PREDICT ME

Predictive Modeling Results

Data Input Summary

1. Total training samples: 300

2. Donation columns: ['Last Gift Amount', 'FY 17 Giving', 'FY 16 Giving', 'FY 15 Giving', 'FY 14 Giving', 'FY 13 Giving']

3. Positive sample count: 103 and Negative sample count: 197

4. Categorical columns: ['Gender', 'Manager', 'Address Type', 'Inclination', 'Alumni Engagement']

5. Train data (80%) used: 240

6. Test data (20%) used: 60

Steps taken to run the model

1. Data Input

2. Data cleaning: Remove null rows and columns

3. Identify columns containing categorical and textual data

4. Assign target value to each sample

5. Train 10 classifiers by spliting dataset for train and test

6. Calculate feature importance for each classifier

7. Plot confusion matrix and classification report

8. Select top 5 classifier using f1-score

9. Receiver Operating Characteristic (ROC) Curve

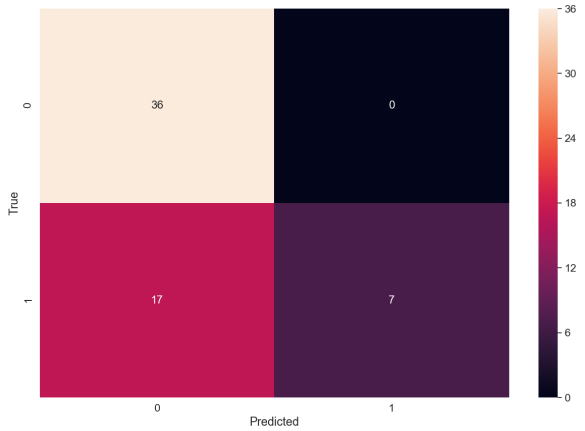
10. Identify optimal threshold and predict

Model Summary

1. Classifier: LogisticRegression

f1-score 0.67 and training time(seconds): 0.278

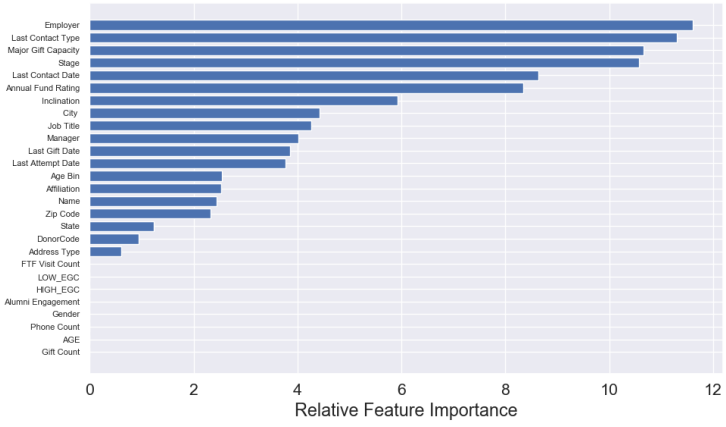
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.68	1.00	0.81	36
1	1.00	0.29	0.45	24
accuracy			0.72	60
macro avg	0.84	0.65	0.63	60
weighted avg	0.81	0.72	0.67	60

