Binary translation of legislation and case law

-

A Deep Learning dive into the possibilities of Natural Language Processing for legislation and case law

Tax & Technology Pt. II

Tilburg University | 14 June 2019

Sylvia de Jong MSc LLM (U530095)
&
mr. Sander Derksen (U756840)

Abstract

Purpose – The purpose of this paper is to discuss the application of machine learning in the field of tax law, and specifically the possibilities of applying Natural Language Processing (NLP) techniques such as text and argumentation mining on the existing data from legislation and case law. The law and rulings of judges are currently written in human language. Is it possible and desirable to come up with a version that can be understood by a computer? What are the possibilities and what are the limitations?

Design/methodology/approach – Based on a literature review, the fundamentals of machine learning have been discussed. It is necessary to first understand the capabilities of machine learning in order to properly analyse the possibilities of its application in legislation and case law. This theoretical framework has led to Natural Language Processing, a field of Artificial Intelligence (AI) which combines machine learning techniques with linguistics. Based on theoretical findings and empirical research studies performed, the possibilities and challenges of for applying NLP techniques in legislation and case law will be discussed.

Findings – The authors recognize the usefulness of machine learning applications in law, but do not expect these to be operational on short notice. However, this field of study is rapidly expanding and creating various applications for different problems. Based on literature review, the decision tree analysis on legislation, combined with the text mining possibilities for case law, will create possibilities to make written human language quite understandable by computers.

Research limitations/implications – The research is solely based on a literature review. Future research may benefit from more survey research and data from real life applications of NLP techniques to be available for analysis. Given the fast development of tax and technology, a follow up of practices in the near future is recommended.

Practical implications – The ultimate objective of the implementation of machine learning techniques would be faster processing of legal cases, and efficient pattern recognition in existing data to apply the same treatment to similar cases. This paper will contribute to the understanding of machine learning applications in the field of tax law.

Originality/value – The research on integrating technology into tax practice is limited and existing literature lacks verification from practice. The gap between both disciplines will benefit from more research, especially in the light of rapidly increasing regulatory and technological developments. The combination of the tax and technology approach makes that this paper is an interdisciplinary contribution to existing academic literature.

Keywords – tax technology, artificial intelligence, machine learning, deep learning, natural language processing, text mining, argumentation mining, case law, legislation, jurisprudence, decision making

Table of Contents

Chapter 1. Introduction

The introduction contains a brief discussion about the background and relevance of the topic. The research gap will be defined, along with the research question and sub questions of this paper.

1.1 Topic Introduction

The current tax environment is characterized by the request for increasing transparency across jurisdictions, including broadened exchange of information and expanded compliance procedures. In addition, the amount of information and data keeps growing. To manage this changing landscape, tax authorities and tax advisors already started to explore the possibilities of data analytics and Artificial Intelligence in tax. Nowadays, Artificial Intelligence has been popping up as a commonly used buzzword when one speaks of Big Data possibilities. While data analytics already has been receiving a lot of attention, the application of Artificial Intelligence in the field of tax law is rather new. However, tax law professionals may benefit from the automation possibilities of artificial intelligence, for example when doing legal research (Sulea et al., 2017).

Machine Learning is a branch of Artificial Intelligence (AI), concentrated around the idea that when the machine is given an existing dataset, it will be able to extract knowledge from this data and by learning, improve its performance over time to ultimately be able to make predictions about future content (Surden, 2014). There are dozens of different Machine Learning methods. Machine Learning techniques have found applications in multiple fields of study, depending on the data it will process. When the machine is given access to a large volumes of tax data, tax codes, jurisprudence and case law, multiple possibilities for tax law may arise. The ability to make informed and useful predictions about potential legal outcomes could be considered one of the primary skills for a lawyer (Surden, 2014). As the quantity and quality of the data increases, so will the predictive accuracy of the machine.

In this paper, the application of machine learning in the field of tax law will be discussed and specifically the possibilities of applying machine learning techniques on the existing data from legislation and case law, for example to automate the legal classification of a contract or to predict the outcome of a current court case. Predicting the outcome of a court case is fundamentally historical and empirical (Brüninghaus & Ashley, 2003). Courts may have come to a decision in many more or less similar cases. Given the existing digital database of court cases, including the argumentation and considerations applied to derive a decision, it may be possible to apply machine learning techniques to induce general lines of reasoning for new cases and, ultimately, predict their outcomes. This paper will thus explore the possibilities of machine learning to support (tax) law professionals. In order to assess these possibilities, emphasis will be laid on Natural Language Processing (NLP), a field of AI which is heavily reliant on machine learning techniques. NLP includes text and argumentation mining approaches, which will be analysed for their applicability in the legal domain.

1.2 Research Questions

Based on the above, the core research question for this paper is as follows:

Research Question

What are the application possibilities for machine learning, and specifically Natural Language Processing approaches, in legislation and case law?

