

**PREDICTING THE CLICK-THROUGH FOR EMAIL
MARKETING CAMPAIGNS IN B2B E-COMMERCE
SECTOR.**

Amila Akalanka

A dissertation submitted in partial fulfilment of the requirement for
Master of Science Business Analytics

**School of Computing
Informatics Institute of Technology, Sri Lanka in Collaboration
with Robert Gordon University**

2022

Abstract

Email is one of the most popular and frequently used marketing channels to communicate with existing and potential customers in B2B and B2C markets. With the increased use of email as a main channel in marketing, many organizations want to adopt computational approaches to optimize the performance of their emails. This way, they can achieve higher success rates with less cost, thus resulting in higher ROI in promotional campaigns. In the case of emails, Click-Through (CT) is a critical performance indicator in email marketing campaigns. CT indicates how many recipients had opened the email and performed a click event. Predicting the Click-through event is important to identify customers who are more likely to be interested in a particular product. This work mainly focused on predicting “Clicks” and “Non-Clicks” events of emails using a machine learning approach. Further, natural language processing (NLP) techniques were used to explore the email content to support prediction.

This research was mainly driven by a two-step process. First, emails were cleaned and normalized. Language detection was performed based on samples of words in the email. Results of the language detection were used to retain emails written in English. Then, email content was analyzed to identify the natural grouping in emails using an unsupervised algorithm and natural language processing (NLP) techniques. For this K-Means clustering was used and 6 groups were identified based on the Elbow plot. A manual approach is taken to identify a description of each email group based on emails in each cluster. Then LDA was used to perform the same task with topics derived for each cluster. Then K-means groupings and LDA topic modeling results were compared while trying to identify an appropriate description for each email grouping. The language used in content offers a reliable source to signal of opening the email and eventually leading to a click event. Therefore, the email groups identified using clustering along with other email features were used in the prediction task as well.

Secondly, persona features and email features were examined together with findings from the first step. The model is based on company, recipient profile features, and features extracted from emails along with grouping identified using clustering.

SVM (Support Vector Machine), Logistic Regression, Decision tree, and Random Forest classification algorithms were used for the classification task and these algorithms were compared in terms of classification performance and training time. The grid search method was applied to determine parameter values to maximize F-1, ROC/AUC, BCR(Balanced correct classification), and accuracy.

The experiments in this study show that SVM and Random Forest have good performance when predicting Clicks in the unforeseen dataset. Out of these 3 performance measures (F-1, ROC/AUC, BCR), SVM outperformed other models in 2 measures (ROC/AUC and BCR) and Random Forest in the F-1 score. SVM had the highest training time while Logistic Regression trained the lowest average training time and out of 16 features used in the study Email type, Company stage in the Sales and Marketing funnel, email word count and length, location, and the identified email group from email content analysis were shown to be important features based on Random Forest feature importance analysis. Further, the Chi-square test for independence was used to check the existence of the relationship between several personas and profile features like Seniority, Department, etc. Test results show that there is a significant relationship between the features and clickthrough.

Keywords: ***B2B, Email Click-Throughs, Classification, Text analysis, Topic Modeling***