Enhancing Regulatory Compliance by Using Artificial Intelligence Text Mining to Identify Penalty Clauses in Legislation

Nachshon Goltz
University of Waikato, School of Law
Hamilton, New Zealand
sean.goltz@waikato.ac.nz

Michael Mayo
University of Waikato, Computer Science
Hamilton, New Zealand
michael.mayo@waikato.ac.nz

ABSTRACT

As regulatory compliance (or compliance governance) becomes ever more challenging, attempts to engage IT solutions and especially artificial intelligence (AI) are on the rise. This paper suggest that regulatory compliance can be enhanced by employing an AI model trained to identify penalty clauses in the regulations. The paper provides the theoretical basis of machine learning for text classification and presents a two stage experiment of (1) training multiple models and selecting the best one; and (2) employing a sliding window detection in order to identify penalty clauses in regulation. Results benchmarked using an algorithm based penalties API suggests further development is needed.

CCS CONCEPTS

• **Artificial Intelligence** → **Regulatory Compliance**; Text Analytics
• **Identifying Penalties** → Legislation

KEYWORDS

Regulatory compliance, Artificial Intelligence, Text mining, Penalties, Machine learning

1 INTRODUCTION

Challenges in regulatory compliance are closely associated with the increased business opportunities resulting from globalization and Information Ccommunication Technology. [1] There is drastic increase in the regulatory requirements with which businesses must comply, not only in sheer number but also in complexity, confronting businesses with the need to adapt to a complex and evolving regulatory environment. [2] In such an environment, organizations are finding compliance with the numerous regulations to be very expensive [3] with estimated $1 US trillion spent worldwide on regulatory compliance.

Compliance is defined as ''adhering to the requirements of legal, industry, and organizational standards and codes, to principles of good governance, and to accepted community and ethical standards''. [4] Regulatory compliance is defined as the set of actives and policies in place in an enterprise to ensure the business activities required to achieve the business goals of the company comply with the relevant normative requirements. [5] Sadiq & Governatori define compliance as ensuring that business processes, operations and practice are in accordance with a prescribed and/or agreed set of norms. [6]

According to El Kharbili et al., [7] regulatory compliance consists of measures and directives, which are implemented by policies, internal controls and procedures and which are modeled for business processes. They identified three classes of compliance rules: regulations, information technology security standards, and quality standards.

When businesses are confronted with evolving and diverse regulatory requirements from multiple sources, it is often suggested that an integrated approach to Governance, Risk, and Compliance (GRC) is useful. [8] At the heart of integrated GRC is the adoption of a risk-based approach to compliance, where resources are allocated to the areas they are most needed based on risk levels.

Mahler [9] defines legal risk as a risk resulting from a set of facts that are assessed under a set of legal norms. By comparison, compliance risk is considered a risk resulting from a failure to comply with laws, regulations, rules, related self regulatory organization standards, and applicable codes of conduct. [4] A common denominator for both risks is legal compliance risk, a risk resulting from a failure to adhere to the requirements of the law.

At a strategic level, compliance is naturally related to the concept of risk. Noncompliant situations expose a company to risks that can often be mitigated. Risk mitigation is the actual driver for internal compliance auditing. The risk entity represents the risks a company wants to monitor; risks are associated with compliance requirements. [10]

KPMG, one of the four big global accounting companies claim that, “The top risk perceived by senior executives is the growing regulatory pressure from governments around the world. C-level executives in almost all industries say this, not just those in Financial Services, where companies are facing arguably the greatest regulatory challenge in their history.” [11]

The market value for compliance related software and services was estimated as over $32 billion in 2008 [12] whereas only the eGRC market size expected to grow from USD 19.42 Billion in 2016 to USD 38.00 Billion by 2021. [13] This boost in business investment is primarily a consequence of regulatory mandates that emerged as a result of events which led to some of the largest scandals in corporate history such as Enron [14], WorldCom (US) [15], HIH (Australia) [16], Societe Generale (France) [17] and above all the global economic crisis of 2008.

