

## Artificial Intelligence Final – Part 1

The objective of my model creation is to predict whether or not a particular client would access a bank term deposit for a Portuguese banking institution based on phone call information only for clients who have successfully given their information. Data collected about these calls includes information about the client's age, employment status, education status, marital status, banking information, and various information about the bank. This data was collected from the UCI Machine Learning Database (<https://archive.ics.uci.edu/ml/datasets/bank+marketing#>).

In order to build various models, the data will have to be cleaned and reset to fit the requirements needed to create a classification model. There are 11 various employment descriptions, 3 marital statuses, 2 types of communication, and 7 education descriptions that will be given integer values (referenced in the table below. Any attributes that contain a 'yes' or 'no' answer will be changed to binary variables. Months and days of the week will also be given integer values where January is 1 and Monday is 1. Finally, whether or not clients were previously contacted will be converted to a binary {0, 1} value instead of how many days, if any, were given between calls. In order to focus only on a dataset that contains known information, any clients with unknown information will be deleted. Making these changes allows for 4,652 instances with 19 predictors. The predictors will be normalized and scaled using the  $l_2$ -norm to center each column and allow the mean to be 0 and standard deviation to 1 for all of the predictors. The response variables will not be centered or scaled due to their binary nature. The training set will be 80% of the full data set to reduce overfitting will allowing for the predictive model to have a high accuracy score.

A classification neural network will be created from the *sklearn* package. The only change made was to set alpha equal to 0.1, while all other parameters will be left at default. For extended information, other *sklearn* machine learning models will be created to compare the success of the neural network – logistic regression, K-nearest neighbors, linear support vector classification, gradient boosting, decision tree, random forest, and Naïve-Bayes. A for-loop will be used to create a prediction model for all of these methods, returning a final table with information on the model, training accuracy score, test accuracy score, and the time required to create the model. This will give information on the best model, as well as the success of the neural network in comparison to these machine learning models.

Employment	
Unemployed	1
Student	2
Retired	3
Self-employed	4
Services	5
Management	6
Housemaid	7
Entrepreneur	8
Blue-collar	9
Administration	10
Technician	11

Education	
Basic.4y	1
Basic.6y	2
Basic.9y	3
High School	4
University	5
Professional	6
Illiterate	7

Marital	
Single	1
Married	2
Divorced	3

Contact	
Cell Phone	1
Telephone	2