Team 31

Κανελλόπουλος Σωτήριος (03117101)

Καραντώνης Αντώνιος (03117439)

Κυριακίδης Δημήτριος (03117077)

Νευρωνικά Δίκτυα

1η Εργασία: Επιβλεπόμενη Μάθηση - Ταξινόμηση

Μελέτη datasets του αποθετηρίου UCI και της πλατφόρμας Kaggle

2ο Μέρος: Kaggle

Το dataset που μας ανατέθηκε είναι το K10 ("Hotel Booking"), το οποίο περιλαμβάνει πληροφορίες σχετικά με κρατήσεις σε δύο ξενοδοχεία, όπως την προγραμματισμένη ημερομηνία άφιξης και την αντίστοιχη διάρκεια παραμονής. Σκοπός της εργασίας μας θα είναι να αποφανθούμε με βάση τα δεδομένα αυτά εάν κάποιος πελάτης θα ακυρώσει την κράτησή του ή όχι. Για τον σκοπό αυτό θα θεωρήσουμε ως label την στήλη is_canceled, η οποία λαμβάνει δυαδική τιμή.

Από τα δεδομένα που μας παρέχει το Kaggle παρατηρούμε τα εξής:

- Το dataset αποτελείται συνολικά απο 119390 δείγματα.
- Εκτός της στήλης is_canceled περιέχει άλλες 35, τις οποίες θα θεωρήσουμε ως features.
- Το dataset είναι σχετικά imbalanced, δηλαδή η κλάση 1 (που αντιστοιχεί σε ακυρωμένη κράτηση) περιέχεται σε ποσοστό 37%.
- Υπάρχουν δύο features με απουσιάζουσες τιμές:
  - Το agent, το οποίο αφορά το ταξιδιωτικό πρακτορείο που έκανε την κράτηση και απουσιάζει στο 14% των δειγμάτων.
  - Το company, το οποίο αφορά την εταιρεία που έκανε την κράτηση ή είναι υπεύθυνη για την πληρωμή και απουσιάζει στο 94% των δειγμάτων.

In [1]:
```
import pandas as pd
import glob
import os
import warnings
```

In [2]:
```
path = r"C:\Users\sotka\Desktop\hotel_booking.csv"
#path = r"C:\Users\anton\OneDrive\Desktop\hotel_booking.csv"
#path = r"C:\Users\KYRIAKOS\Documents\JPNotebooks\NN\Assignment_1\Hotel_Booking\hotel_booking.csv"

df = pd.read_csv(path) ## dataframe
```

Δεν είναι όλα τα features του dataset χρήσιμα για την εξαγωγή του επιθυμητού συμπεράσματος. Συγκεκριμένα επιλέξαμε να αφαιρέσουμε τις εξής στήλες:

- Τα name, email, phone-number και credit_card, επειδή είναι αδιάφορα για την εξαγωγή του συμπεράσματός μας.
- Το company, επειδή όπως αναφέρθηκε παραπάνω περιέχεται μόνο στο 6% των δειγμάτων.
- Το reservation_status και το reservation_status_date, διότι πρακτικά περιέχουν την ίδια πληροφορία με το is_canceled, για το οποίο θέλουμε να αποφανθούμε.
- Το arrival_date_week_number, επειδή είναι πλεονάζουσα πληροφορία. Μπορεί να εξαχθεί από τα arrival_date_month και arrival_date_day_of_month.

Μετά από αυτές τις αλλαγές θα έχουμε 27 features με τα οποία θα εργαστούμε.

In [3]:
```
# remove columns

df.drop(["name", "email", "phone-number", "credit_card", "company", "reservation_status", "reservation_status_date",
         "arrival_date_week_number"], axis = 1, inplace = True)
## axis 0: γραμμές, axis 1: στήλες
```

In [4]:
```
from sklearn.impute import SimpleImputer
import numpy as np
```

```
df.replace("nan", np.NaN, inplace = True)
simp = SimpleImputer(missing_values = np.NaN, strategy = "most_frequent")

simp_df = pd.DataFrame(simp.fit_transform(df))
simp_df.columns = df.columns
simp_df.index = df.index
```

Παραπάνω χρησιμοποιούμε τον Imputer για να γεμίσουμε τις απουσιάζουσες τιμές. Το μόνο πεδίο με απουσιάζουσες τιμές είναι το agent, καθώς το company αφαιρέθηκε όπως εξηγήσαμε πρωτύτερα. Δεδομένου ότι το agent ουσιαστικά είναι κατηγορικό χαρακτηριστικό (δηλαδή περιέχει αριθμό, ο οποίος όμως αναπαριστά το ID ενός πράκτορα), επιλέγουμε να στρατηγική "most_frequent" για το γέμισμα των τιμών.

In [5]:
```
from sklearn.preprocessing import OneHotEncoder

categorical_features = ["hotel", "arrival_date_month", "meal", "country", "market_segment", "distribution_channel",
                        "reserved_room_type", "assigned_room_type", "deposit_type", "customer_type", "agent"]

## "agent", was, in fact, completely dropped, as it was a categorical feature consisting of too many arithmetic values

oh_encoder = OneHotEncoder()

enc_df = pd.DataFrame(oh_encoder.fit_transform(simp_df[categorical_features]).toarray())
final_df = simp_df.join(enc_df)
```

Στο παραπάνω κελί κώδικα χρησιμοποιούμε το OneHotEncoder για την διαχείριση των κατηγορικών χαρακτηριστικών, παράγοντας για καθένα από αυτά dummy στήλες ίσες σε πλήθος με τις δυνατές τιμές του. Έπειτα, θα αφαιρέσουμε τα αρχικά χαρακτηριστικά από το dataframe μας.

Εξαιτίας της παρουσίας ορισμένων κατηγορικών χαρακτηριστικών με μεγάλο πεδίο τιμών (όπως το country), η ανωτέρω τεχνική αυξάνει αρκετά τις διαστάσεις των δεδομένων. Προκειμένου να αποφευχθεί αυτό θα μπορούσαμε να είχαμε επιλέξει άλλη μέθοδο, που να διατηρεί τις ίδιες διαστάσεις, όπως target encoding. Ωστόσο, παρατηρώντας τις κατανομές των τιμών των διαφόρων κατηγορικών χαρακτηριστικών, συμπεράναμε πως οι περισσότερες dummy στήλες επρόκειτο να αφαιρεθούν από τις μεθόδους dimensionality reduction στα επόμενα βήματα, καθιστώντας τα δεδομένα διαχειρίσιμα.

In [6]:
```
final_df.drop(categorical_features, axis = 1, inplace = True)
```

In [8]:
```
## final_df.to_csv(r"C:\Users\anton\OneDrive\Desktop\test5.csv", encoding = "utf-8")
## final_df.to_csv(r"C:\Users\KYRIAKOS\Documents\JPNotebooks\NN\Assignment_1\test5.csv", encoding = "utf-8")
```

Επιλέγουμε test size 20% επειδή το dataset είναι αρκετά μεγάλο.

In [9]:
```
from sklearn.model_selection import train_test_split

features_df = final_df.drop("is_canceled", axis = 1)
labels_series = final_df["is_canceled"]

features = features_df.values
labels = labels_series.values.reshape(119390,)

train_data, test_data, train_labels, test_labels = train_test_split(features, labels, test_size = 0.2)
## of course used random state for reproducible results

train_labels = train_labels.astype(int)
test_labels = test_labels.astype(int)
```

Out of the box

Ως κριτήριο για να βελτιστοποιήσουμε και να αξιολογήσουμε τους ταξινομητές μας θα χρησιμοποιήσουμε την μετρική f1, η οποία είναι πιο κατάλληλη για προβλήματα με imbalanced κλάσεις, όπως το δικό μας. Επιπλέον, αφού το f1 αποτελεί αρμονικό μέσο του precision και του recall, είναι αρκετά καλό για να χρησιμοποιηθεί από μόνο του ως κριτήριο, ενώ αν χρησιμοποιούσαμε κάποια άλλη μετρική (π.χ. precision) μάλλον θα χρειαζόμασταν και μια δεύτερη για να εξασφαλίσουμε την ποιότητα των αποτελεσμάτων μας. Αυτό δεν θα ήταν επιθυμητό, δεδομένου ότι το συγκεκριμένο dataset είναι πολύ μεγάλο και η βελτιστοποίηση ως προς δεύτερη μετρική θα ήταν πολύ χρονοβόρα.

Σημειώνουμε ότι θα προτιμήσουμε το f1 για την κλάση 1, η οποία συναντάται σε μικρότερο ποσοστό στο dataset μας, ούτως ώστε να ελέγχουμε πιο αυστηρά την απόδοση του classifier.

In [10]:
```
ootb_f1scores = {}
opt_f1scores = {}
ootb_train_times = {}
ootb_test_times = {}
opt_train_times = {}
opt_test_times = {}
```

Dummy

In [11]:
```
## https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html
from sklearn.dummy import DummyClassifier
from sklearn.metrics import classification_report, f1_score
from time import perf_counter

## “stratified”: generates predictions by respecting the training set’s class distribution
dc_stratified = DummyClassifier(strategy = "stratified")

start = perf_counter()

model = dc_stratified.fit(train_data, train_labels)

end = perf_counter()
ootb_train_times.update({"Dummy" : end - start})

start = perf_counter()

dc_pred = dc_stratified.predict(test_data)

end = perf_counter()
ootb_test_times.update({"Dummy" : end - start})

dc_f1 = f1_score(test_labels, dc_pred) # returns f1 for class 1 by default
ootb_f1scores.update({"Dummy" : dc_f1})
print(classification_report(test_labels, dc_pred))
print("Dummy f1 score:", dc_f1)
```

```
              precision    recall  f1-score   support

           0       0.63      0.63      0.63     14983
           1       0.38      0.37      0.38      8895

    accuracy                           0.54     23878
   macro avg       0.50      0.50      0.50     23878
weighted avg       0.54      0.54      0.54     23878

Dummy f1 score: 0.37552813925976
```

MLP

In [13]:
```
from sklearn.neural_network import MLPClassifier

mlp = MLPClassifier()

start = perf_counter()

model = mlp.fit(train_data, train_labels)

end = perf_counter()
ootb_train_times.update({"MLP" : end - start})

start = perf_counter()

mlp_pred = mlp.predict(test_data)

end = perf_counter()
ootb_test_times.update({"MLP" : end - start})

mlp_f1 = f1_score(test_labels, mlp_pred)
ootb_f1scores.update({"MLP" : mlp_f1})
print(classification_report(test_labels, mlp_pred))
print("MLP f1 score:", mlp_f1)
```

```
              precision    recall  f1-score   support

           0       0.79      0.97      0.87     14983
           1       0.93      0.56      0.70      8895
```

```
    accuracy                           0.82     23878
   macro avg       0.86      0.77      0.78     23878
weighted avg       0.84      0.82      0.81     23878

MLP f1 score: 0.6969228607559365
```

SVM

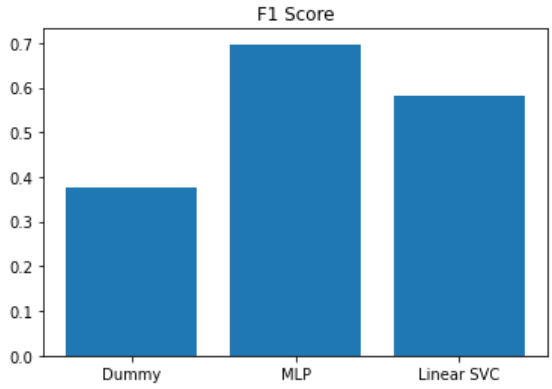
Δοκιμάσαμε τον out-of-the-box SVC και, αφού έτρεξε για περίπου μιάμιση ώρα, τον διακόψαμε γιατί ο χρόνος που απαιτεί θα ήταν εντελώς απαγορευτικός για τη βελτιστοποίηση. Ύστερα, πειραματιστήκαμε μειώνοντας το μήκος του train set και παρατηρήσαμε ότι τα αποτελέσματά του δεν ήταν καθόλου ικανοποιητικά για αρκετά μικρό μήκος (το f1 score του ήταν παρόμοιο με αυτό του dummy).

Για αυτούς τους λόγους στην εργασία **θα ασχοληθούμε με τον Linear SVC**, ο οποίος είναι πολύ ταχύτερος. Τα αποτελέσματά του (όπως φαίνεται παρακάτω) είναι αρκετά χειρότερα σε σύγκριση με τον MLP, ωστόσο με δοκιμές παρατηρήσαμε ότι αυξάνοντας τα max iterations βελτιώνεται αρκετά η επίδοσή του. Έτσι, αναμένουμε ότι ύστερα από τη βελτιστοποίηση θα έχουμε αρκετά καλύτερα αποτελέσματα.

