Machine Learning in Predictive Analytics on Judicial Decision-Making

by

Celia Pienaar

PNRCEL002

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In (partial) fulfilment of the requirements for the degree

MIT (Conversion Master's Degree in Information Technology)

Computer Science Department

UNIVERSITY OF CAPE TOWN

Date of submission:

July 2020

Supervisor:

Dr Geoff Nitschke, Computer Science Department, University of Cape Town

Machine Learning in Predictive Analytics on Judicial Decision-Making

Celia Pienaar

July 2020

Version: 1

Department of Computer Science

Computer Science

Machine Learning in Predictive Analytics on Judicial Decision-Making

Celia Pienaar

Reviewer Dr Geoff Nitschke
Department of Computer Science
University of Cape Town

Supervisor Dr Geoff Nitschke
July 2020

Celia Pienaar

Machine Learning in Predictive Analytics on Judicial Decision-Making

Computer Science, June 2020

University of Cape Town

Department of Computer Science

18 University Avenue

Rondebosch

Cape Town

7700

DECLARATION

I, *Celia Pienaar*, hereby declare that the work on which this dissertation is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Celia Pienaar

Date: 10 July 2020

Abstract

Legal professionals globally are under pressure to provide 'more for less' – not an easy challenge in the era of big data, increasingly complex regulatory and legislative frameworks and volatile financial markets.

Although largely limited to information retrieval and extraction, Machine Learning applications targeted at the legal domain have to some extent become mainstream. The startup market is rife with legal technology providers with many major law firms encouraging research and development through formal legal technology incubator programs.

Experienced legal professionals are expected to become technologically astute as part of their response to the 'more for less' challenge, while legal professionals on track to enter the legal services industry are encouraged to broaden their skill sets beyond a traditional law degree.

Predictive analytics applied to judicial decision-making raise interesting discussions around potential benefits to the general public, over-burdened judicial systems and legal professionals respectively. It is also associated with limitations and challenges around manual input required (in the absence of automatic extraction and prediction) and domain-specific application.

While there is no 'one size fits all' solution when considering predictive analytics across legal domains or different countries' legal systems, this dissertation aims to provide an overview of Machine Learning techniques which could be applied in further research, to start unlocking the benefits associated with predictive analytics on a greater (and hopefully local) scale.

Acknowledgement

I am grateful to the University of Cape Town for responding to the increasingly important role of Information Technology in a corporate and commercial environment, offering an opportunity to learn new (albeit intimidating) skills through such a world-class institution.

I thank my supervisor, Dr Geoff Nitschke, for encouraging me to explore a dissertation related to my (legal) profession and making a contribution to the South African legal community in doing so.

I thank Bowmans for funding this journey. Law degrees are not often associated with an Information Technology degree. My hope is that, in a few years' time, this road will no longer be '*the one less travelled by*'.

Contents

1. INTRODUCTION

1.1 Motivation and problem statement

Since the financial crisis of 2008, much has been advocated on legal service providers being expected to provide 'more for less', placing increasing pressure on legal professionals to work more efficiently and effectively. There is also increasing pressure on the legal fraternity as a whole to become more technologically astute to understand and support the changing nature of clients' businesses.

Arno Lodder tells of how, when informing a chemist in 1995 that he works in the field of IT and law, the reaction was *"Is there any connection between the two at all?"* (Lodder and Oskamp, 2010). In the same year when lawyer Richard Susskind predicted that email would become the dominant way for lawyers to communicate with their clients, the legal profession criticised him and said he was disrespectful towards and did not understand the legal profession. It is worth pointing out that Susskind was not only proven right but is now renowned author and speaker and IT Adviser to the Lord Chief Justice of England and Wales (Susskind, 2020).

In 2013, the American Bar Association updated its Model Rules of Professional Responsibility to amend Rule 1.1 (which deals with lawyers' competence) to include: *"To maintain the requisite knowledge and skill, a lawyer should keep abreast of changes in the law and its practice, including the benefits and risks associated with relevant technology…"* (American Bar Association, 2020).

In 2019 the University of Cape Town announced that it plans to implement a law degree combined with computer science. To quote Professor Danwood Chirwa, Dean of Law: *"Those that came before us long recognised the link between law and humanities and between commerce and law, and introduced combined degree programmes in these fields. Now we see the interconnections between law and* *technology, and we think these links are worth exploring."* (BusinessTech, 2020).

While few would doubt the increasing importance of upskilling on the technological front in order to keep up with legal practice in the 21st century, even fewer would actually understand the underlying workings of Machine Learning (ML) when applied in predictive analytics on judicial decision-making.

1.2 **Methods**

The five areas of legal services most frequently explored for AI application are electronic discovery (eDiscovery), legal search, document generation (such as automated document assembly for the likes of legal brief and memoranda generation), information extraction (for instance extraction of key content used in due diligence transactions) and prediction of case outcomes (Kerikmäe et al., 2018). The first four areas mentioned already have a variety of commercial applications available, many of which are well embedded in law firms and in some instances also in-house legal teams globally. The fifth is not commercially available, yet could add great value to legal professionals, the public and even the judiciary itself.

A lack of understanding by legal professionals of the underlying concepts of ML methods when applied to legal texts (and more specifically court decisions) could slow down or prevent the use of predictive analytics, in turn preventing potential benefits from being realised. This dissertation aims to provide an overview of progress made in research and development to analyse court judgements and opinions given in historic cases and predict and explain outcomes on similar future cases. The focus of this research is not on particular software applications or service providers but rather key concepts playing part in the application of ML for purposes of predictive analytics in judicial decision-making, as researched and developed and applied in other countries. It takes from various technical sources and attempts to serve as an introductory guide for legal professionals wishing to gain a basic understanding of current ML methods for predictive analytics. The specific focus is on the potential benefits as well as challenges and limitations when predictive analytics are applied to judicial decision-making from a South African (and an African) perspective.

To clarify, the purpose of predictive analytics as referred to in this dissertation is to assess to what extent (if at all) ML can replace the role of legal professionals in the processes of: identifying similar or relevant case law; analysing same; assessing validity of and weighing historic arguments; construing new arguments; and predicting outcomes on future matters - based on underlying patterns identified through the use of technology.

1.3 Contributions

This dissertation bridges two worlds: Firstly, that of ML when applied to the legal profession for purposes of predictive analytics; Secondly, that of legal professionals required to build an understanding of the history, current state-of-the-art and future focus areas of predictive analytics when applied to judicial decision-making.

In support of the objective referred to in 1.2, a quick reference map of how concepts such as information retrieval (IR), information extraction (IE) and reasoning relate to one another and their underlying methods have been created.

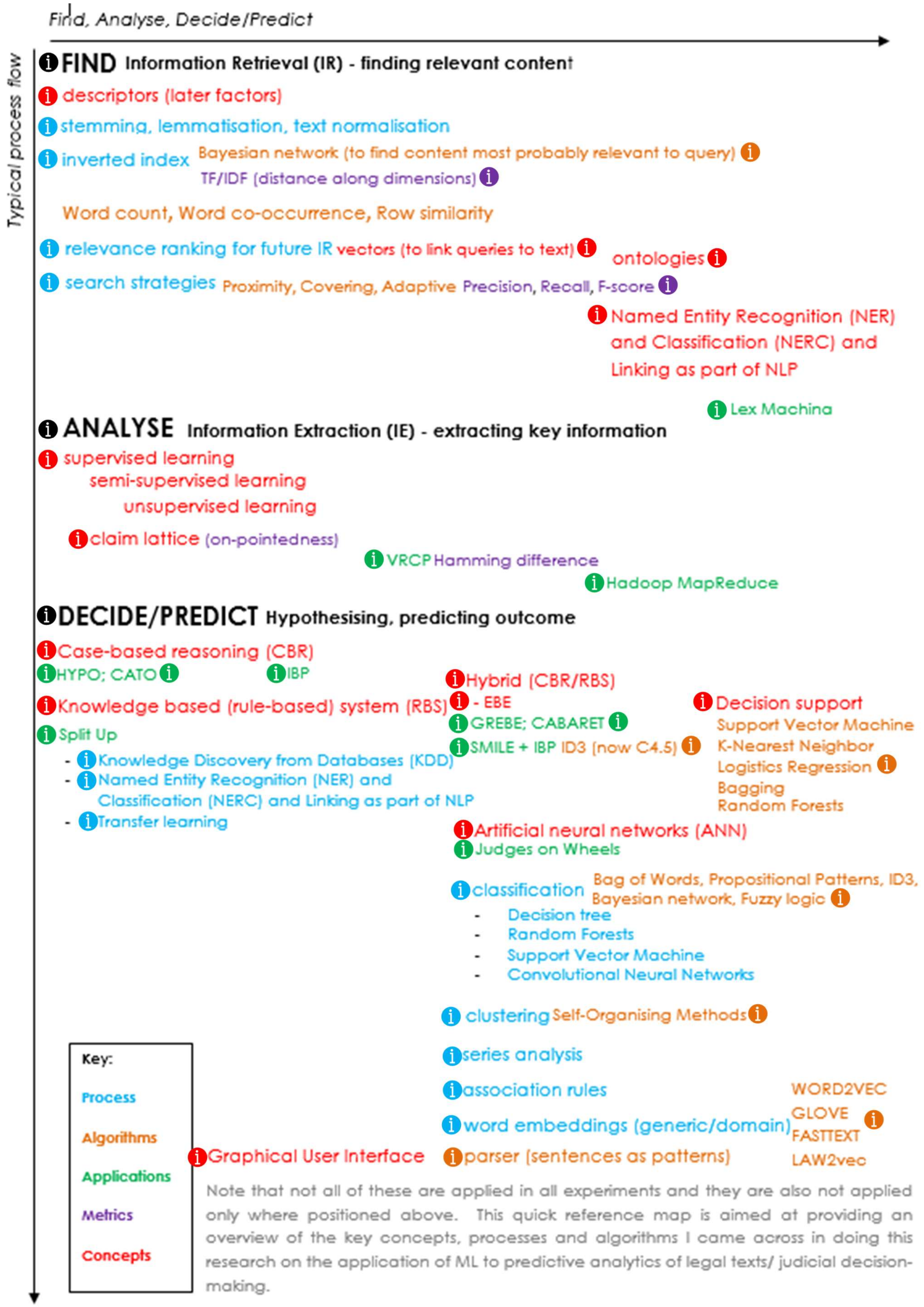


Figure 1: Quick Reference Map *with hyperlinks to further context in this document.*

1.4 Dissertation structure

Chapter 2

This chapter provides an overview of the nature of legal texts and the role of legal informatics along with the benefits of and challenges to the use of predictive analytics on judicial decision-making.

Chapter 3

This chapter provides an overview of the application of IR, IE and reasoning on legal texts. It also briefly touches on the use of artificial neural networks and evolutionary ML.

Chapter 4

This chapter explores the demand for predictive analytics and the role of ML, providing input from surveys highlighting current practices globally as well as in large African law firms specifically.

Chapter 5

This chapter describes experiments and results in a study aiming to create a data-driven legal decision support application.

Chapter 6

This chapter discusses potential injustices and prejudice, the absence of commercial solutions, interesting global developments and thoughts on the potential of predictive analytics to prevent bias or undue influence.

Chapter 7

This chapter concludes with my views on the value of progress to date and whether (and how) predictive analytics should be explored in more detail.

2. BACKGROUND

2.1 Nature of legal texts (specifically court judgements)

The term *case law* refers to historic court judgements which usually consist of court headings, a case summary, case facts, references to legal texts, decisions and opinions and details of the legal representatives (Iftikhar et al., 2019). South Africa has what is called a hybrid legal system in that some parts are based on Roman Dutch law (*civil law*), some principles embedded since the British rule and subsequently followed by our courts (*common law*), and some parts influenced by indigenous law (*customary law*). One of the basic principles of common law systems (and therefore also part of South Africa's judicial system) is that of *stare decisis*, Latin for '*to stand by that which is decided*'. It dictates that the judgement of future cases should follow that of courts at a same or higher level in the court hierarchy. Without the concept of *stare decisis*, predictive analytics would not be possible or even relevant.

When preparing for litigious matters, legal professionals would analyse the matter at hand and search for similar cases supporting their client's case as well as their opponent's, to understand the aspects that could impact a judge's decision-making process. The same process would be undertaken by the judge (or the court assistants) in considering all possible outcomes as part of the decision-making process.

