**Money-laundering detection using Machine Learning**

**Synopsis**



Under the supervision

Of

**Ms. Charu Gupta**

(Asst. Professor)

Computer Science and Engineering Department

by

Karan Rawlley, 34

Kritarth Bisht, 35

Om Amrit, 51

Department of Computer Science & Engineering

Bhagwan Parshuram Institute of Technology

PSP-4, Sec-17, Rohini, Delhi-110089

**Table of Content**

1. INTRODUCTION……………………………………………………
2. PROBLEM STATEMENT…………………………………………....
3. SOFTWARE REQUIREMENT SPECIFICATION…………………..
4. WORKLOAD MATRIX………………………………………………
5. REFERENCES………………………………………………………..

**Chapter 1**

**Introduction**

According to the United Nations Office on Drugs and Crime [1], it’s estimated that 2% to 5% of the global GDP—or $800 billion to $2 trillion, proceeds from drugs and cybercrime to people trafficking—is laundered each year. UK companies have been widely used as vehicles for money-laundering [2][3], but illegal activities leave traces which are particularly well-suited for patterns’ recognition by computers. Our objective is to detect money-money laundering using as input UK corporate data (the nationality and status individual/corporate of its officers and beneficial owner) to perform binary classification of companies. An overview work undertaken in fraud detection using artificial intelligence is presented in [4], with good performance using Deep Autoencoder Networks described in [5], and models adapting to the ever changing nature of fraud in explained [6]. Our work is seminal, as we are the first to use the company registry. Dirty money often comes from predicate (underlying) crimes such as drug trafficking, illicit arms trades, smuggling, prostitution, gambling, corruption and bribery, fraud, piracy, robbery, currency counterfeiting and other organized crimes. In some cases, it may also come from incomes of legal businesses which need to be hidden for evading taxes. ML enables the conversion of cash from the underground (shadow) economy into monetary instruments of the legal economy.

**Chapter 2**

**Problem Statement**

Money Laundering (ML) is a serious problem for the economies and financial institutions around the world. Financial institutions get used by organized criminals and terrorists as vehicles of large-scale money laundering, which presents the institutions with challenges of regulatory compliance, maintaining financial security, preserving goodwill and reputation and avoiding operational risks like liquidity crunch and lawsuits. Hence prevention, detection and control of ML are crucial for the financial security and risk management of financial institutions. UK companies have been widely used as vehicles for money-laundering [2][3], but illegal activities leave traces which are particularly well-suited for patterns’ recognition by computers. Our objective is to detect money-money laundering using as input UK corporate data (the nationality and status individual/corporate of its officers and beneficial owner) to perform binary classification of companies. An overview work undertaken in fraud detection using artificial intelligence is presented in [4], with good performance using Deep Autoencoder Networks described in [5], and models adapting to the ever changing nature of fraud in explained [6]. Our work is seminal, as we are the first to use the company registry.

**Our Solution**

This project propose to perform classification using logistic regression, a fully connected neural network with 2, 5 and 10 layers (using Adam optimization with 512 hidden units, a batch size of 128 and train for 500 Epoch), a convolutional neural network (custom version of LeNet-5 where we only use convolution over one dimension and treated the list of features as a sequence), and a SVM with Gaussian Kernel. Given the relative simplicity of our input data, we do not expect significant better performance by adding more layers to the neural network. We expect the neural networks to perform marginally better than the other models given their capacity to learn more complex functions.

**Chapter 3**

**Software and Modules Requirements**

* Python 3.7
* Jupyter Notebook
* Keras
* TensorFlow
* Pandas
* Numpy
* Scikit-Learn

**Chapter 4**

**Workload Matrix**

|  |  |  |  |
| --- | --- | --- | --- |
| **Person(s) Responsible** | **Task** | **Description** | **Deadline** |
| Om  Karan | Dataset Preprocessing | Processing of data before using and making it efficient for model | Feb 21,2020 |
| Karan  Kritarth | Model Design | Designing of the deep learning model to be used for the project | Mar 2,2020 |
| Kritarth  Om | Training | Training model with the dataset for the project to be | Mar 9,2020 |
| Karan  Om | Parameter Tuning | Tuning algorithm for the specific problem | Mar 25,2020 |
| Karan  Kritarth | Testing and Evaluation | Finding bugs and errors in the project and solving them | Apr 6,2020 |

**References**

[1] United Nations Office on Drugs and Crime: <https://www.unodc.org/unodc/en/money-laundering/globalization.html>

[2] Getting the UK's House in Order. Global Witness: <https://www.globalwitness.org/documents/19717/Getting_the_UKs_House_in_Order_xZZxobR.pdf>

[3] At your service. Investigating how UK businesses and institutions help corrupt individuals and regimes launder their money and reputations. Transparency

International, October 2019: <https://www.transparency.org.uk/wp-content/plugins/download-attachments/includes/download.php?id=9299>

[4] Efstathios Kirkos, Charalambos Spathis, and Yannis Manolopoulos. Data mining techniques for the detection of fraudulent financial statements. Expert systems

with applications, 32(4):995–1003, 2007.

[5] Marco Schreyer, Timur Sattarov, Damian Borth, Andreas Dengel, and Bernd Reimer. Detection of anomalies in large scale accounting data using deep

autoencoder networks. CoRR, abs/1709.05254, 2017.

[6] Maria Jofre and Richard Gerlach. Fighting accounting fraud through forensic data analytics. CoRR, abs/1805.02840, 2018.

[7] Companies House: <https://beta.companieshouse.gov.uk>

[8] Snorkel: Rapid Training Data Creation with Weak Supervision. Alex Ratner, Stephen Bach, Henry Ehrenberg, Jason Fries, Sen Wu, Christopher Ré. VLDB 2018.

[9] Data Programming: Creating Large Training Sets, Quickly. Alex Ratner, Christopher De Sa, Sen Wu, Daniel Selsam, Christopher Ré. NeurIPS 2016