

# A fake news detection and credibility ranking platform for Portuguese online news

Inês Rito Lima<sup>1</sup>, Márcia Pinto<sup>2</sup>, Ivone Amorim<sup>3</sup>, Goreti Marreiros<sup>2</sup>, and Alexandre Ulisses<sup>1</sup>

<sup>1</sup> MOG Technologies, Department of Innovation, Maia, Portugal

<sup>2</sup> ISEP Polytechnic Institute of Porto, Informatics Engineering Department, Portugal

<sup>3</sup> PORTIC Porto Research, Technology Innovation Center, Polytechnic of Porto, Portugal  
{ines.lima,alexandre.ulisses}@mog-technologies.com,  
{1130397,mgt}@isep.ipp.pt, ivone.amorim@portic.ipp.pt

**Abstract.** The growth of social media has enabled the spread of tendentiously fake news content in a disorganized and fast manner. Despite the extensive research on fake news detection methods, algorithms and applications [1], most of the studies focused on Natural Language Processing (NLP) techniques for English content analysis. Assessing online news in Portuguese language and ranking them based on their credibility is still an open issue. Addressing this gap, the present work proposes a platform for European Portuguese online news credibility classification and ranking based on content and publication metadata. A dataset was built from the collection of online news from 36 sources, comprising content-related data (title, lead, body) and supporting metadata (source, category, year, author). A hierarchical architecture was envisioned and implemented for supervised binary classification probabilities extraction aggregated into multi-classes, resorting to three Machine Learning models: i) focused on textual analysis resorting to several NLP techniques; ii) delving into the news publication metadata; and iii) leveraging the balance between the two prior models' probability predictions. This architecture minimizes data entropy, maximizing information retrieved, providing two credibility indexes and a fake news detector with 92% test accuracy and 95% deployed proof-of-concept accuracy. These models were integrated into a web service, enabling users to evaluate and rank online news resorting to the implemented architecture. This solution has several application areas, fostering society's literacy and increased critical thinking, especially amongst younger generations.

**Keywords:** fake news detection; Portuguese language; natural language processing; machine learning models; misinformation on social media