Searching for suspicious companies

Can Machine Learning shine a light on money laundering in the UK?

ANNA MACKENZIE

CAPSTONE PROJECT

AGENDA

01. CONTEXT
Background
Target Audience

02. GOALS
Problem Statement
Success Criteria

03. DATA
Collection & Cleaning
Exploratory Analysis

04. APPROACH
Feature Engineering
Model Selection

05. FINDINGS
Result Analysis
Limitations

06. CONCLUSIONS
Impacts
Next steps

CONTEXT

- Why does it matter?
- Why UK companies?
- How is this happening?
- Who cares?

GOALS

- **Problem statement:** “Can a binary classification model detect suspicious UK registered companies based on public data?”

- **Success metrics:**
  - **Accuracy** $= \dfrac{\text{correct predictions}}{\text{total predictions}}$
  - **Precision** $= \dfrac{\text{true positives}}{(\text{true positives} + \text{false positives})}$
  - **Recall** $= \dfrac{\text{true positives}}{(\text{true positives} + \text{false negatives})}$
  - **F1-score** $= 2 * \dfrac{1}{\dfrac{1}{\text{precision}} + \dfrac{1}{\text{recall}}}$

DATA

SCRAPING

- ❑ Compiling *"positive"* cases
- ❑ Random sampling of *"negative"* cases
- ❑ Collecting data using the Companies House API
    - Companies
    - Officers
    - PSCs and RLEs
    - Filings

DATA

SCRAPING

- Collecting data using the Companies House API
  - 9,786 companies

	name	address	kind	company_number	creation_date	dissolved_date	active_officers	inactive_officers	resigned_officers
4926	jacktown universal llp	61 Bridge Street Kington United Kingdom	llp	OC345842	2009-05-21	NaN	3	0	4
4927	igw worldwide llp	61 Bridge Street Kington United Kingdom	llp	OC375679	2012-05-30	NaN	3	0	4
4928	dartwall systems llp	Enterprise House 82 Whitchurch Road Cardiff	llp	OC339749	2008-08-29	2014-04-08	0	2	2
4929	augela systems llp	Cornwall Buildings 45-51 Newhall Street Office...	llp	OC335847	2008-03-22	2015-11-03	0	2	0

DATA

SCRAPING

- Collecting data using the Companies House API
  - 35,875 officers

	company_number	name	role	type	status	country	address
18645	OC345842	CARGOWEST AG	corporate-llp-designated-member	company	active	Marshall Islands	Ajeltake Road, Ajeltake Island, Majuro, Marsha...
18646	OC345842	CONVEX CREDIT LTD.	corporate-llp-designated-member	company	active	England	John Prince's Street, 4th Floor, London, England
18647	OC345842	OVERLUX AG	corporate-llp-designated-member	company	active	Marshall Islands	Ajeltake Road, Ajeltake Island, Majuro, Marsha...
18648	OC345842	FORMOND INC.	corporate-llp-designated-member	company	resigned	Marshall Islands	Ajeltake Island, Majuro, Marshall Islands
18649	OC345842	IRELAND & OVERSEAS ACQUISITIONS LIMITED	corporate-llp-designated-member	company	resigned	Belize	New Road, Belize City, Belize
18650	OC345842	MILLTOWN CORPORATE SERVICES LIMITED	corporate-llp-designated-member	company	resigned	Belize	No.35, New Road, Belize City, Belize
18651	OC345842	PRIMECROSS INC.	corporate-llp-designated-member	company	resigned	Marshall Islands	Ajeltake Island, Majuro, Marshall Islands

DATA

SCRAPING

- Collecting data using the Companies House API
  - 8,360 PSCs and RLEs

	company_number	name	type	status	country	address
0	OC345842	Convex Credit Ltd.	Limited Liability Company	active	UK	Unit 5, Olympia Industrial Estate, London, Eng...

	name	address	kind	company_number	creation_date	dissolved_date	active_officers	inactive_officers	resigned_officers
9784	convex credit ltd.	2nd Floor, College House 17 King Edwards Road ...	ltd	08916298	2014-02-28	NaN	1	0	4

DATA

SCRAPING

CONVEX CREDIT LTD.

