Instituto Superior de Contabilidade e Administração

Politécnico de Coimbra

Data-Driven Marketing: A sentiment analysis study in Nonprofit Marketing

Marcus Vinicius Estrela Borges

ISCAC | 2022

Instituto Superior de Contabilidade e Administração

Politécnico de Coimbra

Marcus Vinicius Estrela Borges

Data-Driven Marketing: A sentiment analysis study in Nonprofit Marketing

Coimbra, October 2022

Instituto Superior de Contabilidade e Administração
Politécnico de Coimbra
COIMBRA BUSINESS SCHOOL
ISCAC.pt

Marcus Vinicius Estrela Borges

Data-Driven Marketing: A sentiment analysis study in Nonprofit Marketing

Dissertation submitted to the Higher Institute of Accounting and Administration of Coimbra to fulfil the requirements for obtaining a **master's degree in Data Analysis and Decision Support Systems**, conducted under the guidance of Professor Isabel Maria Mendes Pedrosa, Assistant Professor at Coimbra Business School, and co-supervised by Professor Sérgio Miguel Carneiro Moro, Associate Professor with Habilitation at ISCTE.

Coimbra, October 2022

STATEMENT OF RESPONSIBILITY

I declare that I am the author of this Dissertation, which is an original and unpublished work, which has never been submitted to another Higher Education Institution to obtain an academic degree or other qualification. I further certify that all citations are properly identified and that I am aware that plagiarism is a serious lack of ethics, which could result in the annulment of this dissertation.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ACKNOWLEDGEMENTS:

This work closes a crucial stage in my personal and professional development, I could not fail to thank everyone who contributed directly or indirectly to complete this dissertation.

To teachers Dr. Sérgio Moro and Dr. Isabel Pedrosa, I dedicate my most incredible gratitude for this dissertation's guidance and the unconditional support and motivation provided during this stage. I thank you for your scientific support, the availability and promptness with which you always assisted me, but mainly for your patience in the most critical moments and encouraging words.

To Coimbra Business School, in particular to the professors who marked my path. Thank you for the knowledge that formed me and the method that guided me. Thanks also to the rest of the technical staff for the professionalism and courtesy they always treated me.

Equally essential and heartfelt is my gratitude to my wife, Priscilla Pessoa Estrela, who motivated me so much throughout the time I carried out my work. To my parents, my brothers and my friends who contributed directly or indirectly to the realisation of this research.

Finally, I would like to express my deep and sincere thanks to Cision, especially Uriel Oliveira and Júlia Rodrigues, who dedicated some of their precious time to helping me with the Bradwatch platform.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ABSTRACT

The technological expansion in recent years has created opportunities for competitive advantage by applying new Data-driven approaches to Data-driven Marketing practices. Large and small companies from diverse sectors can take advantage of the enormous amount of data generated to develop new business models, open new sources of revenue and initiate disruptive innovations.

This dissertation seeks to analyse the use of different sentiment analysis models for data disclosed on social networks, specifically on Twitter, in a set of disclosures classified as Nonprofit Marketing. This study seeks to innovate in the contribution to knowledge by studying for the first time the relationship between message predictors and user Engagement on Twitter by Nonprofit Organisations.

In this study, the quantitative method is used with the application of Machine Learning techniques for classifying tweets based on Positive, Negative and Neutral sentiments. The collected data were subordinated to statistical studies, namely Spearman's correlation.

This research in the Nonprofit Marketing sector demonstrates that it is possible to predict which sentiment expressed in the message will have the best Engagement, thus generating innovative communications for followers that will lead to increased interaction. It should be noted that Positive messages tend to affect user Engagement negatively.

Keywords: Data-driven Marketing; Nonprofit Marketing; Nonprofit Organisations; Machine Learning; Sentiment Analysis.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

RESUMO

A expansão tecnológica ocorrida nos últimos anos cria oportunidades de vantagem competitiva ao aplicar novas abordagens orientadas a dados às práticas de marketing. Empresas grandes e pequenas de diversos setores podem aproveitar a enorme quantidade de dados gerados, para desenvolver novos modelos de negócios, abrir novas fontes de receita e iniciar inovações disruptivas.

Esta dissertação procura analisar a utilização de diferentes modelos de análises de sentimentos para dados divulgados em redes sociais, mais especificamente no Twitter, em um conjunto de divulgações classificadas como *Nonprofit Marketing*. Este estudo busca inovar na contribuição ao conhecimento ao estudar pela primeira vez a relação entre o sentimento da mensagem e o *Engagement* do utilizador nos perfis do *Twitter* de Organizações sem Fins Lucrativos.

Neste estudo é utilizado o método quantitativo com aplicação de técnicas de *Machine Learning* para a classificação dos *tweets* com base em sentimentos Positivos, Negativos e Neutros. Os dados recolhidos foram subordinados a estudos estatísticos, nomeadamente correlação de *Spearman*.

Esta pesquisa no setor do *Nonprofit Marketing* demonstra que é possível prever qual o sentimento expresso na mensagem terá o melhor *Engagement*, podendo assim gerar comunicações inovadoras para os seus seguidores que levarão ao aumento da interação. Salienta-se que as mensagens Positivas tendem a ter um efeito negativo no *Engagement* do utilizador.

Palavras-chave: Data-driven Marketing; Nonprofit Marketing; Organizações sem Fins Lucrativos; Machine Learning; Análise de sentimentos.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

INDEX

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

TABLES INDEX

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

FIGURES INDEX

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

List of abbreviations and acronyms

AI: Artificial Intelligence

ANN: Artificial Neural Networks

ARFF: Attribute Relation File Format

BERT: Bidirectional Encoder Representation and Transformer

BoW: Bag of Words

CMS: Content Management System

CNN: Convolutional Neural Networks

COVID-19: Coronavirus Disease 2019

DCNN: Dynamic Convolutional Neural Network

eWOM: Electronic Word-of-Mouth

GloVe: Global Vector

KNN: K-Nearest-Neighbor

LSTM: Long Short-Term Memory

ML: Machine Learning

NLP: Natural Language Processing

NLTK: Natural Language Toolkit

NPO: Nonprofit Organizations

RNN: Recurrent Neural Networks

SMLP: Social Media Listening Platform

SVM: Support Vector Machine

UGC: User-Generated Content

URL: Uniform Resource Locator

VSM: Vector Space Model

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

INTRODUCTION

In this first chapter, the introduction to the theme "Data-driven Marketing", the object of this study, refers to the research context, research objectives, and the work structure proposed to follow.

1.1 Contextualization of the study problem

Technological expansion in recent years creates opportunities for competitive advantage by applying new data-driven approaches to marketing practices (Borges et al., 2021). Companies large and small across industries can leverage the massive amount of data generated to develop new business models, open up new revenue streams and initiate disruptive innovations (Ahlemeyer-Stubbe & Müller, 2020). Collecting data from different sources such as the Internet of Things (IoT), social media, and search engines has created significant opportunities for organizations to take an analytical view in developing programmatic marketing approaches (Jabbar et al., 2020). We live in a world of rapid technological advances in the digital realm and an abundance of data generated in which the consequences for marketing practices have been transformative (Shah & Murthi, 2020).

As the role of marketing has grown in importance over time, prominent thought leaders has emphasized the relevance aspect of marketing to a company's business practices (Shah & Murthi, 2020), and so we see more and more using data to produce relevant and timely marketing proposals (Micheaux & Bosio, 2019). Today, companies increasingly need to integrate multiple business functions to acquire, model, analyse and evaluate the information required to understand customer behaviour better and, from there, make data-driven decisions to improve the experience journey (Senvar et al., 2020).

Data-driven analysis, enabled by information systems, technologies, methodologies and practices, such as the use of Big Data, to extract relevant data and its transformation into business insights (Nuortimo & Harkonen, 2019) allow better understanding of consumer needs and competitors (Meng et al., 2018).

Technology-enabled marketing practices can deliver improved operational efficiency, better marketing insights, and more innovative ways to engage with its customers (Shah

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

& Murthi, 2020) . However, these actions depend on information about consumers to perform their actions, such as local variables, momentary geography and a series of similar data (Nadler & McGuigan, 2018).

Data-driven Marketing can be used in several application areas, such as in politics, conceptually dealing with the process of collecting and storing information about voters, volunteers and actual and potential donors in the production of datasets used to guide targeting efforts related to organizing, advertising, and fundraising efforts (Walker & Nowlin, 2018). In this way, this process of collecting complex data through online and offline channels, analysing it to understand individuals' psyche and activity patterns allows the marketing team to develop a highly personalized strategy to connect with the targeting, regardless of the activity area (Grandhi et al., 2020).

1.2 Research objectives

This dissertation is based on a literature review to investigate how adopting new technologies, and data analysis tools helped transform the scope of marketing in companies, specifically in the scope of Nonprofit Marketing. In this sense, the objective of this work is to study how we can use sentiment analysis to improve our understanding of the user and, to achieve this end, we evaluate this approach in a dataset of tweets classified as Nonprofit Marketing, which we pre-process these tweets as described in Go et al. (2009). To the best of our knowledge, no study has synthesised through a review of the literature on Natural Language Processing, the use of a text sentiment classifier using Twitter data for Nonprofit Marketing in the integration of a tokenizer and an embedding layer. BERT with Convolutional Neural Networks (CNN) is compared to a CNN integrated with GloVe.

In this way, this study seeks to analyse the use of different models of sentiment analysis for data released on social networks, namely on Twitter, in a set of disclosures from institutions classified as Nonprofit Marketing to assist the decision-making process of whether or not to use of promotions to increase/recover user engagement.

The project's overall objective is to assess which predictors are most important in a range of Twitter data from Nonprofit Organisations (NPO) to refine the classification proposed

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

by Okazaki et al. (2015) to capture more detailed communication patterns for engagement.

For this, the specific objectives of this work are to:

- Study a supervised learning technique to classify the sentiment of the Twitter audience according to the categories of engagement needs of NPOs,
- Extract posts related to the NPO from more than one source of information;
- Find strengths and points to improve in the communication strategy of the NPO, namely user engagement.

This study seeks to innovate in the contribution to knowledge by studying for the first time the relationship between message predictors and Twitter engagement in the context of Nonprofit Marketing. The research focuses on nonprofit organisations in the United States of America with Twitter accounts.

1.3 Project organization

This dissertation is constituted by five chapters organized as follows to achieve the objectives of the work. The first chapter shows an introduction to the topic, where a contextualization of it is presented, and the research objectives are defined.

The second chapter is dedicated to the literature review, divided into three major themes. The first is Data-driven Marketing and its antecedents, where a historical framework is made, presenting the concept and its importance. Then, variables referring to the fundamentals of Big Data used in this project will be addressed, more specifically its dimensions and respective implications for management and, finally, the topic of Nonprofit Marketing, which will focus on its dimensions and its relationship with o Data-driven Marketing.

The third chapter is reserved for developing the conceptual framework and, for this purpose, the proposed research model and the working hypotheses that support it are presented. This chapter also presents the sample and analyses the quality of the measuring instruments.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The fourth chapter analyses the results and discusses them. In the first step of this chapter, the hypothesis test results will be presented. The results will be discussed in a second phase, namely, the factors determining how users engage with the selected institutions.

Finally, in the fifth chapter, the main conclusions of the investigation and the implications for marketing are highlighted. Some limitations of this investigation and suggestions for future work are also presented.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

2 THEORETICAL FRAMEWORK

In this chapter, which aims to provide a theoretical framework for the investigated topic, the concept and importance of Data-driven Marketing will first be addressed. For this purpose, an approach is applied to the concept of data-driven and its evolution throughout history, to theoretical perspectives concerning marketing and description of the factors that precede the theme. The following is a review of the literature relevant to the big data concepts and tools used in this project. Finally, still in this chapter, a theoretical framework of the nonprofit marketing theme is carried out, highlighting the importance of its development and growth, organization of resources, and works related to the analysis of engagement in Nonprofit Marketing through social networks.

2.1 Data-driven Marketing

2.1.1 Overview

The history of Data-Driven Marketing begins in the 1950s, when there was a conscious effort to make marketing more scientific through quantitative techniques such as linear programming and time series analysis, in addition to Markov chains (Sheth & Kellstadt, 2021). Thus, the marketing decision-making process has gained increasingly significant importance with the development of analytics and the advancement of the digital age (Grandhi et al., 2020).

The promise of leveraging big data and analytics capabilities into a company's strategy has increased pressure on marketing departments to make data-driven analytics central to marketing decision making (Johnson et al., 2021). Thus, the brand can optimize marketing contact with its customers through a coherent and relevant dialogue (Abakouy et al., 2019).

Today marketers have access to an ever-increasing variety and amount of data from customers and other stakeholders (Sleep et al., 2019). Using appropriate data mining techniques, marketers can turn this data into valuable knowledge for their strategies (Bilgic et al., 2021).

The Internet, the World Wide Web, and Google search engines have made connectivity and communication virtually universal (Sheth & Kellstadt, 2021). It is a fact that data

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

drives marketing by helping marketers understand their customers, create effective campaigns and send them to relevant people (Abakouy et al., 2019).

In this way, it is understood that Data-Driven Marketing consists of structuring and analysing data to clarify consumer behaviour and expectations (Abakouy et al., 2019), and its success depends on how well an organization adopts the practice, in which the most crucial indicator of an organization's commitment is the extent of the resources invested (Grandhi et al., 2020).

2.1.2 Data analysis and decision making

Regarding the use of data in decision-making, literacy has produced solutions using social connections to improve business results (Senvar et al., 2020). Through quantitative (Ozgormus & Smith, 2020) and qualitative (J. Li et al., 2018) analyses, performance increases concerning traditional targeted marketing based on sociodemographic data, demographics and prior purchases.

Different companies are implementing measures to improve the personalization of the consumer experience, using marketing tools such as email delivery, ad delivery, content management systems (CMSs), website access analysis and social media (Kawada et al., 2019). However, data-based approaches can help companies to know consumer preferences, and, in this case, a prevalent example is the use of consumer preference analysis that aims to explain which aspects of a particular product affect the decision of consumer purchase (M. Guo et al., 2020) and thus offer increasingly personalized approaches.

An example of this type of approach can be seen in basket/shopping cart data (items purchased per customer visit, market basket analysis), where relationships are established between departments and opportunities to increase impulse purchases are identified. The main objective of this strategy is to find out which products are sold together to develop an adjacency matrix based on the actual sales characteristics of a given store (Ozgormus & Smith, 2020). The use of analytical tools such as this one serves as the primary source of information for marketing analysis. Marketing managers base their strategic decisions since the information helps in decision-making and the development of marketing strategies, focusing on company success indicators (Miklosik et al., 2019).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The literacy further demonstrates how advertisement inventory allocation decisions significantly influence revenue management and significantly affect promotion performance, the balance between supply and demand, and the structure of the programmatic advertising market. Using a data-driven approach can significantly improve maximum revenues compared to traditional strategies (H. Li et al., 2017).

2.2 Big data concepts and tools

Due to their broad reach and ease of use, social networks are increasingly setting trends in topics ranging from entertainment, marketing and technology to politics and public health (Pota et al., 2018). The consecutive increase in the volume of text datasets generated by these means and the countless research directions in the overall ranking have created an excellent opportunity for a comprehensive assessment of current achievements in research on this topic (Xuan Huynh et al., 2019).

The latest theories and practices for marketing have unprecedented investigative opportunities gained from the rapid growth of e-commerce and e-marketplace activities that collect massive data on consumers' online behaviour (Phang et al., 2020). With the rise of artificial intelligence (AI) and accompanying subfields such as machine learning (ML), data mining, and big data, many human tasks can be successfully performed by software (Dalenberg, 2018). Particularly in this work, we will study machine learning (ML), perfectly adapted for deep data-driven marketing analysis based on algorithms that allow specialized AI software to identify patterns in big data and classify them (Capatina et al., 2020).

At the moment, Natural Language Processing (NLP) models such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) have achieved good results for this task, but multiclass text classification and refined sentiment analysis are yet to be performed are challenging (Dong et al., 2020). Primarily utilizing real-time data and continuously improving previous results (Batbaatar et al., 2019) have been widely applied to text sentiment classification and sentiments in commodity analysis due to automatic data extraction and favourable experimental results (Dong et al., 2020).

To understand this project, we first need to understand the topic of Sentiment Analysis. Thus, Hoang et al. presented that sentiment analysis, also known as opinion mining, is a

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

field of NLP that automatically identifies a text's sentiment, often in categories such as negative, neutral and positive. It has become a trendy field in both research and industry due to the large and growing amount of user-generated opinion texts on the Internet, for example, social networks such as Twitter which will be analysed in this project. According to the authors, knowing how users feel or think about a particular brand, product, idea, or topic is a valuable source of information for companies, organisations, and researchers, but it can be a challenging task. Natural language often contains ambiguity and figurative expressions that make automated extraction of information often very complex (Hoang et al., 2019).

According to Dong et al. in 2020, there are two essential methods in sentiment analysis by Natural Language Processing. The first method is based on Machine Learning and imports related features established after data processing into the model. While the performance of this method depends a lot on feature engineering, the second method based on Deep Learning is mainly represented by Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), have been widely applied due to automatic data extraction and favourable experimental results (Dong et al., 2020). In any natural language processing method, the representation of the text (including the destination and its context) is a crucial issue (Gao et al., 2019).

For Devlin et al., machine learning techniques constrain pre-trained representations' power, especially fine-tuning approaches. The main limitation is that standard language models are unidirectional, limiting the choice of architectures that can be used during pre-training (Devlin et al., 2019). According to the authors, these restrictions are suboptimal for sentence-level tasks and can be very harmful when applying fine-tuning-based approaches to token-level tasks, such as answering questions, where it is crucial to incorporate context from both directions.

Li et al. showed that the BERT-based architecture could outperform state-of-the-art models even with a superficial linear classification layer. In addition, they also standardized the comparative study by consistently using a validation development dataset for model selection, which is largely ignored by previous work (Devlin et al., 2019). The authors further explored how to couple the BERT embedding component to various neural models and perform extensive experiments on two reference datasets.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Experimental results demonstrate the superiority of BERT-based models in capturing aspect-based sentiments and their robustness to overfitting.

Given the above, we understand that models with Machine Learning, in general, are successful in well-behaved environments with known limitations, as they tend not to have enough flexibility to perform well outside the context where their initial development took place. However, Deep Learning Models are mostly large neural networks formed by sets (ensemble) of networks with specific structural and functional particularities and thus extend the number and role of the various hidden layers between the input and output layers, achieving a more efficiently model complex nonlinear relationship. The intended results for this analysis are made more appropriate by combining direct and recurrent architectures.