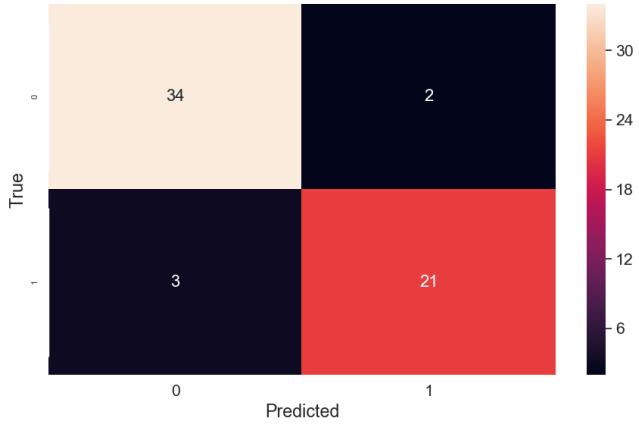
Feature Importance



2. Classifier: GaussianNB

f1-score 0.92 and training time(seconds): 0.008

Confusion Matrix



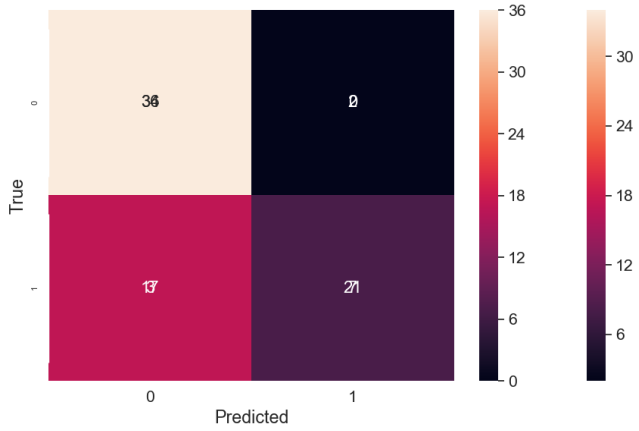
Classification Report

	precision	recall	f1-score	support
0	0.92	0.94	0.93	36
1	0.91	0.88	0.89	24
accuracy			0.92	60
macro avg	0.92	0.91	0.91	60
weighted avg	0.92	0.92	0.92	60

3. Classifier: MultinomialNB

f1-score 0.67 and training time(seconds): 0.003

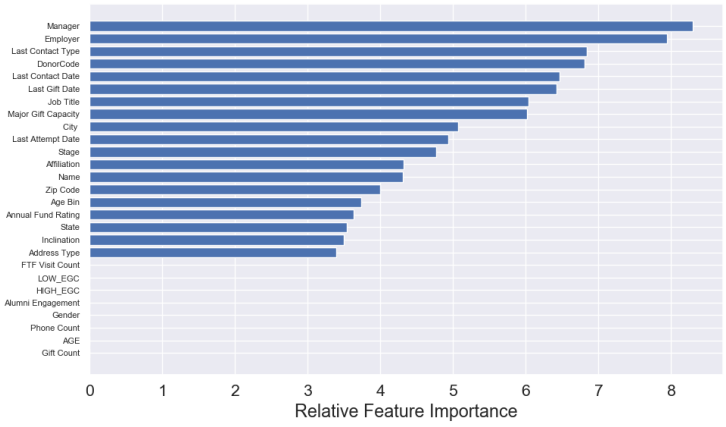
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.68	1.00	0.81	36
1	1.00	0.29	0.45	24
accuracy			0.72	60
macro avg	0.84	0.65	0.63	60
weighted avg	0.81	0.72	0.67	60