Since the global financial crisis of 2008, spending on compliance has broken new records: over $1 US trillion is spent worldwide on regulatory compliance. [18] Over one million people are employed around the world in the regulatory compliance industry. [19]

Several recent works have addressed the issue of (regulatory) compliance in the context of business process management, service computing and cloud computing domain. [20] The general idea is to determine whether the constraints (i.e., norms) imposed by some regulatory framework (ranging from acts, to regulations, to industry standards, to best practices and internal policies) are met by some IT systems. [21]

In order to enable an automated compliance checking, it is necessary that, besides a formal representation of process models, compliance requirements exist in an analyzable format. A formalized compliance requirement is a structural pattern, [22] which defines how the structure of a subsection of a process model has to look like in order to comply with the underlying rule.

Assessing compliance demands for an interpretation and translation of the requirements provided in natural language in an actionable rule description (especially in the case of principle-based regulations). [23] Therefore, artificial intelligence may be suitable for this task.

Artificial Intelligence is the theory and development of computer systems able to perform tasks that would otherwise require human intelligence. Example takes include visual perception, speech recognition, decision making under uncertainty, learning, and translation between languages. [24]

Venture capital investments in companies developing and commercializing AI-related products and technology exceeded $2 billion since 2011. [25] Leading players like IBM, [26] Google, [27] and Facebook [28] have invested heavily in developing their AI capabilities.

No doubt the alleged AI legal application that has received the most public attention is ROSS, a system supported by IBM’s Watson division, claiming to be a “junior associate” [29]. One of the co-founders of the ROSS team describes it as "[b]asically, what we built is a [sic] the best legal researcher available". [30] Even without being made available to the public nor presented in any public demonstration, ROSS has become a symbol of legal AI technology.

Another new application is Global-Regulation.com, [31] the world’s largest search engine of legislation. Global-Regulation.com makes extensive use of both Microsoft and Google’s machine translation to offer laws from China, Mexico and Spain, among many others, in English.

Other fields in which AI has been used within the legal profession are e-discovery (Recommind, Equivio - now part of Microsoft), forecasting outcomes of IP litigation (Lex Machina – now part of Lexis), providing fact and context-specific answers to legal, compliance, and policy questions (Neota Logic) and contract lifecycle software, including discovery, analysis, and due diligence (Kira Systems and KM Standards).

Given the challenging and complex regulatory structure, the cost of compliance and the presumed abilities of AI, the authors’ theoretical assumptions are:

(1) it is possible to train an AI model to identify penalty clauses in regulations; and,

(2) identifying penalty clauses in regulations, can enhance compliance.

This article will first briefly describe the field of machine learning for text classification; further, the proposed AI system will be presented; in the fourth part, an initial exploratory experiments will be described; in part five, a further sliding window detection experiments will be presented and finally, we will summarize our conclusions.

2. MACHINE LEARNING FOR TEXT CLASSIFICATION

In this research we take a machine learning approach to detecting compliance clauses in legislation. Machine learning is one of the dominant paradigms within AI and is concerned with learning from data. The particular style of machine learning we focus on in is classification, i.e. learning rules for assigning class labels to unlabeled examples.

In the classification approach to machine learning the user provides the system with a set of labeled examples $\{(X_1, Y_1), (X_2, Y_2), \ldots (X_n, Y_n)\}$ called the training data where each Xi is a fixed length numeric feature vector of dimensionality D. Each yi is an element of the set of class labels $C = \{c_1, c_2, \ldots, c_K\}$. The aim of a classification-based algorithms is to train a model that can accurately predict class labels for unlabeled examples X_{n+1}, X_{n+2},... etc, which are not present in the training data. The unlabeled examples are called test data. If the true labels for the test data are known (but withheld from the learning algorithm), then the accuracy of any given learning algorithm can be assessed and different algorithms for classification can be compared.