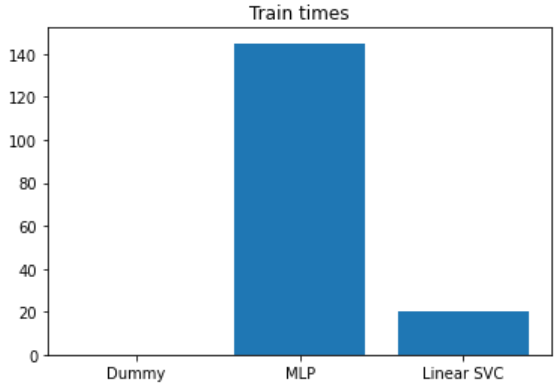
In [16]:
```
from sklearn.svm import LinearSVC

svm = LinearSVC()

start = perf_counter()

model = svm.fit(train_data, train_labels)

end = perf_counter()
ootb_train_times.update({"Linear SVC" : end - start})

start = perf_counter()

svm_pred = svm.predict(test_data)

end = perf_counter()
ootb_test_times.update({"Linear SVC" : end - start})

svm_f1 = f1_score(test_labels, svm_pred)
ootb_f1scores.update({"Linear SVC" : svm_f1})
print(classification_report(test_labels, svm_pred))
print("Linear SVC f1 score:", svm_f1)
```

```
              precision    recall  f1-score   support

           0       0.74      0.99      0.85     14983
           1       0.97      0.42      0.58      8895

    accuracy                           0.78     23878
   macro avg       0.85      0.70      0.71     23878
weighted avg       0.83      0.78      0.75     23878

Linear SVC f1 score: 0.5813148788927336
C:\Users\sotka\anaconda3\lib\site-packages\sklearn\svm\_base.py:976: ConvergenceWarning: Liblinear failed to converge, inc
rease the number of iterations.
  warnings.warn("Liblinear failed to converge, increase "
```

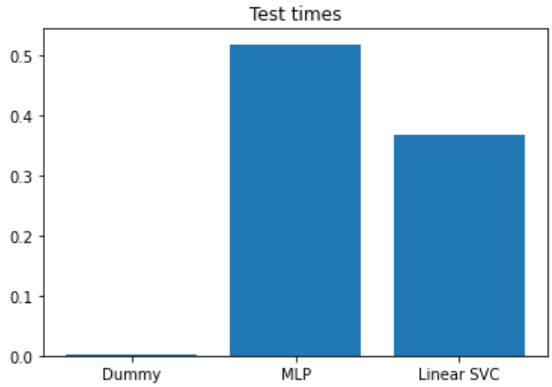
In [17]:
```
import matplotlib.pyplot as plt

plt.title("F1 Score")
plt.bar(*zip(*ootb_f1scores.items()))
plt.show()
```



In [18]:
```
plt.title("Train times")
plt.bar(*zip(*ootb_train_times.items()))
plt.show()
```



In [19]:
```
plt.title("Test times")
plt.bar(*zip(*ootb_test_times.items()))
plt.show()
```



Παρατηρούμε ότι και οι δύο ταξινομητές που μελετήσαμε παράγουν αρκετά καλύτερα αποτελέσματα από το dummy, ωστόσο είναι αισθητά πιο αργοί, το οποίο είναι αναμενόμενο δεδομένου του μεγέθους του train set. Ο ταξινομητής MLP παράγει πιο ικανοποιητικά αποτελέσματα, αλλά είναι εν γένει πιο αργός.

In [20]:
```
# save values
print("F1 scores:\n", ootb_f1scores)
print("Train times:\n", ootb_train_times)
print("Test times:\n", ootb_test_times)
```

```
F1 scores:
 {'Dummy': 0.37552813925976, 'MLP': 0.6969228607559365, 'Linear SVC': 0.5813148788927336}
Train times:
 {'Dummy': 0.003126100000002907, 'MLP': 145.01083319999992, 'Linear SVC': 20.38702610000007}
Test times:
 {'Dummy': 0.0012807999999893127, 'MLP': 0.5190602999999783, 'Linear SVC': 0.36728630000004614}
```

Βελτιστοποίηση

Θα χρησιμοποιήσουμε τη βιβλιοθήκη Optuna για τη βελτιστοποίηση. Αν και θα κάνουμε βελτιστοποίηση ως προς f1, θα τυπώνουμε σε κάθε βήμα και το accuracy απλώς για να έχουμε μια πιο ολοκληρωμένη εικόνα. Έτσι, παρακάτω εντός των κόκκινων μηνυμάτων θα περιέχεται το f1, ενώ στα λευκά (train και validation score) το accuracy.

MLP

In [12]:

```
%%time

import optuna
from imblearn.pipeline import Pipeline
from sklearn.feature_selection import VarianceThreshold
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import RandomOverSampler
from sklearn.decomposition import PCA
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import classification_report, f1_score
import warnings

warnings.filterwarnings("ignore")

def objective(trial):

    x_train,x_val,y_train,y_val = train_test_split(train_data,train_labels,test_size=0.25)

    # Choose pre-processing hyper-parameters
    vthreshold = trial.suggest_float("vthreshold", 0.01, 0.05) #προσαρμόζουμε τις τιμές μας στο variance που παρατηρήσαμε
    n_components = trial.suggest_int("n_components", 5, 50)

    selector = VarianceThreshold(threshold = vthreshold)
    std_scaler = StandardScaler()
    ros = RandomOverSampler()
    pca = PCA(n_components = n_components)

    # Choose classifier hyper-parameteres
    solver = trial.suggest_categorical("solver", ["lbfgs", "sgd", "adam"])
    hlsize = trial.suggest_int("hidden_layer_sizes", 10, 100)
    depth = trial.suggest_int("depth", 1, 2)
    activation = trial.suggest_categorical("activation", ["logistic","relu"])
    alpha = trial.suggest_loguniform("alpha", 1e-5, 1e-3)


    clf = MLPClassifier(solver = solver, alpha = alpha, hidden_layer_sizes = (hlsize, depth), early_stopping = True, activ
    ## random state 31 for reproducible results across multiple function calls

    pipe = Pipeline(steps = [("selector", selector), ("scaler", std_scaler), ("sampler", ros), ("pca", pca), ("mlp", clf)]

    # Train the model
    pipe.fit(x_train,y_train)

    # Evaluate the model
    pred = pipe.predict(x_val)
    metric = f1_score(y_val, pred)

    print("Validation Score:",pipe.score(x_val, y_val))
    print("Train Score:",pipe.score(x_train, y_train))
    print("\n=================")
    return metric

# objective is to maximize
study = optuna.create_study(direction = "maximize")
# start tuning for the hyper-parameters
study.optimize(objective, n_trials = 100)
```

```
[I 2021-11-24 22:26:57,628] A new study created in memory with name: no-name-2520a7c2-751a-4fd7-9078-62716537e115
Validation Score: 0.8340732054610939
Train Score:
[I 2021-11-24 22:29:43,367] Trial 0 finished with value: 0.7830942735136318 and parameters: {'vthreshold': 0.0452594982154
74744, 'n_components': 35, 'solver': 'adam', 'hidden_layer_sizes': 29, 'depth': 1, 'activation': 'relu', 'alpha': 0.000185
15342499507905}. Best is trial 0 with value: 0.7830942735136318.
0.8427841527766144

=================
Validation Score: 0.834701398777117
Train Score: 0.8575815953318257
```

```
=================
[I 2021-11-24 22:32:36,138] Trial 1 finished with value: 0.7904878178247252 and parameters: {'vthreshold': 0.0349704838627
2984, 'n_components': 32, 'solver': 'lbfgs', 'hidden_layer_sizes': 77, 'depth': 2, 'activation': 'logistic', 'alpha': 0.00
018005817384070025}. Best is trial 1 with value: 0.7904878178247252.
Validation Score: 0.8143060557835665
Train Score: 0.8249015830471563

=================
[I 2021-11-24 22:35:11,562] Trial 2 finished with value: 0.7623029913155356 and parameters: {'vthreshold': 0.0466145807319
39026, 'n_components': 15, 'solver': 'adam', 'hidden_layer_sizes': 85, 'depth': 1, 'activation': 'logistic', 'alpha': 0.00
016464145149279772}. Best is trial 1 with value: 0.7904878178247252.
Validation Score: 0.8164000335036435
Train Score: 0.8330262166010554

=================
[I 2021-11-24 22:38:02,629] Trial 3 finished with value: 0.7627705627705628 and parameters: {'vthreshold': 0.0451242563638
75986, 'n_components': 19, 'solver': 'lbfgs', 'hidden_layer_sizes': 94, 'depth': 2, 'activation': 'logistic', 'alpha': 0.0
0022847878973381982}. Best is trial 1 with value: 0.7904878178247252.
Validation Score: 0.8392243906524834
Train Score: 0.85171845771561

=================
[I 2021-11-24 22:41:01,134] Trial 4 finished with value: 0.7910748299319728 and parameters: {'vthreshold': 0.0187981349906
61845, 'n_components': 46, 'solver': 'lbfgs', 'hidden_layer_sizes': 71, 'depth': 2, 'activation': 'relu', 'alpha': 2.18828
55756666177e-05}. Best is trial 4 with value: 0.7910748299319728.
Validation Score: 0.8238964737415194
[I 2021-11-24 22:43:43,469] Trial 5 finished with value: 0.7768402059120099 and parameters: {'vthreshold': 0.0285619386930
9593, 'n_components': 25, 'solver': 'adam', 'hidden_layer_sizes': 79, 'depth': 1, 'activation': 'logistic', 'alpha': 0.000
5636177213484007}. Best is trial 4 with value: 0.7910748299319728.
Train Score: 0.8378423653572326

=================
Validation Score: 0.8278750314096658
[I 2021-11-24 22:46:18,907] Trial 6 finished with value: 0.7852215719063546 and parameters: {'vthreshold': 0.0369310159935
5688, 'n_components': 27, 'solver': 'adam', 'hidden_layer_sizes': 94, 'depth': 1, 'activation': 'logistic', 'alpha': 0.000
2258093066370662}. Best is trial 4 with value: 0.7910748299319728.
Train Score: 0.8430773096574252

=================
Validation Score: 0.8334450121450708
[I 2021-11-24 22:49:00,744] Trial 7 finished with value: 0.7852475835628274 and parameters: {'vthreshold': 0.0433950364233
4802, 'n_components': 35, 'solver': 'sgd', 'hidden_layer_sizes': 96, 'depth': 2, 'activation': 'relu', 'alpha': 3.50078327
6436251e-05}. Best is trial 4 with value: 0.7910748299319728.
Train Score: 0.8449898093084289

=================
Validation Score: 0.7930731217019851
Train Score: 0.8010721165926794

=================
[I 2021-11-24 22:51:51,039] Trial 8 finished with value: 0.7368589231506629 and parameters: {'vthreshold': 0.0230291732724
00526, 'n_components': 9, 'solver': 'lbfgs', 'hidden_layer_sizes': 94, 'depth': 1, 'activation': 'logistic', 'alpha': 1.75
9830136925516e-05}. Best is trial 4 with value: 0.7910748299319728.
Validation Score: 0.8456319624759192
[I 2021-11-24 22:54:37,531] Trial 9 finished with value: 0.8007998270644184 and parameters: {'vthreshold': 0.0424207086528
95635, 'n_components': 48, 'solver': 'lbfgs', 'hidden_layer_sizes': 58, 'depth': 1, 'activation': 'relu', 'alpha': 4.39951
4109356986e-05}. Best is trial 9 with value: 0.8007998270644184.
Train Score: 0.8612809559706285

=================
Validation Score: 0.8353714716475417
[I 2021-11-24 22:57:32,748] Trial 10 finished with value: 0.784520089897495 and parameters: {'vthreshold': 0.0108120845217
88966, 'n_components': 50, 'solver': 'sgd', 'hidden_layer_sizes': 42, 'depth': 1, 'activation': 'relu', 'alpha': 5.5400998
28894736e-05}. Best is trial 9 with value: 0.8007998270644184.
Train Score: 0.8453527654465757

=================
Validation Score: 0.8411089706005528
[I 2021-11-24 23:00:17,111] Trial 11 finished with value: 0.7943631436314361 and parameters: {'vthreshold': 0.018164917655
090734, 'n_components': 49, 'solver': 'lbfgs', 'hidden_layer_sizes': 59, 'depth': 2, 'activation': 'relu', 'alpha': 1.1203
322136848368e-05}. Best is trial 9 with value: 0.8007998270644184.
Train Score: 0.8717787642739481

=================
Validation Score: 0.8399782226317112
```

[I 2021-11-24 23:03:03,144] Trial 12 finished with value: 0.7944483296573244 and parameters: {'vthreshold': 0.010886186533644784, 'n_components': 43, 'solver': 'lbfgs', 'hidden_layer_sizes': 57, 'depth': 2, 'activation': 'relu', 'alpha': 1.0177712635034917e-05}. Best is trial 9 with value: 0.8007998270644184.
Train Score: 0.8595638942401653

==========
Validation Score: 0.8478096993047994
[I 2021-11-24 23:05:36,435] Trial 13 finished with value: 0.8041181543768865 and parameters: {'vthreshold': 0.037322450649964985, 'n_components': 44, 'solver': 'lbfgs', 'hidden_layer_sizes': 52, 'depth': 2, 'activation': 'relu', 'alpha': 6.399993665849823e-05}. Best is trial 13 with value: 0.8041181543768865.
Train Score: 0.8689309545746433

==========
Validation Score: 0.8411508501549544
Train Score: 0.8470698271770388

==========
[I 2021-11-24 23:07:55,347] Trial 14 finished with value: 0.7963927210263568 and parameters: {'vthreshold': 0.037922207598344984, 'n_components': 41, 'solver': 'lbfgs', 'hidden_layer_sizes': 11, 'depth': 2, 'activation': 'relu', 'alpha': 6.261721219613415e-05}. Best is trial 13 with value: 0.8041181543768865.
Validation Score: 0.8344920010051093
Train Score: 0.8491498450456487

==========
[I 2021-11-24 23:10:27,571] Trial 15 finished with value: 0.7850070721357849 and parameters: {'vthreshold': 0.03131413791260672, 'n_components': 40, 'solver': 'lbfgs', 'hidden_layer_sizes': 41, 'depth': 1, 'activation': 'relu', 'alpha': 8.36057317022223e-05}. Best is trial 13 with value: 0.8041181543768865.
Validation Score: 0.834701398777117
[I 2021-11-24 23:13:10,749] Trial 16 finished with value: 0.7895718931598871 and parameters: {'vthreshold': 0.04002356440757324, 'n_components': 39, 'solver': 'sgd', 'hidden_layer_sizes': 46, 'depth': 1, 'activation': 'relu', 'alpha': 3.070624201350249e-05}. Best is trial 13 with value: 0.8041181543768865.
Train Score: 0.8439567802998577

==========
Validation Score: 0.8471396264343748
Train Score: 0.873244548678002

==========
[I 2021-11-24 23:15:54,931] Trial 17 finished with value: 0.8034887477118552 and parameters: {'vthreshold': 0.04942678835431028, 'n_components': 45, 'solver': 'lbfgs', 'hidden_layer_sizes': 65, 'depth': 2, 'activation': 'relu', 'alpha': 0.0005021016847483429}. Best is trial 13 with value: 0.8041181543768865.
Validation Score: 0.834701398777117
Train Score: 0.8565346064717871

==========
[I 2021-11-24 23:18:55,820] Trial 18 finished with value: 0.7856871368844004 and parameters: {'vthreshold': 0.04845662433090212, 'n_components': 31, 'solver': 'lbfgs', 'hidden_layer_sizes': 67, 'depth': 2, 'activation': 'relu', 'alpha': 0.000896892261086702}. Best is trial 13 with value: 0.8041181543768865.
Validation Score: 0.802119105452718
Train Score: 0.8095317865817908