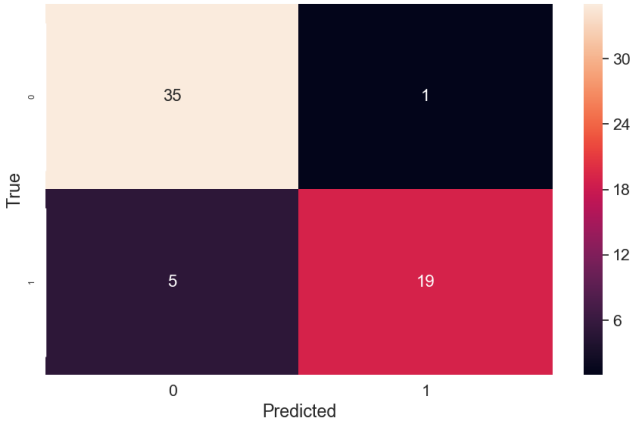
Feature Importance



4. Classifier: ComplementNB

f1-score 0.9 and training time(seconds): 0.003

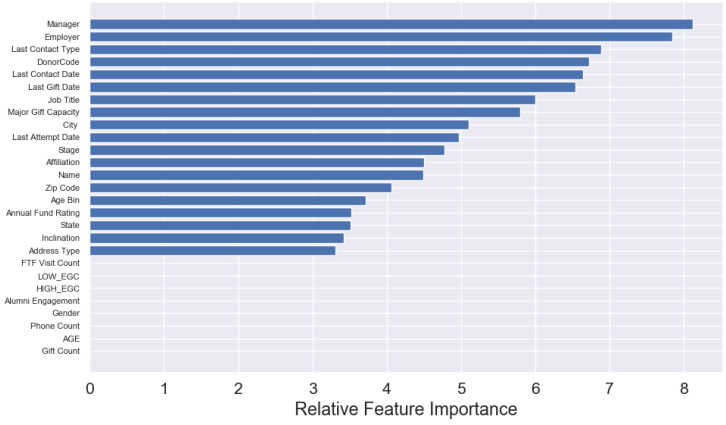
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.88	0.97	0.92	36
1	0.95	0.79	0.86	24
accuracy			0.90	60
macro avg	0.91	0.88	0.89	60
weighted avg	0.90	0.90	0.90	60

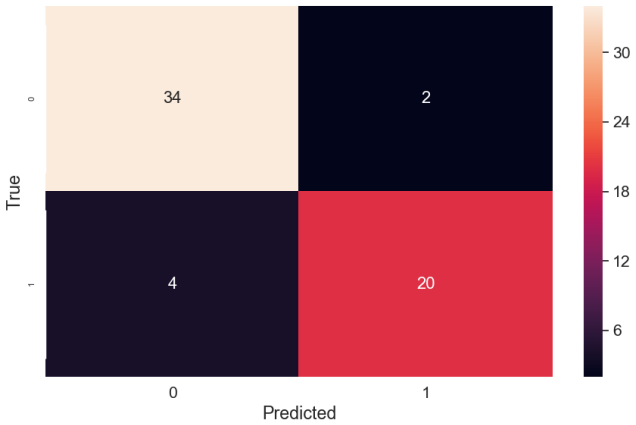
Feature Importance



5. Classifier: BernoulliNB

f1-score 0.9 and training time(seconds): 0.006

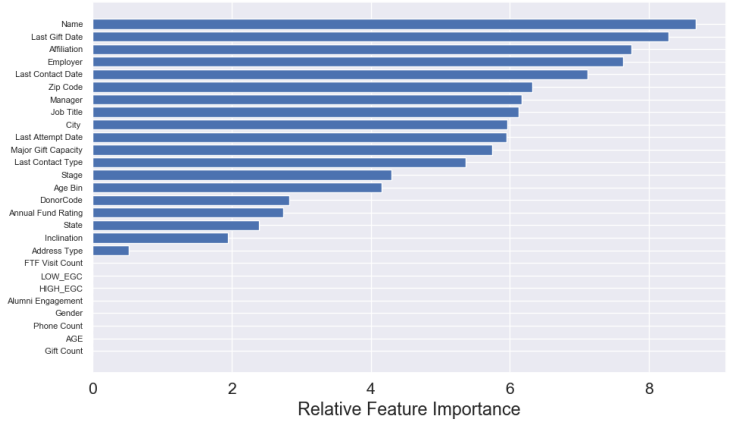
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.89	0.94	0.92	36
1	0.91	0.83	0.87	24
accuracy			0.90	60
macro avg	0.90	0.89	0.89	60
weighted avg	0.90	0.90	0.90	60

