Norwegian School of Economics

Bergen, Fall 2020

Can we unveil the Secrecy of Tax Havens?

Daniyal Arif

Supervisor: Evelina Gavrilova-Zoutman

Master Thesis, Economics & Business Administration

Major: Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Table of Contents

TABLE OF CONTENTS **2**

LIST OF TABLES **4**

LIST OF FIGURES **6**

ACKNOWEDGMENT **7**

1. INTRODUCTION **8**

2. LITERATURE REVIEW **11**

3. SAMPLE AND RESEARCH DESIGN **12**

3.1 DATA COLLECTION & CLEANING 12

3.2 SAMPLE 13

3.3 TAX HAVEN SCORE BY INDUSTRY 14

3.4 HANDLING MISSING VALUES 18

3.5 DESCRIPTIVE STATISTICS 19

3.6 INDUSTRY-WISE ANALYTICS 21

4. RESEARCH DESIGN **28**

4.1 METHODOLOGY 28

4.2 PREDICTING OPERATIONS IN TAX HAVENS 28

4.3 PREDICTING OPERATION INTENSITY IN TAX HAVENS 30

4.4 EMPIRICAL RESULTS 31

4.5 VARIABLE IMPORTANCE 33

4.6 PREDICTION WITH WEIGHED BINS 34

5. FUTURE RESEARCH **37**

6. CONCLUSION **38**

7. REFERENCES & APPENDIX **39**

7.1 REFERENCES 39

7.2 APPENDIX 41

7.3 DATA REPOSITORY 41

List of Tables

Table 1: Companies in each industrial category ... 13

Table 2: Tax haven score per industrial category ... 16

Table 3: Missing data ... 18

Table 4: Top ten tax haven jurisdiction in 2018 ... 20

Table 5: Top ten tax haven jurisdiction per industrial category in 2018 ... 20

Table 6: Most popular tax havens in construction ... 22

Table 7: Most popular tax havens in FIR ... 22

Table 8: Most popular tax havens in manufacturing ... 23

Table 9: Most popular tax havens in mining ... 24

Table 10: Most popular tax havens in public administration ... 24

Table 11: Most popular tax havens in retail trade ... 25

Table 12: Most popular tax havens in services ... 26

Table 13: Most popular tax havens in TCEGS ... 26

Table 14: Most popular tax havens in wholesale trade ... 27

Table 15: Proportion of firms in tax havens in binary classification ... 29

Table 16: GBM confusion table in binary classification ... 29

Table 17: Prediction testing on binary classification ... 30

Table 18: Tax haven intensity ... 30

Table 19: Confusion tables of machine learning predictions ... 33

Table 20: Variable importance across machine learning models ... 34

Table 21: Tax haven range and weighed bins 35

Table 22: Confusion table of GBM with weighed bins... 35

Table 23: Sensitivity and specificity of GBM with weighed bins 35

Table 25: List A.. 41

List of Figures

Fig 1: Density plot for tax haven score 14

Fig 2: Boxplot tax haven score and CIK count per industrial category 17

Fig 3: Predictive accuracy with varying bins 36

Acknowedgment

Working on my master thesis has indeed helped in consolidating the knowledge I have learnt over the course of the degree. The learnings I have accumulated in the completion of this thesis are truly invaluable.

I feel fortunate to have Ms. Evelina Gavrilova-Zoutman as my master thesis supervisor. I owe endless gratitude for suggesting me a topic about which I had little idea initially but got more knowledgeable with time. Data collection looked tough in the start but now when I see in hindsight, I see myself much more capable and skilled. The guidance that my supervisor offered at every step was invaluable.

In addition, I would like to thank Norwegian Tax Administration for their support behind this thesis work. I hope the work will help Norwegian Tax Administration in improving detection of companies that operate in tax haven jurisdictions.

Norges Handelshøyskole

Bergen, Dec 2020

1. Introduction

Gramlich and Whiteaker-Poe (2013) write that in 2010 ninety-eight percent of Google's and ninety-nine percent of Oracle's subsidiary operations were missing in Exhibit 21 disclosure when compared with 2009. They add that these two business giants choose to disclose fewer subsidiaries and conjectures that tax incentives was an important reason behind this. Akamah et al. (2017) present evidence that firms operating in tax havens attempt to aggregate foreign operation disclosure in their financial reporting. Multinational companies have subsidiary operations in many regions of the globe. These companies sometimes chose not to disclose some of the subsidiary operations. In some of the instances, this happens when these companies want to hide operations in tax haven places. This study aims to disclose such instances of dishonest disclosures when companies choose to hide operations in tax havens.

Public listed multinational companies in US are required to disclose all their affiliate operations along with their jurisdictions. These companies often a times aggregate foreign operations or file dishonest financial disclosures by hiding some of their subsidiary operations. It helps them operate in tax havens, enjoy tax exemptions, and avoid public criticism when shifting profits from non-tax havens to tax haven subsidiaries. On the other hand, many of the companies that operate in tax havens may have reasonable grounds for operating in tax havens as well. Not all the companies that operate in tax haven places can be said to be operating for the sole purpose of tax evasion.

Detecting if a company operates in tax havens when no such evidence could be found in the financial disclosure is nearly an impossible task due to ring-fenced taxation. Tax havens often have ring-fenced tax system that provide legal state protection against revealing operations in tax havens. It is extremely difficult to obtain information with respect to a particular company, even upon request for access (Schjelderup, 2016). Therefore, the use of data analytics to uncover instances of tax evasion can prove valuable for tax administrations.

The non-disclosure of subsidiary operations by multinational companies become possible only due to weak regulations from SEC and costly enforcement mechanism. These regulations require companies to provide a transparent discourse of subsidiary operations but the rule is complicated, allowing firms latitude in its interpretation. In addition, the penalties imposed by SEC are extremely insignificant when compared to revenue of the companies and thus the companies choose not to disclose the subsidiary operations. Firms that do not fall into significant category of operations as per the definition by SEC are often aggregated. Firms list these insignificant operations into ‘other countries’ (Gramlich & Whiteaker-Poe, 2013). In addition, it is costly as well as cumbersome for tax authority to identify omissions within company financial statements. Companies may benefit by not declaring all the details but data analytics can help tax authorities in catching such instances of omissions.

In order to address the question if we can unveil hidden operations tax havens, I gathered a novel dataset by web crawling Electronic Data Gathering, Analysis, and Retrieval (EDGAR)[1], which is a digital repository of filings to Security and Exchange Commission (SEC)[2]. The dataset consists of jurisdictions of subsidiary operation for all public listed American companies from 2018. These companies were then categorized into relevant industries using standard industry classification (SIC)[3]. The location of subsidiary operation was crosschecked against a tax haven list prepared that will be termed as ‘List A’ in this study. ‘List A’ is listing of tax havens throughout the world along with the tax haven score. Finally, a haven intensity score was determined for each company in the dataset highlighting the intensity of a firm’s operations in tax haven jurisdictions.