Company number **08916298**

Follow this company | File for this company

Overview | Filing history | People | More

Officers | Persons with significant control

0 active persons with significant control / 1 active statement

Statement ACTIVE

The company knows or has reasonable cause to believe that there is no registrable person or registrable relevant legal entity in relation to the company

Notified on

28 February 2017

DATA

SCRAPING

- Collecting data using the Companies House API
  - 108,185 corporate filings

	company_number	filing_date	description
81677	OC345842	2019-12-07	change-registered-office-address-limited-liabi...
81678	OC345842	2019-07-30	accounts-with-accounts-type-total-exemption-full
81679	OC345842	2019-07-17	appoint-corporate-member-limited-liability-par...
81680	OC345842	2019-07-17	termination-member-limited-liability-partnersh...
81681	OC345842	2019-05-31	confirmation-statement-with-no-updates
81682	OC345842	2019-01-24	accounts-with-accounts-type-total-exemption-full
81683	OC345842	2018-05-22	confirmation-statement-with-no-updates
81684	OC345842	2018-01-03	accounts-with-accounts-type-total-exemption-full
81685	OC345842	2017-06-14	confirmation-statement-with-updates
81686	OC345842	2017-05-24	appoint-corporate-member-limited-liability-par...
81687	OC345842	2016-07-20	accounts-with-accounts-type-total-exemption-full

DATA

CLEANING

- ❑ Convert data to appropriate format
- ❑ Remove incomparable companies
- ❑ Infer missing countries for officers and PSCs
- ❑ Matching and combining similar addresses

DATA

CLEANING

- Matching and combining similar addresses

```
['113-115 george lane london',
 '113-115 george lane, london',
 '113-115  george lane, london',
 '113/115 george lane, london',
 '113-115 george lane, london, england',
 '113/115 george lane, london, england',
 '113-115 george lane london united kingdom',
 '113-115 george lane, london, united kingdom',
 '113 -115 george lane, london, united kingdom',
 '113/115 george lane, london, united kingdom',
 'george lane, 113-115 george lane, london, england',
 'enterprise house, 113 / 115 george lane, london',
 'enterprise house, 113/115 george lane, london',
 'enterprise house`, 113/115 george lane, london',
 'enterprise house, 113-115 george lane, london',
 'enterprise house, 113-115, george lane, london',
 'george lane, london, united kingdom']
```

DATA

EXPLORATORY ANALYSIS

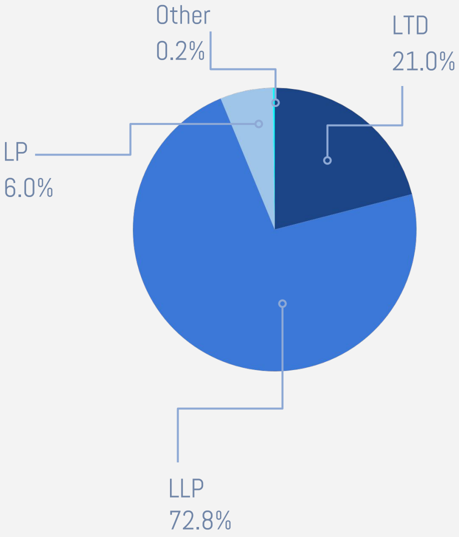
- ❑ Breakdown of companies by type
- ❑ Common addresses
- ❑ Number of active officers
- ❑ Number of non-human officers
- ❑ Officers in secrecy jurisdictions
- ❑ Number of active PSCs