Sub Questions

1. *What are the capabilities of Machine Learning, and specifically Natural Language Processing?*
2. *What are the current applications of Natural Language Processing in legislation and case law?*
3. *What are the possibilities and limitations of the application of Natural Language Processing in law practice?*

1.3 Objective and Structure of the Paper

The ultimate objective for the implementation of machine learning techniques would be faster processing of legal cases, and efficient pattern recognition in existing data to apply the same legal treatment to similar cases. This paper will provide an analysis of the possibilities and desirability of having legislation and case law, written in human language, to be understood by computer algorithms. By doing so, this paper will contribute to the understanding of machine learning applications in the field of tax law. The gap between both disciplines will benefit from more research, especially in the light of rapidly increasing regulatory and technological developments. The combination of tax and technology approaches makes that this paper is an interdisciplinary contribution to existing academic literature.

In Chapter 2, a literature study will identify the theoretical framework and key concepts with regards to the application of machine learning techniques in legislation and case law. In Chapter 3, the current applications of Machine Learning, and specifically Natural Language Processing in legislation and case law will be discussed. The possibilities will be discussed and an overview of challenges of the application of NLP systems in the legal field will be given, based on the theoretical findings. In Chapter 4, the discussion of this research will be given, and the main considerations regarding the application of NLP techniques in practise will be discussed. The concluding Chapter 5 will provide the limitations and conclusions to this research, including answering of the core questions.

Chapter 2. Introduction to Machine Learning – Building the Theoretical Framework

In this chapter, the fundamentals of machine learning will be discussed. It is necessary to first understand the capabilities of machine learning in order to properly analyse the possibilities of its application in legislation and case law. First, the used terminology and general concepts will be elaborated on. Second, most common machine learning strategies will be touched upon to create the context to the functioning of artificial neural networks. After the elaboration on neural networks, the concept of deep learning will be introduced. One of the applications of deep learning is to be found in Natural Language Processing. With NLP, we approach the field of text mining and argumentation mining. This chapter will conclude with an overview of the different mining approaches which may be useful in the field of law.

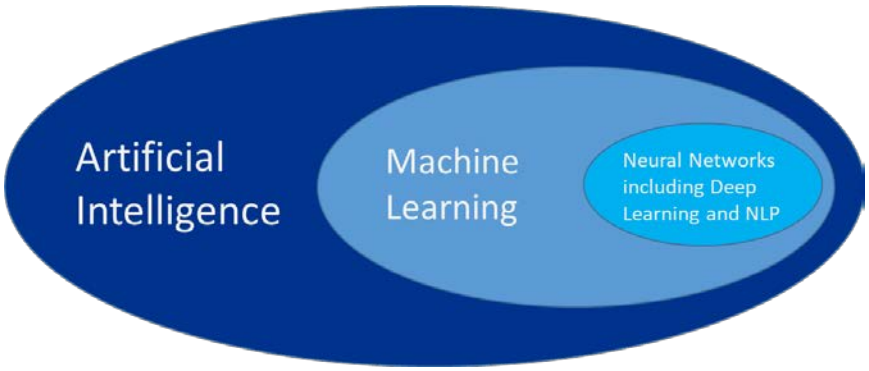


Figure 1. *Overview of the AI field*

2.1 Introduction to machine learning techniques

Big Data has become one of the day-to-day topics. When one speaks about Big Data, often the terms Artificial Intelligence (AI) and Machine Learning are used interchangeably. The concept of Artificial Intelligence is difficult to determine properly (Alarie et al., 2018). Artificial Intelligence may be defined as the name of the entire knowledge field, the extensive concept of machines being able to perform tasks in a smarter and faster way than humans beings could do. AI applications are typically combinations of numerous technologies, such as software, algorithms, statistics and cloud computing. AI includes both cognitive and machine learning, and robotic learning in which a machine mirrors the steps taken by a person to perform a certain task.

At the core of AI lies the growing technical field of machine learning. Where data mining is about finding patterns that help to make sense of existing data, machine learning is all about using available data to automate complex tasks or to make predictions on new incoming data (Surden, 2014). It addresses the question of how to build computers that improve automatically through experience (Jordan & Mitchell, 2015). There are various methods to apply machine learning and the application of data intensive machine learning methods may nowadays be found throughout various fields of study, such as science, technology, education and law.

2.2 Machine learning strategies

The classical machine learning techniques are rooted in statistics. Mathematical tasks and engineering problems were solved by searching for patterns in numbers. Much of the research in the field of

machine learning has focused on pattern classification. The main machine learning strategies applied for this purpose may be classified as supervised, unsupervised, and reinforcement learning (Sathya & Abraham, 2013). In the next subparagraphs, these paradigms will be shortly discussed.

2.2.1 Supervised learning

The most widely used machine learning methods are supervised learning methods (Jordan & Mitchell, 2015). Supervised machine learning can be used to predict a certain category (i.e. classification via Naïve Bayes, Decision Trees or Logistic Regression) or to predict a number (regression analysis) on unseen data (Allahyari et al., 2017). For example, supervised learning is when the algorithm is provided with a series of emails that a human labelled as spam, the algorithm may learn the characteristics of spam mail by analysing the samples (the 'training set') (Surden, 2014). The data is numerical or already manually categorized by its features. The machine will then use existing examples to learn. In supervised learning, the data has available both the inputs and outputs.