Financial variables specific to each company were extracted using Wharton Research Database (WRDS)[4]. Most of these features (property, plant, equipment, log of Assets, log of Liabilities) used as predictors were the ones that showed significant relationship with aggregation (Akamah et al., 2017).

---

[1] https://www.sec.gov/

[2] Security and Exchange Commission (SEC) is a federal government agency in the United States of America that regulates the securities industry by enforcing securities laws. SEC enforces the statutory requirement on the public companies to file quarterly, annual reports, as well as other periodic reports. SEC maintains an online database called EDGAR (the Electronic Data Gathering, Analysis, and Retrieval system) from which investors can access this information filed with SEC.

In this study, I extracted 10 K document from EDGAR. 10 K is a comprehensive annual report that all the public listed companies in US have to file with SEC. This report describes in detail the financial performance of the company. In the report, companies also file Exhibit 21. This document enlists all the domestic as well as foreign operations by the company along with the jurisdiction of operation. Since in this study we intend to investigate operations in tax haven jurisdiction, Exhibit 21 was the primary source of our information.

[3] https://www.naics.com/sic-codes-industry-drilldown

[4] https://wrds-www.wharton.upenn.edu/

The data used in the thesis study can be fetched from the following GitHub repository:

https://github.com/daniyalarif/Master_Thesis

In addition, pre-tax domestic income, pre-tax foreign income, total tax and foreign tax were also used as predictor variables. Gramlich and Whiteaker-Poe (2013) write that pretax income have impact on Oracle and Google in reporting subsidiary operations. They give evidence of Google and Oracle reporting higher proportions of pretax income from foreign operations than revenue from foreign sources in 2011. This happens alongside changing number of subsidiaries disclosed in Exhibit 21. In case a pattern exists and companies use such manipulation tactics for tax planning purposes, these variables might have potential for their predictive ability. However, this needs to be experimented; this study attempts to do so.

The variables mentioned above were used as predictor variables in supervised machine-learning algorithms random forest, gradient boosting machine, k-nearest neighbour, support vector machine and multinomial regression. The aim is to predict operations and intensity of operations for a firm in tax havens.

The study also includes a data analytics approach applied to the same dataset. This involved an industry wide analysis investigating most popular tax haven jurisdictions amongst various industries. Most popular tax haven locations were located across industries and that allowed in isolating the firms that did not operate in those tax havens.

Predicting intensity of operations with supervised machine learning did not give conclusive results with this dataset as majority of the companies had few operations in tax haven jurisdiction. Skewed dataset led to difficulties in getting balanced predictions across the classes made from tax haven intensities. On the other hand, operations within tax havens were predicted with 80% accuracy.

The remainder of the paper is organized as follows. Section 2 provides a brief overview of research literature that could be found in the domain of tax avoidance. Section 3 discusses data collection and provides descriptive data statistics. Section 4 reports the research design and summarizes the empirical results. Finally, section 5, 6 discusses future research and concludes the paper.

2. Literature Review

This study builds on the previous research in the area of tax avoidance and uses machine learning for predicting tax haven aggressiveness on firm level. No similar research or the use of machine learning for predicting tax haven operations could be found in this domain previously. This may be due to lack of publicly available data needed for such a study.

Vast literature can be found on tax evasion and tax avoidance by multi-national firms induced by low taxation rates to operate in tax havens. Transfer pricing, strategic debt location and preferential cost allocation are mainly the means used by these multinational companies in transferring profits from high tax locations to low tax locations (Dyreng and Lindsey, 2009; Richardson and Taylor, 2015). Operation in tax havens often influences multinational companies to hide operation in some of the subsidiaries and present inaccurate disclosures. This helps them in being assessed less critically in the public domain.

In addition, research in tax avoidance examines relationship between a firm’s tax avoidance behaviour and non-disclosure of their geographic earnings (Hope et al., 2013). Tax avoidance influences financial reporting. Firms with operations in tax havens are more likely to aggregate geographic information disclosure in financial reporting (Akamah et al., 2017). Managers of these multinational firms attempting to avoid public criticism make geographic disclosure less transparent (Gramlich & Whiteaker-Poe, 2013). Unveiling the aggregated disclosure or revealing instances of operations in tax havens in dishonest disclosures has the ability to assist tax administration in detecting instances of tax evasion.

In the paper, the term tax haven and secrecy tax haven was used interchangeably. “Offshore financial centre” and “secrecy jurisdiction” are proxy for tax havens. Tax haven have no generally accepted criteria or definition (Guttorm Schjelderup, 2011).

3. Sample and Research Design

3.1 Data collection & cleaning

Foremost source of information for SEC is the investors themselves. Investors submit the company’s financial performance reports but there exists a lot of variability in the way these reports are structurally organized. This leads to inconsistent structure that makes web crawling cumbersome, as it gets harder to generalize a code pattern for the machine to crawl and extract information from the web.

Web addresses for 10-K documents were gathered for the companies over Edgar Server. Through these web addresses, Exhibit 21 webpage of the companies was accessed. List of subsidiary operations for a company filed in a 10-K annual return was extracted henceforth.

List of tax havens was compiled that we will call ‘List A’ in the study. Regular expressions in R programming language was used to compare the locations available in the ‘List A’ with Exhibit 21.

Basic knowledge of HTML code structure will help in understanding why the sample collected was less than ideal. For web scraping to extract information from web it is imperative that the web page from which data is extracted is structured in a consistent fashion. Many variations in the structure of Exhibit 21 were found.

In most occasions, the Exhibit 21 page was structured in table formats that delineated names of subsidiaries in one column while jurisdiction of operation in another column. Even though some tables included much more information (including voting rights and etcetera) but the company’s location of operation was often found in a column named jurisdiction or incorporation.

Certain companies used abbreviation of localities instead of full location names in their Exhibit 21 document. This also caused confusion in data collection. Multiple variants referring to the same place such as UK, England, Britain, British were observed. These inconsistencies led to several trial and error cycles to ensure maximum retrieval of information from the EDGAR server. These inconsistencies would not only confuse a web crawler but also even a casual reader. These inconsistencies also affect SEC contributing to costly enforcement of rules and regulations.

3.2 Sample

Sample consists of all public listed US companies on Edgar server in 2018. Exhibit 21 of each of the company was retrieved and the places of jurisdictions were extracted. In total, there were 10-K documents for 7,093 CIK[5] in Edgar server for 2018. 3,880 CIKs included Exhibit 21 document. Jurisdiction data for 3,140 CIK was extracted making it 44% of total CIK uploaded in 2018.

The variable data (extracted from WRDS) when assimilated with location data (extracted from EDGAR) from Exhibit 21 reduced the number of companies to 2,540 CIK. These 2,540 companies were categorised into industries using standard industry classification (SIC) coding reference.

Tables 1 shows number of companies for each industrial category found in the dataset. Most companies were from manufacturing sector while least number of companies were from Public Administration sector.

