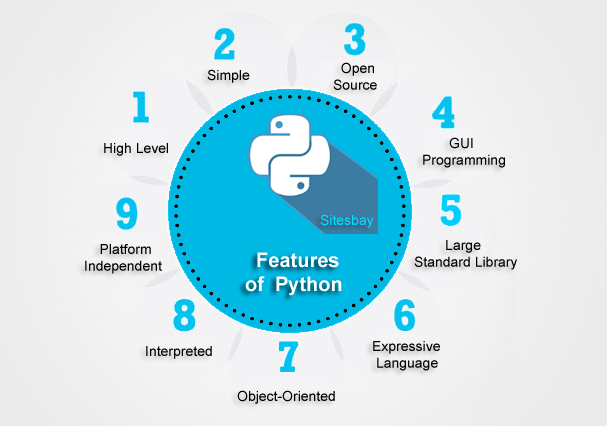
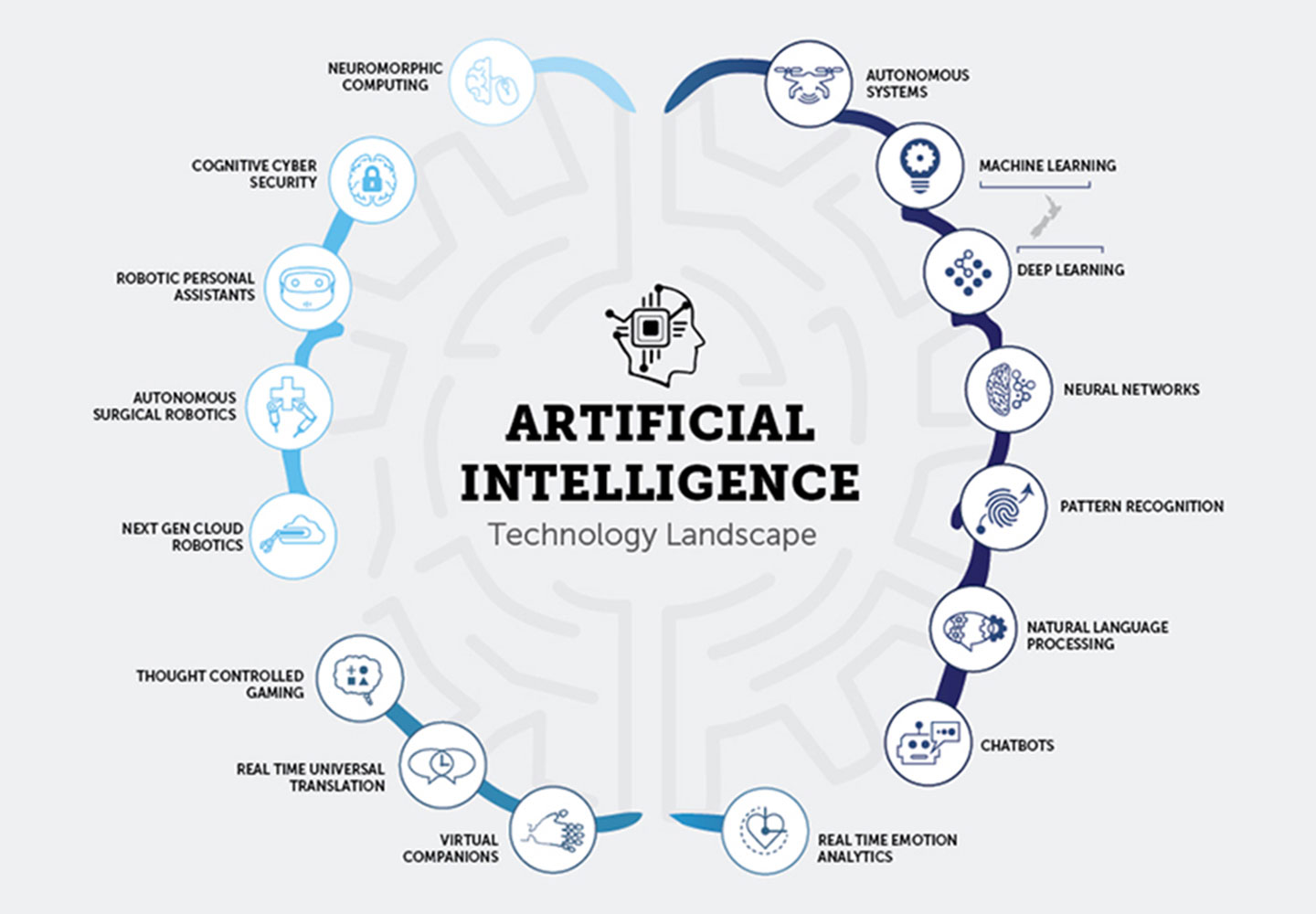
**1. TERM DEPOSIT SUBSCRIPTION PREDICTION**