The challenges raised with IE on legal texts (not necessarily limited to case law) include that legal texts are usually long (the sentence structure as well as actual document length) and complicated. Most cases cite previously decided cases either in support or distinction, which creates complicated citation networks. While it does not happen often, case judgements could be criticised as decided incorrectly or in line with outdated legislation, or even reversed at a later stage. One noted advantage on the nature of case judgements is that texts tend to be well written and error free due to a high level of care and precision taken in producing them (Tran et al., 2013).

In summary, while case judgements might be highly structured documents as far as the formalities and rules around layout, formatting, content and reference style are concerned, the data they contain is generally unstructured. Also, judges do not typically review thousands of cases but probably only a few hundred at most (Lodder and Oskamp, 2010).

2.2 Legal informatics

Legal informatics is defined as "*the discipline which deals with the use of ICT* [Information and Communications Technology] *to process legal information and support legal activities, namely, the creation, the cognition and the application of the law*" (Biasotti et al., 2008).

Legal professionals' roles have been described as finding information, analysing information, and deciding based on such analysis. On the first task described; indices and citator databases are available to assist. On the second; case summaries and relevance rankings offer support. On the third; practice guides assist with the decision-making task but the majority of technology focus on the first task of finding information. Research giant Thomson Reuters suggests that large datasets create patterns from which legal professionals could benefit when identifying correlations between these patterns and outcomes (Conrad and Al-Kofahi, 2017).

An early example of legal informatics is the first legal computer application that was developed at the University of Pittsburg in 1956. It was tasked with finding references to "*retarded child*" in legislation and replacing it with "*exceptional child*". Legislation was transferred to magnetic tape for searching purposes, which lead to the first successful execution of electronic legal text retrieval (Biasotti et al., 2008).

In Italy in 1963 a judge at the Court of Cassation created a database of abstracts of his own decisions, later expanded to also include laws and regulations. This database was subsequently developed into a prominent information system still used today. His example was followed by other nations such as France's Supreme Judge in Administrative cases, Germany's Minister of Justice, Sweden's Directorate of Court Administration and Finland's Supreme Administrative Court; creating a vision of unified storage of legal sources (Biasotti et al., 2008).

2.3 AI in law

The concept of Artificial Intelligence (AI) was officially proposed at the Dartmouth Conference in 1956, meaning research of AI and law is already at least 60 years developed (Zhang et al., 2019).

Edwina Rissland's famous 1990 expression on projects on AI and law still ring true (Rissland, 1990; as cited by Dadgostari et al., 2020):

> *"A unifying theme of the projects is the goal to understand and model legal argument, a keystone of an overarching goal to understand and model legal reasoning. These goals require that we know first how to represent several types of knowledge, such as cases, rules and arguments; second, how to reason with them, such as to manipulate precedents, to apply and make inferences with rules, and to tailor arguments to facts; and third, how to use them ultimately in a computer program that can perform tasks in legal reasoning and argumentation, such as analogizing favorable cases and distinguishing contrary ones, anticipating parries in adversarial argument, and creating artful hypotheticals."*

While the concept is not new, until recently there has not yet been much active research on the application of ML to case law analysis and prediction (Katz, 2012). An example mentioned is the work done by political scientists Martin and Quinn and legal scholars Ruger and Kim during 2003 where three methods of prediction were applied and ultimately, the machine did fare better than the experts in predicting outcomes (Ruger et al., 2004). More recent sources indicate that there has been further research, without comprehensive evaluation models for such methods (Liu and Chen, 2017; Zhang et al., 2019).

2.4 Predictive analytics

An area often written about but not yet part of mainstream legal technology available for commercial use is that of predictive analytics of decision-making; to identify and analyse patterns in historic cases to predict (hopefully with some explanations) future outcomes in similar cases. Already in1970, the growing number of court cases was the focus of research into methods of IE to alleviate the burden on legal professionals (Marx, 1970). As far back as 1993 the vast amount of legal material was described as the "*crisis of law*" (Schweighofer and Winiwarter, 1993).

Although judges may not realise or wish to admit it, their approach to a judgement is a routine in that a mental protocol is followed and it is therefore structured, which lends itself to ML application (Remus and Levy, 2017). Outcome prediction is both historical (focusing on similar previous cases) and empirical (with statistical ML focusing on feature extraction to strengthen or weaken classification) (Raghupathi et al., 2018).

Predictive analytics in this context refers to the process of identifying similar or relevant case law, analysing such, assessing validity of historic arguments, construing new arguments and predicting outcomes on future matters based on underlying patterns. Part of this process is to reduce the number of potentially responsive documents by finding legally similar documents and then extracting key legal concepts or rules from those. Its purpose is not so much the number-crunching often associated with big data analysis, but rather the identification and subsequent analysis of underlying patterns.

2.4.1 Benefits

The application of predictive analytics to judicial decision-making holds various potential benefits:

2.4.1.1 *Improved management of case volumes*

According to a telephone interview with a leading case law publisher in South Africa on 8 October 2019, case volumes are as high as 3,000 per year (not including all courts) with as little as 15% of that at most being reported (in other words, summarised and circulated to the broader legal community as publicly available information).

High volumes of cases not only take a long time to process, thereby placing an enormous burden on the judiciary, but also add to legal professionals' workload in having to stay abreast of legal decisions relevant to their area(s) of practice.

2.4.1.2 *Public interest*

Predictive analytics facilitate public transparency on how the law is interpreted and applied to real-life scenarios. In an ideal world, judgement results should be consistent across similar cases so predictive data models could assist in addressing this challenge.

2.4.1.3 *Reducing the costs of legal services*

One explanation behind the high fees associated with legal services could be the expanding volumes of legal information to be digested by legal professionals. The last two decades have seen many large law firms investing in specialist knowledge officers to assist with this task (many of them former lawyers themselves). Smaller firms and sole practitioners do not necessarily have the luxury of dedicated knowledge professionals tasked with identifying and distributing current awareness.

Reliable legal decision support in the form of a computerised knowledge assistant could enable legal professionals to build better arguments in less time (Cardellino et al., 2017), especially if combined with visual data representation interfaces which would allow them to also explain their arguments more effectively to clients (Conrad and Al-Kofahi, 2017). In addition, legal professionals being able to analyse typical trial length and/or award levels associated with particular types of cases could assist in providing certainty around pricing upfront, and thereby better manage client expectations at the onset of a matter.

2.4.1.4 *Predictability*

Predictive analytics could play an important role in maintaining the rule of law by improving general predictability - providing even the layman with an understanding (albeit basic) of key legal principles. This could facilitate improving access to justice, and perhaps even improving trust in the judicial system. While it is debatable whether this would reduce the number of legal disputes, it should at the very least improve predictability and accuracy in certain types of decisions, which could eventually reduce the burden on courts due to an increase in early settlement.

2.4.1.5 *Training value*

Many commercial legal technology platforms used in other areas of practice such as document creation mentioned above, play a role in educating and guiding the next generation of legal professionals. This is achieved by embedding guidelines and principles as part of the applications' processes. One can argue that predictive analytics available in an understandable format could serve this purpose, potentially levelling the playing field between generations and legal professionals from different demographical or educational backgrounds.

2.4.2 Challenges

The application of predictive analytics is not without its challenges:

2.4.2.1 *Big data, computational requirements and costs*

The legal domain, and specifically case judgements, are characterised by at least three big data characteristics: Volume (the high number of case judgements being released daily); velocity (data being accumulated real-time and rapidly); and variety (data being stored in different formats) (Raghupathi et al., 2018).

Often with ML processes the computationally expensive part is the matching of candidate solutions to the problem at hand (Franco and Bacardit, 2015). That said, a limitation to the use of Knowledge Discovery from Databases techniques (described in 3.1.3.2 below) lies in the fact that large, structured legal data sets are rare and should not be confused with the high numbers of cases reported. Cases fall in different domains so for instance, trade secret misappropriation cases (used in many of the application examples described below) would have little bearing on criminal matters, which in turn would have no relevance to property related disputes, and so forth.

High volumes could lead to computational challenges despite solutions aimed at compacting knowledge bases with spelling correction, stop-word removal or grouping.

Interestingly, the *Hadoop MapReduce* opensource big data analytics framework appears to be increasingly useful in addressing this big data challenge, also in the legal domain (Raghupathi et al., 2018).

2.4.2.2 *Noisy data*

Data sparsity (where not all data is relevant) and data paucity (where data has low usability), dimensionality (where data has many attributes) and heterogeneity (where there are differences between data sets) increase the challenge of IE from unstructured texts. Noisy and low-quality data degrades the performance of IE methods as it complicates identifying semantic relatedness among terms or entities, extracting contextually relevant information, structuring and modelling of the data (Adnan and Akbar, 2019).

2.4.2.3 *Knowledge bases and ontologies*

An ontology is a set of concepts and categories in a subject to demonstrate properties and relations, often created for sub-domains. Even where ontologies exist with sufficient structure and detail for a particular legal domain(s), maintenance becomes challenging where new concepts or examples or rules not previously catered for could come into play in future (Moens and Angheluta, 2003; Priddle-Higson, 2010).

Named entities in the legal domain are not limited to generic concepts such as people or places but also the names of laws or procedures, hence the need for ontologies (Cardellino et al., 2017). It is also "*important to differentiate between words in general, and afterwards to link the occurrences of the same entities*" so Named-Entity Linking (NEL) has been proposed to solve these challenges (Elnaggar et al., 2018).

A further limitation already pointed out is that not all cases are reported, so knowledge bases might not be entirely representative of all matters in a particular domain (Brùninghaus and Ashley, 2003).

2.4.2.4 *Training sets*

Systems such as CATO and SMILE (discussed below) require training sets (184 cases were used in CATO and 146 in SMILE), which are usually manually annotated by law students before being fed into the systems. It is not clear how CATO or SMILE would perform when applied directly to full legal texts. The real value of such applications would lie in them extracting *factors* from legal texts automatically. Over-fitting (when the training set is too closely matched to a particular data set) and bias also come to mind as challenges from a training set perspective (Branting et al., 2017).

2.4.2.5 *Unreliable performance comparison*

There appears to be no realistic performance comparison between ML models for legal decision making (Liu and Chen, 2017; Zhang et al., 2019). They compare five well-known models with the Support Vector Model outperforming the rest (*K*-Nearest Neighbor, Logistic Regression, Bagging and Random Forests). One possible solution suggested is mutation of techniques for evaluating prediction models (Zhang et al., 2019).

2.4.2.6 *Explicability and transparency of methods and results*

As already hinted at, analysis of judicial decisions could improve transparency and predictability (Lodder and Oskamp, 2010) and reduce the cost and uncertainty associated with disputes, in turn increasing settlements and reducing the burden on courts (Raghupathi et al., 2018). Key would be to ensure legal practitioners:

(a) can confidently find the relevant content and context required;

(b) understand the underlying processes enabling predictive analytics; and

(c) can ultimately relay same to colleagues and/or clients.

Legal professionals becoming increasingly familiar with and skilled in the concept and techniques underpinning predictive analytics could improve utilisation of alternative dispute resolution processes. These new skills would equip them to evaluate the likelihood of success and associated financial costs and benefits at an early stage of a matter (Stranieri and Zeleznikow, 2010: 120).

Experiments with the use of knowledge systems show that most users blindly follow the suggestions of the system, even though they (the knowledge systems) do not make the final decision but merely offer possible outcomes (Nieuwenhuis, 1989; and Dijkstra, 1998; as cited in Lodder and Oskamp, 2010).

In practice, juniors often do case research and would need to be able to explain to their seniors what their opinions and arguments have been based upon. This again confirms the importance of predictive analytics playing a supporting role to unlock the benefits, and not having a replacing function as some practicing legal professionals might fear.

3. **MACHINE LEARNING**

3.1 **Data-driven ML**

Technological progress in hardware, ML, Natural Language Processing (NLP) and data science methods, as well as better acceptance by legal professionals of the transformational role of technology, have led to renewed focus on data-centric research and the role of computational systems in legal decision-making. These techniques with the ability to process large data sets such as court cases can offer new insight into citation networks, probability of case outcomes and the evolution of legal doctrines over time (Raghupathi et al., 2018).

IR and IE are high-level tasks forming part of the NLP process to interpret data spoken or written by humans. IE is required in order to start analysing data, perform data mining or Knowledge from Database Discovery (all aimed at extracting structured information from unstructured data) (Adnan and Akbar, 2019). These concepts are described in more detail below.

3.1.1 Information Retrieval

In 1950 Calvin Mooers defined IR as "*the problem of directing a user to stored information, some of which may be unknown to him*" (Dadgostari et al., 2020).

