Machine Learning DevOps Engineer Nanodegree Program Project: A Dynamic Risk Assessment System

## Installation Instructions

Uncompress the file "starter-file.zip".

Install Conda:  $\frac{\text{https://docs.conda.io/projects/conda/en/latest/user-guide/install/index.html}{\text{index.html}}$ 

Create a Conda Environment with Python 3.7 conda create -n MLDOE python=3.7 conda activate MLDOE

Install the requirements:
pip install -r requirements.txt

#### File Structure

```
$ tree --dirsfirst
 — ingesteddata
    - finaldata.csv
    ingestedfiles.txt
  - models
    - apireturns.txt
    - confusionmatrix.png
    ├─ latestscore.txt
   └─ trainedmodel.pkl
  - practicedata
    - dataset1.csv
    dataset2.csv
  - practicemodels
    \vdash apireturns.txt
    - confusionmatrix.png
    latestscore.txt
    └─ trainedmodel.pkl
  - production_deployment
    ingestedfiles.txt
    latestscore.txt

    sourcedata

   ├─ dataset3.csv
└─ dataset4.csv
 — testdata
   └─ testdata.csv
 apicalls.py
 — app.py
  - clear_results.py
 — config.json
 config.json.backup1
  - config.json.backup2
 — crontab.file
deployment.py
 — diagnostics.py
 execute_full_process_twice.sh
 — fullprocess.py
 — ingestion.py
  reporting.py
 - REPORT.odt
 - REPORT.pdf
 — requirements.txt
- scoring.py
```

```
training.py
use_configuration_1.sh
use_configuration_2.sh
wsgi.py

directories, 39 files
```

## Example to Execute (with drift\_must\_improve\_score = False)

Given that the practice model gives a better F1-score than the final model, it was necessary to tweak the Python script fullprocess.py in order to execute the final process twice and to obtain 2 confusion matrices and 2 apireturns.txt files:

```
drift_must_improve_score = False

def check_for_model_drift():
    score0 = scoring.read_f1_score(load_config()['prod_deployment_path'],
    'latestscore.txt')
    print(f'score0={score0}')
    os.system('python training.py')
    os.system('python scoring.py')
    score1 = scoring.read_f1_score(load_config()['output_model_path'], 'latestscore.txt')
    print(f'score0={score0}, score1={score1}')
    return (score1 > score0) if drift_must_improve_score else abs(score1 - score0) >
0.001
```

In this first execution of execute\_full\_process\_twice.sh, drift can be positive and negative, but not zero: (drift\_must\_improve\_score = False) In the second execution of execute\_full\_process\_twice.sh, drift can be only positive, greater than zero: (drift\_must\_improve\_score = True)

Important Note: The Server Side (python app.py) must be executed before executing the script "execute\_full\_process\_twice.sh". Otherwise the RESTful API won't be available.

# \$ ./execute\_full\_process\_twice.sh

```
==== EXECUTING execute_full_process_twice.sh =====
echo "==== EXECUTING execute_full_process_twice.sh ====="
cat execute_full_process_twice.sh
echo "==== EXECUTING python clear_results.py ====="
python clear_results.py
echo "==== EXECUTING python fullprocess.py (FOR THE FIRST TIME) ====="
python fullprocess.py
echo "==== EXECUTING ./use_configuration_2.sh ====="
./use_configuration_2.sh
echo "==== EXECUTING python fullprocess.py (FOR THE SECOND TIME) ====="
python fullprocess.py
==== EXECUTING python clear_results.py =====
==== EXECUTING python fullprocess.py (FOR THE FIRST TIME) =====
new_csv_files=['practicedata/dataset1.csv', 'practicedata/dataset2.csv']
There are new CSV files. Running the script ingestion.py.
_____
===== RUNNING PYTHON SCRIPT ingestion.py =====
csv_files=['practicedata/dataset1.csv', 'practicedata/dataset2.csv']
practicedata/dataset1.csv
  corporation lastmonth_activity lastyear_activity number_of_employees exited
0
                             100
                                              1359
                                                                     1
                                                                              0
         nciw
         lsid
                              68
                                                282
                                                                     14
                                                                              0
2
        pwls
                              71
                                                949
                                                                     40
                                                                              1
                                                                    103
3
        bqlx
                              686
                                               3782
                                                                              0
                                                655
                                                                      7
                                                                              0
4
         zmei
                              45
5
                               0
                                                                     21
                                                                              1
         wosl
                                                18
6
         xcvb
                             189
                                                961
                                                                     18
                                                                              1
7
         dfgh
                              16
                                               1028
                                                                     33
```

