

Text analytics to improve telecommunication customers service management from unstructured data

M. Oliva, C. Amado ^a, P. Soares ^b and B. Oliveira ^c

^a CEMAT, IST, University of Lisbon, Portugal, ^b CEAUL, IST, University of Lisbon, Portugal, ^c Border Innovation, Lisbon, Portugal

Call centers are an important factor of the customer experience and, in their operation, they increasingly incorporate more communication channels, such as bots, chats and social media. This assortment of channels has increased the volume and variety of unstructured data that companies have to deal with. In this context, the current development of text mining techniques may contribute substantially to the analysis of this unstructured information, allowing the improvement of customers' experience and, consequently, the value of companies.

This work presents an application of text mining analysis of telecommunications customers' contacts to improve the management of that information by the development of a predictive model that may help the call center assistant during the process of troubleshooting the client problem.

The available dataset comprises 212473 text messages containing the transcriptions of every customer call during a 3-month period. Each transcription is labelled with the cause of the reported problem that is assigned by the assistant at the time of the call from a set of 153 possible causes. Naturally, this kind of analysis with real data raises several difficulties. The first challenge was to deal with unstructured information, since each call is different in the quality of the text and in the domain of the reported problem (television, internet, cellular or fixed phone). Secondly, modelling was complicated by the severe imbalance among all possible causes. This feature of the problem creates great difficulties to any predictive algorithm: a small number of more frequent causes tend to be overrepresented while rarer causes may never be predicted.

To handle this problem, Naïve Bayes, Random Forests (RF) and Recurrent Neural Networks (RNN) were applied. RF and RNN with an attention mechanism had a noteworthy performance leading to the prediction of the most probable cause in 91% of the cases.

Keywords: Multi-Class Text Classification, Natural Language Processing, Text Mining, Unstructured data, Telecommunication Services