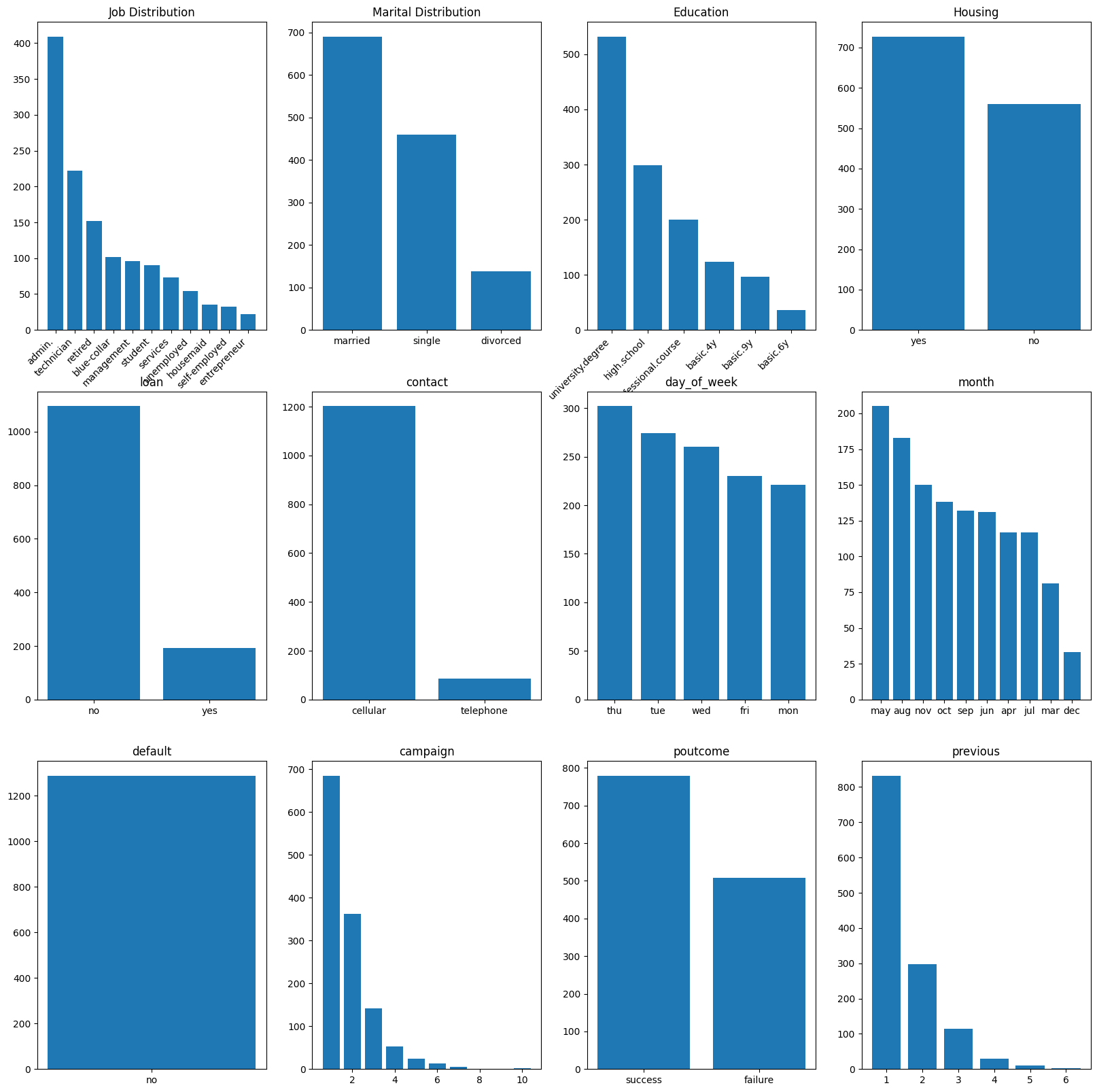
**Executive Summary:**

I compared 4 different classifiers (KNN, SVM, Logistic Regression and Decision Tree) on the input .csv file “bank-additional-full.csv”. Based on my analysis, I prefer to use Decision Tree Classifier *(with Max Depth = 4, Min Samples Split = 2 and Classifier Criterion = gini)*. Overall this classifier works well with Test Accuracy = **86%**. This model also was trained in **11s** compared to SVM as an example that took 371s.