_	,	0	4 -	1	4
8	ngrd	9	45	1	1
9	xful	0	67	14	1
10	kshe	48	986	22	1
11	qqqq	52	650	11	1
12	corp	1090	2452	9	0
13	ekci	6	88	90	1
14	dosk	99	390	99	1
15	endi	75	800	81	1
16	gudj	255	1687	2	0
			1007	2	0
_	cticedata/da				
	corporation	lastmonth_activity		number_of_employees	exited
0	abcd	78	1024	12	1
1	asdf	14	2145	20	0
2	xyzz	182	3891	35	0
3	acme	101	10983	2	1
4	qwer	0	118	42	1
5	tyui	929	1992	1	0
6	ZXCV	19	455	8	1
7	hjkl	94	868	3	1
8	lmno	81	1401	10	0
9		52			
	qqqq		650	11	1
10	corp	1090	2452	9	0
11	ekci	6	88	90	1
12	dosk	99	390	99	1
13	endi	75	800	81	1
14	gudj	255	1687	2	0
15	wosl	0	18	21	1
16	xcvb	189	961	18	1
17	dfgh	16	1028	33	0
18	ngrd	9	45	1	1
	_		43	1	1
_	esteddata/fin		7 1 1 - 1 - 1 - 1		
	corporation	lastmonth_activity	lastyear_activity	number_of_employees	exited
0	corporation nciw	lastmonth_activity 100	1359	1	0
0	corporation nciw lsid	lastmonth_activity 100 68	1359 282	1 14	0
0	corporation nciw	lastmonth_activity 100 68 71	1359 282 949	1 14 40	0 0 1
0	corporation nciw lsid	lastmonth_activity 100 68	1359 282	1 14	0
0 1 2	corporation nciw lsid pwls	lastmonth_activity 100 68 71	1359 282 949	1 14 40	0 0 1
0 1 2 3	corporation nciw lsid pwls bqlx	lastmonth_activity 100 68 71 686	1359 282 949 3782	1 14 40 103	0 0 1 0
0 1 2 3 4	corporation nciw lsid pwls bqlx zmei wosl	lastmonth_activity 100 68 71 686 45	1359 282 949 3782 655 18	1 14 40 103 7 21	0 0 1 0 0
0 1 2 3 4 5	corporation nciw lsid pwls bqlx zmei wosl xcvb	lastmonth_activity 100 68 71 686 45 0	1359 282 949 3782 655 18 961	1 14 40 103 7 21 18	0 0 1 0 0
0 1 2 3 4 5 6 7	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh	lastmonth_activity	1359 282 949 3782 655 18 961 1028	1 14 40 103 7 21 18 33	0 0 1 0 0 1 1
0 1 2 3 4 5 6 7 8	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45	1 14 40 103 7 21 18 33	0 0 1 0 0 1 1 0
0 1 2 3 4 5 6 7 8	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45	1 14 40 103 7 21 18 33 1	0 0 1 0 0 1 1 0
0 1 2 3 4 5 6 7 8 9 10	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe	lastmonth_activity 100 68 71 686 45 0 189 16 9 0 48	1359 282 949 3782 655 18 961 1028 45 67	1 14 40 103 7 21 18 33 1 14	0 0 1 0 0 1 1 0 1
0 1 2 3 4 5 6 7 8 9 10	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986	1 14 40 103 7 21 18 33 1 14 22	0 0 1 0 0 1 1 0 1 1
0 1 2 3 4 5 6 7 8 9 10 11	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452	1 14 40 103 7 21 18 33 1 14 22 11	0 0 1 0 0 1 1 0 1 1 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452	1 14 40 103 7 21 18 33 1 14 22 11	0 0 1 0 0 1 1 1 1 1 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390	1 14 40 103 7 21 18 33 1 14 22 11	0 0 1 0 0 1 1 0 1 1 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452	1 14 40 103 7 21 18 33 1 14 22 11	0 0 1 0 0 1 1 1 1 1 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk	lastmonth_activity 100 68 71 686 45 0 189 16 9 0 48 52 1090 6	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390	1 14 40 103 7 21 18 33 1 14 22 11 9	0 0 1 0 0 1 1 1 1 1 0 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi	lastmonth_activity 100 68 71 686 45 0 189 16 9 0 48 52 1090 6 99 75	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800	1 14 40 103 7 21 18 33 1 14 22 11 9	0 0 1 0 0 1 1 1 1 0 1 1 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd	lastmonth_activity 100 68 71 686 45 0 189 16 9 0 48 52 1090 6 99 75 255 78	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99	0 0 1 0 0 1 1 1 1 0 1 1 1 1 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf	lastmonth_activity 100 68 71 686 45 0 189 16 9 0 48 52 1090 6 99 75 255 78	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2	0 0 1 0 0 1 1 1 1 0 1 1 1 0 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35	0 0 1 0 0 1 1 1 1 0 1 1 1 0 1 1 0 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35	0 0 1 0 0 1 1 1 1 0 1 1 1 0 1 1 0 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme qwer	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983 118	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35 2 42	0 0 1 0 0 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 0 0 0 1 1 0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme qwer tyui	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983 118 1992	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35 2 42	0 0 1 0 0 1 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1 0 0 1 1 0 0 1 0 0 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme qwer tyui zxcv	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983 118 1992 455	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35 2 42 1 8	0 0 1 0 0 1 1 1 1 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 0 0 0 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme qwer tyui zxcv hjkl	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983 118 1992 455 868	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35 2 42 1 8 3	0 0 1 0 0 1 1 1 1 1 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 25 26 27 27 27 27 27 27 27 27 27 27 27 27 27	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme qwer tyui zxcv hjkl lmno	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983 118 1992 455 868 1401	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35 2 42 1 8 3 10	0 0 1 0 0 1 1 1 1 0 1 1 1 0 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 0 0 1 0 0 0 1 0 0 0 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 25 26 27 27 27 27 27 27 27 27 27 27 27 27 27	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme qwer tyui zxcv hjkl lmno	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983 118 1992 455 868 1401	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35 2 42 1 8 3 10	0 0 1 0 0 1 1 1 1 1 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 1 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 Recoil Recoil Reco	corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe qqqq corp ekci dosk endi gudj abcd asdf xyzz acme qwer tyui zxcv hjkl lmno	lastmonth_activity	1359 282 949 3782 655 18 961 1028 45 67 986 650 2452 88 390 800 1687 1024 2145 3891 10983 118 1992 455 868 1401	1 14 40 103 7 21 18 33 1 14 22 11 9 90 99 81 2 12 20 35 2 42 1 8 3 10	0 0 1 0 0 1 1 1 1 1 0 1 1 1 0 0 1 1 1 0 0 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1