2.2.1 Deep learning

As previously presented, sentiment analysis evaluates the dataset from the sentiment aspect by measuring the frequency of words analysed and represented by a sentiment vector. According to Puteh et al (2021). Deep Learning as an application of artificial neural networks (ANN) to learning tasks using multiple layers of neural networks, was initially proposed by G.E. Hinton in 2006 and consisted of a Machine Learning approach then presented as Deep Neural Network (Hinton et al., 2006; Puteh, 2021).

A neural network can be as basic or as complex as needed. Deep "learning" occurs when we create neural network models where multiple inputs with their respective weights pass through multiple hidden layers, supervised or unsupervised phases, feedforward and backpropagation (Goldberg 2016). Over the past ten years, Deep Learning has been revolutionary and has produced cutting-edge results in many application domains, starting with computer vision, then speech recognition, and more recently, natural language processing (NLP) (Collobert et al., 2011; Goldberg, 2016; Zhang et al., 2018).

2.2.2 Natural language processing (NLP)

One of the most exciting applications for sentiment analysis is the so-called natural language processing, which can be defined as the process of recognizing patterns from

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

texts so that we can train a neural network to mine information from themselves ((Gao et al., 2019).

Traditionally, the bag of words (BoW), as the example shown in Figure 1, is used to generate text representations in NLP and text mining, where a document is considered a bag of words (Zhang et al., 2018). This word bag technique is commonly applied to extract features from news articles, measure the occurrence of each word, and then convert text information into vector spaces using those occurrences (Lien Minh et al., 2018).

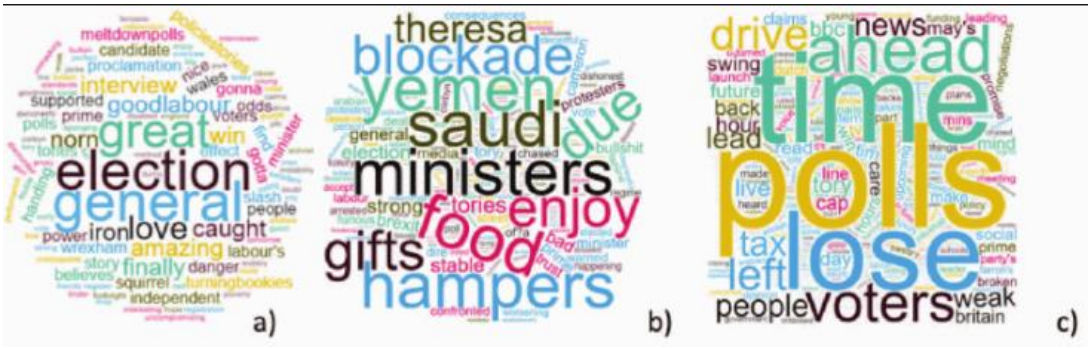


Figure 1. Example bag of words (BoW) from May 22, 2017 based on sentiment extracted from tweets when the UK general election campaign follows different themes, and there is the Manchester terrorist attack: a) positive, b) negative, c) neutral. (Pota et al., 2018).

Many deep learning methods in NLP use a word embedding learning algorithm as a standard approach to feature vector representation, which ignores the sentiment polarity of words (Eke et al., 2021). The model in this work takes advantage of natural language processing (NLP) and deep neural networks and comprises two main stages. The first stage involves the sentiment polarity classifier that classifies tweets as positive and negative. The output of the first stage is then used as input to an emotion classifier that aims to assign a tweet to one of the positive emotion classes (joy and surprise) or one of the negative emotion classes (sadness, disgust, fear, anger) (Imran et al., 2020).

2.2.3 Convolutional neural networks (CNN)

The so-called CNN networks have architectures capable of showing spatial relationships and structures without using many parameters. They are a particular type of feedforward neural network, employed initially in the field of computer vision (Zhang et al., 2018). According to Gao et al. (2019), the first to use CNN to classify feelings at the aspect level

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

were the researchers Huang et al. in 2018 that incorporated aspect information into CNN via parameterized filters and parameterized gates, resulting in good performance metrics in the SemEval-2014 datasets (Gao et al., 2019; Huang et al., 2018).

However, other authors already dealt with CNN, such as Kalchbrenner et al. in 2014, they proposed a convolutional architecture called Dynamic Convolutional Neural Network (DCNN) that they adopted for semantic sentence modelling (Kalchbrenner et al., 2014). Alternatively, Kim, in 2014, reported a series of experiments with convolutional neural networks (CNN) trained on pre-trained word vectors for sentence-level classification tasks and showed that a simple CNN with little hyperparameter tuning and static vectors achieves excellent results. in various benchmarks (Kim, 2014).

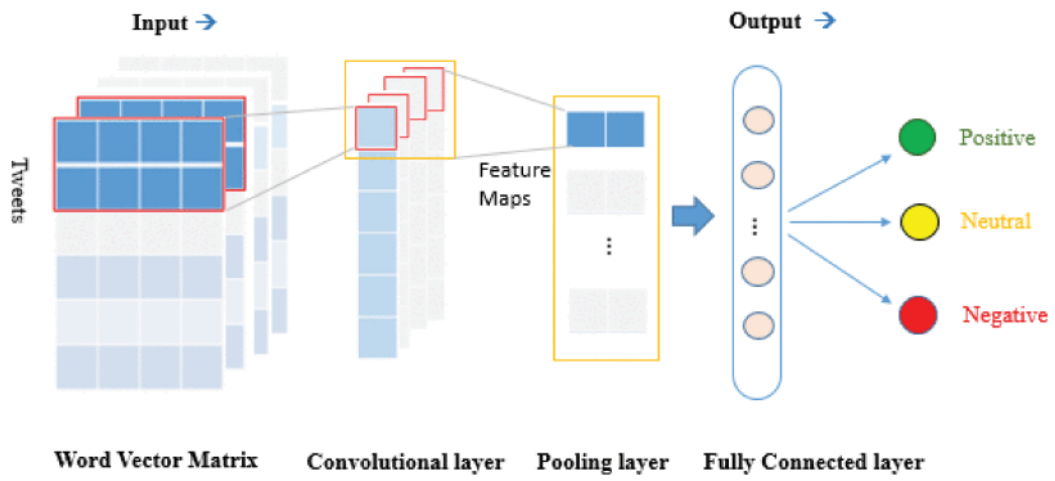


Figure 2. A generic CNN based architecture for text classification (S. Tam et al., 2021).

Figure 2 presents a generic CNN architecture based on text classification (S. Tam et al., 2021), and it is possible to observe that the convolutional layers in CNN play the role of feature extractor, which extracts local features, as they constrain the receptive fields of hidden layers to be local (Zhang et al., 2018). It is observed that current research has inspired the design of several other convolutional networks guided by: hierarchies, chaining of different types of architectures and even more complex topological designs.

2.2.4 Bidirectional encoder representation and transformer (BERT)

Pre-trained language models provide context for words, which have previously learned the occurrence and representations of words from unannotated training data (Hoang et al.,

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

2019). Most previous approaches model the relationship between target words and their context with Long Short-Term Memory (LSTM) and attention. However, LSTMs are challenging to parallelize and truncated backpropagation over time makes it challenging to remember long-term patterns (Gao et al., 2019).

Bidirectional encoder representations from transformers (BERT) originally published by Devlin et al. and it is a pre-trained language model that is designed to consider the context of a word from the left and right sides simultaneously (Devlin et al., 2019) provides dense vector representations for natural language using a deep neural network and pre-trained with transformer architecture (X. Li et al., 2019). As the input representation of BERT can represent a single sentence of text and a pair of sentences of text, we can convert sentiment analysis tasks into a sentence pair classification task and adjust the pre-trained BERT (Sun et al., 2019).

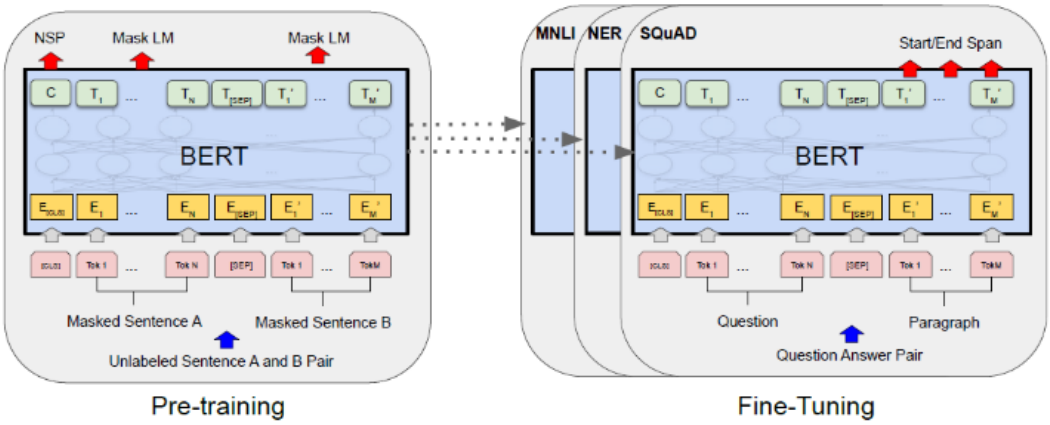


Figure 3. General pre-training and fine-tuning procedures for BERT (Devlin et al., 2019)

In summary, as shown in Figure 3, the BERT architecture combines the two principles (Transformer and bidirectional), as the training method is based on language masking. Uses data: BooksCorpus (800M words) and Wikipedia in English (2,5 M words). For its processing, texts of the "document" type are needed to have continuous sequences, as we will use in this project.

2.2.5 Approach based on CNN integrated with GloVe - CNN

The model used in the approach based on Deep Learning has as its conceptual base the works of Eke et al. (2021) and Tam et al. (2021), who presented CNN architectures consisting of a Deep Learning model with long short-term memory (LSTM) employed

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

with the incorporation of GloVe to build context vectors that represent a semantic word as resources. This model was recreated and applied to document-level sentiment classification in our work. In the case of tweets, it will be compared with the following cases in terms of accuracy.

As suggested by the authors Eke et al. (2021) and Tam et al. (2021), in this first step in the pre-processing, each text data (tweet) is divided into smaller parts in the tokenization step, in words or phrases and these tasks are performed using the NLTK library (Eke et al., 2021; S. Tam et al., 2021).

Then, all items that do not contribute to the classification task are eliminated before the resource extraction process and, therefore, the elimination of unwanted information is carried out since most social network data presents some noise and, as a result, it requires a pre-processing step to eliminating these unwanted items. In the dataset used in this work, tweets contain punctuation, stop words, special characters (such as @, #, for example) and URL links. The processed data was transformed into a matrix representation of the features to simplify model training.

In the next step, the authors suggest that the textual word, generally considered discrete and categorical features, be represented in vector format. The vector representation can be performed by converting the text to the vector space model (VSM) (Eke et al., 2021). However, we employed pre-trained GloVe in this study because of its excellent performance with CNN compared to another pre-trained embedding based on analysis of related studies by the authors (Pennington et al., 2014). The pre-trained Glove file was extracted from the website < https://github.com/stanfordnlp/GloVe> on May 03, 2021.

The convolutional neural network architecture consists of input layers, convolutional layers, pool layers, and output layers (Eke et al., 2021). Input data is fed by the input layer and then passed to the convolutional layer. Feature maps are extracted in the convolutional layer, bypassing the convolutional filter on the input data. However, various filters are used to enter data to extract various features.

The final decision is made by the fully connected layer, connected to the output and previous layers. A *recurrent neural network* is a standard network that uses an edge to feed the slides next time and feed the next layer on a slide of similar timing (S. Tam et

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

al., 2021). It contains a cycle that means there is short memory on the network. On the other hand, the recurrent neural network operates similarly to a hierarchical network that does not require allocating time slides for the input sequence but processes the input in a hierarchical tree structure (Eke et al., 2021).

The model uses the dense layer to connect each input to each output using weights. *Sigmoid* is a function used in the final layer to produce the output (S. Tam et al., 2021). It takes the average of the random results in the forms 1 and 0. The result of the prediction of the sigmoid function is presented in Equation (1). The sentiment result is ranked 0 or 1 using binary cross-entropy. In this study, 0 represents a negative feeling, and 1 represents a positive feeling (S. Tam et al., 2021).

$$Y=sigmoid(wh+b) \quad (1)$$

The Deep Learning Model with Long Short-Term Memory (LSTM) used with the incorporation of GloVe to build context vectors that represent a semantic word as resources is shown in Figure 4. In our work, this model was recreated and will be applied for sentiment classification at the document level as per Annex 2.

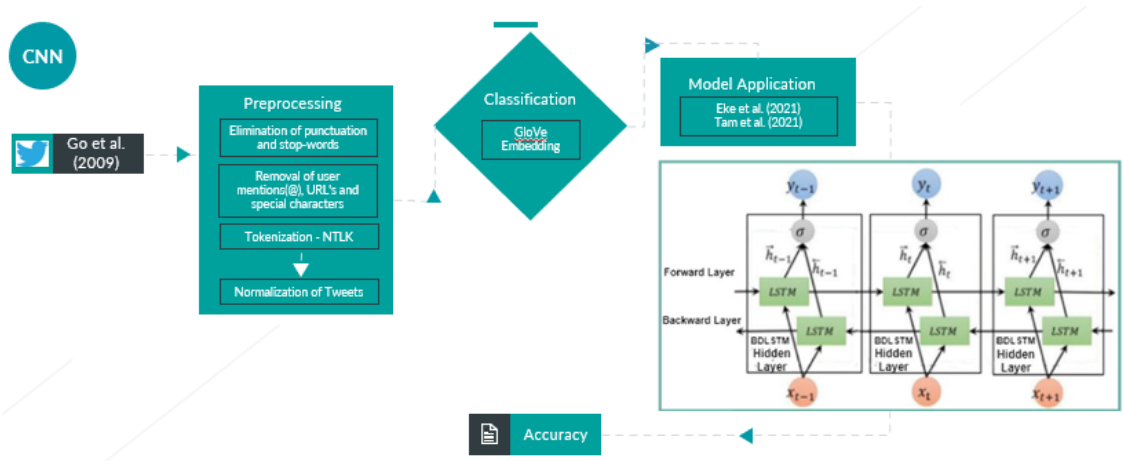


Figure 4. Deep Learning Model with Long Short-Term Memory (LSTM) used with the incorporation of GloVe

2.2.6 BERT-integrated CNN-based approach as Tokenizer - BERT

The model used in this approach has a conceptual baseline of related works the articles by Devlin et al. (2019) and Li et al. (2019). In our model, we use the BERT Large version as Encoders, which has the following configurations: The number of blocks (layers) of

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

the Transformer is L=24, the hidden size is H=1024, the number of self-attention heads is A=16, and the total number of parameters for the pre-trained model is 340M.

A similar pre-processing technique used for the CNN-based approach integrated with GloVe was employed in this section. BERT was used to perform tokenization on processed tweets. Tokenization is applied in splitting strings into subword token strings seeking to convert token strings to ids and vice versa, and encoding/decoding (i.e. tokenization and conversion to integers) (Devlin et al., 2019).

BERT accepts a maximum of 512 token strings as input and represents the token strings of a 768-dimensional array as output (Devlin et al., 2019). According to the authors, in BERT, a maximum of two segments can be inserted into each input string, including [SEP] and [CLS]. As seen in Figure 5, special classification token embedding ([CLS]) is usually the initial input string token. It contains a special classification embedding chosen as the first token in the last hidden layer to represent the complete sequence in a text classification task. However, the last hidden state that correlates with this token is the aggregate string to represent text classification tasks.

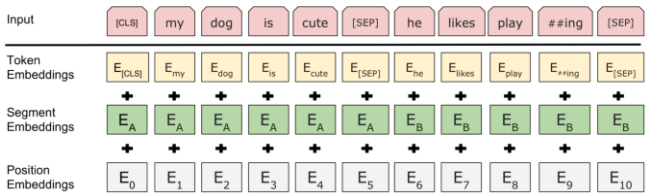


Figure 5. Transformation of inputs (inputs) with BERT. Source Devlin et al. (2019)]

The pairs of a sentence are grouped in a sequence. These phrases can be differentiated into two methods (Devlin et al., 2019). According to the authors, first, a special token embedding ([SEP]) separates the sentences. Second, a learning embedding is added to each token, indicating the phrase to which each phrase belongs (positive or negative) (X. Li et al., 2019). An overview of the classification can be exemplified by Equation (2) below:

$$[CLS]+Sentence\ A+[SEP]+Sentence\ B \qquad (2)$$

The BERT model employed to perform tokenization on processed tweets is shown in Figure 6. In this, tokenization is applied to divide strings into subword token strings

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

seeking to convert token strings into ids and vice versa, and encoding/decoding (i.e., tokenization and conversion to integers). The application of this model is presented in Annex 3.

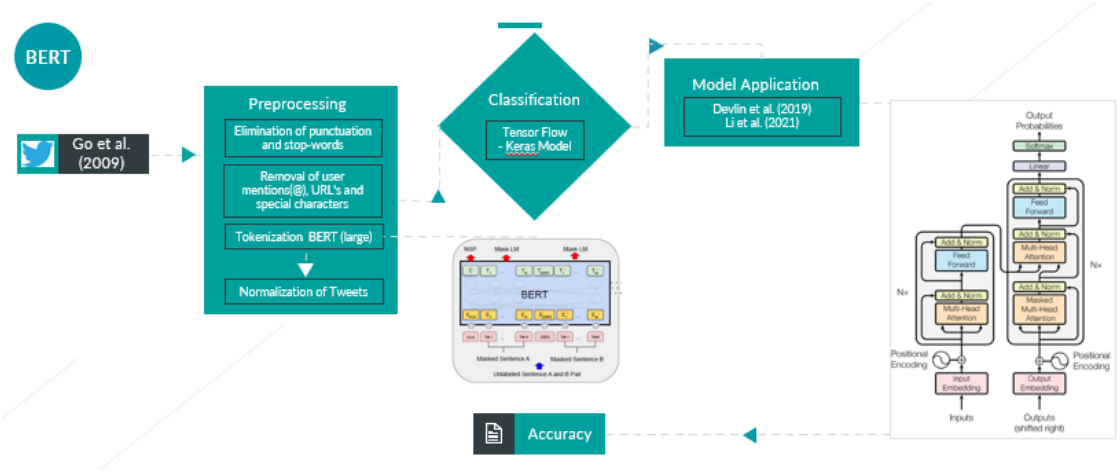


Figure 6. BERT model employed to perform tokenization on processed tweets.

