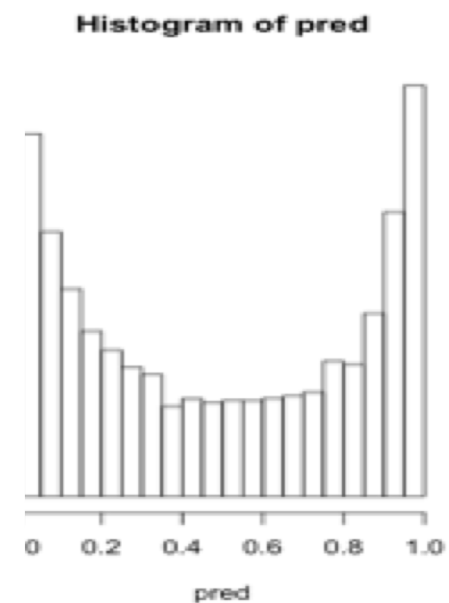
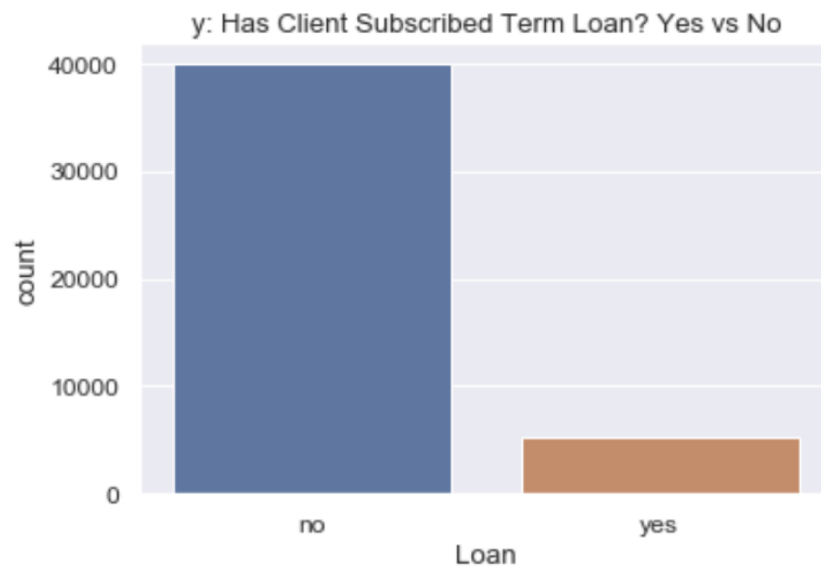
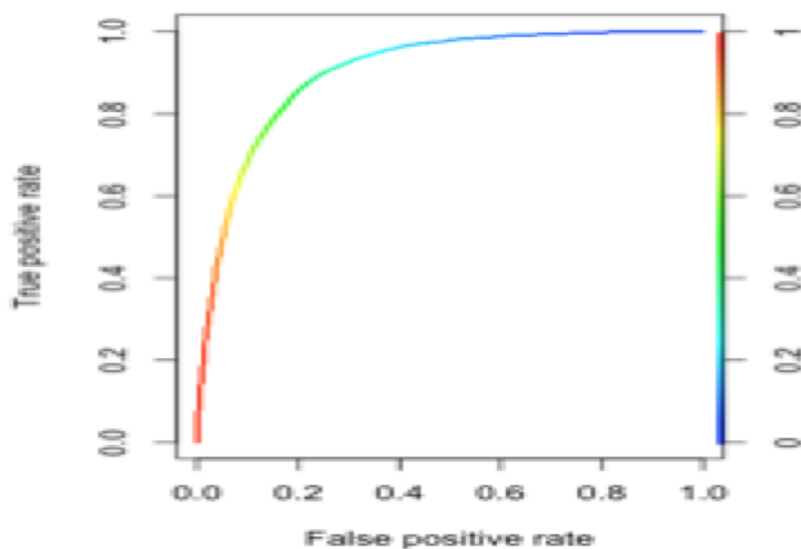


