



Libraries and Archives

POLITesi - Digital archive of degree and doctoral theses

[Home](#) POLIthesis / POLIthesis / Master's Thesis

Scam detection in online housing offers: model ensemble against dataset drifting

[BELLONI, MASSIMO](#)

2017/2018

Abstract

This thesis, developed during an internship at HousingAnywhere, an online accommodation platform based in Rotterdam (Netherlands), describes a six month project related to the automatic detection of scam listings in the online housing domain. Starting point of the analysis is a dataset composed of tens of thousands of advertisers' housing offers and publication behaviors from the past three years that resulted in training and testing state-of-the-art machine learning classifiers with good cross-validation results. A severe drop in performances when testing the models on newly incoming data has resulted in the discovery of how the evolution of a platform over the years forces users to adapt to it, changing their behaviors and the derived features. In the direction of trying to extract the full potential from the dataset, being able to leverage both the knowledge obtainable from old historical behaviors and the information coming from the newly created triggers, an ensemble architecture composed by five different models trained with different features and datasets collected over different time periods has been designed, trained, tested and deployed, resulting in a data product currently used in daily operations by HousingAnywhere employees.

Short sheet **Full sheet**

Speaker

[DANIEL, FLORIAN](#)

Correlator(s)


[VALENTINI, GIANLUCA](#)

School / Dept.*ING - School of Industrial and Information Engineering***Date***16-Apr-2019***Academic year***2017/2018***Abstract in Italian**

This thesis, developed during an internship at HousingAnywhere, an online platform for medium-long term rentals based in Rotterdam (The Netherlands), describes a six-month project on the automatic detection of fraudulent listings in the real estate sector. The starting point of the analysis is a dataset composed of tens of thousands of real estate offers and advertisers' behaviors, collected over the last three years and used for training and testing Machine Learning algorithms with good results in cross-validation. A significant drop in performance when the model was tested with data currently generated by the platform led to the discovery of how the evolution of the interface over the years forces users to adapt to it, changing their behavior and irremediably the data it generates. Moving in the direction of fully exploiting the potential of the dataset, trying to extract as much knowledge as possible from both historical behaviors (and still valid today) and from information obtainable from new systems and notifications, an ensemble architecture was developed, composed of five different models. Each model is trained using subsets of the starting dataset, both in terms of the features used and the size of the dataset itself compared to the time interval considered. The final model has been tested and released, resulting in a software product currently used in daily operations by some HousingAnywhere employees.

Document type*Master's Thesis***It appears in the typologies:**[Master's Thesis](#)

Attached files

File	Size	Format	
Belloni - MSc Thesis.pdf Only authorized users from 03/12/2022	2.39 MB	Adobe PDF	 View/Open
Description: Thesis text			

Documents in POLITesi are protected by copyright and all rights are reserved, unless otherwise indicated.

 ----- Information -----

Please use this identifier to cite or link to this document: <https://hdl.handle.net/10589/147397>

Powered by UNITESI - [about UNITESI](#) - [Use of cookies](#) **CINECA** Copyright © 2025