2.2.7 BERT-based approach as an Embedding Layer – BERT-CNN

This approach follows the same principle presented in section 2.2.6 with the addition of BERT as an Embedding Layer. For this case, we used the articles by Dong et al. (2020) and Sun et al. (2019).

As seen in the previous section, the BERT input representation can explicitly present a pair of text sentences in a sequence of tokens. Its input representation is constructed for a given token by summing the corresponding token, segment, and position embeddings.

Compared to the embedding layer presented in sections 3.3 and 3.4, which provide only a single context-independent representation for each token, the BERT embedding layer takes the phrase as input. It calculates the token-level representations using the information of the entire sentence (Dong et al., 2020).

In this representation, as applied with Sun et al. (2019), to obtain a fixed-dimension clustered representation of the input sequence, we use the final hidden state (i.e. the transformer output) of the first token the input. We denote the vector as $C \in R^H$ (Sun et al., 2019). Finally, the probability of each category (P) is calculated by the softmax function presented in Equation (3) below:

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

$$P=softmax(CW^T) \quad (3)$$

Where T is a target entity set and W is a classification layer whose parameter matrix i $W \in R^{K.H}$ (Sun et al., 2019), whose K is the category number.

As seen in the previous sections, the BERT input representation presents a pair of text sentences in a sequence of tokens. In the model shown in Figure 7, the embedding layer BERT takes the phrase as input and calculates token-level representations using the information from the entire phrase. This model is presented in Annex 4.

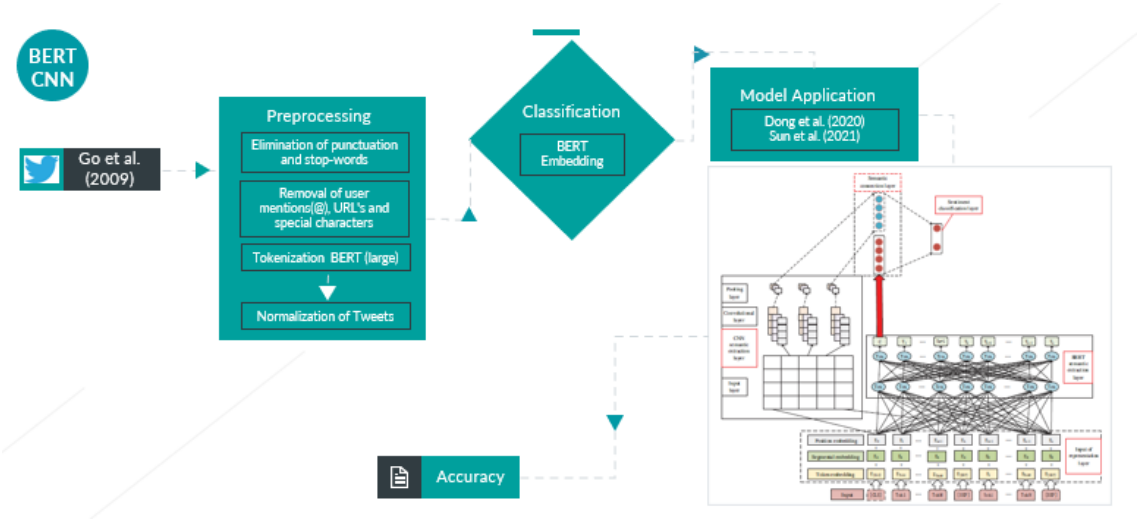


Figure 7. Model with embedding layer BERT.

2.3 Nonprofit Marketing

We review the literature that explores the level of engagement through audience participation in interaction with social media from a Nonprofit Marketing perspective. Considering the relevance of social media to our study, we also included studies on sentiment analysis using social media. As a detailed overview of the Data-Driven Marketing literature would be too extensive, we focused our literature review on studies of Nonprofit Marketing.

2.3.1 The growth and development of nonprofit marketing

Nonprofit organizations (NPOs) are characterized by a solid devotion to those in need, i.e. their end-users (Schetgen et al., 2021), and these organizations have faced fierce competition for charitable donations due to the rapid growth in the number of NPOs and a decrease in government support (Hsu et al., 2021).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Shapiro, in 1974, released an innovative article on nonprofit marketing stating that "the study of marketing in nonprofit organizations is a new field" (Shapiro, 1974). From this moment on, nonprofit marketing grew and evolved strategically and was no longer considered a new field of study (Y.-J. Lee, 2019; McLeish, 2010).

Nonprofit marketing became one of the central topics with the rapid growth of charitable activities (Ha et al., 2022) until, in September 2015, the United Nations General Assembly adopted the Global Agenda 2030 recognizing the business sector as a fundamental agent for achieving the so-called Sustainable Development Goals (SDGs) (Barroso-Méndez et al., 2020).

Different from the objective of traditional marketing, which is to expand the market, the objective of social marketing is to promote the well-being of society (Liao, 2020), referring to the support of a specific cause linked to a nonprofit organization (NPO) or, even, to a charitable donation conditional on the purchase of a product by the consumer (Thomas & Jadeja, 2021).

Nonprofits have gradually adopted the brand-oriented approach to adding value to their stakeholders (Ha et al., 2022), comprised of three different markets of customers, volunteers and donors or funders (Gainer & Padanyi, 2005). Indeed, research suggests that marketing techniques are well-accepted practices in today's nonprofit organizations, and scholars have studied nonprofit marketing strategies and management from various perspectives (Andreasen & Kotler, 2003; Y.-J. Lee, 2019; Pope et al., 2009).

Nonprofit marketing is a process of designing, executing, and controlling programs using a combination of marketing research and marketing mix to enable target audiences to accept certain societal ideals, concepts, and measures (Liao, 2020). Currently, the nonprofit marketing field has a variety of textbooks on the topic and dedicated journals (Y.-J. Lee, 2019). The topic is offered as a significant part of business management and nonprofit curricula at colleges and universities or as a standalone course (Mottner & Wymer, 2011).

2.3.2 Resource development and organization

Nonprofit marketing encourages social behaviour change, and this social behaviour change refers to the behaviour of a certain number of target audiences, not just individual

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

behaviour (Liao, 2020). In recent years, the relationship between nonprofits and donors has been dramatically transformed by the emergence of social media (Schetgen et al., 2021).

Several studies have investigated the relationships between giving behaviours and brand-related factors, for example, brand image (Bennett & Gabriel, 2003; Michel & Rieunier, 2012) and brand personality of nonprofit organizations (Sargeant et al., 2008). According to Meijer (2009), the reputation of charitable organizations is the central construct of donation intention (Ha et al., 2022; Meijer, 2009).

Most nonprofit organizations generate most of their revenue from one or more of three sources: service program revenue, private charitable contributions, and government grants (Y.-J. Lee, 2019). In addition to this traditional donation channel, online donations are a new channel due to the popularity of the Internet, as reaching a more significant share of monetary donation is the main objective of NPOs (Hsu et al., 2021).

Likewise, the impression and reputation of nonprofit institutions had positive effects on donor behaviour (Sargeant et al., 2008). Furthermore, the intention to repeat charitable donations is influenced by fundraising experience, the trustworthiness of the nonprofit agency, and sympathy for volunteer activities (Beldad et al., 2014).

Identifying valuable donors and developing effective marketing strategies can contribute to online donation platforms (Hsu et al., 2021). Four characteristics (individualism, collectivism, hedonism, and utilitarianism) are affected by trust in NPO donation intentions and serve to understand the psychological antecedents of trust in nonprofit marketing campaigns (Thomas & Jadeja, 2021).

2.3.3 Engagement in Nonprofit Marketing through social media

The works related to this dissertation are categorized in Table 1 according to three dimensions: type of research, NPO engagement and social networks. First, two types of research can be distinguished according to Schetgen et al. (2021). While diagnostic research investigates the potential determinants of the behaviour of interest and the underlying reasons, predictive research aims to determine how current information can be used to predict future behaviour (Schetgen et al., 2021).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Second, two surveys related to audience engagement on social media are used as a reference and baseline for the appearance of categorized comments, Okazaki et al. (2015), and Tripathi and Verma (2018). The feedback aspect was then grouped into three NPO engagement needs: promotional, emotional, and informational.

Finally, research on giving behaviour can be categorized according to data type. Social media data includes self-reported information about social media usage, as well as information retrieved directly from social media sites as, for example, can be seen in the work of Enjolras et al. (2013), as well as Farrow and Yuan (2011), who investigated the underlying reasons for the positive impact of social media use on giving time and money.

Table 1. Related works

Authors	**Research Type**	**Engagement**	**Social Media**
H. Farrow, Y.C. Yuan	Diagnostic	Promotional	Facebook
B. Enjolras, K. Steen-Johnsen, D. Wollebæk	Diagnostic	Emotional	Facebook and Twitter
G. D. Saxton, R.D. Waters	Diagnostic	Informational	Facebook
S. Brown, K. Taylor	Diagnostic	Emotional	Blog
A.M. Warren, A. Sulaiman, N.I. Jaafar	Diagnostic	Informational	Facebook
S. Okazaki , A. M. Diaz-Martin, M. Rozano	Diagnostic	Emotional	Twitter
R. Alghrabat et al.	Predictive	Emotional	Facebook
A. Reyes-Menendez, J.R. Saura, C. Alvarez-Alonso	Diagnostic	Informational	Twitter
S. Tripathi, S. Verma	Diagnostic	Emotional	LinkedIn, Facebook, Twitter, Youtube and Blogs

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Authors	Research Type	Engagement	Social Media
S.R. Jordan et al.	Predictive	Promotional	Facebook

2.4 Research gap

The use of social media to reaffirm or transcend socioeconomic divisions in participating in demonstrations was studied by Enjolras et al. (2015), investigating the underlying reasons for the positive impact of social media use on giving time and money. From this study, Schetgen et al. (2021) investigated a decision support system for building an acquisition model based on Facebook data with high predictive performance. Similarly, Ihm (2016) measured the scale and impact of social action using data as a call to action on social media platforms to ensure your data is available to researchers and activists who share your goals for a better society.

Farrow and Yuan (2011) carried out an analysis of structural equation modelling in American universities and non-profit institutions, revealing a change in attitude by showing that the strength of network ties can help ensure consistency between attitude and behaviour. From this study, Febriani and Selamet (2020) showed a significant difference in volunteering intention between participants exposed to NPOs with certain brand personalities and NPOs without any brand personalities. In contrast, Gray et al. (2021) found strong links between e-newsletter receipt and volunteering or donation activity, the usefulness of e-newsletters in building a connection between the organization and recipient was confirmed at a moderate level.

With a strong focus on analysing whether and how nonprofit advocacy organizations use social media, Saxton and Waters (2014) barely touched on the effectiveness of this social media use. More recent research such as Guo and Saxton (2018) has addressed this issue by building and testing a model of the effectiveness of Twitter usage by NPO advocacy organizations. Smith (2018) investigated the use and impact of Facebook and Twitter and revealed that users respond differently to stimuli across platforms.

Brown and Taylor (2015) used self-reported information to determine whether there is a relationship between social media presence and giving behaviour in NPOs. The authors

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

performed a regression analysis to determine the effects of specific individual attributes on charitable behaviour. Based on this article, Capra et al. (2021) investigate the effects of Big Five personality traits on one's willingness to make and keep a promise to volunteer.

Warren et al. (2015) applies social capital theory to build a model to investigate the factors influencing online civic engagement behaviour on Facebook. Analysis of survey data on youth social media use and civic engagement in the US shows that while all three types of online social ties (Facebook Friends and Twitter Followers) are positively associated with civic engagement, there are differences between the different types of connections and the benefits (Y. Lee, 2022) and challenges associated with the use of online communities by different actors in the context of Social Enterprises (Abedin et al., 2021).

In non-transactional customer engagement, such as NPO, Electronic Word of Mouth (eWOM) delivers value creation and empowers to generate new ideas (Fiarni et al., 2020). According to Okazaki et al. (2015), there are three forms of eWOM through social media: objective statement, subjective statement, and knowledge sharing. Discussion and conclusions on this topic have helped further research to evaluate this approach to social media strategy in the NPO (Fiarni et al., 2020) or religious context (Sircar & Rowley, 2016).

While several marketing implications have been considered in social media marketing, less attention has been paid to the influence of consumer brand engagement antecedents (telepresence, social presence and engagement) and its consequences for nonprofit organizations (Algharabat et al., 2018). Algharabat et al. (2018) showed that engagement positively affects consumer engagement with the brand in the context of nonprofit organizations. However, the combination of donor-related factors, system-related factors, and brand-related factors can better understand the factors that affect behavioural intentions (Maleki & Hosseini, 2020). Therefore, there is a need to explore social media tools and technologies for corporate communication and stakeholder engagement (Albanna et al., 2022).

Reyes-Menendez and Saura (2018) performed a sentiment analysis with an algorithm developed in Python and trained with data mining was applied to the sample of tweets to

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

group them according to the sentiments expressed to identify social, economic, environmental factors and cultural issues related to sustainable care for the environment and public health. Research on social media-based sentiment analysis is increasing, and we can explore helpful information to help not only policymakers design their policies but also scholars in developing methodologies and frameworks for analysing social media data (Marzouki et al., 2021) or developing a web-based interface for practical exposure and insights into the COVID-19 crisis (Andhale et al., 2021).

Tripathi et al. (2018), in their research, explores key determinants of NPO engagement and relationship building in social media in India. His research summarized the interest of the NPO audience in fourteen classes, which were grouped based on the relationship, whether it is a personal, organizational or mediating factor (Tripathi & Verma, 2018). Opinion mining is crucial for collecting and ranking opinions based on intended categories and calculating the proportion and impact of the organization's engagement with its potential audience (Fiarni et al., 2020).

In this scenario, nonprofit organizations work to organize their limited and variable resources in marketing with a high potential for returns (Jordan et al., 2019). This strategic relevance is further accentuated by the role of user-generated content (UGC), that is, what is being discussed on these platforms by social media users (Bernardi & Alhamdan, 2022). Thus, conflicting evidence and critical gaps exist in theorizing why affective displays are more effective, leading to practical difficulties for organizations dependent on their ability to generate consumer philanthropy (Mesler & Simpson, 2021).

Given the above, to fill this gap in the literature, we created a decision support system to build an analysis model based on Twitter data. The objective of this model is to analyse the engagement of followers of a nonprofit organization's profile on Twitter based on the type of content and sentiment of the tweet. In the first step of this process, we compared sentiment analysis techniques across three classification algorithms to find the optimal combination of both methods. Second, we use this optimal combination to evaluate different subsets of variables. This step allows us to establish the (added) value of the different types of Twitter features. In addition, we compute variables to find out which characteristics are most important in engagement.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

3 RESEARCH CONCEPTUAL FRAMEWORK

This chapter aims to present the conceptual framework of the research that it proposes to study, presenting the research problems and hypotheses determined by the analysis of the theoretical framework carried out in the previous chapter.

3.1 Research hypotheses

Twitter is a social network widely used in research to obtain information and understand the public opinion exemplified by its users (Reyes-Menendez et al., 2018) since social media is used not only as a tool personal but also for organizations and companies to share content (Jordan et al., 2019). This work intends to carry out an empirical study in nonprofit organizations relating the audience reaction and the characteristics of the posts on Twitter of these organizations with the user engagement, answering the following questions:

- What effect does posting characteristics and, in particular, sentiment factors in the published message have on engagement?
- How do audience reactions moderate the relationship between post characteristics and nonprofit engagement?

As past studies indicate that audience reactions influence engagement (Okazaki et al., 2015). Trust, emotions, and information needs emerge as the primary personal drivers for engagement (Tripathi & Verma, 2018).

The architecture of the research model shown in Figure 8, in which, in a first phase, data collection and treatment are carried out according to Go et al. (2009). In the experimental phase of the study, the proposed strategy ensures that a systematic treatment of the

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

evaluation is carried out according to the type and structure of the data extracted from the sentiment analysis algorithms. Finally, the results are applied to the analysis model.

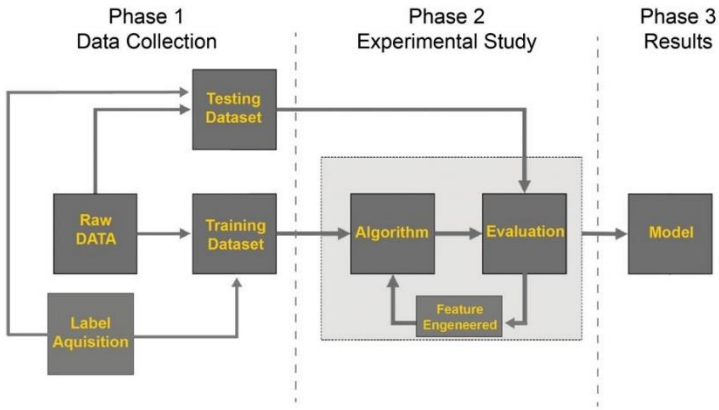


Figure 8. Research conceptual framework

3.1.1 Independent Variables

To hypothesize the messages, we analysed and mapped the engagement needs of NPOs to fit the dataset. In literacy, there are two surveys related to public engagement in social media, Okazaki et al. (2015), and Tripathi and Verma (2018), which were used as a reference and baseline for the appearance of categorized comments from NPOs and are presented in Table 2.

Table 2. Comment categorization, according to Okazaki et al. (2015) and Tripathi and Verma (2018)

Independent Variables	NPO variables	Positive messages
		Neutral messages
		Negative messages
Control Variable		Engagement

Given the above, the hypotheses to be considered in this work for the independent variables referring to Nonprofit Organisations are:

H1. Neutral messages are positively related to user engagement

H2. Positive messages are negatively related to user engagement.

H3. Negative messages are positively related to user engagement.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

3.2 Data set collection

Concerning data collection, we analysed arcolab.org, a website of a university centre that offers research services whose list is presented as a result of a study carried out by the Social Economy Unit of ARCO, in April 2020, on the world's 100 largest philanthropic foundations by total assets <https://www.arcolab.org/en/worlds-100-largest-philanthropic-foundations-list/#>. From this list, information was collected on the five largest US philanthropic foundations that have Twitter accounts, namely the Bill & Melinda Gates Foundation, Howard Hughes Medical Institute, Ford Foundation, Robert Wood Johnson Foundation, and J. Paul Getty Trust.

In this approach, raw data collection is done through Brandwatch (formerly Crimson Hexagon) with the query presented in Annex 1. This is a popular social media listening platform (SMLP) tool for academics and marketers (Breese, 2016). The subscription-based platform provides a social media library of over 1 trillion real-time and historical posts in various social media contexts (Hayes et al., 2021). A dashboard provides surface-level data views of descriptive query results such as date range and platform(s) of interest and AI-generated variables, including emotion analysis, post ratings, and sentiment analysis.