Hundreds of machine learning algorithms for classification have been developed. In the experiments described here, we select a representative sample of algorithms that are commonly used in research and practice, specifically the following:

- Naïve Bayes [32], an efficient algorithm for learning a conditional probability distribution $P(C|X_i)$ over the class labels. The algorithm’s efficiency and simplicity arises from its assumption of the conditional independence of the features, which is often an incorrect assumption but can still yield high classification accuracy.
- The C4.5 decision tree learning algorithm [33]. Decision trees have the advantage that they are often succinct and therefore interpretable by humans. A decision tree consists of a set of tests of feature vector values arranged into a tree structure, with predicted class labels positioned at the leaves of the tree.
- Sequential Minimal Optimization (SMO) [34,35], an algorithm for learning support vector machines. Support vector machines are state-of-the-art classifiers that learn the optimal hyperplane separating examples with

different class labels. When unlabeled examples are to be classified, the example's position with respect to the learned hyperplane is used to predict its class label.

- Random forests [36], an ensemble method in which multiple decision trees are learned, each tree being learned from different random projection of the training data. When new examples need to be classified, the decision trees vote on the class label, and the class label with the majority vote is predicted.

In this work, each training example represents a fragment of text and the set of class labels is {yes, no}, indicating whether or not the fragment contains a compliance clause. For example, one of the clauses labeled yes in our dataset [37] is:

"Section 8(1.2): Fines

A person who contravenes section 7.5 is guilty of an offence and liable

(a) for a first offence, to a fine of not more than $10 000, and

(b) for a 2nd or subsequent offence, to a fine of not more than $100 000".

Machine learning algorithms cannot be applied directly to text fragments like this because the classifier expects examples to be fixed length vectors numbers. We therefore must convert the text fragments into vectors before they can be further processed, and two common sets of approaches exist for achieving this: bag-of-words (BOW)-based approaches, and word2vec (W2V)-based approaches.

The BOW approach is a classical text processing approach in which the order of words in a document is ignored and only the presence/absence of a word is important. For example, the clause above can be converted into the vector shown in Table 1 using the BOW approach. Typically stop words such as "is" and "a" are removed before forming the vector.

contravenes	1
convicted	0
fine	1
guilty	1
liable	1
day	0
continues	0
subsequent	0
… etc …	…

Table 1: Example of a BOW representation of a text document.

One drawback of the BOW approach is that the dimensionality D of the training example vectors is equal to the number of distinct words in the set of training examples. Hence, for large and varied text datasets, the dimensionality can be quite large (in the hundreds or even thousands of words).

A more modern alternative to BOW is the recently proposed W2V approach [38,39]. W2V is inspired by the distributional hypothesis in linguistics [40] which proposes that words occurring in similar contexts have similar meaning. One W2V method uses a neural network to predict the context of a word (i.e. the words preceding and following the word) given the word. Because the neural network contains a hidden layer, the activation values on the hidden layer are used to predict the context whenever a word is presented to the network. Following the distributional hypothesis, words with similar meaning should therefore have similar activation values on the hidden layer of the network. These activation values can therefore be used to "represent" the words as points in a space whose size corresponds to the dimensionality of the neural network's hidden layer. Such a representation is called an embedding.

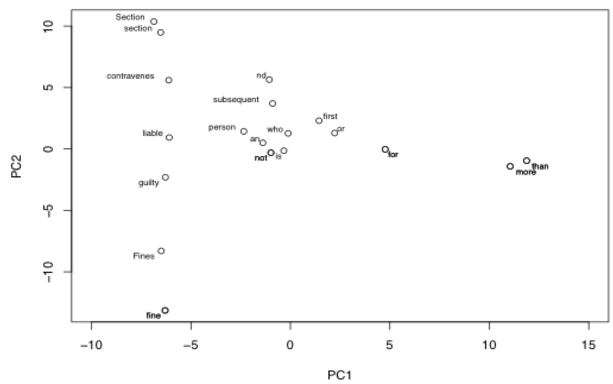

Figure 1: Words vectors from the previous example visualized by plotting their first two principle components.

Figure 1 illustrates word embedding for some of the words in example text fragment shown earlier. Directly visualizing points in a very high (i.e., 300) dimensional space is difficult, so shown in the figure are the first two principle components of the word embedding. The first principle component is measured along the x axis. It can clearly be observed that words related to compliance such as "Fines", "guilty", "liable" and "contravenes" lie close together along this component but are separated along the second principle component. Other pairs of words (e.g. "more" and "than") that are related are also close together in embedding space.

