:24,171] Trial 2 finished with value: 0.6313524108192865 and parameters: {'solver': 'sag', 'C': 0.6668490171508832}. Best is trial 1 with value:
 0.631485691885535.
[I 2021-12-13 17:29:45,635] Trial 3 finished with value: 0.6313288906311251 and parameters: {'solver': 'liblinear', 'C': 0.6562292021435742}. Best is trial 1 with va
lue: 0.631485691885535.
[I 2021-12-13 17:30:02,812] Trial 4 finished with value: 0.631313210505684 and parameters: {'solver': 'saga', 'C': 0.7034896202550367}. Best is trial 1 with value:
 0.631485691885535.
[I 2021-12-13 17:30:19,197] Trial 5 finished with value: 0.6313524108192865 and parameters: {'solver': 'lbfgs', 'C': 0.7197685296802497}. Best is trial 1 with value:
0.631485691885535.
[I 2021-12-13 17:31:29,515] Trial 6 finished with value: 0.6314072912583301 and parameters: {'solver': 'newton-cg', 'C': 0.4162968317742919}. Best is trial 1 with va
lue: 0.631485691885535.
[I 2021-12-13 17:31:46,629] Trial 7 finished with value: 0.6313837710701685 and parameters: {'solver': 'saga', 'C': 0.45572740122819116}. Best is trial 1 with value:
0.631485691885535.
[I 2021-12-13 17:32:04,158] Trial 8 finished with value: 0.6314151313210505 and parameters: {'solver': 'saga', 'C': 0.4131170914789713}. Best is trial 1 with value:
 0.631485691885535.
[I 2021-12-13 17:32:26,437] Trial 9 finished with value: 0.6313210505684046 and parameters: {'solver': 'liblinear', 'C': 0.789271822152566}. Best is trial 1 with val
ue: 0.631485691885535.
[I 2021-12-13 17:33:26,179] Trial 10 finished with value: 0.6313210505684046 and parameters: {'solver': 'newton-cg', 'C': 0.038839105626143844}. Best is trial 1 with
value: 0.631485691885535.
[I 2021-12-13 17:33:45,271] Trial 11 finished with value: 0.6315013720109761 and parameters: {'solver': 'saga', 'C': 0.20663194752199038}. Best is trial 11 with valu
e: 0.6315013720109761.
[I 2021-12-13 17:34:03,176] Trial 12 finished with value: 0.631485691885535 and parameters: {'solver': 'saga', 'C': 0.168295196141211}. Best is trial 11 with value:
 0.6315013720109761.
[I 2021-12-13 17:34:21,116] Trial 13 finished with value: 0.6315170521364171 and parameters: {'solver': 'saga', 'C': 0.21342254963865626}. Best is trial 13 with valu
e: 0.6315170521364171.
[I 2021-12-13 17:34:40,620] Trial 14 finished with value: 0.631172089376715 and parameters: {'solver': 'sag', 'C': 0.016941920740670358}. Best is trial 13 with valu
e: 0.6315170521364171.
[I 2021-12-13 17:34:58,054] Trial 15 finished with value: 0.6314308114464916 and parameters: {'solver': 'lbfgs', 'C': 0.23564256637746384}. Best is trial 13 with val
ue: 0.6315170521364171.
[I 2021-12-13 17:35:17,987] Trial 16 finished with value: 0.6313053704429635 and parameters: {'solver': 'saga', 'C': 0.9932414743438174}. Best is trial 13 with valu
e: 0.6315170521364171.
[I 2021-12-13 17:35:38,429] Trial 17 finished with value: 0.6314229713837711 and parameters: {'solver': 'saga', 'C': 0.1380865293926395}. Best is trial 13 with valu
e: 0.6315170521364171.
[I 2021-12-13 17:35:56,890] Trial 18 finished with value: 0.6314308114464916 and parameters: {'solver': 'saga', 'C': 0.30825183139007606}. Best is trial 13 with valu
e: 0.6315170521364171.
[I 2021-12-13 17:36:35,659] Trial 19 finished with value: 0.6314621716973736 and parameters: {'solver': 'sag', 'C': 0.0859798302036967}. Best is trial 13 with value:
0.6315170521364171.
[I 2021-12-13 17:36:57,934] Trial 20 finished with value: 0.6314151313210505 and parameters: {'solver': 'liblinear', 'C': 0.357885375301218}. Best is trial 13 with v
alue: 0.6315170521364171.
[I 2021-12-13 17:37:16,157] Trial 21 finished with value: 0.6315092120736966 and parameters: {'solver': 'saga', 'C': 0.2567440157756343}. Best is trial 13 with valu
e: 0.6315170521364171.
[I 2021-12-13 17:37:36,533] Trial 22 finished with value: 0.6315013720109761 and parameters: {'solver': 'saga', 'C': 0.19101557903025057}. Best is trial 13 with valu
e: 0.6315170521364171.
[I 2021-12-13 17:37:53,870] Trial 23 finished with value: 0.6315248921991375 and parameters: {'solver': 'saga', 'C': 0.25305676755509275}. Best is trial 23 with valu
e: 0.6315248921991375.
[I 2021-12-13 17:38:11,501] Trial 24 finished with value: 0.6314229713837711 and parameters: {'solver': 'saga', 'C': 0.33459591611069894}. Best is trial 23 with valu
e: 0.6315248921991375.
[I 2021-12-13 17:38:30,957] Trial 25 finished with value: 0.631297530380243 and parameters: {'solver': 'lbfgs', 'C': 0.5311214530379987}. Best is trial 23 with valu
e: 0.6315248921991375.
[I 2021-12-13 17:39:35,089] Trial 26 finished with value: 0.6314072912583301 and parameters: {'solver': 'newton-cg', 'C': 0.10494444636335767}. Best is trial 23 with
value: 0.6315248921991375.
[I 2021-12-13 17:39:53,818] Trial 27 finished with value: 0.631485691885535 and parameters: {'solver': 'saga', 'C': 0.2869459413780808}. Best is trial 23 with value:
0.6315248921991375.
[I 2021-12-13 17:40:11,577] Trial 28 finished with value: 0.6315640925127401 and parameters: {'solver': 'saga', 'C': 0.23216200775438173}. Best is trial 28 with valu
e: 0.6315640925127401.
```

```
[I 2021-12-13 17:40:29,994] Trial 29 finished with value: 0.631375931007448 and parameters: {'solver': 'saga', 'C': 0.569675094227418}. Best is trial 28 with value:
 0.6315640925127401.
[I 2021-12-13 17:41:07,216] Trial 30 finished with value: 0.6315327322618581 and parameters: {'solver': 'saga', 'C': 0.06388184901496352}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:41:45,511] Trial 31 finished with value: 0.6314778518228146 and parameters: {'solver': 'saga', 'C': 0.0527578981823673}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:42:00,570] Trial 32 finished with value: 0.6314229713837711 and parameters: {'solver': 'saga', 'C': 0.13968279418122004}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:42:11,718] Trial 33 finished with value: 0.6315484123872991 and parameters: {'solver': 'saga', 'C': 0.229259702423482}. Best is trial 28 with value:
0.6315640925127401.