Data were collected between January and December 2021 and represent 6,183 mentions of the predefined NPOs. These tweets represent a 16% increase over the same period in 2020 and reached just over 271 million impressions with the audience. The evolution of these mentions can be seen in Figure 9.

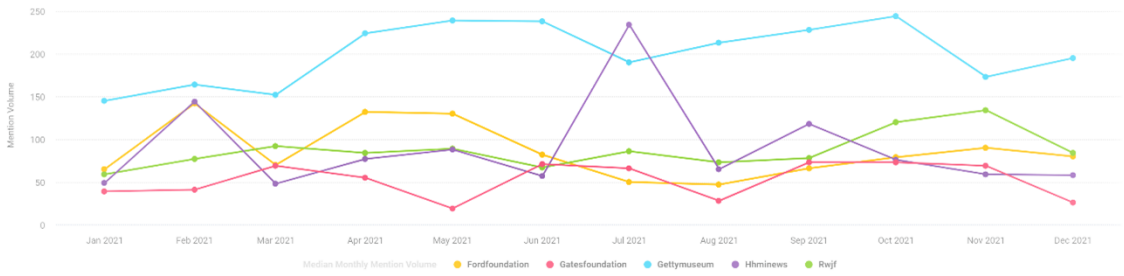


Figure 9. Volume over time

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Regarding trending topics, the most discussed topics involve COVID 19 and related topics. Mentions of topics such as justice and communities are also observed in a bag of words (BoW) with the main ones, as shown in Figure 10.

Figure 10. Trending topics

3.2.1 Online sentiment

According to the platform documentation, Brandwatch uses a NLP algorithm dubbed Sentiment Analysis to provide sentiment polarity analysis for posts (Brandwatch Help Center, 2022). The platform uses a training dataset of over 500,000 human-tagged posts to calculate word frequency distributions that include denied words and emoticons for posts in predefined categories such as: positive, neutral, and negative (Hayes et al., 2021). The frequency distributions are then used to build the algorithm that ranks posts by sentiment and are ranked as shown in Figure 11.

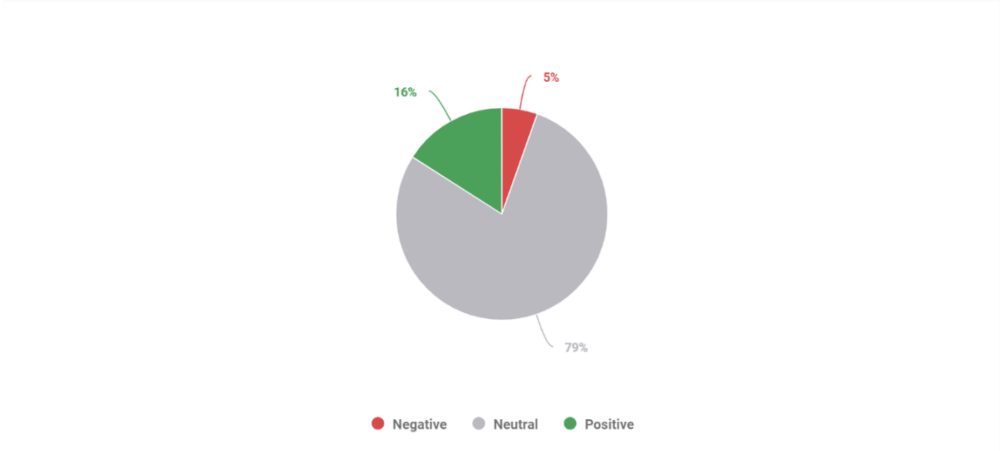


Figure 11. Mention Volume for Queries broken down by Sentiment

A critical factor in our tool is the authors' popularity, which helps the researcher to understand how the sentiment associated with their words and the author's status can

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

influence the opinion of their followers' networks on a particular brand, product or topic (Tedeschi & Benedetto, 2015). We can see in Figure 12 the volume of mentions by sentiment for this factor.

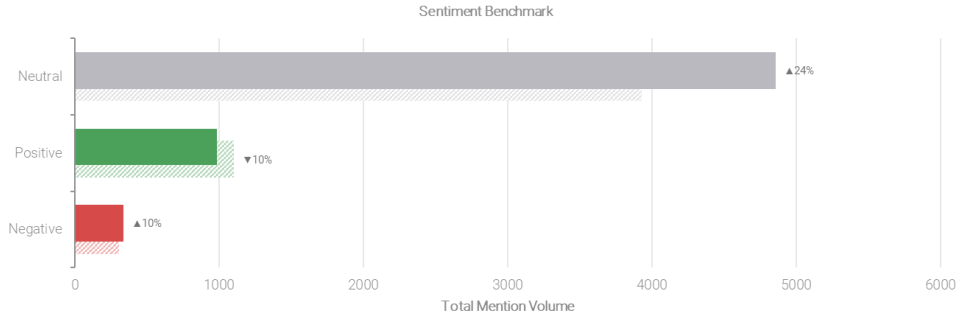


Figure 12. Benchmark Mention Volume for Sentiment

3.2.2 Online Emotion

The AI tools provided by Brandwatch, while potentially useful, like other SMLPs, are mainly black boxes. Sparse information is provided about the construction of the algorithm, leaving researchers with little means of assessing the accuracy of the results (L. Tam & Kim, 2019). However, another loosely documented NLP algorithm, Emotion Analysis, ranks posts based on the six basic emotions Ekman and Friesen (1971) identified through the study of facial expressions: anger, fear, disgust, joy, surprise, and sadness (Ekman & Friesen, 1971; Hayes et al., 2021). The documentation states that the algorithm is trained to identify emotions based on linguistic patterns using a training set of "over 2 million tweets with corresponding emotion labelled". The percentage volume of mentions by emotion collected in our dataset is shown in Figure 13.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

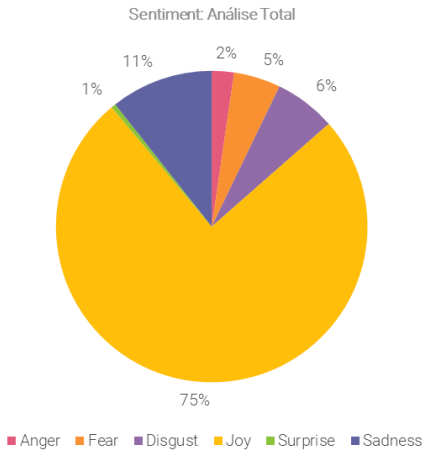


Figure 13. Mention Volume presented by Emotion

After perceiving an emotion, observers infer additional information from the emotional expression. Emotion expressions serve as (social) information in that observers use emotion expressions to draw inferences about characteristics of (a) the expresser, (b) the situation, and/or (c) the self, which can all influence audience behaviour (Lange et al., 2022). To this end, we interleaved the results of emotion by the sentiment of the tweets as shown in Figure 14.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

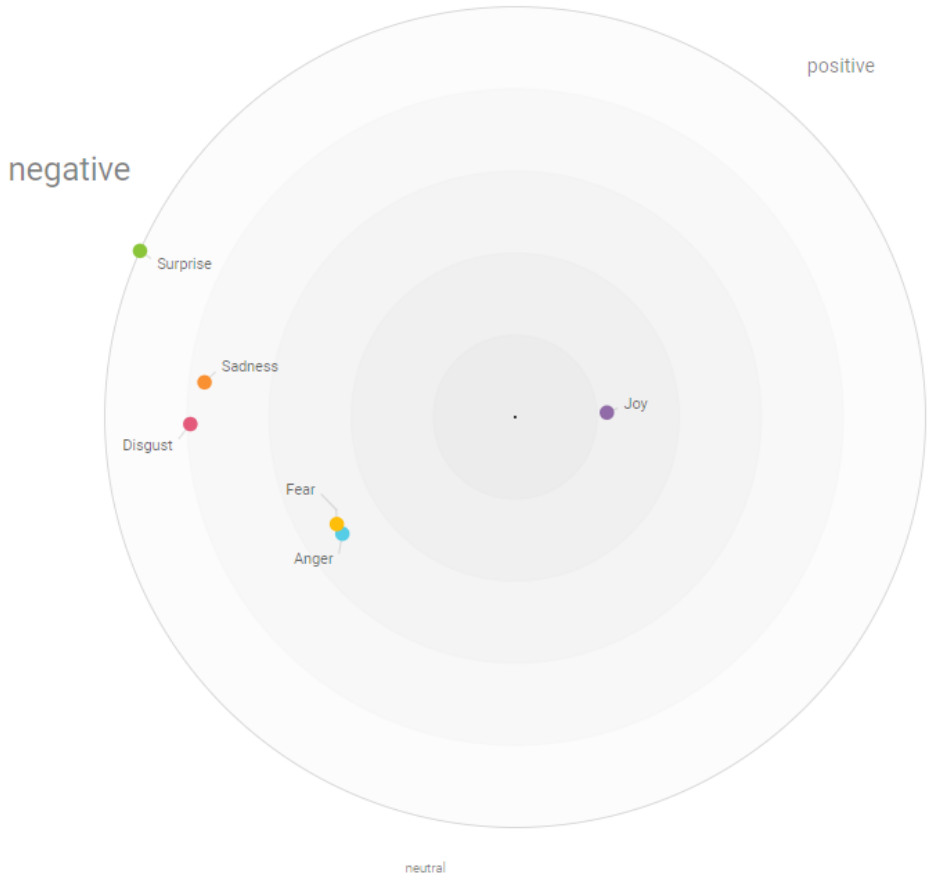


Figure 14. Mention Volume for Emotion broken down by Sentiment.

3.2.3 Online Engagement

Brand tweet volume can be viewed as how engaged stakeholders are with a brand (Robertson et al., 2019). The more discussion a brand can generate, the more engaged it is with its audience (Rust et al., 2021). Figure 15 presents a results framework to examine the Twitter feeds of related nonprofits and examine the effects of messaging features on brand engagement, captured in terms of original Twitter posts, shares and retweets.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

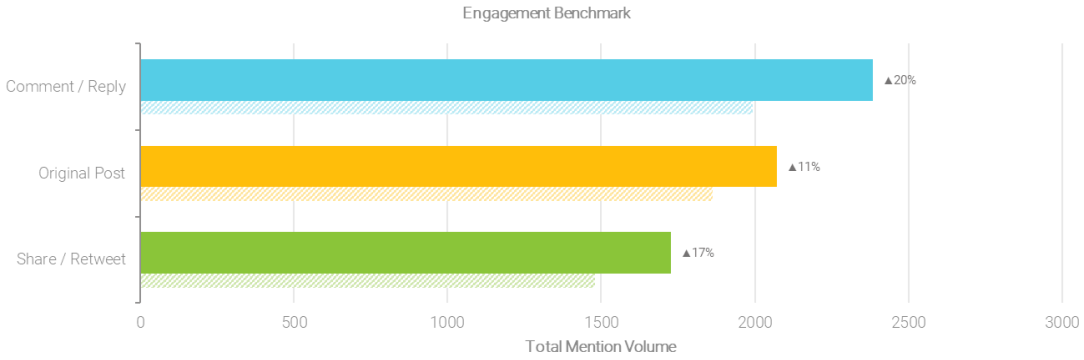


Figure 15. Benchmark Mention Volume for Mention Type

Social media is a significant source of marketing intelligence that can improve brand equity by better understanding and managing brand engagement (Duncan et al., 2019). These results will be possible to draw more significant conclusions concerning the hypotheses raised above.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

4 ANALYSIS AND DISCUSSION OF RESULTS

This chapter aims to analyse the results obtained by pre-processing tweets and their relationship with the user Engagement of NPOs. For this, different comparisons were made to discuss the results.

4.1 Classification Models

In this section, the classification algorithms used to classify each tweet as "neutral", "positive", or "negative" and how each of the algorithms was applied are presented. The classifiers used the Brandwatch labelled dataset and the sampling strategy with cross-validation of ten folders, with the measurement of the ten runs being presented.

The metric used in this work to validate the classifiers was accuracy. It is known that this measure has some shortcomings, such as the possibility that it is not suitable for unbalanced classes, as it tends to privilege the majority class, considering that the rare class is usually more interesting. However, the problem in question presents a few unbalanced classes and, for simplicity and not being the focus of this work, only this measure was used to validate the classifiers.

4.1.1 Naive Bayes

The supervised learning algorithm Naive Bayes multinomial text was chosen to be used in this work because it is widely used in text classification and sentiment analysis tasks and presents good results. To apply it to the Brandwatch labelled dataset, Weka (T. C. Smith & Frank, 2016b) was used, a software written in the Java language that provides a collection of machine learning algorithms to perform tasks related to data mining.

Data from the labelled tweet dataset were converted to ARFF (Attribute Relation File Format), a Weka-specific text file that describes a list of instances that share the same attributes (Paynter et al., 2005). This way, it is possible to use the labelled dataset as a training set for the Naive Bayes algorithm.

After loading the dataset in the Explorer interface of Weka, an overview of the data is shown in the pre-processing tab, as can be seen in Figure 16.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

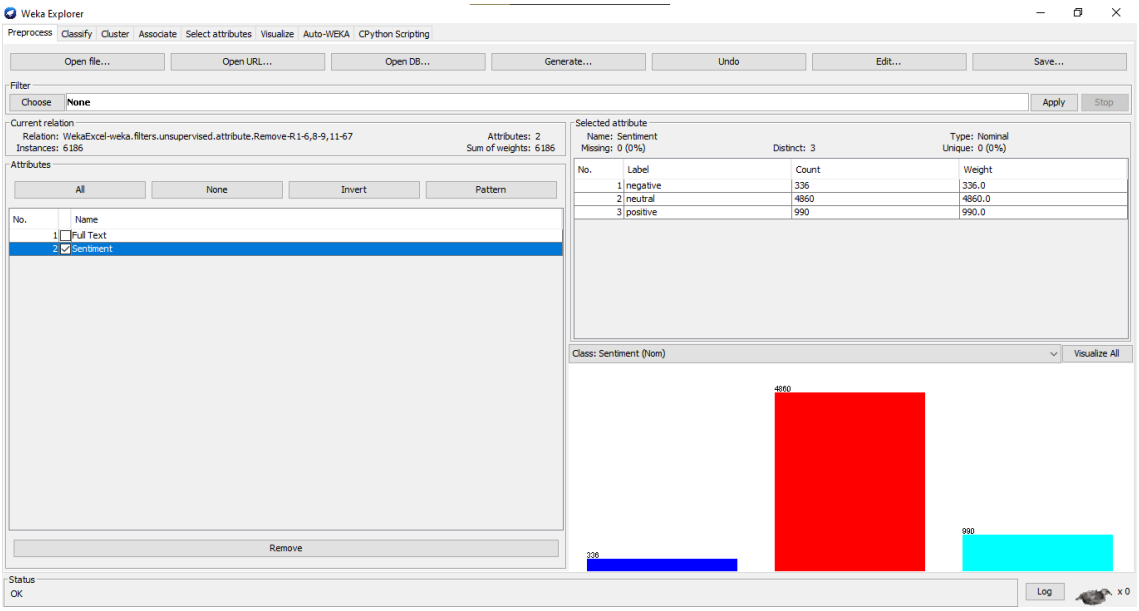


Figure 16. Training set overview in Weka

Then, in the classification tab, the Naive Bayes multinomial text classifier and the 10-fold cross-validation validation strategy are chosen, which consists of dividing the test set into ten random parts, which are individually tested, using the other nine remaining parts as a training set. Weka's default settings were used for Naive Bayes multinomial text.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4860               78.5645 %
Incorrectly Classified Instances      1326               21.4355 %
Kappa statistic                          0
Mean absolute error                      0.2362
Root mean squared error                  0.3436
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances             6186

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,000    0,000    ?          0,000    ?          ?        0,496     0,054     negative
                 1,000    1,000    0,786      1,000    0,880      ?        0,499     0,785     neutral
                 0,000    0,000    ?          0,000    ?          ?        0,500     0,160     positive
Weighted Avg.    0,786    0,786    ?          0,786    ?          ?        0,499     0,646

=== Confusion Matrix ===

    a    b    c   <-- classified as
    0  336    0 |    a = negative
    0 4860    0 |    b = neutral
    0  990    0 |    c = positive
```

Figure 17. Naive Bates multinomial text classifier

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

According to the results presented in Figure 17, after generating the classification model based on the Naive Bayes algorithm by Weka, it is necessary to provide the test sets - in this case, the individual files containing all the tweets collected for each NPO's.

4.1.2 Random Forest

One of these classification and regression techniques used in this work is the random forest approach. These decision tree-based predictors are best known for their excellent computational performance and scalability, and they consist of a combination of tree classifiers where each classifier is generated using a random vector sampled independently of the input vector (Pal, 2005). However, in the case of severely unbalanced training data, as often seen in data from previous studies with large control groups, the training algorithm or sampling process must be changed to improve the prediction quality for minority classes (Amrehn et al., 2019).

In this work, a balanced random forest approach to Weka is proposed. In addition, the predictive quality of the unmodified random forest implementation is shown in Figure 18.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4875               78.807  %
Incorrectly Classified Instances      1311               21.193  %
Kappa statistic                          0.0191
Mean absolute error                      0.2331
Root mean squared error                  0.3414
Relative absolute error                 98.6833 %
Root relative squared error             99.3624 %
Total Number of Instances            6186

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,033    0,000    1,000      0,033    0,063      0,176    0,524     0,089     negative
               1,000    0,989    0,788      1,000    0,881      0,094    0,517     0,799     neutral
               0,004    0,000    1,000      0,004    0,008      0,058    0,514     0,168     positive
Weighted Avg.  0,788    0,777    0,833      0,788    0,697      0,093    0,517     0,659

=== Confusion Matrix ===

    a    b    c   <-- classified as
   11  325    0 |    a = negative
    0 4860    0 |    b = neutral
    0  986    4 |    c = positive
```

Figure 18. Random forest implementation

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

4.1.3 Logistic Regression

Logistic regression aims to produce, from a set of observations, a model that allows the prediction of values taken by a categorical variable, often binary, from a series of continuous and/or binary explanatory variables (Lalitha et al., 2015).

Logistic regression is comparable to supervised techniques proposed in machine learning as a prediction method for categorical variables. Compared to known regression techniques, in particular linear regression, logistic regression is essentially distinguished by the fact that the response variable is categorical (Khalajzadeh et al., 2014). Considering the problem of selecting a learning algorithm and defining its hyperparameters, going beyond previous works that attack these problems separately, we present the results in Figure 19.