DATA

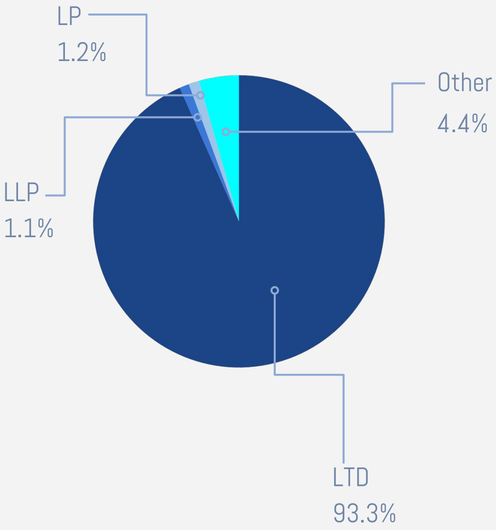
EXPLORATORY ANALYSIS

- Breakdown of companies by type

Suspicious companies



Random sample

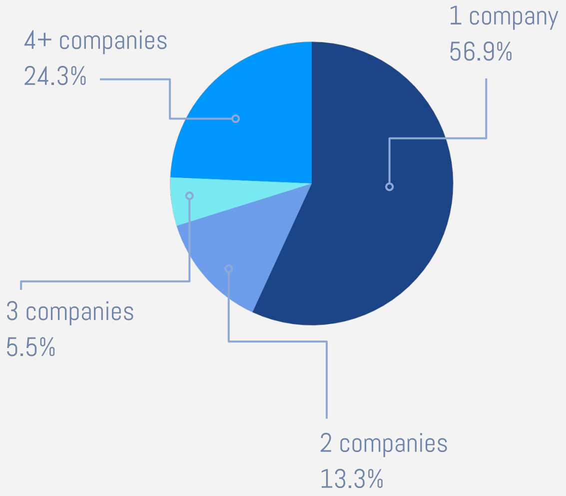


DATA

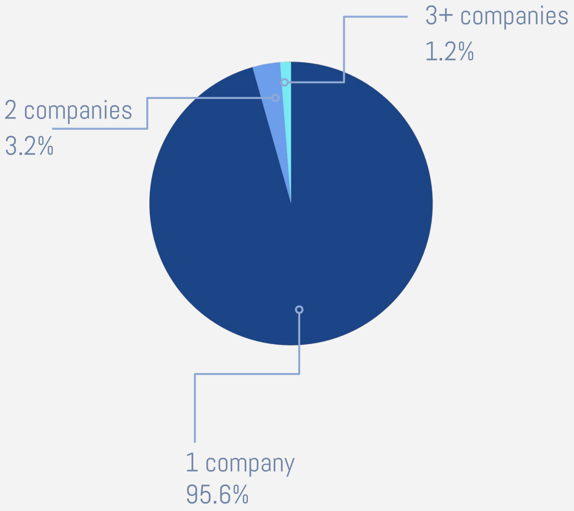
EXPLORATORY ANALYSIS

- Number of companies registered at address