Bank Marketing Prediction



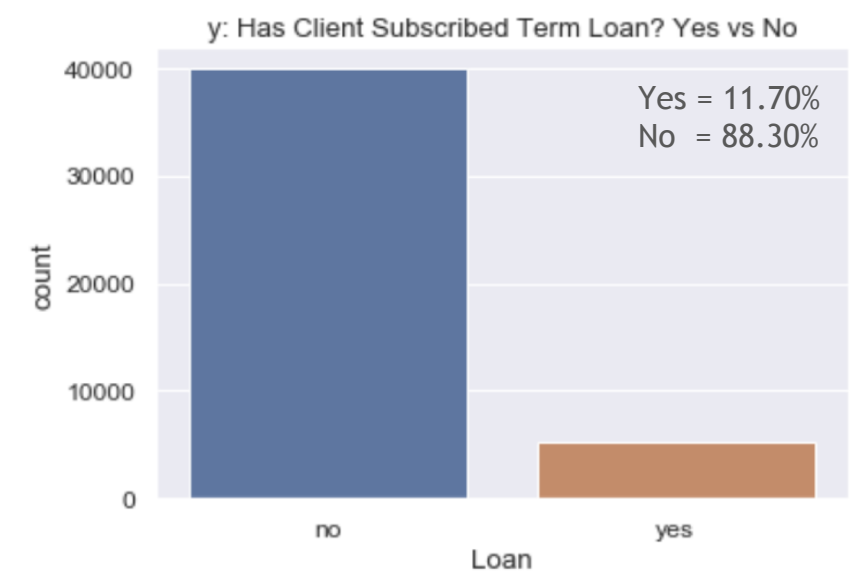
Goal

Our main goal is to predict the probability of customer in subscribing to Term Deposit