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4865               78.6453 %
Incorrectly Classified Instances      1321               21.3547 %
Kappa statistic                          0.0065
Mean absolute error                      0.2357
Root mean squared error                  0.3441
Relative absolute error                 99.7617 %
Root relative squared error            100.1392 %
Total Number of Instances             6186

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,015    0,000    1,000      0,015    0,029      0,119    0,508     0,070     negative
                 1,000    0,996    0,786      1,000    0,880      0,054    0,501     0,786     neutral
                 0,000    0,000    ?          0,000    ?          ?        0,500     0,160     positive
Weighted Avg.    0,786    0,783    ?          0,786    ?          ?        0,501     0,647

=== Confusion Matrix ===

    a    b    c   <-- classified as
    5  331    0 |    a = negative
    0 4860    0 |    b = neutral
    0  990    0 |    c = positive
```

Figure 19. Logistic regression

4.1.4 K-Nearest Neighbour

K-nearest-neighbor (KNN) has been widely used as an effective classification model. We tried to research some improved algorithms and experimentally test their effectiveness (Jiang et al., 2007).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4875               78.807  %
Incorrectly Classified Instances      1311               21.193  %
Kappa statistic                          0.0191
Mean absolute error                      0.2322
Root mean squared error                  0.3412
Relative absolute error                 98.2699 %
Root relative squared error             99.3071 %
Total Number of Instances             6186

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,033    0,000    1,000      0,033    0,063      0,176    0,524     0,087     negative
                 1,000    0,989    0,788      1,000    0,881      0,094    0,516     0,794     neutral
                 0,004    0,000    1,000      0,004    0,008      0,058    0,514     0,167     positive
Weighted Avg.    0,788    0,777    0,833      0,788    0,697      0,093    0,516     0,655

=== Confusion Matrix ===

    a    b    c   <-- classified as
   11  325    0 |    a = negative
    0 4860    0 |    b = neutral
    0  986    4 |    c = positive
```

Figure 20. K-nearest-neighbor

4.1.5 Support Vector Machine

Among the standard classification tools are neural networks and decision trees (from machine learning), logistic regression and discriminant analysis (from traditional statistics); there are emerging tools such as support vector machines (SVMs) (Gandhi et al., 2016).

SVM is a related category of classification tools and rule induction. Unlike a decision tree, in rule induction, the "if-then" statements are induced directly from the training data and do not need to have a hierarchical character (T. C. Smith & Frank, 2016a). The results obtained are presented in Figure 21.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4875               78.807  %
Incorrectly Classified Instances      1311               21.193  %
Kappa statistic                          0.0191
Mean absolute error                      0.281
Root mean squared error                  0.3645
Relative absolute error                118.9413 %
Root relative squared error            106.074  %
Total Number of Instances             6186

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,033    0,000    1,000      0,033    0,063      0,176    0,516     0,085     negative
               1,000    0,989    0,788      1,000    0,881      0,094    0,506     0,788     neutral
               0,004    0,000    1,000      0,004    0,008      0,058    0,503     0,164     positive
Weighted Avg.  0,788    0,777    0,833      0,788    0,697      0,093    0,506     0,650