**1.1 Introduction**

Python is an interpreted, [high-level](https://en.wikipedia.org/wiki/High-level_programming_language), [general-purpose](https://en.wikipedia.org/wiki/General-purpose_programming_language) [programming language](https://en.wikipedia.org/wiki/Programming_language). Created by [Guido van Rossum](https://en.wikipedia.org/wiki/Guido_van_Rossum) and first released in 1991, Python's design philosophy emphasizes [code readability](https://en.wikipedia.org/wiki/Code_readability) with its notable use of [significant whitespace](https://en.wikipedia.org/wiki/Off-side_rule). Its language constructs and [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming) approach aims to help programmers write clear, logical code for small and large-scale projects.Python is [dynamically typed](https://en.wikipedia.org/wiki/Dynamic_programming_language) and [garbage-collected](https://en.wikipedia.org/wiki/Garbage_collection_(computer_science)). It supports multiple [programming paradigms](https://en.wikipedia.org/wiki/Programming_paradigm), including [procedural](https://en.wikipedia.org/wiki/Procedural_programming), object-oriented, and [functional programming](https://en.wikipedia.org/wiki/Functional_programming). Python is often described as a "batteries included" language due to its comprehensive [standard library](https://en.wikipedia.org/wiki/Standard_library).



In [computer science](https://en.wikipedia.org/wiki/Computer_science), artificial intelligence(AI), sometimes called machine intelligence, is [intelligence](https://en.wikipedia.org/wiki/Intelligence) demonstrated by [machines](https://en.wikipedia.org/wiki/Machine), in contrast to the natural intelligence displayed by humans and animals. Colloquially, the term "artificial intelligence" is used to describe machines that mimic "cognitive" functions that humans associate with other [human minds](https://en.wikipedia.org/wiki/Human_mind), such as "learning" and "problem solving".As machines become increasingly capable, tasks considered to require "intelligence" are often removed from the definition of AI, a phenomenon known as the [AI effect](https://en.wikipedia.org/wiki/AI_effect). A quip in Tesler's Theorem says "AI is whatever hasn't been done yet." For instance, [optical character recognition](https://en.wikipedia.org/wiki/Optical_character_recognition) is frequently excluded from things considered to be AI, having become a routine technology.



**1.2 Objective of Research**

Our project is to predict term deposit subscription of clients in a bank. A term deposit is a fixed-term investment that includes the deposit of money into an account at a financial institution.

When an account holder deposits funds at a bank, the bank can use that money to lend to other consumers or businesses. In return for the right to use these funds for lending, they will pay the depositor compensation in the form of interest on the account balance. A customer will deposit, or invest, in one of these accounts agreeing not to withdraw their funds for a fixed period in return for a higher rate of interest paid on the account. The interest earned on a term deposit account is slightly higher than that paid on standard savings or interest-bearing checking accounts.

Our objective is to build a classifier to predict whether or not a client will subscribe a term deposit. If the classifier has high accuracy, the banks can arrange a better management of available resources by focusing on the potential customers “picked” by the classifier, which will improve their efficiency a lot. Besides, we plan to find out which factors are influential to customers’ decision, so that a more efficient and precise campaign strategy can be designed to help to reduce the costs and improve the profits.

**1.3 Problem statement**

Direct Marketing is the practice of delivering promotional messages directly to current or prospective customers on an individual basis rather than using a mass medium. Predictive models are a great tool in analyses used to assess and increase the effectiveness of such marketing campaigns.

Logistic regression remains one of the most popular techniques used to predict customer behavior. Millions of dollars are spent annually on marketing activities that utilize logistic regression models. Therefore, it is essential to build robust logistic models that have strong predictive ability for a successful direct marketing campaign.

**2. REVIEW OF LITERATURE**

Direct Marketing is the practice of delivering promotional messages directly to current or prospective customers on an individual basis rather than using a mass medium. Predictive models are a great tool in analyses used to assess and increase the effectiveness of such marketing campaigns.