Suspicious companies



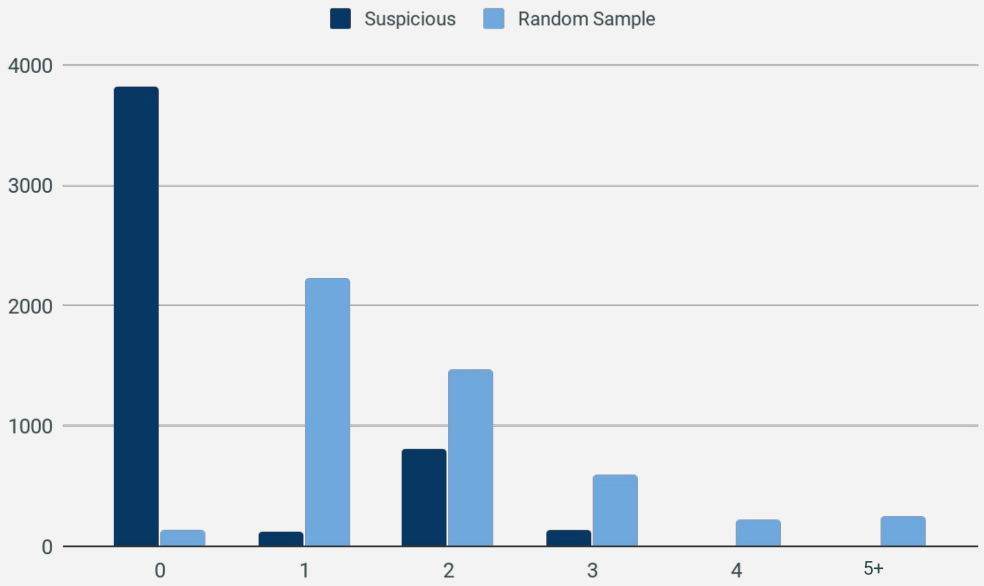
Random sample



DATA

EXPLORATORY ANALYSIS

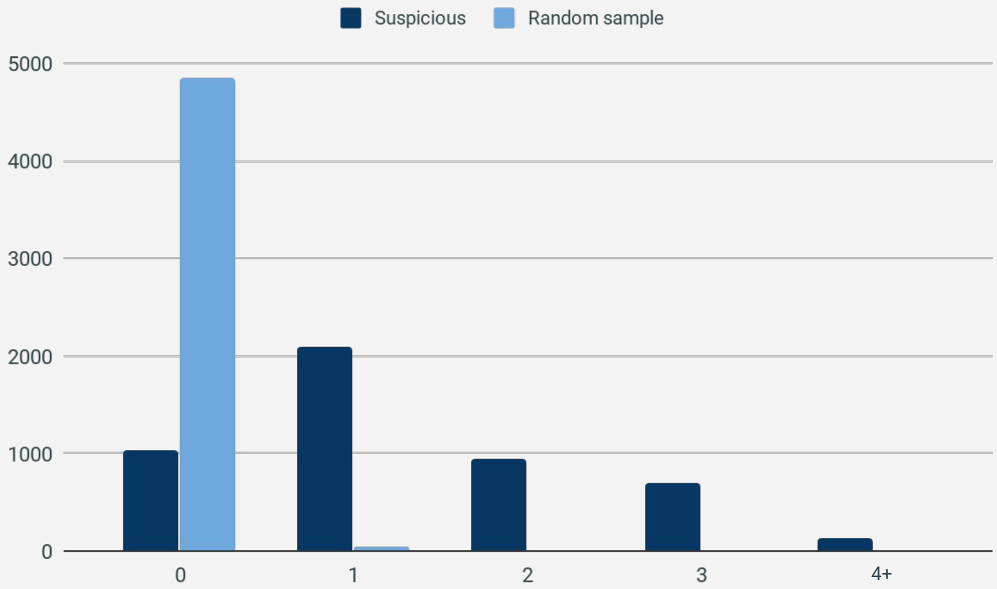
- Number of active officers



DATA

EXPLORATORY ANALYSIS

- Officers based in secrecy jurisdictions[1]

