\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Direct Marketing using logistic regression

Saheer, Madan

\*Major in Data Science, Praxis Business School, Bangalore.

**Abstract**: Direct marketing is a form of [advertising](https://en.wikipedia.org/wiki/Advertising) where organizations communicate directly to [customers](https://en.wikipedia.org/wiki/Customer) through a variety of media including [cell phone](https://en.wikipedia.org/wiki/Mobile_phone) [text messaging](https://en.wikipedia.org/wiki/Text_messaging), [email](https://en.wikipedia.org/wiki/Email), [websites](https://en.wikipedia.org/wiki/Websites), [online adverts](https://en.wikipedia.org/wiki/Online_advertising), [database marketing](https://en.wikipedia.org/wiki/Database_marketing), [fliers](https://en.wikipedia.org/wiki/Flyer_(pamphlet)), [promotional letters](https://en.wikipedia.org/wiki/Advertising_mail), [targeted television](https://en.wikipedia.org/wiki/Television_advertisement), [newspapers](https://en.wikipedia.org/wiki/Newspapers), [magazine advertisements](https://en.wikipedia.org/wiki/Magazine), and [outdoor advertising](https://en.wikipedia.org/wiki/Outdoor_advertising). Among practitioners, it is also known as direct response marketing.

Direct marketing models predict responses to some contact that will be made in the future, helping the organization decide which customers to target is the main objective. So, for this purpose to solve this classification problem we are using logistic regression.

Logistic Regression (also called Logit Regression) is commonly used to estimate the probability that an instance belongs to a particular class (e.g., what is the probability that this email is spam?). If the estimated probability is greater than the threshold probability, then the model predicts that the instance belongs to that class (called the positive class, labeled “1”), or else it predicts that it does not (i.e., it belongs to the negative class, labeled “0”). This makes it a binary classifier.

Using data from a company, we evaluate our model empirically by testing whether our assumptions hold, examine the extent of variation in predicted values across models build from various previous data, evaluate the amount of improvement over extant models in terms of prediction error and performance as measured by a gains table. Hence, we target the customers who are in the top section of gains table.