Best Model Selection and Hyperparamter Tunning

In [1]:
```
# Import the loan data as a data frame and ensure that the data is loaded properly.

# load package first
import numpy as np
import pandas as pd

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler

# load the data frame
loanset = pd.read_csv('Loan_Train.csv')

#Check load is successful
loanset.head()
```

Out[1]:

	Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Tern
0	LP001002	Male	No	0	Graduate	No	5849	0.0	NaN	360.
1	LP001003	Male	Yes	1	Graduate	No	4583	1508.0	128.0	360.
2	LP001005	Male	Yes	0	Graduate	Yes	3000	0.0	66.0	360.
3	LP001006	Male	Yes	0	Not Graduate	No	2583	2358.0	120.0	360.
4	LP001008	Male	No	0	Graduate	No	6000	0.0	141.0	360.

In [2]:
```
# prep the data
# drop column "Load_ID"
loan = loanset.drop(['Loan_ID'], axis=1)

#Find out rows with missing data
#noempty = pd.notnull(loan)
#cleanloan = loan[noempty]
#cleanloan
cleanloan = loan.dropna()
```

In [3]:
```
print('variables with NA values', cleanloan.isna().sum())
cleanloan.shape
```

```
variables with NA values Gender              0
Married              0
Dependents           0
Education            0
Self_Employed        0
ApplicantIncome      0
CoapplicantIncome    0
LoanAmount           0
Loan_Amount_Term     0
Credit_History       0
Property_Area        0
Loan_Status          0
dtype: int64
```

Out[3]:
```
(480, 12)
```

In [4]:
```
#Convert the categorical features into dummy variables.
cat = cleanloan.select_dtypes(exclude=np.number)
print(cat.keys())
```

```
Index(['Gender', 'Married', 'Dependents', 'Education', 'Self_Employed',
       'Property_Area', 'Loan_Status'],
      dtype='object')
```

In [5]:
```
newdf = pd.get_dummies(cleanloan, columns=cat.keys(), drop_first=True)
newdf.shape
```

Out[5]:
```
(480, 15)
```

In [6]:
```
newdf.head()
```

Out[6]:

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Gender_Male	Married_Yes	Dependents_1	Depende
1	4583	1508.0	128.0	360.0	1.0	1	1	1	
2	3000	0.0	66.0	360.0	1.0	1	1	0	
3	2583	2358.0	120.0	360.0	1.0	1	1	0	
4	6000	0.0	141.0	360.0	1.0	1	0	0	
5	5417	4196.0	267.0	360.0	1.0	1	1	0	

In [6]:
```
newdf.rename(columns = {'Loan_Status_Y':'Loan_Status'}, inplace = True)
```

In [8]:
```
newdf
```

Out[8]:

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Gender_Male	Married_Yes	Dependents_1	Deper
1	4583	1508.0	128.0	360.0	1.0	1	1	1	
2	3000	0.0	66.0	360.0	1.0	1	1	0	
3	2583	2358.0	120.0	360.0	1.0	1	1	0	
4	6000	0.0	141.0	360.0	1.0	1	0	0	
5	5417	4196.0	267.0	360.0	1.0	1	1	0	
...	...	...	...	...	...	...	...	...	
609	2900	0.0	71.0	360.0	1.0	0	0	0	
610	4106	0.0	40.0	180.0	1.0	1	1	0	
611	8072	240.0	253.0	360.0	1.0	1	1	1	
612	7583	0.0	187.0	360.0	1.0	1	1	0	
613	4583	0.0	133.0	360.0	0.0	0	0	0	

480 rows × 15 columns

In [7]:
```
# Split the data into a training and test set, where the “Loan_Status” column is the target.
X=newdf.drop(columns=['Loan_Status'],axis=1)
y=newdf['Loan_Status']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)
```

In [9]:
```
#Create a pipeline with a minmax scaler and a KNN classifer
from sklearn.pipeline import Pipeline
from sklearn.neighbors import KNeighborsClassifier

scaler = MinMaxScaler(feature_range=(0, 1))
pipe = Pipeline([('std', scaler), ('classifier', KNeighborsClassifier(n_neighbors=2))], verbose = True)
```

In [37]:
```
# Fit a default KNN classifier to the data with this pipeline
model = pipe.fit(X_train, y_train)
```

```
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
```

In [21]:
```
# scoring data
from sklearn.metrics import accuracy_score
print('accuracy of first KNN classifier is:', accuracy_score(y_test, model.predict(X_test)))
```

```
accuracy of first KNN classifier is: 0.6041666666666666
```