==========
[I 2021-11-24 23:21:51,880] Trial 19 finished with value: 0.7513027001421129 and parameters: {'vthreshold': 0.03228770186408448, 'n_components': 21, 'solver': 'sgd', 'hidden_layer_sizes': 29, 'depth': 2, 'activation': 'relu', 'alpha': 0.00038260400528035253}. Best is trial 13 with value: 0.8041181543768865.
Validation Score: 0.8412346092637575
Train Score: 0.8574699165200882

==========
[I 2021-11-24 23:24:34,507] Trial 20 finished with value: 0.7958865019113767 and parameters: {'vthreshold': 0.04873743061448713, 'n_components': 45, 'solver': 'lbfgs', 'hidden_layer_sizes': 30, 'depth': 2, 'activation': 'relu', 'alpha': 0.0001087106480323155 6}. Best is trial 13 with value: 0.8041181543768865.
Validation Score: 0.851955775190552
Train Score: 0.8712064103637938

==========
[I 2021-11-24 23:27:38,410] Trial 21 finished with value: 0.8087017695762757 and parameters: {'vthreshold': 0.04119377069009062, 'n_components': 50, 'solver': 'lbfgs', 'hidden_layer_sizes': 63, 'depth': 2, 'activation': 'relu', 'alpha': 4.354925798366319e-05}. Best is trial 21 with value: 0.8087017695762757.
Validation Score: 0.8419884412429852
Train Score: 0.8636122511656475

==========
[I 2021-11-24 23:30:40,300] Trial 22 finished with value: 0.7979002624671917 and parameters: {'vthreshold': 0.04067898065490589, 'n_components': 38, 'solver': 'lbfgs', 'hidden_layer_sizes': 67, 'depth': 2, 'activation': 'relu', 'alpha': 0.0001013164454974 68}. Best is trial 21 with value: 0.8087017695762757.
Validation Score: 0.8501968339056872

```
Train Score: 0.8708574140771143

==================
[I 2021-11-24 23:33:35,814] Trial 23 finished with value: 0.8068261597450992 and parameters: {'vthreshold': 0.034763013300
624326, 'n_components': 45, 'solver': 'lbfgs', 'hidden_layer_sizes': 49, 'depth': 2, 'activation': 'relu', 'alpha': 6.8905
90863144156e-05}. Best is trial 21 with value: 0.8087017695762757.
Validation Score: 0.8439567802998577
Train Score: 0.8633051344333696

==================
[I 2021-11-24 23:36:36,154] Trial 24 finished with value: 0.7989206691851053 and parameters: {'vthreshold': 0.027542769841
867133, 'n_components': 43, 'solver': 'lbfgs', 'hidden_layer_sizes': 50, 'depth': 2, 'activation': 'relu', 'alpha': 7.0808
29661111818e-05}. Best is trial 21 with value: 0.8087017695762757.
Validation Score: 0.8521232934081582
Train Score: 0.8753106066951447

==================
[I 2021-11-24 23:39:31,325] Trial 25 finished with value: 0.8101510833915802 and parameters: {'vthreshold': 0.034942650621
52232, 'n_components': 50, 'solver': 'lbfgs', 'hidden_layer_sizes': 51, 'depth': 2, 'activation': 'relu', 'alpha': 2.60585
06087871053e-05}. Best is trial 25 with value: 0.8101510833915802.
Validation Score: 0.8460088784655331
Train Score: 0.8613088756735628

==================
[I 2021-11-24 23:42:25,909] Trial 26 finished with value: 0.7993232549255034 and parameters: {'vthreshold': 0.034428619967
40335, 'n_components': 49, 'solver': 'lbfgs', 'hidden_layer_sizes': 36, 'depth': 2, 'activation': 'relu', 'alpha': 2.16937
98764517295e-05}. Best is trial 25 with value: 0.8101510833915802.
Validation Score: 0.8406064159477343
[I 2021-11-24 23:45:08,419] Trial 27 finished with value: 0.7948911403319681 and parameters: {'vthreshold': 0.024196930669
225573, 'n_components': 50, 'solver': 'adam', 'hidden_layer_sizes': 18, 'depth': 2, 'activation': 'relu', 'alpha': 3.28418
3934083465e-05}. Best is trial 25 with value: 0.8101510833915802.
Train Score: 0.8496942792528688

==================
Validation Score: 0.8353714716475417
Train Score: 0.8502805930144903

==================
[I 2021-11-24 23:48:06,981] Trial 28 finished with value: 0.7862773881367912 and parameters: {'vthreshold': 0.033468377077
51249, 'n_components': 35, 'solver': 'lbfgs', 'hidden_layer_sizes': 47, 'depth': 2, 'activation': 'relu', 'alpha': 1.57712
4019847402e-05}. Best is trial 25 with value: 0.8101510833915802.
Validation Score: 0.8280844291816735
[I 2021-11-24 23:51:00,752] Trial 29 finished with value: 0.778956437456249 and parameters: {'vthreshold': 0.0394079678697
05865, 'n_components': 36, 'solver': 'sgd', 'hidden_layer_sizes': 34, 'depth': 2, 'activation': 'relu', 'alpha': 0.0001305
0033254896815}. Best is trial 25 with value: 0.8101510833915802.
Train Score: 0.8333193734818661

==================
Validation Score: 0.8452969260407069
Train Score: 0.8574559566686211

==================
[I 2021-11-24 23:53:48,420] Trial 30 finished with value: 0.8007766152518607 and parameters: {'vthreshold': 0.029658814676
341547, 'n_components': 47, 'solver': 'adam', 'hidden_layer_sizes': 24, 'depth': 2, 'activation': 'relu', 'alpha': 4.51187
4708145934e-05}. Best is trial 25 with value: 0.8101510833915802.
Validation Score: 0.8450875282686993
Train Score: 0.8677583270514002

==================
[I 2021-11-24 23:56:46,338] Trial 31 finished with value: 0.8016090104585679 and parameters: {'vthreshold': 0.036299923508
37173, 'n_components': 42, 'solver': 'lbfgs', 'hidden_layer_sizes': 53, 'depth': 2, 'activation': 'relu', 'alpha': 5.01874
8213993713e-05}. Best is trial 25 with value: 0.8101510833915802.
Validation Score: 0.8478515788592009
[I 2021-11-24 23:59:43,275] Trial 32 finished with value: 0.8048348106365834 and parameters: {'vthreshold': 0.042708141115
20758, 'n_components': 45, 'solver': 'lbfgs', 'hidden_layer_sizes': 62, 'depth': 2, 'activation': 'relu', 'alpha': 2.53536
3364274354e-05}. Best is trial 25 with value: 0.8101510833915802.
Train Score: 0.8759667197141022

==================
Validation Score: 0.8528771253873859
Train Score: 0.8699919032861491

==================
[I 2021-11-25 00:02:49,302] Trial 33 finished with value: 0.8123297184678668 and parameters: {'vthreshold': 0.042326407214
433884, 'n_components': 47, 'solver': 'lbfgs', 'hidden_layer_sizes': 62, 'depth': 2, 'activation': 'relu', 'alpha': 3.1492
63977958106e-05}. Best is trial 33 with value: 0.8123297184678668.
```

```
Validation Score: 0.846385794455147
[I 2021-11-25 00:06:04,415] Trial 34 finished with value: 0.8003700881680635 and parameters: {'vthreshold': 0.044704027377
78651, 'n_components': 47, 'solver': 'lbfgs', 'hidden_layer_sizes': 74, 'depth': 2, 'activation': 'logistic', 'alpha': 3.9
83704072799154e-05}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8763715554066505

=================
Validation Score: 0.7762794203869671
Train Score: 0.7804394561241869

=================
[I 2021-11-25 00:09:11,035] Trial 35 finished with value: 0.7059341627215677 and parameters: {'vthreshold': 0.046746461273
45469, 'n_components': 5, 'solver': 'lbfgs', 'hidden_layer_sizes': 82, 'depth': 2, 'activation': 'relu', 'alpha': 1.489796
3512156194e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8336125303626769
[I 2021-11-25 00:12:12,742] Trial 36 finished with value: 0.7876423111871292 and parameters: {'vthreshold': 0.035526374540
87821, 'n_components': 31, 'solver': 'lbfgs', 'hidden_layer_sizes': 72, 'depth': 2, 'activation': 'logistic', 'alpha': 2.8
493841606591495e-05}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8527794064271156

=================
Validation Score: 0.8487310495016333
Train Score: 0.866432141162018

=================
[I 2021-11-25 00:14:49,773] Trial 37 finished with value: 0.8020604997807979 and parameters: {'vthreshold': 0.039537392301
0737, 'n_components': 50, 'solver': 'adam', 'hidden_layer_sizes': 42, 'depth': 2, 'activation': 'relu', 'alpha': 2.0851378
607157538e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8519138956361504
[I 2021-11-25 00:17:52,073] Trial 38 finished with value: 0.8066914498141263 and parameters: {'vthreshold': 0.026756796779
161846, 'n_components': 47, 'solver': 'lbfgs', 'hidden_layer_sizes': 60, 'depth': 2, 'activation': 'logistic', 'alpha': 0.
00014331270542738067}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8739704609542954

=================
Validation Score: 0.8154786833068096
[I 2021-11-25 00:20:59,372] Trial 39 finished with value: 0.762377305576529 and parameters: {'vthreshold': 0.0459706182895
9758, 'n_components': 15, 'solver': 'lbfgs', 'hidden_layer_sizes': 87, 'depth': 2, 'activation': 'relu', 'alpha': 8.612556
34969488e-05}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8275958343803222

=================
Validation Score: 0.846385794455147
Train Score: 0.8620347879498562

=================
[I 2021-11-25 00:23:39,592] Trial 40 finished with value: 0.8048728588147676 and parameters: {'vthreshold': 0.041600776153
86702, 'n_components': 41, 'solver': 'adam', 'hidden_layer_sizes': 77, 'depth': 2, 'activation': 'logistic', 'alpha': 2.58
92995112921896e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8473490242063825
[I 2021-11-25 00:26:45,303] Trial 41 finished with value: 0.8021494870542257 and parameters: {'vthreshold': 0.026761662643
595103, 'n_components': 47, 'solver': 'lbfgs', 'hidden_layer_sizes': 63, 'depth': 2, 'activation': 'logistic', 'alpha': 0.
00013955290407692021}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8728955523913225

=================
Validation Score: 0.8443336962894715
[I 2021-11-25 00:29:45,169] Trial 42 finished with value: 0.7985038217596356 and parameters: {'vthreshold': 0.031377060232
5436, 'n_components': 47, 'solver': 'lbfgs', 'hidden_layer_sizes': 55, 'depth': 2, 'activation': 'logistic', 'alpha': 0.00
030011572661970933}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8629002987408214

=================
Validation Score: 0.8506156294497027
[I 2021-11-25 00:32:47,133] Trial 43 finished with value: 0.8072412861388814 and parameters: {'vthreshold': 0.020632097505
194333, 'n_components': 50, 'solver': 'lbfgs', 'hidden_layer_sizes': 47, 'depth': 2, 'activation': 'logistic', 'alpha': 3.
787053811449478e-05}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.868595918139431

=================
Validation Score: 0.8414021274813636
Train Score: 0.8588659016668063

=================
[I 2021-11-25 00:35:53,015] Trial 44 finished with value: 0.7966929725667042 and parameters: {'vthreshold': 0.018065005319
180025, 'n_components': 50, 'solver': 'lbfgs', 'hidden_layer_sizes': 48, 'depth': 2, 'activation': 'logistic', 'alpha': 3.
```

796906864455109e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8409414523829466
Train Score: 0.8610855180500879

==========

[I 2021-11-25 00:38:58,089] Trial 45 finished with value: 0.7937886849820828 and parameters: {'vthreshold': 0.014711676650808435, 'n_components': 45, 'solver': 'lbfgs', 'hidden_layer_sizes': 44, 'depth': 2, 'activation': 'logistic', 'alpha': 5.222730704839288e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8389312337716727
Train Score: 0.849945556579278

==========

[I 2021-11-25 00:42:04,587] Trial 46 finished with value: 0.7923550372529963 and parameters: {'vthreshold': 0.022603574388860365, 'n_components': 48, 'solver': 'sgd', 'hidden_layer_sizes': 38, 'depth': 2, 'activation': 'logistic', 'alpha': 7.607299651481057e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8359996649635648
Train Score: 0.8596895329033699

==========

[I 2021-11-25 00:45:04,678] Trial 47 finished with value: 0.7905658359182799 and parameters: {'vthreshold': 0.038304745519898487, 'n_components': 37, 'solver': 'lbfgs', 'hidden_layer_sizes': 56, 'depth': 2, 'activation': 'relu', 'alpha': 1.7803118009228077e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8350783147667309
Train Score: 0.8440963788145294

==========

[I 2021-11-25 00:48:06,362] Trial 48 finished with value: 0.7870200108166576 and parameters: {'vthreshold': 0.043906127514060614, 'n_components': 43, 'solver': 'lbfgs', 'hidden_layer_sizes': 69, 'depth': 1, 'activation': 'relu', 'alpha': 3.527331189636898e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8254041376999749
Train Score: 0.8394058687215568

==========

[I 2021-11-25 00:50:59,856] Trial 49 finished with value: 0.7730909486746856 and parameters: {'vthreshold': 0.033510066891673312, 'n_components': 23, 'solver': 'lbfgs', 'hidden_layer_sizes': 51, 'depth': 2, 'activation': 'relu', 'alpha': 1.3415399641463422e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8276237540832565
Train Score: 0.845841360247927

==========

[I 2021-11-25 00:54:00,997] Trial 50 finished with value: 0.7795865909821141 and parameters: {'vthreshold': 0.020571356811527358, 'n_components': 27, 'solver': 'lbfgs', 'hidden_layer_sizes': 60, 'depth': 2, 'activation': 'logistic', 'alpha': 5.4758620387940476e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8475584219783902
Train Score: 0.8710109724432532

==========

[I 2021-11-25 00:57:09,984] Trial 51 finished with value: 0.8048048048048049 and parameters: {'vthreshold': 0.025383260678005902, 'n_components': 48, 'solver': 'lbfgs', 'hidden_layer_sizes': 59, 'depth': 2, 'activation': 'logistic', 'alpha': 4.512683341123288e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8436636234190468

[I 2021-11-25 01:00:09,754] Trial 52 finished with value: 0.8012776151184455 and parameters: {'vthreshold': 0.029880470240229315, 'n_components': 46, 'solver': 'lbfgs', 'hidden_layer_sizes': 63, 'depth': 2, 'activation': 'logistic', 'alpha': 0.0001956892736061278 3}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8712343300667281