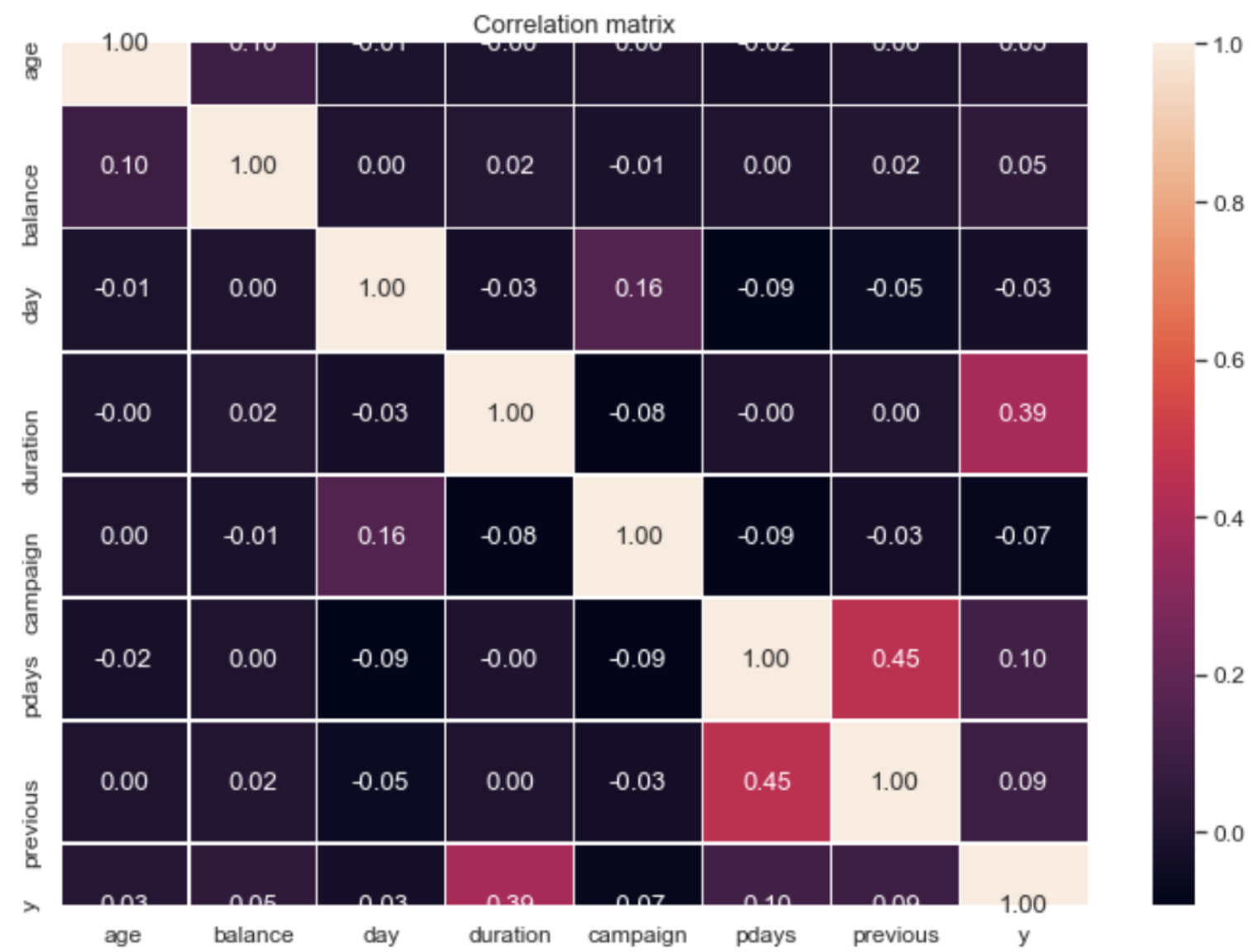
Approach

Built Logistic Regression Model

Correlation & Chi Squared Test for Independent Variables in Relation to The Dependent variables(y)



x_var	Chisquared	df	pvalue
job	836.110	11	< 2.2e-16
marital	196.500	2	< 2.2e-16
education	238.920	3	< 2.2e-16
default	22.724	1	1.871e-06
housing	875.690	1	< 2.2e-16
contact	1035.700	2	< 2.2e-16
loan	210.190	1	< 2.2e-16



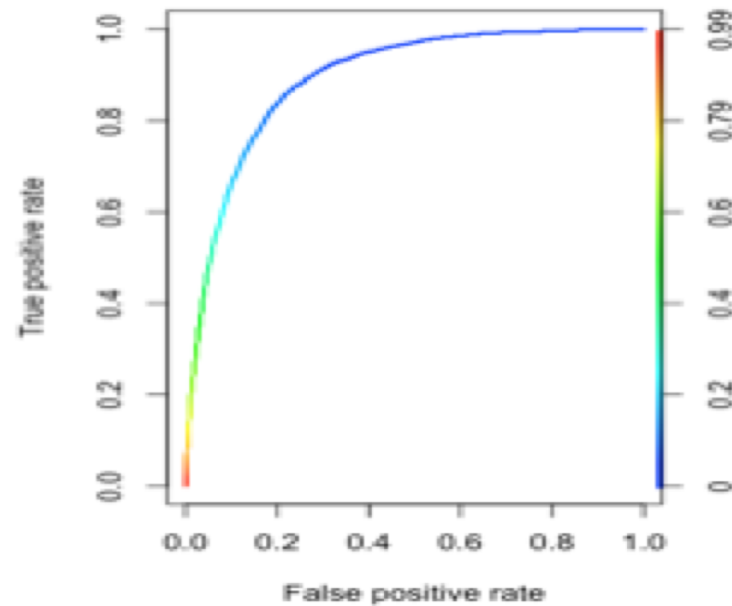
MODEL 1

Significant Variables:

