```
 """
Maria Teresa Luna Plascencia
Class: CS 677
Date: 04/23/2024
Final Project
"""
```

Main.py output

Loading loan file into a pandas dataframe

	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History
count	381.000000	381.000000	381.000000	370.000000	351.000000
mean	3579.845144	1277.275381	104.986877	340.864865	0.837607
std	1419.813818	2340.818114	28.358464	68.549257	0.369338
min	150.000000	0.000000	9.000000	12.000000	0.000000
25%	2600.000000	0.000000	90.000000	360.000000	1.000000
50%	3333.000000	983.000000	110.000000	360.000000	1.000000
75%	4288.000000	2016.000000	127.000000	360.000000	1.000000
max	9703.000000	33837.000000	150.000000	480.000000	1.000000

Loan_ID	object
Gender	object
Married	object
Dependents	object
Education	object
Self_Employed	object
ApplicantIncome	int64
CoapplicantIncome	float64
LoanAmount	float64
Loan_Amount_Term	float64
Credit_History	float64
Property_Area	object

Loan_Status object

dtype: object

Missing data

Gender 5

Married 0

Dependents 8

Education 0

Self_Employed 21

ApplicantIncome 0

CoapplicantIncome 0

LoanAmount 0

Loan_Amount_Term 11

Credit_History 30

Property_Area 0

Loan_Status 0

dtype: int64

Total missing data after pre-processing

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

Correlation matrix

	Gender	Married	Dependents	...	Credit_History	Property_Area	Loan_Status
Gender	1.000000	0.340476	2.145019e-01	...	-1.625894e-02	0.001534	-0.014497
Married	0.340476	1.000000	3.908320e-01	...	3.167619e-02	0.034601	0.092473
Dependents	0.214502	0.390832	1.000000e+00	...	-2.898939e-17	0.022785	0.013701
Education	0.102832	0.040532	1.158341e-01	...	6.783810e-03	-0.086582	-0.055586
Self_Employed	-0.067022	-0.031636	4.936785e-02	...	7.795180e-02	-0.071441	0.035915
ApplicantIncome	0.032050	-0.043796	1.432384e-01	...	-2.988279e-03	-0.064762	-0.010167
CoapplicantIncome	0.101598	0.081431	-6.753563e-02	...	2.505787e-02	0.010931	0.009017
LoanAmount	0.061304	0.142164	8.703642e-02	...	-4.644290e-02	-0.165424	0.041220
Loan_Amount_Term	-0.102015	-0.137874	-1.024416e-01	...	-2.388432e-02	-0.094657	-0.049748
Credit_History	-0.016259	0.031676	-2.898939e-17	...	1.000000e+00	0.037731	0.609630
Property_Area	0.001534	0.034601	2.278546e-02	...	3.773095e-02	1.000000	0.057834
Loan_Status	-0.014497	0.092473	1.370145e-02	...	6.096304e-01	0.057834	1.000000

[12 rows x 12 columns]

	Gender	Married	Dependents	Education	...	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
0	1	1	1	0	...	360.0	1.0	0	0
1	1	1	0	0	...	360.0	1.0	2	1
2	1	1	0	1	...	360.0	1.0	2	1
3	1	0	0	0	...	360.0	1.0	2	1
4	1	1	0	1	...	360.0	1.0	2	

1

[5 rows x 12 columns]

X TRAIN

(266, 11)

X TEST

(115, 11)

	d1	d2	d3	d4
N1	0.860870	0.860870	0.834783	0.843478
N2	0.860870	0.817391	0.834783	0.834783
N3	0.860870	0.860870	0.860870	0.834783
N4	0.860870	0.860870	0.852174	0.843478
N5	0.826087	0.860870	0.843478	0.834783
N6	0.800000	0.860870	0.834783	0.826087
N7	0.730435	0.852174	0.852174	0.843478
N8	0.721739	0.826087	0.852174	0.826087
N9	0.808696	0.860870	0.860870	0.843478
N10	0.860870	0.860870	0.860870	0.843478
N11	0.826087	0.860870	0.860870	0.843478
N12	0.808696	0.860870	0.860870	0.843478
N13	0.747826	0.834783	0.860870	0.852174
N14	0.730435	0.826087	0.860870	0.843478

Best combination: N=1, d=1, Acc: 0.8608695652173913

	d1	d2	d3	d4
N1	0.860870	0.756522	0.704348	0.660870
N2	0.860870	0.765217	0.713043	0.713043
N3	0.860870	0.869565	0.713043	0.782609
N4	0.860870	0.808696	0.756522	0.756522
N5	0.860870	0.817391	0.747826	0.773913

N6 0.860870 0.817391 0.756522 0.730435

N7 0.860870 0.808696 0.800000 0.739130

N8 0.843478 0.800000 0.808696 0.739130

N9 0.773913 0.817391 0.765217 0.739130

N10 0.713043 0.817391 0.756522 0.739130

N11 0.704348 0.817391 0.791304 0.765217

N12 0.773913 0.817391 0.817391 0.756522

N13 0.730435 0.808696 0.808696 0.739130

N14 0.773913 0.826087 0.826087 0.756522

Best combination: N=3, d=2, Acc: 0.8695652173913043

Q3.2 Plotting a graph showing the different accuracies

Evaluating LinearSVC

C:\Users\marlu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\svm_classes.py:31: FutureWarning: The default value of `dual` will change from `True` to `'auto'` in 1.5. Set the value of `dual` explicitly to suppress the warning.