score0=0.0

csv\_file=ingesteddata/finaldata.csv
X=[[ 100 1359 1]
 [ 68 282 14]
 [ 71 949 40]
 [ 686 3782 103]
 [ 45 655 7]

```
0
        18
               21]
 ſ
              18]
  189
       961
   16 1028
              33]
 [
    9
        4.5
               11
 Γ
    0 67 14]
 [
   48 986 22]
 Γ
    52 650 11]
 [
 [ 1090 2452
               9]
   6 88 90]
 [
 [ 99
        390 99]
    75
        800
             81]
 Γ
 [ 255 1687
               21
[
    78
        1024
               121
    14 2145
 [
               20]
 [ 182 3891
              35]
 [ 101 10983
               21
             42]
    Ω
       118
 [
 [ 929 1992
               11
   19 455
               8]
Γ
   94 868
               31
[
[ 81 1401 10]]
Model was saved in the file "practicemodels/trainedmodel.pkl".
Test data loaded from file "testdata/testdata.csv".
X=[[ 234 3
               101
[ 14 2145 99]
[ 34 333 1000]
   101 12346
              2]
   0 675
[
               2511
Y = [1 \ 0 \ 0 \ 1 \ 1]
Model was loaded from file "practicemodels/trainedmodel.pkl".
f1 score: 0.5714285714285715
F1-score 0.5714285714285715 saved in file "practicemodels/latestscore.txt".
score0=0.0, score1=0.5714285714285715
There is drift. Running the scripts deployment.py, apicalls.py, and reporting.py.
______
===== RUNNING PYTHON SCRIPT deployment.py =====
File "practicemodels/trainedmodel.pkl" was successfully copied to
"production_deployment".
File "practicemodels/latestscore.txt" was successfully copied to "production_deployment".
File "ingesteddata/ingestedfiles.txt" was successfully copied to "production_deployment".
===== RUNNING PYTHON SCRIPT apicalls.py =====
INPUT 1: {'location': 'testdata/testdata.csv'}
OUTPUT 1: {"predictions": [0, 1, 1, 1, 1]}
OUTPUT 2: {'f1_score': 0.5714285714285715}
OUTPUT 3: {'exited': {'mean': 0.5769230769230769, 'median': 1.0, 'stdev':
0.4940474068717357}, 'lastmonth_activity': {'mean': 165.65384615384616, 'median': 73.0,
'stdev': 278.5174959713127}, 'lastyear_activity': {'mean': 1502.923076923077, 'median':
955.0, 'stdev': 2150.065274913888}, 'number_of_employees': {'mean': 26.884615384615383,
'median': 14.0, 'stdev': 30.745014509018585}}
OUTPUT 4: {'na_percentages': [0.0, 0.0, 0.0, 0.0], 'outdated_packages': {'Pillow':
['8.1.0', '8.1.0', '8.1.2'], 'pandas': ['1.2.2', '1.2.2', '1.2.3']}, 'times':
[1.0436301231384277, 0.46016383171081543]
The API returns were saved in file "practicemodels/apireturns.txt".
==== RUNNING PYTHON SCRIPT reporting.py =====
Test data loaded from file "testdata/testdata.csv".
X = [[234 3 10]
[ 14 2145 99]
  34 333 1000]
[
[ 101 12346
               2]
[ 0 675
               25]]
Y = [1 \ 0 \ 0 \ 1 \ 1]
Predictions:
[0, 1, 1, 1, 1]
tn=0, fp=2, fn=1, tp=2
Confusion matrix plot saved to the file "practicemodels/confusionmatrix.png".
```

```
==== EXECUTING ./use_configuration_2.sh =====
```

==== EXECUTING python fullprocess.py (FOR THE SECOND TIME) ==== new\_csv\_files=['sourcedata/dataset3.csv', 'sourcedata/dataset4.csv']

\_\_\_\_\_

There are new CSV files. Running the script ingestion.py.

===== RUNNING PYTHON SCRIPT ingestion.py =====

csv\_files=['sourcedata/dataset3.csv', 'sourcedata/dataset4.csv']