[I 2021-12-13 17:42:22,377] Trial 34 finished with value: 0.6314151313210505 and parameters: {'solver': 'saga', 'C': 0.3855739753054221}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:42:48,895] Trial 35 finished with value: 0.6314229713837711 and parameters: {'solver': 'sag', 'C': 0.09985295367135301}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:43:01,973] Trial 36 finished with value: 0.631109368874951 and parameters: {'solver': 'liblinear', 'C': 0.016512103852271226}. Best is trial 28 with
value: 0.6315640925127401.
[I 2021-12-13 17:43:13,006] Trial 37 finished with value: 0.6315013720109761 and parameters: {'solver': 'saga', 'C': 0.27216893174449597}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:43:24,899] Trial 38 finished with value: 0.631313210505684 and parameters: {'solver': 'lbfgs', 'C': 0.46842789059036777}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:43:36,948] Trial 39 finished with value: 0.6314151313210505 and parameters: {'solver': 'saga', 'C': 0.14468623959222282}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:44:12,358] Trial 40 finished with value: 0.6314308114464916 and parameters: {'solver': 'newton-cg', 'C': 0.3444907385797983}. Best is trial 28 with
 value: 0.6315640925127401.
[I 2021-12-13 17:44:23,687] Trial 41 finished with value: 0.6315484123872991 and parameters: {'solver': 'saga', 'C': 0.22114137732095895}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:44:34,980] Trial 42 finished with value: 0.6315640925127401 and parameters: {'solver': 'saga', 'C': 0.23180877186234763}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:44:46,306] Trial 43 finished with value: 0.6315013720109761 and parameters: {'solver': 'saga', 'C': 0.1789596585249374}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:45:14,903] Trial 44 finished with value: 0.6314700117600941 and parameters: {'solver': 'saga', 'C': 0.06970439647332966}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:45:25,814] Trial 45 finished with value: 0.6314151313210505 and parameters: {'solver': 'saga', 'C': 0.31241240553059696}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:45:37,859] Trial 46 finished with value: 0.6314151313210505 and parameters: {'solver': 'saga', 'C': 0.41471476891886294}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:45:49,167] Trial 47 finished with value: 0.6315170521364171 and parameters: {'solver': 'saga', 'C': 0.21215788966190618}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:46:05,066] Trial 48 finished with value: 0.6314464915719326 and parameters: {'solver': 'liblinear', 'C': 0.12443291331195848}. Best is trial 28 with
value: 0.6315640925127401.
[I 2021-12-13 17:46:15,986] Trial 49 finished with value: 0.6313994511956096 and parameters: {'solver': 'saga', 'C': 0.4623321889705078}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:46:35,975] Trial 50 finished with value: 0.6313053704429635 and parameters: {'solver': 'sag', 'C': 0.8537853763331291}. Best is trial 28 with value:
0.6315640925127401.
[I 2021-12-13 17:46:48,625] Trial 51 finished with value: 0.6314935319482556 and parameters: {'solver': 'saga', 'C': 0.2618101050476938}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:46:59,931] Trial 52 finished with value: 0.6315640925127401 and parameters: {'solver': 'saga', 'C': 0.23150423611955934}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:47:11,233] Trial 53 finished with value: 0.631485691885535 and parameters: {'solver': 'saga', 'C': 0.17387524321300313}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:47:22,525] Trial 54 finished with value: 0.6315092120736966 and parameters: {'solver': 'saga', 'C': 0.21581937086086203}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:47:33,328] Trial 55 finished with value: 0.6314151313210505 and parameters: {'solver': 'saga', 'C': 0.3860063462522674}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:48:21,245] Trial 56 finished with value: 0.631454331634653 and parameters: {'solver': 'newton-cg', 'C': 0.2972336940778451}. Best is trial 28 with v
alue: 0.6315640925127401.
[I 2021-12-13 17:48:32,530] Trial 57 finished with value: 0.6315248921991375 and parameters: {'solver': 'saga', 'C': 0.21874548883791752}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:48:44,854] Trial 58 finished with value: 0.6313916111328891 and parameters: {'solver': 'lbfgs', 'C': 0.17720251004516163}. Best is trial 28 with val
ue: 0.6315640925127401.
[I 2021-12-13 17:49:17,619] Trial 59 finished with value: 0.6313210505684046 and parameters: {'solver': 'saga', 'C': 0.040364038465462396}. Best is trial 28 with val
ue: 0.6315640925127401.
[I 2021-12-13 17:49:28,954] Trial 60 finished with value: 0.6315327322618581 and parameters: {'solver': 'saga', 'C': 0.2503863913752739}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:49:39,781] Trial 61 finished with value: 0.6314151313210505 and parameters: {'solver': 'saga', 'C': 0.3223366806518014}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:49:51,095] Trial 62 finished with value: 0.6315170521364171 and parameters: {'solver': 'saga', 'C': 0.2419893551273464}. Best is trial 28 with valu
e: 0.6315640925127401.
[I 2021-12-13 17:50:02,026] Trial 63 finished with value: 0.6313994511956096 and parameters: {'solver': 'saga', 'C': 0.36343712697564384}. Best is trial 28 with valu
e: 0.6315640925127401.
```



Out[3]:
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay at 0x7f4a1d2ab160>

<Figure size 432x288 with 0 Axes>
```



Resultados da Regressão Logística

In [4]:
```
optuna.visualization.plot_optimization_history(study_3)
```

In [5]:
```
optuna.visualization.plot_slice(study_3)
```

In [6]:
```
optuna.visualization.plot_param_importances(study_3)
```

Comentários acerca dos resultados da Regressão Logística

Vimos que há uma tendência de resultados melhores na região de 0.2 até 0.26 para a variável C. Vamos afunilar e gerar mais testes nesse intervalo.

In [7]:
```python
def rl(trial):
    # penalty = trial.suggest_categorical("penalty", ["none", "l2", "l1", "elasticnet"])
    solver = trial.suggest_categorical("solver", ["newton-cg", "lbfgs", "liblinear", "sag", "saga"])
    C = trial.suggest_float("C", 0.19, 0.26)
    multi_class = "ovr"

    regressao_log = LogisticRegression(max_iter=1000,
                                       random_state=1,
                                       solver=solver,
                                       C=C,
                                       multi_class=multi_class).fit(df_treino, [0]*len(treino_ina)+[1]*len(treino_adi))

    best_rl.append(regressao_log)

    rl_pred_class = regressao_log.predict(df_validacao)
    rl_pred_scores = regressao_log.predict_proba(df_validacao)

    accuracy, recall, precision, f1, auroc, aupr = compute_performance_metrics_sem_plot([0]*len(valid_ina)+[1]*len(valid_adi), rl_pred_class, rl_pred_scores)

    return accuracy

study_3 = optuna.create_study(direction="maximize")
study_3.optimize(rl, n_trials=64)
```

```
[I 2021-12-13 17:53:19,003] A new study created in memory with name: no-name-ce42977e-cad4-43c8-a88c-c1c38867c807
[I 2021-12-13 17:53:30,829] Trial 0 finished with value: 0.6315327322618581 and parameters: {'solver': 'saga', 'C': 0.24683243924552206}. Best is trial 0 with value:
0.6315327322618581.