2.2.2 Unsupervised learning

Unsupervised learning methods are techniques trying to find hidden structure out of unlabelled data (Allahyari et al., 2017). These methods have to work with the input provided. In contract to supervised learning, there are no explicit target outputs associated with the input (Dayan et al., 1999). Unsupervised learning may result in clustering (data divided by similarity), association (by identification of sequences), or generalization. In this case, the machine will try to find any patterns on its own based on only the input. They do not need any training phase. The larger class of unsupervised learning methods consists of maximum likelihood (ML) density estimation methods. In such an estimation, it is included that the output should convey most of the information in the input, and that they should report on the underlying causes of the input (Dayan et al., 1999).

2.2.3 Reinforcement learning

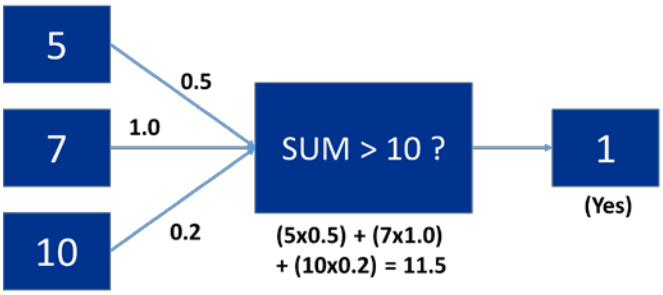
Reinforcement learning is used in circumstances where a problem is not related to the data. It is about an agent interacting with its environment. The goal is for the agent to learn the behaviour that maximizes the expected cumulative reward over time (Lison, 2015). In reinforcement learning, the information available in the training data is intermediate between supervised and unsupervised learning (Jordan & Mitchell, 2015). The training data will only provide an indication if an action is correct or not and when the action is incorrect, the problem remains to be solved for the right action. The agent will thus need to gather its own experience, receiving inputs and trying out different actions. For example, a self-driving car will also need experience, not only knowledge of the road rules and recognition of traffic signs. Reinforcement learning takes place in a virtual learning environment. In the car example, a virtual city may be built to let the self-driving car learn.

2.3 Neural Networks

The machine learning processing models rely on the concept of (artificial) neural networks. Neural networks are adaptive learning systems that give answers to problems after having been trained for similar problems (Krizhevsky, 2012). It is a collection of neurons, and connections between them. Its task is to take all numbers from the input, perform a function on these numbers, and provide the results in the output. In essence, it works on calculating probabilities. Connections are necessary to connect the output of one neuron with the input to another. Each connection has only one

parameter/weight. This weight tells the neuron to respond more to one input and less to another. A simplification of such a process is depicted below:

Figure 2. *A simplified representation of a one layer neural network*



During the training, weights may be adjusted. This training is necessary to develop the neural network learning system. However, problems that are solved via neural networks are usually more complex than the aforementioned example. The neurons are linked by layers. Within a layer, neurons are not connected, but they are connected to the neurons of the next and previous layers. For example, when a certain input is being applied, the associated neurons will be activated, which will activate the next layer, and so on until the output is ultimately reached. These networks are usually represented as matrices and calculated via matrix multiplication.

Neural networks are key in teaching computers to think and understand like human beings. Neural networks are complex algorithms that operate in a way similar to the human brain's decision making process to solve problems and to model relationships in data. For example, a neural network can be taught to recognize images and classify them according to elements they contain. An example of this is the Google reCAPTCHA (the abbreviation of Completely Automated Public Turing test to tell Computers and Humans Apart), which is the feature across the web to verify whether visitors are robots or human begins. Each time one completes a certain security task (e.g. 'select all images containing street lanterns'), it is training Google's machine learning datasets.

2.4 Deep Learning

A method to build, train and use large neural networks is via deep learning, which can be considered a new architecture (Skymind, 2019). When the structure of the system consists of more than two layers of neural networks, it is referred to as a deep neural network. Where machine learning makes use of an input of which it extracts features in order to perform a classification and ultimately an output, deep learning combines the process of feature extraction and classification. Deep learning is based on supervised learning. The feature extraction is done by the computer instead of manually labelling. Deep networks are considered a high-impact area of progress in supervised learning in recent years (Jordan & Mitchell, 2015).

The outcome of a deep neural network system is primarily data centric, meaning that the outcome is rather a reflection of the data the network is trained against than the outcome of a designed algorithm. In other words, the outcome is a reflection of the patterns that are recognized by the machine in the data used for the training. The two most applied kinds of deep learning are Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).