Developed in the mid-80s, Ashley named the first computer program comparing cases based on their facts VRCP (Visual Representation of Case Patterns) (Ashley, 2010: 31). *Figure 2* depicts what this looked like, with the diagram projecting 60 Canadian tax cases decided over a 10-year period onto a two-dimensional space (13 of them in favour of (*pro)* the taxpayer and 47 not in favour (*con)*).

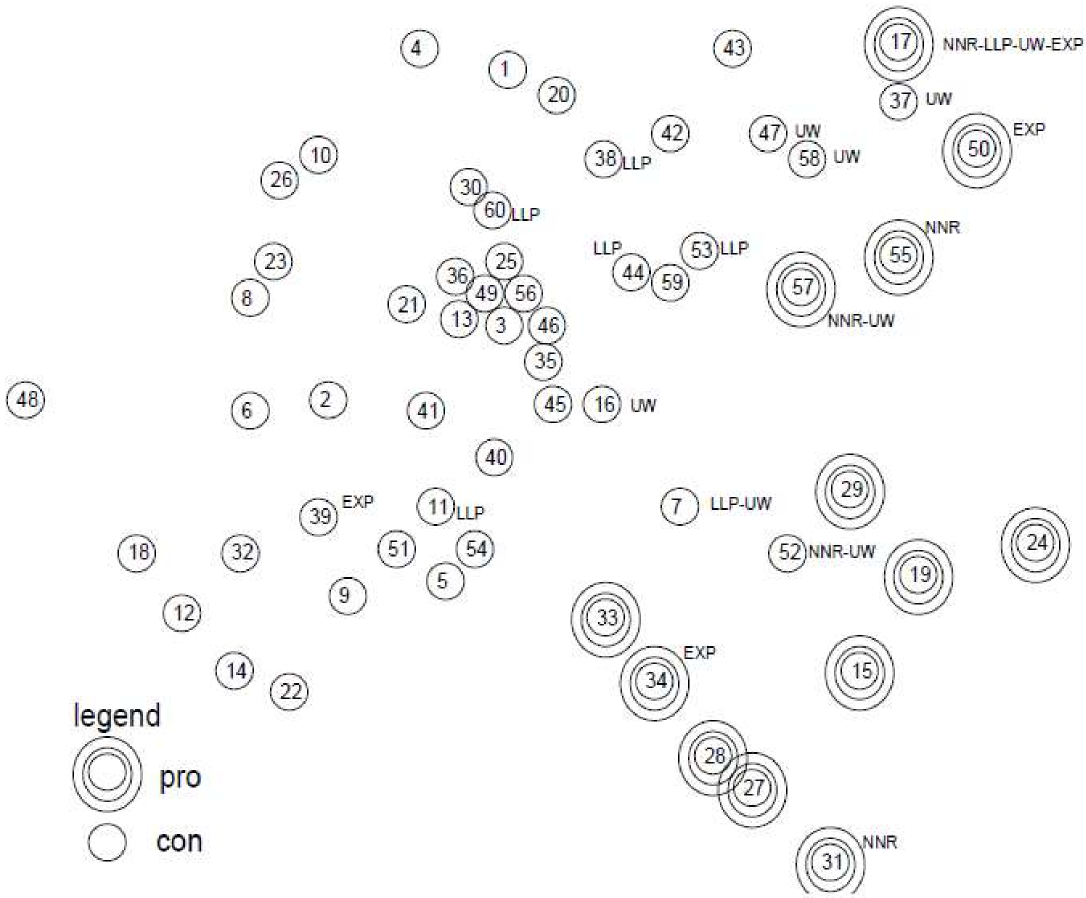


Figure 2: VRCP output *(from 'Case-based Reasoning' (Ashley, 2010: 32, based on Mackaay and Robillard, 1974: Figure 3: 318)) where labels indicate the following: incorrectly decided (EXP), nearest neighbor approach (NNR), linear programming method to compute weights of fact descriptors (LLP), unit weighing approach (UW).*

In the above VRCP analysis, each case's facts have been summarised in terms of 46 *descriptors* to cover legally relevant factual aspects. A '*1*' indicated a descriptor was present whereas a '*0*' meant it was absent. The two examples given are '*1*' where the *"private party is a company"* and '*0*' if not; and '*1*' where *"purchase was not followed by sale within a short period thereafter"* and '*0*' where the opposite. The diagram above also captured dissimilarity, represented by the distance between the pairs of cases (with the measure of dissimilarity being described as the *Hamming difference,* being the number of fact *descriptors* for which the case pair differed). The fewer descriptors in common, the greater the distance between case pairs. As a result, looking at *nearest neighbors* could assist in inferring future outcomes or identify anomalies, and the distinction between so-called *pro* or *con* cases assists in identifying uncertainty on a new matter (for instance those close to the border between *pro* and *con* cases or those with labels). Ashley viewed the VRCP computer program as the first to not only compare facts of new cases to historic cases automatically (except for capturing *descriptors*, which was still a manual process) but also visually represent the information (Ashley, 2010: 33).

IR processes usually use software programs to remove stop-words (such as *"a"*, *"the", "and"*) and stem words (e.g. removing *"-ing", "-s"*). Legal IR applications also identify citation networks (references to legislation or other cases) and count word appearances, thereby creating an inverted index (*hashmaps* of content and their locations in documents) to retrieve cases by in future (Ashley, 2010: 35).

Ashley (Ashley, 2010: 36) describes two methods for IR applications to link queries to texts: The first method is a term *vector* approach representing each word (or feature) as a vector in a space of cases through trigonometric calculations, thereby drawing an arrow from the origin along each dimension to the text (0,0,0,...0). The distance along each dimension is called its TF/IDF weight (*Term Frequency* being how often the term appears in the text, inversely related to the *Inverse Document Frequency* (the number of times the term appears in the data set)). This approach finds the most similar document to the query by computing the cosine of angles between corresponding term vectors. The second method is found in some commercial systems such as Westlaw and LexisNexis that use a *Bayesian inference network* to find documents most probably relevant to a query. *Figure 3* shows an inference network as example for computing to what extent a query need has been satisfied (Ashley, 2010: 37).

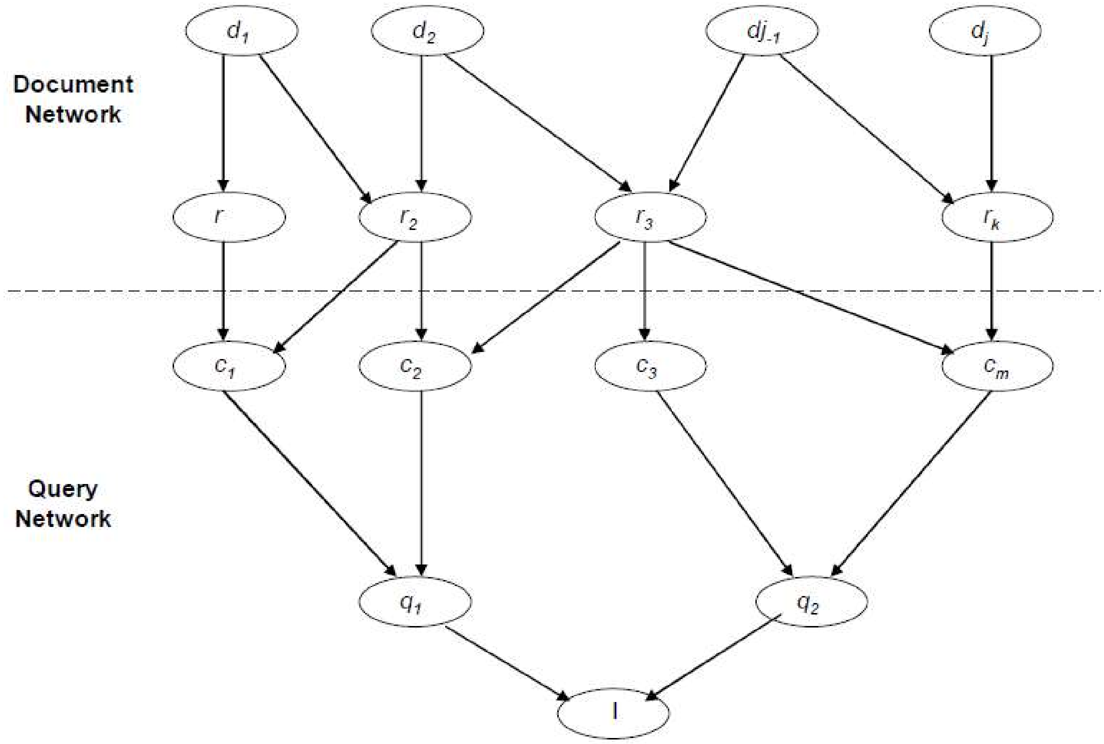


Figure 3: Bayesian inference network retrieval model *(from 'Case-based Reasoning' (Ashley, 2010: 38)) capturing probabilistic dependencies of whether an information need has been met if a document has been observed.*

Interesting to note from an IR perspective is how this works in practice: The process starts with the court submitting judgements in electronic form to the commercial IR providers, where the creation of inverted indices happens automatically and stop-word removal and stemming take place. In response to queries, the IR applications will retrieve indexed documents and rank them according to statistical criteria (TF/IDF) automatically (Ashley and Brùninghaus, 2009).

Ashley reckons it is relatively easy to create and maintain full-text legal IR systems such as Westlaw or LexisNexis in that the inverted index for new cases are constructed automatically, and the TF/IDF-based similarity measures or conditional probabilities can be computed and applied automatically. While they do assist in promoting texts most probably relevant to a query through the presence of the frequency-weighted terms, a shortcoming is that it does not focus on legally relevant similarities or differences between cases. IR applications do not understand the query and cannot explain why a retrieved case is legally relevant. For this reason, graphical representation of output would probably not be very useful as it would not point out *descriptors* as is the case with VRCP. In working towards a solution, the SCALIR program used TF/IDF weights to generate pictures of networks of cases sharing substantive concepts where the weights were used to position cases along a vertical line, with those closer to the line more likely to be relevant (Rose and Belew, 1991; as cited by Ashley, 2010: 39).

IR performance is assessed in terms of *precision* (relevant retrieved documents as a percentage of all retrieved documents), *recall* (relevant retrieved documents as a percentage of all relevant documents) and *f-score* (a combined metric of precision and recall). Three popular search strategies applied for IR (specifically in citations but perhaps with potential for further use when applied to *descriptors* or *factors*) are (Dadgostari et al., 2020):

- The *proximity algorithm* which starts with a source document and then selects all documents closest to the source within the space;

- The *covering algorithm* which starts with a source document and then selects the most proximate document. Using fixed parameters, it then determines whether to return to the source document or to continue along that line of documents; and

- The *adaptive algorithm* which is similar to the covering algorithm but instead of fixed parameters, the parameters are learned from the data set using reinforcement learning.

3.1.2 Information Extraction

The aim of IE is to prepare and improve data retrieved through IR methods for further analysis. IE is described as a type of IR that automatically extracts *"structured or semi-structured information from unstructured or semi-structured machine-readable documents"*, for instance, recognising the names or entities such as people, organisations, or products (Jackson and Moulinier, 2007; as cited by Ashley and Brüninghaus, 2009). Its purpose is to *"extract instances of predefined categories from unstructured data, building a structured and unambiguous representation of the entities and the relations between them"* (Adnan and Akbar, 2019).

Learning-based approaches to assist with IE are divided into supervised, unsupervised and semi-supervised techniques. A drawback of supervised learning techniques is that it requires manually trained (labeled) data which is time-consuming to create (and possibly open to inconsistency or bias). This approach is effective for domain-specific IE (Adnan and Akbar, 2019). On the other side, unsupervised learning does not require labeled data, for example automatic clustering of documents or concepts. Pre-processing of data is required to avoid missing values or noise often associated with big data. Semi-supervised learning requires less supervision than supervised learning techniques and can use both labeled and unlabeled data.

3.1.3 Reasoning systems applied to the legal domain

Lex Machina (Lex Machina, 2020) is probably the best-known example of a commercial data-driven system used by legal professionals, applying NLP and ML to the legal field (specialising in the intellectual property domain). A limitation is that it does not predict outcomes based on similarity factors but rather particular courts or opponents' previous records.

Unlike the Lex Machina example, true reasoning systems do more than just store and process information but also connect stored information with case facts and reason with it. Three types of reasoning applications are case-based reasoning, knowledge-based systems and artificial neural networks, each briefly described below (Lodder and Oskamp, 2010):

3.1.3.1 *Case-based reasoning*

Case-based reasoning (CBR) is "*…the process of using previous experience to analyse or solve a new problem, explain why previous experiences are or are not similar to the present problem and adapting past solutions to meet the requirements of the present problem*" (Nissan, 2015).