[I 2021-12-13 17:53:49,486] Trial 1 finished with value: 0.6315327322618581 and parameters: {'solver': 'liblinear', 'C': 0.24644369495349933}. Best is trial 0 with v
alue: 0.6315327322618581.
[I 2021-12-13 17:54:47,903] Trial 2 finished with value: 0.6315405723245786 and parameters: {'solver': 'newton-cg', 'C': 0.23096729387561515}. Best is trial 2 with v
alue: 0.6315405723245786.
[I 2021-12-13 17:55:00,556] Trial 3 finished with value: 0.6315327322618581 and parameters: {'solver': 'saga', 'C': 0.2502997486409192}. Best is trial 2 with value:
 0.6315405723245786.
[I 2021-12-13 17:55:48,734] Trial 4 finished with value: 0.6315092120736966 and parameters: {'solver': 'newton-cg', 'C': 0.20775547948277212}. Best is trial 2 with v
alue: 0.6315405723245786.
[I 2021-12-13 17:56:00,787] Trial 5 finished with value: 0.6315170521364171 and parameters: {'solver': 'saga', 'C': 0.2124200236063185}. Best is trial 2 with value:
 0.6315405723245786.
[I 2021-12-13 17:56:13,063] Trial 6 finished with value: 0.6315092120736966 and parameters: {'solver': 'saga', 'C': 0.1963930960061364}. Best is trial 2 with value:
 0.6315405723245786.
[I 2021-12-13 17:56:28,126] Trial 7 finished with value: 0.6313367306938455 and parameters: {'solver': 'lbfgs', 'C': 0.21235618670109935}. Best is trial 2 with valu
e: 0.6315405723245786.
[I 2021-12-13 17:56:50,338] Trial 8 finished with value: 0.6315092120736966 and parameters: {'solver': 'sag', 'C': 0.25813895016555877}. Best is trial 2 with value:
 0.6315405723245786.
[I 2021-12-13 17:58:04,165] Trial 9 finished with value: 0.6315092120736966 and parameters: {'solver': 'newton-cg', 'C': 0.20316898685680146}. Best is trial 2 with v
alue: 0.6315405723245786.
[I 2021-12-13 17:58:54,780] Trial 10 finished with value: 0.6315405723245786 and parameters: {'solver': 'newton-cg', 'C': 0.22778793145687184}. Best is trial 2 with
 value: 0.6315405723245786.
[I 2021-12-13 17:59:42,598] Trial 11 finished with value: 0.6315405723245786 and parameters: {'solver': 'newton-cg', 'C': 0.2336088773707623}. Best is trial 2 with v
alue: 0.6315405723245786.
[I 2021-12-13 18:00:40,429] Trial 12 finished with value: 0.6315405723245786 and parameters: {'solver': 'newton-cg', 'C': 0.2279261916805474}. Best is trial 2 with v
alue: 0.6315405723245786.
[I 2021-12-13 18:01:22,170] Trial 13 finished with value: 0.6315405723245786 and parameters: {'solver': 'newton-cg', 'C': 0.23512507970331903}. Best is trial 2 with
 value: 0.6315405723245786.
[I 2021-12-13 18:02:11,794] Trial 14 finished with value: 0.6315248921991375 and parameters: {'solver': 'newton-cg', 'C': 0.2212772126531317}. Best is trial 2 with v
alue: 0.6315405723245786.
[I 2021-12-13 18:02:29,754] Trial 15 finished with value: 0.6315327322618581 and parameters: {'solver': 'liblinear', 'C': 0.22220047554576147}. Best is trial 2 with
 value: 0.6315405723245786.
[I 2021-12-13 18:02:41,781] Trial 16 finished with value: 0.6313680909447276 and parameters: {'solver': 'lbfgs', 'C': 0.23870672131520754}. Best is trial 2 with valu
e: 0.6315405723245786.
[I 2021-12-13 18:03:07,009] Trial 17 finished with value: 0.6315640925127401 and parameters: {'solver': 'sag', 'C': 0.23832960730630523}. Best is trial 17 with valu
e: 0.6315640925127401.
[I 2021-12-13 18:03:28,454] Trial 18 finished with value: 0.6315092120736966 and parameters: {'solver': 'sag', 'C': 0.25890860331921356}. Best is trial 17 with valu
e: 0.6315640925127401.
[I 2021-12-13 18:03:51,667] Trial 19 finished with value: 0.6315484123872991 and parameters: {'solver': 'sag', 'C': 0.2401800322848446}. Best is trial 17 with value:
0.6315640925127401.
[I 2021-12-13 18:04:15,591] Trial 20 finished with value: 0.6315405723245786 and parameters: {'solver': 'sag', 'C': 0.24104428209663428}. Best is trial 17 with valu
e: 0.6315640925127401.
[I 2021-12-13 18:04:37,834] Trial 21 finished with value: 0.6315405723245786 and parameters: {'solver': 'sag', 'C': 0.2417431880495569}. Best is trial 17 with value:
0.6315640925127401.
[I 2021-12-13 18:05:01,687] Trial 22 finished with value: 0.6315484123872991 and parameters: {'solver': 'sag', 'C': 0.22379859313737488}. Best is trial 17 with valu
e: 0.6315640925127401.
[I 2021-12-13 18:05:25,730] Trial 23 finished with value: 0.6315248921991375 and parameters: {'solver': 'sag', 'C': 0.22026105693489417}. Best is trial 17 with valu
e: 0.6315640925127401.
[I 2021-12-13 18:05:47,359] Trial 24 finished with value: 0.6315092120736966 and parameters: {'solver': 'sag', 'C': 0.2536171285242527}. Best is trial 17 with value:
0.6315640925127401.
[I 2021-12-13 18:06:10,405] Trial 25 finished with value: 0.6315484123872991 and parameters: {'solver': 'sag', 'C': 0.23775937954220988}. Best is trial 17 with valu
e: 0.6315640925127401.
[I 2021-12-13 18:06:35,660] Trial 26 finished with value: 0.6315248921991375 and parameters: {'solver': 'sag', 'C': 0.2175890470774522}. Best is trial 17 with value:
0.6315640925127401.
[I 2021-12-13 18:06:59,570] Trial 27 finished with value: 0.6315092120736966 and parameters: {'solver': 'sag', 'C': 0.24573443801061115}. Best is trial 17 with valu
e: 0.6315640925127401.
[I 2021-12-13 18:07:23,476] Trial 28 finished with value: 0.6315484123872991 and parameters: {'solver': 'sag', 'C': 0.2340819591755565}. Best is trial 17 with value:
0.6315640925127401.
[I 2021-12-13 18:07:47,802] Trial 29 finished with value: 0.6315719325754606 and parameters: {'solver': 'sag', 'C': 0.2360967624268411}. Best is trial 29 with value:
0.6315719325754606.
[I 2021-12-13 18:08:10,189] Trial 30 finished with value: 0.6315092120736966 and parameters: {'solver': 'sag', 'C': 0.25197453496046585}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:08:35,262] Trial 31 finished with value: 0.6315484123872991 and parameters: {'solver': 'sag', 'C': 0.23282664125706914}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:08:52,585] Trial 32 finished with value: 0.6315484123872991 and parameters: {'solver': 'liblinear', 'C': 0.2312379023723466}. Best is trial 29 with
 value: 0.6315719325754606.
