Email Campaign Effectiveness Prediction

Karan Rawat, Bhaskar Mendhe
Data science trainees,
AlmaBetter, Bangalore

Abstract:

Email marketing is one
of the most modern means and marketing
methods in the world. The modern
information and
communication technologies also facilitate
the circulation and sending of electronic
messages with the
highest quality and accuracy. In addition,
Email marketing campaigns help to
increase the sale of
products in electronic shops and to target
customers efficiently and legally.
However, despite the
benefits of Email marketing, there are the
firm chances that the emails may not be
acknowledged by the clients. In this
project we are going to see how much
percent does the clients and customers
acknowledge the emails by using various
regression models and what are the
reasons that they are not taken into
consideration. Some of the inferences are
also drawn from the data that may lead to
low effectiveness of the Email Campaign.

***Keywords: Machine Learning, Email
Campaign Success, Email Marketting***

1.Problem Statement

Most of the small to medium business
owners are making effective use of Gmail-
based Email marketing Strategies for offline
targeting of converting their prospective
customers into leads so that they stay with
them in Business. The main objective is to
create a machine learning model to
characterize the mail and track the mail that
is ignored, read, acknowledged by the
reader.

Email ID - This column contains the email
ids of individuals.

Email type - Email type contains 2
categories 1 and 2. We can assume that the
types are like promotional email or sales
email.

Subject Hotness Score - It is the subject-line
effectiveness score.

Email Source - It represents the source of
the email like sales marketing or product
type email.

Email Campaign Type - Campaign type

Total Past Communications - This column
contains the previous mails from the source.

Customer Location - Categorical data which
explains the different demographic location
of the customers.

Time Email sent Category - It has 3
categories: 1,2 and 3 which are considered
as morning, evening and night time slot.

Word Count - It contains the number of
words in the mail.

Total Links - Total links in the email body.

Total Images - The banner images in the
email body.

Email Status - It is the target variable which
contains the characterization of the mail that
is ignored; read; acknowledged by the
reader.

2. Introduction

Electronic mail (or e-mail or email) is an Internet service that allows people who have an e-mail address (accounts) to send and receive electronic letters. Those are much like postal letters, except that they are delivered much faster than snail mail when sending over long distances, and are usually free.

The very first email was sent in 1971 by a computer engineer named Ray Tomlinson. The message he sent was just a string of numbers and letters, but it was the beginning of a new era of communication. Tomlinson was also the person who introduced the usage of the "@" symbol in email addresses.

In 1978, a marketing manager at Digital Equipment Corp named Gary Thuerk used this new method of direct communication to send out the first commercial email to let people know about a new product.

By the '90s, the internet had become commercially available to the masses. The way people communicated with one another began to change dramatically, and marketers discovered that email could be an effective way to advertise. The emergence of marketing emails also ushered in the need for regulatory updates; the U.K.'s Data Protection Act, for example, was adjusted to require an "opt out" option for all marketing emails.

Email has become such a popular marketing tool for businesses partly because it forces the user to take some kind of action; an email will sit in the inbox until it's read, deleted, or archived.

3. Email Marketing and Types

There are many different types of email marketing. Each one serves a different purpose and takes a different avenue to engage with your audience. We are going to look at some of the many different types, so you can create the best email marketing campaign for your company

1. **Welcome emails** :
   This type of email welcomes customers and encourages them to learn more about your product or service. They often offer a trial or other bonus. It is used to introduce a potential new customer to the business.

2. **Newsletter emails :**
   Newsletter emails are very popular, and they often highlight new products and services. They may also include articles, blogs, and customer reviews. Usually, there will be a call to action to move the reader to do something, whether that is reading a new blog post or checking out a new product.

3. **Lead nurturing emails :**
   This type of email targets a specific audience through a series of emails in the hope of eventually converting them. Typically, lead nurturing emails focus on a group that is interested in a specific product or service and then build their interest through more emails that offer additional information or

relevant promotions. The goal is to push users from the consideration stage to the purchasing stage.

4. **Confirmation emails :**
   Those that have recently signed up for emails or newsletters, or have purchased an item online for the first time may get a confirmation email. This ensures the prospect that the information has been received and they are on the list to receive additional information. These are also a way to let users know that their purchase has been received or that their sign-up was successful and can include more actions for them to take.