==========

Validation Score: 0.8486472903928303
Train Score: 0.870843454225647

==========

[I 2021-11-25 01:03:09,704] Trial 53 finished with value: 0.8050701186623517 and parameters: {'vthreshold': 0.028491426562212475, 'n_components': 50, 'solver': 'lbfgs', 'hidden_layer_sizes': 55, 'depth': 2, 'activation': 'logistic', 'alpha': 0.00015207138442975391}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8362928218443756
Train Score: 0.8599128905268448

==========

[I 2021-11-25 01:06:21,053] Trial 54 finished with value: 0.7905705866595231 and parameters: {'vthreshold': 0.015709356888235663, 'n_components': 44, 'solver': 'lbfgs', 'hidden_layer_sizes': 50, 'depth': 2, 'activation': 'logistic', 'alpha': 0.0001162342735359064 8}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8424909958958037
Train Score: 0.862174386464528

==========

[I 2021-11-25 01:09:22,635] Trial 55 finished with value: 0.7955867166693842 and parameters: {'vthreshold': 0.020364208939

03769, 'n_components': 40, 'solver': 'lbfgs', 'hidden_layer_sizes': 54, 'depth': 2, 'activation': 'logistic', 'alpha': 2.8331696655205732e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8388893542172712
Train Score: 0.8440126197057264

==========

[I 2021-11-25 01:12:20,420] Trial 56 finished with value: 0.7929159713624374 and parameters: {'vthreshold': 0.037886884915 0052, 'n_components': 48, 'solver': 'sgd', 'hidden_layer_sizes': 45, 'depth': 2, 'activation': 'relu', 'alpha': 6.664419411933583e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8341988441242986

[I 2021-11-25 01:15:12,443] Trial 57 finished with value: 0.7851988497639846 and parameters: {'vthreshold': 0.041639412221 33994, 'n_components': 33, 'solver': 'lbfgs', 'hidden_layer_sizes': 39, 'depth': 1, 'activation': 'relu', 'alpha': 9.30889 5151896589e-05}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.849331323114722

==========
Validation Score: 0.8531702822681967
Train Score: 0.8742356981321718

==========

[I 2021-11-25 01:18:22,591] Trial 58 finished with value: 0.8117482817869415 and parameters: {'vthreshold': 0.035016283270 72911, 'n_components': 46, 'solver': 'lbfgs', 'hidden_layer_sizes': 66, 'depth': 2, 'activation': 'relu', 'alpha': 0.00026 02982578312142}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8505737498953011
Train Score: 0.8742356981321718

==========

[I 2021-11-25 01:21:16,511] Trial 59 finished with value: 0.8098081023454158 and parameters: {'vthreshold': 0.035007223396 365666, 'n_components': 42, 'solver': 'adam', 'hidden_layer_sizes': 70, 'depth': 2, 'activation': 'relu', 'alpha': 0.00057 3724684204751}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8536309573666137
Train Score: 0.8709272133344501

==========

[I 2021-11-25 01:24:02,560] Trial 60 finished with value: 0.8103532475988929 and parameters: {'vthreshold': 0.036756076132 04065, 'n_components': 43, 'solver': 'adam', 'hidden_layer_sizes': 77, 'depth': 2, 'activation': 'relu', 'alpha': 0.000737 7755347088214}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8476840606415947
Train Score: 0.8693637099701259

==========

[I 2021-11-25 01:26:50,311] Trial 61 finished with value: 0.805038863575449 and parameters: {'vthreshold': 0.0366374864947 6221, 'n_components': 42, 'solver': 'adam', 'hidden_layer_sizes': 74, 'depth': 2, 'activation': 'relu', 'alpha': 0.0009295 634495649429}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8537984755842197

[I 2021-11-25 01:29:47,993] Trial 62 finished with value: 0.8087647219939743 and parameters: {'vthreshold': 0.032913606028 729905, 'n_components': 49, 'solver': 'adam', 'hidden_layer_sizes': 87, 'depth': 2, 'activation': 'relu', 'alpha': 0.00069 71473276959144}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8823324119831365

==========
Validation Score: 0.8527096071697797
Train Score: 0.8790937264427506

==========

[I 2021-11-25 01:32:44,514] Trial 63 finished with value: 0.8109851131294674 and parameters: {'vthreshold': 0.033043662660 095854, 'n_components': 44, 'solver': 'adam', 'hidden_layer_sizes': 83, 'depth': 2, 'activation': 'relu', 'alpha': 0.00074 61278961091028}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8396431861964989
Train Score: 0.8626211017114778

==========

[I 2021-11-25 01:35:38,732] Trial 64 finished with value: 0.7919587068731323 and parameters: {'vthreshold': 0.032132170632 188845, 'n_components': 39, 'solver': 'adam', 'hidden_layer_sizes': 87, 'depth': 2, 'activation': 'relu', 'alpha': 0.00070 97062349495664}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8548873439986598

[I 2021-11-25 01:38:44,153] Trial 65 finished with value: 0.8099391146947508 and parameters: {'vthreshold': 0.034104223419 56182, 'n_components': 43, 'solver': 'adam', 'hidden_layer_sizes': 82, 'depth': 2, 'activation': 'relu', 'alpha': 0.000574 4168551307116}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8794148030264958

==========
Validation Score: 0.8503643521232934
Train Score: 0.871150570957925

==========

[I 2021-11-25 01:41:42,962] Trial 66 finished with value: 0.8072919475756432 and parameters: {'vthreshold': 0.035581758319290994, 'n_components': 43, 'solver': 'adam', 'hidden_layer_sizes': 82, 'depth': 2, 'activation': 'relu', 'alpha': 0.00047659495966894197}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8531284027137951
Train Score: 0.8760085992685038

==========

[I 2021-11-25 01:44:33,838] Trial 67 finished with value: 0.8053288925895088 and parameters: {'vthreshold': 0.03449031095401985, 'n_components': 41, 'solver': 'adam', 'hidden_layer_sizes': 91, 'depth': 2, 'activation': 'relu', 'alpha': 0.0003772111055485339}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8396013066420973
Train Score: 0.858070190133177

==========

[I 2021-11-25 01:47:13,633] Trial 68 finished with value: 0.7910529187124932 and parameters: {'vthreshold': 0.0370683987947196, 'n_components': 38, 'solver': 'adam', 'hidden_layer_sizes': 78, 'depth': 2, 'activation': 'relu', 'alpha': 0.0007881447573357275}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8499036770248765
Train Score: 0.8818996565876539

==========

[I 2021-11-25 01:50:10,062] Trial 69 finished with value: 0.8050902762671308 and parameters: {'vthreshold': 0.031165855981121544, 'n_components': 44, 'solver': 'adam', 'hidden_layer_sizes': 98, 'depth': 2, 'activation': 'relu', 'alpha': 0.0005332864242668614}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8351201943211324
Train Score: 0.8541893514253008

==========

[I 2021-11-25 01:52:53,359] Trial 70 finished with value: 0.7819199025092783 and parameters: {'vthreshold': 0.0386415818774403, 'n_components': 29, 'solver': 'adam', 'hidden_layer_sizes': 81, 'depth': 2, 'activation': 'relu', 'alpha': 0.0005965743436791446}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8333193734818661
Train Score: 0.8509785855878493

==========

[I 2021-11-25 01:55:35,769] Trial 71 finished with value: 0.7985422150232842 and parameters: {'vthreshold': 0.03414060502451334, 'n_components': 45, 'solver': 'adam', 'hidden_layer_sizes': 91, 'depth': 2, 'activation': 'relu', 'alpha': 0.0006621187605879859}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8501968339056872
Train Score: 0.8723231984811681

==========

[I 2021-11-25 01:58:22,947] Trial 72 finished with value: 0.8029309679907443 and parameters: {'vthreshold': 0.033195158526441206, 'n_components': 46, 'solver': 'adam', 'hidden_layer_sizes': 85, 'depth': 2, 'activation': 'relu', 'alpha': 0.0004321612732441757}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8467208308903593
Train Score: 0.8698243850685429

==========

[I 2021-11-25 02:01:14,096] Trial 73 finished with value: 0.7985912392692055 and parameters: {'vthreshold': 0.032737307078756335, 'n_components': 41, 'solver': 'adam', 'hidden_layer_sizes': 74, 'depth': 2, 'activation': 'relu', 'alpha': 0.0002782113277394114}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8466370717815562
Train Score: 0.8680933634866125

==========

[I 2021-11-25 02:04:04,554] Trial 74 finished with value: 0.802842683320771 and parameters: {'vthreshold': 0.035850870886761685, 'n_components': 43, 'solver': 'adam', 'hidden_layer_sizes': 68, 'depth': 2, 'activation': 'relu', 'alpha': 0.0008104809532773215}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8434542256470391
Train Score: 0.8697825055141414

==========

[I 2021-11-25 02:06:50,799] Trial 75 finished with value: 0.8004058094831268 and parameters: {'vthreshold': 0.030804819024642494, 'n_components': 48, 'solver': 'adam', 'hidden_layer_sizes': 71, 'depth': 2, 'activation': 'relu', 'alpha': 0.0006165278917270133}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8420303207973867
Train Score: 0.8778652595136388

==========

[I 2021-11-25 02:09:53,231] Trial 76 finished with value: 0.7987622705932566 and parameters: {'vthreshold': 0.02901945688524666, 'n_components': 46, 'solver': 'adam', 'hidden_layer_sizes': 77, 'depth': 2, 'activation': 'relu', 'alpha': 0.0007528806060071966}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8469721082167686

[I 2021-11-25 02:12:48,993] Trial 77 finished with value: 0.8026145203111495 and parameters: {'vthreshold': 0.040330440894054034, 'n_components': 44, 'solver': 'adam', 'hidden_layer_sizes': 85, 'depth': 2, 'activation': 'relu', 'alpha': 0.00097

62473977198516}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8657620682915934

==========
Validation Score: 0.8452131669319038
Train Score: 0.875799201496496

==========
[I 2021-11-25 02:15:44,643] Trial 78 finished with value: 0.7988680888114932 and parameters: {'vthreshold': 0.035077399986
04398, 'n_components': 49, 'solver': 'adam', 'hidden_layer_sizes': 91, 'depth': 2, 'activation': 'relu', 'alpha': 0.000351
70042153264704}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8456319624759192
[I 2021-11-25 02:18:38,507] Trial 79 finished with value: 0.8027400192657604 and parameters: {'vthreshold': 0.032099251420
298874, 'n_components': 39, 'solver': 'adam', 'hidden_layer_sizes': 80, 'depth': 2, 'activation': 'relu', 'alpha': 0.00045
840866742054574}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8612809559706285

==========
Validation Score: 0.8421559594605913
[I 2021-11-25 02:21:26,818] Trial 80 finished with value: 0.7949736169286841 and parameters: {'vthreshold': 0.037149767348
83386, 'n_components': 35, 'solver': 'adam', 'hidden_layer_sizes': 71, 'depth': 2, 'activation': 'relu', 'alpha': 0.000569
4251509091397}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8580981098361113

==========
Validation Score: 0.8471396264343748
Train Score: 0.873705223776419

==========
[I 2021-11-25 02:24:26,674] Trial 81 finished with value: 0.8078947368421052 and parameters: {'vthreshold': 0.039355585003
69629, 'n_components': 49, 'solver': 'adam', 'hidden_layer_sizes': 64, 'depth': 2, 'activation': 'relu', 'alpha': 0.000864
3238692217691}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8485635312840272
Train Score: 0.8677164474969986

==========
[I 2021-11-25 02:27:10,275] Trial 82 finished with value: 0.8021232351975484 and parameters: {'vthreshold': 0.041100014679
15322, 'n_components': 49, 'solver': 'adam', 'hidden_layer_sizes': 65, 'depth': 2, 'activation': 'relu', 'alpha': 0.000239
71549739942776}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.840271379512522
[I 2021-11-25 02:30:07,964] Trial 83 finished with value: 0.7924240775008162 and parameters: {'vthreshold': 0.033839417172
60107, 'n_components': 46, 'solver': 'sgd', 'hidden_layer_sizes': 75, 'depth': 2, 'activation': 'relu', 'alpha': 0.0006907
929370184466}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8522070525169613

==========
Validation Score: 0.8489823268280425
Train Score: 0.8687913560599715

==========
[I 2021-11-25 02:32:48,674] Trial 84 finished with value: 0.8056483777083109 and parameters: {'vthreshold': 0.035939652174
036434, 'n_components': 42, 'solver': 'adam', 'hidden_layer_sizes': 84, 'depth': 2, 'activation': 'relu', 'alpha': 0.00040
695003715069734}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8139710193483541
[I 2021-11-25 02:35:27,813] Trial 85 finished with value: 0.7572942847776198 and parameters: {'vthreshold': 0.043280475191
6388, 'n_components': 17, 'solver': 'adam', 'hidden_layer_sizes': 88, 'depth': 2, 'activation': 'relu', 'alpha': 0.0005209
511929235747}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8206019487952648

==========
Validation Score: 0.8413602479269621
[I 2021-11-25 02:38:05,682] Trial 86 finished with value: 0.8005895978100653 and parameters: {'vthreshold': 0.047063008769
76543, 'n_components': 47, 'solver': 'adam', 'hidden_layer_sizes': 67, 'depth': 2, 'activation': 'relu', 'alpha': 0.000333
1023991364874}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8606108831002038

==========
Validation Score: 0.8295920931401289
[I 2021-11-25 02:40:59,245] Trial 87 finished with value: 0.7801610027554163 and parameters: {'vthreshold': 0.038439402229
288835, 'n_components': 46, 'solver': 'sgd', 'hidden_layer_sizes': 70, 'depth': 2, 'activation': 'relu', 'alpha': 2.377522
7386995973e-05}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8442638970321356

==========
Validation Score: 0.8420722003517882
[I 2021-11-25 02:43:40,493] Trial 88 finished with value: 0.799702554841451 and parameters: {'vthreshold': 0.0304620880894

17517, 'n_components': 48, 'solver': 'adam', 'hidden_layer_sizes': 79, 'depth': 2, 'activation': 'relu', 'alpha': 0.000616
5693297665688}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8571069603819416