- housing_yes
- loan_yes
- campaign
- balance
- duration

Accuracy = 89.99%

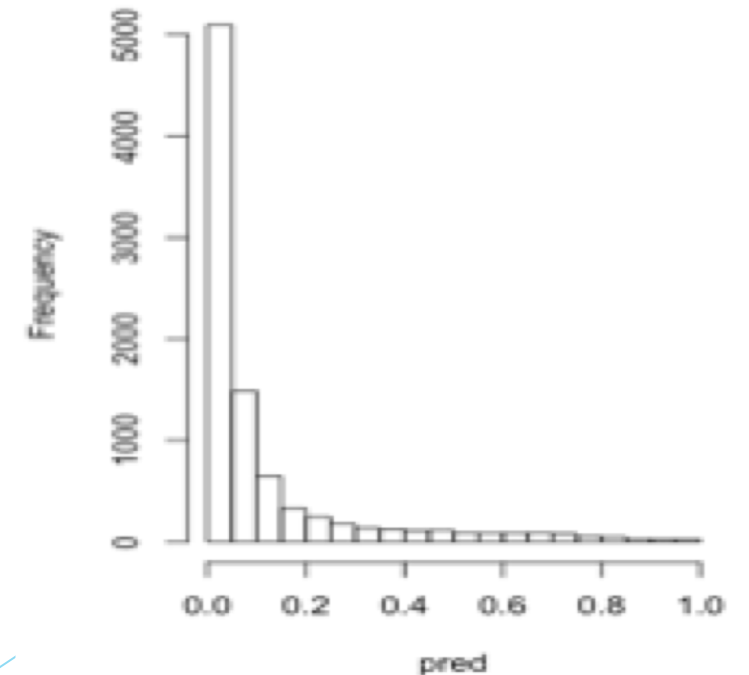
ROC



	y_actual = No	y_actual = Yes
y_pred = No	7764	684
y_pred = Yes	221	374

Histogram of predicted values

Histogram of pred



MODEL 2: Generate Synthetic Samples to Handle Imbalance Data

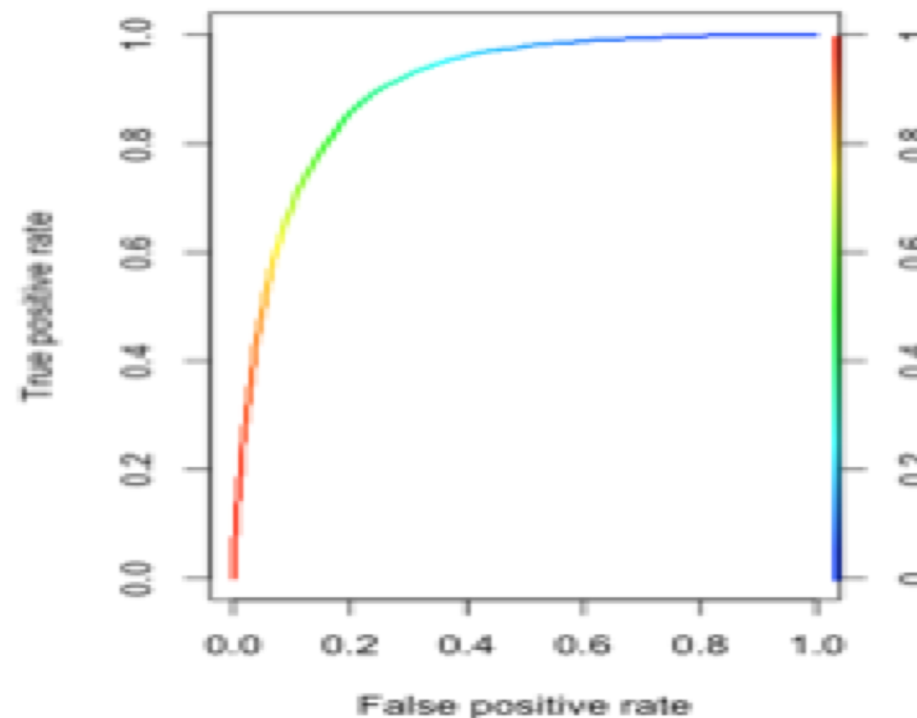
Significant Variables:

- job_student
- marital_married
- job_admin
- job_retired
- balance
- duration

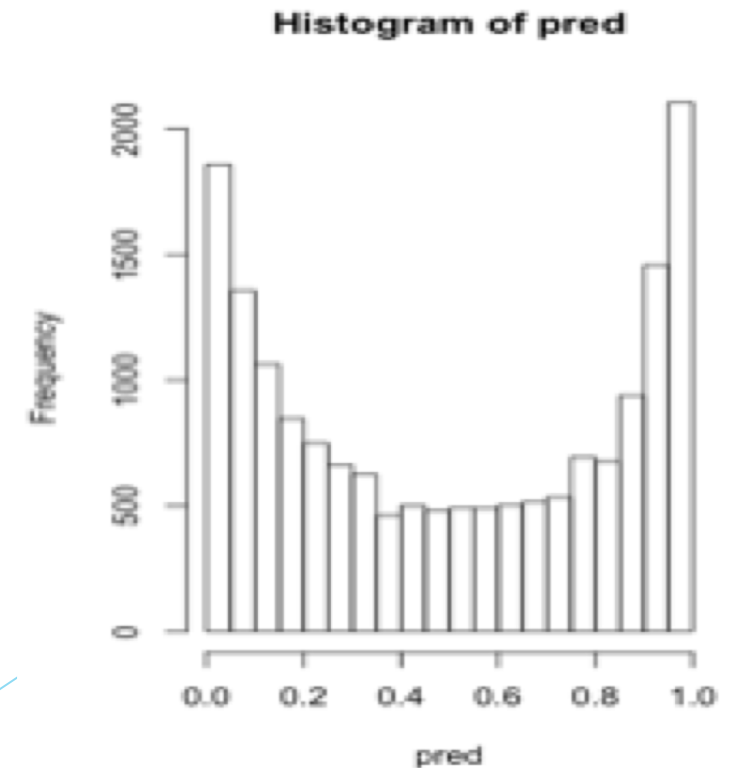
Accuracy = 82.27%

	y_actual = No	y_actual = Yes
y_pred = No	7043	1557
y_pred = Yes	1457	6943

ROC



Histogram of predicted values



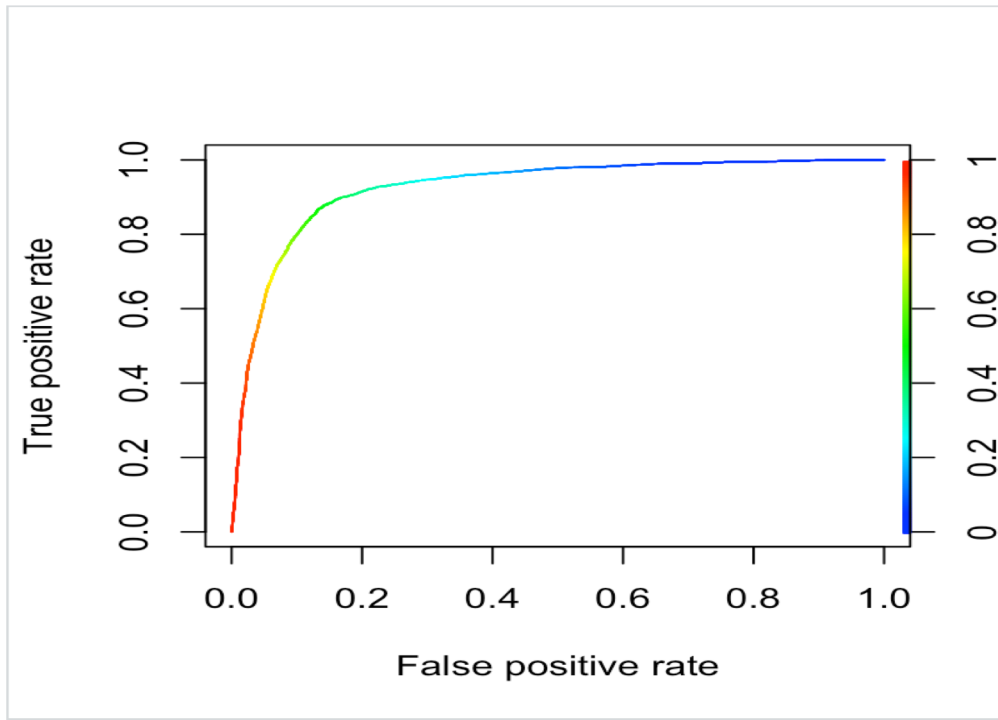
MODEL 3: New Customers

Significant Variables:

- Age
- Balance
- Duration
- jobs - retired, student, blue_collar
- marital_married