- ***Convolutional Neural Networks*** are used to search for certain objects in photos and videos, and to apply face recognition and enhancing images. This is a generally applied technique in

mobile phones nowadays. However, where text may be split into sentences and word attributes, images have to be labelled to detect a certain object. The machine will need to learn the certain features of an object and does this by dividing images in smaller pixel blocks to analyse them against other pixel blocks.

- ***Recurrent Neural Networks*** are best used for sequential data such as voice memos, text or music fragments. How far this has progressed may be visible from digital voice assistants such as Apple's Siri, who is able to speak whole phrases with even the right accents. RNN uses text as input and the accompanying audio as the desired output.

Deep neural network methods are being actively pursued in a variety of additional applications. In the following paragraph, the concept of Natural Language Processing will be discussed.

2.5 Natural Language Processing

Katz et al. (2017) noted that the combination of both human intelligence and computer based analytics will likely prove superior to that of human analysis alone for a variety of tasks which involve prediction of legal cases. To this end, Natural Language Processing (NLP), another field of AI, has been an upcoming field over the previous years.

Where AI is used to build intelligent system and Machine Learning is used for building systems that can learn from experience, a next step in this concept is to combine AI with linguistics to automate identification of what the meaning of a certain text or selection of words is. NLP examines the use of words and phrases to draw connections within and across written or spoken language (Alarie et al., 2018). Within the field of NLP, computational algorithms are built to automatically analyse and represent human language. NLP applies to both retrieving information (e.g. determining the relevance of a certain text or document) and extracting information (e.g. identifying the key terms from a text or document) (Alarie et al., 2018). It looks for structure by tagging words or drawing connections between references.

NLP is heavily reliant on machine learning techniques, especially the supervised learning methods as previously discussed (Surden, 2014). A NLP system based on machine learning creates a system that is able to learn how to understand language. Machine learning is used to help the machine understand the nuances in the human language, and to learn the NLP application to respond in a manner that is comprehensible. NLP tools have largely replaced the traditional keyword searches (Alarie et al., 2018). The idea behind NLP applications was to find a way to interact and communicate with electronic devices – for example, let Siri or a chat bot respond when a question is being asked.

2.5.1 Text Mining and Argumentation Mining

The most common machine learning techniques for text analysis include summarization, classification, language identification, feature selection and clustering. Text mining approaches are related to data mining (Allahyari et al., 2017). Text mining is all about finding meaningful patterns in natural language text. Many of the text mining algorithms extensively make use of NLP techniques, such as part of speech tagging (POG), syntactic parsing and other types of linguistic analysis (Allahyari et al., 2017). Text mining is one of the most common applications in NLP, ranging from plagiarism detection to estimating the period in which a text was published (Sulea et al., 2017). Textual data is roughly 80% of all potentially useful data of businesses (Gharehchopogh & Khalifelu, 2011). Text analyses, data mining

combined with machine learning, provides the possibility to eliminate or at least significantly reduce the need for expert human resources for manual analyses of textual information.

However, text mining on a large collection of documents may be considered a complex task. Therefore, it is of critical importance to have a data structure for text which facilitates further analysis (Allahyari et al., 2017; Sulea et al., 2017). The most common way to represent documents is as a bag of words (BOW), which considers the number of occurrences of each word or phrase, but ignores the order (Allahyari et al., 2017).

Text mining is often used for classification purposes. Based on a data set consisting of written or spoken language, the computer is trained to be able to find meaningful patterns in the data. In order to make the computer understand written or spoken language, feature extraction, feature selection and the classification algorithm are of significant importance (Allahyari et al., 2017). The pre-processing of the textual components may influence the success of the classification process. It is of viable importance to select the right features, i.e. the variables or parameters the machine needs to look at.
A first step in the pre-processing stage before actual text mining takes place may be tokenization. Tokenization is breaking up a continuous sequence of language into smaller pieces, e.g. words or phrases, which are called tokens. The second step is to take a token and relate it to its base concept, which is referred to as stemming. The grouping the various forms of a words (tokens and roots) together to analyse them as a single item is referred to as lemmatization.

The availability of ever-growing amounts of information and data in which arguments can be found, together with the advances in linguistics and machine learning, gave rise to the research area of argumentation mining (Lippi & Torroni, 2016). Argumentation mining aims at automatically extracting structured arguments from unstructured textual documents, to ultimately provide structured data for computational models or argument and reasoning engines (Lippi & Torroni, 2016). The most complex stage of this process is to ultimately predict the links between the arguments, as it requires the algorithm to understand relationships and connections between the detected arguments (Palau & Moens, 2009). Argumentative sentence detection is fundamentally a sentence classification task (Lippi & Torroni, 2016), a supervised machine learning strategy. The second step in argumentation mining would be the component boundary selection, and the final step the prediction of the argument structure. Law has been the pioneering domain in the application for argumentation mining (Lippi & Torroni, 2016).