[I 2021-12-13 18:09:09,639] Trial 33 finished with value: 0.6315327322618581 and parameters: {'solver': 'liblinear', 'C': 0.22844020304631202}. Best is trial 29 with
value: 0.6315719325754606.
[I 2021-12-13 18:09:32,935] Trial 34 finished with value: 0.6315092120736966 and parameters: {'solver': 'sag', 'C': 0.24489411551753681}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:09:49,699] Trial 35 finished with value: 0.6315484123872991 and parameters: {'solver': 'liblinear', 'C': 0.23157341024317407}. Best is trial 29 with
value: 0.6315719325754606.
[I 2021-12-13 18:10:06,543] Trial 36 finished with value: 0.6315327322618581 and parameters: {'solver': 'liblinear', 'C': 0.23601376815604988}. Best is trial 29 with
value: 0.6315719325754606.
[I 2021-12-13 18:10:22,526] Trial 37 finished with value: 0.6315248921991375 and parameters: {'solver': 'liblinear', 'C': 0.24723329622367238}. Best is trial 29 with
value: 0.6315719325754606.
[I 2021-12-13 18:10:47,445] Trial 38 finished with value: 0.6315248921991375 and parameters: {'solver': 'sag', 'C': 0.24206920159005874}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:10:58,955] Trial 39 finished with value: 0.6315092120736966 and parameters: {'solver': 'saga', 'C': 0.21477802999147932}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:11:11,393] Trial 40 finished with value: 0.6315013720109761 and parameters: {'solver': 'lbfgs', 'C': 0.2233535218428334}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:11:36,181] Trial 41 finished with value: 0.6315719325754606 and parameters: {'solver': 'sag', 'C': 0.23695225760055608}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:11:52,887] Trial 42 finished with value: 0.6315484123872991 and parameters: {'solver': 'liblinear', 'C': 0.23016013645390984}. Best is trial 29 with
value: 0.6315719325754606.
[I 2021-12-13 18:12:09,310] Trial 43 finished with value: 0.6315327322618581 and parameters: {'solver': 'liblinear', 'C': 0.22985042634483688}. Best is trial 29 with
value: 0.6315719325754606.
[I 2021-12-13 18:12:32,569] Trial 44 finished with value: 0.6315719325754606 and parameters: {'solver': 'sag', 'C': 0.23656835024485606}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:12:44,052] Trial 45 finished with value: 0.6315327322618581 and parameters: {'solver': 'saga', 'C': 0.2489927140039956}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:13:08,158] Trial 46 finished with value: 0.6315484123872991 and parameters: {'solver': 'sag', 'C': 0.22624251586718486}. Best is trial 29 with valu
e: 0.6315719325754606.
[I 2021-12-13 18:13:32,783] Trial 47 finished with value: 0.6315797726381811 and parameters: {'solver': 'sag', 'C': 0.23708588227231753}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:13:55,442] Trial 48 finished with value: 0.6315248921991375 and parameters: {'solver': 'sag', 'C': 0.24322517198264107}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:14:18,722] Trial 49 finished with value: 0.6315719325754606 and parameters: {'solver': 'sag', 'C': 0.2372060807103527}. Best is trial 47 with value:
0.6315797726381811.
[I 2021-12-13 18:14:42,986] Trial 50 finished with value: 0.6315797726381811 and parameters: {'solver': 'sag', 'C': 0.23664838991030696}. Best is trial 47 with valu
e: 0.6315797726381811.
```

```
[I 2021-12-13 18:15:07,265] Trial 51 finished with value: 0.6315640925127401 and parameters: {'solver': 'sag', 'C': 0.2354383883269074}. Best is trial 47 with value:
0.6315797726381811.
[I 2021-12-13 18:15:30,491] Trial 52 finished with value: 0.6315719325754606 and parameters: {'solver': 'sag', 'C': 0.23610408821306403}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:15:55,283] Trial 53 finished with value: 0.6315640925127401 and parameters: {'solver': 'sag', 'C': 0.2389454478740418}. Best is trial 47 with value:
0.6315797726381811.
[I 2021-12-13 18:16:09,191] Trial 54 finished with value: 0.6313367306938455 and parameters: {'solver': 'lbfgs', 'C': 0.25576384551400083}. Best is trial 47 with val
ue: 0.6315797726381811.
[I 2021-12-13 18:16:33,287] Trial 55 finished with value: 0.6315484123872991 and parameters: {'solver': 'sag', 'C': 0.22594710511617722}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:16:55,513] Trial 56 finished with value: 0.6315092120736966 and parameters: {'solver': 'sag', 'C': 0.24407676624418756}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:17:16,827] Trial 57 finished with value: 0.6315248921991375 and parameters: {'solver': 'sag', 'C': 0.24859897970991518}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:17:41,445] Trial 58 finished with value: 0.6315797726381811 and parameters: {'solver': 'sag', 'C': 0.23733509116013135}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:18:05,986] Trial 59 finished with value: 0.6315562524500196 and parameters: {'solver': 'sag', 'C': 0.24027629709865264}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:18:17,875] Trial 60 finished with value: 0.6315640925127401 and parameters: {'solver': 'saga', 'C': 0.23347016311973762}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:18:45,128] Trial 61 finished with value: 0.6315405723245786 and parameters: {'solver': 'sag', 'C': 0.23674017787132007}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:19:12,200] Trial 62 finished with value: 0.6315640925127401 and parameters: {'solver': 'sag', 'C': 0.23920829130306664}. Best is trial 47 with valu
e: 0.6315797726381811.
[I 2021-12-13 18:19:38,259] Trial 63 finished with value: 0.6315640925127401 and parameters: {'solver': 'sag', 'C': 0.23483146826742954}. Best is trial 47 with valu
e: 0.6315797726381811.
```

Comentários acerca dos resultados da Regressão Logística 2

Os resultados foram pouco distintos dos anteriores, dado que o algoritmo melhorou apenas 0.001%, ou seja, o teste mostrou que atingimos o seu melhor com os parâmetros anteriores.

