

# Automatic processing and classification of citizens' reports

**Giulio Angiani**

Department of Engineering and Architecture  
University of Parma, Italy  
giulio.angiani@unipr.it

**Paolo Fornacciari**

Department of Engineering and Architecture  
University of Parma, Italy  
paolo.fornacciari@unipr.it

**Gianfranco Lombardo**

Department of Engineering and Architecture  
University of Parma, Italy  
gianfranco.lombardo@unipr.it

**Monica Mordonini**

Department of Engineering and Architecture  
University of Parma, Italy  
monica.mordonini@unipr.it

**Umberto Pietroni**

Department of Engineering and Architecture  
University of Parma, Italy  
umberto.pietroni@studenti.unipr.it

**Michele Tomaiuolo**

Department of Engineering and Architecture  
University of Parma, Italy  
michele.tomaiuolo@unipr.it

## ABSTRACT

We present a comprehensive framework for managing reports sent to the local government by citizens through well-known instant messaging apps. It leverages a combined use of Web systems and automated bots, based on Machine Learning techniques. This project has been developed in collaboration with the local administration of Montecchio Emilia (Italy). The results show that an automatic classification system of this kind can reach very good levels of accuracy, also above 90%.

## KEYWORDS

Text analysis, Image classification, Government 2.0, Chatbot.

## ACM Reference Format:

Giulio Angiani, Paolo Fornacciari, Gianfranco Lombardo, Monica Mordonini, Umberto Pietroni, and Michele Tomaiuolo. 2018. Automatic processing and classification of citizens' reports. In *International Conference on Smart Objects and Technologies for Social Good (Goodtechs '18)*, November 28–30, 2018, Bologna, Italy. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3284869.3284930>

## 1 INTRODUCTION AND LITERATURE REVIEW

The widespread diffusion of mobile devices in recent years has allowed citizens to access and send information at any time of the day. This kind of interactions could be interesting for public administrations, in order to get reports on different

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*Goodtechs '18, November 28–30, 2018, Bologna, Italy*

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6581-9/18/11.

<https://doi.org/10.1145/3284869.3284930>

public services [3]. Recently, several e-Government services have been introduced by government and administrations in the form of conversational chatbots (e.g [7]), but often without a NLP processing of contents, exploiting the state of art in Social Media Analysis, to discover knowledge about users' sentiment [5], [6] and topic detection [1]. At the same time, image analysis and object-recognition could be useful for understanding what a user sends to the municipality. In our project we propose a system based on free and widely used apps (Whatsapp, Telegram and Messenger) to send information to local public administration, including both texts and images. Reports are also automatically categorized in one of four predefined classes (environment, lighting, maintenance and security), by a classification system based on text-analysis, image-recognition and object-detection.

## 2 METHODOLOGY

The overall architecture of the system is shown in Figure 1(a). It is realized over ActoDES, which is a software framework which adopts the actor model for the development of complex distributed systems [4].

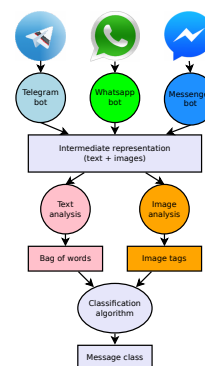


Figure 1: Data flow representation of system.

The first layer includes specific bots related to the different messaging systems and the appropriate components to make the representations of the various messages homogeneous in terms of content format, user data, geolocation, data and time. Reports are then managed in a Web-based system, which we have developed for playing the role of an ad-hoc Customer Relationship Management (CRM) system.

### Dataset collection

For realizing the system, about 800 instances containing both text and images have been downloaded from the institutional websites of several Italian municipalities. In order to have a balanced training-set we used 200 instances for each selected category: Environment, Lighting, Maintenance, Security, corresponding to the administrative offices that will manage the various kinds of citizens' reports.