Because training a neural network to compute the word embedding for millions of words is computationally expensive (typically requiring billions of examples for robustness and therefore significant resources) we make use of a pre-trained set of three million word embedding provided by Google [39]. Prior research has shown that this set of embedding is surprisingly robust even when applied to text classification problems that are not news-related [41]. The dimensionality of each word embedding is 300 (significantly lower than that arising from the BOW method), and given all the word vectors in a document, a single "document vector" for an entire text fragment can be produced by a simple averaging procedure. This process is outlined in Table 2, and has been shown in several prior works (e.g. [42]) to produce robust and accurate feature vectors for classification.

Word	x_1	x_2		x_{300}
A	0.044241	0.089298	…	0.014781
person	0.120826	-0.108401	…	0.040490
who	0.050564	0.061801	…	0.031993
contravenes	0.062897	-0.111759	…	0.048083
… etc …	…	…	…	…
Average	0.016457	-0.033592	…	-0.002725

Table 2: Algorithm for computing a 300-dimensional representation of a document.

A disadvantage of both approaches to generating feature vectors from text is that they typically ignore any parts of the text that are non-alphabetical, such as punctuation symbols, formatting and numbers. This may impact on the performance of the system when the presence or absence of these features (e.g. fine amounts) in the text is important.

3. INITIAL EXPLORATORY EXPERIMENTS

Our initial exploratory experiments tested the theoretical assumptions brought in the introduction, according to which the use of machine learning will enable to identify compliance clauses in the regulations and hence, enhance compliance. Given the initial exploratory nature of this experiment, we avoided using analysis based on Formal Concept Analysis and Distributional Semantic Techiniques (such as Latent Semantic Analysis and Topic Modeling). These will be considered in a follow up research.

3.1 Curated Canadian Dataset

In order to test the hypothesis, we identified 40 negative and 40 positive penalty clauses within Canadian federal and provincial legislation (see the dataset here - https://github.com/mmayo888/MIREL2017Supplementary). The negative examples are clauses from laws or regulations that does not contain penalties. The positive examples are clauses taken from Canadian federal and provincial legislation and regulation that contains penalties. Both clauses contain dollar amounts.

	naïve Bayes	C4.5	SMO	Random Forest
BOW	97.50	92.50	96.25	97.50
W2V	91.25	78.75	100.00	92.50

For example, the following clause taken from the Canada Deposit Insurance Corporation Act, [43] is a positive penalty clause:

"50.1 Every member institution or other person who commits an offence under this Act is liable on summary conviction

(a) in the case of a natural person, to a fine not exceeding $100,000 or to imprisonment for a term not exceeding twelve months, or to both; or

(b) in any other case, to a fine not exceeding $500,000".

And a negative example would be the following clause taken from Alberta: Election Finances and Contribution Disclosure Act s 13(2) Exemptions: [44]

"Money or goods provided by any person that do not exceed $50 in the aggregate are not a contribution for the purposes of this Act but shall be recorded as to the gross amount by the chief financial officer of the recipient unless the donor specifically requests that the amount be considered a contribution".

3.2 Results

Given the dataset of examples described in the previous section, we next performed a set of experiments to determine which machine learning algorithm would classify the examples with the highest accuracy. As described in Section 2, we focused on four different machine learning algorithms (naïve Bayes, C4.5, SMO and random forests) along with two different methods for converting documents into feature vectors (BOW and W2V).

The approach taken to evaluate the machine learning in this initial set of experiments is leave-one-out-cross-validation, a methodology in which a given machine learning algorithm is trained on all but one of the labeled examples in the dataset (79 in this case) and tested on the single remaining held-out example. This train and test procedure is then repeated for each of the remaining 79 examples until each example has been tested, so that an overall accuracy metric for algorithm can be computed.

