**Data Preprocessing explanatory note**

**Target 1: find missing value "unknown", and handle it.**

|  |
| --- |
| job unknown percent: 1.19%  marital unknown percent: 1.50%  education unknown percent: 4.43%  default unknown percent: 21.52%  housing unknown percent: 3.95%  loan unknown percent: 3.98%  Others 0%  If we try to delete the columns with unknown data, we need to delete 10977 columns, about 1/3 of our whole dataset (test+train).  Those unknown percent between the test dataset and train dataset are the same.  So we can use the same method to handle unknown data in the test&train dataset.  These 2 methods are the most commonly used. |

* **Method 1: Use Mode to fill those unknown data.**

'job' mode: admin.

'marital'mode: married

'education' mode: university.degree

'default' mode: no

'housing' mode: yes

'loan' mode: no

* **Method 2: Just turn unknown to 0, no need to handle it.**

**Target 2 : Convert descriptive variables to numeric ones**

1. Job

|  |  |
| --- | --- |
| Label | Code |
| unknown | 0 |
| admin. | 1 |
| blue-collar | 2 |
| entrepreneur | 3 |
| housemaid | 4 |
| management | 5 |
| retired | 6 |
| self-employed | 7 |
| services | 8 |
| student | 9 |
| technician | 10 |
| unemployed | 11 |

1. Marital

|  |  |
| --- | --- |
| Label | Code |
| unknown | 0 |
| single | 1 |
| married | 2 |
| divorced | 3 |

1. Education

|  |  |
| --- | --- |
| Label | Code |
| unknown | 0 |
| illiterate | 1 |
| basic.4y | 2 |
| basic.6y | 3 |
| basic.9y | 4 |
| high.school | 5 |
| university.degree | 6 |
| professional.course | 7 |

1. Housing/loan/default

|  |  |
| --- | --- |
| Label | Code |
| unknown | 0 |
| no | 1 |
| yes | 2 |

1. Contact

|  |  |
| --- | --- |
| Label | Code |
| cellular | 0 |
| telephone | 1 |

1. Poutcome

|  |  |
| --- | --- |
| Label | Code |
| nonexistent | 0 |
| failure | 1 |
| success | 2 |

1. Day of week

|  |  |
| --- | --- |
| Label | Code |
| mon | 1 |
| tue | 2 |
| wed | 3 |
| thu | 4 |
| fri | 5 |

1. Month

|  |  |
| --- | --- |
| Label | Code |
| mar | 3 |
| apr | 4 |
| may | 5 |
| jun | 6 |
| jul | 7 |
| aug | 8 |
| sep | 9 |
| oct | 10 |
| nov | 11 |
| dec | 12 |

1. Subscribe

|  |  |
| --- | --- |
| Label | Code |
| no | 1 |
| yes | 2 |