### Text analysis

For the text analysis branch we have build a four-output classifier. The message is pre-processed and vectorized according the Bag of Words approach. As proposed in [2], the Multinomial Naive-Bayes classification algorithm is chosen for the analysis of natural language. However, we also compare it with other well known automatic classification algorithms, namely: Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) with K=1.

### Image classification

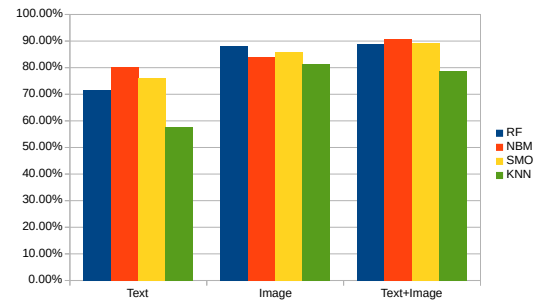
To improve the correct classification of reports, an and image-recognition step is carried out using Clarifai<sup>1</sup>. For each image associated with a message, Clarifai returns a list of entities and concepts associated to the image with a confidence value. After analyzing these responses, we build a sequence of words using only the entities with a confidence value greater than 0.9. This sequence of entities is analyzed in the same way as the bag of words obtained from the text.

## 3 RESULTS

We have performed a comparison among three cases: (i) Classification through text analysis only (using the bag of words approach), (ii) Classification through image analysis only (using the image entities), (iii) Classification through both text and image analyses (concatenating their lists of features). For this evaluations we have adopted the well-known ten-fold Cross Validation technique.

In Figure 2, it can be observed that all algorithms perform better on image entities than on text. Moreover, all algorithms improve their results using all available features, with the exception of KNN, which is known to work better on a limited set of features.

<sup>1</sup><https://clarifai.com/>



**Figure 2: Accuracy of classifiers, using different features and different algorithms.**

Overall, the best classification accuracy is obtained by the NBM algorithm, using the features of both text and images. In this case, the classification is correct in over 90% of cases. However, using the RF algorithm, a very close value of accuracy can be obtained even using only the image features.

## 4 CONCLUSIONS AND FUTURE WORKS

The accuracy of the final classification can achieve very good results, also above 90%, using features from both text and associated images. However, results also show that using appropriate machine learning algorithms, it is possible to collect citizens' report in a very simplified way, receiving just geolocalized images, which can be classified automatically in most cases.

## REFERENCES

- [1] Charu C Aggarwal and ChengXiang Zhai. 2012. *Mining text data*. Springer Science & Business Media.
- [2] Giulio Angiani, Laura Ferrari, Tomaso Fontanini, Paolo Fornacciari, Eleonora Iotti, Federico Magliani, and Stefano Manicardi. 2016. A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter.. In *KDWeb*.
- [3] Maged N Kamel Boulos et al. 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. *International journal of health geographics* 10, 1 (2011), 67.
- [4] Paolo Fornacciari, Monica Mordonini, Agostino Poggi, Laura Sani, and Michele Tomaiuolo. 2018. A holistic system for troll detection on Twitter. *Computers in Human Behavior* 89 (2018), 258–268. <https://doi.org/10.1016/j.chb.2018.08.008>
- [5] Paolo Fornacciari, Monica Mordonini, and Michele Tomaiuolo. 2015. Social Network and Sentiment Analysis on Twitter: Towards a Combined Approach.. In *KDWeb*. 53–64.
- [6] Gianfranco Lombardo, Paolo Fornacciari, Monica Mordonini, Laura Sani, and Michele Tomaiuolo. 2018. A combined approach for the analysis of support groups on Facebook - The case of patients of hidradenitis suppurativa. *Multimedia Tools and Applications* (2018), 1–19. <https://doi.org/10.1007/s11042-018-6512-5>
- [7] Andreas Lommatzsch. 2018. A Next Generation Chatbot-Framework for the Public Administration. In *Innovations for Community Services*. Springer International Publishing, Cham, 127–141.