**POLITECNICO DI MILANO**

Department of Electronics, Informatics and Bioengineering

M.Sc. programme in Computer Science and Engineering



Automated Analysis of Social Data using

Machine Learning Techniques

Advisors:

Prof. Letizia TANCA

Prof Maristella MATERA

Ing. Ricardo MEDANA

Master Thesis of:

Petar KORDA

Pavle VIDANOVIC

Academic Year 2017-2018

*We would like to express our sincere appreciation to our supervisors and our*

*families for all their support. We would also like to extend a special thank you to*

*Prof.ssa Maristella Matera for her time, guidance and understanding.*

**Abstract**

In today’s society everything is happening on the internet, in particular on social networks. Social networks play central role in everyday life of average person. So naturally, companies recognize the opportunity and try to make use of that by changing their business plans and focus to potential customers on social networks. Business is realized by company’s presence on web and producing a content that will take customer’s attention. In return users share their opinion about particular products by leaving comments on them and reacting on company’s posts.

Taking that into consideration, it is useful to have an automated way to check user reactions on products that company is offering. Also knowing types of people fallowing and leaving opinion on products can be turned into advantage for creating future business plans. For example, to predict which products can be attractive for specific user groups or to determine best time when to lunch products. We recognized the potential of that and that’s why we were eager to examine sentiment analysis tools and machine learning algorithms to achieve that goal.

In this thesis we have built automated for calculating sentiment of users who commented on specific company’s post along with intelligent spam filter. Sentiment analysis was done separately on text and on the emojis. For the evaluating text sentiment, we used open source API and for emojis we used table of evaluation for each emoji. Spam filter was designed using supervised machine learning techniques to determine spam, not just by searching URL patterns in comments, but also to determine the spam by checking text content. We have also built clustering module which uses unsupervised learning techniques on user data and visualizing characteristics for each discovered group. Finally, previous models together were used in complete API to evaluate success of company’s posts.