sourceda	ta/dat	taset 3	CSV

sou	ırcedata/data	set3.csv			
	corporation	lastmonth_activity	lastyear_activity	number_of_employees	exited
0	nciw	45	0	99	1
1	lsid	36	234	541	0
2	pwls	23	555	23	0
3	bqlx	15	11	190	1
4	zmei	100	2929	999	1
5	wosl	2	1	1359	0
6	xcvb	0	14	282	1
7		500	40	949	0
	dfgh				
8	ngrd	1234	103	3782	1
9	xful	98765	7	655	0
10	kshe	34	2345	18	1
sou	rcedata/data				
	corporation	lastmonth_activity	lastyear_activity	number_of_employees	exited
0	abcd	99	871	3	0
1	asdf	1243	0	10	0
2	xyzz	0	25	11	1
3	acme	813	129	9	1
4	qwer	2989	9982	90	0
5	tyui	395	190	99	0
6	ZXCV	19028	999	81	0
7	hjkl	345	78	2	1
8	lmno	1024	14	3	1
9		2145	182	110	1
	qqqq				
10	corp	3891	101	998	0
11	ekci	10983	0	1200	0
12	dosk	118	929	81	1
13	endi	1992	19	2	0
14	gudj	455	94	298	1
	5 5				
	gesteddata/fin				
			lastyear_activity	number_of_employees	exited
	gesteddata/fin	aldata.csv			
ing	gesteddata/fin corporation	aldata.csv lastmonth_activity	lastyear_activity	number_of_employees	exited
ing 0	gesteddata/fin corporation nciw lsid	aldata.csv lastmonth_activity 45	lastyear_activity 0	number_of_employees	exited
ing 0 1	gesteddata/fin corporation nciw lsid pwls	aldata.csv lastmonth_activity 45 36	lastyear_activity 0 234	number_of_employees 99 541 23	exited 1 0
ing 0 1 2	gesteddata/fin corporation nciw lsid	aldata.csv lastmonth_activity 45 36 23	lastyear_activity 0 234 555	number_of_employees 99 541	exited 1 0
ing 0 1 2 3 4	gesteddata/fin corporation nciw lsid pwls bqlx zmei	aldata.csv lastmonth_activity 45 36 23 15	lastyear_activity 0 234 555 11 2929	number_of_employees 99 541 23 190 999	exited
ing 0 1 2 3 4 5	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl	aldata.csv lastmonth_activity 45 36 23 15 100 2	lastyear_activity 0 234 555 11 2929	number_of_employees 99 541 23 190 999 1359	exited 1 0 0 1 1
ing 0 1 2 3 4 5 6	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb	aldata.csv lastmonth_activity 45 36 23 15 100 2 0	lastyear_activity 0 234 555 11 2929 1	number_of_employees 99 541 23 190 999 1359 282	exited
ing 0 1 2 3 4 5 6 7	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500	lastyear_activity 0 234 555 11 2929 1 14	number_of_employees 99 541 23 190 999 1359 282 949	exited
ing 0 1 2 3 4 5 6 7 8	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234	lastyear_activity 0 234 555 11 2929 1 14 40 103	number_of_employees 99 541 23 190 999 1359 282 949 3782	exited  1 0 0 1 1 0 1 0 1 1 0 1
ing 0 1 2 3 4 5 6 7 8 9	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765	lastyear_activity 0 234 555 11 2929 1 14 40 103 7	number_of_employees 99 541 23 190 999 1359 282 949 3782 655	exited  1 0 0 1 1 0 1 0 1 0 1 0
ing 0 1 2 3 4 5 6 7 8 9 10	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18	exited  1 0 0 1 1 0 1 0 1 0 1 0 1
ing 0 1 2 3 4 5 6 7 8 9 10 11	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10	exited  1 0 0 1 1 0 1 0 1 0 1 0 0 0 0
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 1 0 1 1 1 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 1 1 1 1 1 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 1 0 0 1 0 0 0 1 0
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982 190	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 1 1 1 1 1 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 0 1 0 0 1 0 0 1 0 0 0 1 0
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982 190	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982 190 999	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 1 0 1 1 0 1 1 1 0 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv hjkl lmno	aldata.csv lastmonth_activity  45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028 345	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982 190 999 78	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99 81	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 1 1 0 0 1 1 1 1 1 1 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv hjkl lmno	aldata.csv lastmonth_activity  45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028 345 1024 2145	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 255 129 9982 190 999 78 14 182	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99 81 2 3 110	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 1 1 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv hjkl lmno qqqq corp	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028 345 1024 2145 3891	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982 190 999 78 14 182 101	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99 81 22 3 110 998	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 1 1 0 0 1 1 1 0 0 0 1 1 0 0 0 0 1 1 1 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv hjkl lmno qqqq corp ekci	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028 345 1024 2145 3891 10983	lastyear_activity 0 234 555 11 2929 1 144 40 103 7 2345 871 0 255 129 9982 190 999 78 14 182 101 0	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99 81 22 3 110 998 1200	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 1 0 0 1 1 1 0 0 0 0 0 0 1 1 1 0
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv hjkl lmno qqqq corp ekci dosk	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028 345 1024 2145 3891 10983 118	lastyear_activity 0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982 190 999 78 14 182 101 0 929	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99 81 22 3 110 998 1200 81	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 1 0 0 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 1 0 1 1 1 1 1 0 1