Industry	**No. of Companies**	**Prop. of Companies**
Construction	34	1.3%
FIR	615	24.2%
Manufacturing	865	34.1%
Mining	134	5.3%
PA	6	0.2%
Retail Trade	134	5.3%
Services	403	15.9%
TCEGS	264	10.4%
WT	85	3.3%
	2,540	

Table 1: Companies in each industrial category

[5] CIK: Central index key is a unique identifier assigned by securities exchange commission (SEC) to identify corporations and help investors get information about companies that have filed disclosure with the SEC.

(FIR: Finance, Insurance, Real Estate; PA: Public Administration; TCEGS: Transportation, Communications, Electric, Gas, And Sanitary Services; WT: Wholesale Trade)

Mean tax haven score was calculated for each company. Secrecy Index from Tax Justice Network helped in enlisting whether a location shared in Exhibit 21 list is a tax haven. The Financial Secrecy Index ranks jurisdictions according to of their offshore financial activities. The list is helpful in determining illicit financial flows or capital flight[6]. It scores locations from 1 – 100. 100 is the maximum score for a place amongst tax havens. Cayman Island had a score of 76 while Ireland had 48.

The dataset of mean tax haven score for 2,540 companies was right-skewed. There were more observations lying on the lower end of mean tax haven score. Density plot in Figure 1 depicts the distribution of mean tax haven score. The size of bars represent the count of observations. Firms with low mean tax haven score dominate the dataset. Mean tax haven score increases as we go towards the right. The bars decrease in size meaning the observations are low or very few companies had high mean tax haven score.

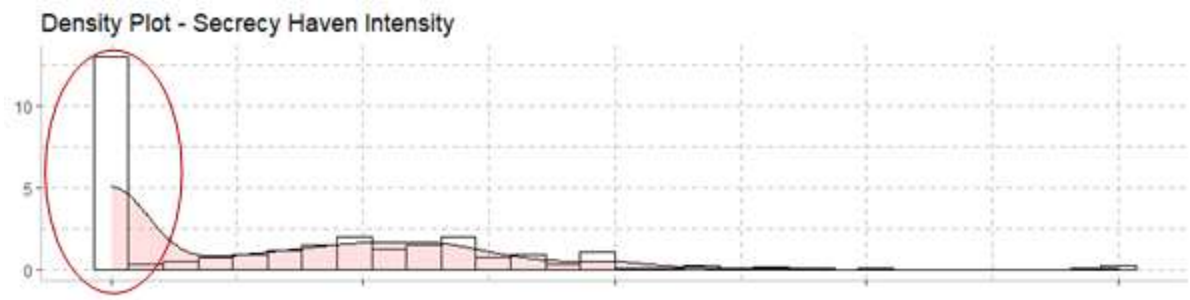

Fig 1: Density plot for Tax Haven Score

3.3 Tax Haven Score by Industry

Figure 3 shows a summary of tax haven score per industrial category. Box plot visualization below will also help in understanding the distribution of tax haven score across various industries. Companies operating in public administration industry had the highest mean tax haven score but only 6 out of 2,540 companies were from public administration. Low

[6] https://fsi.taxjustice.net/en/

observations of companies from public administration sector made the comparison with other industries harder.

On average, companies had 10 – 20 % operations in tax havens. Companies belonging to manufacturing and service sector looked to have a higher tendency to be working in tax havens. Companies from these industries had higher median and mean tax haven score compared to the other industries.

Agios, which is a US healthcare research firm, shared subsidiary operations in Massachusetts, Bermuda and Switzerland. Two out of three places were tax havens as per our methodology used within this study. The mean tax-haven aggressiveness score for this firm was 0.67 meaning that 67% of the operation for this company link to tax haven places.

KBR Inc. is an American engineering, procurement, and construction company. It was public listed on EDGAR database in 2018 and belongs to the construction sector. The tax haven intensity score was 0.4. This means that 40% of the operations were declared to be within tax havens in the Exhibit 21 filed by the company in 2018. KBR Inc. had operations in Alabama, Mexico, Texas, England, Delaware, Saudi Arabia, Canada, Cayman Islands, Netherlands, Norway, Singapore, India, Panama and Indonesia. KBR Inc. had operations in the following tax haven jurisdiction: Saudi Arabia, Cayman Islands, Netherlands, Singapore, Panama and Indonesia.

Boeing Co. is an American multinational corporation that designs, manufactures, and sells airplanes, rotorcraft, rockets, satellites, telecommunications equipment, and missiles worldwide. It was public listed on EDGAR database in 2018 and belongs to the manufacturing sector. The tax haven intensity score was 0.375. This means that 37.5% of the operations (or subsidiaries) were declared to be within tax havens in the Exhibit 21 filed by the company in 2018. Boeing Co. had operations in Delaware, Germany, United Kingdom, Netherlands, Bermuda, Singapore, Washington and Canada. Boeing Co. had operations in the following tax haven jurisdiction: Netherlands, Bermuda and Singapore.

Table 2: Tax haven score per industrial category

Construction (Cons)			
Min	Mean	Max	St. Dev
0	0.071	0.4	0.108
Finance, Insurance, Real Estate (FIR)			
Min	Mean	Max	St. Dev
0	0.0945	1	0.17
Manufacturing (Manu)			
Min	Mean	Max	St. Dev
0	0.242	1	0.202
Mining (Min)			
Min	Mean	Max	St. Dev
0	0.111	0.571	0.162
Public Administration (PA)			
Min	Mean	Max	St. Dev
0.167	0.2713	0.4	0.084
Retail Trade (RT)			
Min	Mean	Max	St. Dev
0	0.107	0.750	0.157
Services (Serv)			
Min	Mean	Max	St. Dev
0	0.199	1	0.185
Transportation, Communications, Electric, Gas, And Sanitary			
Min	Mean	Max	St. Dev
0	0.116	0.800	0.17
Wholesale Trade (WT)			
Min	Mean	Max	St. Dev
0	0.131	1	0.172