With CBR, knowledge is represented by the relevant factors found in case precedents, which allows the application to point out differences between similar cases with similar or different outcomes.

The *Jurimetrics Era* (1950s through 1970s) is described as the period where computerised CBR systems were built to help legal professionals predict outcomes of disputes (without explaining their predictions). The subsequent *AI and Law Era* focused on assisting legal professionals to build arguments for and against proposed outcomes. In the *Era of Convergence* programs can explain predictions and make reasonable legal arguments for both sides to the argument. CBR systems focus on specific domains, often consisting of only a few hundred cases. In an ideal world a CBR system would be able to analyse a claim, find a list of relevant cases and rules, predict an outcome based on those while also explaining the reasoning and potential legal arguments that can be used (Ashley, 2010: 27, 64).

Two mechanisms have been developed to address the question of "*[w]hen is it reasonable to infer that because a court decided a similar precedent, the same or different outcome should apply to the problem?*" (Ashley, 2010: 42):

a) Dimensional comparison which compares cases on their respective strengths to draw inferences from the strongest (for example in the HYPO system described below); and
b) Matching Exemplar-Based Explanations (EBE) which draws inferences on the extent to which an explanation from one case maps onto another's facts (for example in the GREBE system (Branting, 1991, 2000; as cited by Ashley, 2010: 42)).

HYPO is a CBR system created by Edwina Rissland and Kevin Ashley in the trade secrets domain (Ashley, 2010: 43). HYPO analyses a new matter, retrieves relevant cases from its case base; sorts these by on-pointedness (referring to a *claim lattice*, see below); selects the best case(s) for each side of the dispute; and generates arguments and strengths and weaknesses for each side by way of hypotheses (Rissland et al., 2006).

HYPO was followed by CATO (also a CBR system), developed by Kevin Ashley and Vincent Aleven in 1997 as a tutoring system to teach law students to reason with precedents. CATO used *factors* (binary, and not as detailed as dimensions as it was either present or not, whereas dimensions indicated benefiting either side to a varying degree (Priddle-Higson, 2010)). CATO was also able to point out alternatives or distinctions (where factors were or were not present but with similar outcomes in the cases). Both HYPO and CATO order relevant cases based on their on-pointedness to the problem in a claim lattice data structure, where the closer it is to the root node, the more on point it is. An example is shown in *Figure 4* (Ashley, 1987; 1990; as cited by Ashley, 2010: 47).

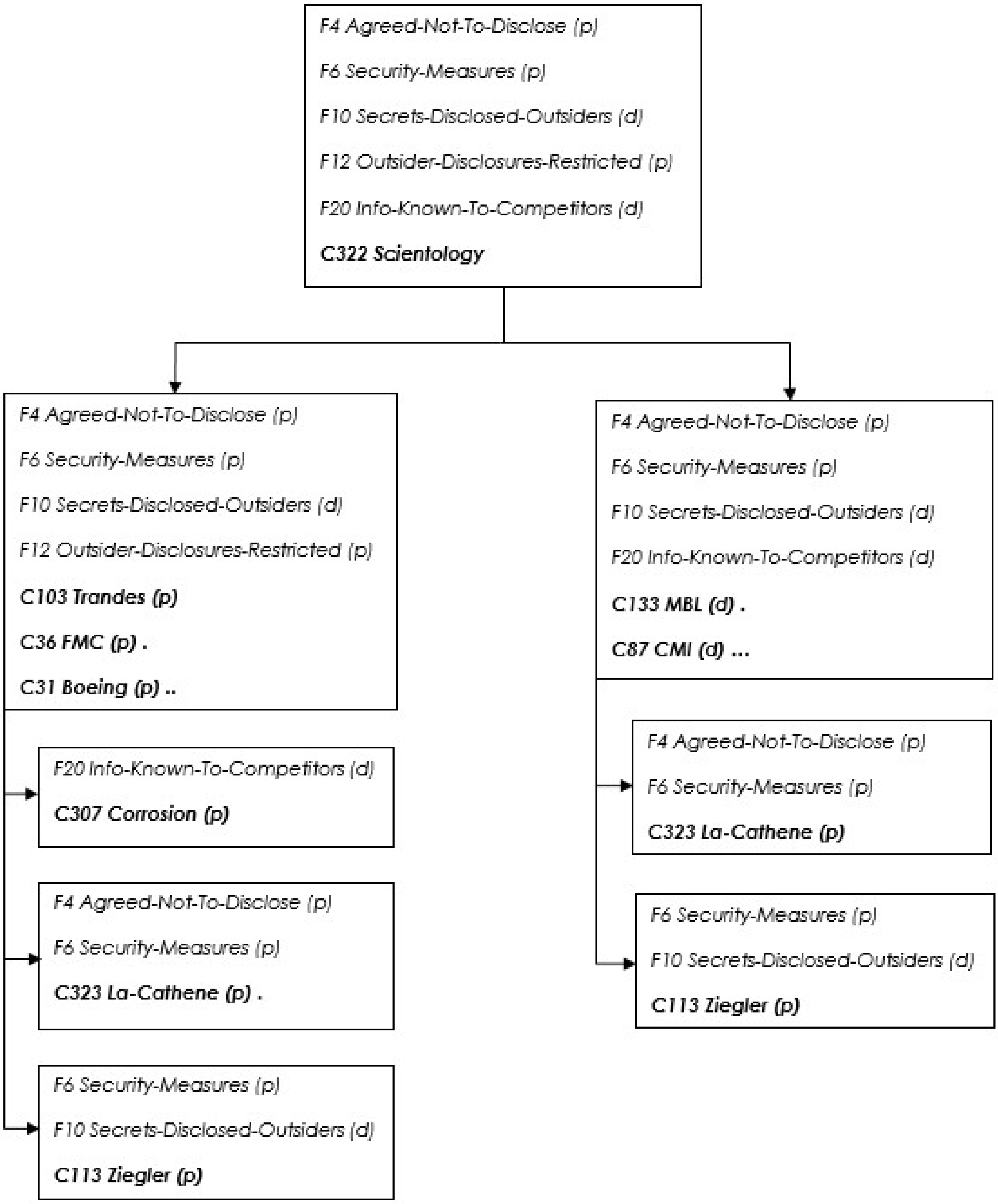


Figure 4: Claim Lattice Example *(from 'Case-based Reasoning' (Ashley, 2010: 48) as an example of representing a trade law case (Religious Technology Center v. Netcom on-Line Communication Services, Inc. et al. 923 F.Supp. 1231 N.D.Cal., 1995 referred to as the "Scientology case") with factors. CATO retrieves the most relevant cases either side can cite without the opponent being able to respond with a more relevant counterexample (e.g. C103, C36 etc.).*

EBEs represent case facts as listed as relevant by the court in reaching its legal conclusion. It forms a semantic network where nodes represent objects, concepts or events, and arcs represent relations e.g. "consequent of" or "has part" (Ashley, 2010: 49).

GREBE was a hybrid CBR/RBR (rule-based reasoning) system developed by Karl Branting in 2000, which related case facts to conclusions representing argument strength through heuristic measures. It used backward-chaining by reasoning backwards from a rule to other rules to create written arguments (Branting, 1991; as cited by Ashley, 2010: 52). In testing Branting's evaluation of GREBE's analysis, a domain expert compared GREBE's analysis favourably to those of law students (Rissland et al., 2006).

Another hybrid CBR/RBR system example is CABARET, which applied dimensions to sub-issues across a dataset of 23 income tax cases (Priddle-Higson, 1995). CABARET was a landmark system as it was the first hybrid system to interleave CBR and RBR dynamically (previously, hybrid CBR/RBR systems used CBR to sense-check results or when all else failed (Rissland et al., 2006)).

A limitation up until this point was the manual input required for case representation, as the above systems were all based on a small number of cases with manual annotation (Priddle-Higson, 2010).

From the above overview of progress on CBR systems to date, it appears that these applications can produce an argument based on historic or new cases, but they cannot predict outcomes. For this a more recent approach called Issue-Based Prediction (IBP) is preferred to test hypotheses on which party is likely to win an argument and explain the reasoning behind such hypotheses (Brùninghaus & Ashley, 2003; as cited by Ashley, 2010: 59). As another example of application in the trade secret domain, IBP identified two main issues and five sub-issues on trade secret misappropriation claims. For each issue in a domain, if all the sub-issues favour the same party, IBP predicts that party will win the case. In the event of conflicts, IBP retrieves cases indexed by the relevant factors to examine their outcomes, hypothesising based on the majority of outcomes from the retrieved cases. The hypothesis is deemed confirmed if all retrieved cases are consistent. If not, it aims to explain away the counterexamples found by distinguishing them from the current case by looking for so-called *'knock out' factors*. Where IBP cannot explain away all the counterexamples, it abstains from making predictions on outcome. IBP also identifies *'weak' factors* where the probability of the favoured side winning is below 20%. Where the only factor concerning an issue in a counterexample is *a* weak factor, IBP will disregard the issue (Ashley & Brǜninghaus, 2009).

Where IBP finds a hypothesis too narrow to retrieve similar cases, it broadens the query by dropping one or more of the factors favoring the majority side. Ashley concludes that combining IBP's predictions with CATO's arguments could be very useful as, in addition to predicting and explaining an outcome and showing arguments consistent with that prediction, it could also make the strongest arguments it knows how *against* that predicted outcome. *Figure 5* shows the algorithm for a case represented as factors. IBP identifies the issues raised and on each one, retrieves cases with similar issues and finally predicts which side should win along with an explanation of how the decision was made (Ashley & Brǜninghaus, 2009).

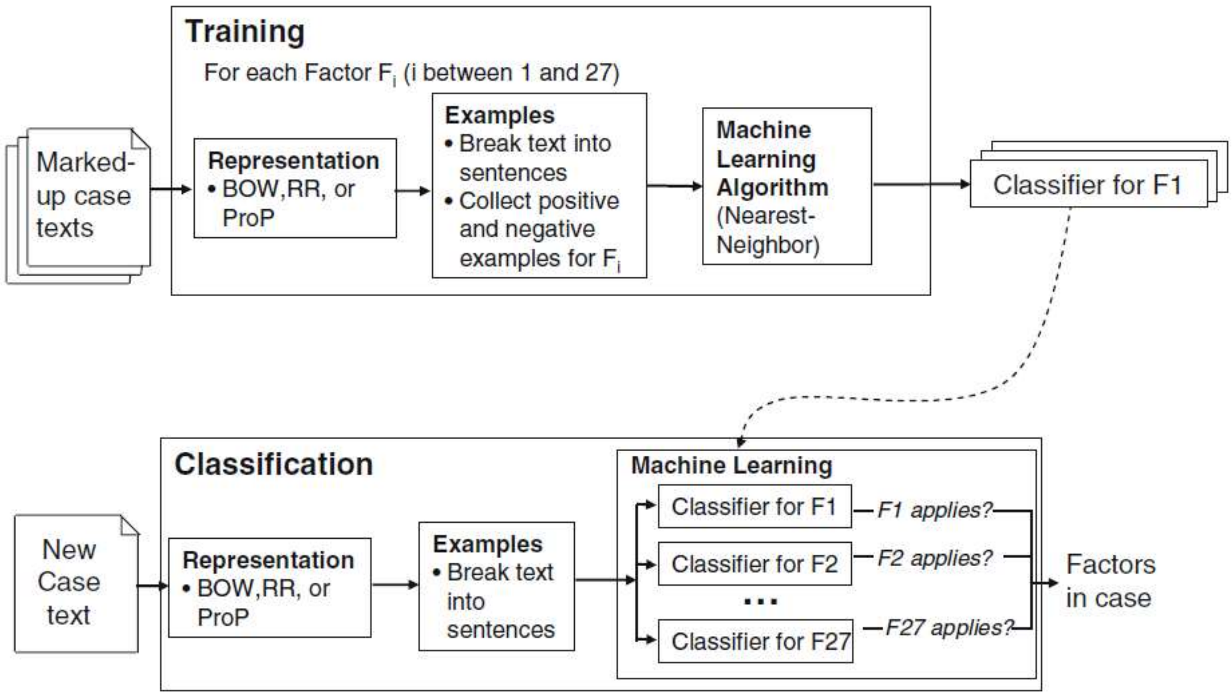
Input: Current fact situation (*cfs*)
A. Identify issues raised by *cfs* Factors
B. For each issue raised, determine the side favored for that issue using Theory-Testing:
1. if all issue-related Factors favor the same side, then return that side,
2. else retrieve issue-related cases in which all issue-related Factors apply
a. if there are such issue-and-factor-related cases, then form hypothesis that same side *s* will win that won majority of cases
i. if all issue-and-factor-related cases favor side *s*, then return side *s*,
ii. else try to explain away exceptions with outcomes contrary to hypothesis
(a) if all exceptions can be explained away, then return side *s* favored by hypothesis
(b) else, return "abstain"
b. if no issue-and-factor-related cases are found, then call Broaden-Query
i. if query can be broadened, then call Theory-Testing for each subset of issue-related Factors and combine predictions for each set.
ii. else return "abstain"
C. Combine prediction for each issue
Output: Predicted outcome for *cfs* and explanation

Figure 5: IBP algorithm *(from 'Automatically classifying case texts and predicting outcomes' (Ashley & Brùninghaus, 2009: 134)) as the Issue-Based Prediction algorithm to retrieve cases sharing the issue and issue-related factors with the problem issues raised.*

Despite the progress above, a hurdle was still the absence of automatic or semi-automatic methods applied across a larger number of cases and more legal domains of cases to index for factors. If that was possible, it would be easier to integrate IBP or CATO with full-text legal IR systems such as Westlaw or LexisNexis (Ashley, 2010: 64).