=== Confusion Matrix ===

    a    b    c   <-- classified as
   11  325    0 |    a = negative
    0 4860    0 |    b = neutral
    0  986    4 |    c = positive
```

Figure 21. Support vector machines

4.1.6 Analysis of classification models

From the data from the previous sections, Table 3 was constructed, where it can be observed that the Random Forest classifier obtained greater accuracy among the evaluated ones. Therefore, this will be the ranking algorithm to rank the complete tweet base.

MODEL	ACCURACY
NAIVE BAYES	78.5645%
RANDOM FOREST	78.807%
LOGISTIC REGRESSION	78.6453%
K-NEAREST-NEIGHBOUR	78.807%
SUPPOR TVECTOR MACHINE	78.807%

Table 3. Ranking algorithm

This process was carried out individually for each of the five NPOs in question, resulting in the following Table 4, where we can observe the feelings of Twitter users concerning the NPOs according to the classification made from the Random Forest model.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

NPO	**Positive**		**Negative**		**Neutral**	
Gates	40	6%	64	10%	525	84%
Ford	141	14%	61	6%	827	80%
RWJF	68	6%	103	10%	870	84%
HHMI	419	39%	7	1%	645	60%
Getty	317	13%	101	4%	1987	83%
Total	985	16%	336	5%	4854	79%

Table 4. NPOs according to the classification made from the Random Forest model

From the analysis of Table 4, it can be noted that, for all NPOs, the number of tweets classified as "positive" is greater than the number of those classified as "negative". This is indicative that NPOs chose to spend more time posting positive tweets on social issues than posting negative reviews on other subjects they did not like.

There is also a high rate of tweets classified as "neutral", which does not add relevant information to the study. It is known that the NPO posted it, but it is not known about the nature of this comment, that is, whether the sentiment expressed in the tweet is positive or negative. This difficulty happens a lot in the Sentiment Analysis task and is one of the challenges in the area.

4.2 Validation of pre-processing

Several pre-processing steps described above were carried out to treat the tweet base to discard useless terms for classification. By observing some clouds of words, it is possible to notice that the pre-processing, for the most part, generated promising results since most of the words that appear in the clouds make references to the themes of the area of activity of the respective NPOs and their relationship with the period lived in 2021.

In Figure 22, which contains the word cloud referring to the tweets published by the Ford Foundation, it can be seen that words of a positive nature, such as "justice", "work", "support", and "join", were quite present among the tweets. Some keywords about the

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

pandemic crisis are mentioned, and several terms related to human rights issues (e.g. "black " and "indigenous") also appear, as this is a topic of interest to the NPO.

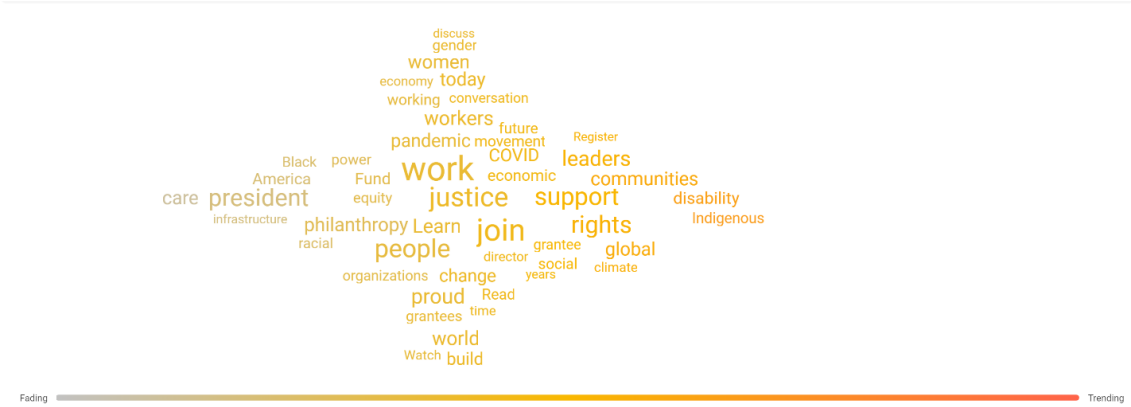


Figure 22. Word cloud contained in tweets collected from the Ford Foundation

The word cloud for the Gates Foundation tweets is shown in Figure 23, and it is possible to see that the NPO has published a lot about the pandemic crisis ("COVID-19" and "pandemic") and other quotes about the health and empowerment of feminine ("women", "girls" and "empowered").

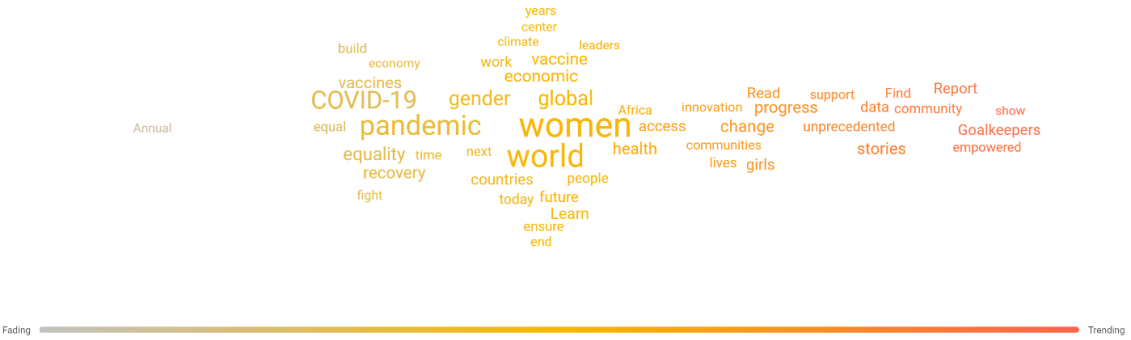


Figure 23. Word cloud contained in tweets collected from the Bill and Melinda Gates Foundation

However, the Getty Museum's word cloud features non-pandemic keywords such as "art", "exhibition", and "photographs". This presents a possible option for the NPO to focus on topics related to its area of expertise and on "positive" words.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 24. Word cloud contained in tweets collected from the Getty Museum

It is also possible to notice that all the words that appear in the clouds are complete; that is, no slang or abbreviations appeared, which means that the queries built to analyse these tweets worked well.

4.3 Comparison of tweet sentiment with CNN algorithms

To analyse the results, the approach based on CNN integrated with GloVe will henceforth be called CNN, the approach based on CNN integrated with BERT as Tokenizer will be called BERT and finally, the approach based on BERT as Embedding Layer will be BERT-CNN.

The BERT-CNN model is compared with the original CNN model and the BERT model. In order to ensure the accuracy and objectivity of the experimental results, the three models are run five times on the same training set and test set, respectively, to solve for accuracy as the results of the final model, as shown in Table 5.

Model	***Training Data***	***Test Data***
CNN	*0.7217*	*0.7142*
BERT	*0.7890*	*0.7591*
BERT-CNN	*0.8214*	*0.8032*

Table 5. Accuracy as the results of the final model

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

As shown in Table 5, the accuracy values of both the BERT-CNN and BERT models are higher than those of the CNN model in the training set and test set, mainly because both models are unsupervised pre-training models for sets. Large-scale text messages that do not need artificial participation, and its attention engine can precisely connect semantic contexts to explore precise sentence meanings from social media review texts like Twitter and semantic logic between sub-phrases in each NPO's tweet. Furthermore, the difference between the accuracy value of the BERT-CNN model in the training set and the test set is minimal, so the model has favourable generalization capability.

4.4 Analysis of the correlation between the results

From the data obtained with the application of BERT-CNN model, we calculated the Engagement rate through the methodology based by Okazaki et al. (2015), presented in Equation (4):

$$\frac{Engaged}{Outreach} \tag{4}$$

In which, according to the aforementioned author, Engaged is the sum of interactions (likes, comments, retweets), and Outreach is the reach that the publication (Tweet) has in its audience.

In order to establish the correlation between Engagement and the sentiment of tweets in Nonprofit Marketing, represented by five different NPOs, the non-parametric correlation, rho of Spearman, was used with SPSS software.

Spearman's correlation uses ranks rather than the absolute values of the variables (Malhotra, 2010), and we know that the analysed variables do not have interval scaling or ratio properties and do not assume a normal distribution (Fávero & Belfiore, 2017).

This way, Spearman's rho correlation coefficient between user Engagement and the sentiment was calculated. The results are shown in Table 6 to find out if there is any significant association between the variables.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

	GATES	HHMI	FORD	RWJF	GETTY
NEUTRAL	-0.052	-0.144**	0.049	0.048	0.092**
POSITIVE	-0.038	-0.153**	-0.072*	-0.154**	-0.118**
NEGATIVE	0.094*	0.052	0.024	0.067*	0.025

Table 6. Correlation between Engagement Rate and the sentiment of tweets in Nonprofit Marketing

** The correlation is significant at the 0.01 level (2 ends)

* The correlation is significant at the 0.05 level (2 ends)

4.5 Discussion

In order to relate directly to the questions posed in the Introduction of this dissertation and to position the results obtained within the context of the literature review, an analysis by NPO is presented below.

It can be seen that, although none of the coefficients found to represent a correlation very close to 1 or -1 between the NPOs, it is still quite likely to predict what would be the tone of the message that tends to have the worst Engagement, in this case, positive, and which NPO's would be among the least prestigious by their audience, Howard Hughes Medical Institute and Robert Wood Johnson Foundation, for example. In other words, from this methodology, it is possible to predict which feeling expressed in the message will have the best Engagement, despite the low correlation. We can say that using different models of sentiment analysis for data disclosed on social networks, namely on Twitter, in a set of disclosures by institutions classified as Non-Profit Marketing can help the decision-making process of whether or not to use promotions to increase/recover user engagement.

4.5.1 Bill & Melinda Gates Foundation (@gatesfoundation)

The results obtained for the correlation analysis referring to the Gates Foundation's Twitter profile are statistically significant for tweets classified as Negative, so hypotheses H2 and H3 are verified.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The greater effect observed in negative publications can be supported by the characteristic of this NPO, which, according to the foundation's website, fights against poverty, disease and inequality, and can be seen in the example below:

Figure 25. Gates Foundation Tweet of April 15, 2021

Figure 25 presents the publication with the highest engagement of the Gates Foundation in 2021 and exemplifies the results presented in the previous section.

4.5.2 Howard Hughes Medical Institute (@HHMINEWS)

As can be seen in Table 6, the results referring to the Howard Hughes Medical Institute show a negative and statistically significant relationship in the use of positive and neutral publications. Although negative publications are not statistically significant, we can state that hypotheses H2 and H3 are also met for this NPO.

4.5.3 Ford Foundation (@fordfoundation)

The Ford Foundation, an institution created to fund programs to promote democracy, reduce racism and poverty, has a negative and significant correlation with Positive messages. The results presented in table 6 demonstrate that the message descriptions indicate the core of the intention and seriousness of supporting a cause and confirm hypotheses H1, H2 and H3 for this NPO. This conclusion is in line with what has already been concluded by Okazaki et al. (2015) and Tripathi and Verma (2018).

4.5.4 Robert Wood Johnson Foundation (@RWJF)

The Robert Wood Johnson Foundation is an American philanthropic organisation focused exclusively on health. The results obtained in the previous section demonstrate that the effort of the relationship between message sentiment and engagement has an equally positive impact on Neutral and Negative messages. Meanwhile, Positive messages impact negatively and statistically significantly.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

All these findings are duly aligned with the hypotheses presented, demonstrating that the same continues to be verifiable for the situation, particularly for the Robert Wood Johnson Foundation.

4.5.5 J. Paul Getty Trust – Getty Museum (@gettymuseum)

The Getty Museum Twitter profile represents communication between the Paul Getty Museum and the Getty Research Institute. It is a cultural and philanthropic institution dedicated to presenting, conserving, and interpreting the world's artistic legacy.

The results for this NPO support hypothesis H1, given that there is a positive and statistically significant relationship between the Neutral message and Engagement. The relationship between the Negative message and Engagement, although positive, is not statistically significant, and a positive relationship cannot be clearly assessed in hypothesis H3. However, when analysing the effects between the Positive message and Engagement, there is a negative and statistically significant relationship, whereby both the use of Positive and Neutral messages influences the Engagement of the audience.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

5 CONCLUSION

Data-driven Marketing is understood as data-driven analysis, enabled by information systems, technologies, methodologies, and practices, such as the use of Big Data, that allows the extraction of relevant data and its transformation into business insights (Nuortimo & Harkonen, 2019) through data collection - which are often publicly available - and online research allowing to understand better consumer needs and competitors' current offerings (Meng et al., 2018).

Within this context, this study falls within management, and in particular marketing, a perspective that considers the importance of impact in using individual-level data to identify Twitter touchpoint sequences and determine the relative role of each touchpoint in Nonprofit Marketing profile Engagement. This work directed our contribution towards the influence of the sentiments analysed in owner tweets on engagement, as well as on the respective interaction with followers.

In particular, the study revealed that sentiment analysis of tweets influences engagement and provides some evidence, albeit limited, that followers interact with profiles on specific features of tweets per Nonprofit segment. Next, we present the contributions of our work, its implications for managers, and its limitations. We conclude by adducing new directions of research.

5.1 Main contributions

The study that is now concluded brings contributions to academic and professional research because, synthetically, it contributes to the investigation of sentiment analysis and its implications for follower participation and the moderating effects of this phenomenon on the relationship between tweet characteristics and engagement. At the same time, it also serves as an asset to Marketing professionals in general, particularly those whose objective is to implement customer-focused strategies in Nonprofit Marketing effectively.

The present study found traces of feelings in tweets that effectively influence engagement, particularly in the Nonprofit area. Tweets are related to the formation of individuals' perceptions about their goals in participating in the production of nonprofit services, and

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

these perceived characteristics have consequences on their engagement. At the same time, this study confirmed that several features of Natural Language Processing (NPL) applications, namely Convolutional Neural Networks (CNN), in particular with the use of BERT and Glove, have a relevant effect on sentiment analysis. It should also be noted that the relationship between Twitter data in Nonprofit Marketing, collected from Brandwatch, and validation with classification models were tested for the first time in this work.

This study contributed to research in the Nonprofit Marketing area, considering the importance sentiment analysis has in interpreting followers' interactions with the tweet. In this domain, it is believed to be innovative. In summary, the most innovative contribution that we brought with this work was to analyse the extent to which the feelings of those targeted in the tweets contribute to engagement and the respective interaction with the characteristics of NPL tools.

5.2 Work limitations

This work has some limitations. It is worth mentioning them here.

First, the sample is not representative of the universe of individuals involved in this service sector, so the data may not be generalizable to the national reality or the Nonprofit sector in general.

Secondly, we considered the moderating effect of the feelings of the tweets in the relationship with engagement, not considering other contextual variables, that is, linked to the organization providing the service that can also interact with engagement.

Thirdly, when analysing only the Brandwatch sample, a higher percentage of tweets with a neutral sentiment rating was obtained, which is consistent with the fact that Nonprofit companies do not show feelings about specific facts. However, this more significant number of neutral tweets concerning positive and negative may have had consequences on the result of training the dataset for sentiment analysis. Thus, it is recommended that future studies consider the interaction between personal values and/or cultural values.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

5.3 Directions for future investigations

This study aimed to determine the influence of the sentiments of tweets in the context of Nonprofit Marketing and the influence of tweets on the relationship between audience engagement.

This investigation shows a relationship between some of the dimensions of feelings and engagement, arousing the curiosity of researchers to continue with complementary studies. There is still significant scope for investigation in the field of audience behaviour studies; as our approach focused on neutral, positive, and negative feelings, future research may consider other variables for the formation of antecedent factors of engagement such as individual, social and organizational factors.

Finally, this study also offers a platform for future research. Concerning feelings, this variable can be combined with other personality variables (e.g., big five personality traits, etc.) to determine the individual contribution of each of them, as well as the contribution resulting from moderating effects.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

REFERENCES

Abakouy, R., En-naimi, E. M., Haddadi, A. E., & Lotfi, E. (2019). Data-driven marketing: How machine learning will improve decision-making for marketers. *Proceedings of the 4th International Conference on Smart City Applications*, 1–5. https://doi.org/10.1145/3368756.3369024

Abedin, B., Maloney, B., & Watson, J. (2021). Benefits and Challenges Associated with Using Online Communities by Social Enterprises: A Thematic Analysis of Qualitative Interviews. *Journal of Social Entrepreneurship*, *12*(2), 197–218. https://doi.org/10.1080/19420676.2019.1683879

Ahlemeyer-Stubbe, A., & Müller, A. (2020). How to leverage internet of things data to generate benefits for sales and marketing. *Applied Marketing Analytics*, *5*(3), 233–242.

Albanna, H., Alalwan, A. A., & Al-Emran, M. (2022). An integrated model for using social media applications in non-profit organizations. *International Journal of Information Management*, *63*, 102452. https://doi.org/10.1016/j.ijinfomgt.2021.102452

Algharabat, R., Rana, N. P., Dwivedi, Y. K., Alalwan, A. A., & Qasem, Z. (2018). The effect of telepresence, social presence and involvement on consumer brand engagement: An empirical study of non-profit organizations. *Journal of Retailing and Consumer Services*, *40*, 139–149. https://doi.org/10.1016/j.jretconser.2017.09.011

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Amrehn, M., Mualla, F., Angelopoulou, E., Steidl, S., & Maier, A. (2019). *The Random Forest Classifier in WEKA: Discussion and New Developments for Imbalanced Data* (arXiv:1812.08102). arXiv. https://doi.org/10.48550/arXiv.1812.08102

Andhale, S., Mane, P., Vaingankar, M., Karia, D., & Talele, K. T. (2021). Twitter Sentiment Analysis for COVID-19. *2021 International Conference on Communication Information and Computing Technology (ICCICT)*, 1–12. https://doi.org/10.1109/ICCICT50803.2021.9509933

Andreasen, A. R., & Kotler, P. (2003). *Strategic marketing for nonprofit organizations* (6th ed). Prentice Hall.

Barroso-Méndez, M. J., Galera-Casquet, C., Valero-Amaro, V., & Nevado-Gil, M. T. (2020). Antecedents of Relationship Learning in Business-Non-Profit Organization Collaboration Agreements. *Sustainability*, *12*(1), Article 1. https://doi.org/10.3390/su12010269

Batbaatar, E., Li, M., & Ryu, K. H. (2019). Semantic-Emotion Neural Network for Emotion Recognition From Text. *IEEE Access*, *7*, 111866–111878. https://doi.org/10.1109/ACCESS.2019.2934529

Beldad, A., Snip, B., & van Hoof, J. (2014). Generosity the Second Time Around: Determinants of Individuals' Repeat Donation Intention. *Nonprofit and Voluntary Sector Quarterly*, *43*(1), 144–163. https://doi.org/10.1177/0899764012457466

Bennett, R., & Gabriel, H. (2003). Image and Reputational Characteristics of UK Charitable Organizations: An Empirical Study. *Corporate Reputation Review*, *6*(3), 276–289. https://doi.org/10.1057/palgrave.crr.1540206

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Bernardi, C. L., & Alhamdan, N. (2022). Social media analytics for nonprofit marketing:# Downsyndrome on Twitter and Instagram. *Journal of Philanthropy and Marketing*, e1739.

Bilgic, E., Cakir, O., Kantardzic, M., Duan, Y., & Cao, G. (2021). Retail analytics: Store segmentation using Rule-Based Purchasing behavior analysis. *The International Review of Retail, Distribution and Consumer Research*, *31*(4), 457–480. https://doi.org/10.1080/09593969.2021.1915847

Borges, M., Bernardino, J., & Pedrosa, I. (2021). Data-driven decision making strategies applied to marketing. *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–7. https://doi.org/10.23919/CISTI52073.2021.9476506

Brandwatch Help Center. (2022). *Sentiment Analysis*. Consumer Research. https://consumer-research-help.brandwatch.com/hc/en-us/articles/360013739958-Sentiment-Analysis

Breese, E. B. (2016). When Marketers and Academics Share a Research Platform: The Story of Crimson Hexagon. *Journal of Applied Social Science*, *10*(1), 3–7. https://doi.org/10.1177/1936724415569953

Capatina, A., Kachour, M., Lichy, J., Micu, A., Micu, A.-E., & Codignola, F. (2020). Matching the future capabilities of an artificial intelligence-based software for social media marketing with potential users' expectations. *Technological Forecasting and Social Change*, *151*, 119794. https://doi.org/10.1016/j.techfore.2019.119794

Capra, C. M., Jiang, B., & Su, Y. (2021). Altruistic self-concept mediates the effects of personality traits on volunteering: Evidence from an online experiment. *Journal*

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

of Behavioral and Experimental Economics, *92*, 101697. https://doi.org/10.1016/j.socec.2021.101697

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, *12*(ARTICLE), 2493–2537.

Dalenberg, D. J. (2018). Preventing discrimination in the automated targeting of job advertisements. *Computer Law & Security Review*, *34*(3), 615–627. https://doi.org/10.1016/j.clsr.2017.11.009

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv:1810.04805 [Cs]*. http://arxiv.org/abs/1810.04805

Dong, J., He, F., Guo, Y., & Zhang, H. (2020). A Commodity Review Sentiment Analysis Based on BERT-CNN Model. *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, 143–147. https://doi.org/10.1109/ICCCS49078.2020.9118434

Duncan, S. Y., Chohan, R., & Ferreira, J. J. (2019). What makes the difference? Employee social media brand engagement. *Journal of Business & Industrial Marketing*, *34*(7), 1459–1467. https://doi.org/10.1108/JBIM-09-2018-0279

Eke, C. I., Norman, A. A., & Shuib, L. (2021). Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model. *IEEE Access*, *9*, 48501–48518. https://doi.org/10.1109/ACCESS.2021.3068323

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. https://doi.org/10.1037/h0030377

Fávero, L. P., & Belfiore, P. (2017). *Manual de Análise de Dados: Estatística e Modelagem Multivariada com Excel®, SPSS® e Stata®*. Elsevier Brasil.

Febriani, D. M., & Selamet, J. (2020). College Students' Intention to Volunteer for Non-profit Organizations: Does Brand Image Make a Difference? *Journal of Nonprofit & Public Sector Marketing*, *32*(2), 166–188. https://doi.org/10.1080/10495142.2019.1656136

Fiarni, C., Maharani, H., & Wisastra, G. R. (2020). Opinion Mining Model System For Indonesian Non Profit Organization Using Multinomial Naive Bayes Algorithm. *2020 8th International Conference on Information and Communication Technology (ICoICT)*, 1–7. https://doi.org/10.1109/ICoICT49345.2020.9166391

Gainer, B., & Padanyi, P. (2005). The relationship between market-oriented activities and market-oriented culture: Implications for the development of market orientation in nonprofit service organizations. *Journal of Business Research*, *58*(6), 854–862. https://doi.org/10.1016/j.jbusres.2003.10.005

Gandhi, N., Armstrong, L. J., Petkar, O., & Tripathy, A. K. (2016). Rice crop yield prediction in India using support vector machines. *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 1–5. https://doi.org/10.1109/JCSSE.2016.7748856

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-Dependent Sentiment Classification With BERT. *IEEE Access*, *7*, 154290–154299. https://doi.org/10.1109/ACCESS.2019.2946594

Goldberg, Y. (2016). A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, *57*, 345–420.

Grandhi, B., Patwa, N., & Saleem, K. (2020). Data-driven marketing for growth and profitability. *EuroMed Journal of Business*. Scopus. https://doi.org/10.1108/EMJB-09-2018-0054

Gray, F. E., Murray, N., & Hopkins, K. (2021). Affordances of e-Newsletters for NPO General-Public Stakeholders. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *32*(5), 1165–1181. https://doi.org/10.1007/s11266-021-00374-2

Guo, C., & Saxton, G. D. (2018). Speaking and Being Heard: How Nonprofit Advocacy Organizations Gain Attention on Social Media. *Nonprofit and Voluntary Sector Quarterly*, *47*(1), 5–26. https://doi.org/10.1177/0899764017713724

Guo, M., Liao, X., Liu, J., & Zhang, Q. (2020). Consumer preference analysis: A data-driven multiple criteria approach integrating online information. *Omega*, *96*, 102074. https://doi.org/10.1016/j.omega.2019.05.010

Ha, Q.-A., Pham, P. N. N., & Le, L. H. (2022). What facilitate people to do charity? The impact of brand anthropomorphism, brand familiarity and brand trust on charity support intention. *International Review on Public and Nonprofit Marketing*. https://doi.org/10.1007/s12208-021-00331-1

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Hayes, J. L., Britt, B. C., Evans, W., Rush, S. W., Towery, N. A., & Adamson, A. C. (2021). Can Social Media Listening Platforms' Artificial Intelligence Be Trusted? Examining the Accuracy of Crimson Hexagon's (Now Brandwatch Consumer Research's) AI-Driven Analyses. *Journal of Advertising*, *50*(1), 81–91. https://doi.org/10.1080/00913367.2020.1809576

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-Based Sentiment Analysis using BERT. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 187–196. https://www.aclweb.org/anthology/W19-6120

Hsu, C.-W., Chang, Y.-L., Chen, T.-S., Chang, T.-Y., & Lin, Y.-D. (2021). Who Donates on Line? Segmentation Analysis and Marketing Strategies Based on Machine Learning for Online Charitable Donations in Taiwan. *IEEE Access*, *9*, 52728–52740. https://doi.org/10.1109/ACCESS.2021.3066713

Huang, B., Ou, Y., & Carley, K. M. (2018). Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks. In R. Thomson, C. Dancy, A. Hyder, & H. Bisgin (Eds.), *Social, Cultural, and Behavioral Modeling* (pp. 197–206). Springer International Publishing. https://doi.org/10.1007/978-3-319-93372-6_22

Ihm, J. (2016). *More than a moral person: How do communication and identity influence online and offline engagement*.

Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

19 Related Tweets. *IEEE Access*, *8*, 181074–181090. https://doi.org/10.1109/ACCESS.2020.3027350

Jabbar, A., Akhtar, P., & Dani, S. (2020). Real-time big data processing for instantaneous marketing decisions: A problematization approach. *Industrial Marketing Management*, *90*, 558–569. https://doi.org/10.1016/j.indmarman.2019.09.001

Jiang, L., Cai, Z., Wang, D., & Jiang, S. (2007). Survey of Improving K-Nearest-Neighbor for Classification. *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, *1*, 679–683. https://doi.org/10.1109/FSKD.2007.552

Johnson, D. S., Sihi, D., & Muzellec, L. (2021). Implementing Big Data Analytics in Marketing Departments: Mixing Organic and Administered Approaches to Increase Data-Driven Decision Making. *Informatics*, *8*(4), Article 4. https://doi.org/10.3390/informatics8040066

Jordan, S. R., Rudeen, S., Hu, D., Diotalevi, J. L., Brown, F. I., Miskovic, P., Yang, H., Colonna, M., & Draper, D. (2019). The Difference a Smile Makes: Effective Use of Imagery by Children's Nonprofit Organizations. *Journal of Nonprofit & Public Sector Marketing*, *31*(3), 227–248. https://doi.org/10.1080/10495142.2018.1526739

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. *ArXiv:1404.2188 [Cs]*. http://arxiv.org/abs/1404.2188

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Kawada, K., Miyake, T., Akiyama, A., & Mugita, T. (2019). Data-driven marketing to accelerate decision making. *Fujitsu Scientific and Technical Journal*, *55*(4), 50–56. Scopus.

Khalajzadeh, H., Mansouri, M., & Teshnehlab, M. (2014). Face Recognition Using Convolutional Neural Network and Simple Logistic Classifier. In V. Snášel, P. Krömer, M. Köppen, & G. Schaefer (Eds.), *Soft Computing in Industrial Applications* (pp. 197–207). Springer International Publishing. https://doi.org/10.1007/978-3-319-00930-8_18

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. *ArXiv:1408.5882 [Cs]*. http://arxiv.org/abs/1408.5882

Lalitha, S., Mudupu, A., Nandyala, B. V., & Munagala, R. (2015). Speech emotion recognition using DWT. *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 1–4. https://doi.org/10.1109/ICCIC.2015.7435630

Lange, J., Heerdink, M. W., & van Kleef, G. A. (2022). Reading emotions, reading people: Emotion perception and inferences drawn from perceived emotions. *Current Opinion in Psychology*, *43*, 85–90. https://doi.org/10.1016/j.copsyc.2021.06.008

Lee, Y. (2022). Social media capital and civic engagement: Does type of connection matter? *International Review on Public and Nonprofit Marketing*, *19*(1), 167–189. https://doi.org/10.1007/s12208-021-00300-8

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Lee, Y.-J. (2019). Nonprofit Marketing Expenses: Who Spends More than Others? *Journal of Nonprofit & Public Sector Marketing*, *33*(3), 385–402. https://doi.org/10.1080/10495142.2019.1707743

Li, H., Zhang, Z., & Liu, Z. (2017). Application of Artificial Neural Networks for Catalysis: A Review. *Catalysts*, *7*(10), 306. https://doi.org/10.3390/catal7100306

Li, J., Ni, X., Yuan, Y., & Wang, F.-Y. (2018). A hierarchical framework for ad inventory allocation in programmatic advertising markets. *Electronic Commerce Research and Applications*, *31*, 40–51. Scopus. https://doi.org/10.1016/j.elerap.2018.09.001

Li, X., Bing, L., Zhang, W., & Lam, W. (2019). Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. *ArXiv:1910.00883 [Cs]*. http://arxiv.org/abs/1910.00883

Liao, C.-H. (2020). Evaluating the Social Marketing Success Criteria in Health Promotion: A F-DEMATEL Approach. *International Journal of Environmental Research and Public Health*, *17*(17), Article 17. https://doi.org/10.3390/ijerph17176317

Lien Minh, D., Sadeghi-Niaraki, A., Huy, H. D., Min, K., & Moon, H. (2018). Deep Learning Approach for Short-Term Stock Trends Prediction Based on Two-Stream Gated Recurrent Unit Network. *IEEE Access*, *6*, 55392–55404. https://doi.org/10.1109/ACCESS.2018.2868970

Maleki, F., & Hosseini, S. M. (2020). Charity donation intention via m-payment apps: Donor-related, m-payment system-related, or charity brand-related factors, which

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

one is overkill? *International Review on Public and Nonprofit Marketing*, *17*(4), 409–443. https://doi.org/10.1007/s12208-020-00254-3

Malhotra, N. K. (2010). *Marketing research: An applied orientation* (6th ed). Pearson.

Marzouki, A., Chouikh, A., Mellouli, S., & Haddad, R. (2021). From Sustainable Development Goals to Sustainable Cities: A Social Media Analysis for Policy-Making Decision. *Sustainability*, *13*(15), Article 15. https://doi.org/10.3390/su13158136

McLeish, B. J. (2010). *Successful Marketing Strategies for Nonprofit Organizations: Winning in the Age of the Elusive Donor*. John Wiley & Sons.

Meijer, M.-M. (2009). The Effects of Charity Reputation on Charitable Giving. *Corporate Reputation Review*, *12*(1), 33–42. https://doi.org/10.1057/crr.2009.5

Meng, H. (Meg), Zamudio, C., & Jewell, R. D. (2018). Unlocking competitiveness through scent names: A data-driven approach. *Business Horizons*, *61*(3), 385–395. https://doi.org/10.1016/j.bushor.2018.01.004

Mesler, R. M., & Simpson, B. (2021). How Affective Displays and Self-Construal Impact Consumers' Generosity. *Journal of Nonprofit & Public Sector Marketing*, *0*(0), 1–26. https://doi.org/10.1080/10495142.2021.1939225

Micheaux, A., & Bosio, B. (2019). Customer Journey Mapping as a New Way to Teach Data-Driven Marketing as a Service. *Journal of Marketing Education*, *41*(2), 127–140. https://doi.org/10.1177/0273475318812551

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Michel, G., & Rieunier, S. (2012). Nonprofit brand image and typicality influences on charitable giving. *Journal of Business Research*, *65*(5), 701–707. https://doi.org/10.1016/j.jbusres.2011.04.002

Miklosik, A., Kuchta, M., Evans, N., & Zak, S. (2019). Towards the Adoption of Machine Learning-Based Analytical Tools in Digital Marketing. *IEEE Access*, *7*, 85705–85718. https://doi.org/10.1109/ACCESS.2019.2924425

Mottner, S., & Wymer, W. (2011). Nonprofit Education: Course Offerings and Perceptions in Accredited U.S. Business Schools. *Journal of Nonprofit & Public Sector Marketing*, *23*(1), 1–19. https://doi.org/10.1080/10495142.2010.508207

Nadler, A., & McGuigan, L. (2018). An impulse to exploit: The behavioral turn in data-driven marketing. *Critical Studies in Media Communication*, *35*(2), 151–165. https://doi.org/10.1080/15295036.2017.1387279

Nuortimo, K., & Harkonen, J. (2019). Establishing an automated brand index based on opinion mining: Analysis of printed and social media. *Journal of Marketing Analytics*, *7*(3), 141–151. https://doi.org/10.1057/s41270-019-00060-9

Okazaki, S., Díaz-Martín, A., Suplet, M., & Menéndez, H. (2015). *Using twitter to engage with customers: A data mining approach*. *25*, 416–434. https://doi.org/10.1108/IntR-11-2013-024

Ozgormus, E., & Smith, A. E. (2020). A data-driven approach to grocery store block layout. *Computers & Industrial Engineering*, *139*, 105562. https://doi.org/10.1016/j.cie.2018.12.009

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Paynter, G., Trigg, L., Frank, E., & Kirkby, R. (2005). Attribute-Relation File Format (ARFF), 2002. *URL Http://Www. Cs. Waikato. Ac. Nz/~ Ml/Weka/Arff. Html. Access Date*, *4*(11).

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Phang, D. C. W., Wang, K., Wang, Q., Kauffman, R. J., & Naldi, M. (2020). A 2020 perspective on "How to derive causal insights for digital commerce in China? A research commentary on computational social science methods". *Electronic Commerce Research and Applications*, *41*, 100975. https://doi.org/10.1016/j.elerap.2020.100975

Pope, J. A., Isely, E. S., & Asamoa-Tutu, F. (2009). Developing a Marketing Strategy for Nonprofit Organizations: An Exploratory Study. *Journal of Nonprofit & Public Sector Marketing*, *21*(2), 184–201. https://doi.org/10.1080/10495140802529532

Pota, M., Esposito, M., Palomino, M. A., & Masala, G. L. (2018). A Subword-Based Deep Learning Approach for Sentiment Analysis of Political Tweets. *2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, 651–656. https://doi.org/10.1109/WAINA.2018.00162

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Puteh, N. (2021). Sentiment Analysis with Deep Learning: A Bibliometric Review. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(3), 1509–1519.

Reyes-Menendez, A., Saura, J. R., & Alvarez-Alonso, C. (2018). Understanding #WorldEnvironmentDay User Opinions in Twitter: A Topic-Based Sentiment Analysis Approach. *International Journal of Environmental Research and Public Health*, *15*(11), Article 11. https://doi.org/10.3390/ijerph15112537

Robertson, J., Lord Ferguson, S., Eriksson, T., & Näppä, A. (2019). The brand personality dimensions of business-to-business firms: A content analysis of employer reviews on social media. *Journal of Business-to-Business Marketing*, *26*(2), 109–124. https://doi.org/10.1080/1051712X.2019.1603354

Rust, R. T., Rand, W., Huang, M.-H., Stephen, A. T., Brooks, G., & Chabuk, T. (2021). Real-Time Brand Reputation Tracking Using Social Media. *Journal of Marketing*, *85*(4), 21–43. https://doi.org/10.1177/0022242921995173

Sargeant, A., Ford, J. B., & Hudson, J. (2008). Charity Brand Personality: The Relationship With Giving Behavior. *Nonprofit and Voluntary Sector Quarterly*, *37*(3), 468–491. https://doi.org/10.1177/0899764007310732

Schetgen, L., Bogaert, M., & Van den Poel, D. (2021). Predicting donation behavior: Acquisition modeling in the nonprofit sector using Facebook data. *Decision Support Systems*, *141*, 113446. https://doi.org/10.1016/j.dss.2020.113446

Senvar, O., Akburak, D., & Yel, N. (2020). Customer oriented intelligent DSS based on two-phased clustering and integrated interval type-2 fuzzy AHP and hesitant

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

fuzzy TOPSIS. *Journal of Intelligent & Fuzzy Systems*, *39*(5), 6121–6143. https://doi.org/10.3233/JIFS-189084

Shah, D., & Murthi, B. P. S. (2020). Marketing in a data-driven digital world: Implications for the role and scope of marketing. *Journal of Business Research*. Scopus. https://doi.org/10.1016/j.jbusres.2020.06.062

Shapiro, B. P. (1974). Marketing in Nonprofit Organizations. *Journal of Voluntary Action Research*, *3*(3–4), 1–16. https://doi.org/10.1177/089976407400300301

Sheth, J., & Kellstadt, C. H. (2021). Next frontiers of research in data driven marketing: Will techniques keep up with data tsunami? *Journal of Business Research*, *125*, 780–784. https://doi.org/10.1016/j.jbusres.2020.04.050

Sircar, A., & Rowley, J. (2016). Social Media and Megachurches. In Y. K. Dwivedi, M. Mäntymäki, M. N. Ravishankar, M. Janssen, M. Clement, E. L. Slade, N. P. Rana, S. Al-Sharhan, & A. C. Simintiras (Eds.), *Social Media: The Good, the Bad, and the Ugly* (pp. 695–700). Springer International Publishing. https://doi.org/10.1007/978-3-319-45234-0_62

Sleep, S., Hulland, J., & Gooner, R. A. (2019). The Data Hierarchy: Factors influencing the adoption and implementation of data-driven decision making. *AMS Review*, *9*(3), 230–248. https://doi.org/10.1007/s13162-019-00146-8

Smith, J. N. (2018). The Social Network?: Nonprofit Constituent Engagement Through Social Media. *Journal of Nonprofit & Public Sector Marketing*, *30*(3), 294–316. https://doi.org/10.1080/10495142.2018.1452821

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Smith, T. C., & Frank, E. (2016a). Introducing Machine Learning Concepts with WEKA. In E. Mathé & S. Davis (Eds.), *Statistical Genomics: Methods and Protocols* (pp. 353–378). Springer. https://doi.org/10.1007/978-1-4939-3578-9_17

Smith, T. C., & Frank, E. (2016b). Statistical genomics: Methods and protocols, chapter introducing machine learning concepts with WEKA. *Statistical Genomics: Methods and Protocols*, 353–378.

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence. *ArXiv:1903.09588 [Cs]*. http://arxiv.org/abs/1903.09588

Tam, L., & Kim, J.-N. (2019). Social media analytics: How they support company public relations. *Journal of Business Strategy*.

Tam, S., Said, R. B., & Tanriöver, Ö. Ö. (2021). A ConvBiLSTM Deep Learning Model-Based Approach for Twitter Sentiment Classification. *IEEE Access*, *9*, 41283–41293. https://doi.org/10.1109/ACCESS.2021.3064830

Tedeschi, A., & Benedetto, F. (2015). A cloud-based big data sentiment analysis application for enterprises' brand monitoring in social media streams. *2015 IEEE 1st International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI)*, 186–191. https://doi.org/10.1109/RTSI.2015.7325096

Thomas, S., & Jadeja, A. (2021). Psychological antecedents of consumer trust in CRM campaigns and donation intentions: The moderating role of creativity. *Journal of Retailing and Consumer Services*, *61*, 102589. https://doi.org/10.1016/j.jretconser.2021.102589

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Tripathi, S., & Verma, S. (2018). Social media, an emerging platform for relationship building: A study of engagement with nongovernment organizations in India. *International Journal of Nonprofit and Voluntary Sector Marketing*, *23*. https://doi.org/10.1002/nvsm.1589

Walker, D., & Nowlin, E. L. (2018). Data-Driven Precision and Selectiveness in Political Campaign Fundraising. *Journal of Political Marketing*, *0*(0), 1–20. https://doi.org/10.1080/15377857.2018.1457590

Xuan Huynh, H., Nguyen, V. T., Duong-Trung, N., Pham, V.-H., & Phan, C. T. (2019). Distributed Framework for Automating Opinion Discretization From Text Corpora on Facebook. *IEEE Access*, *7*, 78675–78684. https://doi.org/10.1109/ACCESS.2019.2922427

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*, *8*(4), e1253. https://doi.org/10.1002/widm.1253

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ANNEXES

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ANNEX 1. BRANDWATCH QUERY

Nonprofit_MKT

FILTERS Language: English Location: Worldwide Content sources: Twitter Exclusions: All

Write your query