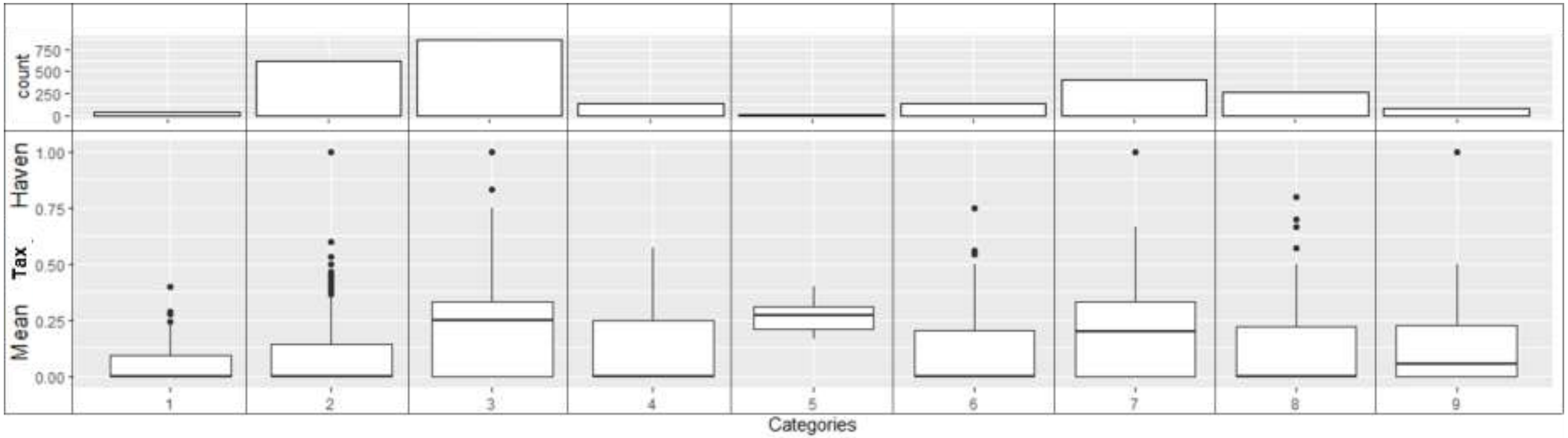

Fig 2: Boxplot tax haven score and CIK count per industrial category

1 : Construction; 2 : Finance, Insurance, And Real Estate; 3 : Manufacturing; 4 : Mining; 5 : Public Administration;

6 : Retail Trade; 7 : Services; 8 : Transportation, Communications, Electric, Gas, And Sanitary Services; 9 : Wholesale Trade;

3.4 Handling Missing Values

Company specific variables used as independent variables for this predictive study contained missing values. These needed to be addressed before progressing to predictions. There are three major approaches to handle missing values (Saar-Tsechansky & Provost, 2007). Simplest of all is to discard the observations with missing values. Doing this leaves us with 1,258 CIK, which means 50% of the observations are lost in the process and cannot be used for the prediction set.

The second one is to rely on the learning algorithm to deal with the missing values in the training phase. The third one is to impute the missing values before training the prediction method (Valdiviezo & Aelst, n.d.).

The aim is to test different machine learning algorithms on the data set. In order to test different machine learning algorithms we need to have a dataset free of missing values and thus imputing the missing values served us best. Third technique was applied to this dataset and dataset was imputed with bagged decision trees. This predictive value imputation technique is an ensemble of classification tree based imputation that has shown to produce accurate and well-calibrated probability compared to single tree-based imputation. However, this comes at a cost of over fitting (Valdiviezo & Aelst, n.d.). Table 3 lists missing data in each variable. Most of the missing data was from ‘Pre-tax Income Domestic’ and ‘Pre-tax Income Foreign’.

Variable	**Missing Values**
Assets	1
Liabilities	3
Pretax Income Domestic (PI Dom)	1,212
Pretax Income Foreign (PI For)	1,225
Property, Plant, Equipment (PPE)	139
Tax Foreign	482
Tax Total	4

Table 3: Missing data

3.5 Descriptive Statistics

Financial secrecy index from Tax Justice Network was used to determine the places that fell in the category of tax haven. The Financial Secrecy Index ranks jurisdictions according to the scale of offshore financial activities [7].The index issues a secrecy score for jurisdictions that helps rank tax havens globally; in this study, a cut-off of 60 was determined above which a place was ranked a tax haven.

Netherlands was the most popular tax haven location of jurisdiction for foreign operation. Netherland having FSI value 66 had frequency of 792, which means that 792 / 2,540 or 31% of the companies had operations in Netherland. China was second most popular tax haven with 743 occurrences and FSI value 60. This means that 743 / 2,540 or 29% of the companies had operations in China. Singapore was third most popular tax haven with 671 occurrences and FSI value 67. This means that 671 / 2,540 or 26% of the companies had operations in Singapore. Table 4 shows top ten tax haven locations, frequency of occurrence and FSI value amongst US companies in 2018.

Rank	**Tax Havens**	**Frequency**	**FSI Value[8]**
1	Netherlands	792	66.0
2	China	743	60.1
3	Singapore	671	67.1
4	Hong Kong	641	71.1
5	Japan	518	60.5
6	Switzerland	477	76.5
7	Cayman Islands	357	72.3
8	Bermuda	314	73.1
9	Jersey	314	65.5

[7] https://fsi.taxjustice.net/en/

[8] Financial Secrecy Index

In the course of the study, I realized that the approach failed to distinguish overlapping names such as Jersey and New Jersey. Therefore, the analysis excludes Jersey

10	Thailand	248	79.9

Table 4: Top ten tax haven jurisdiction in 2018

Table 5 shows companies' preferred tax haven locations in each industry. Companies were categorized using standard industry classification (SIC) into Construction; Finance, Insurance, And Real Estate (FIR); Manufacturing; Mining; Public Administration (PA); Retail Trade; Services; Transportation, Communications, Electric, Gas (TCEGS); Wholesale Trade (WT).

#	Construction	FIR	Manufacturing	Mining	PA	Retail Trade	Services	TCEGS	WT
1	Jersey	Jersey	China	Netherlands	Switzerland	Hong Kong	Netherlands	Netherlands	Netherlands
2	Netherlands	Cayman Islands	Netherlands	Cayman Islands	Japan	China	Singapore	Jersey	Singapore
3	Chile	Hong Kong	Singapore	Singapore	Netherlands	Puerto Rico	China	Cayman Islands	China
4	Panama	Bermuda	Hong Kong	Bermuda	Singapore	Netherlands	Hong Kong	Bermuda	Switzerland
5	Indonesia	Singapore	Japan	British Virgin	China	Bermuda	Japan	Hong Kong	Hong Kong
6	Singapore	China	Switzerland	Bahamas	Bermuda	Jersey	Switzerland	Singapore	Japan
7	Bermuda	Netherlands	Thailand	Venezuela	Cayman Islands	Japan	Jersey	Japan	Indonesia
8	Cayman Islands	Japan	Turkey	Indonesia	Hong Kong	Singapore	Israel	China	Puerto Rico
9	Japan	Switzerland	Cayman Islands	Switzerland	Jersey	Cayman Islands	Cayman Islands	Switzerland	Turkey
10	Puerto Rico	Puerto Rico	Taiwan	Ghana	Aruba	Switzerland	Philippines	Chile	Jersey

Table 5: Top ten tax haven jurisdiction per industrial category in 2018

Netherland, China and Singapore came out to be the top tax haven jurisdictions. Netherlands and Hong Kong are also tax havens, but they foster a lot of inland real activity. Operations in jurisdictions like Cayman Islands and Bermuda are doubtful when we mention real activity. This is because the mentioned places have low population and the market size needed for business operations is not the main reason for business operations. Tax evasion related activity could thus be said to be a more plausible reason of presence in such places.

3.6 Industry-wise Analytics

In this section, we will dig into industry wise subsidiary operations in tax haven jurisdictions and look for instances of individual prediction error. Individual prediction error happens when most of the companies operate in certain tax haven jurisdiction while some do not disclose operations in those tax havens.