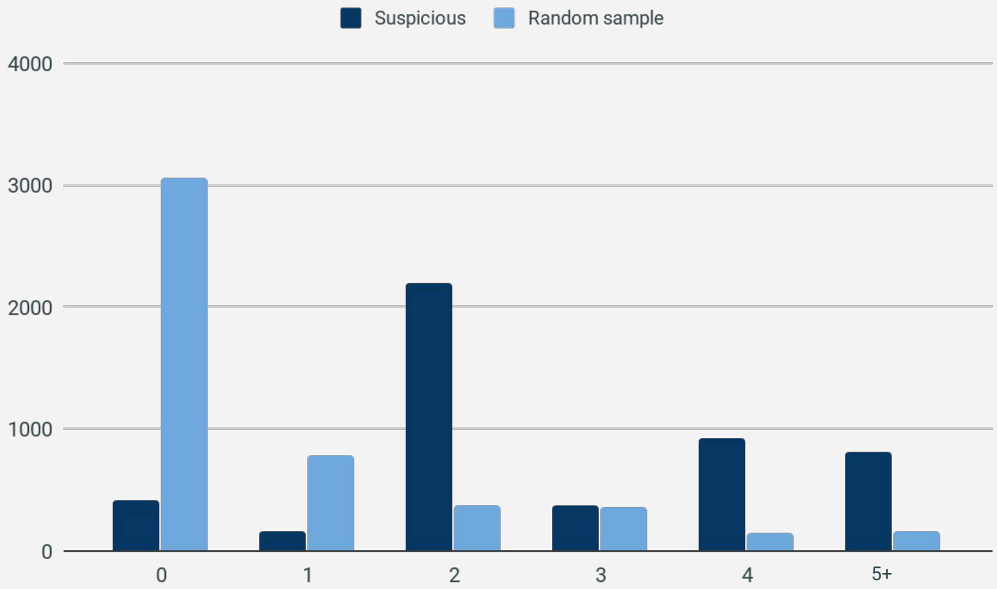


[1] Countries with a Tax Justice Network "Financial Secrecy Score" greater than 60

DATA

EXPLORATORY ANALYSIS

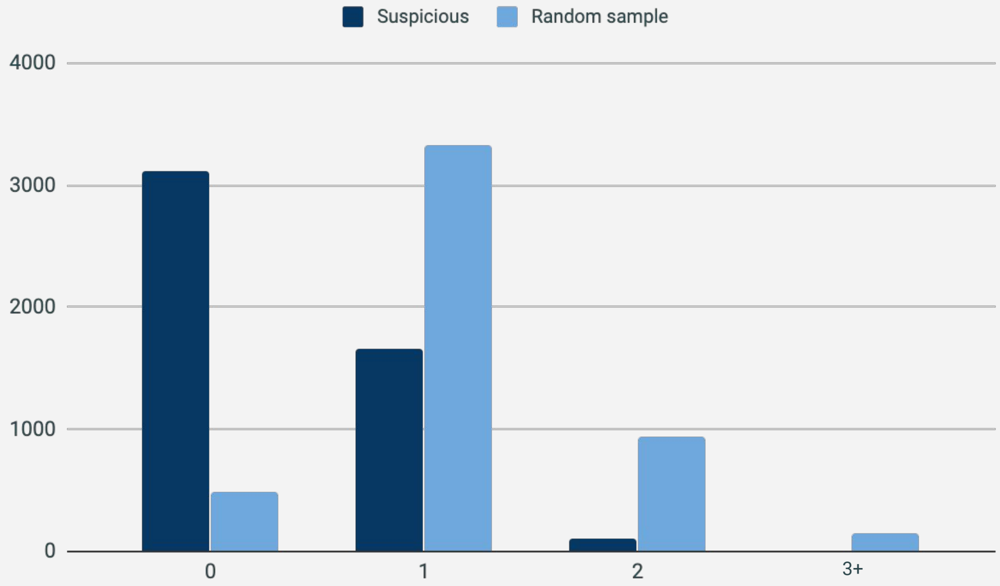
- Number of non-human officers



DATA

EXPLORATORY ANALYSIS

- Number of active PSCs



OVERALL APPROACH

01
DATA PREPARATION

02
EXPLORATORY DATA ANALYSIS

03
FEATURE ENGINEERING

04
MODEL SELECTION

05
EVALUATION OF RESULTS AND CONCLUSIONS

OVERALL APPROACH

01 DATA PREPARATION

02 EXPLORATORY DATA ANALYSIS

03 FEATURE ENGINEERING

04 MODEL SELECTION

05 EVALUATION OF RESULTS AND CONCLUSIONS

- DATA SCRAPING
- DATA CLEANING
- CREATION OF RELATIONAL DATABASES

OVERALL APPROACH

- 01 DATA PREPARATION
- 02 EXPLORATORY DATA ANALYSIS
- 03 FEATURE ENGINEERING
- 04 MODEL SELECTION
- 05 EVALUATION OF RESULTS AND CONCLUSIONS

