SparkCognition Data Science Assignmnet

Data Set:

You are working for SparkCognition as a Data Scientist. SparkCognition has been commissioned by an insurance company to develop a tool to optimize their marketing efforts. They have given us a data set as a result of an email marketing campaign. The data set includes customer information, described below, as a well as whether the customer responded to the marketing campaign or not.

Here are the descriptions of each column in the data set:

Type	Name	Description
Input Variables	custAge	The age of the customer (in years)
Input Variables	profession	Type of job
Input Variables	marital	Marital status
Input Variables	schooling	Education level
Input Variables	default	Has a previous defaulted account?
Input Variables	housing	Has a housing loan?
Input Variables	loan	Has a personal loan?
Input Variables	contact	Preferred contact type
Input Variables	month	Last contact month
Input Variables	day_of_weel	Last contact day of the week
Input Variables	campaign	Number of times the customer was contacted
Input Variables	pdays	Number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
Input Variables	previous	Number of contacts performed before this campaign and for this client
Input Variables	poutcome	Outcome of the previous marketing campaign
Input Variables	emp.var.rate	Employment variation rate - quarterly indicator
Input Variables	cons.price.id	Consumer price index - monthly indicator
Input Variables	cons.conf.idx	Consumer confidence index - monthly indicator
Input Variables	euribor3m	Euribor 3 month rate - daily indicator
Input Variables	nr.employed	Number of employees - quarterly indicator
Input Variables	pmonths	Number of months that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
Input Variables	pastEmail	Number of previous emails sent to this client
Target Variables	responded	Did the customer respond to the marketing campaign and purchase a policy?

Task:

Design a model that will be able to predict whether a customer will respond to the marketing campaign based on his/her information. In other words, predict the 'responded' target variable described above based on all the input variables provided.

Files:

marketing_training.csv - contains the training set that you will use to build the model. The target variable is 'responded'

marketing_test.csv – contains testing data where the input variables are provided but not the 'responded' target column.

Readme.pdf – this document.

Deliverables:

Provide the following:

- The source code you used to build the model and make predictions. (You are free to use any language and any open-source package/library)
- A .csv file containing the predictions of the test data. You can add the target column ('responded') to the test data or simply provide it alone with the 'id' column.
- Briefly answer the following questions:
    - Describe your model and why did you choose this model over other types of models?
    - Describe any other models you have tried and why do you think this model preforms better?
    - How did you handle missing data?
    - How did you handle categorical (string) data?
    - How did you handle unbalanced data?
    - How did you test your model?