Earlier we observed top tax haven jurisdictions within each industrial category. With industry specific information about total number of companies, total tax haven operation occurrences, unique tax jurisdictions, and total number of companies operating in tax haven, companies that did not disclose operations in top tax haven locations were investigated. This helped isolate companies that did not disclose operations in the top ranked tax haven in Exhibit 21. It can also be highly likely that Exhibit 21 is wrong and can open possibilities for further investigation by the tax authorities.

- **Construction**

In the construction industry, there were in total 34 companies. There were 73 instances of operations within tax haven jurisdictions with operations found in 32 unique tax haven locations. Out of 34 companies, 15 had operations in tax haven jurisdictions. Jersey, Netherland, Chile were the most popular tax haven locations (data on Jersey was inaccurate and we will not focus on it). Netherland was the top tax haven location. 7 (21%) out of 34 companies in construction industry had operations in Netherland.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	Jersey	7	6	Singapore	4
2	Netherlands	7	7	Bermuda	3
3	Chile	5	8	Cayman Islands	3
4	Panama	5	9	Japan	3
5	Indonesia	4	10	Puerto Rico	3

Table 6: Most popular tax havens in construction

- **Finance, Insurance, and Real Estate (FIR)**

In the finance, insurance, and real estate industry, there were in total 615 companies. There were 729 instances of operations within tax haven jurisdictions with operations found in 53 unique tax haven locations. Out of 615 companies, 206 had operations in tax haven jurisdictions. Jersey, Cayman Island, Hong Kong were the most popular tax haven locations (data on Jersey was inaccurate and we will not focus on it). Cayman Island was the top tax haven location and 72 (12%) out of 615 companies in finance, insurance, and real estate industry had operations in Cayman Islands.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	Jersey	82	6	China	43
2	Cayman Islands	72	7	Netherlands	42
3	Hong Kong	62	8	Japan	37
4	Bermuda	53	9	Switzerland	28
5	Singapore	51	10	Puerto Rico	26

Table 7: Most popular tax havens in FIR

- **Manufacturing**

In the manufacturing industry, there were in total 865 companies. There were 3,817 instances of operations within tax haven jurisdictions with operations found in 61 unique tax haven locations. Out of 865 companies, 636 had operations in tax haven jurisdictions. China, Netherlands, Singapore were the most popular tax haven locations. 421 (49%) companies from the manufacturing sector out of 853 had operations in China.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	China	421	6	Switzerland	269
2	Netherlands	405	7	Thailand	156
3	Singapore	323	8	Turkey	126
4	Hong Kong	311	9	Cayman Islands	121
5	Japan	275	10	Taiwan	121

Table 8: Most popular tax havens in manufacturing

- **Mining**

In the mining industry, there were in total 134 companies. There were 175 total instances of operations within tax haven jurisdictions with operations found in 31 unique tax haven locations. Out of 134 companies, 50 had operations in tax haven jurisdictions. Netherlands, Cayman Island, Singapore were the most popular tax haven locations. 25 (19%) companies from the mining sector out of 134 had operations in Netherlands.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	Netherlands	25	6	Bahamas	9
2	Cayman Islands	22	7	Venezuela	9
3	Singapore	14	8	Indonesia	8
4	Bermuda	12	9	Switzerland	8

5	British Virgin Islands	11	10	Ghana	6

Table 9: Most popular tax havens in mining

- **Public Administration**

In the public administration industry, there were in total 6 companies. There were 42 instances of operations within tax haven jurisdictions with operations found in 23 unique tax haven locations. All of these 6 companies had operations in tax haven jurisdictions. Switzerland, Japan, Netherlands were the most popular tax haven locations. 5 (83%) companies from the public administration sector out of 6 had operations in Switzerland. CCUR Holdings Inc. was the only company from public administration that did not list operations in Switzerland in its financial disclosure. It may be a possibility that this company failed to declare operations in Switzerland when rest of the companies within this industrial category claimed to have operations in Switzerland.

Rank	**Tax Haven Jurisdiction**	**Companies Operating**	**Rank**	**Tax Haven Jurisdiction**	**Companies Operating**
1	Switzerland	5	6	Bermuda	2
2	Japan	4	7	Cayman Islands	2
3	Netherlands	4	8	Hong Kong	2
4	Singapore	4	9	Jersey	2
5	China	3	10	Aruba	1

Table 10: Most popular tax havens in public administration

Retail Trade

In the retail trade industry, there were in total 134 companies. There were 178 instances of operations within tax haven jurisdictions with operations found in 29 unique tax haven locations. Out of 134 companies, 56 had operations in tax haven jurisdictions. Hong Kong, China, Puerto Rico were the most popular tax haven locations. 25 (19%) companies from the retail trade sector out of 134 had operations in Hong Kong.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	Hong Kong	25	6	Jersey	12
2	China	20	7	Japan	11
3	Puerto Rico	16	8	Singapore	8
4	Netherlands	15	9	Cayman Islands	7
5	Bermuda	12	10	Switzerland	7

Table 11: Most popular tax havens in retail trade

Services:

In the services industry, there were in total 403 companies. There were 1,451 instances of operations within tax haven jurisdictions with operations found in 58 unique tax haven locations. Out of 403 companies, 274 had operations in tax haven jurisdictions. Netherlands, Singapore, China were the most popular tax haven locations. 153 (38%) companies from the services sector out of 403 had operations in Netherlands.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	Netherlands	153	6	Switzerland	83

2	Singapore	146	7	Jersey	58
3	China	121	8	Israel	57
4	Hong Kong	113	9	Cayman Islands	53
5	Japan	103	10	Philippines	42

Table 12: Most popular tax havens in services

- **Transportation, Communications, Electric, Gas, Sanitary Services (TCEGS)**

In the transportation, communications, electric, gas, sanitary services industry, there were in total 264 companies. There were 453 instances of operations within tax haven jurisdictions with operations found in 51 unique tax haven locations. Out of 264 companies, 116 had operations in tax haven jurisdictions. Netherlands, Jersey, Cayman Islands were the most popular tax havens. 40 (15%) companies from transportation, communications, electric, gas, sanitary services sector out of 264 had operations in Netherlands.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	Netherlands	40	6	Singapore	27
2	Jersey	35	7	Japan	26
3	Cayman Islands	34	8	China	19
4	Bermuda	29	9	Switzerland	17
5	Hong Kong	29	10	Chile	14

Table 13: Most popular tax havens in TCEGS

- **Wholesale Trade**

In the wholesale trade industry, there were in total 85 companies. There were 247 instances of operations within tax haven jurisdictions with operations found in 38 unique tax haven locations. Out of 85 companies, 43 had operations in tax haven jurisdictions. Netherlands, Singapore, and China were the most popular tax havens. 24 (28%) companies from the wholesale trade sector out of 85 had operations in Netherlands.