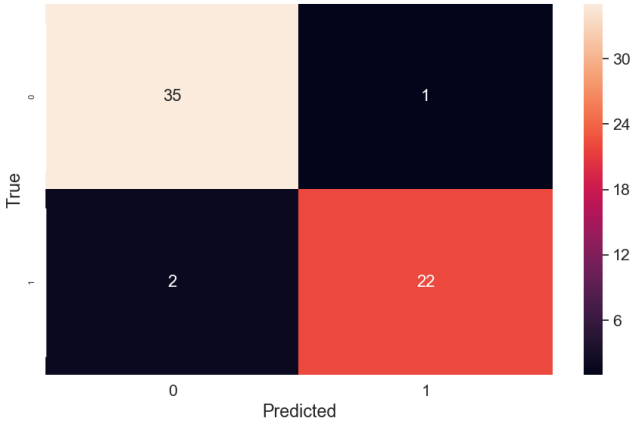
Feature Importance



6. Classifier: DecisionTreeClassifier

f1-score 0.95 and training time(seconds): 0.018

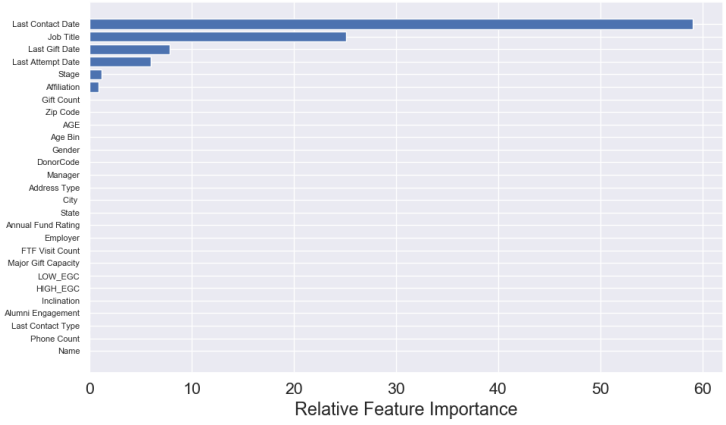
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.95	0.97	0.96	36
1	0.96	0.92	0.94	24
accuracy			0.95	60
macro avg	0.95	0.94	0.95	60
weighted avg	0.95	0.95	0.95	60

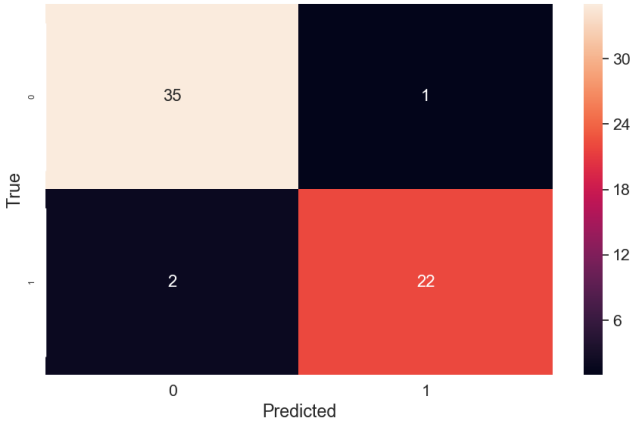
Feature Importance



7. Classifier: SGDClassifier

f1-score 0.95 and training time(seconds): 0.013

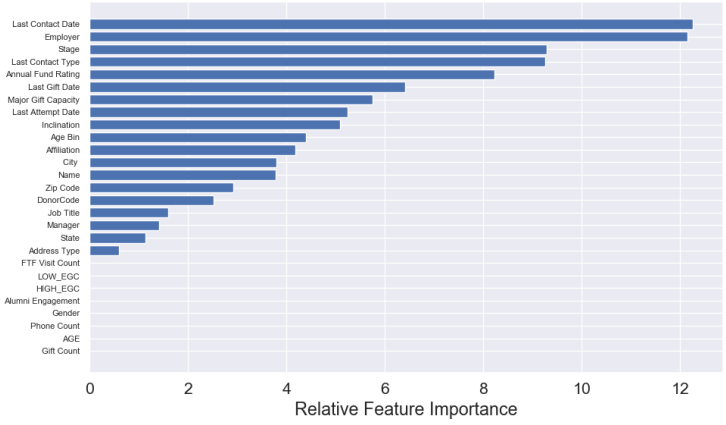
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.95	0.97	0.96	36
1	0.96	0.92	0.94	24
accuracy			0.95	60
macro avg	0.95	0.94	0.95	60
weighted avg	0.95	0.95	0.95	60

