

Documents

Mukhopadhyay, M., Ghosh, K.

Panama Papers: How Data Science Fought Corruption
(2020) *SSRN*, .

DOI: 10.2139/ssrn.3644821

Indian Institute of Technology (IIT), Vinod Gupta School of Management, Kharagpur, India

Abstract

The Panama Papers are eleven million leaked electronic documents that detail financial and attorney–client information for more than two hundred thousand offshore entities. The documents were leaked in April 2016 by an anonymous whistle-blower from the database of Panamanian law firm and corporate service provider Mossack Fonseca. In this case study, we discuss how a team of international network of journalists collaborated using data mining tools to unearth vital financial fraud information from this large chunk of unstructured data. The case study begins with a prologue of dialogue between the whistle-blower and protagonist. The case is divided into nine mini chapters where we start with how a data team were formed followed by the tools used to clean and annotate the unstructured data into a graph-based database and finally using expert network of journalists to generate financial insights from the various nodes and links of the graph. We conclude the case by providing word clouds that highlights Indian connections. © 2020, The Authors. All rights reserved.

Author Keywords

Data Mining; Financial Fraud; Graph Database; Panama Paper; Shell Corporation; Tax Haven; Whistle Blowing

Index Keywords

C55, F38, K34, K42

Correspondence Address

Mukhopadhyay M.; Indian Institute of Technology (IIT), India; email: mayukhmukhopadhyay@iitkgp.ac.in

Publisher: SSRN

ISSN: 15565068

Language of Original Document: English

2-s2.0-85110253372

Document Type: Preprint

Publication Stage: Final

Source: Scopus