The problem statement:

There are three stages,

Stage-1: Machine learning

Stage-2: supervised learning

Stage-3: classification

1.logisticRegression:

```
the report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        45
           1       1.00      0.99      0.99        75

    accuracy                           0.99       120
   macro avg       0.99      0.99      0.99       120
weighted avg       0.99      0.99      0.99       120
```

The logisticRegression represent the accuracy level 0.99 in 'penalty': 'l2', 'solver': 'newton-cg': 0.9916844900066377

2.SVM:

```
the report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98        45
           1       1.00      0.97      0.99        75

    accuracy                           0.98       120
   macro avg       0.98      0.99      0.98       120
weighted avg       0.98      0.98      0.98       120
```

The SVM (SVC) represent the accuracy level 0.98 in 'C': 10, 'gamma': 'auto', 'kernel': 'sigmoid': 0.9834018801410106

3.Decision Tree:

```
the report:
              precision    recall  f1-score   support

           0       0.96      0.98      0.97        45
           1       0.99      0.97      0.98        75

    accuracy                           0.97       120
   macro avg       0.97      0.98      0.97       120
weighted avg       0.98      0.97      0.98       120
```

The Decision Tree represent the accuracy level 0.97 in 'criterion': 'log_loss', 'max_features': 'log2', 'splitter': 'random': 0.975053470019913

4.Random Forest:

```
the report:
              precision    recall  f1-score   support

           0       1.00      0.98      0.99        45
           1       0.99      1.00      0.99        75

    accuracy                           0.99       120
   macro avg       0.99      0.99      0.99       120
weighted avg       0.99      0.99      0.99       120
```

The Random Forest represent the accuracy level 0.99 in 'class_weight': 'balanced', 'criterion': 'entropy', 'max_features': 'log2': 0.9916474440062505

5.K-Nearest Neighbor(knn):

the report:

	precision	recall	f1-score	support
0	0.88	1.00	0.94	45
1	1.00	0.92	0.96	75
accuracy			0.95	120
macro avg	0.94	0.96	0.95	120
weighted avg	0.96	0.95	0.95	120

The K-Nearest Neighbor(knn) represent the accuracy level 0.95 in 'algorithm': 'auto', 'metric': 'minkowski', 'weights': 'distance': 0.9505208333333334

6.Navie bayes:

1(a) GaussianNB:

	precision	recall	f1-score	support
0	0.96	1.00	0.98	45
1	1.00	0.97	0.99	75
accuracy			0.98	120
macro avg	0.98	0.99	0.98	120
weighted avg	0.98	0.98	0.98	120

The GaussianNB represent the accuracy level 0.98

2(b) MultinomialNB:

	precision	recall	f1-score	support
0	0.67	0.98	0.79	45
1	0.98	0.71	0.82	75
accuracy			0.81	120
macro avg	0.82	0.84	0.81	120
weighted avg	0.86	0.81	0.81	120

The MultinomialNB represent the accuracy level 0.81

3(b) BernoulliNB:

	precision	recall	f1-score	support
0	0.85	1.00	0.92	45
1	1.00	0.89	0.94	75
accuracy			0.93	120
macro avg	0.92	0.95	0.93	120
weighted avg	0.94	0.93	0.93	120

The BernoulliNB represent the accuracy level 0.93

4(b) ComplementNB:

	precision	recall	f1-score	support
0	0.67	0.98	0.79	45
1	0.98	0.71	0.82	75
accuracy			0.81	120
macro avg	0.82	0.84	0.81	120
weighted avg	0.86	0.81	0.81	120

The ComplementNB represent the accuracy level 0.81

The finalized model:

The best model is LinearRegression for the given dataset

REASONS TO SELECT THIS AS A BEST MODEL:

- ✓ Compared to other models LinearRegression has the best model accuracy level.
- ✓ The accuracy level is 0.99
- ✓ In 'penalty': 'l2', 'solver': 'newton-cg': 0.9916844900066377.

Info about the dataset:

There are 399 rows × 25 columns

Preprocessing methods:

There are the two methods ;

- One is standardscaler.
- the another one is converting the string to nominal data.Because the dataset has same repeated string so we decided to convey the language to computer language (nominal data).