In [8]:
```
# Create a search space where the n_neighbors vareis from 1 to 10
from sklearn.model_selection import GridSearchCV

scaler = MinMaxScaler()

knn = KNeighborsClassifier(n_neighbors=5, n_jobs=-1)

pipe2 = Pipeline([('std', scaler), ('classifier', knn)], verbose = True)

# Create search space
search_space = [{"classifier__n_neighbors": [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]}]

# Fit a grid search with the pipeline, with 5 fold and find the best value for n_neighbors parameter
#Create grid search
classifier = GridSearchCV(pipe2, search_space, cv=5, verbose=0).fit(X_train, y_train)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[8], line 6
      2 from sklearn.model_selection import GridSearchCV
      4 scaler = MinMaxScaler()
----> 6 knn = KNeighborsClassifier(n_neighbors=5, n_jobs=-1)
      8 pipe2 = Pipeline([('std', scaler), ('classifier', knn)], verbose = True)
     10 # Create search space

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

In [69]:
```
# # Best neighborhood size (k)

print ('best value is:', classifier.best_estimator_.get_params()["classifier__n_neighbors"])
```

```
best value is: 5
```

In [10]:
```
# repeat steps 6 and 7 with the same pipeline, but expand your search space
# include logistic regression and random forest models
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.pipeline import Pipeline

#build pipe
scaler = MinMaxScaler()

pipe3 = Pipeline([('std', scaler), ("classifier", RandomForestClassifier())])

# Create disctionary for 2 classifier
search_space = [{"classifier": [LogisticRegression()],
                 "classifier__penalty": ['l1', 'l2'],
                 "classifier__C": np.logspace(0, 4, 10)},
                {"classifier": [RandomForestClassifier()],
                 "classifier__n_estimators": [10, 100, 1000],
                 "classifier__max_features": [1, 2, 3]}]
```

In [11]:
```
# What are the best model and hyperparameters found in the grid search?
gridsearch = GridSearchCV(pipe, search_space, cv=5, verbose=0)

#Fit grid search
best_model = gridsearch.fit(X_train, y_train)
```

```
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.1s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
```

```
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] ............... (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
```

```
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
```

```
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.1s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.1s
```

```
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.3s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.3s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
```

```
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.2s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.1s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   2.0s
[Pipeline] .............. (step 1 of 2) Processing std, total=   0.0s
[Pipeline] ........ (step 2 of 2) Processing classifier, total=   0.0s
```

```
C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\model_selection\_validation.py:425: FitFailedWarning:
50 fits failed out of a total of 145.
The score on these train-test partitions for these parameters will be set to nan.
If these failures are not expected, you can try to debug them by setting error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
50 fits failed with the following error:
Traceback (most recent call last):
  File "C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\model_selection\_validation.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\base.py", line 1151, in wrapper
    return fit_method(estimator, *args, **kwargs)
           ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\pipeline.py", line 420, in fit
    self._final_estimator.fit(Xt, y, **fit_params_last_step)
  File "C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\base.py", line 1151, in wrapper
    return fit_method(estimator, *args, **kwargs)
           ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\linear_model\_logistic.py", line 1168, in fit
    solver = _check_solver(self.solver, self.penalty, self.dual)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\linear_model\_logistic.py", line 56, in _check_solver
    raise ValueError(
ValueError: Solver lbfgs supports only 'l2' or 'none' penalties, got l1 penalty.

  warnings.warn(some_fits_failed_message, FitFailedWarning)
C:\Users\Daisy\anaconda3\Lib\site-packages\sklearn\model_selection\_search.py:976: UserWarning: One or more of the test
scores are non-finite: [       nan 0.79955571        nan 0.79692413        nan 0.79692413
        nan 0.79692413        nan 0.79429255        nan 0.79429255
        nan 0.79429255        nan 0.79429255        nan 0.79429255
        nan 0.79429255 0.7213944  0.77091593 0.77881066 0.7291866
 0.78133971 0.78920027 0.74227614 0.78393712 0.7917635 ]
  warnings.warn(
```

In [12]:
```
#View best model
best_model.best_estimator_.get_params()["classifier"]
```

Out[12]:

▾ LogisticRegression

LogisticRegression()

In [17]:
```
#Find the accuracy of this model on the test set.
predicted = best_model.predict(X_test)
```

```
from sklearn.metrics import accuracy_score
print('accuracy of Logistic Regression classifier is:', round(accuracy_score(y_test, predicted),2))
```

```
accuracy of Logistic Regression classifier is: 0.82
```

Summary

In this exercise, I tried to use hyperparamter grid search to fine tune the KNN classifier, as well as compare 3 different models (KNN, Logistic regression and Random forest classifiers). The winner is logistic regression, its accuracy is 0.82 vs the KNN classifier is only 0.60

Another observation is the build in gridsearch make the model comparision and selection much easier and visible. different models with different parameters can be easily tried out for model selection purpose.

In [ ]:

```
```