

Project Name:

No title yet.

Problem Statement:

I will be trying to predict whether a customer will sign up for a long-term deposit.

Data Science Hypotheses & Solutions:

Because I will be predicting whether or not a customer will sign up for a long-term deposit this will be a classification project. I may employ one or more classification algorithms to include logistic regression, random forests, naive bayes, SVM, or KNN. One challenge I will face is determining which method will yield the best results. I may need to compare the results of several methods or employ an ensemble methods.

Data Sets:

I will be using the banking dataset available from UCI at <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>. The dataset contains the response variable which indicates whether or not the contact signed up for a long-term deposit and predictor variables indicating age, employment type, marital status, education, the existence of credit in default, average yearly account balance, the existence of a home loan, the existence of a personal loan, contact communication type, the last contact day of the month, the last contact month of the year, the duration of the last contact, the number of contacts performed during the current campaign, the number of days that passed after the customer was last contacted in connection with a prior campaign, and the number of contacts performed prior to the present campaign, and the outcome of the prior campaign.

Business Cases:

The outcome of this analysis could enable the bank to more efficiently direct its resources in future marketing campaigns by targeting customers who are the most likely to sign up for a long-term deposit. This would in turn improve ROI for marketing campaigns by increasing the number of subscriptions per unit cost.

The results may also assist in marketing other products such as credit cards or loans.

It may be possible to create an score for loans.

Stakeholders:

A bank would be the ideal client for this project.