**Machine Learning Project - Bank Marketing**

**1. Data Source**

<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

The data for this project has been selected from the bank marketing industry. In this project, I have used a datasource composed of twenty predictors that will help to determine if a client will agree to open a term deposit with an unnamed Portuguese banking institution. A term deposit is a cash investment held at a financial institution where money is invested for an agreed rate of interest over a fixed amount of time called the term. Term deposits can be invested into a bank, building society, or credit union.

This dataset represents data related to direct marketing campaigns of the Portuguese banking institution seeking to promote their term deposit accounts via telephone marketing campaigns. Often more than one call to the same client was required in order to determine if the client would agree to subscribe to a term deposit account ('yes') or not ('no').

**2. Number of instances of the dataset**

41,188

**3. Number of attributes of the dataset**

20 Input features + Output feature.

**4. Description of attributes**

|  |  |  |  |
| --- | --- | --- | --- |
| SNo | Input variable | Data Type | Description |
| 01 | age | Numerical | Age of the customer |
| 02 | job | Categorical | Type of job |
| 03 | marital | Categorical | Marital Status |
| 04 | education | Categorical | Highest degree of customer |
| 05 | default | Categorical | Has credit in default |
| 06 | housing | Categorical | Has housing loan |
| 07 | loan | Categorical | Has personal loan |
| 08 | contact | Categorical | Contact communication type |
| 09 | month | Categorical | Last contact month of year |
| 10 | day\_of\_week | Categorical | Last contact day of the week |
| 11 | duration | Numerical | Last contact duration, in seconds |
| 12 | campaign | Numerical | Number of contacts performed during this campaign and for this client |
| 13 | pdays | Numerical | Number of days that passed by after the client was last contacted from a previous campaign |
| 14 | previous | Numerical | Number of contacts performed before this campaign and for this client |
| 15 | poutcome | Categorical | Outcome of the previous marketing campaign |
| 16 | emp.var.rate | Numerical | Employment variation rate - quarterly indicator |
| 17 | cons.price.idx | Numerical | Consumer price index - monthly indicator |
| 18 | cons.conf.idx | Numerical | Consumer confidence index - monthly indicator |
| 19 | euribor3m | Numerical | Euribor 3 month rate - daily indicator |
| 20 | nr.employed | Numerical | Number of employees - quarterly indicator |
|  | **Y** – Output variable | Categorical | As the client subscribed a term deposit |

**5. Useful Tools/Methods for the study**

**Matlab** : Predictive Analytics [Possible algorithms include but not limited to: Logistic Regression, SVM, Naive Bayes, Linear Regression, Random Forest, Neural Network, etc.]

**Excel** : for managing, coordinating column definition and selection, and for data load.

**Matlab** : for dataset profiling and data analysis

**6. Exactly what problems/questions does the project plans to predict/study**

I am using the data to build a model that will predict the client’s decision to purchase a term deposit subscription. The classification goal is to predict future clients for deposit depending on if the client will subscribe to a term deposit in the bank, which will be indicated by ‘yes’ or ‘no’.

**7. What was the nature of the experiment?**

This project will experiment with the following models to explore the dataset:

1. **Logistical Regression:** Explore the model with and without Lasso regularization to determine which predictors are important.
2. **SVM:** Explore the model with and without the predictors that were eliminated by Lasso.
3. **SVM with kernel**: Explore the SVM model with the ‘rbf’ kernel, with and without the predictors that were eliminated by Lasso.
4. **Neural Networks:** Explore the classification problem using two neural networks of different architectures.

**8. What were the results of the experiment?**

Based on the Lasso experiment, I have determined the following predictors to improve the likelihood of the customer opening an account:

|  |  |  |  |
| --- | --- | --- | --- |
| SNo | Input variable | Data Type | Description |
| 09 | Month (August) | Categorical | Last contact month of year |
| 11 | duration | Numerical | Last contact duration, in seconds |
| 15 | poutcome | Categorical | Outcome of the previous marketing campaign |

I determine the following predictors to decrease the likelihood of the customer opening an account:

|  |  |  |  |
| --- | --- | --- | --- |
| SNo | Input variable | Data Type | Description |
| 09 | Month (September) | Categorical | Last contact month of year |
| 13 | pdays | Numerical | Number of days that passed by after the client was last contacted from a previous campaign |
| 16 | emp.var.rate | Numerical | Employment variation rate - quarterly indicator |
| 20 | nr.employed | Numerical | Number of employees - quarterly indicator |

Also determined that the Neural Network is the best classification model for making predictions based on our dataset. This decision was based on the Neural Network having the highest accuracy, recall, precision, and F1 scores amongst the models. While making this decision, I decided to emphasize the scores of the minority class (opens account), which is most important for the business.