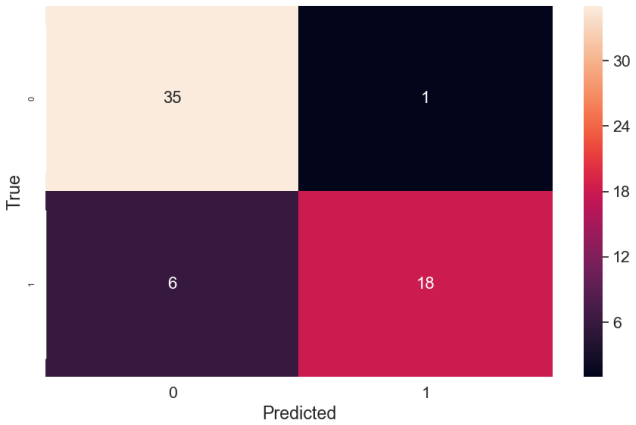
Feature Importance



8. Classifier: PassiveAggressiveClassifier

f1-score 0.88 and training time(seconds): 0.022

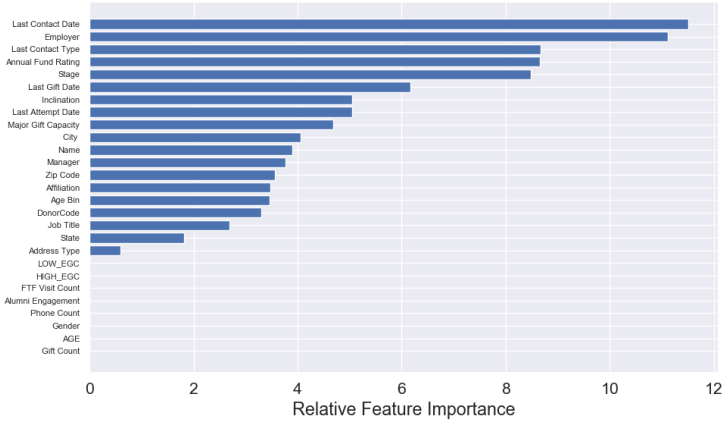
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.85	0.97	0.91	36
1	0.95	0.75	0.84	24
accuracy			0.88	60
macro avg	0.90	0.86	0.87	60
weighted avg	0.89	0.88	0.88	60

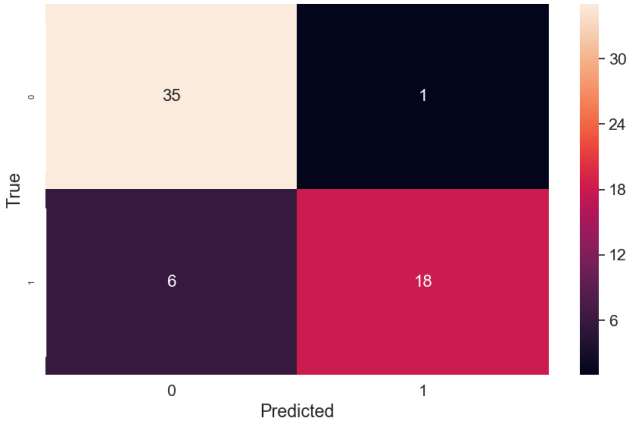
Feature Importance



9. Classifier: LinearSVC

f1-score 0.88 and training time(seconds): 0.015

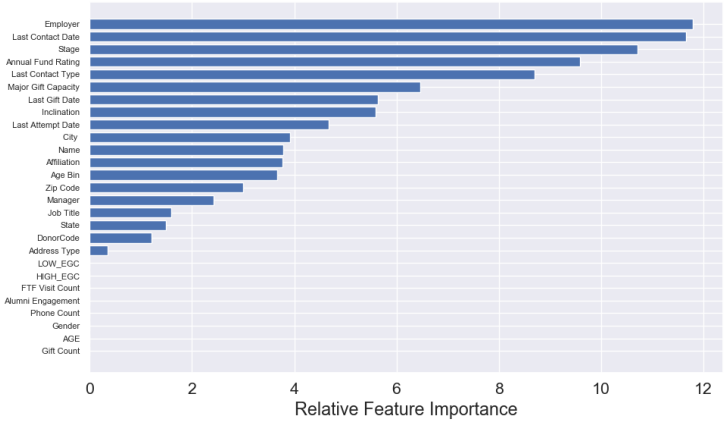
Confusion Matrix



Classification Report

	precision	recall	f1-score	support
0	0.85	0.97	0.91	36
1	0.95	0.75	0.84	24
accuracy			0.88	60
macro avg	0.90	0.86	0.87	60
weighted avg	0.89	0.88	0.88	60