Ensemble de MLP

In [4]:

```python
from numpy import mean
from numpy import std
from sklearn.datasets import make_classification
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedStratifiedKFold
from sklearn.neighbors import KNeighborsClassifier
from sklearn.ensemble import VotingClassifier
from matplotlib import pyplot
import scikitplot as skplt
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score, roc_auc_score, average_precision_score, plot_confusion_matrix
from sklearn.model_selection import cross_val_predict

ds_dropado = ds.drop(labels=['IND_BOM_1_2', 'IND_BOM_1_1'], axis=1)
df_treino = treino_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(treino_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao = valid_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(valid_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_validacao.drop(columns=['INDEX'], axis=1, inplace=True)
df_teste = teste_ina.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1).append(teste_adi.drop(['IND_BOM_1_2','IND_BOM_1_1'], axis=1))
df_teste.drop(columns=['INDEX'], axis=1, inplace=True)

def get_dataset():
        X, y = df_treino, [0]*len(treino_ina)+[1]*len(treino_adi)
        return X, y

def get_voting():
        models = list()
        models.append(('mlp0', MLPClassifier(verbose=True, max_iter=10000, early_stopping=True, hidden_layer_sizes=(13), solver='lbfgs', learning_rate='constant', ac
        models.append(('mlp1', MLPClassifier(verbose=True, max_iter=10000, early_stopping=True, hidden_layer_sizes=(3,), solver='sgd', learning_rate='adaptive', acti
        models.append(('mlp2', MLPClassifier(verbose=True, max_iter=10000, early_stopping=True, hidden_layer_sizes=(5,), solver='sgd', learning_rate='constant', acti
        ensemble = VotingClassifier(verbose=True, estimators=models, voting='soft').fit(get_dataset()[0], get_dataset()[1])
        return ensemble

def get_models():
        models = dict()
        models['mlp0'] = MLPClassifier(max_iter=10000, early_stopping=True)
        models['mlp1'] = MLPClassifier(max_iter=10000, early_stopping=True)
        models['mlp2'] = MLPClassifier(max_iter=10000, early_stopping=True)
        models['soft_voting'] = get_voting()
        return models


def compute_performance_metrics_sem_plot2(y, y_pred_class, y_pred_scores, rede_trial):
    accuracy = accuracy_score(y, y_pred_class)
    recall = recall_score(y, y_pred_class)
    precision = precision_score(y, y_pred_class)
    f1 = f1_score(y, y_pred_class)
    performance_metrics = (accuracy, recall, precision, f1)
    if y_pred_scores is not None:
        skplt.metrics.plot_ks_statistic(y, y_pred_scores)
        # plt.show()
        y_pred_scores = y_pred_scores[:, 1]
        auroc = roc_auc_score(y, y_pred_scores)
        aupr = average_precision_score(y, y_pred_scores)
        performance_metrics = performance_metrics + (auroc, aupr)
        plt.title(label=rede_trial, y=0.9)
        plt.suptitle('Acurácia: {:3.3f}\nRecall: {:3.3f}\nPrecision: {:3.3f}\nF1: {:3.3f}\nAUROC: {:3.3f}\nAURP: {:3.3f}'.format(accuracy, recall, precision, f1, aur
        plt.savefig(rede_trial, dpi=100)
        plt.close()
    return performance_metrics

def evaluate_model(model, X, y):
        cv = RepeatedStratifiedKFold(n_splits=10, n_repeats=3, random_state=1)
        scores = cross_val_predict(model, X, y, cv=cv, n_jobs=-1, method='predict')
        scores_proba = cross_val_predict(model, X, y, cv=cv, n_jobs=-1, method='predict_proba')
        compute_performance_metrics_sem_plot2(y, scores, scores_proba, 'Ensamble/Ensamble')
        return scores

# Dataset de treino
X, y = get_dataset()

# Modelos de ensamble
models = get_models()

# Avaliar cada modelo e armazenar seus resultados
results, names = list(), list()

pyplot.boxplot(evaluate_model(get_voting(), ds.drop(axis=1, labels=['IND_BOM_1_1', 'IND_BOM_1_2']), ds['IND_BOM_1_1']), labels=names, showmeans=True)
pyplot.show()
```

```
[Voting] ..................... (1 of 3) Processing mlp0, total= 5.5min
Iteration 1, loss = 0.65859254
Validation score: 0.627871
Iteration 2, loss = 0.64850178
Validation score: 0.610937
Iteration 3, loss = 0.64736500
Validation score: 0.630184
Iteration 4, loss = 0.64552217
Validation score: 0.612544
Iteration 5, loss = 0.64443723
Validation score: 0.630341
Iteration 6, loss = 0.64354886
Validation score: 0.615876
Iteration 7, loss = 0.64290993
Validation score: 0.623167
```

```
Iteration 8, loss = 0.64191137
Validation score: 0.620306
Iteration 9, loss = 0.64178602
Validation score: 0.630459
Iteration 10, loss = 0.64099773
Validation score: 0.630694
Iteration 11, loss = 0.64020231
Validation score: 0.633007
Iteration 12, loss = 0.63987681
Validation score: 0.632654
Iteration 13, loss = 0.63906045
Validation score: 0.632419
Iteration 14, loss = 0.63838819
Validation score: 0.633987
Iteration 15, loss = 0.63788467
Validation score: 0.635947
Iteration 16, loss = 0.63764156
Validation score: 0.635751
Iteration 17, loss = 0.63718784
Validation score: 0.635163
Iteration 18, loss = 0.63702467
Validation score: 0.636652
Iteration 19, loss = 0.63649882
Validation score: 0.637162
Iteration 20, loss = 0.63606402
Validation score: 0.636260
Iteration 21, loss = 0.63606826
Validation score: 0.638299
Iteration 22, loss = 0.63557745
Validation score: 0.637436
Iteration 23, loss = 0.63544389
Validation score: 0.637593
Iteration 24, loss = 0.63519591
Validation score: 0.637907
Iteration 25, loss = 0.63495646
Validation score: 0.639945
Iteration 26, loss = 0.63475987
Validation score: 0.634614
Iteration 27, loss = 0.63441440
Validation score: 0.637358
Iteration 28, loss = 0.63430366
Validation score: 0.638965
Iteration 29, loss = 0.63432867
Validation score: 0.638338
Iteration 30, loss = 0.63407835
Validation score: 0.640416
Iteration 31, loss = 0.63383961
Validation score: 0.639396
Iteration 32, loss = 0.63387912
Validation score: 0.638377
Iteration 33, loss = 0.63358630
Validation score: 0.636574
Iteration 34, loss = 0.63363580
Validation score: 0.639044
Iteration 35, loss = 0.63374128
Validation score: 0.639240
Iteration 36, loss = 0.63368041
Validation score: 0.639828
Iteration 37, loss = 0.63339755
Validation score: 0.638495
Iteration 38, loss = 0.63364066
Validation score: 0.638573
Iteration 39, loss = 0.63335142
Validation score: 0.639436
Iteration 40, loss = 0.63327994
Validation score: 0.638965
Iteration 41, loss = 0.63319188
Validation score: 0.640416
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.011737
Iteration 42, loss = 0.63088604
Validation score: 0.641004
Iteration 43, loss = 0.63070783
Validation score: 0.638887
Iteration 44, loss = 0.63058488
Validation score: 0.637436
Iteration 45, loss = 0.63056632
Validation score: 0.640886
Iteration 46, loss = 0.63058969
Validation score: 0.637946
Iteration 47, loss = 0.63058279
Validation score: 0.638338
Iteration 48, loss = 0.63048137
Validation score: 0.639867
Iteration 49, loss = 0.63048117
Validation score: 0.641278
Iteration 50, loss = 0.63053991
Validation score: 0.638573
Iteration 51, loss = 0.63051911
Validation score: 0.639828
Iteration 52, loss = 0.63044897
Validation score: 0.638299
Iteration 53, loss = 0.63038055
Validation score: 0.640376
Iteration 54, loss = 0.63044285
Validation score: 0.638887
Iteration 55, loss = 0.63032460
Validation score: 0.639710
Iteration 56, loss = 0.63033801
Validation score: 0.637593
Iteration 57, loss = 0.63033046
Validation score: 0.640024