 warnings.warn(

Evaluating SVC (rbf kernel)

Evaluating SVC (poly kernel)

C:\Users\marlu\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\svm_classes.py:31: FutureWarning: The default value of `dual` will change from `True` to `'auto'` in 1.5. Set the value of `dual` explicitly to suppress the warning.

 warnings.warn(

Model to evaluate:RandomForestClassifier

Accuracy for RandomForestClassifier(criterion='entropy', max_depth=1, n_estimators=1,

                random_state=42): , 0.8608695652173913

Classification report for RandomForestClassifier

          precision   recall f1-score   support

       0   0.95   0.55   0.69   33

       1   0.84   0.99   0.91   82

accuracy			0.86	115
macro avg	0.90	0.77	0.80	115
weighted avg	0.87	0.86	0.85	115

Model to evaluate:BalancedRandomForestClassifier

Accuracy for BalancedRandomForestClassifier(bootstrap=False, criterion='entropy',
max_depth=2, n_estimators=3, random_state=42,
replacement=True, sampling_strategy='all'): , 0.8695652173913043

Classification report for BalancedRandomForestClassifier

	precision	recall	f1-score	support
0	0.95	0.58	0.72	33
1	0.85	0.99	0.92	82
accuracy			0.87	115
macro avg	0.90	0.78	0.82	115
weighted avg	0.88	0.87	0.86	115

Model to evaluate:KNeighborsClassifier

Accuracy for KNeighborsClassifier(metric='euclidean', n_neighbors=7): , 0.8608695652173913

Classification report for KNeighborsClassifier

	precision	recall	f1-score	support
0	0.95	0.55	0.69	33
1	0.84	0.99	0.91	82
accuracy			0.86	115
macro avg	0.90	0.77	0.80	115

weighted avg 0.87 0.86 0.85 115

Model to evaluate:LogisticRegression

Accuracy for LogisticRegression(max_iter=1000): , 0.8521739130434782

Classification report for LogisticRegression

	precision	recall	f1-score	support
0	0.90	0.55	0.68	33
1	0.84	0.98	0.90	82
accuracy			0.85	115
macro avg	0.87	0.76	0.79	115
weighted avg	0.86	0.85	0.84	115

Model to evaluate:LinearSVC

Accuracy for LinearSVC(C=10, loss='hinge'): , 0.8608695652173913

Classification report for LinearSVC

	precision	recall	f1-score	support
0	0.95	0.55	0.69	33
1	0.84	0.99	0.91	82
accuracy			0.86	115
macro avg	0.90	0.77	0.80	115
weighted avg	0.87	0.86	0.85	115

Model to evaluate:VotingClassifier

Accuracy for VotingClassifier(estimators=[('rf',

RandomForestClassifier(criterion='entropy',

max_depth=1,

n_estimators=1,

random_state=42)),

('brf',

BalancedRandomForestClassifier(bootstrap=False,

criterion='entropy',

max_depth=2,

n_estimators=3,

random_state=42,

replacement=True,

sampling_strategy='all')),

('knn',

KNeighborsClassifier(metric='euclidean',

n_neighbors=7)),

('svm', LinearSVC(C=10, loss='hinge'))]): , 0.8608695652173913

Classification report for VotingClassifier

	precision	recall	f1-score	support
0	0.95	0.55	0.69	33
1	0.84	0.99	0.91	82
accuracy			0.86	115
macro avg	0.90	0.77	0.80	115
weighted avg	0.87	0.86	0.85	115

Model to evaluate:DecisionTreeClassifier

Accuracy for DecisionTreeClassifier(): , 0.7913043478260869

Classification report for DecisionTreeClassifier

	precision	recall	f1-score	support

0	0.60	0.79	0.68	33
1	0.90	0.79	0.84	82
accuracy			0.79	115
macro avg	0.75	0.79	0.76	115
weighted avg	0.82	0.79	0.80	115

	Model	TP	FP	TN	FN	Accuracy	TPR	TNR	F1[0]	F1[1]
0	RandomForestClassifier	81	15	18	1	0.861	0.988	0.545	0.692	0.910
1	BalancedRandomForestClassifier	81	14	19	1	0.870	0.988	0.576	0.717	0.915
2	KNeighborsClassifier	81	15	18	1	0.861	0.988	0.545	0.692	0.910
3	LogisticRegression	80	15	18	2	0.852	0.976	0.545	0.679	0.904
4	LinearSVC	81	15	18	1	0.861	0.988	0.545	0.692	0.910
5	VotingClassifier	81	15	18	1	0.861	0.988	0.545	0.692	0.910
6	DecisionTreeClassifier	65	7	26	17	0.791	0.793	0.788	0.684	0.844