Rank	Tax Haven Jurisdiction	Companies Operating	Rank	Tax Haven Jurisdiction	Companies Operating
1	Netherlands	24	6	Japan	12
2	Singapore	24	7	Indonesia	10
3	China	19	8	Puerto Rico	9
4	Switzerland	15	9	Turkey	9
5	Hong Kong	14	10	Jersey	8

Table 14: Most popular tax havens in wholesale trade

4. Research Design

4.1 Methodology

Tax haven aggressiveness score was used as the independent variable while a company's assets, liabilities, pre-tax-domestic-income, pre-tax-foreign-income, property-plant-equipment, tax-foreign, tax-total were used as dependent variables to predict the tax haven intensity for a firm.

The predictor variable tax haven intensity score lies between zero and one. Zero means none of its subsidiary operates in tax haven location while one means that the firm operates solely in places that our considered to be tax havens. First step includes predicting operations in tax havens while in the later step intensity of operations in tax havens was predicted.

4.2 Predicting operations in Tax Havens

Multinational firms have tendency to hide operations from declaring in financial statements especially when these companies operate in tax havens. Predicting if a company operates in a tax haven jurisdiction may help tax authorities to unveil aggregation or even catch operations in tax havens when Exhibit 21 fails to say so.

Two bins were created from tax haven score: one signifies no presence in tax haven locations while the other contains all the companies operating in tax haven jurisdictions. When tax haven score equals zero a firm belongs to bin zero. This means that the firm had no operation in tax haven jurisdictions. On the contrary, if a firm had tax haven score greater than zero this means that the firm had operations in tax havens and all such firms were a part of bin one. This binary method of classification formulates nearly equal distribution or number of companies across the two bins.

70-30 train-test split was used in this study. 70%, of the sample, which amounts to 1,780 CIK, was used for train sample and rest of 30%, which amounts to 760 CIK were used as test sample. The test set contained values that were not used to train the algorithm. These set of values were new for the algorithm and determines how accurately the model predicts unknown values.

Bins	Firms Proportion	Tax Haven Range
1	55.2%	(0.00 – 1.00]
0	44.8%	0.0 – 0.0

Table 15: Proportion of firms in tax havens in binary classification

Confusion matrix of the prediction can be seen below:

GBM	Actual	
Prediction	0	1
0	**273**	85
1	68	**335**

Table 16: GBM confusion table in binary classification

The highlighted values on the diagonal in all the confusion matrices refer to prediction being accurate. Actual refers to the real tax haven category while prediction refers to the category predicted by the algorithm. Model's accuracy was 79.8%. It was determined by dividing values on diagonals by total observations.

This model had an accuracy of almost 80% which means that it is possible to predict operations in tax havens with 80% certainty. The model had a sensitivity of 80.0%. This means that places that are not operating in tax havens will be correctly identified as not operating in tax havens 80% of the times. Specificity for the model was 79.8%. This means that firms that operate in tax havens will be correctly identified to be operating in tax havens 79.8% of the times.

- **Prediction Testing**

The model for predictions was tested in real time to see the model's performance. Ten randomly chosen companies from the test dataset were compared against actual tax haven category.

Name	Tax Haven Score	Category Distribution	Predicted Category	Prediction Status
Baker Hughes	0.143	2	2	✕
Cheniere Energy	0.429	2	1	✔
Adam Resources and Energy	0	1	1	✔
EP Energy Corp	0	1	1	✔
Superior Energy Services Inc.	0.125	2	2	✔
Chaparral Energy Inc.	0	1	1	✔
National Oil Well Varco	0.323	2	2	✔
Parsley Energy	0	1	1	✔
PBF Energy	0	1	1	✔
Berkshire Hathway Energy Co	0	1	2	✕

Table 17: Prediction testing on binary classification

The model predicted eight out of ten companies (randomly selected) accurately.

4.3 Predicting operation intensity in tax havens

The predictor variable was made into a categorical variable with five equally spaced categories. This helped gauge the level of activity by a firm in tax haven jurisdictions. If a firm had 0.34 tax haven score that means 34% of operations were in tax havens. This place would get '2' in such a categorization. '2' refers to those companies who had 20-40% operations in tax havens.

Haven Category	Haven Score Range	% of Companies	# of Companies
1	0 – 0.2	60%	1,529
2	0.2 – 0.4	30%	763
3	0.4 – 0.6	8%	194
4	0.6 – 0.8	1%	27
5	0.8 – 1.0	1%	27

Table 18: Tax haven intensity

Earlier we observed that the sample is skewed. Tax haven categories three, four, five have low observations compared to one, two. Random sampling might have led to a test set without any observation from tax haven bins three, four or five. In that case, algorithm might not be tested against all tax haven categories. Random sampling would thus have led to unrepresentative train/test sample splits and predictions from the model would thus be inaccurate.

Therefore, stratified sampling was used. In this sampling, population is divided into sub-populations or strata for tax haven categories. Each strata is divided into (70/30) train/test split ensuring that all tax haven categories get represented proportionately across the training and test set [9].

4.4 Empirical Results

Five machine learning models including Random Forest, Gradient Boosting Machine, Support Vector Machine, K Nearest Neighbor, and Multinomial Logistic Regression were implemented to determine which machine-learning model works best for this data set and had the highest prediction rate. Confusion Tables listed below shows the performance results for each of the models.

Random Forest	**Actual (Accuracy: 68.7%)**				
Prediction	1	2	3	4	5
1	**387**	92	31	5	6
2	70	**135**	27	3	2
3	1	1	**0**	0	0
4	0	0	0	**0**	0
5	0	0	0	0	**0**

[9] http://essedunet.nsd.uib.no/cms/topics/weight/2/5.html

GBM[10]	Actual (Accuracy: 71.2%)				
Prediction	1	2	3	4	5
1	**409**	93	25	5	6
2	47	**132**	32	3	1
3	0	3	**0**	0	0
4	1	0	1	**0**	1
5	1	0	0	0	**0**

KNN[11]	Actual (Accuracy: 61.5%)				
Prediction	1	2	3	4	5
1	**436**	197	56	8	8
2	22	**31**	2	0	0
3	0	0	**0**	0	0
4	0	0	0	**0**	0
5	0	0	0	0	**0**

SVM[12]	Actual (Accuracy: 63.3%)				
Prediction	1	2	3	4	5
1	**456**	203	54	8	8
2	1	**24**	3	0	0
3	1	1	**1**	0	0
4	0	0	0	**0**	0
5	0	0	0	0	**0**

[10] Gradient Boosting Machine

[11] K Nearest Neighbour

[12] Support Vector Machine

Multi-nomial	**Actual (Accuracy: 62.4%)**				
Prediction	1	2	3	4	5
1	**453**	204	56	8	8
2	4	**21**	2	0	0
3	1	1	**0**	0	0
4	0	0	0	**0**	0
5	0	2	0	0	**0**

Table 19: Confusion tables of machine learning predictions

Prediction accuracy of Random Forest Model, Gradient Boosting Machine, K Nearest Neighbor, Support vector machine and Multinomial Regression was 68.7%, 71.2%, 61.5%, 63.3%, and 62.4% respectively.