Iteration 58, loss = 0.63037905
Validation score: 0.638887
Iteration 59, loss = 0.63023855
Validation score: 0.639828
Iteration 60, loss = 0.63031291
Validation score: 0.640494
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.002347
Iteration 61, loss = 0.62950877
Validation score: 0.639396
Iteration 62, loss = 0.62949857
Validation score: 0.639161
Iteration 63, loss = 0.62950009
Validation score: 0.639200
Iteration 64, loss = 0.62950112
Validation score: 0.639867
Iteration 65, loss = 0.62946247
Validation score: 0.640102
Iteration 66, loss = 0.62944914
Validation score: 0.639122
Iteration 67, loss = 0.62946002
Validation score: 0.639475
Iteration 68, loss = 0.62943881
Validation score: 0.640612
Iteration 69, loss = 0.62945152
Validation score: 0.639632
Iteration 70, loss = 0.62946220
Validation score: 0.640220
```

```
Iteration 71, loss = 0.62946774
Validation score: 0.639200
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000469
Iteration 72, loss = 0.62926121
Validation score: 0.640141
Iteration 73, loss = 0.62923734
Validation score: 0.640651
Iteration 74, loss = 0.62925218
Validation score: 0.639318
Iteration 75, loss = 0.62925183
Validation score: 0.639632
Iteration 76, loss = 0.62924657
Validation score: 0.639475
Iteration 77, loss = 0.62923922
Validation score: 0.639004
Iteration 78, loss = 0.62923556
Validation score: 0.639357
Iteration 79, loss = 0.62924380
Validation score: 0.639200
Iteration 80, loss = 0.62922735
Validation score: 0.639632
Iteration 81, loss = 0.62923839
Validation score: 0.639553
Iteration 82, loss = 0.62924091
Validation score: 0.639436
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000094
Iteration 83, loss = 0.62919729
Validation score: 0.639436
Iteration 84, loss = 0.62919330
Validation score: 0.639279
Iteration 85, loss = 0.62919404
Validation score: 0.639318
Iteration 86, loss = 0.62919359
Validation score: 0.639632
Iteration 87, loss = 0.62919304
Validation score: 0.639632
Iteration 88, loss = 0.62919404
Validation score: 0.639592
Iteration 89, loss = 0.62919433
Validation score: 0.639671
Iteration 90, loss = 0.62919442
Validation score: 0.639396
Iteration 91, loss = 0.62919088
Validation score: 0.639592
Iteration 92, loss = 0.62919317
Validation score: 0.639475
Iteration 93, loss = 0.62919261
Validation score: 0.639592
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000019
Iteration 94, loss = 0.62918210
Validation score: 0.639514
Iteration 95, loss = 0.62918251
Validation score: 0.639514
Iteration 96, loss = 0.62918197
Validation score: 0.639514
Iteration 97, loss = 0.62918214
Validation score: 0.639436
Iteration 98, loss = 0.62918225
Validation score: 0.639357
Iteration 99, loss = 0.62918163
Validation score: 0.639514
Iteration 100, loss = 0.62918187
Validation score: 0.639436
Iteration 101, loss = 0.62918215
Validation score: 0.639514
Iteration 102, loss = 0.62918184
Validation score: 0.639514
Iteration 103, loss = 0.62918171
Validation score: 0.639436
Iteration 104, loss = 0.62918158
Validation score: 0.639357
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000004
Iteration 105, loss = 0.62917949
Validation score: 0.639436
Iteration 106, loss = 0.62917945
Validation score: 0.639396
Iteration 107, loss = 0.62917942
Validation score: 0.639357
Iteration 108, loss = 0.62917941
Validation score: 0.639357
Iteration 109, loss = 0.62917946
Validation score: 0.639357
Iteration 110, loss = 0.62917939
Validation score: 0.639357
Iteration 111, loss = 0.62917939
Validation score: 0.639279
Iteration 112, loss = 0.62917937
Validation score: 0.639318
Iteration 113, loss = 0.62917940
Validation score: 0.639318
Iteration 114, loss = 0.62917939
Validation score: 0.639318
Iteration 115, loss = 0.62917945
Validation score: 0.639318
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000001
Iteration 116, loss = 0.62917890
Validation score: 0.639279
Iteration 117, loss = 0.62917890
Validation score: 0.639279
Iteration 118, loss = 0.62917889
Validation score: 0.639279
Iteration 119, loss = 0.62917889
Validation score: 0.639279
Iteration 120, loss = 0.62917889
Validation score: 0.639279
Iteration 121, loss = 0.62917888
Validation score: 0.639279
Iteration 122, loss = 0.62917888
Validation score: 0.639279
Iteration 123, loss = 0.62917888
Validation score: 0.639279
Iteration 124, loss = 0.62917887
Validation score: 0.639279
Iteration 125, loss = 0.62917888
Validation score: 0.639279
Iteration 126, loss = 0.62917887
Validation score: 0.639279
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Learning rate too small. Stopping.
[Voting] ..................... (2 of 3) Processing mlp1, total= 1.0min
Iteration 1, loss = 0.65874036
Validation score: 0.628185
Iteration 2, loss = 0.64593450
Validation score: 0.630419
Iteration 3, loss = 0.64272669
Validation score: 0.633791
Iteration 4, loss = 0.64038335
Validation score: 0.633438
Iteration 5, loss = 0.63855599
```

```
Validation score: 0.634379
Iteration 6, loss = 0.63713540
Validation score: 0.636417
Iteration 7, loss = 0.63587003
Validation score: 0.638965
Iteration 8, loss = 0.63523851
Validation score: 0.639279
Iteration 9, loss = 0.63459328
Validation score: 0.638769
Iteration 10, loss = 0.63398924
Validation score: 0.638769
Iteration 11, loss = 0.63355689
Validation score: 0.640298
Iteration 12, loss = 0.63289817
Validation score: 0.639788
Iteration 13, loss = 0.63276923
Validation score: 0.637123
Iteration 14, loss = 0.63234321
Validation score: 0.639945
Iteration 15, loss = 0.63210204
Validation score: 0.640024
Iteration 16, loss = 0.63189010
Validation score: 0.640455
Iteration 17, loss = 0.63164956
Validation score: 0.639553
Iteration 18, loss = 0.63127024
Validation score: 0.637358
Iteration 19, loss = 0.63107236
Validation score: 0.638534
Iteration 20, loss = 0.63097774
Validation score: 0.639475
Iteration 21, loss = 0.63085636
Validation score: 0.637828
Iteration 22, loss = 0.63057968
Validation score: 0.640572
Iteration 23, loss = 0.63053893
Validation score: 0.640376
Iteration 24, loss = 0.63038481
Validation score: 0.640180
Iteration 25, loss = 0.63016074
Validation score: 0.640455
Iteration 26, loss = 0.63006402
Validation score: 0.639279
Iteration 27, loss = 0.63003233
Validation score: 0.639357
Iteration 28, loss = 0.62990412
Validation score: 0.638652
Iteration 29, loss = 0.62987235
Validation score: 0.641631
Iteration 30, loss = 0.62956052
Validation score: 0.639828
Iteration 31, loss = 0.62955315
Validation score: 0.641239
Iteration 32, loss = 0.62936805
Validation score: 0.639906
Iteration 33, loss = 0.62938160
Validation score: 0.638652
Iteration 34, loss = 0.62924181
Validation score: 0.639553
Iteration 35, loss = 0.62918646
Validation score: 0.640063
Iteration 36, loss = 0.62920010
Validation score: 0.642101
Iteration 37, loss = 0.62910583
Validation score: 0.641356
Iteration 38, loss = 0.62888516
Validation score: 0.641474
Iteration 39, loss = 0.62877433
Validation score: 0.639710
Iteration 40, loss = 0.62877448
Validation score: 0.640141
Iteration 41, loss = 0.62883059
Validation score: 0.638848
Iteration 42, loss = 0.62864917
Validation score: 0.641631
Iteration 43, loss = 0.62865143
Validation score: 0.641827
Iteration 44, loss = 0.62857450
Validation score: 0.639945
Iteration 45, loss = 0.62848227
Validation score: 0.640768
Iteration 46, loss = 0.62843994
Validation score: 0.641121
Iteration 47, loss = 0.62854992
Validation score: 0.640768
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Stopping.