❑ CORRELATION ANALYSIS

❑ PATTERN EXPLORATION

❑ SUMMARY STATISTICS

OVERALL APPROACH

01 DATA PREPARATION

02 EXPLORATORY DATA ANALYSIS

03 FEATURE ENGINEERING

04 MODEL SELECTION

05 EVALUATION OF RESULTS AND CONCLUSIONS

- CONVERTING CATEGORICAL FEATURES
- CREATING NEW FEATURES
- REMOVING IRRELEVANT INFORMATION

OVERALL APPROACH

01 DATA PREPARATION

02 EXPLORATORY DATA ANALYSIS

03 FEATURE ENGINEERING

04 MODEL SELECTION

05 EVALUATION OF RESULTS AND CONCLUSIONS

- RESCALING DATA
- SPLIT DATA FOR TRAIN AND TEST
- SET UP APPROPRIATE MODELS
- OPTIMISE MODELS
- CROSS VALIDATE AND COMPARE SCORES

OVERALL APPROACH

01 DATA PREPARATION

02 EXPLORATORY DATA ANALYSIS

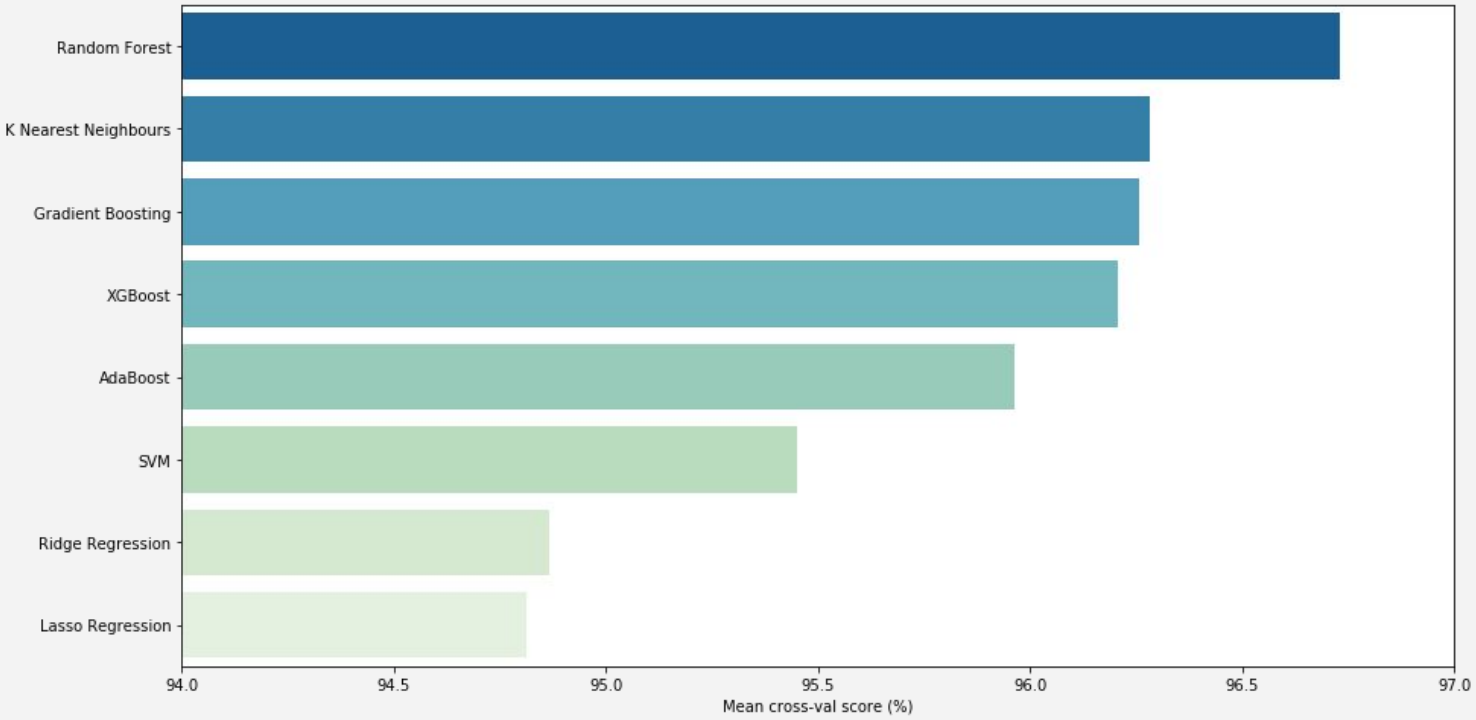
03 FEATURE ENGINEERING

04 MODEL SELECTION

05 EVALUATION OF RESULTS AND CONCLUSIONS