```
author:gatesfoundation
OR (gates AND foundation) OR #gatesfoundation OR @gatesfoundation OR "gates foundation" OR "Bill and Melinda Gates Foundation"

OR author:fordfoundation
OR (ford AND foundation) OR #fordfoundation OR @fordfoundation OR "ford foundation"

OR author:gettymuseum
OR (getty AND museum) OR #gettymuseum OR @gettymuseum OR "getty museum"

OR author:HHMINEWS
OR (Howard AND Hughes AND (Medical OR Institute)) OR #HHMINEWS OR @HHMINEWS OR "Howard Hughes Medical Institute"

OR author:RWJF
OR (Robert AND Wood AND Johnson AND Foundation) OR #RWJF OR @RWJF OR "Robert Wood Johnson Foundation"
```

Suggest AI-powered keywords: Off | Find location codes | Keyboard shortcuts 81905 characters left

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ANNEX 2. CODE OF THE APPROACH BASED ON CNN INTEGRATED WITH GLOVE – PYTHON

```
#Approach based on CNN integrated with GloVe
import numpy as np
import math
import re
import pandas as pd
from bs4 import BeautifulSoup
import random
import seaborn as sns
import matplotlib.pyplot as plt
import nltk
from nltk.corpus import stopwords
nltk.download('stopwords')
import tensorflow as tf
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from google.colab import drive

#1. Dataset
drive.mount('/content/drive')
cols = ['sentiment', 'id', 'user', 'text']
data = pd.read_csv('/content/drive/MyDrive/training.mentions.csv',
                   header = None,
                   names = cols,
                   engine='python',
                   encoding='latin1')

data.drop(['id', 'user'],
          axis = 1, inplace=True)

#2. Pre-processing

def clean_tweet(tweet):
  tweet = BeautifulSoup(tweet, 'lxml').get_text()
  tweet = re.sub(r"@[A-Za-z0-9]+", ' ', tweet)
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
  tweet = re.sub(r"https?://[A-Za-z0-9./]+", ' ', tweet)
  tweet = re.sub(r"[^a-zA-Z.!?']", ' ', tweet)
  tweet = re.sub(r" +", ' ', tweet)
  return tweet

test = '99 ' + data.text[0]
test
result = clean_tweet(test)
result
data_clean = [clean_tweet(tweet) for tweet in data.text]
data_labels = data.sentiment.values
#data_labels[data_labels == 4] = 1

sentences = data.text
labels = data_labels

embedding_dim = 100
max_length = 16
trunc_type='post'
padding_type='post'
oov_tok = "<OOV>"
training_size=1280000

tokenizer = Tokenizer()
tokenizer.fit_on_texts(sentences)
word_index = tokenizer.word_index
vocab_size=len(word_index)
sequences = tokenizer.texts_to_sequences(sentences)
padded = pad_sequences(sequences, maxlen=max_length, padding=padding_type,
truncating=trunc_type)

#3. Creation of the New Dataset
data_with_len = [[sent, data_labels[i], len(sent)]
                 for i, sent in enumerate(sequences)]

random.shuffle(data_with_len)
data_with_len.sort(key=lambda x: x[2])
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
sorted_all = [(sent_lab[0], sent_lab[1])
              for sent_lab in data_with_len if sent_lab[2] > 7]

all_dataset = tf.data.Dataset.from_generator(lambda: sorted_all,
                                             output_types     =     (tf.int32,
tf.int32))

next(iter(all_dataset))

BATCH_SIZE = 32
all_batched  =  all_dataset.padded_batch(BATCH_SIZE,  padded_shapes=((None,  ),
()))
next(iter(all_batched))
NB_BATCHES = len(sorted_all) // BATCH_SIZE
NB_BATCHES

NB_BATCHES_TEST = NB_BATCHES // 10
NB_BATCHES_TEST

all_batched.shuffle(NB_BATCHES)
test_dataset = all_batched.take(NB_BATCHES_TEST)
train_dataset = all_batched.skip(NB_BATCHES_TEST)

next(iter(test_dataset))

next(iter(train_dataset))

#4. Training

train_padded = padded[0:training_size]
test_padded = padded[training_size:]
train_labels = labels[0:training_size]
test_labels = labels[training_size:]

train_padded = np.array(train_padded)
train_labels = np.array(train_labels)
test_padded = np.array(test_padded)
test_labels = np.array(test_labels)
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
embeddings_index = {};
with open('/content/drive/MyDrive/Glove/glove.6B.100d.txt') as f:
    for line in f:
        values = line.split();
        word = values[0];
        coefs = np.asarray(values[1:], dtype='float32');
        embeddings_index[word] = coefs;

embeddings_matrix = np.zeros((vocab_size+1, embedding_dim));
for word, i in word_index.items():
    embedding_vector = embeddings_index.get(word);
    if embedding_vector is not None:
        embeddings_matrix[i] = embedding_vector;

model = tf.keras.Sequential([
    tf.keras.layers.Embedding(vocab_size+1,                    embedding_dim,
input_length=max_length,
                              weights=[embeddings_matrix], trainable=False),
    tf.keras.layers.Dropout(0.2),
    tf.keras.layers.Conv1D(64, 5, activation='relu'),
    tf.keras.layers.MaxPooling1D(pool_size=4),
    tf.keras.layers.LSTM(64),
    tf.keras.layers.Dense(1, activation='sigmoid')

])

model.compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy'
])
model.summary()

checkpoint_path = '/content/drive/MyDrive/Checkpoints'
ckpt = tf.train.Checkpoint(model=model)

ckpt_manager     =     tf.train.CheckpointManager(ckpt,     checkpoint_path,
max_to_keep=1)

if ckpt_manager.latest_checkpoint:
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
  ckpt.restore(ckpt_manager.latest_checkpoint)
  print('Latest checkpoint restored!')

class MyCustomCallBack(tf.keras.callbacks.Callback):
  def on_epoch_end(self, epoch, logs=None):
    ckpt_manager.save()
    print("Checkpoint savet at {}".format(checkpoint_path))

num_epochs = 5
history = model.fit(train_padded,
                    train_labels,
                    epochs=num_epochs,
                    steps_per_epoch = 100,
                    validation_data=(test_padded, test_labels),
                    verbose=2
                   )
print("Training Complete")
history.history.keys()
plt.plot(history.history['loss'])
plt.title('Loss progress');

plt.plot(history.history['accuracy'])
plt.title('Accuracy progress');

results = model.evaluate(test_dataset)
print(results)

results = model.evaluate(train_dataset)
print(results)
```

ANNEX 3. CODE OF THE BERT-INTEGRATED CNN-BASED APPROACH AS TOKENIZER - BERT

```
#1. Import of libraries
import numpy as np
import math
import re
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
import pandas as pd
from bs4 import BeautifulSoup
import random
import seaborn as sns
import matplotlib.pyplot as plt
from google.colab import drive

  # 1.1 Architecture installation
!pip install bert-for-tf2
!pip install sentencepiece #Token que será utilizado

import tensorflow_hub as hub
from tensorflow.keras import layers
import bert

#2. Dataset Loading
drive.mount('/content/drive')

cols = ['sentiment', 'id','date', 'query', 'user', 'text']
data = pd.read_csv('/content/drive/MyDrive/training.mentions.csv',
                   header = None,
                   names = cols,
                   engine='python',
                   encoding='latin1')
data.drop(['id', 'user'], #'date', 'query',
          axis = 1, inplace=True)

#2.1 Text cleaning
def clean_tweet(tweet):
  tweet = BeautifulSoup(tweet, 'lxml').get_text()
  tweet = re.sub(r"@[A-Za-z0-9]+", ' ', tweet)
  tweet = re.sub(r"https?://[A-Za-z0-9./]+", ' ', tweet)
  tweet = re.sub(r"[^a-zA-Z.!?']", ' ', tweet)
  tweet = re.sub(r" +", ' ', tweet)
  return tweet
test = '99 ' + data.text[0]
test
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
result = clean_tweet(test)
result
data_clean = [clean_tweet(tweet) for tweet in data.text]
data_labels = data.sentiment.values
data_labels

#3 Tokenizer
# Using BERT features to apply a token to the dataset
FullTokenizer = bert.bert_tokenization.FullTokenizer
bert_layer = hub.KerasLayer('https://tfhub.dev/tensorflow/bert_en_uncased_L-24_H-1024_A-16/1', trainable=False)
vocab_file = bert_layer.resolved_object.vocab_file.asset_path.numpy()
do_lower_case = bert_layer.resolved_object.do_lower_case.numpy()
tokenizer = FullTokenizer(vocab_file, do_lower_case)
vocab_file
len(tokenizer.vocab)
def encode_sentence(sent):
  return tokenizer.convert_tokens_to_ids(tokenizer.tokenize(sent))
data_inputs = [encode_sentence(sentence) for sentence in data_clean]

#3.1 Creation of the new dataset
data_with_len = [[sent, data_labels[i], len(sent)]
                 for i, sent in enumerate(data_inputs)]

random.shuffle(data_with_len)
data_with_len.sort(key=lambda x: x[2])
sorted_all = [(sent_lab[0], sent_lab[1])
              for sent_lab in data_with_len if sent_lab[2] > 7]

all_dataset = tf.data.Dataset.from_generator(lambda: sorted_all,
                                             output_types = (tf.int32, tf.int32))

next(iter(all_dataset))

BATCH_SIZE = 32
all_batched = all_dataset.padded_batch(BATCH_SIZE, padded_shapes=((None, ), ()))
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
next(iter(all_batched))
len(sorted_all)

NB_BATCHES = len(sorted_all) // BATCH_SIZE
NB_BATCHES

NB_BATCHES_TEST = NB_BATCHES // 10
NB_BATCHES_TEST

all_batched.shuffle(NB_BATCHES)
test_dataset = all_batched.take(NB_BATCHES_TEST)
train_dataset = all_batched.skip(NB_BATCHES_TEST)

next(iter(test_dataset))

next(iter(train_dataset))

#4 DCNN with BERT
class CNN(tf.keras.Model):

  def __init__(self,
               vocab_size,
               emb_dim=128,
               nb_filters = 50,
               FFN_units=512,
               nb_classes=2,
               dropout_rate=0.1,
               training=False,
               name="dcnn"):
    super(CNN, self).__init__(name=name)
    self.embedding = layers.Embedding(vocab_size, emb_dim)
    self.bigram = layers.Conv1D(filters = nb_filters,
                                kernel_size = 2,
                                padding='valid',
                                activation='relu')
    self.trigram = layers.Conv1D(filters = nb_filters,
                                kernel_size = 3,
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
                                padding='valid',
                                activation='relu')
    self.fourgram = layers.Conv1D(filters = nb_filters,
                                kernel_size = 4,
                                padding='valid',
                                activation='relu')
        self.pool = layers.GlobalMaxPool1D()

    self.dense_1 = layers.Dense(units = FFN_units, activation='relu')
    self.dropout = layers.Dropout(rate=dropout_rate)
    if nb_classes == 2:
      self.last_dense = layers.Dense(units=1, activation='sigmoid')
    else:
      self.last_dense = layers.Dense(units=nb_classes, activation='softmax')
  def call(self, inputs, training):
    x = self.embedding(inputs)
    x_1 = self.bigram(x)
    x_1 = self.pool(x_1)
    x_2 = self.trigram(x)
    x_2 = self.pool(x_2)
    x_3 = self.fourgram(x)
    x_3 = self.pool(x_3)
    merged = tf.concat([x_1, x_2, x_3], axis = -1)
    merged = self.dense_1(merged)
    merged = self.dropout(merged, training)
    output = self.last_dense(merged)
    return output