[Voting] ................... (3 of 3) Processing mlp2, total=  24.5s
[Voting] ................... (1 of 3) Processing mlp0, total= 5.2min
Iteration 1, loss = 0.65061413
Validation score: 0.630655
Iteration 2, loss = 0.64105214
Validation score: 0.616190
Iteration 3, loss = 0.63792377
Validation score: 0.632536
Iteration 4, loss = 0.63648794
Validation score: 0.632536
Iteration 5, loss = 0.63531031
Validation score: 0.633987
Iteration 6, loss = 0.63501860
Validation score: 0.631282
Iteration 7, loss = 0.63447364
Validation score: 0.632967
Iteration 8, loss = 0.63420964
Validation score: 0.635868
Iteration 9, loss = 0.63402057
Validation score: 0.634339
Iteration 10, loss = 0.63384759
Validation score: 0.627911
Iteration 11, loss = 0.63364203
Validation score: 0.633163
Iteration 12, loss = 0.63347399
Validation score: 0.634222
Iteration 13, loss = 0.63337491
Validation score: 0.634927
Iteration 14, loss = 0.63334128
Validation score: 0.637123
Iteration 15, loss = 0.63321087
Validation score: 0.636770
Iteration 16, loss = 0.63307627
Validation score: 0.631870
Iteration 17, loss = 0.63317692
Validation score: 0.629087
Iteration 18, loss = 0.63280983
Validation score: 0.634143
Iteration 19, loss = 0.63267721
Validation score: 0.637475
Iteration 20, loss = 0.63276467
Validation score: 0.634888
```

```
Iteration 21, loss = 0.63279217
Validation score: 0.634927
Iteration 22, loss = 0.63274359
Validation score: 0.633046
Iteration 23, loss = 0.63260057
Validation score: 0.635515
Iteration 24, loss = 0.63250837
Validation score: 0.633438
Iteration 25, loss = 0.63248891
Validation score: 0.631948
Iteration 26, loss = 0.63253837
Validation score: 0.635241
Iteration 27, loss = 0.63243423
Validation score: 0.635555
Iteration 28, loss = 0.63248129
Validation score: 0.634692
Iteration 29, loss = 0.63234613
Validation score: 0.637475
Iteration 30, loss = 0.63225407
Validation score: 0.635163
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.011737
Iteration 31, loss = 0.63004894
Validation score: 0.632967
Iteration 32, loss = 0.62979048
Validation score: 0.636182
Iteration 33, loss = 0.62971191
Validation score: 0.635947
Iteration 34, loss = 0.62965640
Validation score: 0.635319
Iteration 35, loss = 0.62980506
Validation score: 0.636652
Iteration 36, loss = 0.62958131
Validation score: 0.636378
Iteration 37, loss = 0.62958892
Validation score: 0.633634
Iteration 38, loss = 0.62962837
Validation score: 0.636025
Iteration 39, loss = 0.62955482
Validation score: 0.636025
Iteration 40, loss = 0.62952462
Validation score: 0.636378
Iteration 41, loss = 0.62954064
Validation score: 0.636417
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.002347
Iteration 42, loss = 0.62881666
Validation score: 0.635084
Iteration 43, loss = 0.62874961
Validation score: 0.637201
Iteration 44, loss = 0.62876921
Validation score: 0.637162
Iteration 45, loss = 0.62874752
Validation score: 0.635123
Iteration 46, loss = 0.62874847
Validation score: 0.636535
Iteration 47, loss = 0.62872245
Validation score: 0.637672
Iteration 48, loss = 0.62872585
Validation score: 0.636182
Iteration 49, loss = 0.62870224
Validation score: 0.636770
Iteration 50, loss = 0.62870830
Validation score: 0.636966
Iteration 51, loss = 0.62869033
Validation score: 0.635986
Iteration 52, loss = 0.62868779
Validation score: 0.637005
Iteration 53, loss = 0.62867082
Validation score: 0.635515
Iteration 54, loss = 0.62868147
Validation score: 0.636535
Iteration 55, loss = 0.62869993
Validation score: 0.636652
Iteration 56, loss = 0.62870455
Validation score: 0.635241
Iteration 57, loss = 0.62871333
Validation score: 0.636809
Iteration 58, loss = 0.62867150
Validation score: 0.636848
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000469
Iteration 59, loss = 0.62850356
Validation score: 0.636417
Iteration 60, loss = 0.62848692
Validation score: 0.636848
Iteration 61, loss = 0.62848992
Validation score: 0.636770
Iteration 62, loss = 0.62848032
Validation score: 0.636691
Iteration 63, loss = 0.62848520
Validation score: 0.637475
Iteration 64, loss = 0.62847999
Validation score: 0.636966
Iteration 65, loss = 0.62848233
Validation score: 0.637436
Iteration 66, loss = 0.62847731
Validation score: 0.636574
Iteration 67, loss = 0.62848919
Validation score: 0.637319
Iteration 68, loss = 0.62847670
Validation score: 0.637358
Iteration 69, loss = 0.62847639
Validation score: 0.637397
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000094
Iteration 70, loss = 0.62843565
Validation score: 0.637123
Iteration 71, loss = 0.62843205
Validation score: 0.637240
Iteration 72, loss = 0.62843303
Validation score: 0.637397
Iteration 73, loss = 0.62843293
Validation score: 0.637397
Iteration 74, loss = 0.62843096
Validation score: 0.637358
Iteration 75, loss = 0.62843141
Validation score: 0.637279
Iteration 76, loss = 0.62843019
Validation score: 0.637319
Iteration 77, loss = 0.62843179
Validation score: 0.637123
Iteration 78, loss = 0.62842975
Validation score: 0.637358
Iteration 79, loss = 0.62843183
Validation score: 0.637397
Iteration 80, loss = 0.62843076
Validation score: 0.637358
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000019
Iteration 81, loss = 0.62842078
Validation score: 0.637358
Iteration 82, loss = 0.62842026
```

```
Validation score: 0.637279
Iteration 83, loss = 0.62842020
Validation score: 0.637240
Iteration 84, loss = 0.62842039
Validation score: 0.637279
Iteration 85, loss = 0.62842038
Validation score: 0.637279
Iteration 86, loss = 0.62842039
Validation score: 0.637319
Iteration 87, loss = 0.62842017
Validation score: 0.637319
Iteration 88, loss = 0.62842027
Validation score: 0.637319
Iteration 89, loss = 0.62842020
Validation score: 0.637319
Iteration 90, loss = 0.62842027
Validation score: 0.637319
Iteration 91, loss = 0.62842017
Validation score: 0.637319
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000004