Note that since the single test example in each round is not present in the training data for that round, then the algorithm is always being tested on an example that it has not "seen" in its training dataset. Therefore, the outcome of the experiment should give an indication of how well each algorithm performs on unlabeled data.

To carry out our experiments, we used implementations of the algorithms available in Weka 3.8.0 (http://www.cs.waikato.ac.nz/~ml/weka/) with default settings. [45] As mentioned previously, the word vector data is obtained from Google (https://code.google.com/archive/p/word2vec/) after being trained on a corpus of news items. [39]

Table 3: Leave-one-out-cross-validation percent correct across the dataset of 80 examples, by feature vector type and machine learning algorithm.

The results of the experiment are summarized in Table 3. As can be clearly observed, the SMO algorithm in conjunction with W2V achieves the greatest performance with 100% accuracy. Unfortunately, all of the other W2V approaches underperform the BOW approach, especially the C4.5/W2V combination which achieves only 79%. The average accuracy however for all approaches is above 90%.

Curious as to the reason why the accuracy for most methods was so high, we next explored the accuracy of a very simple classification method that classifies an example as positive if it contains the word "fine" and negative otherwise. The accuracy of this simple approach is 96.25% on our dataset of 80 examples. This fact happens to explains why the BOW approach performs well compared to the W2V approach: the BOW feature vector has an explicit feature for the presence of the word "fine" (see Figure 1), and a classifier only has to detect this single important feature in order to perform well with high accuracy.

However, W2V-based algorithms do not have this shortcut – instead the word vector for "fine" is averaged into one overall document vector, as illustrated in Table 2. The fact that the SMO/W2V algorithm came out as the top performer therefore makes this approach all the more impressive.

Next, we decided to perform further tests of two of the algorithms on more challenging and realistic test data. The results of that experiment are described in the next section.

4. SLIDING WINDOW DETECTION EXPERIMENTS

In the next set of experiments, we took two of the classifiers used in the previous experiment and tested them on some more challenging and realistic test documents. The first classifier we tested was SMO/W2V, which was selected because it had the highest accuracy in the previous experiment. The second classifier we selected was random forest/W2V which, while not the most accurate method, is interesting to examine further because it gives reasonable probability estimates when it makes a prediction (e.g. 0.63 probability of a yes), whereas SMO typically does not produce proper probabilities.

The methodology we used for this experiment was as follows. Firstly, one of the machine learning algorithms (either SMO/W2V or random forest/W2V) was trained on all 80 examples in our labeled dataset. Next, we obtained a selection of further test documents - four Canadian federal and provincial Acts and one US. These documents had not been labeled. These documents were much longer than our training examples since they are complete acts rather than small fragments.

To account for the size disparity between the training examples and the longer test documents, we adopted a sliding window approach and considered only 50 words of each test document at a time. For each possible position of the sliding window, a prediction was made (yes or no) using both of our models. We then advanced the sliding window by 25 words (to ensure that each window would overlap) and repeated the process.

This meant that the number of predictions made per test document varied depending on its length, from 38 predictions (for the shortest document) to 109 predictions (for the longest document). [MM1]

4.1 Results

The Ferries Act [46] - This Act contain one penalty clause which was identified positively by both methods.

Motor Vehicle Fuel Consumption Standards Act [47] - Both methods missed the first penalty clause and identified it as negative. However, the second penalty clause was clearly identified by both methods as positive. Few additional clauses discussing 'inspector', 'order' and other enforcement related subject were identified positive by the SMO/W2V and scored relatively high (>=40) in the random forest/W2V.

The Partnership Act [48] - This Act does not contain any penalties. SMO/W2V yielded 10 positive results and the random forest/W2V yielded 14 results with yes probability equal or higher than 50%. Two positive results in the SMO/W2V yielded less than %50 in the random forest/W2V and two of the positive results (>=50%) in the random forest/W2V yielded negative results in the SMO/W2V. It is interesting to note that although this legislation did not contain any penalties, the majority of the clauses identified by both methods as positive dealt with related subject of liability.

The Pay Equity Act [49] - This Act does not contain any penalties. The SMO/W2V indicated three cases of positive penalty clauses whereas only the middle one was positive in the random forest/W2V as well. These clauses discussed related subject of court order.