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv hjkl lmno qqqq corp ekci dosk endi	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028 345 1024 2145 3891 10983 118 1992	lastyear_activity  0 234 555 11 2929 1 14 40 103 7 2345 871 0 25 129 9982 190 999 78 14 182 101 0 929 19	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99 81 2 3 110 998 1200 81	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 0 1 1 1 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0 1 0
ing 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25	gesteddata/fin corporation nciw lsid pwls bqlx zmei wosl xcvb dfgh ngrd xful kshe abcd asdf xyzz acme qwer tyui zxcv hjkl lmno qqqq corp ekci dosk endi gudj	aldata.csv lastmonth_activity 45 36 23 15 100 2 0 500 1234 98765 34 99 1243 0 813 2989 395 19028 345 1024 2145 3891 10983 118	lastyear_activity  0 234 555 11 2929 1 144 40 103 7 2345 871 0 25 129 9982 190 999 78 14 182 101 0 929 19	number_of_employees 99 541 23 190 999 1359 282 949 3782 655 18 3 10 11 9 90 99 81 2 3 110 998 1200 81 2 298	exited  1 0 0 1 1 0 1 0 1 0 1 0 1 0 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 1 1 0 0 1 1 1 1 1 0 1

```
==== TESTING FOR DRIFT =====
score0=0.5714285714285715
csv_file=ingesteddata/finaldata.csv
X = [[ 45 0 99]
[ 36 234 541]
[ 23 555 23]
  15 11 190]
[
[ 100 2929
            999]
[ 2 1 1359]
[ 0
       14 282]
[ 500
        40
            9491
       103 3782]
[ 1234
[98765
        7
 [ 34 2345
             18]
 [
   99
       871
              3]
        0
 [ 1243
              101
        25 11]
 0 ]
 [ 813 129
              91
[ 2989 9982 90]
[ 395 190 99]
[19028 999 81]
[ 345 78 2]
[ 1024
        14
              31
[ 2145 182 110]
       101
            998]
[ 3891
        0 1200]
[10983
             81]
        929
[ 118
              2]
[ 1992
        19
       94 298]]
[ 455
Y=[1 0 0 1 1 0 1 0 1 0 1 0 1 0 0 1 1 0 0 0 1 1 1 0 0 1 0 1]
Model was saved in the file "models/trainedmodel.pkl".
Test data loaded from file "testdata/testdata.csv".
X = [[234 3 10]
[ 14 2145 99]
[ 34 333 1000]
[ 101 12346 2]
[ 0 675
              25]]
Y = [1 \ 0 \ 0 \ 1 \ 1]
Model was loaded from file "models/trainedmodel.pkl".
f1 score: 0.333333333333333333
______
There is drift. Running the scripts deployment.py, apicalls.py, and reporting.py.
==== RUNNING PYTHON SCRIPT deployment.py =====
File "models/trainedmodel.pkl" was successfully copied to "production_deployment".
File "models/latestscore.txt" was successfully copied to "production_deployment".
File "ingesteddata/ingestedfiles.txt" was successfully copied to "production_deployment".
==== RUNNING PYTHON SCRIPT apicalls.py =====
INPUT 1: {'location': 'testdata/testdata.csv'}
OUTPUT 1: {"predictions": [0, 1, 1, 0, 1]}
OUTPUT 3: {'exited': {'mean': 0.5, 'median': 0.5, 'stdev': 0.5}, 'lastmonth_activity':
{'mean': 5625.923076923077, 'median': 425.0, 'stdev': 19067.170236829497},
'lastyear_activity': {'mean': 763.5384615384615, 'median': 97.5, 'stdev':
1977.4481658240022}, 'number_of_employees': {'mean': 457.46153846153845, 'median': 99.0,
'stdev': 785.0576280057029}}
OUTPUT 4: {'na_percentages': [0.0, 0.0, 0.0, 0.0], 'outdated_packages': {'Pillow':
['8.1.0', '8.1.0', '8.1.2'], 'pandas': ['1.2.2', '1.2.2', '1.2.3']}, 'times':
[1.0206544399261475, 0.46021509170532227]}
The API returns were saved in file "models/apireturns.txt".
===== RUNNING PYTHON SCRIPT reporting.py =====
Test data loaded from file "testdata/testdata.csv".
X=[[ 234 3 10]
[ 14 2145
   34 333 1000]
```

```
[ 101 12346
   0 675
                25]]
 [
Y = [1 \ 0 \ 0 \ 1 \ 1]
Predictions:
[0, 1, 1, 0, 1]
tn=0, fp=2, fn=2, tp=1
Confusion matrix plot saved to the file "models/confusionmatrix.png".
Server Side
$ python app.py
 * Serving Flask app "app" (lazy loading)
 * Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
 * Debug mode: on
 * Running on http://0.0.0.0:8000/ (Press CTRL+C to quit)
* Restarting with stat
 * Debugger is active!
* Debugger PIN: 166-233-122
X = [[234 3 10]
[ 14 2145
                991
   34 333 1000]
Γ
[ 101 12346
                 2]
    0 675
[
                25]]
Y = [1 \ 0 \ 0 \ 1 \ 1]
127.0.0.1 - - [16/Mar/2021 05:27:00] "POST /prediction HTTP/1.1" 200 -
Test data loaded from file "testdata/testdata.csv".
X = [[234 3 10]
[ 14 2145 99]
Γ
   34 333 1000]
[ 101 12346
                 21
 [ 0 675
                25]]
Y = [1 \ 0 \ 0 \ 1 \ 1]
Model was loaded from file "practicemodels/trainedmodel.pkl".
f1 score: 0.5714285714285715
F1-score 0.5714285714285715 saved in file "practicemodels/latestscore.txt".
127.0.0.1 - - [16/Mar/2021 05:27:00] "GET /scoring HTTP/1.1" 200 -
Test data loaded from file "ingesteddata/finaldata.csv".
127.0.0.1 - - [16/Mar/2021 05:27:00] "GET /summarystats HTTP/1.1" 200 -
Test data loaded from file "ingesteddata/finaldata.csv".
WARNING: pip is being invoked by an old script wrapper. This will fail in a future
version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the underlying
To avoid this problem you can invoke Python with '-m pip' instead of running pip
directly.
WARNING: pip is being invoked by an old script wrapper. This will fail in a future
version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the underlying
issue.
To avoid this problem you can invoke Python with '-m pip' instead of running pip
directly.
127.0.0.1 - - [16/Mar/2021 05:27:11] "GET /diagnostics HTTP/1.1" 200 -
X=[[ 234 3
                101
[ 14 2145 99]
   34 333 1000]
Γ
[ 101 12346
                21
     0 675
                25]]
Y = [1 \ 0 \ 0 \ 1 \ 1]
127.0.0.1 - - [16/Mar/2021 05:27:19] "POST /prediction HTTP/1.1" 200 -
Test data loaded from file "testdata/testdata.csv".
X=[[ 234 3
                101
[ 14 2145
```