==========
Validation Score: 0.8492336041544518

[I 2021-11-25 02:46:22,278] Trial 89 finished with value: 0.7989500726013626 and parameters: {'vthreshold': 0.034843550399
526114, 'n_components': 44, 'solver': 'adam', 'hidden_layer_sizes': 61, 'depth': 2, 'activation': 'relu', 'alpha': 1.99986
29559741064e-05}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8673814110617863

==========
Validation Score: 0.8513694614289304
Train Score: 0.8708574140771143

==========

[I 2021-11-25 02:49:26,648] Trial 90 finished with value: 0.8068571428571429 and parameters: {'vthreshold': 0.031731160909
14114, 'n_components': 50, 'solver': 'lbfgs', 'hidden_layer_sizes': 57, 'depth': 1, 'activation': 'relu', 'alpha': 3.19806
8812357939e-05}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8523745707345673

[I 2021-11-25 02:52:21,458] Trial 91 finished with value: 0.8078285994657363 and parameters: {'vthreshold': 0.040137613006
58483, 'n_components': 49, 'solver': 'adam', 'hidden_layer_sizes': 64, 'depth': 2, 'activation': 'relu', 'alpha': 0.000850
9186504336661}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8743194572409749

==========
Validation Score: 0.8523326911801659
Train Score: 0.874473015607114

==========

[I 2021-11-25 02:55:09,578] Trial 92 finished with value: 0.8090751570283734 and parameters: {'vthreshold': 0.038980074417
13803, 'n_components': 48, 'solver': 'adam', 'hidden_layer_sizes': 65, 'depth': 2, 'activation': 'relu', 'alpha': 0.000978
2649610839218}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.7992294161990117
Train Score: 0.8059301449032582

==========

[I 2021-11-25 02:57:59,254] Trial 93 finished with value: 0.7470184696569921 and parameters: {'vthreshold': 0.036414329478
64339, 'n_components': 11, 'solver': 'adam', 'hidden_layer_sizes': 66, 'depth': 2, 'activation': 'relu', 'alpha': 0.000981
3514219985191}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.850489990786498

[I 2021-11-25 03:00:40,836] Trial 94 finished with value: 0.8041044776119404 and parameters: {'vthreshold': 0.042026595327
25997, 'n_components': 47, 'solver': 'adam', 'hidden_layer_sizes': 75, 'depth': 2, 'activation': 'relu', 'alpha': 0.000703
5874639840817}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8703269397213613

==========
Validation Score: 0.8525420889521735

[I 2021-11-25 03:03:31,742] Trial 95 finished with value: 0.8071002027064044 and parameters: {'vthreshold': 0.037537603247
04164, 'n_components': 45, 'solver': 'adam', 'hidden_layer_sizes': 83, 'depth': 2, 'activation': 'relu', 'alpha': 0.000502
8240922665295}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8796800402043722

==========
Validation Score: 0.8498199179160734
Train Score: 0.8760085992685038

==========

[I 2021-11-25 03:06:56,755] Trial 96 finished with value: 0.8052780191138141 and parameters: {'vthreshold': 0.038973917786
54409, 'n_components': 48, 'solver': 'lbfgs', 'hidden_layer_sizes': 89, 'depth': 2, 'activation': 'relu', 'alpha': 0.00018
021506475464447}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.842909791439819

[I 2021-11-25 03:10:06,812] Trial 97 finished with value: 0.7987768896518427 and parameters: {'vthreshold': 0.033061797388
401, 'n_components': 42, 'solver': 'lbfgs', 'hidden_layer_sizes': 72, 'depth': 2, 'activation': 'relu', 'alpha': 0.0007623
827619769263}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8684842393276936

==========
Validation Score: 0.8475584219783902
Train Score: 0.8619929083954547

==========

[I 2021-11-25 03:12:49,596] Trial 98 finished with value: 0.8055139987176747 and parameters: {'vthreshold': 0.044780453940
03676, 'n_components': 49, 'solver': 'adam', 'hidden_layer_sizes': 58, 'depth': 2, 'activation': 'relu', 'alpha': 0.000638
8583620187282}. Best is trial 33 with value: 0.8123297184678668.
Validation Score: 0.8467627104447608

```
[I 2021-11-25 03:15:50,229] Trial 99 finished with value: 0.7955066227016152 and parameters: {'vthreshold': 0.042801965308
661316, 'n_components': 40, 'solver': 'lbfgs', 'hidden_layer_sizes': 62, 'depth': 2, 'activation': 'relu', 'alpha': 0.0008
672148474117009}. Best is trial 33 with value: 0.8123297184678668.
Train Score: 0.8720300416003574

=================
Wall time: 4h 48min 52s
```

Αυτή τη φορά χρησιμοποιήσαμε ποσοστό 25% για το validation set, επειδή το σύνολο με το οποίο ασχολούμαστε είναι μικρότερο συγκριτικά με πριν και θέλουμε να εξασφαλίσουμε ότι το validation set θα είναι επαρκώς μεγάλο.

Χρησιμοποιούμε βελτιστοποίηση με τη βιβλιοθήκη Optuna, κάνοντας 100 trials για διαφορετικές τιμές των παραμέτρων. Οι παράμετροι που αποφασίσαμε να επηρεάσουμε μαζί με τα αντίστοιχα διαστήματα τιμών φαίνονται παραπάνω στον κώδικα. Πριν αποφασίσουμε για τα διαστήματα των τιμών, πραγματοποιήσαμε μερικές δοκιμές ώστε να κατανοήσουμε περίπου τι διάστημα είναι επιθυμητό για την κάθε παράμετρο.

Συγκεκριμένα, για το πλήθος των hidden layers ορίσαμε δυνατές τιμές 1 ή 2, δεδομένου ότι θεωρητικά δύο hidden layers επαρκούν για την προσέγγιση οποιασδήποτε συνάρτησης με MLP και ότι συνήθως χρησιμοποιείται ένα layer για ταξινόμηση γραμμικά διαχωρίσιμων δεδομένων. Όσον αφορά το πλήθος των νευρώνων ανά layer, στο input και το output καθορίζονται από τις διαστάσεις των δεδομένων και των κλάσεων αντιστοίχως, ενώ στα hidden layers επιλέξαμε διάστημα από 10 ως 100, ούτως ώστε να είναι συγκρίσιμο με το μέγεθος του input, αλλά όχι τόσο μεγάλο που να προκαλείται overfitting.

Επιπλέον, για να αποφύγουμε το overfitting ενεργοποιήσαμε την παράμετρο early stopping, ώστε να σταματάει νωρίτερα όταν δεν παρατηρείται σημαντική βελτίωση στην απόδοση.

Σημειώνουμε επίσης ότι αυτή τη φορά χρησιμοποιήσαμε μόνο ένα pipeline (αντί για τα τρία διαφορετικά με τα οποία πειραματιστήκαμε στο UCI dataset), επειδή η εκπαίδευση διαρκεί πολύ περισσότερη ώρα και επειδή στο UCI παρατηρήσαμε ότι όλα τα pipelines που δοκιμάσαμε παρήγαγαν παρόμοια αποτελέσματα.

Βλέπουμε ότι το trial 33 παρήγαγε τα καλύτερα αποτελέσματα. Θα κρατήσουμε τις τιμές των παραμέτρων του και θα τις χρησιμοποιήσουμε παρακάτω για να το δοκιμάσουμε στο test set και να εξάγουμε τα τελικά αποτελέσματά μας.

Παρόμοια διαδικασία θα ακολουθήσουμε και για τον Linear SVC παρακάτω.

SVM (Linear SVC)