A further version of IBP is the SMart Index LEarner (SMILE) system which *"bridges case-based reasoning and extracting information from texts"* through a combination of IE tools and ML (specifically the ID3 learning algorithm as a predecessor of the C4.5 algorithm). It has a training set of manually marked-up sentences that are either positive or negative instances of factors, thereby learning decision trees for factors (again applied to the trade secret misappropriation domain). SMILE uses a parser to recognise sentences as patterns (learning from that training set) as either positive or negative instances of factors. IE is then applied to generalise those instances (such as substituting

Sub-figure 6(A)



Sub-figure 6(B)

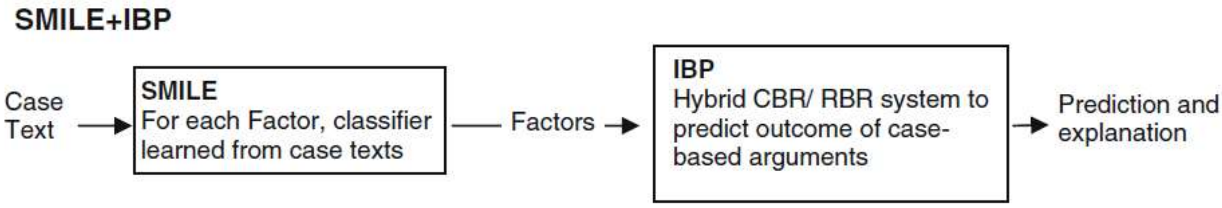


Figure 6: Overview of SMILE (Figure 6(A)) and SMILE + IBP (Figure 6(B)) *(From 'Automatically classifying case texts and predicting outcomes' (Ashley & Brùninghaus, 2009: 128, 140). The SMILE part of SMILE + IBP identifies applicable factors. IBP then relates factors to legal issues, automatically compares the problem and cases on their facts; tests hypotheses; and explains predictions.)*

names). A ML algorithm is then applied to automatically match similarly patterned sentences of new cases to those in the training set (Ashley & Brùninghaus, 2009). *Figure 6(A)* provides an overview of the SMILE system. Some success was proven in extracting information (SMILE) from text-based case descriptions and combining it with IBP to predict outcomes of future cases (Ashley & Brùninghaus, 2009). It appears to be a first (in 2009) to start the IE process on legal texts (as opposed to more structured data) and to provide explanations that legal professionals can understand. SMILE + IBP attempts to automatically classify the cases' text according to a scheme of classification concepts (factors), as shown in *Figure 6(B)*.

Although the so-called knock-out factors concept would initially have to be explained to legal professionals as users of an application, the IBP hypothesis-testing are intuitively accessible to them. *Figure 7* is an example of IBP's prediction algorithm. For each factor, IBP identifies the issues raised and retrieves cases that share those issues or factors and predicts which side should win on each issue, accumulating in a combined prediction for the case and supported with an explanation (Ashley & Brùninghaus, 2009).

Ashley and Brùninghaus conclude their SMILE + IBP experiment by saying future researchers will probably address automatically classifying cases by factors through the use of unsupervised ML, instead of relying on a small set of annotated cases (Ashley and Brùninghaus, 2009). Unfortunately to date, this has not yet been achieved as far as can be publicly ascertained.

IBP Analysis for *National Rejectors* Case as Input by:	
Human	**SMILE**
1. Prediction for NATIONAL-REJECTORS Factors favoring plaintiff: (F18 F15 F7) Factors favoring defendant: (F27 F19 F16 F10)	**1'. Prediction for NATIONAL-REJECTORS** Factors favoring plaintiff: (F18 F7 F6) Factors favoring defendant: (F25 F19 F16 F10)
2. Issue raised in this case is INFO-USED Relevant factors in case: F18(P) F7(P) The issue-relate factors all favor the outcome PLAINTIFF.	**2'. Issue raised in this case is INFO-USED** Relevant factors in case: F25(D) F18(P) F7(P) Theory testing did not retrieve any cases, broadening the query. For INFO-USED, the query can be broadened for PLAINTIFF. Each of the pro-P Factors (F7 F18) is dropped for new theory testing. Theory testing with Factors {F27 F25} still does not retrieve any cases. Theory testing with Factors {F18 F25} gets the following cases: (KG PLAINTIFF F6 F14 F15 F16 F18 F21 F25) (MINERAL-DEPOSITS PLAINTIFF F1 F16 F18 F25) In this broadened query, PLAINTIFF is favored. By a-fortiori argument, PLAINTIFF is favored for INFO-USED.
3. Issue raised in this case is SECURITY-MEASURES Relevant factor in case: F19(D) F10(D) The issue-relate factors all favor the outcome DEFENDANT.	**3'. Issue raised in this case is SECURITY-MEASURES** Relevant factors in case: F19(D) F10(D) F6(P) Theory testing did not retrieve any cases, broadening the query. For SECURITY-MEASURES, the query can be broadened for DEFENDANT. Each of the pro-D Factors (F10 F19) is dropped for new theory testing. Theory testing with Factors {F10 F6} gets the following cases: *[11 cases won by plaintiff, 2 cases won by defendant]* Trying to explain away the exceptions favoring DEFENDANT MBL can be explained away with unshared ko-factors(s) (F20), CMI can be explained away with unshared ko-factors(s) (F27 F20 F17). Therefore, PLAINTIFF is favored for the issue. In this broadened query, PLAINTIFF is favored. Theory testing with Factors {F19 F6} still does not retrieve any cases. There is no resolution for SECURITY-MEASURES, even when broadening the query.
4. Issue raised in this case is INFO-VALUABLE Relevant factors in case: F27(D) F16(D) F15(P) Theory testing did not retrieve any cases, broadening the query. For INFO-VALUABLE, the query can be broadened for DEFENDANT. Each of the pro-D actors (F16 F27) is dropped for new theory testing. Theory testing with Factors {F16 F15} gets the following cases: [8 cases won by plaintiff] In this broadened query, PLAINTIFF is favored. Theory testing with factors {F27 F15} gets the following cases: (DYNAMICS DEFENDANT F4 F5 F6 F15 F27) In this broadened query, DEFENDANT is favored. There is no resolution for INFO-VALUABLE, even when broadening the query.	**4'. Issue raised in this case is INFO-VALUABLE** Relevant factors in case: F16(D) The case has only one weak factor related to the issue, which is not sufficient evidence to include this issue in the prediction.
5. Outcome of the issue-based analysis: For issue INFO-VALUABLE, ABSTAIN is favored. For issue SECURITY-MEASURES, DEFENDANT is favored. For issue INFO-USED, PLAINTIFF is favored. => Predicted outcome for NATIONAL-REJECTORS is DEFENDANT	**5'. Outcome of the issue-based analysis:** For issue INFO-USED, PLAINTIFF is favored. For issue SECURITY-MEASURES, ABSTAIN is favored. => Predicted outcome for ANTIONAL-REJECTORS is ABSTAIN

Figure 7: IBP analysis (left column) versus SMILE + IBP analysis (right column) *(From 'Automatically classifying case texts and predicting outcomes' (Ashley & Brüninghaus, 2009: 137)).*

3.1.3.2 *Knowledge based system*

A Knowledge Based System (KBS) is a RBR system where domain knowledge is represented as *IF-THEN* rules, reasoning with these rules with forward or backward chaining mechanisms. With forward chaining mechanisms the system starts with the conditions of the rules and if those conditions are satisfied, it chains the conclusion of the rule with the conditions of another rule in that knowledge base. This approach is useful where the outcome of a case is not known. Where the outcome of a case is known, backward chaining could perform better. Backward chaining looks at conclusions of rules and where the conditions are satisfied, chains it to the conclusions of another rule in that knowledge base. Most systems use a combination of forward and backward chaining (Lodder and Oskamp, 2010).

Knowledge Discovery from Databases (KDD) refers to the analysis of data contained in a database which then produces new knowledge. KDD techniques have been applied to the legal domain in useful ways especially in those fields of law that allow more judicial discretion, for example in the *Split Up* project (Stranieri and Zeleznikow, 2010: 92). The five phases of the knowledge discovery process are data selection, data pre-processing, data transformation, data mining and interpretation (Fayyad et al., 1996; as cited by Stranieri and Zeleznikow, 2010: 100). *Figure 8* shows an example of the data transformation phase using a partial tree in the *Split Up* project, where data was collected from Australian family court cases dealing with post-divorce property distribution. The model provides a framework for decomposing a task to integrate domain knowledge. *Figure 8* shows 6 of 94 variables used in this domain. Specialist lawyers determined that 3 variables usually determine the asset split, namely (i) the future needs; (ii) past contributions; and (iii) level of marital wealth (collectively referred to as the *child nodes*). The heuristics *"much more", "more", "about the same", "less"*

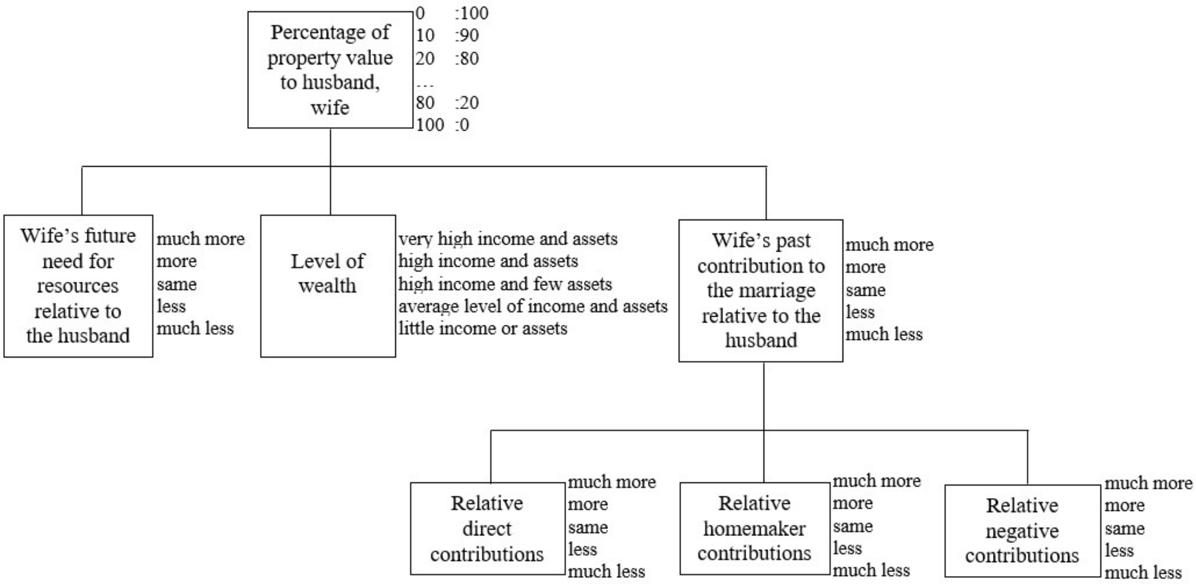


Figure 8: Partial tree for Split Up *(From 'Knowledge Discovery from Legal Databases – using neural networks and data mining to build legal decision support systems' (Stranieri and Keleznikow, 2010: 103)).*

and "*much less*" were used to assign values to these 3 variables (Lodder and Oskamp, 2010).

The four categories of KDD techniques are as follows (Fayyad et al., 1996; as cited by Stranieri and Zeleznikow, 2010: 116):

- ***Classification***; *IF-THEN* type rules automatically extracted through rule induction or neural networks. Examples of application of classifiers in the legal domain are the ID3 rule induction algorithm (which automatically induces rules from large data sets, see *Figure* 9 below), Bayesian belief networks (which assigns numerical values to propositions based on the degree of belief accorded to them) and fuzzy logic (which deals with undefined terms and requires interpretation on the degree of membership to the set) (Chen, 2001; as cited by Stranieri and Keleznikow, 2010: 121-122).

Case	Is Property Asset Rich	Children	Wife-Works	Equal split
50	Yes	Yes	Yes	Yes
51	No	Yes	No	No
52	No	Yes	No	No
53	Yes	No	Yes	Yes
54	Yes	Yes	No	No
55	No	No	Yes	Yes
56	No	Yes	Yes	No

Table 1: Table of data for property split in family law *(From 'Knowledge Discovery from Legal Databases – using neural networks and data mining to build legal decision support systems' (Stranieri and Keleznikow, 2010: 117)).*

An example of how the ID3 algorithm automatically induces rules from a dataset can be seen in *Figure* 9 where the rules were extracted from the table shown in *Table 1* and represented as a decision tree.

Four examples of popular text classification models applied to legal texts are (Sangkeettrakarn et al., 2019):

- *Decision Tree*; used to classify entries to tree roots, repeated until a representative class is found;
- *Random Forest*; where multiple decision trees are construed;
- *Support Vector Machine (SVM);* as supervised learning model determining separators between different classes targeted in a search space;
- *Convolutional Neural Network (CNN);* as a type of deep neural network inspired by biology often used for image classification but also applied to text classification.

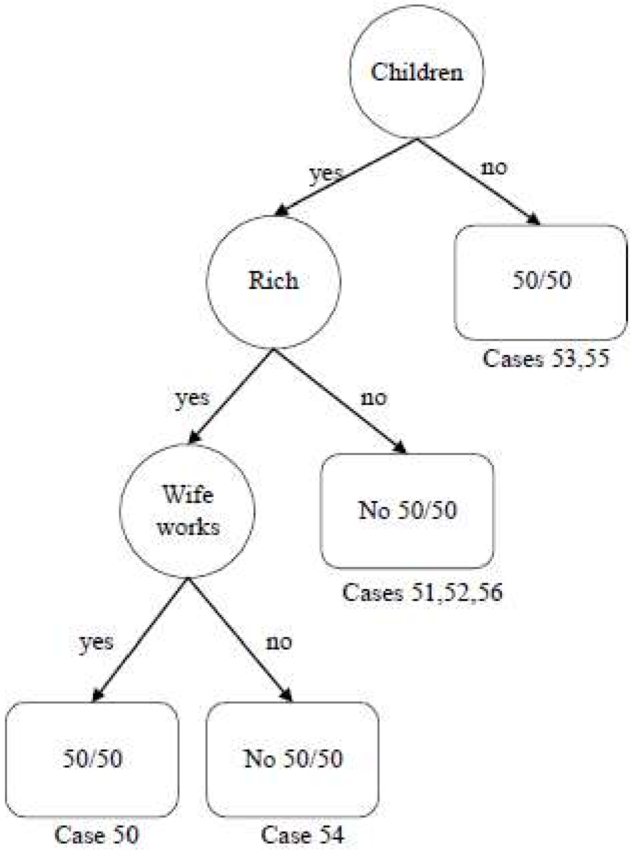


Figure 9: ID3 decision tree on property split in Table 1 *(From 'Knowledge Discovery from Legal Databases – using neural networks and data mining to build legal decision support systems' (Stranieri and Keleznikow, 2010: 118)).*

- ***Clustering***; Unsupervised learning by the grouping of documents based on similarity, for example as used to group European Parliament cases into clusters based on case similarity through the use of Self-Organising Methods (SOM) as a type of neural network (Merkl et al., 1999; as cited by Stranieri and Keleznikow, 2010: 125).

- ***Series Analysis***; Time-series databases (storing information in pairs of time and value) that could assist in tracking and reporting on changes in decision-making patterns over time;

- ***Association rules****;* A famous example is the association rule of *"if beer, then nappies"*, which is not necessarily causal but drawn purely from data. It infers that in 80% of transactions where nappies were bought, so was beer, giving an 80% confidence level (Stranieri and Zeleznikow, 2010: 127). In an experiment where association rules were applied across a data set consisting of 300,000 Australian legal aid applications, it picked up an association between applicants' ages and the categories of aid applied for. Results do not explain much but can be used in the formulation of hypotheses (Avkovic et al., 2003; as cited by Stranieri and Zeleznikow, 2010: 127).

Ashley and Bruninghaus (Ashley and Bruninghaus, 2009) experimented with three representation schemes for text classification on the sample sentence "*Diekman signed a nondisclosure agreement*" which are included below for purposes of understanding examples of text classification algorithms in practice:

- The *Bag of Words* (BOW) method removes punctuation, numbers and duplicate words and represents the sentence as "*a agreement Diekman nondisclosure signed*";
- The *Role Replaced* (RR) method replaces case-specific names with their roles and represents the sentence as "*a agreement defendant nondisclosure signed";*
- The *Propositional Patterns* (ProPs) method represents words within syntactic relationships as "*(defendant sign) (sign nondisclosure agreement)*".

Named Entity Recognition (NER) is a method to identify generic entities such as location, persons, organisations; or domain-specific entities such as courts, laws, concepts etc. (Adnan and Akbar, 2019). Named Entity Recognition and Classification (NERC) forms an important part of IE. Named Entity Linking (NEL) as a form of NERC enables linking entities to knowledge bases or other resources, thereby adding classes to the entities. See *Figure 10* and *Figure 11* as examples of application in practice of generic NER and domain-specific NEL respectively (Cardellino et al., 2017).

Named entities in the legal domain are best represented through ontologies (Cardellino et al., 2017). The creation and subsequent maintenance of ontologies are time-consuming and still very much sub-domain specific, which creates a barrier when it comes to successful IE on legal texts. Another challenge is presented by complex nested entities or noise such as homonyms or language variability (Adnan and Akbar, 2019). A solution with some success demonstrated is transfer learning (Elnaggar et al., 2018), where knowledge gained through solving one problem can be stored and applied to solve another related problem. An example is algorithms applied to identify cars that could also be used to identify trucks (Wikipedia, 2020). This can be applied to legal documents to transfer knowledge from large datasets to smaller datasets, although further research in this field is suggested.

Cardellino in closing mentioned that NERC and NEL will be used to "*speed up manual annotations of judgements*" forming part of data sets although fully automated NEL is not yet satisfactory (Cardellino et al., 2017).

Example 1.1. The [Court]$_{organization}$ is not convinced by the reasoning of the [combined divisions of the Court of Cassation]$_{organization}$, because it was not indicated in the [judgment]$_{abstraction}$ that [Eğitim-Sen]$_{person}$ had carried out [illegal activities]$_{abstraction}$ capable of undermining the unity of the [Republic of Turkey]$_{person}$.

Figure 10: Domain-specific NER *(From 'A Low-cost, High-coverage Legal Named Entity Recognizer, Classifier and Linker' (Cardellino et al., 2017: 9)).*

Example 1.2. The [Court]$_{\text{European\_Court\_of\_Human\_Rights}}$ is not convinced by the reasoning of the [combined divisions of the Court of Cassation]$_{\text{Yargıtay Hukuk Genel Kurulu}}$, because it was not indicated in the [judgment]$_{\text{Court\_of\_Cassation's\_judgment\_of\_22\_May\_2005}}$

Figure 11: NEL *(From 'A Low-cost, High-coverage Legal Named Entity Recognizer, Classifier and Linker' (Cardellino et al., 2017: 9)).*

3.1.3.3 *Artificial neural networks*

Artificial neural networks (ANNs) differ from CBR and KBS because they use knots and links to mimic the way synapses in human brains (neural networks) work. The input side represents relevant factors and the output side possible outcomes, with adjustable hidden layers in between the input and output. A training set of cases is used to teach the ANN (with an initial optimised setting of the hidden layers) on how to decide outcomes. Every new case fed to the ANN indicates whether the outcome suggested by the network is correct or not, which allows the network to gradually improve accuracy of outcomes.

As another form of AI, deep learning has been used to replace rule-based approaches in the legal domain as they perform better with NLP, possibly due to reduced manual engineering required and the introduction of so-called *word embeddings*. Word embeddings are defined as "*low-dimensional dense vectors employed as word feature representations*" (Chalkidis and Kampas, 2018). These dense vectors incorporate semantic and syntactic correlations of words as opposed to manual metrics such as TF/IDF. Chalkidis and Kampas describe the following popular unsupervised algorithms to learn word embeddings (Chalkidis and Kampa, 2018):

- *WORD2VEC;* which uses the *skip-gram* and *Continuous Bag of Words* (CBOW) algorithms to form pairs of words found in the same *sliding window*. Skip-gram predicts a word based on the central word in the sliding window, whereas CBOW predicts the central word based on the other words in the sliding window (Mikolov et al., 2013; as cited by Chalkidis and Kampas, 2018: 173);

- *GLOVE*; which uses two sets of vectors to create word pairs, one for the words and one for the context. Each word is then represented as the sum of its corresponding vectors (Pennington et al., 2014; as cited by Chalkidis and Kampas, 2018: 174); and

- *FASTTEXT*; upgraded from WORD2VEC to deal with out-of-vocabulary issues as opposed to the previous two algorithms' fixed vocabulary (Bojanowski et al., 2016; as cited by Chalkidis and Kampas, 2018: 174).

The major drawback of ANNs compared to CBR and KBS is that an ANN cannot explain its outcome (Lodder and Oskamp, 2010). Even if one receives the desired output, it is of little use if the reasoning is not clear. On a practical level, a court would not be satisfied to hear that "*in 92% of similar cases the outcome was X…*" without a convincing underlying argument based on facts and applicable law (Lui and Chen, 2017). One should also consider the impact if case decisions are reversed in future; in the absence of an application explaining its predictions, legal professionals would not be able to assess whether those decisions still make up reliable sources to base arguments on.

An example of the use of ANN in law is *Judge on Wheels*, which uses a KDD approach that involves ANNs. It was created by a judge in Brazil to alleviate decision-making on traffic accident disputes at the scene of an accident with a program called *The Electronic Judge*. The application is now used on 68% of accidents in the state of Espirito Santo, enhancing the consistency and speed of judicial decision-making (Stranieri and Keleznikow, 2010: 114).

ANN can support KDD provided underlying processes and assumptions are clearly articulated to address the concern around absence of explanations (Stranieri and Keleznikow, 2010: 112).

3.2 **Evolutionary ML**

ANNs and Evolutionary Algorithms (EA) are examples of technologies that draw inspiration from principles of nature, with application in problems in optimisation, system identification and data mining.

Evolutionary Computing (EC) (Eiben and Smith, 2015) is described as the "*task of a collection of algorithms based on the evolution of a population toward a solution of a certain problem… The population of possible solutions evolves from one generation to the next, ultimately arriving at a satisfactory solution to the problem*" (Stranieri and Keleznikow, 2010: 124). Each domain has to find the algorithms that work best for it as there is no one-size-fits-all for best performance, confirming again why the extraction of domain knowledge is essential (Nissan, 2015). That could explain why a lot of emphasis is placed on the automatic acquisition of domain knowledge.

EA have a broad range of applications and have also been successfully applied in the legal domain with the creation of models of ontological evolution in legal reasoning and automatic extraction of specific domain knowledge (Priddle-Higson, 2010).

Five mainstream forms of EA for EC are listed below. The various approaches differ in solution representation, sequence of operations, implementation and parameters but they all imitate a population approach and principles of replication, variation and selection from evolution theory (Biethahn and Nissen, 2012):

- Genetic Algorithms (GA); one of the oldest optimisation techniques where a population of individuals (fixed length bit strings) represents a potential solution to a problem, which in turn is defined by its objective function. Each individual has a fitness attribute as evaluating factor. GAs have three operators: selection, crossover and mutation (Slowik and Kwasnicka, 2020);

- Genetic Programming (GP); relatively new and a specialised form of GAs which operates on very specific types of solutions (programs as opposed to bits or numbers as is the case with GAs) using modified operators (Slowik and Kwasnicka, 2020);

- Differential Evolution (DE); with efficient memory utilisation and lower computation complexity, used mainly for function optimisation in a continuous search space (Slowik and Kwasnicka, 2020);

- Evolution Strategies (ES); unlike with GAs where the next generation is created from the parental population, a temporary population is created. ES operate on floating point number vectors instead of the binary vectors used by GAs (Slowik and Kwasnicka, 2020); and

- Evolutionary Programming (EP); created to discover grammar of the unknown language and now popular due to its use as numerical optimisation technique (Slowik and Kwasnicka, 2020).

Advantages of EA include:

- their ability to continue with optimisation for as long as resources are available;
- flexible customisation for high quality solutions;
- a good fit for complex search spaces;
- reliability;
- applicable even where limited domain knowledge is available;
- they can be used in combination with other techniques; and
- efficient use of parallel hardware (Biethahn, J. & Nissen, V. 2012).

Disadvantages include:

- their heuristic character in that there is no guarantee for it to reach optimisation within a limited time;
- high processing requirements; and
- their theories are still in its infancy (Biethahn, J. & Nissen, V. 2012).

4. DEMAND FOR PREDICTIVE ANALYTICS

4.1 User experience

As part of this research process I conducted a brief survey with participants across large African law firms to (a) assess current case research practices; and (b) opinions on the use and potential value of commercially available predictive analytics tools. 71.4% of survey participants agreed predictive analytics on case outcomes would be useful in practice with 69.2% of those saying their answer would be different if no explanations for such predictions are provided.

Various case retrieval systems are available to legal professionals in Africa with Juta, LexisNexis, Practical Law, Sabinet, SAFLII and Westlaw being the most frequently mentioned in the survey. Some of these databases rate cases for relevance, but such relevance rankings (relied upon by only 42.8% of survey participants) are based on the underlying citation networks and not based on a comparison to new case facts.

Case law searches are typically done based on keywords entered into online case databases (often using variations or combinations of keywords); text book references; and indices on cases dealing with particular topics or legislative aspects. Survey responses relating to typical time spent performing case searches per matter were 35.7% saying less than 5 hours; 42.9% saying between 5 and 10 hours; and 21.4% saying more than 10 hours. Research is typically done by associates (50% of responses) or candidate (trainee) attorneys (35.7%).

LexisNexis shared results of their 2018 survey (LexisNexis, 2020) completed by legal analytics users of top 200 US law firms when asked about the value of such analytics:

- 90% said it adds value to their practice;
- 29% called it invaluable;
- 98% said it is valuable in determining strategies with respect to particular courts or judges;
- 96% said it adds value in predicting likely outcomes of strategies or arguments; and
- 94% said it adds value in performing case assessments.

Drivers of the use of these legal analytics were cost savings (according to 84% of participants) and winning cases and attracting new business cited by 71% respectively.

Interestingly, on being asked why their firms are not using legal analytics, 27% cited no one in their firm being trained on legal analytics as a reason. This highlights the need for upskilling legal professionals to improve their understanding of underlying methods of data-driven legal analytics applications.

4.2 **Human *versus* machine**

A statistical model looking at 628 US Supreme Court of Appeal cases over a term assigned each case equal weight in constructing classification trees to generate predictions. It focused on 6 characteristics of each case which were easily observable without specific training, namely (1) circuit of origin; (2) issue area of the case; (3) type of petitioner (e.g. the state or an employer); (4) type of respondent; (5) ideological direction of the lower court ruling (liberal or conservative); and (6) whether the petitioner argued that a law or practice was unconstitutional. They compared this system's predictions with legal experts' predictions which were based on a more limited review of the cases (simply due to human limitations) yet took into account actual facts of the cases and applicable legal doctrines. Although the experiment did not focus on opinions but rather just the outcomes of cases (so binary output as either affirmed or reversed), the statistical model performed better by predicting 75% of case outcomes correctly, with the experts scoring 59.1% correctly (Ruger et al., 2004).

The above is just one example where the question is raised whether future judicial models should perhaps explore the possibility of performing statistical analysis prior to deep diving into the facts of each case, in the hopes of improving consistency in decision-making while also creating an objective assessment process. Even a model as basic as VRCP could at least assist in pointing out anomalies, which could speed up the decision-making process by focusing judges' attention on the outliers upfront.

Relationships between legal problems and outcomes vary and is indicative of the nuances of legal reasoning (Conrad and Kofahi, 2017). Of course, that does not mean humans would perform consistently but perhaps, for the time being and rightly or wrongly so, the public still places more trust in humans than machines. This is ironic seeing that a 2011 study conducted on the sequence of parole decisions of experienced judges showed extraneous factors could influence decision-making on repetitive matters (Danziger et al., 2011). The Israeli study assessed 1,112 judgements collected over 50 days (over a 10-month period), consisting of parole requests (78.2% of the matters) or requests to change parole terms. The study found that the likelihood of favourable rulings was higher after each food break than later in a session. This led to the conclusion that experts are not immune to the influence of extraneous factors (and also, tongue in cheek, that the parody *"justice is what the judge ate for breakfast"* might just be appropriate when it comes to human decision-making…).

4.3 The role of ML

There is a twofold argument for using computer applications to model legal decision-making: Firstly, to drive a cohesive, multi-disciplinary study in this field; Secondly, to improve our knowledge of computational methods (Raghupathi et al., 2018).

No information is available on the cost-benefit analysis of the input required (manual annotation, domain-specific knowledge, application development etc.) for existing applications to predict case outcomes and whether it would be justified when looking at the current limitations of the output (especially where not used to its full potential by legal professionals).

Magic Circle firms Clifford Chance and Linklaters are among the growing list of global law firms who have introduced coding courses for their lawyers. In Africa, leading law firm Bowmans has implemented a program called *BowBots* as an introductory coding learning experience for its staff (Bowmans, 2019). Internet search results are rife with statistics on the impact of coding, programming and AI on the roles of lawyers, with some legal technology providers having to ensure lawyers that they are not aiming to create robot lawyers but rather to take the robot out of lawyers. As concluded by

Conrad and Al-Kofahi (Conrad and Al-Kofahi, 2017), *"The solution is not to take away the discretion of judges; rather, it is to make them aware of the data, to ensure their decisions are as informed as they can be".*

One could argue that it is not so much the roles of specialised legal professionals under threat, but rather those focusing on administrative tasks such as finding information, extracting key content or representing information. Virtual legal assistants such as ROSS (Ross Intelligence, 2020) are already used by some law firms in the United States to assist with legal research, with the American Bar Association quoted as saying *"ROSS Intelligence is an example of how artificial intelligence can be used to improve the delivery of legal services"* – frightening words from the perspectives of legal secretaries, paralegals and trainee lawyers.

5. ML IN LAW – CASE STUDY

Thomson Reuters products, specifically its time-keeping and billing system (3E) and research platform (Practical Law) are widely used by law firms, globally and in South Africa (and Africa). As such, the research below is of particular interest to demonstrate progress in research and development in predictive analytics on judicial decision-making.

5.1 Scenario Analytics

In their study *Scenario Analytics* (*Analyzing Jury Verdicts to Evaluate Legal Case Outcomes, 2015)*, Conrad and Al-Kofahi evaluate different scenarios' merits and consequences to address questions such as how long litigation is likely to take, whether it is warranted and should be pursued before a judge or jury and strategies for the most favourable outcome. By answering key questions, they sequentially built a data-driven legal decision support system. Their use of data mining, NLP and ML focus on the tasks of analyse and decide/predict, as outlined in *Figure 1's Quick Reference Map*. Their study claims to differ from previous in that it doesn't focus on providing statistics of judges' rulings but rather unearthing deeper patterns in the dataset. The aim was to identify, organise and analyse underlying fact patterns and legal strategies of similar cases, subsequently evaluating which strategies were most, or least, effective (assessed based on award levels and/or trial lengths) to support (and not replace) legal professionals in considering strategic options. Such data could potentially be used to build predictive models to forecast award levels and trial duration based on chosen strategies (to be refined per jurisdiction and litigation type).

5.2 **Experiments**

The study was conducted on Thomson Reuters' databases consisting of approximately 400,000 US jury verdict records across all 50 states, specifically chosen for consisting of shorter paragraphs than case law reports. The records contained more than 25 information fields such as:

a) date of activity (accident, filing, trial or settlement);
b) event-type (rear-end collision, sexual harassment etc.);
c) docket number;
d) jurisdiction (county, state, court);
e) case-type (liability, discrimination, malpractice etc.);
f) description (general and specific);
g) injury type (award category, award range, exact award);
h) damage summary (plaintiff profile); and
i) unstructured textual description of the event, plaintiff's claim and defendant's claim (these summaries were authored by Thomson Reuters employees trained to use a standard, semi-closed vocabulary to describe facts and claims of cases).

The following series of experiments supported by NLP methods was designed to create a legal decision support application.

5.2.1 Automatic topic classification

Thomson Reuters' Key Number System (KNS) is a legal taxonomy consisting of c.100,000 leaf nodes and c.200,000 total nodes (of which the depth of the taxonomic tree ranges from 3 to 11 with an average depth of 6). A key number assigner classification tool trained on editorially produced and KNS-classified points of law (headnotes) was applied to unstructured text containing factual descriptions of cases from three different litigation domains (namely premises liability, medical malpractice and racial discrimination). The objective was to explore leveraging KNS classification of unstructured textual descriptions of the facts and plaintiff claims and grouping these in *classes* (topics, being the three domains mentioned above). A 5-point Likert scale of "*on point*" (5), "*highly relevant*" (4), "*correct*" (3), "close to topic" (2) and "*poor classification*" (1) was used. Compared with the set as classified by legal domain experts, the automatic key number assignments were found to be reliable to capture essential features. Mean scores for identifying the correct domain for each case were 4.4/5.0 for premises liability, 4.1/5.0 for medical malpractice and 3.9/5.0 for racial discrimination. It is worth adding that an out-of-the-box key number assigner tool was used without having been trained on jury verdicts, indicating room for further improvement in accuracy.

5.2.2 Clustering plaintiff claims

Subsequent to automatically classify jury verdicts into classes as achieved through the previous experiment, the next experiment focused on differentiating one set of cases from another (within the classes) and based on underlying legal principles or strategies). Jury verdicts were segmented into four sections namely background facts, plaintiff claims, defendant claims and remaining details. The *NLTK 3.0 toolkit* was used to apply a *k*-means clustering algorithm over plaintiff claims. To differentiate between more effective and less effective claims, a metric based on award behavior for a given cluster *('award_quotient'* (AQ), as the ratio of a cluster's non-zero awards to its zero awards) was used to identify when clusters have a high degree of awards:

$$award\_quotient = \frac{(\text{cases with non} - \text{zero award})}{(\text{cases with zero award})}$$

A particular cluster under premises liability had a higher AQ and on closer examination, it appeared that the plaintiff's attorneys emphasised certain details such as the permanent nature of injuries or the fact that multiple injuries were sustained. Empirical examination of the language patterns corroborated such cluster-specific patterns. It was noted that these patterns are to be used as exemplar material in legal decision support and are not necessarily definitive results.

5.2.3 Associating language patterns with award distributions

Once cases have been clustered on topic (domain) and differentiated from others based on AQ and language patterns in cluster sets, the next step was to present the award distributions associated with each of the cluster sets. For example, this allowed a comparison of patterns in awards under the domain of racial discrimination in different clusters such as sexual harassment, disparate treatment or sex discrimination. Result showed that these data-driven patterns could guide legal professionals in using specific legal principles to construe arguments to produce similar strategies aimed at a particular award.

5.2.4 Analysing relations between trial length and award level

The last in this series of experiments was to examine the relationships between trial duration and award level. Two competing hypotheses were highlighted:

a) Short case hypotheses:
    i. No award: the case was dismissed quickly as it had no or little merit.

ii. Award: the case against the defendant was so strong with few mitigating circumstances that the award was determined promptly.

b) Long case hypotheses:

i. No significant award: the case was complex and blame was difficult to assign so a large award or the requested award was not granted.

ii. Significant award: the case was complex and took time to present all the issues, but blame was sufficiently decided thereby granting a significant award.

The data analysis indicated that short cases tended to have lower awards, but the opposite was not true and longer cases were associated with no, modest and large awards. In conclusion, there appeared to be no reliable correlation between trial length and award level.

5.3 **Results and conclusion**

The study concluded in saying that a system that automatically classifies and clusters case summaries (without manual annotation) and identifies language patterns to provide legal decision support is possible. While patterns and relationships vary by topic, it would still assist legal professionals in determining the most promising approach on litigation.

