Cyprus University of Technology
Faculty of Communication and Media Studies

Bachelor's Thesis

"IDENTIFYING SOCIAL ACTORS ON TWITTER THROUGH PROFILE AUGMENTATION AND MACHINE LEARNING CLASSIFICATION"

Christos Christodoulou

Limassol 2021

CYPRUS UNIVERSITY OF TECHNOLOGY

FACULTY OF COMMUNICATION AND MEDIA STUDIES

DEPARTMENT COMMUNICATION AND INTERNET STUDIES

Bachelor's Thesis

"IDENTIFYING SOCIAL ACTORS ON TWITTER THROUGH PROFILE AUGMENTATION AND MACHINE LEARNING CLASSIFICATION"

Christos Christodoulou

Supervisor

Dr. Constantinos Djouvas

Limassol, June 2021

Copyrights

Copyright© Christos Christodoulou, 2021

All rights reserved.

The approval of the thesis by the Department of Communication and Internet Studies of Cyprus University of Technology does not imply necessarily the approval by the Department of the views of the writer.

“No one who achieves success does so without acknowledging the help of others. The wise and confident acknowledge this help with gratitude”

Alfred North Whitehead.

Now, finishing my bachelor's degree, I try to look back and gather all my experiences as a student. The joy and emotion for what I have experienced and what the Cyprus University of Technology has offered me are inexplicable. I remain assured and positive, that what my experience has taught me is to continue to wholeheartedly engage with every opportunity I have to gain and share knowledge.

Firstly, I’d like to express my sincere gratitude to my patient and supportive supervisor, Dr. Constantinos Djouvas, who has supported me throughout my undergraduate studies and my thesis project. I am incredibly grateful for the guidance provided to me, the friendly chats, and the long uninterrupted tireless conversations to complete this project. Now, finalizing this project, I can confidently say that without his help, beneficial ideas, and all the advice provided to me, such a project could not be completed. I also owe to express my gratitude to my professor Dr. Vicky Triga, who was always there to help me with whatever I needed for my thesis, and the knowledge transmitted to me throughout my studies. Closing, the most overwhelming key to my success throughout my studies is the positive involvement of my parents Joseph and Andri, and I would like to thank them for their continued support throughout my life.

ABSTRACT

The Web and the Online Social Network platforms (OSNs) brought a new era in politics. Introduced by Barack Obama's election campaign in 2008, we observe a rapidly increasing election campaigns to take place on OSNs. Citizens, Academics, Journalists, Influencers, and many more actors, on the other hand, talk about politics on the online media public sphere. This interaction creates enormous and extremely valuable data to be collected and analyzed. Recent advances in Natural Language Processing, allowed the deep contextual understanding of these data, sheeting light on the question of "what" is discussed on OSNs. However, there is huge gap in the literature, and consequently on the available techniques, on approaches that would allow someone to answer the question of "who", i.e., different groups of users, is talking on OSNs.

The aim of this research is to break new ground in public opinion interpretation using Big Data, by answering the question of "who" is discussing on OSNs and Twitter in particular. This will be achieved by proposing and developing an automated process that classifies Twitter's users into different social actors. The methodology adopted derives from the fields of Machine Learning and Natural Language processing (for tweets classification) accompanied with political communication theories (for identifying categories of social actors). Through the combination of these disciplines, this study proposes a theoretically sound approach, unlike most existing related work in the literature, where classification categories were defined in a random manner. Moreover, the proposed methodology, in contrast to other approaches in the literature, does not rely on the tweet's content for the classification process; this can introduce bias in cases where tweets will undergo other types of analyses, e.g., sentiment analysis. Given the above, this work proposes a novel and sound approach for classifying tweets into different social actor classes.

Keywords: Machine Learning, Natural Language Processing, Text Classification, Crowdsourcing, Social Actors Classification