

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Detecção de sarcasmo em redes sociais
utilizando DeBERTa**

Lucas Paiolla Forastiere

MONOGRAFIA FINAL

MAC 499 — TRABALHO DE
FORMATURA SUPERVISIONADO

Supervisor: Prof. Dr. Ricardo Marcondes Marcacini

São Paulo
2017

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0
(Creative Commons Attribution 4.0 International License)*

Resumo

Lucas Paiolla Forastiere. **Deteção de sarcasmo em redes sociais utilizando DeBERTa**. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2017.

[illegible]

Palavras-chave: Palavra-chave1. Palavra-chave2. Palavra-chave3.

Abstract

Lucas Paiolla Forastiere. **Sarcasm detection in social media using decoding-enhanced BERT with disentangled attention**. Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2017.

[illegible]

Keywords: Keyword1. Keyword2. Keyword3.

Lista de Abreviaturas

CFT	Transformada contínua de Fourier (<i>Continuous Fourier Transform</i>)
DFT	Transformada discreta de Fourier (<i>Discrete Fourier Transform</i>)
EIIP	Potencial de interação elétron-íon (<i>Electron-Ion Interaction Potentials</i>)
STFT	Transformada de Fourier de tempo reduzido (<i>Short-Time Fourier Transform</i>)
ABNT	Associação Brasileira de Normas Técnicas
URL	Localizador Uniforme de Recursos (<i>Uniform Resource Locator</i>)
IME	Instituto de Matemática e Estatística
USP	Universidade de São Paulo

Lista de Símbolos

ω	Frequência angular
ψ	Função de análise <i>wavelet</i>
Ψ	Transformada de Fourier de ψ

Lista de Figuras

Lista de Tabelas

Lista de Programas

Sumário

1	Introduction	1
----------	---------------------	----------

Capítulo 1

Introduction

Sarcasm detection is an important aspect of many natural language processing (NLP) systems, with many implications in natural language understanding, dialog systems, and data mining. However, sarcasm detection is difficult because it is infrequent in many conversations and, many times, it is difficult even for humans to discern.

Many studies have been made in the area and many datasets have been proposed with either *balanced* or *unbalanced* data. Also many of these datasets use humans to annotate sarcastic statements.

In this paper, we use the Self-Annotated Reddit Corpus (SARC), which is a large corpus for sarcasm detection created using Reddit posts to get labels automatically, to train, evaluate, and compare many NLP models.