The increasing number of marketing campaigns over time has reduced their effects on the general public. First, due to competition, positive response rate to mass campaigns are typically very low, according to a recent study, less than 1% of the contacts will subscribe a term deposit. Second, direct marketing has drawbacks, such as causing negative attitude towards banks due to intrusion of privacy. In order to save costs and time, it is important to filter the contacts but keep a certain success rate.

**3. DATA COLLECTION**

The data was obtained from the UCI Machine Learning Repository, and relates to the direct marketing campaigns of a Portuguese banking institution attempting to get its clients to subscribe to a term deposit. The marketing campaigns were conducted by making multiple phone calls to the clients. The client responses and predictor variable information are used to assess whether the subject will subscribe to the bank term deposit or not.

The dataset has 10 predictor variables like age, duration, campaign, pdays, previous, emp.var.rate, cons.price.idx, cons.conf.idx, euribor3m and nr.employed.

Attribute information:

Input variables:

1. age – age of the client
2. duration - last contact duration, in seconds
3. campaign - number of contacts performed during this campaign and for this client
4. pdays - number of days that passed by after the client was last contacted from a previous campaign
5. previous - number of contacts performed before this campaign and for this client
6. emp.var.rate - employment variation rate
7. cons.price.idx - consumer price index - monthly indicator
8. cons.conf.idx - consumer confidence index - monthly indicator
9. euribor3m - euribor 3 month rate - daily indicator
10. nr.employed - number of employees - quarterly indicator

Output variable:

subscribe - has the client subscribed a term deposit or not

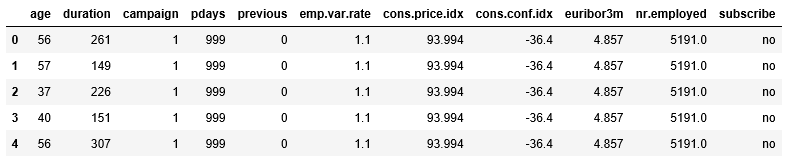


Fig. 3.1: Dataset of our project

**4. METHODOLOGY**

**4.1 Exploratory data analysis**

**4.1.1 Figures and tables**

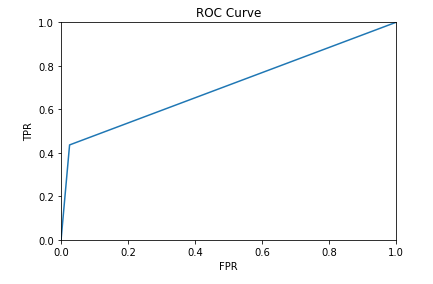


Fig. 4.1.1: Plot

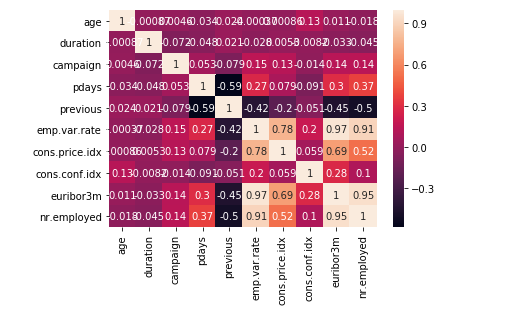


Fig. 4.1.2: Heat Map

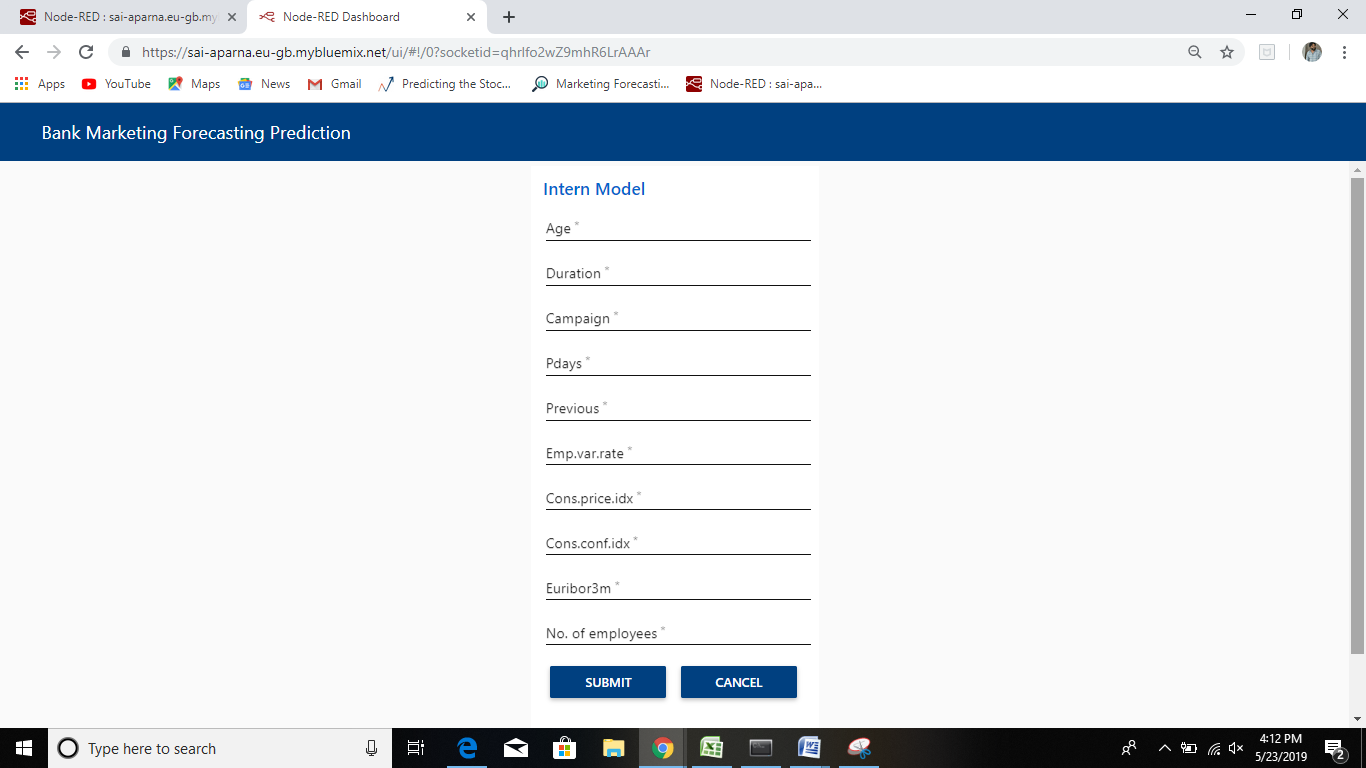


Fig. 4.1.3: User Interface

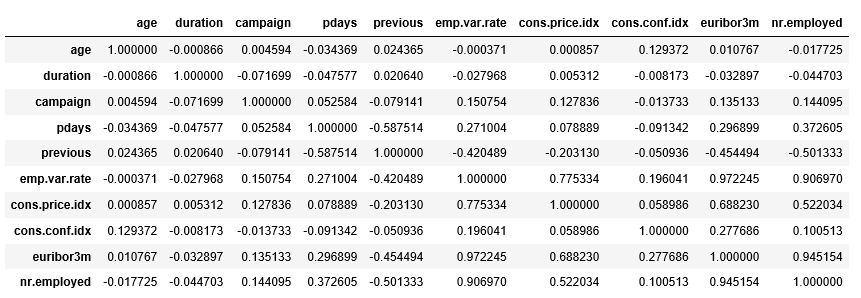


Fig. 4.1.4: Correlation

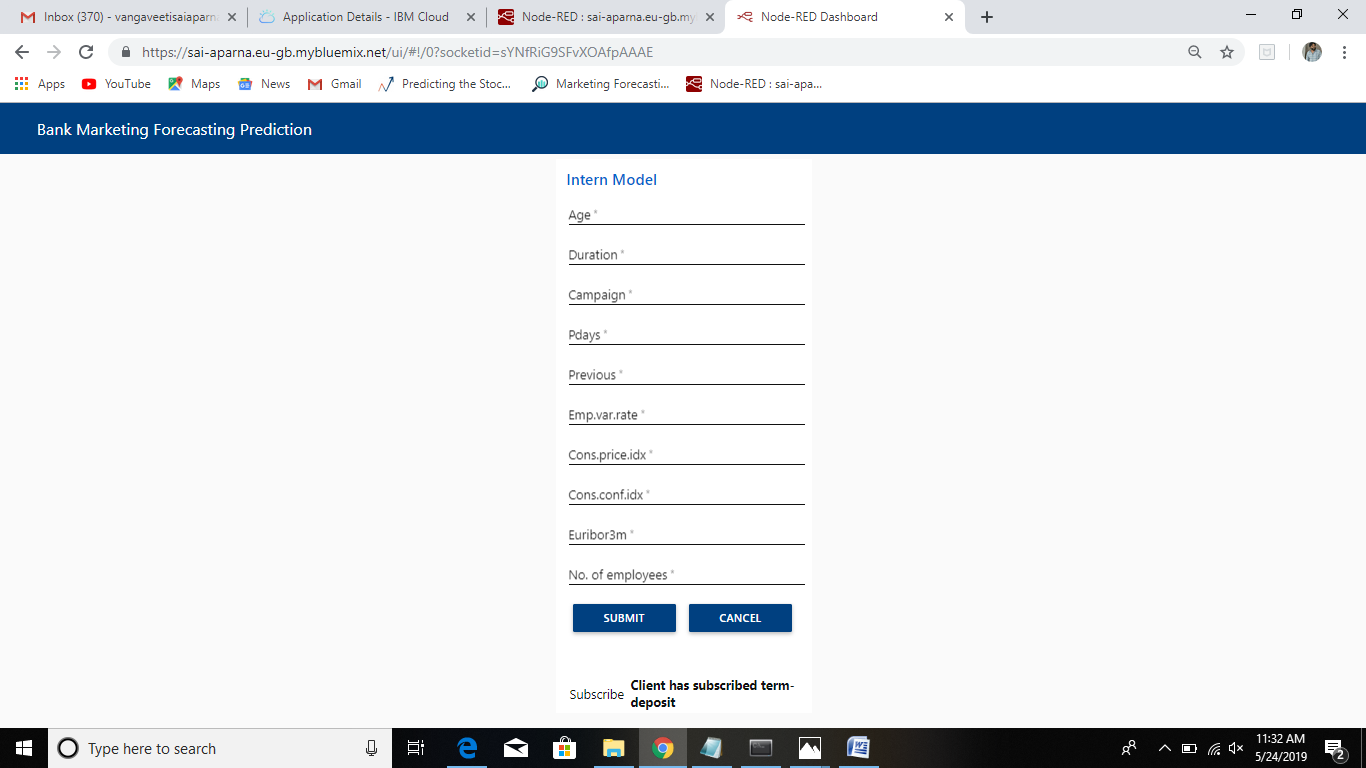


Fig. 4.1.5: Client has subscribed

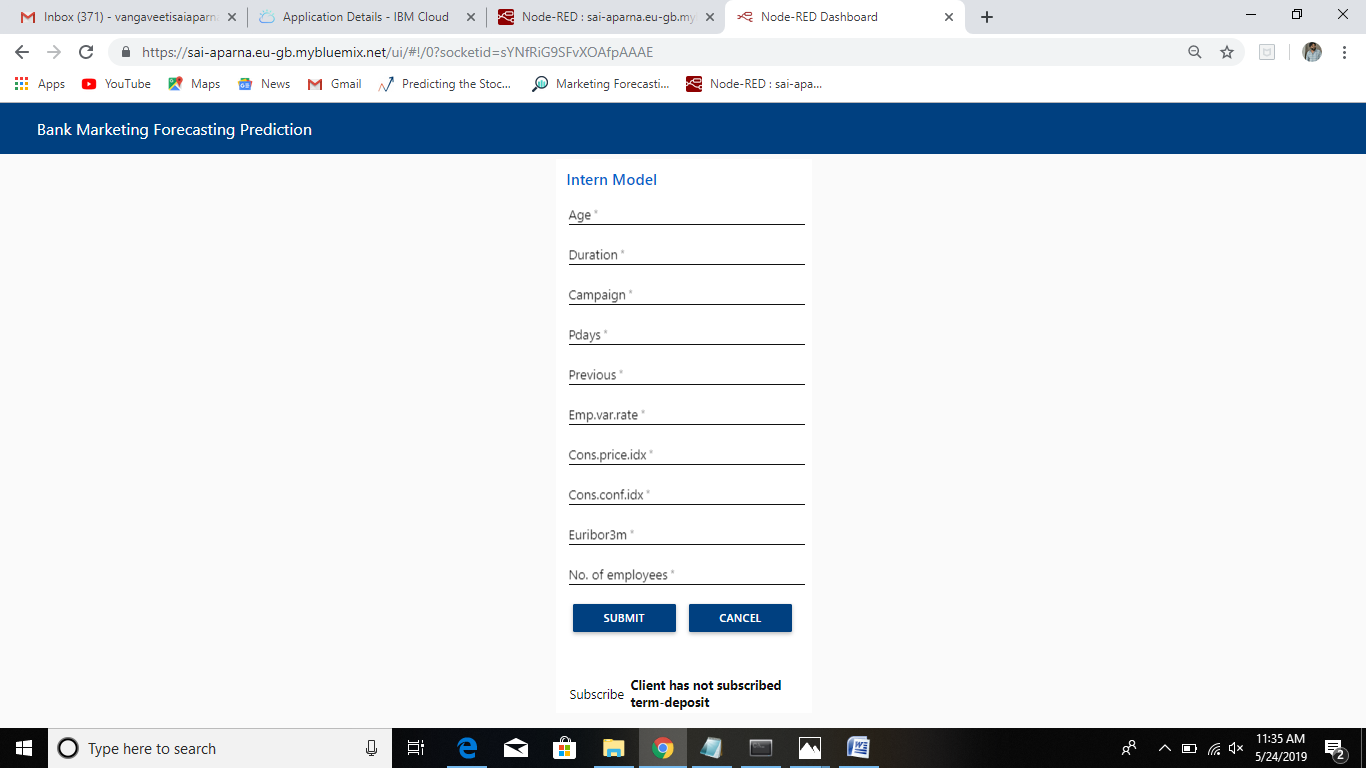


Fig. 4.1.6: Client has not subscribed

**4.2 Data Modeling**

In our project, we aim to increase campaign efficiency by identifying the main factors that affect the success of a campaign and predicting whether the campaign will be successful to a certain client, namely, whether the client will subscribe a term deposit. As our data are imbalanced, we use resampling methods before building models. After preprocessing the data, we build logistic regression model.

Logistic regression:

Logistic regression uses an equation as the representation, very much like linear regression. Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). A key difference from linear regression is that the output value being modeled is a binary value (0 or 1) rather than a numeric value. Below is an example logistic regression equation:

y = e^ (b0 + b1\*x) / (1 + e^ (b0 + b1\*x))

