

Dissertation press release**13.9.2019**

Machine Learning Methods for Classification of Unstructured Data applied to Malware Detection and Soft Skill Mining

Title of the dissertation Machine Learning Methods for Classification of Unstructured Data

Contents of the dissertation Machine learning has been recently one of the rapidly developing fields within artificial intelligence and computer science in general. Digitalization and the usage of the Internet have resulted in large amounts of digital data that people want to analyze. Free-form text is an example of unstructured data where no clear structure is defined. Being able to make some analysis of unstructured data can pose challenges. For example, we might be interested in extracting basic information about a candidate from a resume that often does not have a predefined structure. At the same time, Android applications that consist of a set of files also belong to unstructured data since every application consists of the various number of files and no clear structure is defined beforehand.

In this dissertation, we cover two application fields: Android malware classification and soft skill mining. Increased risk by mobile malware motivates researchers to develop robust and accurate methods to detect malicious applications. Android malware analysis is built upon decades of classical static and dynamic code analysis, however, to cope with more and more sophisticated malware, we need to fully exploit its architecture specifics. The goal of my research was to develop a fully machine learning-based classification system following industrial requirements and fair assessment practices.

Second application deals with the analysis of soft skills in the context of the job market needs. We aimed not only at creating a large resource of soft skills list to facilitate future research but also showed how to mine those soft skills with more precision. We analyzed the relation between soft skills and salary, as well as the implications on occupational gender segregation.

Field of the dissertation Computer Science

Doctoral candidate M.Sc. (Tech) Luiza Sayfullina, born in Kazan, Russia

Time of the defence 27.9.2019 at 13

Place of the defence Aalto University School of Science, lecture hall AS2, Maarintie 2, Espoo

Opponent Assistant professor Isabelle Augenstein, University of Copenhagen, Denmark

Custos Professor Juho Kannala, Aalto University School of Science, Department of Computer Science

Electronic dissertation <http://urn.fi/URN:ISBN:978-952-60-8675-0>

Doctoral candidate's contact information Luiza Sayfullina, Department of Computer Science, +358 50 4304 643, sayfullina.luiza@gmail.com
