Accepted Manuscript

Classifying online job advertisements through machine learning

Roberto Boselli, Mirko Cesarini, Fabio Mercorio, Mario Mezzanzanica

PII: S0167-739X(17)32183-0
DOI: https://doi.org/10.1016/j.future.2018.03.035
Reference: FUTURE 4047

To appear in: *Future Generation Computer Systems*

Received date : 27 September 2017
Revised date : 21 February 2018
Accepted date : 17 March 2018

Please cite this article as: R. Boselli, M. Cesarini, F. Mercorio, M. Mezzanzanica, Classifying online job advertisements through machine learning, *Future Generation Computer Systems* (2018), https://doi.org/10.1016/j.future.2018.03.035

Classifying Online Job Advertisements through Machine Learning

Roberto Boselli[a,b], Mirko Cesarini[a,b], Fabio Mercorio[a,b], Mario Mezzanzanica[a,b]

[a] *Department of Statistics and Quantitative Methods, University of Milan-Bicocca, Italy*
[b] *CRISP Research Centre, University of Milan-Bicocca, Italy*

Abstract

The rapid growth of Web usage for advertising job positions provides a great opportunity for real-time labour market monitoring. This is the aim of Labour Market Intelligence (LMI), a field that is becoming increasingly relevant to EU Labour Market policies design and evaluation. The analysis of Web job vacancies, indeed, represents a competitive advantage to labour market stakeholders with respect to classical survey-based analyses, as it allows for reducing the time-to-market of the analysis by moving towards a fact-based decision making model. In this paper, we present our approach for automatically classifying million Web job vacancies on a standard taxonomy of occupations. We show how this problem has been expressed in terms of text classification via machine learning. We also show how our approach has been applied to certain real-life projects and we discuss the benefits provided to end users.

Keywords: Machine Learning, Text Classification, Big Data, NLP

1. Introduction

In the last few years, the diffusion of web-centric services is growing exponentially, and this allows a significant part of the European Labour demand to be communicated through specialised web portals and services. This has also led to the introduction of the term "Labour Market Intelligence" (LMI), which refers to the use and design of AI algorithms and frameworks for Labour Market Data to support decision-making.

Motivating Example. In the on-line job market, a *job vacancy* is a document containing two main text fields: a *title* and a *full description*. The title shortly