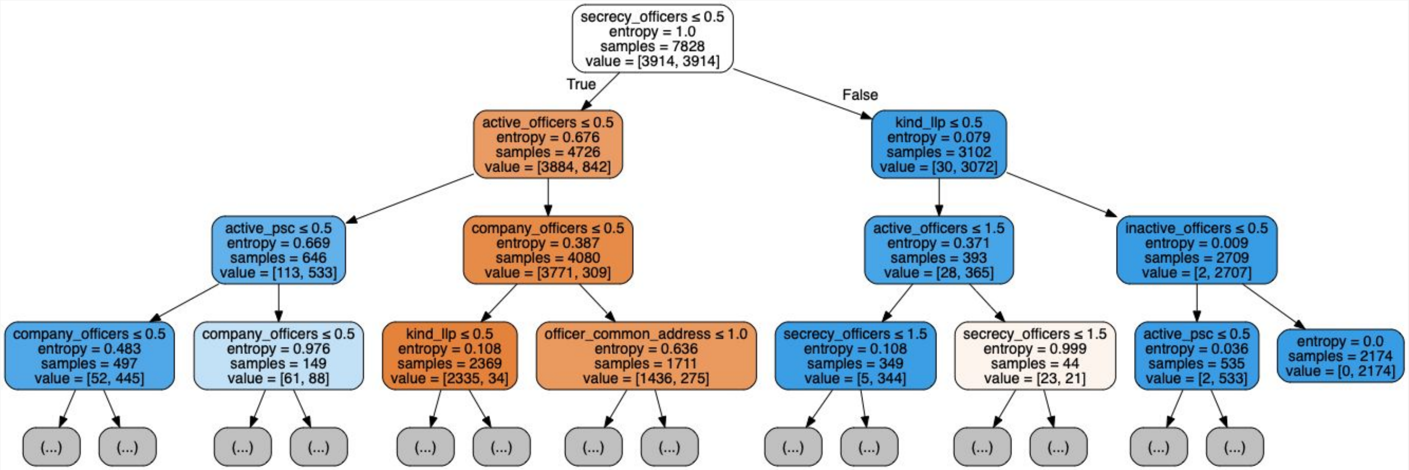
- ❑ CONTEXTUALISE RESULTS
- ❑ IDENTIFY “RED FLAGS”
- ❑ REVIEW SUCCESS OF MODEL

MODEL OVERVIEW



MODEL OVERVIEW

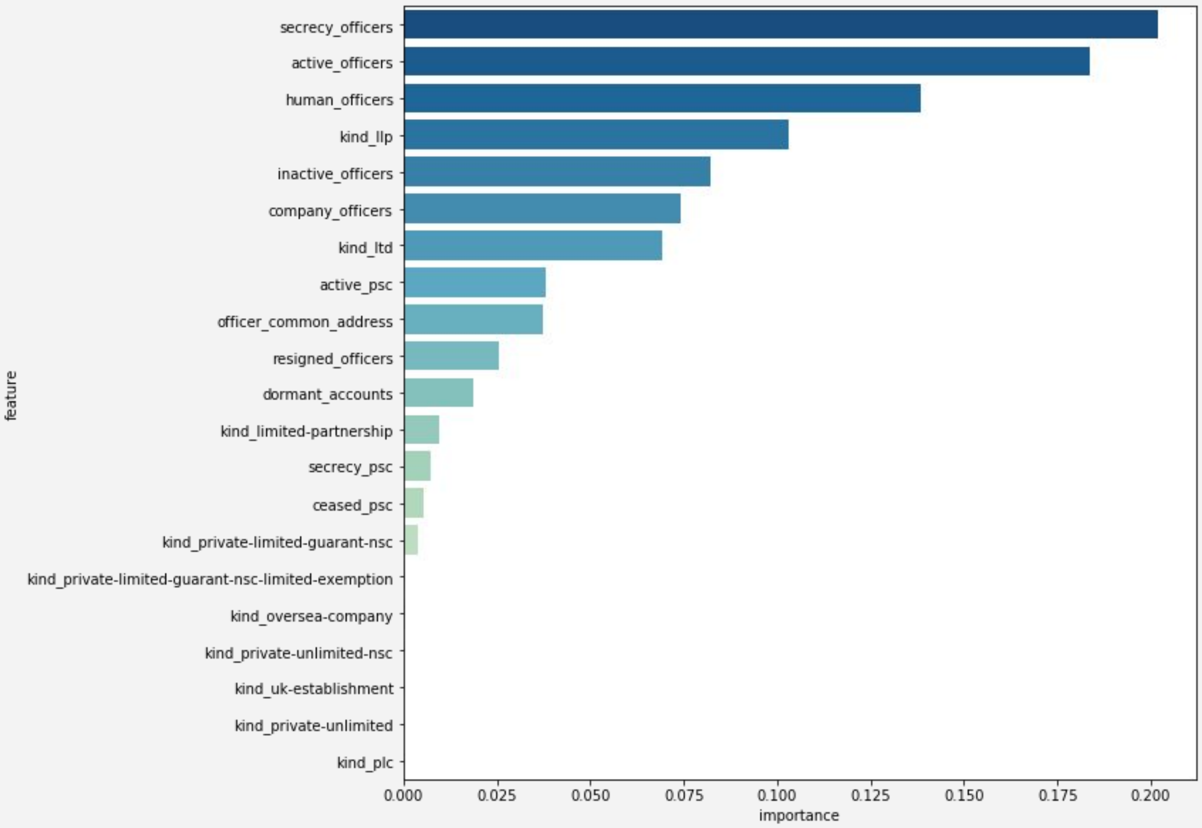
- ❑ 600 individual decision trees
- ❑ 600 randomly sampled training sets
- ❑ 1 “vote” per model
- ❑ Class with most votes wins



FINDINGS

	Training set	Testing set
Accuracy	98.5%	96.9%
Precision	98.7%	97.2%
Recall	98.1%	96.7%
F1-score	98.4%	97.0%

FINDINGS


secrecy_officers
active_officers
human_officers
kind_llp
inactive_officers
company_officers
kind_ltd
active_psc
officer_common_address
resigned_officers
dormant_accounts
kind_limited-partnership
secrecy_psc
ceased_psc
kind_private-limited-guarant-nsc
kind_private-limited-guarant-nsc-limited-exemption
kind_oversea-company
kind_private-unlimited-nsc
kind_uk-establishment
kind_private-unlimited
kind_plc
feature
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
importance

FINDINGS

- **Red flags:**
  - No active officers
  - Officers based in secrecy jurisdictions
  - Corporate entities as officers
  - No active PSC or RLE

- **Limitations:**
  - Positive samples taken from specific laundromats and leaks, may not generalise to other cases
  - Undetected false negatives in random sample
  - Can only raise the flag for risky companies

CONCLUSIONS

- **Impact:**
  - Enhanced screening and early detection
  - Highlights system vulnerabilities
- **Areas for improvement:**
  - Network analysis
  - Feature interaction
  - Further feature creation
- **Productionisation:**
  - Deployment in industry
  - Feedback

Q & A

THANK YOU FOR LISTENING