In [10]:

```
%%time

import optuna
from imblearn.pipeline import Pipeline
from sklearn.feature_selection import VarianceThreshold
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import RandomOverSampler
from sklearn.decomposition import PCA
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report, f1_score
import warnings

warnings.filterwarnings("ignore")

def objective(trial):

    x_train,x_val,y_train,y_val = train_test_split(train_data,train_labels,test_size=0.25)

    vthreshold = trial.suggest_float("vthreshold", 0.01, 0.05) #προσαρμόζουμε τις τιμές μας στο variance που παρατηρήσαμε
    n_components = trial.suggest_int("n_components", 5, 50)

    selector = VarianceThreshold(threshold = vthreshold)
    std_scaler = StandardScaler()
    ros = RandomOverSampler()
    pca = PCA(n_components = n_components)

    C = trial.suggest_loguniform('C', 1e-4, 10)
    max_iter = trial.suggest_int('max_iter',1000, 20000)
    penalty = trial.suggest_categorical("penalty", ["l1", "l2"])


    # Construct the model
    # Select the algorithm to either solve the dual or primal optimization problem. Prefer dual=False when n_samples > n_f
    clf = LinearSVC(C=C, dual=False, max_iter=max_iter, penalty=penalty)

    pipe = Pipeline(steps=[('selector', selector), ('scaler', std_scaler), ('sampler', ros), ('pca', pca), ('svc', clf)],
```

```
    # Train the model
    pipe.fit(x_train,y_train)

    # Evaluate the model
    pred = pipe.predict(x_val)
    metric = f1_score(y_val, pred)

    print("Validation Score:",pipe.score(x_val, y_val))
    print("Train Score:",pipe.score(x_train, y_train))
    print("\n=================")
    return metric

study = optuna.create_study(direction='maximize')
# start tuning for the hyper-parameters
study.optimize(objective, n_trials=100)
```

```
[I 2021-11-25 16:53:52,333] A new study created in memory with name: no-name-dc7e48f2-4ad7-43df-bc5d-4cf09b9353a5
Validation Score: 0.8033336125303627
Train Score: 0.8031660943127565

=================
[I 2021-11-25 16:56:21,146] Trial 0 finished with value: 0.7425156267134555 and parameters: {'vthreshold': 0.0457675677472
475, 'n_components': 50, 'C': 0.0016241487526447568, 'max_iter': 9534, 'penalty': 'l2'}. Best is trial 0 with value: 0.742
5156267134555.
Validation Score: 0.7693274143563112
[I 2021-11-25 16:58:55,312] Trial 1 finished with value: 0.6944752607055691 and parameters: {'vthreshold': 0.0215048597941
96814, 'n_components': 32, 'C': 0.000556500712998072, 'max_iter': 5279, 'penalty': 'l2'}. Best is trial 0 with value: 0.74
25156267134555.
Train Score: 0.7674149147053075

=================
Validation Score: 0.794999581204456
Train Score: 0.8013373537705559

=================
[I 2021-11-25 17:01:31,734] Trial 2 finished with value: 0.7331698010357046 and parameters: {'vthreshold': 0.0184095154207
90085, 'n_components': 49, 'C': 1.8239507425879995, 'max_iter': 19568, 'penalty': 'l2'}. Best is trial 0 with value: 0.742
5156267134555.
Validation Score: 0.7692855348019098
Train Score: 0.7684758634168133

=================
[I 2021-11-25 17:04:01,871] Trial 3 finished with value: 0.6988136241867585 and parameters: {'vthreshold': 0.0158171017809
5378, 'n_components': 37, 'C': 0.059460620372265134, 'max_iter': 12435, 'penalty': 'l2'}. Best is trial 0 with value: 0.74
25156267134555.
Validation Score: 0.7617053354552308
Train Score: 0.7607002261495938

=================
[I 2021-11-25 17:06:28,088] Trial 4 finished with value: 0.6500615006150062 and parameters: {'vthreshold': 0.0318731820244
40325, 'n_components': 10, 'C': 0.00041867631097779466, 'max_iter': 19812, 'penalty': 'l1'}. Best is trial 0 with value:
 0.7425156267134555.
Validation Score: 0.7651394589161571
Train Score: 0.7659910098556552

=================
[I 2021-11-25 17:08:48,912] Trial 5 finished with value: 0.6885828520657485 and parameters: {'vthreshold': 0.0401721113877
3092, 'n_components': 18, 'C': 1.1746912267499214, 'max_iter': 17091, 'penalty': 'l1'}. Best is trial 0 with value: 0.7425
156267134555.
Validation Score: 0.7732222129156546
Train Score: 0.7755255884077393

=================
[I 2021-11-25 17:11:17,347] Trial 6 finished with value: 0.6981100518481351 and parameters: {'vthreshold': 0.0464474428576
1667, 'n_components': 30, 'C': 0.024857168522304057, 'max_iter': 9279, 'penalty': 'l1'}. Best is trial 0 with value: 0.742
5156267134555.
Validation Score: 0.803668648965575
[I 2021-11-25 17:13:44,453] Trial 7 finished with value: 0.7484168723838145 and parameters: {'vthreshold': 0.0310227748441
53903, 'n_components': 47, 'C': 1.2965326400701036, 'max_iter': 4465, 'penalty': 'l2'}. Best is trial 7 with value: 0.7484
168723838145.
Train Score: 0.802663539659938

=================
Validation Score: 0.7818075215679705
[I 2021-11-25 17:16:15,696] Trial 8 finished with value: 0.7103302568664517 and parameters: {'vthreshold': 0.0215438668778
93295, 'n_components': 35, 'C': 0.6810354900921363, 'max_iter': 2712, 'penalty': 'l2'}. Best is trial 7 with value: 0.7484
168723838145.
```

```
Train Score: 0.7766004969707122

=================
Validation Score: 0.7785827958790519
[I 2021-11-25 17:18:43,837] Trial 9 finished with value: 0.713115198871344 and parameters: {'vthreshold': 0.04252680412176
596, 'n_components': 33, 'C': 2.6434623955920227, 'max_iter': 7167, 'penalty': 'l2'}. Best is trial 7 with value: 0.748416
8723838145.
Train Score: 0.7775637267219476

=================
Validation Score: 0.8000251277326409
[I 2021-11-25 17:21:36,487] Trial 10 finished with value: 0.7412625304795448 and parameters: {'vthreshold': 0.034524872105
96533, 'n_components': 43, 'C': 0.09772049824922562, 'max_iter': 1302, 'penalty': 'l1'}. Best is trial 7 with value: 0.748
4168723838145.
Train Score: 0.8011837954044169

=================
Validation Score: 0.8038780467375827
Train Score: 0.8028868972834129

=================
[I 2021-11-25 17:24:00,208] Trial 11 finished with value: 0.7446425650253558 and parameters: {'vthreshold': 0.049177224311
59557, 'n_components': 49, 'C': 0.0030611033942632565, 'max_iter': 12984, 'penalty': 'l2'}. Best is trial 7 with value: 0.
7484168723838145.
Validation Score: 0.7759862635061563
[I 2021-11-25 17:26:33,428] Trial 12 finished with value: 0.7063409278067526 and parameters: {'vthreshold': 0.010890467960
346747, 'n_components': 42, 'C': 0.007384693184784823, 'max_iter': 10789, 'penalty': 'l2'}. Best is trial 7 with value: 0.
7484168723838145.
Train Score: 0.7810397297372756

=================
Validation Score: 0.7699556076723344
Train Score: 0.7677499511405199

=================
[I 2021-11-25 17:29:12,285] Trial 13 finished with value: 0.6954930982870448 and parameters: {'vthreshold': 0.027230913404
243218, 'n_components': 22, 'C': 7.722673571492917, 'max_iter': 14092, 'penalty': 'l2'}. Best is trial 7 with value: 0.748
4168723838145.
Validation Score: 0.8050506742608259
[I 2021-11-25 17:31:37,808] Trial 14 finished with value: 0.7450013694878116 and parameters: {'vthreshold': 0.036329379857
73022, 'n_components': 44, 'C': 0.0030931852598399044, 'max_iter': 14671, 'penalty': 'l2'}. Best is trial 7 with value: 0.
7484168723838145.
Train Score: 0.8031242147583549

=================
Validation Score: 0.8015327916910964
Train Score: 0.8029985760951504

=================
[I 2021-11-25 17:34:08,237] Trial 15 finished with value: 0.7408824976761986 and parameters: {'vthreshold': 0.037146666128
31764, 'n_components': 42, 'C': 0.20732937069628604, 'max_iter': 15894, 'penalty': 'l2'}. Best is trial 7 with value: 0.74
84168723838145.
Validation Score: 0.7653488566881649
[I 2021-11-25 17:36:39,525] Trial 16 finished with value: 0.6911416129210076 and parameters: {'vthreshold': 0.026312335542
614276, 'n_components': 25, 'C': 0.016992468346488192, 'max_iter': 6225, 'penalty': 'l2'}. Best is trial 7 with value: 0.7
484168723838145.
Train Score: 0.7709048775721027

=================
Validation Score: 0.787880056956194
Train Score: 0.7882290532428735

=================
[I 2021-11-25 17:39:09,844] Trial 17 finished with value: 0.7205825564075687 and parameters: {'vthreshold': 0.030299945143
049754, 'n_components': 40, 'C': 0.00010465513051839594, 'max_iter': 3563, 'penalty': 'l2'}. Best is trial 7 with value:
 0.7484168723838145.
Validation Score: 0.8032917329759611
Train Score:
[I 2021-11-25 17:42:08,247] Trial 18 finished with value: 0.7419656100642751 and parameters: {'vthreshold': 0.037711108689
5971, 'n_components': 45, 'C': 0.15694933428797905, 'max_iter': 7805, 'penalty': 'l1'}. Best is trial 7 with value: 0.7484
168723838145.
0.8023983024820616

=================
Validation Score: 0.7476756847307144
Train Score: 0.7523941145266214
```

```
=================
[I 2021-11-25 17:44:37,749] Trial 19 finished with value: 0.6513511949539956 and parameters: {'vthreshold': 0.026132128898
997194, 'n_components': 9, 'C': 0.42783434031614, 'max_iter': 15988, 'penalty': 'l2'}. Best is trial 7 with value: 0.74841
68723838145.
Validation Score: 0.7859954770081247
Train Score: 0.7904486696261552

=================
[I 2021-11-25 17:47:07,341] Trial 20 finished with value: 0.723005203816132 and parameters: {'vthreshold': 0.0343136877847
20846, 'n_components': 38, 'C': 0.0073070905322427625, 'max_iter': 4760, 'penalty': 'l2'}. Best is trial 7 with value: 0.7
484168723838145.
Validation Score: 0.805678867576849
Train Score: 0.8013373537705559

=================
[I 2021-11-25 17:49:41,870] Trial 21 finished with value: 0.7500807928471399 and parameters: {'vthreshold': 0.048712171334
555736, 'n_components': 47, 'C': 0.002100698245062733, 'max_iter': 13123, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Validation Score: 0.8050087947064243
Train Score:
[I 2021-11-25 17:52:07,571] Trial 22 finished with value: 0.7473410028217929 and parameters: {'vthreshold': 0.042561149178
76047, 'n_components': 45, 'C': 0.001278227008904285, 'max_iter': 11999, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
 0.802817098026077

=================
Validation Score: 0.800653321048664
Train Score: 0.8027891783231427

=================
[I 2021-11-25 17:54:31,708] Trial 23 finished with value: 0.741051028179741 and parameters: {'vthreshold': 0.0494959099282
87576, 'n_components': 47, 'C': 0.0006436083741883895, 'max_iter': 12039, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Validation Score: 0.8035848898567719
[I 2021-11-25 17:56:55,869] Trial 24 finished with value: 0.7434354485776806 and parameters: {'vthreshold': 0.042825378540
33068, 'n_components': 45, 'C': 0.00011811067505141862, 'max_iter': 11024, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Train Score: 0.8032917329759611

=================
Validation Score: 0.7884244911634141
Train Score: 0.7910349833877768

=================
[I 2021-11-25 17:59:22,348] Trial 25 finished with value: 0.7239042518308012 and parameters: {'vthreshold': 0.044088025600
98683, 'n_components': 39, 'C': 0.0011915312522500791, 'max_iter': 8652, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.804673758271212
Train Score: 0.8026077002540692

=================
[I 2021-11-25 18:01:54,404] Trial 26 finished with value: 0.7443262800131565 and parameters: {'vthreshold': 0.040844981244
13622, 'n_components': 50, 'C': 0.009578088466246235, 'max_iter': 14198, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.7687829801490912
Train Score: 0.7700254069296703

=================
[I 2021-11-25 18:04:24,085] Trial 27 finished with value: 0.6896745545500533 and parameters: {'vthreshold': 0.046727168877
11735, 'n_components': 28, 'C': 9.923166177150227, 'max_iter': 10971, 'penalty': 'l1'}. Best is trial 21 with value: 0.750
0807928471399.
Validation Score: 0.7682385459418711
Train Score: 0.7689784180696317
[I 2021-11-25 18:06:51,192] Trial 28 finished with value: 0.683445829996568 and parameters: {'vthreshold': 0.0396675603797
8203, 'n_components': 15, 'C': 0.00021697843263314355, 'max_iter': 17572, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.

=================
Validation Score: 0.8001926459502471
Train Score: 0.8041851634698607

=================
[I 2021-11-25 18:09:15,525] Trial 29 finished with value: 0.7395884504120955 and parameters: {'vthreshold': 0.045157589866
76377, 'n_components': 46, 'C': 0.002095318836370198, 'max_iter': 9881, 'penalty': 'l2'}. Best is trial 21 with value: 0.7
500807928471399.
```

```
Validation Score: 0.8057207471312505
Train Score: 0.8041991233213278

=================
[I 2021-11-25 18:11:44,303] Trial 30 finished with value: 0.7463502651867243 and parameters: {'vthreshold': 0.049676289849
92181, 'n_components': 50, 'C': 0.048445948270252656, 'max_iter': 12909, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.8030823352039534
Train Score: 0.8037524080743781

=================
[I 2021-11-25 18:14:10,671] Trial 31 finished with value: 0.7453147004658217 and parameters: {'vthreshold': 0.047777324339
52334, 'n_components': 48, 'C': 0.04862682561724629, 'max_iter': 13238, 'penalty': 'l2'}. Best is trial 21 with value: 0.7
500807928471399.
Validation Score: 0.8058882653488567
Train Score: 0.8025239411452663

=================
[I 2021-11-25 18:16:37,889] Trial 32 finished with value: 0.7484669235361154 and parameters: {'vthreshold': 0.044926125264
78925, 'n_components': 50, 'C': 0.0010427648340292565, 'max_iter': 12142, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Validation Score: 0.7984755842197839
Train Score: 0.8024541418879303

=================
[I 2021-11-25 18:19:03,933] Trial 33 finished with value: 0.7380796864794252 and parameters: {'vthreshold': 0.044850424390
0171, 'n_components': 41, 'C': 0.0009368325020981573, 'max_iter': 11640, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.7475081665131083
[I 2021-11-25 18:21:27,785] Trial 34 finished with value: 0.6414084339499198 and parameters: {'vthreshold': 0.042856540697
83723, 'n_components': 5, 'C': 0.0002760196292949644, 'max_iter': 15350, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Train Score: 0.7472568891866991

=================
Validation Score: 0.772175224055616
Train Score: 0.7800764999860401

=================
[I 2021-11-25 18:24:03,406] Trial 35 finished with value: 0.7041227020559121 and parameters: {'vthreshold': 0.022486032362
10217, 'n_components': 36, 'C': 0.003444644002358164, 'max_iter': 10347, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.7993131753078148
[I 2021-11-25 18:26:27,826] Trial 36 finished with value: 0.7378842577398534 and parameters: {'vthreshold': 0.039183885634
65979, 'n_components': 47, 'C': 0.0004902167789677302, 'max_iter': 8357, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Train Score: 0.8037105285199766

=================
Validation Score: 0.7932825194739928
[I 2021-11-25 18:29:02,337] Trial 37 finished with value: 0.7333909473911634 and parameters: {'vthreshold': 0.015545129112
1633, 'n_components': 50, 'C': 0.0019493262689891369, 'max_iter': 18124, 'penalty': 'l1'}. Best is trial 21 with value: 0.
7500807928471399.
Train Score: 0.7966049641231817

=================
Validation Score: 0.8050925538152274
[I 2021-11-25 18:31:29,123] Trial 38 finished with value: 0.7466521502449646 and parameters: {'vthreshold': 0.032199381054
59355, 'n_components': 45, 'C': 0.013360590789635116, 'max_iter': 9458, 'penalty': 'l2'}. Best is trial 21 with value: 0.7
500807928471399.
Train Score: 0.8024960214423319

=================
Validation Score: 0.7789597118686657
[I 2021-11-25 18:33:54,962] Trial 39 finished with value: 0.7101910828025477 and parameters: {'vthreshold': 0.041912195169
417896, 'n_components': 34, 'C': 0.0009234598442942891, 'max_iter': 5820, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Train Score: 0.781695842756233

=================
Validation Score: 0.7783315185526426
[I 2021-11-25 18:36:18,225] Trial 40 finished with value: 0.7021551966687299 and parameters: {'vthreshold': 0.047224597628
139985, 'n_components': 31, 'C': 0.004179864230712423, 'max_iter': 13833, 'penalty': 'l1'}. Best is trial 21 with value:
 0.7500807928471399.
Train Score: 0.7758327051400173
```

```
==================
Validation Score: 0.8037942876287796
[I 2021-11-25 18:38:45,320] Trial 41 finished with value: 0.744728382280826 and parameters: {'vthreshold': 0.0323079753053827, 'n_components': 44, 'C': 0.01672237833271665, 'max_iter': 9408, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Train Score: 0.8021749448585868

==================
Validation Score: 0.8045062400536058
[I 2021-11-25 18:41:13,419] Trial 42 finished with value: 0.7427532238509864 and parameters: {'vthreshold': 0.02882080681849701, 'n_components': 47, 'C': 0.005512405490904135, 'max_iter': 12062, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Train Score: 0.8037803277773125

==================
Validation Score: 0.8006952006030655
Train Score: 0.8031102549068878

==================
[I 2021-11-25 18:43:40,014] Trial 43 finished with value: 0.7412321244086781 and parameters: {'vthreshold': 0.035607914170280204, 'n_components': 43, 'C': 0.0013681476871712814, 'max_iter': 4136, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Validation Score: 0.7830639082000167
[I 2021-11-25 18:46:09,385] Trial 44 finished with value: 0.7182945399173374 and parameters: {'vthreshold': 0.024046810722713634, 'n_components': 40, 'C': 0.010709217790596998, 'max_iter': 2237, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Train Score: 0.7862607141860011

==================
Validation Score: 0.803207973867158
[I 2021-11-25 18:48:36,311] Trial 45 finished with value: 0.7423935091277891 and parameters: {'vthreshold': 0.03247065978595822, 'n_components': 48, 'C': 4.532754706419532, 'max_iter': 6834, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Train Score: 0.8048133567858838

==================
Validation Score: 0.7987268615461931
Train Score: 0.7971493983304018

==================
[I 2021-11-25 18:51:09,586] Trial 46 finished with value: 0.7382067763372917 and parameters: {'vthreshold': 0.029913804710819607, 'n_components': 45, 'C': 0.022799354765264537, 'max_iter': 13180, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Validation Score: 0.7856185610185108
Train Score: 0.7861490353742636

==================
[I 2021-11-25 18:53:36,415] Trial 47 finished with value: 0.7206243519074388 and parameters: {'vthreshold': 0.03816785814835145, 'n_components': 36, 'C': 0.0002840482325585395, 'max_iter': 1091, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Validation Score: 0.7877125387385878
Train Score: 0.7871681045313678

==================
[I 2021-11-25 18:56:04,833] Trial 48 finished with value: 0.7284221805518349 and parameters: {'vthreshold': 0.01899879054290686, 'n_components': 43, 'C': 0.09499173796149793, 'max_iter': 15272, 'penalty': 'l1'}. Best is trial 21 with value: 0.7500807928471399.
Validation Score: 0.7993969344166179
[I 2021-11-25 18:58:26,605] Trial 49 finished with value: 0.7375342465753425 and parameters: {'vthreshold': 0.047843306585821144, 'n_components': 41, 'C': 1.399800284739482, 'max_iter': 11488, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Train Score: 0.8029287768378145

==================
Validation Score: 0.7876706591841863
Train Score: 0.7875171008180473

==================
[I 2021-11-25 19:00:48,277] Trial 50 finished with value: 0.7252330370691524 and parameters: {'vthreshold': 0.04431542085144956, 'n_components': 38, 'C': 0.390719917236961 53, 'max_iter': 10083, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Validation Score: 0.8026216601055365
[I 2021-11-25 19:03:09,612] Trial 51 finished with value: 0.744234004449992 and parameters: {'vthreshold': 0.049963304287087035, 'n_components': 49, 'C': 2.6608004565034706, 'max_iter': 12302, 'penalty': 'l2'}. Best is trial 21 with value: 0.7500807928471399.
Train Score: 0.8036407292626406
```

```
=================
[I 2021-11-25 19:05:32,213] Trial 52 finished with value: 0.7485892729962198 and parameters: {'vthreshold': 0.045688706419
22693, 'n_components': 50, 'C': 0.059982811996686806, 'max_iter': 13161, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.8078147248513275
Train Score: 0.8029427366892816

=================
Validation Score: 0.8037942876287796
Train Score: 0.8037942876287796

=================
[I 2021-11-25 19:07:54,046] Trial 53 finished with value: 0.7454219420746617 and parameters: {'vthreshold': 0.046157973077
78689, 'n_components': 46, 'C': 0.6051655302875785, 'max_iter': 14671, 'penalty': 'l2'}. Best is trial 21 with value: 0.75
00807928471399.
Validation Score: 0.8023285032247257
[I 2021-11-25 19:10:15,440] Trial 54 finished with value: 0.7457444516268045 and parameters: {'vthreshold': 0.041364750543
302596, 'n_components': 49, 'C': 0.002184743643817788, 'max_iter': 13618, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Train Score: 0.803738448222911

=================
Validation Score: 0.8009464779294748
Train Score: 0.8047575173800151

=================
[I 2021-11-25 19:12:40,034] Trial 55 finished with value: 0.7409103297901335 and parameters: {'vthreshold': 0.034062888505
50372, 'n_components': 47, 'C': 0.0006348191567386251, 'max_iter': 12645, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Validation Score: 0.7998576095150347
[I 2021-11-25 19:15:02,698] Trial 56 finished with value: 0.7389523133227727 and parameters: {'vthreshold': 0.043907178115
823293, 'n_components': 44, 'C': 0.15272832292956515, 'max_iter': 9088, 'penalty': 'l2'}. Best is trial 21 with value: 0.7
500807928471399.
Train Score: 0.8039897255493201

=================
Validation Score: 0.8022866236703241
[I 2021-11-25 19:17:29,640] Trial 57 finished with value: 0.7427668501062495 and parameters: {'vthreshold': 0.027745075722
878862, 'n_components': 48, 'C': 0.013011887284200526, 'max_iter': 7444, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Train Score: 0.8035848898567719

=================
Validation Score: 0.7671077979730295
Train Score: 0.7687411005946897

=================
[I 2021-11-25 19:19:51,934] Trial 58 finished with value: 0.6854460093896714 and parameters: {'vthreshold': 0.048604288716
035546, 'n_components': 20, 'C': 0.03537468731937294, 'max_iter': 16809, 'penalty': 'l1'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.8002764050590502
Train Score: 0.8019795069380462

=================
[I 2021-11-25 19:22:15,116] Trial 59 finished with value: 0.7430080293150831 and parameters: {'vthreshold': 0.045646944448
548554, 'n_components': 42, 'C': 0.007435869866039599, 'max_iter': 10813, 'penalty': 'l2'}. Best is trial 21 with value:
 0.7500807928471399.
Validation Score: 0.8025379009967334
Train Score: 0.8035430103023704

=================
[I 2021-11-25 19:24:35,718] Trial 60 finished with value: 0.7403205375337336 and parameters: {'vthreshold': 0.030516786137
395658, 'n_components': 45, 'C': 0.081626116560459 98, 'max_iter': 14473, 'penalty': 'l2'}. Best is trial 21 with value: 0.
7500807928471399.
Validation Score: 0.8098249434626016
Train Score: 0.8027333389172739

=================
[I 2021-11-25 19:27:03,984] Trial 61 finished with value: 0.7526823157780077 and parameters: {'vthreshold': 0.048310157848
46824, 'n_components': 50, 'C': 0.06035689862313827, 'max_iter': 12891, 'penalty': 'l2'}. Best is trial 61 with value: 0.7
526823157780077.
Validation Score: 0.8011977552558841
[I 2021-11-25 19:29:36,056] Trial 62 finished with value: 0.7436963446898117 and parameters: {'vthreshold': 0.046832375534
61912, 'n_components': 50, 'C': 0.033709689424692345, 'max_iter': 11399, 'penalty': 'l2'}. Best is trial 61 with value: 0.
7526823157780077.
```

```
Train Score: 0.8044085210933356

=================
Validation Score: 0.8044643604992043
[I 2021-11-25 19:31:59,863] Trial 63 finished with value: 0.7492885142028674 and parameters: {'vthreshold': 0.043356719030
50683, 'n_components': 46, 'C': 0.21006220125864836, 'max_iter': 12264, 'penalty': 'l2'}. Best is trial 61 with value: 0.7
526823157780077.
Train Score: 0.802272663818857

=================
Validation Score: 0.8033336125303627
[I 2021-11-25 19:34:23,751] Trial 64 finished with value: 0.7456672443674177 and parameters: {'vthreshold': 0.043076665857
04406, 'n_components': 48, 'C': 0.2874042301593087, 'max_iter': 12400, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Train Score: 0.8031660943127565

=================
Validation Score: 0.8002764050590502
[I 2021-11-25 19:36:49,338] Trial 65 finished with value: 0.7389852772152592 and parameters: {'vthreshold': 0.048270481511
29445, 'n_components': 50, 'C': 0.8184944150628654, 'max_iter': 13920, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Train Score: 0.8031940140156909

=================
Validation Score: 0.8054694698048412
[I 2021-11-25 19:39:14,561] Trial 66 finished with value: 0.7467146518348874 and parameters: {'vthreshold': 0.046193879152
10111, 'n_components': 46, 'C': 0.17583993083504307, 'max_iter': 13269, 'penalty': 'l2'}. Best is trial 61 with value: 0.7
526823157780077.
Train Score: 0.8018678281263087

=================
Validation Score: 0.8043387218359996
[I 2021-11-25 19:41:40,829] Trial 67 finished with value: 0.7455614856769415 and parameters: {'vthreshold': 0.039872075147
26948, 'n_components': 48, 'C': 0.06909815502374052, 'max_iter': 11701, 'penalty': 'l2'}. Best is trial 61 with value: 0.7
526823157780077.
Train Score: 0.8027891783231427

=================
Validation Score: 0.8022866236703241
Train Score: 0.8042968422815981

=================
[I 2021-11-25 19:44:00,881] Trial 68 finished with value: 0.744077627798558 and parameters: {'vthreshold': 0.0450010117036
6801, 'n_components': 46, 'C': 0.10797510443653213, 'max_iter': 15265, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Validation Score: 0.7628779629784739
[I 2021-11-25 19:46:20,367] Trial 69 finished with value: 0.6807982861652948 and parameters: {'vthreshold': 0.043365100467
5189, 'n_components': 15, 'C': 0.05078561617103085, 'max_iter': 12689, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Train Score: 0.7651673786190915

=================
Validation Score: 0.7693692939107127
[I 2021-11-25 19:48:41,676] Trial 70 finished with value: 0.6937322729547856 and parameters: {'vthreshold': 0.041732855449
69769, 'n_components': 25, 'C': 0.33108627392253515, 'max_iter': 15935, 'penalty': 'l1'}. Best is trial 61 with value: 0.7
526823157780077.
Train Score: 0.7680570678727978

=================
Validation Score: 0.8009883574838764
[I 2021-11-25 19:51:02,791] Trial 71 finished with value: 0.7398445198729882 and parameters: {'vthreshold': 0.045969869975
62067, 'n_components': 46, 'C': 0.2215721892968297, 'max_iter': 13496, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Train Score: 0.8033615322332971

=================
Validation Score: 0.8016165507998995
Train Score: 0.8036407292626406

=================
[I 2021-11-25 19:53:24,644] Trial 72 finished with value: 0.7428478367081048 and parameters: {'vthreshold': 0.048380041854
098575, 'n_components': 43, 'C': 0.12667176440204767, 'max_iter': 13128, 'penalty': 'l2'}. Best is trial 61 with value: 0.
7526823157780077.
Validation Score: 0.8031660943127565
[I 2021-11-25 19:55:44,865] Trial 73 finished with value: 0.7464393612429867 and parameters: {'vthreshold': 0.047016120385
95106, 'n_components': 49, 'C': 0.2040401862054987, 'max_iter': 10544, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
```

```
26823157780077.
Train Score: 0.8040874445095905

=================
Validation Score: 0.8009045983750733
[I 2021-11-25 19:58:06,603] Trial 74 finished with value: 0.7404455121205503 and parameters: {'vthreshold': 0.045333800226
10954, 'n_components': 47, 'C': 0.5620448654127024, 'max_iter': 12001, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Train Score: 0.8043108021330653

=================
Validation Score: 0.8023285032247257
Train Score: 0.8045201999050731

=================
[I 2021-11-25 20:00:28,797] Trial 75 finished with value: 0.7432270699597433 and parameters: {'vthreshold': 0.040714834963
4901, 'n_components': 50, 'C': 0.8459669514000678, 'max_iter': 14193, 'penalty': 'l2'}. Best is trial 61 with value: 0.752
6823157780077.
Validation Score: 0.8029566965407489
[I 2021-11-25 20:02:50,666] Trial 76 finished with value: 0.7458268056831073 and parameters: {'vthreshold': 0.048927206346
12254, 'n_components': 44, 'C': 0.02486983592638244, 'max_iter': 11424, 'penalty': 'l2'}. Best is trial 61 with value: 0.7
526823157780077.
Train Score: 0.803975765697853

=================
Validation Score: 0.8014071530278918
Train Score:
[I 2021-11-25 20:05:14,663] Trial 77 finished with value: 0.7443114418203387 and parameters: {'vthreshold': 0.038692802835
41284, 'n_components': 46, 'C': 1.8206367848595004, 'max_iter': 14894, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
 0.802202864561521

=================
Validation Score: 0.7797973029566966
[I 2021-11-25 20:07:46,476] Trial 78 finished with value: 0.7099514563106795 and parameters: {'vthreshold': 0.010791195719
535279, 'n_components': 41, 'C': 0.0008865408250141619, 'max_iter': 12845, 'penalty': 'l2'}. Best is trial 61 with value:
 0.7526823157780077.
Train Score: 0.7745204791021023

=================
Validation Score: 0.8031660943127565
[I 2021-11-25 20:10:11,312] Trial 79 finished with value: 0.746521410851041 and parameters: {'vthreshold': 0.0369600290789
3337, 'n_components': 48, 'C': 0.06985283039726803, 'max_iter': 16531, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Train Score: 0.803975765697853

=================
Validation Score: 0.7846553312672753
Train Score: 0.7905045090320238

=================
[I 2021-11-25 20:12:36,241] Trial 80 finished with value: 0.7168502202643172 and parameters: {'vthreshold': 0.042753511478
02794, 'n_components': 39, 'C': 0.00038841831434172314, 'max_iter': 11032, 'penalty': 'l2'}. Best is trial 61 with value:
 0.7526823157780077.
Validation Score: 0.8035848898567719
Train Score: 0.8027612586202083

=================
[I 2021-11-25 20:15:00,120] Trial 81 finished with value: 0.7436878347360367 and parameters: {'vthreshold': 0.047423622291
73778, 'n_components': 45, 'C': 0.002430666542150064, 'max_iter': 9830, 'penalty': 'l2'}. Best is trial 61 with value: 0.7
526823157780077.
Validation Score: 0.8025379009967334
Train Score:
[I 2021-11-25 20:17:29,610] Trial 82 finished with value: 0.7444028839377677 and parameters: {'vthreshold': 0.024517280590
250136, 'n_components': 49, 'C': 0.0014267130184234015, 'max_iter': 1928, 'penalty': 'l2'}. Best is trial 61 with value:
 0.7526823157780077.
 0.8042410028757294

=================
Validation Score: 0.802119105452718
Train Score: 0.8028310578775442

=================
[I 2021-11-25 20:19:53,728] Trial 83 finished with value: 0.7435271128480703 and parameters: {'vthreshold': 0.031091291854
58086, 'n_components': 43, 'C': 0.18052485577530009, 'max_iter': 5149, 'penalty': 'l2'}. Best is trial 61 with value: 0.75
26823157780077.
Validation Score: 0.7993550548622163
```