2.5.2 Sentiment Analysis

Since early 2000, sentiment analysis has grown to be one of the most active research areas in NLP. It is also widely studied in data mining and text mining (Liu, 2012). The growth of this field of study is in line with the increasing recordings of opinionated data such as reviews, social media posts and online discussions. Sentiment analysis is a field that build systems that try to perform text classification based on the automated process of understanding an opinion about a given subject from written or spoken language. Sentiment analysis has many practical implications, as it may structure the data on any topic one can express an opinion about. Sentiment analysis can be modelled as a classification problem. Sentiment analysis may be applied on different levels: document level, sentence level and even sub-sentence level. Algorithms to implement a sentiment analysis system may be rule-based, based on

machine learning techniques to learn from data, or a hybrid version of the previous two system designs.

2.6 Chapter Summary

In this chapter, an overview has been provided of the various possibilities within the field of machine learning. First, the most common machine learning techniques have been discussed (supervised, unsupervised, reinforcement). Second, a more in-depth paragraph has been included on the functioning of neural networks, in order to introduce the deep learning concept. This theoretical framework has led to the field of natural language procession, a field of AI which combines machine learning techniques with linguistics. In the light of the research question, emphasis has been put on text mining and argumentation mining as concepts which have been identified as useful techniques to support law practitioners. Also, sentiment analysis has been touched upon. In the next chapter, the application of NLP in the law domain will be explored more in-depth.

Chapter 3. Machine learning in the legal landscape

In Chapter 3, the current application of Machine Learning, and specifically Natural Language Processing in legislation and case law will be discussed more in-depth. First, the theoretical framework from chapter 2 will serve as the basis to provide the possibilities for applying NLP techniques in the case and legislation law domain. Second, the challenges for these applications will be discussed. In summary, this chapter will provide an overview of strengths and weaknesses of machine learning systems in the legal field, based on theoretical findings and the empirical research studies performed.

3.1 The application of NLP techniques on case law

Already in 1963, Lawlor presumed that one day, computers would become able to analyse and predict the outcomes of judicial decisions. Nowadays, the law library has long been replaced by electronic search mechanisms. Applications of this nature already have entered the marketplace, and are likely to improve over the next years. It is already possible to find cases and material related to them using NLP, allowing researchers to perform question answering (which is the process of producing accurate answers to questions posed by humans in a natural language) or a coherent summarization of the relevant information (Allahyari et al., 2017).

3.1.1 The possibilities: pattern recognition in case law

Current developments in the field of machine learning and NLP provide for tools which may be used to build predictive models that can be applied to unveil patterns driving judicial decisions (Aletras et al, 2016). In the environment of growing complexity in regulatory frameworks and legal rulings, it may be very useful for both lawyers and judges to identify cases and extract the patterns that have led to a certain decision. To support legal decision making, relevant information is often retrieved by searching through both legislation and case law. As the judgments of the Court have a distinctive structure, this makes them particularly suitable for a text-based analysis (Aletras et al., 2016). Automated text mining tools that can perform information extraction from the existing case law have many advantages, as it is less time consuming compared to a manual search through historical documents (Wyner et al., 2010). A certain model may be simplified as follows:

Figure 3. *Simplified model of current pattern recognition in case law*

In 2004, scholars from the Washington University tested their statistical model on every argued Supreme Court case (628) in 2002. The model predicted 75% of the Court's results correctly, while the legal experts collectively forecasted 59.1 % right (Ruger et al., 2004). These results were to be considered notable, given that the statistical model disregarded information about the specific law or facts of the cases, but mostly took into account the political stances of the judges. In 2015, Katz et al., (2017) also constructed a model designed to predict the behaviour of the Supreme Court of the United States, but included a time-evolving random forest classifier that leverages unique feature engineering to predict more than 240,000 justice votes and 28,000 cases outcomes over the timespan from 1816 to 2015. Their research obtained 70.2% accuracy at the case outcome level and 71.9% at the justice vote level.

In 2016, Aletras et al. analysed the cases tried by the European Court of Human Rights, solely on textual content. In this study, the classifiers have been the textual content from the case and the actual judgment has been used as the target output. Their models have been able to predict the Court's decision making with a 79% accuracy. As the scholars have based their model primarily on textual information, this study stipulates the importance of formal facts, as these have been the most predictive factor.

Pattern recognition may also be used by advisors and advocate-generals when preparing for a case. This will contribute to the opinion of the advocate-general or the advice from a lawyer or advisor, as they may be able to predict the outcome of a case based on precedents. Another interesting argumentation analysis that may be performed in order to increase accuracy of predictions is to investigate to what extent the reasoning of the advocate-general has been followed by the judge. Especially in countries with legal systems based on common law, for example the United Kingdom, case law is of vital importance in legal reasoning and judicial decisions.

3.1.2. The challenges of applying NLP techniques in Case Law

The most important challenges in the application of NLP techniques on case law are the unstructured nature of law data, the possible rise of machine learning bias and the occurrence of black box situations.