Iteration 92, loss = 0.62841778
Validation score: 0.637319
Iteration 93, loss = 0.62841780
Validation score: 0.637319
Iteration 94, loss = 0.62841781
Validation score: 0.637279
Iteration 95, loss = 0.62841783
Validation score: 0.637319
Iteration 96, loss = 0.62841774
Validation score: 0.637319
Iteration 97, loss = 0.62841780
Validation score: 0.637319
Iteration 98, loss = 0.62841775
Validation score: 0.637319
Iteration 99, loss = 0.62841771
Validation score: 0.637319
Iteration 100, loss = 0.62841772
Validation score: 0.637319
Iteration 101, loss = 0.62841769
Validation score: 0.637319
Iteration 102, loss = 0.62841770
Validation score: 0.637319
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Setting learning rate to 0.000001
Iteration 103, loss = 0.62841725
Validation score: 0.637319
Iteration 104, loss = 0.62841725
Validation score: 0.637319
Iteration 105, loss = 0.62841724
Validation score: 0.637319
Iteration 106, loss = 0.62841725
Validation score: 0.637319
Iteration 107, loss = 0.62841725
Validation score: 0.637319
Iteration 108, loss = 0.62841723
Validation score: 0.637319
Iteration 109, loss = 0.62841724
Validation score: 0.637319
Iteration 110, loss = 0.62841723
Validation score: 0.637319
Iteration 111, loss = 0.62841723
Validation score: 0.637319
Iteration 112, loss = 0.62841724
Validation score: 0.637319
Iteration 113, loss = 0.62841723
Validation score: 0.637319
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Learning rate too small. Stopping.
[Voting] ..................... (2 of 3) Processing mlp1, total=  55.1s
Iteration 1, loss = 0.65815106
Validation score: 0.627087
Iteration 2, loss = 0.64655583
Validation score: 0.630263
Iteration 3, loss = 0.64419775
Validation score: 0.630968
Iteration 4, loss = 0.64220745
Validation score: 0.633791
Iteration 5, loss = 0.64030094
Validation score: 0.633438
Iteration 6, loss = 0.63849581
Validation score: 0.636025
Iteration 7, loss = 0.63688180
Validation score: 0.637123
Iteration 8, loss = 0.63570633
Validation score: 0.637907
Iteration 9, loss = 0.63474405
Validation score: 0.637475
Iteration 10, loss = 0.63402728
Validation score: 0.637750
Iteration 11, loss = 0.63322733
Validation score: 0.636613
Iteration 12, loss = 0.63277643
Validation score: 0.638260
Iteration 13, loss = 0.63264798
Validation score: 0.637946
Iteration 14, loss = 0.63225570
Validation score: 0.635555
Iteration 15, loss = 0.63186203
Validation score: 0.637828
Iteration 16, loss = 0.63167257
Validation score: 0.637789
Iteration 17, loss = 0.63128129
Validation score: 0.639749
Iteration 18, loss = 0.63111886
Validation score: 0.639161
Iteration 19, loss = 0.63110669
Validation score: 0.641396
Iteration 20, loss = 0.63085328
Validation score: 0.639357
Iteration 21, loss = 0.63076563
Validation score: 0.639396
Iteration 22, loss = 0.63055059
Validation score: 0.639710
Iteration 23, loss = 0.63042406
Validation score: 0.640024
Iteration 24, loss = 0.63018616
Validation score: 0.641513
Iteration 25, loss = 0.63019343
Validation score: 0.639044
Iteration 26, loss = 0.63017958
Validation score: 0.640063
Iteration 27, loss = 0.63003342
Validation score: 0.637868
Iteration 28, loss = 0.62985011
Validation score: 0.640651
Iteration 29, loss = 0.62984504
Validation score: 0.638848
Iteration 30, loss = 0.62967966
Validation score: 0.642454
Iteration 31, loss = 0.62959930
```

```
Validation score: 0.640259
Iteration 32, loss = 0.62959973
Validation score: 0.639710
Iteration 33, loss = 0.62947148
Validation score: 0.640533
Iteration 34, loss = 0.62949737
Validation score: 0.638887
Iteration 35, loss = 0.62959642
Validation score: 0.637358
Iteration 36, loss = 0.62947785
Validation score: 0.641239
Iteration 37, loss = 0.62926852
Validation score: 0.640337
Iteration 38, loss = 0.62912497
Validation score: 0.640612
Iteration 39, loss = 0.62917163
Validation score: 0.638181
Iteration 40, loss = 0.62913082
Validation score: 0.638142
Iteration 41, loss = 0.62903024
Validation score: 0.638808
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Stopping.
[Voting] ..................... (3 of 3) Processing mlp2, total=  21.0s
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-4-0cc2ceb51439> in <module>
     75 results, names = list(), list()
     76 
---> 77 pyplot.boxplot(evaluate_model(get_voting(), ds.drop(axis=1, labels=['IND_BOM_1_1', 'IND_BOM_1_2']), ds['IND_BOM_1_1']), labels=names, showmeans=True)
     78 pyplot.show()

<ipython-input-4-0cc2ceb51439> in evaluate_model(model, X, y)
     60 def evaluate_model(model, X, y):
     61         cv = RepeatedStratifiedKFold(n_splits=10, n_repeats=3, random_state=1)
---> 62         scores = cross_val_predict(model, X, y, cv=cv, n_jobs=-1, method='predict')
     63         scores_proba = cross_val_predict(model, X, y, cv=cv, n_jobs=-1, method='predict_proba')
     64         compute_performance_metrics_sem_plot2(y, scores, scores_proba, 'Ensamble/Ensamble')

~/.local/lib/python3.8/site-packages/sklearn/utils/validation.py in inner_f(*args, **kwargs)
     61             extra_args = len(args) - len(all_args)
     62             if extra_args <= 0:
---> 63                 return f(*args, **kwargs)
     64 
     65             # extra_args > 0

~/.local/lib/python3.8/site-packages/sklearn/model_selection/_validation.py in cross_val_predict(estimator, X, y, groups, cv, n_jobs, verbose, fit_params, pre_dispatch, method)
    843     test_indices = np.concatenate([test for _, test in splits])
    844     if not _check_is_permutation(test_indices, _num_samples(X)):
--> 845         raise ValueError('cross_val_predict only works for partitions')
    846 
    847     # If classification methods produce multiple columns of output,

ValueError: cross_val_predict only works for partitions
```

Utilizando o sistema de Voting Classifier com 3 MLPs, obtivemos o resultado de validação médio: 63,88%.