333 1000]

34

```
[ 101 12346
    0
       675
                25]]
 [
Y = [1 \ 0 \ 0 \ 1 \ 1]
Model was loaded from file "models/trainedmodel.pkl".
f1 score: 0.333333333333333333
127.0.0.1 - - [16/Mar/2021 05:27:19] "GET /scoring HTTP/1.1" 200 -
Test data loaded from file "ingesteddata/finaldata.csv".
127.0.0.1 - - [16/Mar/2021 05:27:19] "GET /summarystats HTTP/1.1" 200 -
Test data loaded from file "ingesteddata/finaldata.csv".
WARNING: pip is being invoked by an old script wrapper. This will fail in a future
version of pip.
Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the underlying
To avoid this problem you can invoke Python with '-m pip' instead of running pip
```

directly.

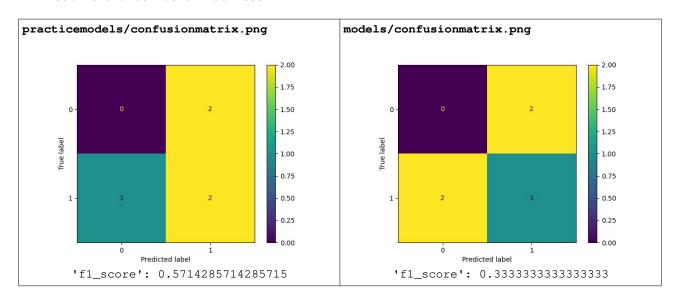
WARNING: pip is being invoked by an old script wrapper. This will fail in a future version of pip.

Please see https://github.com/pypa/pip/issues/5599 for advice on fixing the underlying

To avoid this problem you can invoke Python with '-m pip' instead of running pip

127.0.0.1 - - [16/Mar/2021 05:27:30] "GET /diagnostics HTTP/1.1" 200 -

### API Returns and Confusion Matrices



## \$ cat practicemodels/apireturns.txt

```
INPUT 1: {'location': 'testdata/testdata.csv'}
OUTPUT 1: {"predictions": [0, 1, 1, 1, 1]}
OUTPUT 2: {'f1_score': 0.5714285714285715}
OUTPUT 3: {'exited': {'mean': 0.5769230769230769, 'median': 1.0, 'stdev':
0.4940474068717357}, 'lastmonth_activity': {'mean': 165.65384615384616, 'median': 73.0,
'stdev': 278.5174959713127}, 'lastyear_activity': {'mean': 1502.923076923077, 'median':
955.0, 'stdev': 2150.065274913888}, 'number_of_employees': {'mean': 26.884615384615383,
'median': 14.0, 'stdev': 30.745014509018585}}
OUTPUT 4: {'na_percentages': [0.0, 0.0, 0.0, 0.0], 'outdated_packages': {'Pillow':
['8.1.0', '8.1.0', '8.1.2'], 'pandas': ['1.2.2', '1.2.2', '1.2.3']}, 'times':
[1.0436301231384277, 0.46016383171081543]}
```

## \$ cat models/apireturns.txt

```
INPUT 1: {'location': 'testdata/testdata.csv'}
OUTPUT 1: {"predictions": [0, 1, 1, 0, 1]}
OUTPUT 3: {'exited': {'mean': 0.5, 'median': 0.5, 'stdev': 0.5}, 'lastmonth_activity':
{'mean': 5625.923076923077, 'median': 425.0, 'stdev': 19067.170236829497},
```

```
'lastyear_activity': {'mean': 763.5384615384615, 'median': 97.5, 'stdev': 1977.4481658240022}, 'number_of_employees': {'mean': 457.46153846153845, 'median': 99.0, 'stdev': 785.0576280057029}}

OUTPUT 4: {'na_percentages': [0.0, 0.0, 0.0, 0.0], 'outdated_packages': {'Pillow': ['8.1.0', '8.1.0', '8.1.2'], 'pandas': ['1.2.2', '1.2.2', '1.2.3']}, 'times': [1.0206544399261475, 0.46021509170532227]}
```

## Example to Execute (with drift\_must\_improve\_score = True)

In the first execution of execute\_full\_process\_twice.sh, drift could be positive and negative, but not zero: (drift\_must\_improve\_score = False) In this second execution of execute\_full\_process\_twice.sh, drift can be only positive, greater than zero: (drift\_must\_improve\_score = True). As a result, only in the first execution of "python fullprocess.py", drift is generated. And drift is not generated in the second execution of "python fullprocess.py". In other words, there is only 1 apireturns.txt file and 1 confusion matrix, not 2 pairs.

```
drift_must_improve_score = True

def check_for_model_drift():
    score0 = scoring.read_f1_score(load_config()['prod_deployment_path'],
    'latestscore.txt')
    print(f'score0={score0}')
    os.system('python training.py')
    os.system('python scoring.py')
    score1 = scoring.read_f1_score(load_config()['output_model_path'], 'latestscore.txt')
    print(f'score0={score0}, score1={score1}')
    return (score1 > score0) if drift_must_improve_score else abs(score1 - score0) >
0.001
```

```
$ ./execute_full_process_twice.sh
==== EXECUTING execute_full_process_twice.sh =====
echo "===== EXECUTING execute_full_process_twice.sh ====="
cat execute_full_process_twice.sh
echo "===== EXECUTING python clear_results.py ======"
python clear_results.py
echo "==== EXECUTING python fullprocess.py (FOR THE FIRST TIME) ====="
python fullprocess.py
echo "==== EXECUTING ./use_configuration_2.sh ====="
./use_configuration_2.sh
echo "==== EXECUTING python fullprocess.py (FOR THE SECOND TIME) ====="
python fullprocess.py
==== EXECUTING python clear_results.py =====
==== EXECUTING python fullprocess.py (FOR THE FIRST TIME) =====
new_csv_files=['practicedata/dataset1.csv', 'practicedata/dataset2.csv']
______
There are new CSV files. Running the script ingestion.py.
===== RUNNING PYTHON SCRIPT ingestion.py =====
csv_files=['practicedata/dataset1.csv', 'practicedata/dataset2.csv']
practicedata/dataset1.csv
  corporation lastmonth_activity lastyear_activity number_of_employees exited
0
         nciw
                             100
                                               1359
                                                                      1
                                                                              0
                                                                     14
                                                                              Ω
1
         lsid
                              68
                                                282
                              71
                                                949
                                                                     40
2
         pwls
                                                                              1
                              686
                                               3782
                                                                    103
3
                                                                              0
         balx
4
                              45
                                                655
                                                                      7
         zmei
5
         wosl
                               0
                                                 18
                                                                     21
6
         xcvb
                              189
                                                961
                                                                     18
                                                                              1
                                               1028
                                                                     33
7
         dfgh
                              16
                                                                              Ω
8
                               9
         ngrd
                                                 45
                                                                      1
                                                                              1
9
                               0
                                                 67
                                                                     14
                                                                              1
         xful
```