In order to benchmark the results, we ran the same legislation, through Global-Regulation.com publically available algorithm based Penalties API (https://www.global-regulation.com/penalties.php).

With regards to the first Act (The Ferries Act), the algorithm based API reported that, "Found penalty clauses but not dollar figures". The penalty clause in this Act (positively identified by the sliding window method) contain the words 'dollar', 'two' and 'eight', but not the signs '$', '2' and '8'. This could be the reason for the algorithm based API to miss this clause while the sliding window correctly identified it. This was also the case with the second Act - Motor Vehicle Fuel Consumption Standards Act.

The partnership Act does not contain penalty clauses. However, both our experiment and the algorithm based API identified positive penalty clauses. The API was not specific and just reported: "Found penalty clauses but not dollar figures".

Finally, with the Pay Equity Act, the algorithm based API identified correctly that there are no penalty clauses while our experiment identified three cases of false positive penalty clauses (dealing with compliance relayed matters).

5. CONCLUSIONS

We have conducted two experiments with the purpose of using artificial intelligence text mining in order to identify penalty clauses in legislation.

From a practical perspective, the second experiment results are more important since the testing was done using 'real data' (i.e., legislation), whereas in the first experiment the samples were tested against each other.

While the results of the first experiment (BOW/W2V) were impressive with above 90% accuracy (see table 2), the results of the second experiment (Sliding Window Detection) showed less accuracy (see section 4.1). Nonetheless, the use of an algorithm based penalties API as a benchmark shows potential in further development of the AI based method.

In light of the lessons learned from this experiment, there is a need to consider replacing two basic elements used in the current method:

i) the basis for the current transformation system from words to vector was based on Google's news dataset. [39] Further development of the AI based method would benefit from creating a specific legal dataset that may generate more accurate transformation and thus improve the system overall accuracy;
ii) the current system used Google's W2V tool [38,39] to transform the text into a bag of word vectors. This tool as we have used it (pertained with from Google's new corpus) ignores punctuation makes and numbers, since it is only concerned with vectors for English words. This may not be adequate for the purpose of identifying penalty clauses where symbols such as dollar signs and figures may be important. Further development of the AI based method would benefit from using Facebook's 'fastText' system [50], which builds word vectors indirectly from n-gram character vectors and can therefore produce vectors for punctuation symbols and numbers properly.

Employing these elements along with a bigger dataset of examples may produce better results. Moreover, using the two methods, algorithm based system and AI based system, in tandem, could produce the most accurate results and hence enhance regulatory compliance.

ACKNOWLEDGMENT

The authors would like to thank Addison Cameron-Huff and Antonino Rotolo for their illuminating comments and Liam Whittaker for his research assistance.