# 5. Training
VOCAB_SIZE = len(tokenizer.vocab)
EMB_DIM = 200
NB_FILTERS = 100
FFN_UNITS = 256
NB_CLASSES = 2
DROPOUT_RATE = 0.2
NB_EPOCHS = 5
Cnn = CNN(vocab_size=VOCAB_SIZE,
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
            emb_dim=EMB_DIM,
            nb_filters = NB_FILTERS,
            FFN_units = FFN_UNITS,
            nb_classes = NB_CLASSES,
            dropout_rate = DROPOUT_RATE)
if NB_CLASSES == 2:
  Cnn.compile(loss='binary_crossentropy',                    optimizer='adam',
metrics=['accuracy'])
else:
  Cnn.compile(loss='sparse_categorical_crossentropy',        optimizer='adam',
metrics=['sparse_categorical_accuracy'])


checkpoint_path = '/content/drive/MyDrive/Checkpoints'


ckpt = tf.train.Checkpoint(Cnn=Cnn)
ckpt_manager     =     tf.train.CheckpointManager(ckpt,     checkpoint_path,
max_to_keep=1)


if ckpt_manager.latest_checkpoint:
  ckpt.restore(ckpt_manager.latest_checkpoint)
  print('Latest checkpoint restored!')
class MyCustomCallBack(tf.keras.callbacks.Callback):
  def on_epoch_end(self, epoch, logs=None):  #definição padrão
    ckpt_manager.save()                      #código personalizado
    print("Checkpoint savet at {}".format(checkpoint_path))


history = Cnn.fit(train_dataset,
                  epochs=NB_EPOCHS,
                  steps_per_epoch = 100,
                  callbacks=[MyCustomCallBack()])


# 6.0 Model evaluation
history.history.keys()
plt.plot(history.history['loss'])
plt.title('Loss progress');


plt.plot(history.history['accuracy'])
plt.title('Accuracy progress');
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
results = Cnn.evaluate(test_dataset)
print(results)

results = Cnn.evaluate(train_dataset)
print(results)
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ANNEX 4. CODE OF THE BERT-BASED APPROACH AS AN EMBEDDING LAYER – BERT-CNN

```
import numpy as np
import math
import re
import pandas as pd
from bs4 import BeautifulSoup
import random
import matplotlib.pyplot as plt

from google.colab import drive

!pip install bert-for-tf2
!pip install sentencepiece

import tensorflow as tf
import tensorflow_hub as hub
from tensorflow.keras import layers
import bert

drive.mount("/content/drive")

cols = ["sentiment", "id", "date", "query", "user", "text"]
data = pd.read_csv(
    "/ content/drive/MyDrive/training.mentions.csv ",
    header=None,
    names=cols,
    engine="python",
    encoding="latin1"
)

data.drop(["id", "date", "query", "user"],
          axis=1,
          inplace=True)

def clean_tweet(tweet):
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
    tweet = BeautifulSoup(tweet, "lxml").get_text()
    tweet = re.sub(r"@[A-Za-z0-9]+", ' ', tweet)
    tweet = re.sub(r"https?://[A-Za-z0-9./]+", ' ', tweet)
    tweet = re.sub(r"[^a-zA-Z.!?']", ' ', tweet)
    tweet = re.sub(r" +", ' ', tweet)
    return tweet
data_clean = [clean_tweet(tweet) for tweet in data.text]
data_labels = data.sentiment.values
FullTokenizer = bert.bert_tokenization.FullTokenizer
bert_layer  =  hub.KerasLayer("https://tfhub.dev/tensorflow/bert_en_uncased_L-
24_H-1024_A-16/1",
                              trainable=False)
vocab_file = bert_layer.resolved_object.vocab_file.asset_path.numpy()
do_lower_case = bert_layer.resolved_object.do_lower_case.numpy()
tokenizer = FullTokenizer(vocab_file, do_lower_case)

def encode_sentence(sent):
  return ["[CLS]"] + tokenizer.tokenize(sent) + ["[SEP]"]

data_inputs = [encode_sentence(sentence) for sentence in data_clean]

def get_ids(tokens):
  return tokenizer.convert_tokens_to_ids(tokens)

np.char.not_equal("[PAD]", "[PAD]")

def get_mask(tokens):
  return np.char.not_equal(tokens, "[PAD]").astype(int)

def get_segments(tokens):
  seg_ids = []
  current_seg_id = 0
  for tok in tokens:
    seg_ids.append(current_seg_id)
    if tok == "[SEP]":
      current_seg_id = 1 - current_seg_id
  return seg_ids
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
my_sent = ["[CLS]"] + tokenizer.tokenize("She goes for a walk.") + ["[SEP]"]
my_sent

bert_layer([
            tf.expand_dims(tf.cast(get_ids(my_sent), tf.int32), 0),
            tf.expand_dims(tf.cast(get_mask(my_sent), tf.int32), 0),
            tf.expand_dims(tf.cast(get_segments(my_sent), tf.int32), 0)
           ])

data_with_len = [[sent, data_labels[i], len(sent)]
                 for i, sent in enumerate(data_inputs)]
random.shuffle(data_with_len)
data_with_len.sort(key = lambda x: x[2])
sorted_all = [([get_ids(sent_lab[0]),
                get_mask(sent_lab[0]),
                get_segments(sent_lab[0])],
               sent_lab[1])
               for sent_lab in data_with_len if sent_lab[2] > 7]

all_dataset = tf.data.Dataset.from_generator(lambda: sorted_all,
                                             output_types=(tf.int32,
tf.int32))

BATCH_SIZE = 32
all_batched = all_dataset.padded_batch(BATCH_SIZE,
                                       padded_shapes=((3, None), ()),
                                       padding_values=(0, 0))

NB_BATCHES = len(sorted_all) // BATCH_SIZE
NB_BATCHES_TEST = NB_BATCHES // 10
all_batched.shuffle(NB_BATCHES)
test_dataset = all_batched.take(NB_BATCHES_TEST)
train_dataset = all_batched.skip(NB_BATCHES_TEST)

class CNNBERTEmbedding(tf.keras.Model):
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
    def __init__(self,
                 nb_filters=50,
                 FFN_units=512,
                 nb_classes=2,
                 dropout_rate=0.1,
                 name="dcnn"):
        super(CNNBERTEmbedding, self).__init__(name=name)

        self.bert_layer                                                    =
hub.KerasLayer("https://tfhub.dev/tensorflow/bert_en_uncased_L-24_H-1024_A-
16/1",
                                        trainable = False)

        self.bigram = layers.Conv1D(filters=nb_filters,
                                    kernel_size=2,
                                    padding="valid",
                                    activation="relu")
        self.trigram = layers.Conv1D(filters=nb_filters,
                                     kernel_size=3,
                                     padding="valid",
                                     activation="relu")
        self.fourgram = layers.Conv1D(filters=nb_filters,
                                      kernel_size=4,
                                      padding="valid",
                                      activation="relu")
        self.pool = layers.GlobalMaxPool1D()
        self.dense_1 = layers.Dense(units=FFN_units, activation="relu")
        self.dropout = layers.Dropout(rate=dropout_rate)
        if nb_classes == 2:
            self.last_dense = layers.Dense(units=1,
                                           activation="sigmoid")
        else:
            self.last_dense = layers.Dense(units=nb_classes,
                                           activation="softmax")

    def embed_with_bert(self, all_tokens):
      _, embs = self.bert_layer([all_tokens[:, 0, :],
                                 all_tokens[:, 1, :],
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
                                    all_tokens[:, 2, :]])
        return embs

    def call(self, inputs, training):
        x = self.embed_with_bert(inputs)

        x_1 = self.bigram(x)
        x_1 = self.pool(x_1)
        x_2 = self.trigram(x)
        x_2 = self.pool(x_2)
        x_3 = self.fourgram(x)
        x_3 = self.pool(x_3)

        merged = tf.concat([x_1, x_2, x_3], axis=-1)
        merged = self.dense_1(merged)
        merged = self.dropout(merged, training)
        output = self.last_dense(merged)

        return output

NB_FILTERS = 100
FFN_UNITS = 256
NB_CLASSES = 2
DROPOUT_RATE = 0.2
BATCH_SIZE = 32
NB_EPOCHS = 5

Dcnn = CNNBERTEmbedding(nb_filters=NB_FILTERS,
                        FFN_units=FFN_UNITS,
                        nb_classes=NB_CLASSES,
                        dropout_rate=DROPOUT_RATE)

if NB_CLASSES == 2:
    Dcnn.compile(loss="binary_crossentropy",
                 optimizer="adam",
                 metrics=["accuracy"])
else:
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
    Dcnn.compile(loss="sparse_categorical_crossentropy",
                 optimizer="adam",
                 metrics=["sparse_categorical_accuracy"])

checkpoint_path = "/content/drive/MyDrive/Checkpoints"

ckpt = tf.train.Checkpoint(Dcnn=Dcnn)

ckpt_manager = tf.train.CheckpointManager(ckpt, checkpoint_path,
max_to_keep=1)

if ckpt_manager.latest_checkpoint:
    ckpt.restore(ckpt_manager.latest_checkpoint)
    print("Latest checkpoint restored!!")

class MyCustomCallback(tf.keras.callbacks.Callback):

    def on_epoch_end(self, epoch, logs=None):
        ckpt_manager.save()
        print("Checkpoint saved at {}.".format(checkpoint_path))

history = Dcnn.fit(train_dataset,
                   epochs=NB_EPOCHS,
                   steps_per_epoch = 100,
                   callbacks=[MyCustomCallback()])

history.history.keys()

plt.plot(history.history['loss'])
plt.title('Loss progress');

plt.plot(history.history['accuracy'])
plt.title('Accuracy progress');

# Analyse test results
results = Dcnn.evaluate(test_dataset)
print(results)
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

```
results = Dcnn.evaluate(train_dataset)
print(results)
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ANNEX 5. PAPER PUBLISHED IN CISTI 2021

Estratégias de tomada de decisão orientada por dados aplicadas ao marketing

Data-driven decision making strategies applied to marketing

Marcus Borges
Instituto Politécnico de Coimbra,
Coimbra Business School, ISCAC,
Coimbra, Portugal
a2020106438@alumni.iscac.pt

Jorge Bernardino
Instituto Politécnico de Coimbra,
Instituto de Investigação Aplicada,
Coimbra, Portugal
bernardino@ipc.pt

Isabel Pedrosa
Instituto Politécnico de Coimbra,
Coimbra Business School, ISCAC,
Coimbra ISTAR-IUL, Lisboa, Portugal
ipedrosa@iscac.pt

***Resumo* — A tomada de decisão orientada por dados torna-se cada vez mais um componente essencial da prática do marketing em todos os setores de negócio e recebe atenção sem precedentes em termos de políticas e apoio financeiro. O objetivo deste artigo é melhorar a compreensão das estratégias de tomada de decisão orientada por dados, com ênfase numa visão abrangente sobre aplicações no marketing. Deste modo, pretendemos estabelecer uma perspetiva sobre o planeamento e gestão estratégica de marketing baseado em dados, complementando o entendimento através da literatura atual. Ao adotar a perspetiva de "data-driven", os profissionais de marketing se beneficiam sobre uma abordagem de mais e melhor controle sobre os dados e um conhecimento mais amplo dos dados e dos modelos estratégicos da organização para poder obter decisões relevantes.**

Palavras Chave* – *marketing; estratégia: tomada de decisão; orientado por dados.

***Abstract* — Data-driven decision making is increasingly becoming an essential component of marketing practice across all business sectors and receives unprecedented attention in terms of policies and financial support. The purpose of this article is to improve our understanding of data-driven decision-making strategies and focus on a comprehensive view of applications in marketing and related research. As such, we intend to establish a perspective on strategic planning and management of data-based marketing, complementing the understanding through the current literature. By adopting the "data-driven" perspective, marketers benefit from an approach of more and better control over data and a broader knowledge of the organization's data and strategic models in order to obtain relevant decisions.**

Keywords* – *marketing; strategy; decision making; data-driven.

I. INTRODUÇÃO

A expansão tecnológica ocorrida nos últimos anos cria oportunidades de vantagem competitiva ao aplicar novas abordagens orientadas a dados às práticas de marketing. Empresas grandes e pequenas de diversos setores podem aproveitar a enorme quantidade de dados gerados, para desenvolver novos modelos de negócios, abrir novas fontes de receita e iniciar inovações disruptivas [1]. A recolha de dados de diferentes fontes, como a Internet das Coisas (IoT), media sociais e mecanismos de pesquisa, criaram oportunidades significativas para que as organizações tenham uma visão analítica no desenvolvimento de abordagens de marketing programático [2]. Vivemos num mundo de rápidos avanços tecnológicos no âmbito digital e abundância de dados gerados no qual as consequências para as práticas de marketing foram transformadoras [3].

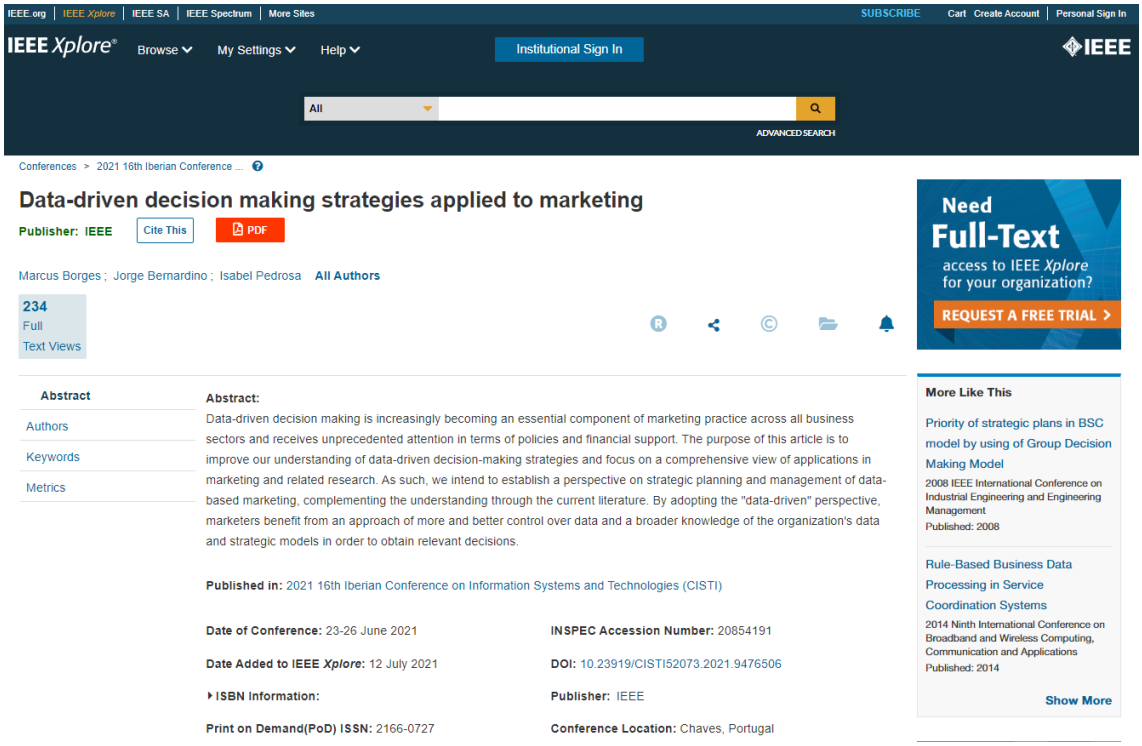
Este trabalho é baseado numa revisão da literatura para investigar como a adoção de novas tecnologias e ferramentas de análise de dados ajudaram a transformar o âmbito do marketing nas empresas e assim elaborar uma revisão atualizada dos trabalhos existentes sobre *Data-driven Marketing*. Nesse sentido, pensamos que é muito útil analisar o estado da arte do tema, fornecendo uma estrutura e orientações claras para direcionar os gestores e identificar futuras direções de investigação. Que seja do nosso conhecimento, nenhum estudo até ao momento sintetizou através de uma revisão da literatura o conhecimento atual sobre o planeamento e gestão estratégica de marketing baseado em dados, o que seria muito relevante para investigadores e para os gestores, uma vez que *Data-driven Marketing* não é simplesmente uma variação do marketing tradicional.

Para tanto, será realizada uma revisão sistemática da literatura relevante sobre *Data-driven Marketing*, orientada por três questões principais de pesquisa:

- RQ1: Quais são as principais definições de *data-driven marketing* (e conceitos relacionados)?
- RQ2: Quais são as principais implicações da tomada de decisão de forma analítica em *data-driven marketing*?
- RQ3: Quais são os fatores que auxiliam o planeamento estratégico de forma a obter vantagem competitiva utilizando *data-driven marketing*?

O foco deste trabalho será em artigos que tratam de *Data-driven Marketing* de produtos e serviços, publicados em revistas científicas com revisão por pares em inglês, entre 2016 e 2020, inclusive.

https://ieeexplore.ieee.org/abstract/document/9476506



IEEE.org | IEEE Xplore | IEEE SA | IEEE Spectrum | More Sites SUBSCRIBE Cart Create Account | Personal Sign In

IEEE Xplore® Browse My Settings Help Institutional Sign In IEEE

All

ADVANCED SEARCH

Conferences > 2021 16th Iberian Conference ...

Data-driven decision making strategies applied to marketing

Publisher: IEEE Cite This PDF

Marcus Borges ; Jorge Bernardino ; Isabel Pedrosa All Authors

234 Full Text Views

Abstract
Authors
Keywords
Metrics

Abstract:
Data-driven decision making is increasingly becoming an essential component of marketing practice across all business sectors and receives unprecedented attention in terms of policies and financial support. The purpose of this article is to improve our understanding of data-driven decision-making strategies and focus on a comprehensive view of applications in marketing and related research. As such, we intend to establish a perspective on strategic planning and management of data-based marketing, complementing the understanding through the current literature. By adopting the "data-driven" perspective, marketers benefit from an approach of more and better control over data and a broader knowledge of the organization's data and strategic models in order to obtain relevant decisions.

Published in: 2021 16th Iberian Conference on Information Systems and Technologies (CISTI)

Date of Conference: 23-26 June 2021
INSPEC Accession Number: 20854191
Date Added to IEEE *Xplore*: 12 July 2021
DOI: 10.23919/CISTI52073.2021.9476506
ISBN Information:
Publisher: IEEE
Print on Demand(PoD) ISSN: 2166-0727
Conference Location: Chaves, Portugal

Need Full-Text access to IEEE *Xplore* for your organization?
REQUEST A FREE TRIAL >

More Like This

Priority of strategic plans in BSC model by using of Group Decision Making Model
2008 IEEE International Conference on Industrial Engineering and Engineering Management
Published: 2008

Rule-Based Business Data Processing in Service Coordination Systems
2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications
Published: 2014

Show More

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt