Data Mining for Business Dr. Kamesam

# **Group Project Dataset Summary Sheet**

1. **Section and Team Members**: 001: Akshita Khandelwal, Natani Drati, Souritya Saha

2. **Dataset Name**: Bank Marketing dataset

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

4. **Source**: UCI data mining Archive

5. **No of records in the dataset:** 41,188

(**pick a dataset that has at least 1000 records, preferably a lot more)**

6. T**arget column in dataset**: y-has the client subscribed a term deposit? **Datatype:** binary:yes/no

7. **No of Features** 19 (Excluding the 11th attribute,”Duration” with the intention of having a realistic predictive model)

8. **Brief Description of the dataset**: This is related to the direct marketing campaigns of a Portuguese banking institution. The entire marketing campaigns were based on reaching clients via phone call. During the campaigns, often more than one contact to the same client was required in order to assess if the bank term deposit would be subscribed. The classification goal is to predict if the client will subscribe(yes/no) to a term deposit(variable y).

9. A **table showing every column in the dataset**, a brief description of the data, data type

Please see the attached table in the last

10. **Data Mining Task(s):** Classification

11. **If Classification, how many classes in the data:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Attribute (Feature)** | **Data type** | **Range of Values** |
| 2 | job | Categorical | 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown’ |
|  | marital | Categorical | 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed |
|  | education | Categorical | 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown' |
|  | default | Categorical | 'no','yes','unknown' |
|  | housing | Categorical | 'no','yes','unknown' |
|  | loan | Categorical | 'no','yes','unknown' |
|  | contact | Categorical | 'cellular','telephone' |
|  | month | Categorical | ‘jan’,’feb’,’mar’.’apr’,’may’,’jun’,’jul’,’aug’,’sep’,’oct’,’nov’,’dec’ |
|  | day\_of\_week | Categorical | 'mon','tue','wed','thu','fri |
|  | poutcome | Categorical | 'failure','nonexistent','success' |
|  | **y** | **Categorical** | **binary: 'yes','no'** |

12. **What ML/DM techniques do you plan to apply?** We plan to use Decision Trees, Artificial Neural Networks, and Logistic Regression.

13. **What do you hope to learn** (understand) or **predict from the data mining**

From the data mining we hope to develop the best model that predicts whether a client will subscribe to a term deposit via the marketing campaign efforts of the Portuguese banking institution using the data we have on clients and the campaign itself.

14. **How can the data mining results be made use of?**

By identifying the trends of clients who subscribe to a term deposit, the Portugal banking institution can tailor their marketing campaigns to reach clients who align with these trends. Doing so will increase the marketing campaign efficiency and productivity and consequently increase the business revenue of the banking institution by acquiring more subscribed clients.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Attribute (Feature)** | **Data type** | **Range of Values** |
| 1 | age | Numeric | 17-98 |
| 2 | job | Categorical | 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown’ |
| 3 | marital | Categorical | 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed |
| 4 | education | Categorical | 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown' |
| 5 | default | Categorical | 'no','yes','unknown' |
| 6 | housing | Categorical | 'no','yes','unknown' |
| 7 | loan | Categorical | 'no','yes','unknown' |
| 8 | contact | Categorical | 'cellular','telephone' |
| 9 | month | Categorical | ‘Jan’,’feb’,’march’.’april’,’may’,’june’,’july’,’august’,’september’,’october’,’november’,’december’ |
| 10 | day\_of\_week | Categorical | 'mon','tue','wed','thu','fri |
| 11 | duration | Numeric | 0 - 4918 (in seconds) |
| 12 | campaign | Numeric | 1-56 (number of contacts performed during this campaign and for this client) |
|  | pdays | Numeric | 0-999 (999 means client was not previously contacted) |
|  | previous | Numeric | 0-7 (number of contacts performed before this campaign and for this client) |
|  | poutcome | Categorical | 'failure','nonexistent','success' |
|  | emp.var.rate | Numeric | -3.400 - 1.400 (quarterly indicator) |
|  | cons.price.idx | Numeric | 92.201- 94.767 (monthly indicator) |
|  | cons.conf.idx | Numeric | -50.8 - -26.9 (monthly indicator) |
|  | euribor3m | Numeric | 0.634 - 5.045 (daily indicator) |
|  | nr.employed | Numeric | 4963 - 5228 (quarterly indicator) |
|  | **y** | **Categorical** | **binary: 'yes','no'** |

1

Data Mining for Business