REFERENCES

[1] Samson Esayas & Tobias Mahler, Modelling compliance risk: a structured approach, Artif Intell Law (2015) 23:271–300.
[2] Giblin C, Liu AY, Muller S, Pfitzmann B, Zhou X (2005) Regulations expressed as logical models (REALM). In: Proceedings on legal knowledge and information systems: the eighteenth annual conference, IOS Press, pp 37–48
[3] Ponemon Institute (2011) The role of governance, risk management and compliance in organizations study of GRC practitioners. Research Report, Sponsored by RSA, Security Division of EMC, Bedford
[4] Standards Australia (2006) Compliance programs AS 3806-2006, 2nd edn. Australia, Sydney
[5] Governatori, Guido, Business Process Compliance: An Abstract Normative Framework, Information Technology 55(6) (2013) 231-238 at 231.
[6] Shazia Wasim Sadiq, Guido Governatori: Managing Regulatory Compliance in Business Processes. Handbook on Business Process Management (2) 2015: 265-288
[7] El Kharbili, Marwane, Sebastian Stein, Ivan Markovic, and Elke Pulvermueller (2008): Towards a Framework for Semantic Business Process Compliance Management, CEUR Workshop Proceedings, 339: 1-15.
[8] Bonazzi R, Hussami L, Pigneur Y (2010) Compliance management is becoming a major issue in IS design. In: D'Atri A, Sacca` D (eds) Information systems: people, organizations, institutions, and technologies. Springer, Berlin, pp 391–398
[9] Mahler T (2010) Legal risk management: developing and evaluating elements of a method for proactive legal analyses, with a particular focus on contracts. Dissertation, University of Oslo
[10] Silveira, Patricia et al., On the Design of Compliance Governance Dashboards for Effective Compliance and Audit Management, January 2009, Conference: Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops - International Workshops, ICSOC/ServiceWave 2009, Stockholm, Sweden, November 23-27, 2009, Revised Selected Papers
[11] https://assets.kpmg.com/content/dam/kpmg/pdf/2013/08/expectations-risk-management-survey-v3.pdf
[12] Hagerty, J, J Hackbush, D Gaughan and S Jacobson (2008). The governance, risk management, and compliance spending report, 2008–2009: Inside the $32B GRC Market. AMR Research, Boston, USA, 25th Mar. 2008.
[13] Enterprise Governance, Risk, and Compliance (eGRC) Market by Component (Software (Type & Usage), Service), Deployment Model (Cloud & On-Premise), Organization Size, Business Function (Finance, IT, Legal, & Operations), & Vertical - Global Forecast to 2021, http://www.marketsandmarkets.com/PressReleases/enterprise-governance-risk-compliance.asp
[14] Enron's Collapse: The Overview; Enron Collapses as Suitor Cancels Plans for Merger, http://www.nytimes.com/2001/11/29/business/enron-s-collapse-the-overview-enron-collapses-as-suitor-cancels-plans-for-merger.html
[15] Worldcom's Collapse: The Overview; Worldcom Files for Bankruptcy; Largest U.S. Case, http://www.nytimes.com/2002/07/22/us/worldcom-s-

collapse-the-overview-worldcom-files-for-bankruptcy-largest-us-case.html

[16] HIH : The Inside Story Of Australia's Biggest Corporate Collapse http://www.smh.com.au/articles/2003/03/14/1047583693489.html

[17] Wikipedia, (On January 24, 2008, the bank announced that a single futures trader at the bank had fraudulently lost the bank €4.9 billion (an equivalent of US$7.2 billion), the largest such loss in history). https://en.wikipedia.org/wiki/Soci%C3%A9t%C3%A9_G%C3%A9n%C3%A9rale#cite_note-18

[18] http://www.blplaw.com/media/know-how/speed-of-business/BLP_The-Speed-of-Business-impact-of-regulation.pdf

[19] https://www.linkedin.com/vsearch/p?keywords=regulatory+compliance&queryRewriteParams=1

[20] Becker J, Delfmann P, Eggert M, Schwittay S (2012) Generalizability and Applicability of Model-Based Business Process Compliance-Checking Approaches – A State-of-the-Art Analysis and Research Roadmap. BuR - Business Research Journal 5(2):221–247; Fellmann M, Zasada A (2014) State-of-the-Art of Business Process Compliance Approaches. In: 22st European Conference on Information Systems, ECIS 2014, Tel Aviv, Israel, June 9-11, 2014, URL http://aisel.aisnet.org/ecis2014/ proceedings/track06/8

[21] Mustafa Hashmi, Guido Governatori, Moe Thandar Wynn:Normative requirements for regulatory compliance: An abstract formal framework. Information Systems Frontiers 18(3): 429-455 (2016)

[22] Ghose, Aditya and George Koliadis (2007): Auditing Business Process Compliance, in: Bernd J. Krämer, Kwei-Jay Lin, and Priya Narasimhan (eds.): Service-Oriented Computing (ICSOC 2007), Springer: Berlin et al., 169-180.

[23] C. Giblin, S. Müller, B. Pfitzmann. From Regulatory Policies to Event Monitoring Rules: Towards Model-Driven Compliance Automation. IBM Research Report RZ 3662, Zurich, October, 2006. 10. K. Namiri, N. Stojanovic. A Semantic-based Approach for Compliance Management of Internal Controls in Business Processes. CAISE'07, pp. 61-64, 2007.