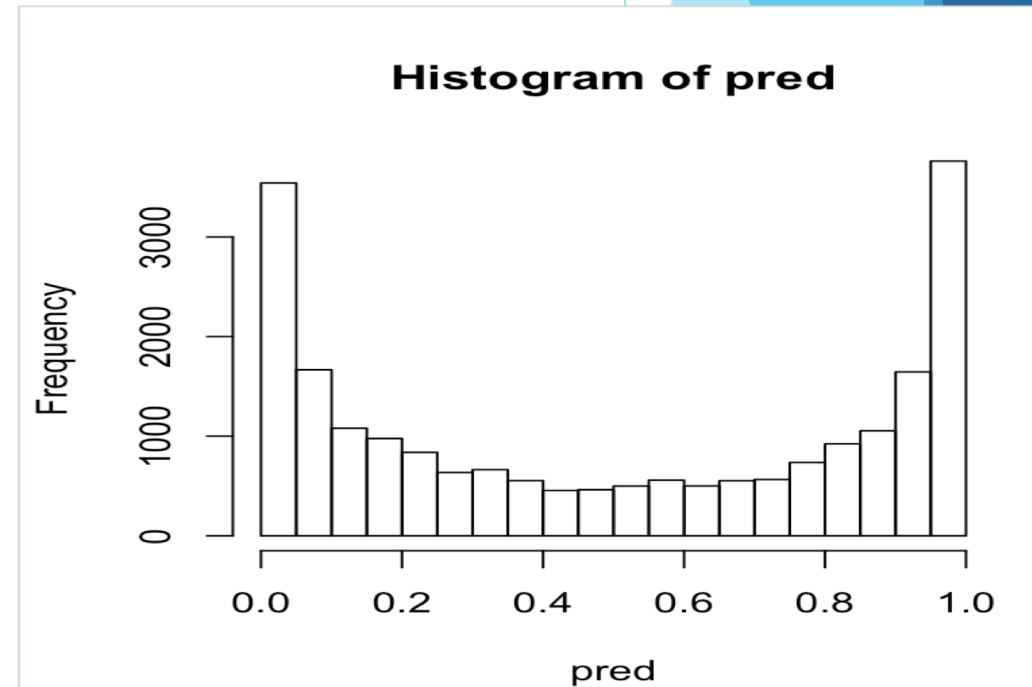
Accuracy = 86.63%

ROC



	y_actual = No	y_actual = Yes
y_pred = No	9403	1468
y_pred = Yes	1429	9364

Histogram of predicted values



MODEL 4: Old Customer

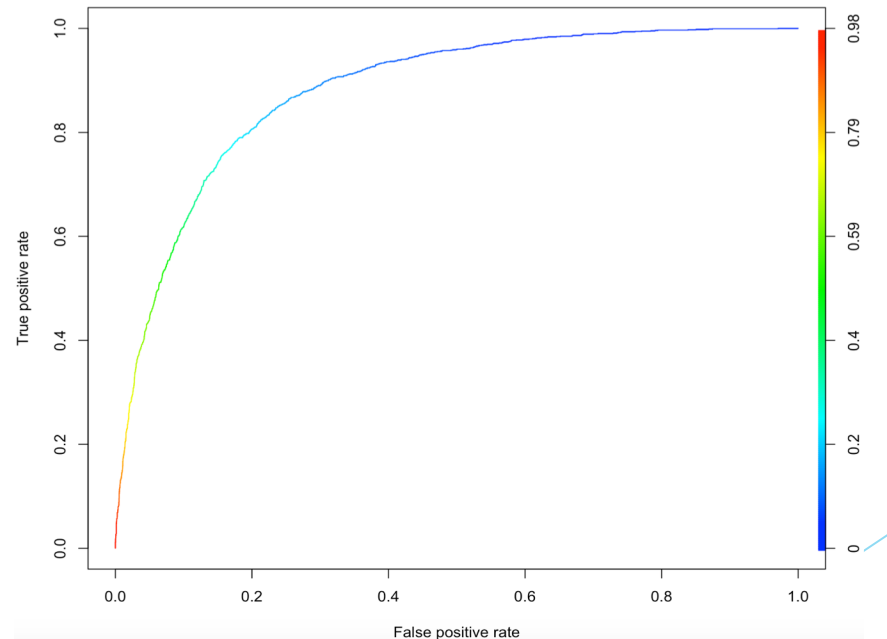
Significant Variables:

- housing_yes
- loan_yes
- campaign
- balance
- duration

- **Accuracy = 83.69%**

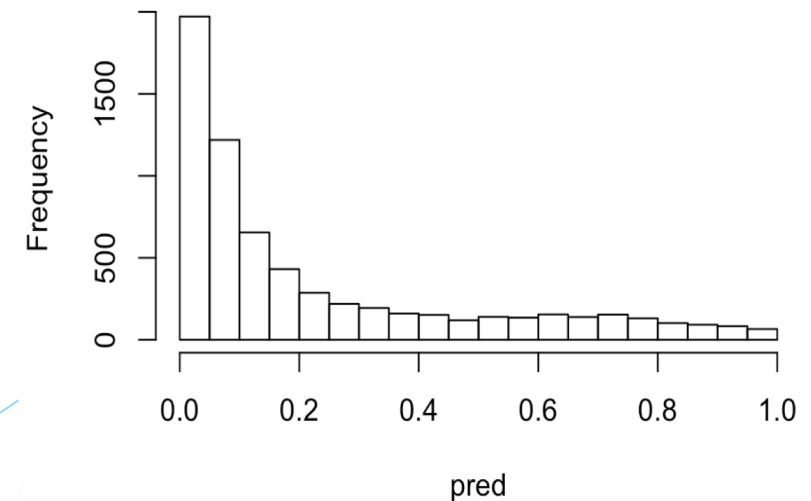
	y_actual = No	y_actual = Yes
y_pred = No	4703	705
y_pred = Yes	372	825

ROC



Histogram of predicted values

Histogram of pred



Key Takeaways

- Accuracy is not the only measure of how good the model is
- Should assess the quality of the model and improvise
- Ensure that the data is not imbalanced
- Create segments of similar customers using affinity clustering or k-means
- Create separate model for each segment to capture different customer behaviour
 - there is no such thing as an average customer and personalisation is the key!