[I 2021-11-25 20:22:16,765] Trial 84 finished with value: 0.7404518121241 67 and parameters: {'vthreshold': 0.044167445087737796, 'n_components': 47, 'C': 0.004055668027555821, 'max_iter': 12064, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Train Score: 0.8040316051037217

==========
Validation Score: 0.8047156378256135
[I 2021-11-25 20:24:38,744] Trial 85 finished with value: 0.749179710612662 and parameters: {'vthreshold': 0.033484095343336 37, 'n_components': 45, 'C': 0.0446456092507219, 'max_iter': 19231, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Train Score: 0.8038640868861155

==========
Validation Score: 0.8026216601055365
Train Score: 0.802817098026077

==========
[I 2021-11-25 20:27:00,096] Trial 86 finished with value: 0.7446220536439989 and parameters: {'vthreshold': 0.033401213227576505, 'n_components': 50, 'C': 0.037300490585687185, 'max_iter': 19336, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Validation Score: 0.79755423402295
Train Score: 0.7998994890694363

==========
[I 2021-11-25 20:29:21,130] Trial 87 finished with value: 0.7175081813931743 and parameters: {'vthreshold': 0.046000703349 84958, 'n_components': 42, 'C': 0.00015681747496025427, 'max_iter': 19360, 'penalty': 'l1'}. Best is trial 61 with value: 0.7526823157780077.
Validation Score: 0.8005276823854595
[I 2021-11-25 20:31:41,435] Trial 88 finished with value: 0.7407608991454852 and parameters: {'vthreshold': 0.049109198084 17105, 'n_components': 44, 'C': 0.12830738426350657, 'max_iter': 18388, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Train Score: 0.8044783203506715

==========
Validation Score: 0.8007789597118686
Train Score: 0.8045062400536058

==========
[I 2021-11-25 20:34:03,401] Trial 89 finished with value: 0.7379785183145139 and parameters: {'vthreshold': 0.042167180545376354, 'n_components': 46, 'C': 0.023699040921878953, 'max_iter': 13329, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Validation Score: 0.8017003099087026
Train Score: 0.8039618058463858

==========
[I 2021-11-25 20:36:28,357] Trial 90 finished with value: 0.7408461496360353 and parameters: {'vthreshold': 0.035376631067 614406, 'n_components': 49, 'C': 0.0544003340957473, 'max_iter': 3050, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Validation Score: 0.8020353463439149
[I 2021-11-25 20:38:58,660] Trial 91 finished with value: 0.7434185528958366 and parameters: {'vthreshold': 0.029780882377 061637, 'n_components': 45, 'C': 4.81167863562718, 'max_iter': 6438, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Train Score: 0.8026495798084708