5. **Dedicated emails :**
    If you want to reach out to only a portion of your email list, this is called a dedicated email. Its list may be based on recent purchases, inactive clients, new members, and other specific types of criteria.

6. **Invite emails :**
   These types of emails often announce upcoming events, new product launches, and seminars. Most companies use these types of emails when there is something special going on to gain attention and increase awareness about special events.

7. **Promotional emails :**
   These types of marketing emails are very common and tend to be generic and go out to a large audience. They are usually used to maintain awareness and may tease new products and services.

8. **Survey emails :**
   Feedback from customers is one of the best tools for a business. Sending out these emails communicates to your customers that you value their opinion and want to create an experience, product, or whatever you're offering that they'll enjoy. Businesses can also take the feedback from these surveys and apply them to their offerings, creating what is hopefully a better product.

9. **Seasonal marketing emails :**
   Many companies take advantage of the holiday season or special occasions to reach out to their customers and prospects with information on upcoming sales and promotions. They are often tied to holidays like Christmas, Valentine's Day, Mother's, and Father's Day.

4. Why Is Email Campaign Effective?

- **Cost**

You don't have to spend a single dollar to start an email marketing campaign. All you need are a free email account and a form on your website to collect email addresses.

You can increase ROI by working with email services. Many of them offer free

versions of their software. As you build your email list, you might have to pay a small fee. However, more contacts can translate to more sales.

Email marketing does come with one cost: Time. However, it's the best option when you have a shoestring marketing budget (and even if you don't).

There are times when so many people are requesting rides that there aren't enough cars on the road to help take them all. Bad weather, rush hour, and special events, for instance, may cause unusually large numbers of people to want to request a ride with Sigma all at the same time.

● Success Rates

Segment your audience, so you send the ideal message to each consumer. Email campaigns let you customize each missive for greater impact.

For instance, let's say you run an e-commerce shoe company. You sell boots, flats, heels, sandals, sneakers, loafers, and dozens of other types of shoes.

Women might visit your website to buy a pair of the hottest new wedges. They represent a distinct segment of your audience. They're not looking for the same thing as a man who needs flip-flops to wear around the house.

● Measurability

Data has become the most valuable tool for any business owner. If you keep track of key metrics, you know when to adjust your strategy.For instance, you can measure open rates. If most of your subscribers delete emails unread, maybe you need to change

your subject lines. The same goes for engagement levels. When nobody clicks on your links, you might want to change your content strategy.

Measurability makes an email marketing strategy more malleable. Adjust your strategy as you go without much effort or any cost.

5. Steps involved:

- **Exploratory Data Analysis**
  After loading the dataset we performed this method by comparing our target variable that is **Email Status** with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

- **Null values Treatment**
  Our dataset contains a large number of null values which might tend to disturb our accuracy hence we plot distribution for each feature and after analyzing we imputed null values by their means and modes. In Customer Locations we imputed null values using "fillna" command with the phrase "missing value.

● **Encoding of categorical columns**

We used One Hot Encoding to produce binary integers of 0 and 1 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

● **Feature Engineering**

In these, By generating polynomial features, we uncovered potential new relationships between the features and the target and improvised the model's performance.

● **Standardization of features**

Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it.
The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

● **Fitting different models**

For modelling we tried various classification algorithms like:

1. **Logistic Regression**
2. **SVM Classifier**
3. **Random Forest Classifier**
4. **XGBoost classifier**

● **Tuning the hyperparameters for better accuracy**

Tuning the hyperparameters of respective algorithms is necessary for getting better accuracy and to avoid overfitting in case of tree based models like Random Forest Classifier and XGBoost classifier.
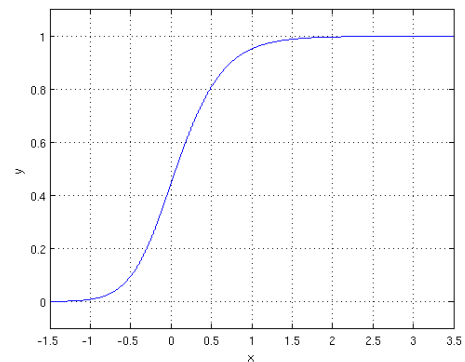
6. Algorithms:

1. Logistic Regression:

Logistic Regression is actually a classification algorithm that was given the name regression due to the fact that the mathematical formulation is very similar to linear regression.
The function used in Logistic Regression is sigmoid function or the logistic function given by:
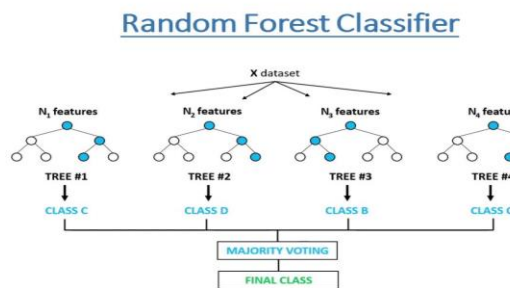
$$f(x) = 1/1 + e^{-x}$$

The optimization algorithm used is: Maximum Log Likelihood. We mostly take log likelihood in Logistic:

$$\ln L(\mathbf{y}, \beta) = \ln \prod_{i=1}^{n} f_i(y_i) = \sum_{i=1}^{n} \left[y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) \right] + \sum_{i=1}^{n} \ln(1-\pi_i)$$

2. Random Forest Classifier:

Random Forest is a bagging type of Decision Tree Algorithm that creates a number of decision trees from a randomly selected subset of the training set, collects the labels from these subsets and then averages the final prediction depending on the most number of times a label has been predicted out of all.

Random Forest Classifier

3. XGBoost :

To understand XGBoost we have to know gradient boosting beforehand.

- **Gradient Boosting-**
  Gradient boosted trees consider the special case where the simple model is a decision tree

In this case, there are going to be 2 kinds of parameters P: the weights at each leaf, w, and the number of leaves T in each tree (so that in the above example, T=3 and w=[2, 0.1, -1]).
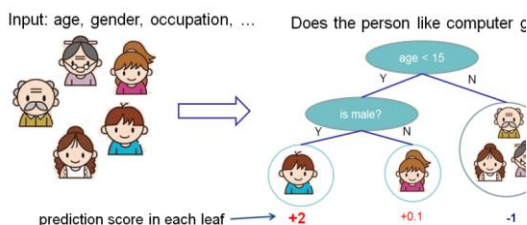
When building a decision tree, a challenge is to decide how to split a current leaf. For instance, in the above image, how could I add another layer to the (age > 15) leaf? A 'greedy' way to do this is to consider every possible split on the remaining features (so, gender and occupation), and calculate the new loss for each split; you could then pick the tree which most reduces your loss.

XGBoost is one of the fastest implementations of gradient boosting. trees. It does this by tackling one of the major inefficiencies of gradient boosted trees: considering the potential loss for all possible splits to create a new branch (especially if you consider the case where there are thousands of features, and therefore thousands of possible splits). XGBoost tackles this inefficiency by looking at the distribution of features across all data points in a leaf and using this information to reduce the search space of possible feature splits.

4. Decision Tree :

Decision Trees are a type of Supervised Machine Learning (that is you explain what the input is and what the corresponding output is in the training data) where the data is continuously split according to a certain parameter. The tree can be explained by two entities, namely decision nodes and leaves. The leaves are the decisions or the final outcomes. And the decision nodes are where the data is split.
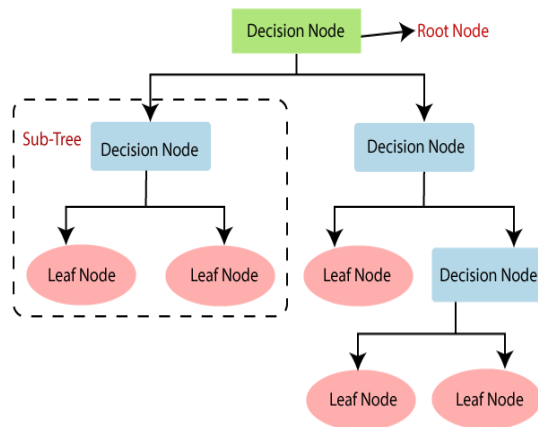