986

650

22

11

1

1

48

52

10

11

kshe

qqqq

1.0		1000	0.450		0		
12	corp	1090	2452	9	0		
13	ekci	6	88	90	1		
14	dosk	99	390	99	1		
15	endi	75	800	81	1		
16	gudj	255	1687	2	0		
practicedata/dataset2.csv							
_	corporation	lastmonth_activity	lastyear_activity	number_of_employees	exited		
0	abcd	78	1024	12	1		
1	asdf	14	2145	20	0		
2	xyzz	182	3891	35	0		
3	acme	101	10983	2	1		
4		0	118	42	1		
	qwer	929			0		
5	tyui		1992	1			
6	ZXCV	19	455	8	1		
7	hjkl	94	868	3	1		
8	lmno	81	1401	10	0		
9	qqqq	52	650	11	1		
10	corp	1090	2452	9	0		
11	ekci	6	88	90	1		
12	dosk	99	390	99	1		
13	endi	75	800	81	1		
14	gudj	255	1687	2	0		
15	wosl	0	18	21	1		
16	xcvb	189	961	18	1		
17	dfgh	16	1028	33	0		
18	ngrd	9	45	1	1		
	ngid esteddata/fin		45	1	_		
_			1				
	corporation	lastmonth_activity	lastyear_activity		exited		
0	nciw	100	1359	1	0		
1	lsid	68	282	14	0		
2	pwls	71	949	40	1		
3	bqlx	686	3782	103	0		
4	zmei	45	655	7	0		
5	wosl	0	18	21	1		
6	xcvb	189	961	18	1		
7	dfgh	16	1028	33	0		
8	ngrd	9	45	1	1		
9	xful	0	67	14	1		
10	kshe	48	986	22	1		
11	qqqq	52	650	11	1		
12	corp	1090	2452	9	0		
13	ekci	6	88	90	1		
14	dosk	99	390	99	1		
15	endi	75	800	81	1		
16		255	1687	2	0		
	gudj						
17	abcd	78	1024	12	1		
18	asdf	14	2145	20	0		
19	xyzz	182	3891	35	0		
20	acme	101	10983	2	1		
21	qwer	0	118	42	1		
22	tyui	929	1992	1	0		
23	ZXCV	19	455	8	1		
24	hjkl	94	868	3	1		
25	lmno	81	1401	10	0		
		tion was saved in fil					
		OR DRIFT =====	-				

==== TESTING FOR DRIFT =====

score0=0.0

csv\_file=ingesteddata/finaldata.csv

	-5 v_	1116-1	riigesce	daaca, iii
2	K=[[	100	1359	1]
	[	68	282	14]
	[	71	949	40]
	[	686	3782	103]
	[	45	655	7]
	[	0	18	21]
	[	189	961	18]
	[	16	1028	33]
	[	9	45	1]

```
0
         67
               14]
 ſ
        986
    48
               221
   52 650
              11]
 [
 [ 1090 2452
                91
   6 88 90]
 Γ
   99 390 99]
 ſ
    75 800 81]
 [
 [ 255 1687
               2]
 [ 78 1024 12]
 [ 14 2145 20]
 [ 182 3891 35]
 [ 101 10983
                21
 [
    0
        118
              421
 [ 929 1992
                1]
 [
    19
         455
                8]
        868
 [
    94
                31
    81 1401 10]]
 [
Model was saved in the file "practicemodels/trainedmodel.pkl".
Test data loaded from file "testdata/testdata.csv".
X = [[234 3 10]
[ 14 2145 99]
[ 34 333 1000]
[ 101 12346
              21
[ 0 675
                25]]
Y = [1 \ 0 \ 0 \ 1 \ 1]
Model was loaded from file "practicemodels/trainedmodel.pkl".
f1 score: 0.5714285714285715
F1-score 0.5714285714285715 saved in file "practicemodels/latestscore.txt".
score0=0.0, score1=0.5714285714285715
There is drift. Running the scripts deployment.py, apicalls.py, and reporting.py.
==== RUNNING PYTHON SCRIPT deployment.py =====
File "practicemodels/trainedmodel.pkl" was successfully copied to
"production_deployment".
File "practicemodels/latestscore.txt" was successfully copied to "production_deployment".
File "ingesteddata/ingestedfiles.txt" was successfully copied to "production_deployment".
==== RUNNING PYTHON SCRIPT apicalls.py =====
INPUT 1: {'location': 'testdata/testdata.csv'}
OUTPUT 1: {"predictions": [0, 1, 1, 1, 1]}
OUTPUT 2: {'f1_score': 0.5714285714285715}
OUTPUT 3: {'exited': {'mean': 0.5769230769230769, 'median': 1.0, 'stdev':
0.4940474068717357}, 'lastmonth_activity': {'mean': 165.65384615384616, 'median': 73.0,
'stdev': 278.5174959713127}, 'lastyear_activity': {'mean': 1502.923076923077, 'median':
955.0, 'stdev': 2150.065274913888}, 'number_of_employees': {'mean': 26.884615384615383,
'median': 14.0, 'stdev': 30.745014509018585}}
OUTPUT 4: { 'na_percentages': [0.0, 0.0, 0.0, 0.0], 'outdated_packages': { 'Pillow':
['8.1.0', '8.1.0', '8.1.2'], 'pandas': ['1.2.2', '1.2.2', '1.2.3']}, 'times':
[1.035158634185791, 0.4606971740722656]}
The API returns were saved in file "practicemodels/apireturns.txt".
==== RUNNING PYTHON SCRIPT reporting.py =====
Test data loaded from file "testdata/testdata.csv".
X=[[ 234 3
                101
[ 14 2145 99]
   34
       333 1000]
Γ
[ 101 12346
               2.1
   0 675
                2511
Γ
Y = [1 \ 0 \ 0 \ 1 \ 1]
Predictions:
[0, 1, 1, 1, 1]
tn=0, fp=2, fn=1, tp=2
Confusion matrix plot saved to the file "practicemodels/confusionmatrix.png".
==== EXECUTING ./use_configuration_2.sh =====
==== EXECUTING python fullprocess.py (FOR THE SECOND TIME) ====
new_csv_files=['sourcedata/dataset3.csv', 'sourcedata/dataset4.csv']
```