Unstructured data

A central challenge for using AI applications in law is the unstructured nature of legal data (Alarie et al., 2018). Text data is typical unstructured information, which is easily processed and perceived by humans, but significantly harder for machines to understand (Allahyari et al., 2017). Evaluating if a current case fits into the existing set of published court decisions may give rise to serval challenges, e.g. certain legal issues may have been shaped over time by numerous judicial decisions. Moreover, while opinions often follow a certain structure, the application of case law may be based on exceptions or based on a specific fact pattern. Moreover, the different judicial instances may have their own way of presenting information (Alarie et al., 2018). This may be an obstacle when an algorithm needs to extract the relevant information.

Machine learning bias

An important challenge in the application of machine learning algorithms, is that these algorithms may become biased. Bias in machine learning may occurs when an algorithm produces results that are systematically prejudiced due to erroneous assumptions in the machine learning process. In other words, the machine learning method is unable to capture true relationships. A large part of the bias may also be related to the data. The data set could lack diversity and be limited, which can result in a dataset which does not reflect the entire population.

If the data has inherent biases, the machine learning model will not only learn those biases but will end up amplifying them. With biased data, the outcomes of the model will end up in self-fulfilling prophecies, which in many instances can lead to unwanted outcomes.

Black Box situations

A weakness of the application of machine learning on case law, may be that this form of analysis does not include the legal argumentation and reasoning how the fact pattern has led to a certain outcome. That this is often not incorporated into the analysis is apparent from previous work on predicting judicial decisions, which has largely focused on the analysis and prediction of judges' votes given political or economic circumstances (Aletras et al., 2016).

This weakness is due to the complexity of the system. Therefore, in deep learning situations, it is often unclear how the system uses the data to reach a certain conclusion. The mystery regarding to the how or why the system has reached a certain conclusion is referred to as a black box, in the sense that one cannot inspect a trained model to say which input features were most important. This is contrasting with a regression model or decision tree where the learned coefficients or branches provide that information.

A black box situation occurs when the system finds and uses relations and patterns in the underlying data that are unknown to the developers and users to come to a conclusion. In the end, the essential process that the machine uses to come to the conclusion, is thus unknown. This use of deep neural network is therefore unsuitable for cases in which the decision making process (i.e. the why) that leads to a conclusion is of vital importance.

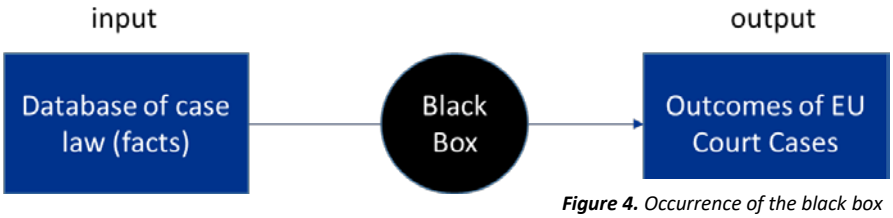


Figure 4. *Occurrence of the black box*

The ideal scenario would include analysis of the underlying legislation/legal framework to substantiate the generated output. This desired flow may be visualized as follows:

Figure 5. *Desired simplified model for legislation analysis*

3.2 The application of NLP techniques in legislation

When applying NLP techniques to derive patterns from existing case law, will be difficult to extract the specific features of the applied legislation and the accompanying argumentation, as there are a lot of exceptions in the application of law and each jurisdiction applies its own legislation, which makes it difficult for a machine to perform tax mining and to derive patterns. Therefore, the possibilities of a NLP approach which involves decision tree analysis when analysing legislation will be discussed in the following paragraphs.

3.1.2 The possibilities: decision tree analysis for legislation

When an analysis is based on text, it is difficult to derive argumentation structures via text mining possibilities. This challenge has been described in the previous paragraph. However, to include an

analysis of the applicable legislation, it may be possible to apply decision tree methodology. A decision tree is a hierarchical tree of the training set, in which a condition on the attribute value is used to divide the data hierarchically (Allahyari et al., 2017). A decision tree partitions the training data set into smaller subdivisions based on a set of test defined at each node or branch. This may be compared to the analysis of legislation, which is also often hierarchically structured (e.g. only if sub article 1 applies, then one may proceed to sub article 2).

Another example would be the analysis of contracts, which are also prepared based on a standard format. This structure and wording may give an indication of the type of contract that is being analysed. When such decision tree analysis is being included, this opens possibilities to binary translation of legislation. Reference may be made to the explanation of the neural network in paragraph 2.3, where the input is considered the criteria of a certain article of the law, the neuron performs the test if the article is applicable, and the outcome will be a binary yes or no. In that sense, decision tree analysis will map the decision process of reaching a certain outcome. In this process, the features and the accompanying labels that are applicable are of vital importance.

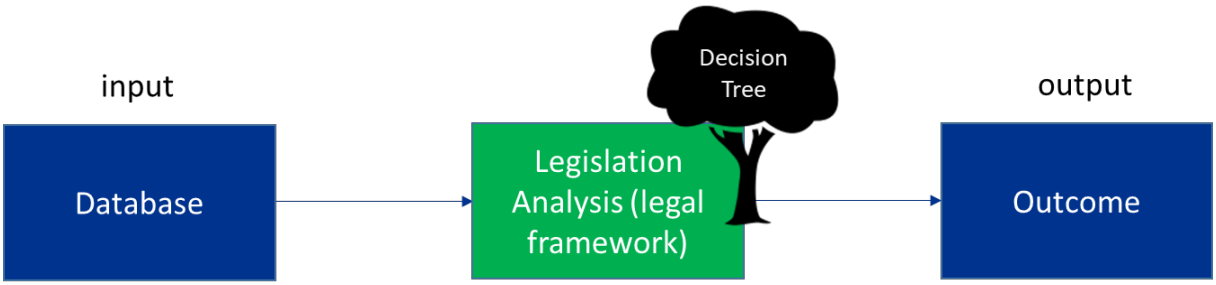


Figure 6. *Decision tree included analysis model*

Inclusion of decision tree analysis will diminish the risks of a black box occurrence, as long as the decision tree is able to provide a clear overview of its structure.

3.2.2 The challenges of applying decision tree analysis on legislation

The most important challenges in the application of NLP techniques on legislation is that the application of decision tree analysis will always lead to a binary outcome and that the complex nature of legislation makes it difficult to derive a suitable decision tree.

The answer is always binary

A computer does not know anything other than binary. This means that any word or character is encoded by the computer in a binary string. Therefore, the output of a computerized analysis will always be binary. However, in the field of law, the answer often is not black or white, but either a shade of grey. Therefore, for the computer to approach the human interpretation of facts and circumstances, sentiment and applicable legislation, includes multiple levels of feature analysis and labelling of concepts. This is an extensive operation and although its results will aim to mimic human decision making, it will not be able to completely emulate it. Adding the case law dimension could reduce this grey area, but not entirely.

Complex sentences

In addition, a general challenge for the NLP system is to understand what a certain sentence means. Tax documents – especially legislation - often have a high level of complexity, consisting of complex sentences, exceptions and exceptions to the exception, references to other parts or documents and

lengthy lists of conditions and qualifiers. In order for a NLP system to understand the meaning of certain text, it needs to learn through processing the meaning of internal and external references and tax data. The more tax documents are integrated in the system, the more the depth of the meaning will grow. In order to do so, the meaning of certain text fragments needs to be identified by labelling them – which is an extensive manual task – and secondly, an algorithm needs to be deployed to recognize similar meanings in references and relating subjects. Finally, an algorithm needs to combine and/or structure the hierarchy of related meanings. Detailed document searches could be enabled by parsing results, in which similar phrases and meanings may be associated.

3.3 Chapter Summary

In this chapter, the application of NLP in the law domain has been explored more in-depth. A distinction has been made in the application of NLP techniques for case law and legislation, as the suitable approaches may differ. The possibilities for applying NLP techniques in the case law and legislation domain have been discussed, accompanied with the possible challenges in applying these techniques. The following chapter will be the concluding chapter of this paper, consisting of the conclusion, practical implications, limitations and suggestions for further research.

Chapter 4. Conclusion, Limitations and Further Research

In this section, the importance of the study and the main research findings will be discusses. The theoretical and practical implications of these findings will be elaborated on, including the limitations and suggestions for further research. This section will end with the conclusions.

4.1 Main Findings and implications

As machine learning has found its application in multiple fields of study, it may also provide possibilities in the analysis of case law and legislation. The ultimate objective of the implementation of machine learning techniques in law would be faster processing of legal cases, and efficient pattern recognition in existing data to apply the same legal treatment to similar cases. In the previous chapters, the theoretical possibilities of applying machine learning techniques, specifically NLP approaches, were discussed.

This paper will contribute to the understanding of machine learning applications in the field of tax law. The authors recognize the usefulness of machine learning applications in law, but do not expect these to be operational on short notice. However, this field of study is rapidly expanding and creating various applications for different problems. Based on literature review, the decision tree analysis on legislation, combined with the text mining possibilities for case law, will create possibilities to make written human language quite understandable by computers.

4.2 Limitations and Suggestions for Further Research

Machine learning applications may provide various possibilities for the field or tax law. However, at this moment, certain limitations must be taken into consideration as well. First, the human factor is being dismissed when an intelligent automation system will prepare outcomes. Where in practise the reasoning of the judges is very important, this argumentative structure mining is a difficult exercise, meaning that the logic of the NLP system is not transparent. Also, where a computer will always provide a binary result (say yes/no), the outcomes of law cases are often not so clear. When decisions will be made based solely on NLP systems, this thus could lead to an unjust outcome and may diminish trust in legal certainty. Moreover, the risk of machine learning biases in the data and the black box will still exist. The decision tree approach may reduce the risk of the black box scenario, but only to the extent that the tree is well-structured.

This research has been based on a literature review solely. Future research may contain survey research and more data from real life applications of NLP techniques to analyse. Moreover, to determine the practical application of machine learning on case law and legislation, future study could benefit from implementing the suggested decision tree approach. This should be tested. In addition, given the fast development of tax and technology, a follow up of practices in the near future is recommended.