Gradient Boosting Machine gave the highest accuracy for this dataset. A pattern could be observed across all the confusion tables; that is, low prediction rate for companies that had high tax haven scores on tax haven categorization. It is because the companies with higher scores on tax haven intensity were much fewer in number and thus the observations of such companies were rare in the dataset. Models were trained poorly with few observations and thus there was low prediction accuracy at category 3, 4, 5 or companies having above 40% operations in tax havens.

4.5 Variable Importance

Out of the seven-predictor variables used to predict tax haven intensity, Foreign Tax emerged as the top predictor across the board for nearly all the models. Variable importance for the other variables differed across the models.

	Models				
	RF	**GBM**	**KNN**	**SVM**	**Multinomial**
Prediction Accuracy	68.7%	71.2%	61.5%	63.3%	62.4%
	Variable by Importance				

1	Foreign Tax		Foreign Tax	Foreign Tax	Foreign Tax
2	PI For[13]		PPE	PPE	PPE
3	PI Dom[14]		Assets	Assets	PI Dom
4	Total Tax		Liabilities	Total Tax	Total Tax
5	Liabilities		PI Dom	Liabilities	PI For
6	Assets		Total Tax	PI Dom	Liabilities
7	PPE[15]		PI For	PI For	Assets

Table 20: Variable importance across machine learning models

4.6 Prediction with weighed bins

It was clear from the confusion tables that prediction accuracy majorly came from company's belonging to bin one or two. Therefore, such a methodology would not do well when predicting a firm belonging to bins three or four or five. 90% of the observations belong to bin one or two. The results showed more accuracy in predicting companies with lower tax-haven intensity score. `

Instead, the bins were formulated again based on proportion of firms instead of tax haven score. By doing so, tax haven score may not be equally spaced but all the bins had ample number of observations that would facilitate the model training. Details of the bins created can be seen below:

Bins	**Proportion of Firms**	**Haven Score Range**
5	14.0%	0.364 – 1.0
4	13.9%	0.286 – 0.36
3	14.3%	0.2 – 0.282
2	13.0%	0.021 – 0.196

[13] PI For: Pretax income foreign

[14] Pre Dom: Pretax income domestic

[15] PPE: Property, plant, equipment

1	44.8%	[0.0 – 0.021)

Table 21: Tax haven range and weighed bins

GBM model was trained with these newly constructed bins. Confusion table in table 22 illustrates the results.

GBM	**Actual (Accuracy: 49.7%)**				
Prediction	1	2	3	4	5
1	**276**	36	26	27	38
2	31	**22**	11	10	10
3	16	18	**30**	23	15
4	5	20	20	**28**	20
5	13	9	18	17	**22**

Table 22: Confusion table of GBM with weighed bins

	Class 1	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Sensitivity	80.9%	21.0%	28.6%	26.7%	21.0%
Specificity	70.0%	90.5%	89.0%	90.0%	91.3%

Table 23: Sensitivity and specificity of GBM with weighed bins

Prediction accuracy decreased from 71.2% to 49.7%. Bin 1 had relatively higher number amongst the values on the diagonal of the confusion table. This bin contained all the firms having negligible operation in tax havens; it contained nearly 45% of all the observations from the dataset. Higher number of observations led to more accurate predictions for bin 1.

The drop in the prediction accuracy of weighed bin model is attributable to prediction being made across the five tax haven categories. Previously, the predictions were mainly from bin 1 and 2. This resulted in higher predictive accuracy with less inaccurate predictions from bin 3, 4, 5. Accuracy dropped with weighed bins as less observation to train in each of the bins resulted in predictions that are more inaccurate.

To confirm the observation stated earlier, the bins size was varied and GBM model was trained. All the models with varying bin sizes were tested for predictive accuracy. As number of bins increased, predictive accuracy of the model decreased. Predictive accuracy for 2, 3, 4, 5 weighed bins was 79.8%, 58.2%, 53.9%, 49.7% respectively.

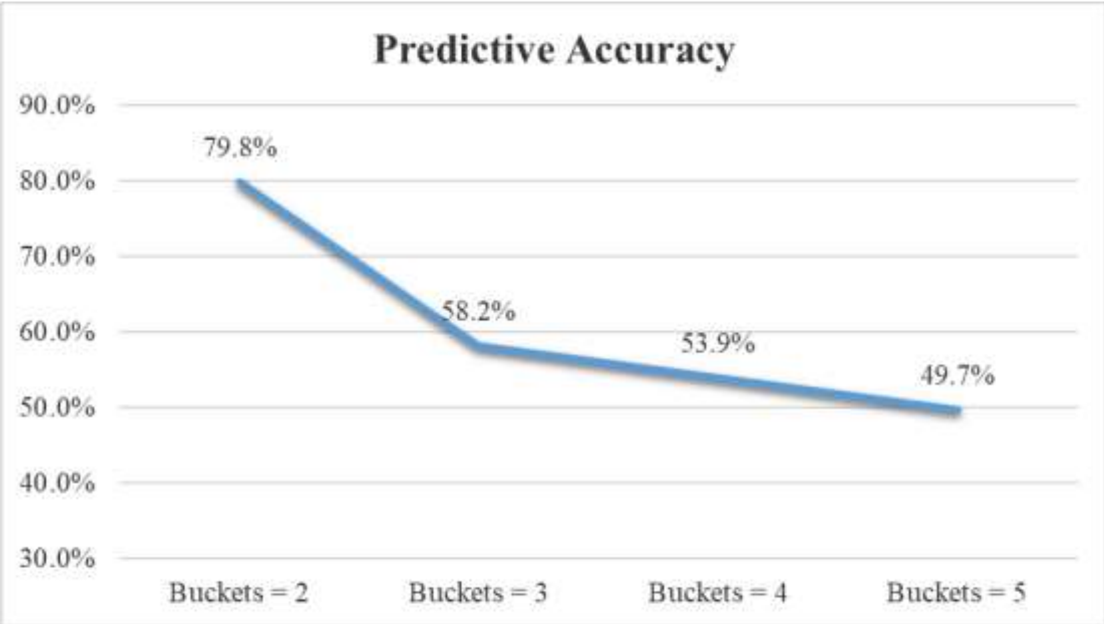

Fig 3: Predictive accuracy with varying bins

5. Future Research

The extraction code used in web crawling was able to collect jurisdiction data for 3,140 CIK out of 3,880 CIK having Exhibit 21 in 10 K report. The code mainly gathered data that was listed in HTML table format and remaining jurisdiction data for 740 CIK was not gathered thorough the approach used in the study. Web scraping the rest of the CIK jurisdiction data is possible but required greater time and effort in the phase of data collection.

During the course of the study, it was observed that the code script struggled to distinguish overlapping names such as Jersey and New Jersey. Having realized this as an area of opportunity, later attempted research in the area can elicit better prediction results if mindful of this highlighted observation.

The focus in this study was primarily on US companies from 2018. Training dataset could be improved with inclusion of jurisdiction data from companies such as Canada or Europe. This would increase the number of observations with companies belonging to various tax haven categories.