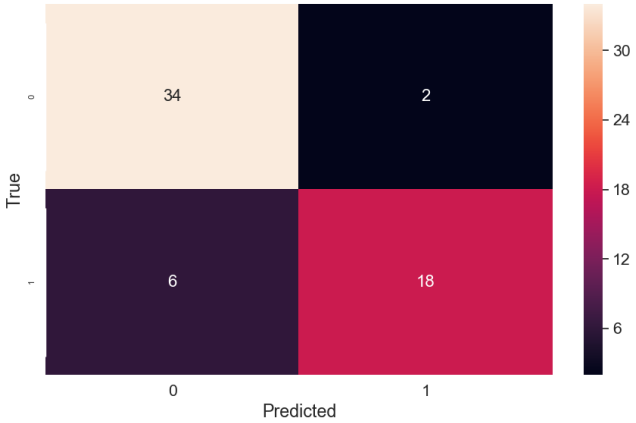
Feature Importance



10. Classifier: RandomForestClassifier

f1-score 0.86 and training time(seconds): 0.019

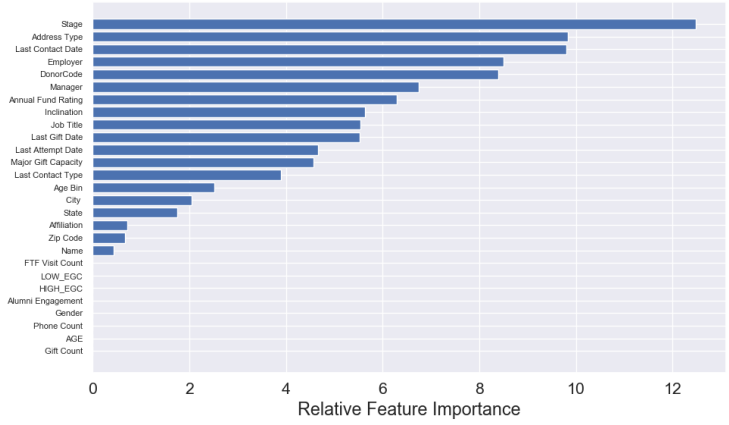
Confusion Matrix



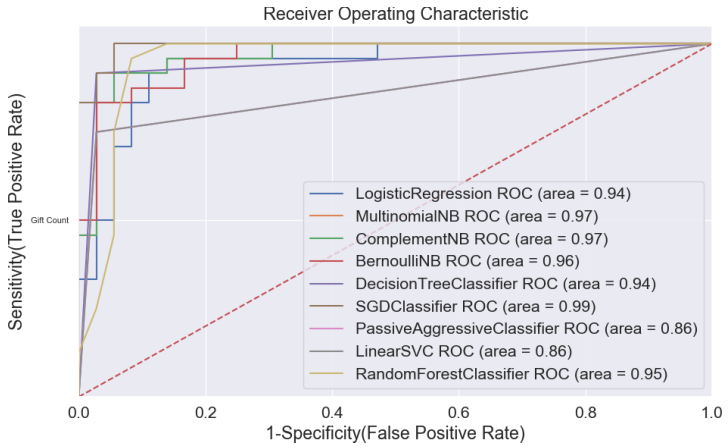
Classification Report

	precision	recall	f1-score	support
0	0.85	0.94	0.89	36
1	0.90	0.75	0.82	24
accuracy			0.87	60
macro avg	0.88	0.85	0.86	60
weighted avg	0.87	0.87	0.86	60

Feature Importance



Receiver Operating Characteristic (ROC) Curve



Top 5 models used to predict

1. DecisionTreeClassifier

F1-score: 0.95

Threshold used: 0.8

Donor predicted: 102

Non-Donor predicted: 198

2. SGDClassifier

F1-score: 0.95

Threshold used: 0.75

Donor predicted: 105

Non-Donor predicted: 195

3. GaussianNB

F1-score: 0.92

Threshold used: 0.8

Donor predicted: 107

Non-Donor predicted: 193

4. ComplementNB

F1-score: 0.9

Threshold used: 0.55

Donor predicted: 111

Non-Donor predicted: 189

5. BernoulliNB

F1-score: 0.9

Threshold used: 0.5

Donor predicted: 102

Non-Donor predicted: 198