**Business Insight:**

Clients that have highest probability of subscribing to deposit includes,

* Job of Admin and Technician
* Married
* University or High School Degree
* Has Housing Loan
* No Personal Loan
* Contact communication through Cellular
* Days contacted on Thursday or Tuesday
* Month contacted in May or August
* Number of campaigns is 1 or 2
* Outcome of previous marketing was a success
* Number of contacts before the campaign is 1 or 2
* Does NOT have credit in default



1. **KNN Classifier Report:**

**Train Accuracy:** 0.82

**Test Accuracy:** 0.85

**Model Train Time:** 8s

**Confusion matrix:**

[[606 78]

[ 90 157]]

**Classification report:**

precision recall f1-score support

no 0.87 0.89 0.88 684

yes 0.67 0.64 0.65 247

accuracy 0.82 931

macro avg 0.77 0.76 0.76 931

weighted avg 0.82 0.82 0.82 931

1. **Decision Tree Classifier Report:**

**Best Params:**

{'classifier\_\_criterion': 'gini', 'classifier\_\_max\_depth': 4, 'classifier\_\_min\_samples\_split': 2}

**Best Score:**

0.84

**Confusion Matrix:**

[[629 55]

[ 75 172]]

**Classification Report:**

precision recall f1-score support

no 0.89 0.92 0.91 684

yes 0.76 0.70 0.73 247

accuracy 0.86 931

macro avg 0.83 0.81 0.82 931

weighted avg 0.86 0.86 0.86 931

**Feature Importance Value:**

[0.0089482 0.3302555 0. 0.52903671 0. 0.

0. 0.06220005 0.04160725 0.02795229 0. 0.

0. 0. 0. 0. 0. 0.

0. 0. 0. 0. 0. 0.

0. 0. 0. 0. 0. 0.

0. 0. 0. 0. 0. 0.

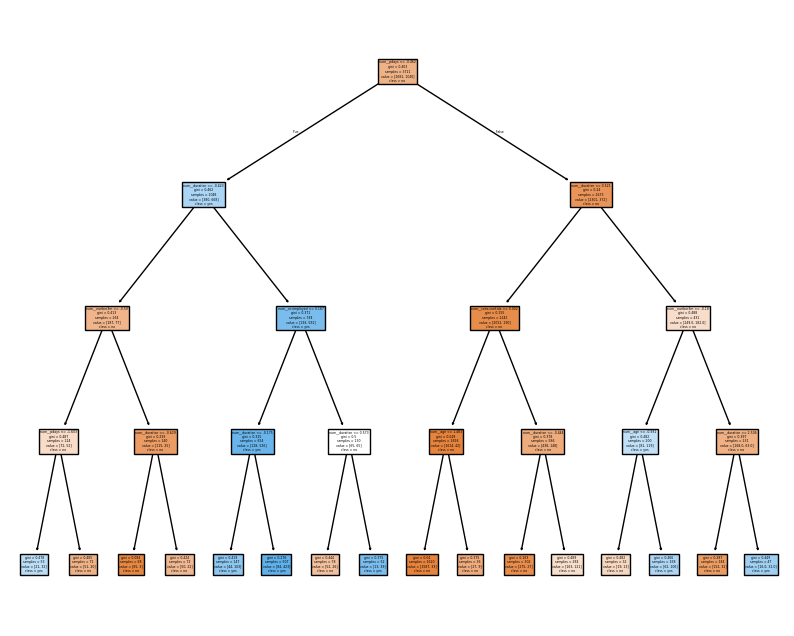
0. 0. 0. 0. 0. 0.

0. 0. 0. 0. 0. 0.

0. 0. 0. 0. 0. 0.

0. ]

**Classification Tree:**



**Decision Path:**

(0, 0) 1

(0, 1) 1

(0, 9) 1

(0, 10) 1

(0, 12) 1

**Maximum Depth of the Tree:**

4

**Number of Leaf Nodes:**

16

**Decision Path:**

{'ccp\_alpha': 0.0, 'class\_weight': None, 'criterion': 'gini', 'max\_depth': 4, 'max\_features': None, 'max\_leaf\_nodes': None, 'min\_impurity\_decrease': 0.0, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'min\_weight\_fraction\_leaf': 0.0, 'monotonic\_cst': None, 'random\_state': None, 'splitter': 'best'}

**Train Accuracy:** 0.85

**Test Accuracy:** 0.86

**Time to train the model:** 11s

1. **SVM Classifier:**

**Train Accuracy:** 0.87

**Test Accuracy:** 0.85

**Time to train the model:** 371s

**Confusion Matrix:**

[[628 56]

[ 81 166]]

**Classification Report:**

precision recall f1-score support

no 0.89 0.92 0.90 684

yes 0.75 0.67 0.71 247

accuracy 0.85 931

macro avg 0.82 0.80 0.80 931

weighted avg 0.85 0.85 0.85 931

**Best Parameters:**

{'classifier\_\_C': 10, 'classifier\_\_gamma': 0.01, 'classifier\_\_kernel': 'rbf'}

1. **Logistic Regression Classifier:**

**Train Accuracy:** 0.84

**Test Accuracy:** 0.85

**Time to train the model:** 36s

**Confusion Matrix:**

[[629 55]

[ 81 166]]

**Classification Report:**

precision recall f1-score support

no 0.89 0.92 0.90 684

yes 0.75 0.67 0.71 247

accuracy 0.85 931

macro avg 0.82 0.80 0.81 931

weighted avg 0.85 0.85 0.85 931

**Best Params:**

{'classifier\_\_C': 0.1, 'classifier\_\_penalty': 'l1', 'classifier\_\_solver': 'saga'}

**Best Score:**

0.84