Cyprus University of Technology

Faculty of Engineering and Technology

Doctoral Dissertation

Robust Financial Crime Detection in Big Data via Uncertainty-Aware Deep Learning Techniques

Christos Kleanthous

Limassol, February 2021

CYPRUS UNIVERSITY OF TECHNOLOGY

FACULTY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING, COMPUTER ENGINEERING AND INFORMATICS

Doctoral Dissertation

Robust Financial Crime Detection in Big Data via Uncertainty-Aware Deep Learning Techniques

Christos Kleanthous

Limassol, February 2021

Approval Form

Doctoral Dissertation

Robust Financial Crime Detection in Big Data via Uncertainty-Aware Deep Learning Techniques

Presented by

Christos Kleanthous

Supervisor: Dr. Sotirios Chatzis, Assistant Professor, Cyprus University of Technology

Signature _____

Member of the committee: Dr. Dimitrios Kosmopoulos, Associate Professor, University of Patras

Signature _____

Member of the committee: Dr. Stelios Z. Xanthopoulos, Associate Professor, University of the Aegean

Signature _____

Cyprus University of Technology

Limassol, February 2021

Copyrights

The approval of the dissertation by the Department of Electrical Engineering, Computer Engineering and Informatics does not imply necessarily the approval by the Department of the views of the writer.

I would like to thank ....

My supervisor Dr. Sotirios Chatzis for his invaluable support, the Commissioner of the Cyprus Tax Department, Mr. Yiannis Tsangaris, for his investment in the pursued state-of-the-art technology, my wife and children who have been a pillar of strength.

ABSTRACT

Taxation is one of the most important sources of revenue for the European Union and Value Added Tax (VAT) accounts [1] to EUR 1,2T and as such it is prevalent target for tax evasion. The European commission has estimated the difference between the estimated and collected VAT (VAT GAP) to be EUR 147B or 12.3% of the VAT revenue [2].

It is unfortunate that many EU Tax departments rely on outdated technology like rules-based systems to target high-yield taxpayers for audit in their effort to decrease the VAT GAP. In addition, the absence of research in state of the art technology by the Tax Departments is surprising, meaning that they have not benefited from advancements in intelligent systems.

This thesis draws inspiration from the most recent machine learning advances in areas like visual recognition and speech perception. We seek to introduce cutting edge technology in the tax departments arsenal against tax evasion. Specifically, we target the selection of high-yield taxpayers for audit. In our work, we rely on intelligently processed raw data obtained from available tax returns. The high-dimensional nature of the available data calls for the development of machine learning techniques that can learn to extract meaningful lower-dimensional representations to drive the predictive inference process. We address these needs in a comprehensive manner, yielding a novel a novel set of supervised and semi-supervised techniques. In all cases, we take special care mitigating the epistemic

uncertainty our problem is fraught with, as a result of the limited number of audited (labelled) data.

The success of this thesis would not have been possible without the wholeheartedly assistance of the Cyprus Tax Department and the inspired mentoring of the Taxation Commissioner Mr Yiannis Tsangaris. Specifically, with their approval, we were given anonymized access to over a million submitted VAT returns and the tax audit results, pertaining to the period 2013-2019. This availability of a large corpus of real-world data was a crucial factor that allowed for us to successfully pursue our research goals.

Keywords: Value Added Tax, audit selection, representation learning, epistemic uncertainty.