4.3 Conclusions

To conclude this paper, the research question will be answered. So, is it possible and desirable to come up with a version of the written law and rulings or judges which can be understood by computers? First, the authors are of the opinion that it is desirable in the field of law to be able to have written

language being analysed by computers. The data being generated keeps increasing and machine learning techniques will be able to find patterns in all these data. It may also benefit the process of doing legal research for both lawyers and judges.

Secondly, the authors conclude that this is possible by applying NLP techniques, specifically techniques which perform text/argumentation mining. The theoretical framework substantiates the possibilities of performing supervised deep learning methods on texts such as legislation and case law. The authors have made a distinction in the analysis of case law, which has been done in the past by other scholars by primarily looking at the facts and patterns, and analysis of legislation, which is suitable for decision tree analysis due to its hierarchical structure. The limitations to NLP is that it will try to mimic human decision making, but will not be able to emulate this process completely. Factors that make this more difficult are that due to the complexity of the cases, the reasoning for a certain outcome may often be unclear or non-transparent (black box).

To make a computer understand written language as best as possible, considered that the output will always be binary, the most suitable approach will be the combination of both analyses.

Chapter 6. List of References

Alarie, B., Niblett, A., & Yoon, A. H. (2018). How artificial intelligence will affect the practice of law. *University of Toronto Law Journal*, *68*(supplement 1), 106-124.

Aletras N., Tsarapatsanis D., Preoţiuc-Pietro D., Lampos V. (2016). Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective. *PeerJ Computer Science* 2:e93

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). *A brief survey of text mining: Classification, clustering and extraction techniques.*

Altman, E. R., Ebcioglu, K., Gschwind, M., & Sathaye, S. (2001). Advances and future challenges in binary translation and optimization. *Proceedings of the IEEE*, 89(11), 1710-1722.

Brüninghaus, S., & Ashley, K. D. (2003, June). Predicting outcomes of case based legal arguments. In *Proceedings of the 9th international conference on Artificial intelligence and law* (pp. 233-242). ACM.

Dayan, P., Sahani, M., & Deback, G. (1999). Unsupervised learning. *The MIT encyclopedia of the cognitive sciences*.

Gharehchopogh, F. S., & Khalifelu, Z. A. (2011, October). Analysis and evaluation of unstructured data: text mining versus natural language processing. In *2011 5th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1-4). IEEE.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

Kantardzic, M. (2011). Data mining: concepts, models, methods, and algorithms. John Wiley & Sons.

Katz DM, Bommarito MJ II, Blackman J (2017) A general approach for predicting the behavior of the Supreme Court of the United States. *PLoS ONE 12(4):* e0174698.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* (pp. 1097-1105).

Lawlor, R. C. (1963). What computers can do: Analysis and prediction of judicial decisions. *ABAJ*, *49*, 337.

Lippi, M., & Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, *16*(2), 10.

Lison, P. (2015). *An introduction to machine learning.*

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, *5*(1), 1-167.

Marr, B. (2016). *What Is The Difference Between Artificial Intelligence And Machine Learning?* Retrieved on 12-06-2019 from: https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#83d78422742b

Mochales, R., & Moens, M. F. (2011). Argumentation mining. *Artificial Intelligence and Law*, *19*(1), 1-22.

Palau, R. M., & Moens, M. F. (2009, June). Argumentation mining: the detection, classification and structure of arguments in text. *Proceedings of the 12th international conference on artificial intelligence and law* (pp. 98-107). ACM.

Ruger, T. W., Kim, P. T., Martin, A. D., & Quinn, K. M. (2004). The Supreme Court forecasting project: Legal and political science approaches to predicting Supreme Court decisionmaking. *Columbia Law Review*, 1150-1210.

Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. *International Journal of Advanced Research in Artificial Intelligence*, *2*(2), 34-38.

Skymind (2019). *A Beginner's Guide to Neural Networks and Deep Learning.* Retrieved on 11-06-2019 from: https://skymind.ai/wiki/neural-network

Sulea, O. M., Zampieri, M., Malmasi, S., Vela, M., Dinu, L. P., & van Genabith, J. (2017). *Exploring the use of text classification in the legal domain.*

Sulea, O. M., Zampieri, M., Vela, M., & van Genabith, J. (2017). *Predicting the law area and decisions of french supreme court cases*.

Surden, H. (2014). Machine learning and law. *Wash. L. Rev.*, *89*, 87.

Wyner, A., Mochales-Palau, R., Moens, M. F., & Milward, D. (2010). Approaches to text mining arguments from legal cases. *Semantic processing of legal texts* (pp. 60-79). Springer, Berlin, Heidelberg.

Zuo, F., Li, X., Zhang, Z., Young, P., Luo, L., & Zeng, Q. (2018*). Neural machine translation inspired binary code similarity comparison beyond function pairs.*