There are new CSV files. Running the script ingestion.py.

\_\_\_\_\_

==== RUNNING PYTHON SCRIPT ingestion.py =====

csv\_files=['sourcedata/dataset3.csv', 'sourcedata/dataset4.csv']

sourcedata/dataset3.csv

corporation	lastmonth_activity	lastvear activity	number_of_employees	exited
0 nciw	45	0	99	1
1 lsid	36	234	541	0
2 pwls	23	555	23	0
3 bqlx	15	11	190	1
4 zmei	100	2929	999	1
5 wosl	2	1	1359	0
6 xcvb	0	14	282	1
7 dfgh	500	40	949	0
8 ngrd	1234	103	3782	1
9 xful	98765	7	655	0
10 kshe	34	2345	18	1
sourcedata/data		2313	10	_
corporation	lastmonth_activity	lastyear_activity	number_of_employees	exited
0 abcd	99	871	3	0
1 asdf	1243	0	10	0
2 xyzz	0	25	11	1
3 acme	813	129	9	1
4 qwer	2989	9982	90	0
5 tyui	395	190	99	0
6 zxcv	19028	999	81	0
7 hjkl	345	78	2	1
8 lmno	1024	14	3	1
9 qqqq	2145	182	110	1
10 corp	3891	101	998	0
11 ekci	10983	0	1200	0
12 dosk	118	929	81	1
13 endi	1992	19	2	0
14 gudj	455	94	298	1
ingesteddata/fir		31	230	_
corporation	lastmonth_activity	lastyear_activity	number_of_employees	exited
0 nciw	45	0	99	1
1 lsid	36	234	541	0
2 pwls	23	555	23	0
3 bqlx	15	11	190	1
4 zmei	100	2929	999	1
5 wosl	2	1	1359	0
6 xcvb	0	14	282	1
7 dfgh	500	40	949	0
8 ngrd	1234	103	3782	1
9 xful	98765	7	655	0
10 kshe	34	2345	18	1
11 abcd	99	871	3	0
12 asdf	1243	0	10	0
13 xyzz	0	25	11	1
14 acme	813	129	9	1
15 qwer	2989	9982	90	0
16 tyui	395	190	99	0
17 zxcv	19028	999	81	0
18 hjkl	345	78	2	1
19 lmno	1024	14	3	1
20 qqqq	2145	182	110	1
21 corp	3891	101	998	0
22 ekci	10983	0	1200	0
23 dosk	118	929	81	1
24 endi	1992	19	2	0
25 gudj	455	94	298	1
		e "ingesteddata/ing		-

==== TESTING FOR DRIFT =====

score0=0.5714285714285715

csv\_file=ingesteddata/finaldata.csv

X = [[ 45 0 99]

```
[ 36
      234
          541]
  23
      555
          23]
[
  15
     11
          190]
[
[ 100 2929
          999]
  2
      1 1359]
[
  0
     14
[
          282]
[ 500
     40
          949]
[98765
      7
          655]
[ 34 2345
          18]
[ 99
     871
           3]
[ 1243
       0
           10]
[ 0
[ 813
          11]
       25
      129
            9]
          90]
[ 2989 9982
          99]
[ 395
      190
          81]
     999
[19028
          2]
[ 345
      78
[ 1024
      14
           3]
[ 2145  182  110]
[ 3891 101 998]
[10983
      0 1200]
[ 118 929 81]
[ 1992 19
           2]
[ 455 94
         298]]
Model was saved in the file "models/trainedmodel.pkl".
Test data loaded from file "testdata/testdata.csv".
X=[[ 234 3 10]
[ 14 2145 99]
[ 34 333 1000]
[ 101 12346
          2]
[ 0 675
           25]]
Y = [1 \ 0 \ 0 \ 1 \ 1]
Model was loaded from file "models/trainedmodel.pkl".
fl score: 0.333333333333333333
There is no drift. Process ended.
```