**Business Understanding**

Assume that you are an analyst working for a local bank. You are tasked with helping the marketing team to understand factors impacting customer decisions and directly target individual customers who are more likely to subscribe to the bank's new term-deposit financial product.

The training data is generated as a result of this bank's ongoing direct marketing campaign and includes the results of the campaign this campaign up until now.

You are specifically asked to, first, help the marketing manager to understand what is (are) the most important attribute(s) in determining whether a customer is going to subscribe to the new product;  and, second, to help the marketing team identify customers who are most likely to subscribe to the product by providing the team with best performing predictive model.

You are not expected to deploy your model but only to identify the best-performing one.

Data Understanding

**Data Dictionary:**

 age (numeric)

 job : type of job (categorical: "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services")

 marital : marital status (categorical: "married", "divorced", "single"; note: "divorced" means divorced or widowed)

 education (categorical: "unknown", "secondary", "primary", "tertiary")

 default: has credit in default? (binary: "yes", "no")

 balance: average yearly balance, in euros (numeric)

 housing: has housing loan? (binary: "yes", "no")

 loan: has personal loan? (binary: "yes", "no")

 contact: contact communication type (categorical: "unknown", "telephone", "cellular")

 pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)

 previous: number of contacts performed before this campaign and for this client (numeric)

 Subscribed - has the client subscribed to a term deposit? (binary: "yes", "no")

1. **Considering the data specifications and stated business goal(s), what are possible data modeling methodologies that you can use to achieve stated goals?**

***Instructions:***

*You should select all modeling techniques that apply. Wrong selections will result in negative points!*

|  |  |
| --- | --- |
|  | Clustering |
|  | Random Forest |
|  | Association Rules Analysis |
|  | Gradient Boosted Trees |
|  | Decision Tree |
|  | Linear Regression |
|  | Logistic Regression |
|  | Gradient Decent |
|  | Neural Nets |
|  | Support Vector Machines |

1. **Explain your reasoning on why you selected those specific methodologies in question 1.**

***Hint:****You should focus on data specification (e.g., target variable type, independent variables, and analysis goals)*

1. **Do you have to prepare your data in order to be used with any of the previously identified methodologies?  If so, explain what needs to be done and prepare your data accordingly.**

*Instructions:*

*Explain the issues that you identified in data and how you plan to deal with them*

*Upload a screenshot of your data tab showing that you have fixed the unidentified issues*

1. **Of all methodologies that can be used on this problem (your answer to question 1) which one(s) is(are) better suited to provide a good understanding of what attribute(s) affect your target variable the most.**

*Instruction:*

*You should select all modeling techniques that applies. Wrong selections will result in negative points!*

Options:

|  |  |
| --- | --- |
|  | Gradient Boosted Trees |
|  | Random Forest |
|  | Linear Regression |
|  | Association Rules Analysis |
|  | Decision Tree |
|  | Gradient Decent |
|  | Clustering |
|  | Neural Nets |
|  | Support Vector Machines |
|  | Logistic Regression |

1. **Explain your reasoning on why you selected those specific methodologies in question 4.**

*Hint: You should focus on the capabilities and characteristics of selected methodologies versus those that were not suitable for the stated purpose of question 4.*

1. **Relying on your response to the previous questions, build a model to better understand what attribute(s) affect the target variable and answer the following questions**

* Which modeling technique did you end up choosing for this purpose, and why did you go with that modeling technique
* What is(are) the most important attribute(s) affecting the target variable according to the results of your modeling effort.

1. **Which modeling technique gives you the best performance in predicting the desired outcome? How does your chosen modeling technique compare to other candidates? Report your final model performance measure.**

*Hint: Need to state specific performance measures and demonstrate why your chosen technique is the best one in technical/numerical terms.*

1. **Upload screenshots of your RapidMiner process window (Design tab) pertaining to the previous question where you constructed and compared different predictive models.**
2. **Upload a screenshot of your "final" RapidMiner process window (Design tab), showing your best-performing model at work.**