5. Naive Bayes Classifier :

Naïve Bayes Classifiers are a collection of classification algorithms based on Bayes Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

The fundamental Naïve Bayes assumption is that each feature makes an:

- Independent
- Equal

Contribution to the outcome.
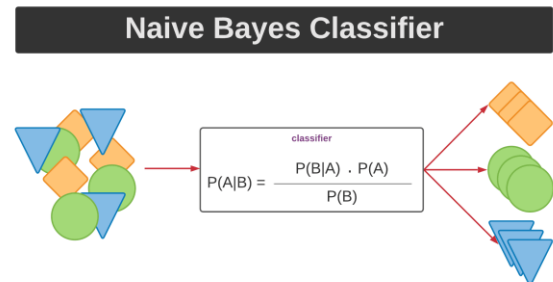
6.2. Model performance:

Model can be evaluated by various metrics such as:

1. Confusion Matrix-

The confusion matrix is a table that summarizes how successful the classification model is at predicting examples belonging to various classes. One axis of the confusion matrix is the label that the model predicted, and the other axis is the actual label.

2. Precision/Recall-

Precision is the ratio of correct positive predictions to the overall number of positive predictions : TP/TP+FP

Recall is the ratio of correct positive predictions to the overall number of positive examples in the set: TP/FN+TP

3. **Accuracy**-
   Accuracy is given by the number of correctly classified examples divided by the total number
   of classified examples. In terms of the confusion matrix, it is given by: TP+TN/TP+TN+FP+FN

4. **Area under ROC Curve(AUC)**-
   ROC curves use a combination of the true positive rate (the proportion of positive examples predicted correctly, defined exactly as recall) and false positive rate (the proportion of negative examples predicted incorrectly) to build up a summary picture of the classification performance.

6.3. Hyper parameter tuning:

Hyperparameters are sets of information that are used to control the way of learning an algorithm. Their definitions impact parameters of the models, seen as a way of learning, change from the new hyperparameters. This set of values affects performance, stability and interpretation of a model. Each algorithm requires a specific hyperparameters grid that can be adjusted according to the business problem. Hyperparameters alter the way a model learns to trigger this training algorithm after parameters to generate outputs.

We used Grid Search CV, Randomized Search CV and Bayesian Optimization for hyperparameter tuning. This also results in cross validation and in our case we divided the dataset into different folds. The best performance improvement among the three was by Bayesian Optimization.

1. **Grid Search CV-**Grid Search combines a selection of hyperparameters established by the scientist and runs through all of them to evaluate the model's performance. Its advantage is that it is a simple technique that will go through all the programmed combinations. The biggest disadvantage is that it traverses a specific region of the parameter space and cannot understand which movement or which region of the space is important to optimize the model.

2. **Randomized Search CV-** In Random Search, the hyperparameters are chosen at random within a range of values that it can assume. The advantage of this method is that there is a greater chance of finding regions of the cost minimization space with more suitable hyperparameters, since the choice for each iteration is random. The disadvantage of this method is that the combination of hyperparameters is beyond the scientist's control

7. Conclusion:

That's it! We reached the end of our exercise.
Starting with loading the data so far we have done EDA , null values treatment, encoding of categorical columns, feature Engineering and then model building.

The final model selected is Logistic Regression.

Taking Scores into consideration, XGBoost outperforms all models , but it hardly classify 1 or 2 points correctly from the minority class. At other hand Logistic Regression being simplest model able to correctly classify most number of points from minority class other than any model. Also Logistic Regression is very easy to interpret , as it fits a hyperplane for a classification.

Also We can derive feature importance from the coefficient of Logistic Regression very easily. To get the features which are important to classify points of class 0, we will look at the first array of coefficient. The more larger values of coefficients corresponding to features , then more the feature is important. Similarly we fetched the features which are important for classification of points belong to class 1 and class 2.

The "Email Campaign" as a whole feature is the most important individual feature for classification. The other features "Total_Past_Communications","Word_Count","Subject_Hotness_Score","Total Link Images" which is combination of "Total Links" and 'Total Images", are next important features which are used in combinations of polynomials for classifications. So we can look upon these features to improve the business work.

References-

1. Towards Data Science
2. GeeksforGeeks
3. Analytics Vidhya