[24] Oxford Dictionaries, "Definition of artificial intelligence," http://www.oxforddictionaries.com/us/definition/american_english/artificialintelligence.

[25] Demystifying artificial intelligence What business leaders need to know about cognitive technologies, Deloitte Uni. Press.

[26] IBM, "IBM Watson," http://www-03.ibm.com/press/us/en/presskit/27297. (IBM has committed $1 billion to commercializing Watson, its cognitive computing platform).

[27] Amit Chowdhry, "Google to acquire artificial intelligence company DeepMind," Forbes, January 27, 2014, http://www.forbes.com/sites/amitchowdhry/2014/01/27/googleto-acquire-artificial-intelligence-companydeepmind/. (Google has made major investments in AI in recent years, including acquiring eight robotics companies and a machine learning company).

[28] Josh Constine, "NYU 'Deep Learning' Professor LeCun will head Facebook's new artificial intelligence lab," TechCrunch, December 9, 2013, http://techcrunch.com/2013/12/09/facebook-artificial-intelligence-lablecun/ (Facebook hired AI luminary Yann LeCun to create an AI laboratory with the goal of bringing major advances in the field).

[29] University of Toronto's next lawyer: A computer program named Ross, http://www.theglobeandmail.com/report-on-business/industry-news/the-law-page/university-of-torontos-next-lawyer-a-computer-program-named-ross/article22054688/

[30] Addison Cameron-Huff, Why I'm Skeptical That Watson Will Be Your Next Junior Associate, March 2, 2015, https://www.cameronhuff.com/blog/why-i-am-skeptical-of-watson/

[31] www.global-regulation.com; disclaimer: Nachshon Goltz is the website's editor-in-chief

[32] G. John, P. Langley. Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345, 1995.

[33] J. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[34] J. Platt: Fast Training of Support Vector Machines using Sequential Minimal Optimization. In B. Schoelkopf and C. Burges and A. Smola, editors, Advances in Kernel Methods - Support Vector Learning, 1998.

[35] S. Keerthi, S. Shevade, C. Bhattacharyya, K. Murthy (2001). Improvements to Platt's SMO Algorithm for SVM Classifier Design. Neural Computation. 13(3):637-649.

[36] L. Breiman (2001). Random Forests. Machine Learning. 45(1):5-32.

[37] Canada Alberta Tobacco and Smoking Reduction Act 2.

[38] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. Technical report, arXiv preprint arXiv:1301.3781.

[39] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems 26 (NIPS 2013), 2013.

[40] S. Magnus (2008). The Distributional Hypothesis. Rivista di Linguistica. 20 (1): 33–53.

[41] M. Kusner, Y. Sun, N. Kolkin, and K. Weinberger. From word embeddings to document distances. In 32nd International Conference on Machine Learning (ICML), Proc. of, 2015.

[42] M. Iyyer, P. Enns, J. Boyd-Graber, and P. Resnik. Political ideology detection using recursive neural networks. In 52nd Annual Meeting of the Association for Computational Linguistics (ACL), Proc. of, pages 1113–1122, 2014.

[43] http://laws.justice.gc.ca/eng/acts/C-3/FullText.html

[44] http://www.qp.alberta.ca/1266.cfm?page=E02.cfm&leg_type=Acts&isbncln=9780779787241&display=html
[45] E. Frank, M. Hall, and I. Witten (2016). The WEKA Workbench. Online Appendix for Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, Fourth Edition, 2016.
[46] http://nslegislature.ca/legc/statutes/ferries.htm
[47] http://laws-lois.justice.gc.ca/eng/acts/M-9/FullText.html
[48] https://www.ontario.ca/laws/statute/90p05
[49] http://nslegislature.ca/legc/statutes/payequit.htm
[50] A. Joulin, E. Grave, P. Bojanowski, T. Mikolov, Bag of Tricks for Efficient Text Classification, arXiv preprint arXiv:1607.01759, 2016

[MM1] These numbers will change when the Canadian docs are added to the experiment