==========
Validation Score: 0.8039618058463858
Train Score: 0.8042828824301309

==========
[I 2021-11-25 20:41:24,839] Trial 92 finished with value: 0.7477229857181353 and parameters: {'vthreshold': 0.033255095865 64556, 'n_components': 47, 'C': 0.017553455115657982, 'max_iter': 7881, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Validation Score: 0.8012396348102856
[I 2021-11-25 20:43:49,206] Trial 93 finished with value: 0.7425130208333331 and parameters: {'vthreshold': 0.033014898730 35402, 'n_components': 48, 'C': 0.0179733407350821, 'max_iter': 8365, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Train Score: 0.8035290504509032

==========
Validation Score: 0.8030823352039534
Train Score: 0.8040595248066561

==========
[I 2021-11-25 20:46:13,751] Trial 94 finished with value: 0.7417902251510159 and parameters: {'vthreshold': 0.031210979264 6711, 'n_components': 47, 'C': 0.042112319273584976, 'max_iter': 4330, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526823157780077.
Validation Score: 0.7918167350699389

```
[I 2021-11-25 20:48:47,350] Trial 95 finished with value: 0.7290269828291087 and parameters: {'vthreshold': 0.012407110353
444117, 'n_components': 49, 'C': 0.0009043459264445361, 'max_iter': 12445, 'penalty': 'l2'}. Best is trial 61 with value:
 0.7526823157780077.
Train Score: 0.7968004020437223

=================
Validation Score: 0.8011977552558841
Train Score: 0.8034732110450344

=================
[I 2021-11-25 20:51:11,085] Trial 96 finished with value: 0.744633923287966 and parameters: {'vthreshold': 0.0349734567011
0014, 'n_components': 44, 'C': 0.2746958240257374, 'max_iter': 8853, 'penalty': 'l2'}. Best is trial 61 with value: 0.7526
823157780077.
Validation Score: 0.8006114414942626
[I 2021-11-25 20:53:35,208] Trial 97 finished with value: 0.7416010854816825 and parameters: {'vthreshold': 0.028363010716
84807, 'n_components': 47, 'C': 0.0011196521156070455, 'max_iter': 13885, 'penalty': 'l2'}. Best is trial 61 with value:
 0.7526823157780077.
Train Score: 0.8045201999050731

=================
Validation Score: 0.7727196582628361
Train Score: 0.7766982159309825

=================
[I 2021-11-25 20:55:56,900] Trial 98 finished with value: 0.6998174677802975 and parameters: {'vthreshold': 0.044570687794
98448, 'n_components': 29, 'C': 0.0017351778578197485, 'max_iter': 10309, 'penalty': 'l2'}. Best is trial 61 with value:
 0.7526823157780077.
Validation Score: 0.8033336125303627
Train Score: 0.8017421894631042

=================
[I 2021-11-25 20:58:53,018] Trial 99 finished with value: 0.7443101382990308 and parameters: {'vthreshold': 0.046779834282
26952, 'n_components': 46, 'C': 0.07745918352136648, 'max_iter': 11019, 'penalty': 'l1'}. Best is trial 61 with value: 0.7
526823157780077.
Wall time: 4h 5min 1s
```

Ακολουθούμε παρόμοια διαδικασία με την προηγούμενη περίπτωση. Σημειώνουμε ότι επιλέξαμε η παράμετρος dual να έχει σταθερά την τιμή false επειδή έχουμε n_samples > n_features. Για την παράμετρο C προτιμήσαμε να δοκιμάσουμε κυρίως μικρές τιμές για να μην οδηγηθούμε σε overfitting.

Παρατηρούμε ότι το trial 61 παρήγαγε τα καλύτερα αποτελέσματα. Οι παράμετροί του θα χρησιμοποιηθούν παρακάτω για την εκπαίδευση του βέλτιστου ταξινομητή.

Εκ των υστέρων παρατηρήσαμε ότι πολλά από τα μοντέλα που είχαν καλές μετρικές είχαν μεγάλο αριθμό PCA components, το οποίο μας οδηγεί στο συμπέρασμα ότι ενδεχομένως να καταλήγαμε σε καλύτερα αποτελέσματα εάν αυξάναμε τις δυνατές τιμές της παραμέτρου n_components. Ωστόσο, δεδομένων των αρχικών διαστάσεων του dataset, και βλέποντας ότι η βελτιστοποίηση χρειάζεται περίπου 4 ώρες για να ολοκληρωθεί και οι αποκλίσεις ως προς τη μετρική f1 δεν είναι πολύ μεγάλες ανάμεσα στα trials, δεν κρίθηκε σκόπιμο να επαναλάβουμε όλη την διαδικασία για μια τέτοια αλλαγή. Επίσης, στις δοκιμές που πραγματοποιήσαμε πριν από την βελτιστοποίηση παρατηρήσαμε ότι κάποιες μεγαλύτερες τιμές της παραμέτρου n_components οδηγούσαν σε error.

Εκπαίδευση βέλτιστων ταξινομητών

Σε αυτό το σημείο θα εκπαιδεύσουμε τους δύο ταξινομητές μας, χρησιμοποιώντας τις βέλτιστες παραμέτρους που βρήκαμε από τα παραπάνω optimizations. Στη συνέχεια θα τους δοκιμάσουμε στο test set μας για να εξάγουμε τα τελικά μας αποτελέσματα, τα οποία θα παρουσιάσουμε και σε διαγράμματα.

MLP

In [15]:

```
from imblearn.pipeline import Pipeline
from sklearn.feature_selection import VarianceThreshold
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import RandomOverSampler
from sklearn.decomposition import PCA
from sklearn.svm import LinearSVC
from sklearn.metrics import classification_report, f1_score
import warnings
from sklearn.neural_network import MLPClassifier
from time import perf_counter

warnings.filterwarnings("ignore")

opt_selector = VarianceThreshold(threshold = 0.042326407214433884)
std_scaler = StandardScaler()
```

```
ros = RandomOverSampler()
opt_pca = PCA(n_components = 47)

opt_solver = "lbfgs"
opt_hlsize = 62
opt_depth = 2
opt_activation = "relu"
opt_alpha = 3.149263977958106e-05

opt_clf = MLPClassifier(solver = opt_solver, alpha = opt_alpha, activation = opt_activation, hidden_layer_sizes = (opt_hls
## random state for MLP

pipe = Pipeline(steps = [("selector", opt_selector), ("scaler", std_scaler), ("sampler", ros), ("pca", opt_pca), ("mlp", c

start = perf_counter()
pipe.fit(train_data, train_labels)
end = perf_counter()
opt_train_times.update({"MLP" : end - start})

start = perf_counter()
optMLPpred = pipe.predict(test_data)
end = perf_counter()
opt_test_times.update({"MLP" : end - start})

opt_f1scores.update({"MLP" : f1_score(test_labels, optMLPpred)})

print(classification_report(test_labels, optMLPpred))
```

```
              precision    recall  f1-score   support

           0       0.90      0.86      0.88     14983
           1       0.78      0.85      0.81      8895

    accuracy                           0.85     23878
   macro avg       0.84      0.85      0.85     23878
weighted avg       0.86      0.85      0.85     23878
```

SVM (Linear SVC)

In [16]:

```
from sklearn.svm import LinearSVC

warnings.filterwarnings("ignore")

opt_selector = VarianceThreshold(threshold = 0.04831015784846824)
std_scaler = StandardScaler()
ros = RandomOverSampler()

opt_pca = PCA(n_components = 50)

opt_C = 0.06035689862313827
opt_max_iter = 12891
opt_penalty = "l2"

clf = LinearSVC(C = opt_C, dual = False, max_iter = opt_max_iter, penalty = opt_penalty)

pipe = Pipeline(steps = [("selector", opt_selector), ("scaler", std_scaler), ("sampler", ros), ("pca", opt_pca), ("mlp", c

start = perf_counter()
pipe.fit(train_data, train_labels)
end = perf_counter()
opt_train_times.update({"Linear SVC" : end - start})

start = perf_counter()
optSVMpred = pipe.predict(test_data)
end = perf_counter()
opt_test_times.update({"Linear SVC" : end - start})

opt_f1scores.update({"Linear SVC" : f1_score(test_labels, optSVMpred)})

print(classification_report(test_labels, optSVMpred))
```

```
              precision    recall  f1-score   support

           0       0.91      0.84      0.88     14983
           1       0.77      0.86      0.81      8895

    accuracy                           0.85     23878
   macro avg       0.84      0.85      0.84     23878
```

```
weighted avg       0.86      0.85      0.85     23878
```

Αποτελέσματα

Συγκρίσεις

Παρακάτω φαίνονται τα αποτελέσματα μας από τις βελτιστοποιήσεις που κάναμε, καθώς και συγκρίσεις με τους αντίστοιχους out-of-the-box ταξινομητές.

F1 score για κάθε ταξινομητή, out of the box και βελτιστοποιημένο.

Classifier	F1-OotB	F1-optimal
Dummy	0.376	-
MLP	0.697	0.812
Linear SVC	0.581	0.810

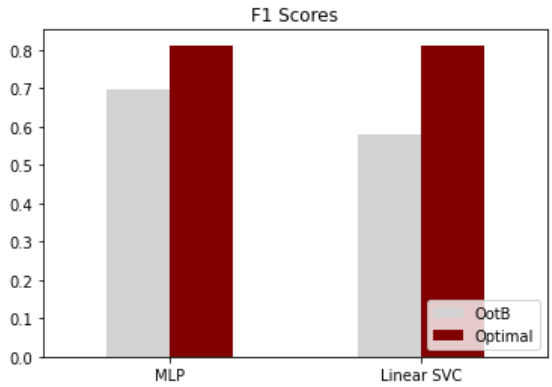
Χρόνοι training και testing για κάθε ταξινομητή, σε δευτερόλεπτα.

Classifier	Train time-OotB	Train time-optimal	Test time-OotB	Test time-optimal
Dummy	0.003	-	0.001	-
MLP	145.011	228.167	0.519	0.303
Linear SVC	20.387	223.365	0.367	0.233

In [19]:
```
final_f1 = {}
for i in ["MLP", "Linear SVC"]:
    final_f1.update({i : []})
    final_f1[i].append(ootb_f1scores[i])
    final_f1[i].append(opt_f1scores[i])
```

In [20]:
```
labels = list(final_f1.keys())
ootbScores = [i[0] for i in final_f1.values()]
optScores = [i[1] for i in final_f1.values()]
scoresDF = pd.DataFrame({"OotB" : ootbScores, "Optimal" : optScores}, index = labels)

## https://matplotlib.org/stable/gallery/color/named_colors.html

ax = scoresDF.plot.bar(rot = 0, color = {"OotB" : "lightgray", "Optimal" : "maroon"}, title = "F1 Scores")
ax.legend(loc = "lower right")
```

Out[20]:
```
<matplotlib.legend.Legend at 0x1e75a5d8850>
```



Συμπεράσματα και παρατηρήσεις

Παρατηρούμε ότι και οι δύο ταξινομητές που μελετήσαμε παρουσίασαν αρκετά βελτιωμένη απόδοση μετά τα optimizations μας. Η μεγαλύτερη βελτίωση παρατηρείται στον Linear SVC, ο οποίος ωστόσο έγινε πολύ πιο αργός σε σύγκριση με τον αντίστοιχο out-of-the-box ταξινομητή.

Μετά τις βελτιστοποιήσεις οι δύο ταξινομητές έχουν πολύ παρόμοιες επιδόσεις και ως προς τη μετρική f1 και ως προς ταχύτητα. Ο MLP παράγει ελάχιστα καλύτερα αποτελέσματα, αλλά είναι ελάχιστα πιο αργός. Αυτό έρχεται σε σύγκρουση με την επίδοση των αντίστοιχων out-of-the-box ταξινομητών, στην οποία περίπτωση η αντίστοιχη διαφορά ως προς f1 και ως προς ταχύτητα ήταν πάρα πολύ πιο αισθητή, όπως φαίνεται από τους παραπάνω πίνακες τιμών και το αντίστοιχο διάγραμμα.

Συνεπώς, δεν καταλήγουμε σε κάποιο ισχυρό συμπέρασμα σχετικά με το ποιος ταξινομητής είναι προτιμότερος. Και οι δύο φαίνεται να είναι σχεδόν εξίσου καλοί για το συγκεκριμένο πρόβλημα.

Παρακάτω δείχνουμε και τα confusion matrices των ταξινομητών στο test set μας, για να έχουμε μια πιο ολοκληρωμένη εικόνα για το πόσο ικανοποιητικά είναι τα αποτελέσματά μας.

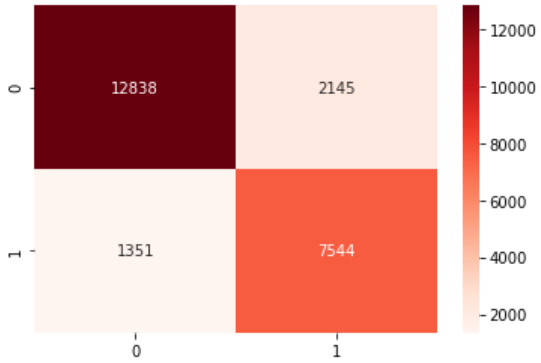
Βλέπουμε ότι και για τους δύο ταξινομητές η συντριπτική πλειοψηφία των δειγμάτων τοποθετείται στην διαγώνιο. Τα αποτελέσματα και πάλι φαίνονται πολύ παραπλήσια και για τους δύο, με τον MLP να τείνει να βγάζει λίγο περισσότερα false negatives και τον linear SVC λίγο περισσότερο false positives.

Confusion matrices

MLP

In [29]:
```
from sklearn.metrics import confusion_matrix
import seaborn as sns

## https://seaborn.pydata.org/generated/seaborn.heatmap.html
## https://matplotlib.org/stable/tutorials/colors/colormaps.html

cnf_matrix = confusion_matrix(test_labels, optMLPpred)

sns.heatmap(cnf_matrix, annot = True, cmap = "Reds", fmt = 'd')
```

Out[29]:
```
<AxesSubplot:>
```



SVM (Linear SVC)

In [30]:
```
cnf_matrix = confusion_matrix(test_labels, optSVMpred)

sns.heatmap(cnf_matrix, annot = True, cmap = "Reds", fmt = 'd')
```

Out[30]:
```
<AxesSubplot:>
```