Company specific variables had plenty of missing values. In this study, bagged decision trees was used to impute the missing values. There is plenty of room to locate variables with lesser missing values and thus the impact of imputation minimal. In addition, this study did not focus on experimentation with various imputation methods or their impact on the prediction accuracy.

6. Conclusion

This study attempted to predict operations as well as intensity of operations of companies in tax havens. A novel dataset was extracted for that purpose which consisted of all the public listed US companies from 2018. The dataset extracted was right-skewed and almost 90% of the companies belonged to tax haven category one or two having less than 40% operations in tax haven jurisidiction.

In the first step, industry wide analytics was performed and instances of individual prediction error was located. All the companies in public administration had operations in the tax haven jurisdiction Switzerland, except CCUR Holdings. This raises some doubts and provides ample evidence for tax authorities to investigate the instance of tax evasion or incomplete filing of Exhibit 21.

Supervised machine learning techniques were used to predict operations and intensity of operations in tax havens. The study led to conclusion that in instances of aggregation or emission from declaring operations in tax havens, it is possible to predict operations in tax havens with 80% confidence.

Predicting intensity of operations in tax havens was tough with this dataset. Even though the prediction accuracy was around 70% but almost all the true predictions were from bin one or two. To address the skewed data set weighed bins were created that were dependent on number of companies instead of tax haven score. Prediction accuracy dropped to 50% but predictions came from all tax haven categories.

Unveiling disclosure aggregation in financial statements using machine-learning techniques and predicting operations in tax havens may help tax authorities in discovering tax evasion and other illicit activities. If tax authorities become capable of predicting operation in tax havens, multinational companies would no longer be incentivized to aggregate financial disclosures or hide operations in tax havens.

7. References & Appendix

7.1 References

Akamah, H., Hope, O., & Thomas, W. B. (2017). Tax havens and disclosure aggregation. *Journal of International Business Studies, 49*(1), 49-69

Dharmapala, D., & Hines, J. R. (2009). Which countries become tax havens? *Journal of Public Economics*, *93*(9-10), 1058-1068. doi:10.1016/j.jpubeco.2009.07.005

Dyreng, S.D., & Lindsey B.P. (2009). Using financial accounting data to examine the effect of foreign operations located in tax havens and other countries on U.S. multinational firms' tax rates. *Journal of Accounting Research 47(5)*

Fast Answers. (2009, June 26). Retrieved October 21, 2020, from https://www.sec.gov/fast-answers/answers-form10khtm.html

Gramlich, J. D., & Whiteaker-Poe, J. (2013). Disappearing Subsidiaries: The Cases of Google and Oracle. *SSRN Electronic Journal*.

Hope, O., Ma, M., & Thomas, W. B. (2013). Tax avoidance and geographic earnings disclosure. *Journal of Accounting and Economics*, *56*(2-3), 170-189.

Machine Learning: What it is and why it matters. (n.d.). Retrieved October 21, 2020, from https://www.sas.com/en_id/insights/analytics/machine-learning.html

Mara, E. R. (2015). Determinants of tax havens. *Procedia Economics and Finance*, *32*, 1638-1646.

Levy, P. S., & Lemeshow, S. (2008). *Sampling of populations: Methods and applications*. Hoboken, NJ: Wiley.

Richardson, G., & Taylor, G. (2015). Income Shifting Incentives and Tax Haven Utilization: Evidence from Multinational U.S. Firms. *The International Journal of Accounting, 50*, 458-485.

Saar-Tsechansky, M., & Provost, F. (2007). Handling Missing Values when Applying Classification Models. *Journal of Machine Learning Research*, *8*, 1625-1657.

Schjelderup, G. (2016). Secrecy Jurisdictions. *International Tax and Public Finance, 23*(1), 168-89

Search SIC Codes by Industry. (n.d.). Retrieved October 21, 2020, from https://www.naics.com/sic-codes-industry-drilldown

Services, W. (n.d.). Wharton Research Data Services. Retrieved October 21, 2020, from https://wrds-www.wharton.upenn.edu/

Stratified Sampling. (n.d.). Retrieved October 21, 2020, from http://essedunet.nsd.uib.no/cms/topics/weight/2/5.html

User, S. (n.d.). Retrieved October 21, 2020, from https://fsi.taxjustice.net/en/

Valdiviezo, H. C., & Aelst, S. V. (2015). Tree-based prediction on incomplete data using imputation or surrogate decisions. *Information Sciences, 311*, 163-181.

What We Do. (2020, February 19). Retrieved October 21, 2020, from https://www.sec.gov/Article/whatwedo.html

Why choose Singapore for Business: Rikvin. (2020, October 07). Retrieved October 21, 2020, from https://www.rikvin.com/incorporation/why-singapore-is-preferred-by-foreign-companies/

7.2 Appendix

List A (Secrecy/Tax Haven Score > 60)
Andorra (66.1), Anguilla (77.5), Antigua and Barbuda (86.9), Aruba (76)
Bahamas (84.5), Bahrain (77.8), Barbados (73.9), Belize (75.2)
Bermuda (73.1), Bolivia (80.4), Botswana (68.7), British Virgin Islands (68.7)
Brunei (84.1), Cayman Islands (72.3), Chile (61.6), China (60.1)
Cook Islands (74.6), Costa Rica (68.7), Curacao (74.8), Dominica (77.3)
Dominican Republic (71.6), Dubai (83.9), Gambia (76.6), Ghana (61.8)
Gibraltar (70.8), Grenada (77.1), Guatemala (73.1), Guernsey (72.5)
Hong Kong (71.1), Indonesia (61.5), Isle of Man (63.6), Israel (63.3)
Japan (60.5), Kenya (80.1), Labuan (71.9), Lebanon (72),
Liberia (79.7), Liechtenstein (78.3), Macao (68.3), Maldives (81.1)
Malta (60.5), Marshall Islands (72.9), Mauritius (72.4), Monaco (77.5)
Montenegro (63.2), Montserrat (77.5), Nauru (66.7), Netherlands (66)
Panama (76.6), Paraguay (84.3), Philippines (65.4), Puerto Rico (77.2)
Romania (65.5), Russia (64), Samoa (77.6), San Marino (64)
Saudi Arabia (69.9), Seychelles (75.2), Singapore (67.1), St. Kitts and Nevis (76.7)
St. Lucia (78.3), St. Vincent and The Grenadines (70), Switzerland (76.5), Taiwan (75.8)
Tanzania (73.4), Thailand (79.9), Trinidad and Tobago (65.3), Turkey (68)
Turks and Caicos Islands (76.8), Ukraine (69.2), Uruguay (60.8), US Virgin Islands (73.1)
Vanuatu (88.6), Venezuela (68.5)

Table 24: List A

7.3 Data Repository

The jurisdiction data, company specific variable data and “List A” can be found on the following GitHub repository: https://github.com/daniyalarif/Master_Thesis

A link to shiny web application can also be found that includes visualisation of the tax haven, tax haven frequency, and FSI score. A second visualisation is company specific and shows box plot of mean tax haven score for a public listed company on SEC in 2018.