

# *Artificial Intelligence with Python*

## *Assignment 6: Banking predictions*

---

Consider the Bank Marketing Dataset from UCI at

<https://archive.ics.uci.edu/ml/datasets/bank+marketing>

0) Go to above web site and read the description of the dataset.

Download [bank.csv](#).

1) Read in the CSV file using pandas. Pay attention to the file delimiter.

Inspect the resulting dataframe with respect to the column names and the variable types.

2) Pick data from the following columns to a second dataframe 'df2': y, job, marital, default, housing, poutcome.

3) Convert categorical variables to dummy numerical values using the command

```
df3 = pd.get_dummies(df2, columns=['job', 'marital', 'default', 'housing', 'poutcome'])
```

4) Produce a heat map of correlation coefficients for all variables in df3. Describe the amount of correlation between the variables in your own words.

5) Select the column called 'y' of df3 as the target variable y, and all the remaining columns for the explanatory variables X.

6) Split the dataset into training and testing sets with 75/25 ratio.

7) Setup a logistic regression model, train it with training data and predict on testing data.

8) Print the confusion matrix (or use heat map if you want) and accuracy score for the logistic regression model.

9) Repeat steps 7 and 8 for k-nearest neighbors model. Use k=3, for example, or experiment with different values.

10) Compare the results between the two models.

Type your findings in code comments or between triple quotes `"""..."""`.

Each step 1-10 above is worth a point. You need 6 points to complete this problem.