Where, y is the predicted output, b0 is the bias or intercept term and b1 is the coefficient for the single input value (x). Each column in your input data has an associated b coefficient (a constant real value) that must be learned from your training data.

The actual representation of the model that you would store in memory or in a file are the coefficients in the equation (the beta value or b’s).

The logistic regression model takes real-valued inputs and makes a prediction as to the probability of the input belonging to the default class (class 0). If the probability is > 0.5 we can take the output as a prediction for the default class (class 0), otherwise the prediction is for the other class (class 1). For this dataset, the logistic regression has three coefficients just like linear regression, for example:

Output = b0 + b1\*x1+b2\*x2+b3\*x3

The job of the learning algorithm will be to discover the best values for the coefficients (b0, b1 and b2) based on the training data. Unlike linear regression, the output is transformed into a probability using the logistic function:

P (class=0) = 1 / (1 + e^ (-output))

This would be written as:

P (class=0) = 1 / (1 + EXP (-output))

Logistic regression does NOT assume a linear relationship between the dependent variable and the independent variables, but it does assume linear relationship between the logit of the explanatory variables and the response.

Independent variables can be even the power terms or some other nonlinear transformations of the original independent variables. The dependent variable does NOT need to be normally distributed, but it typically assumes a distribution from an exponential family (e.g. binomial, Poisson, multinomial, normal…) binary logistic regression assume binomial distribution of the response.

Steps to implement logistic regression algorithm:

* Load the data set.
* Understanding the data.
* Split the data into training and test dataset.
* Use the training dataset to model the logistic regression model.
* Calculate the accuracy of the trained model on the training dataset.
* Calculate the accuracy of the model on the test dataset.

**5. REFERENCES**

1. <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>
2. <http://www.columbia.edu/~jc4133/ADA-Project.pdf>
3. <http://dataaspirant.com/2017/04/15/implement-logistic-regression-model-python-binary-classification>
4. [https://www.kaggle.com](https://www.kaggle.com/)
5. [https://data.gov.in](https://data.gov.in/)
6. [https://www.analyticsvidhya.com](https://www.analyticsvidhya.com/)

**6. CONCLUSION**

Finally in our project by using logistic regression model we predict whether the client has subscribed the term deposit or not. In order to implement this various input variables to get the output.

Whenever program takes the input data it gives the output in the form of binary i.e., either zero or one. If the output is 1 then we get the message ‘client has subscribed the term deposit’ and if the output is 0 then we get the message ‘client has not subscribed the term deposit’.