5.4 **Own conclusion**

This research focused on jury verdicts typically consisting of shorter paragraphs than other case law reports. Further research is required to assess how the methods described above would fare when applied to full case opinions.

The case study raised the fact that US states have different approaches to negligence and arguments may have to take this into account to adapt accordingly. One advantage to the South African jurisdiction (and most other African countries') is that this additional challenge does not apply as provinces do not operate as separate states with different laws. Legal professionals admitted in one court can practice in all provinces (with law society consent) without having to pass conversion exams, as is required in some of the US states.

A limitation to the above study from a South African perspective is that the KNS applied in the first experiment (to automatically classify cases by topic) does not apply to South African cases and its equivalent would have to be sourced or developed locally for future progress in this field.

6. DISCUSSION

6.1 Potential injustice and prejudice

An aspect of irony to predictive analytics is that it has the potential to provide two sides to a dispute with the same support and competitive advantage. If we look at the other four areas of application of AI mentioned in the introductory section above: applications to these fields are typically aimed at providing a party to a dispute or transaction with a competitive advantage in that the input they provide and analysis they apply would be unique to each party. As first example; with a system such as Kira used for contract content extraction, a party gradually trains the system based on their approach and knowledge base. One party's version of Kira could look very different to another based on its own learnings systematically applied to the system over a period of use. A further example is that of technology assisted review (TAR) or continuous active learning (CAL) used in eDiscovery processes, whereby a system could pick up on patterns in the human classification and tagging of documents forming part of evidence to a particular dispute. One party's review team would certainly not have the same approach as the other's in terms of relevance or grounds for argument. In other words, many commercial systems are aimed at providing a distinct competitive advantage, which parties could use to help them respond to this the 'more for less' challenge mentioned as introduction.

On the other hand, the benefits discussed in the second chapter are very much focused on public benefit. Where an application such as that focusing on predictive analytics in judicial decision-making is not based on one party's own cases but rather publicly available information (reported cases), the benefit goes to the party able to afford such analytical systems and the party not able to, potentially suffers the injustice.

It is predicted that data will be increasingly collected in a structured fashion as society becomes more information-based (Stranieri and Keleznikow, 2010: 129). An example of the benefits of a standardised approach to data collection is that of the Italian *Norme in Rete* project which requires all bodies to store its data according to a shared standard. This is done through the Extensible Markup Language (XML) which allows tagging of structures and adding of metadata to documents to make it readable by humans and machines. This enables ease of retrieval for further public use and distribution, with automatically created central indices (Biasotti et al., 2008). Should this type of standardised approach become embedded in courts' publications of legal decisions with some degree of legal relevance indicators, it would theoretically be possible to automate the analysis and prediction process. This could assist in leveling the playing field to some degree, as the underlying data would be available to all, as opposed to giving one party with access to resources to extract such data an unfair advantage.

6.2 Absence of commercial solutions

It is clear that there is not yet a commercially available solution to remove the need for construing arguments by legal professionals and the subsequent decision-making by skilled judges. While some applications can find information, extract content, hypothesise on and predict outcomes, this is not yet available to the extent that it will perform legal professionals' work of assessing argument validity, convincing courts of its correctness and ultimately making a binding decision.

In the field of eDiscovery and its version of predictive analytics (TAR, as referred to under 5.1 above), English courts (in the *Pyrrho* and *Brown v BCA Trading* cases) have been quite outspoken in promoting (and in some cases even demanding) its use to ensure costs of legal disputes are reasonable and proportionate to the matters at hand. Many providers of eDiscovery software now offer TAR functionality, as increasingly high volumes of electronically stored information require smarter methods of review for purposes of disclosing evidence in litigious disputes or regulatory investigations. It does make one wonder who will drive the research and development required to develop commercially available predictive analytics systems to be used for decision-making prediction. Australia's law society allows for listing of law firms, as done by Slater & Gordon Ltd in 2007 when it became the first law firm to list on a stock exchange (The Sunday Morning Herald, 2020). In 2012, England's Solicitors Regulation Authority introduced what is referred to as Alternative Business Structures (ABS) allowing non-lawyer ownership in law firms (Thomson Reuters, 2020). In the absence of allowing external investment into African law firms, the burden would lie on a law firm partnership to invest profits into research and development of advanced applications which, in turn, would:

i. be limited to a chosen domain(s) of law (due to the manual input currently required as already discussed earlier); and
ii. not necessarily be susceptible for broader commercial application outside that particular law firm.

This, combined with a legal industry not yet as comfortable with and sophisticated in the application of and reliance on advanced legal technologies as its European or American counterparts, certainly does place further research and development in the hands of commercial service providers.

6.3 Interesting global developments

In 2019, France has made it illegal to publish statistical information about judges' decisions, punishable with a five-year prison sentence (Artificial Lawyer, 2019). This approach is a first of its kind globally.

The English translation of this new Article 33 of the Justice Reform Act reads:

> *'The identity data of magistrates and members of the judiciary cannot be reused with the purpose or effect of evaluating, analysing, comparing or predicting their actual or alleged professional practices.'*

While some argue that this applies only to the personal information of court officials, that data would inevitably form part of publicly available court cases forming part of predictive analytics data sets. It would be interesting to follow the effects of and challenges to this new Article 33 to see how it plays out, and whether other countries follow suit. To date, the US and UK seem to have accepted the application of NLP and ML in analysing individual judges' decision-making patterns and behaviour.

6.4 Quis custodiet ipsos custodes?

Latin for '*who will guard the guards themselves*', it brings to mind the possibility of using predictive analytics as part of judicial clerks' research and preparation processes for decision making. South Africa's Constitution enshrines judicial impartiality in order to uphold the rule of law. Predictive analytics could be applied as supporting tool to avoid bias or extraneous influences such as that highlighted by the Israeli study discussed under 4.2 by providing a quick and effective overview of historic decision making to guide the decision-makers. Note that my suggestion is not to replace the role of judicial decision-makers but rather support it with relevant and accurate data. On the note of accuracy, South Africa (and as far can be ascertained, Africa) is still a long way from having an application with clearly explainable results and with legal professionals and the judiciary trained and equipped to use such an application(s). Should we manage to reach that stage and as discussed under benefits, predictive analytics applied to judicial decision-making could lead to an improved understanding of the application of law in practice, which in turn could reduce the class and gender divide found in many African countries.

Article 22 of the European Union's General Data Protection Regulation (GDPR, which came into effect in 2018) covers the aspect of citizen's rights to receive an explanation for algorithmic decisions (Goodman and Flaxman, 2017). By using ML in predictive analytics only in a supporting role in the judicial decision-making process (as opposed to allowing the ML application to make the final decision), one can avoid the conflict of rights between owing an explanation to applications' decision-making processes and improving access to justice.

To date, no advanced research has been published from a South African or African perspective as far as such predictive analytics on judicial decision-making are concerned. The legal community relies either on legal professionals to take a keen interest in legal informatics to progress research and development in this area, or for major research or intelligence providers to take a keen interest in the legal domain.

7. CONCLUSION

"Law is important, maybe critical, for the future of global justice and prosperity. Knowledge technology, appropriately managed, is important, maybe critical, for the future of law. Those of us who know

and care about both things need to exert disciplined and energetic effort if we expect positive change." (Lodder and Oskamp, 2010).

In closing and in support of Lodder's remark quoted above, I trust this dissertation will contribute to an improved understanding of and keen interest in key concepts in the application of ML for purposes of predictive analytics; the potential benefits; and challenges and limitations when applied to judicial decision-making from a South African (and more broadly an African) perspective.

In 2014, McGinnis and Pearce predicted that perfecting semantics in evaluating precedents will happen in the next "*ten to fifteen years*" (McGinnis and Pearce, 2014). I believe they might be correct and if so, legal professionals need to ensure they can adapt accordingly by incorporating tools such as predictive analytics into their practices to ensure continued relevance.

8. **BIBLIOGRAPHY**

Adnan, K. & Akbar, R. 2019. An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data*, (2019)6:9. [6 March 2020].

American Bar Association. 2013. *Publications*. Available from: https://www.americanbar.org/groups/business_law/publications/blt/2013/03/keeping_current/. [16 June 2020].

Artificial Lawyer. 2019. *Litigation Prediction*. Available from: https://www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers/ [13 April 2020].

Ashley, K.D. 2010. 'Case-based Reasoning.' In Lodder, A. & Oskamp, A. (eds) Information Technology & Lawyers. Available from: *Springer Netherlands.* [6 March 2020].

Ashley, K.D. 2017. Artificial Intelligence and Legal Analytics. *Cambridge University Press*. [6 July 2020].

Ashley, K.D. & Brùninghaus, S. 2009. Automatically classifying case texts and predicting outcomes. *Artificial Intelligence Law*, 17:125-165. [6 March 2020].

Biasotti, M., Francesconi, E., Palmirani, M., Sartor, G. & Vitali, F. 2008. Legal Informatics and Management of Legislative Documents. *Global Centre for ICT in Parliament*. [6 March 2020].

Biethahn, J. & Nissen, V. 2012. Evolutionary Algorithms in Management Applications. *Springer Science & Business Media*. [22 May 2020.]

Bowmans. 2019. *Press releases*. Available from: https://www.bowmanslaw.com/press_releases/bowmans-introducesbowbots/. [26 May 2020].

Branting, K., Yeh, A., Weiss, B., Merkhofer, E & Brown, B. 2017. Inducing Predictive Models for Decision Support in Administrative Adjudication. *The MITRE Corporation*, 17-1719. [29 April 2020].

Brùninghaus, S. & Ashley, K.D. 2003. Predicting Outcomes of Case based Legal Arguments. *ICAIL 2003*, Edinburgh, Scotland, UK. [6 March 2020].

Brown v BCA Trading Ltd [2016] EWHC 1464 (Ch) (17 May 2016).

BusinessTech. 2019. Technology. Available from: https://businesstech.co.za/news/technology/347680/uct-wants-to-launch-a-new-combined-law-and-technology-degree/. [26 May 2020].

Cardellino, C., Teruel, M., Alemany, L.A. & Villata, S. 2017. A Low-cost, High-coverage Legal Named Entity Recognizer, Classifier and Linker. *ICAIL 2017*, London, UK. [6 March 2020].

Chalkidis, I. & Kampas, D. 2018. Deel learning in law: early adaptation and legal word embeddings trained on large corpora. *Artificial Intelligence in Law (2019) 27:171-198.* [6 March 2020].

Conrad, J.G., & Al-Kofahi, K. 2017. Scenario Analytics: Analyzing Jury Verdicts to Evaluate Legal Case Outcomes. *ICAIL 2017,* London, UK. [17 June 2020].

Dadgostari, F., Guim, M., Beling, P.A., Livermore, A. & Rockmore. 2020. Modeling law search as prediction. *Artificial Intelligence and Law*. [29 April 2020].

Danziger, S., Levav, J. & Avnaim & Pesso, L. 2011. Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences of the United States of America.* [29 June 2020].

Eiben, A. & Smith, J. 2015. Introduction to Evolutionary Computing. 2nd ed. *Springer Publishing Company, Incorporated.* [6 July 2020].

Elnaggar, A., Otto, R. & Matthes, F. 2018. Named-Entity Linking Using Deep Learning For Legal Documents: A Transfer Learning Approach. [18 February 2020].

Franco, M.A. and Bacardit, J. 2015. Large-scale experimental evaluation of GPU strategies for evolutionary machine learning. *Information Sciences,* 330 (2016) 385-402. [18 February 2020].

Goodman, B. & Flaxman, S. 2017. European Union Regulation on Algorithmic Decision Making and a "Right to Explanation". *AI Magazine*. [1 July 2020].

Iftikhar, A., Jaffry, S.W.U.Q. & Malik, M.K. 2019. Information Mining From Criminal Judgments of Lahore High Court. *IEEE* (7) 2019 2169-3536. [10 January 2020].

Katz, D.M. 2012. Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal*. [17 October 2019].

Kerikmäe, T., Hoffmann, T. & Chochia, A. 2018. Legal technology for law firms: determining roadmaps for innovation. *Croatian International Relations Review*, XXIV (81) 2018, 91-112 [19 June 2020].

Lex Machina. 2020. *How it works*. Available from: https://lexmachina.com/what-we-do/how-it-works/. [17 June 2020].

LexisNexis. 2020. *Press Releases.* Available from: https://www.lexisnexis.com/en-us/about-us/media/press-release.page?id=1517245567230645&y=2018. [16 June 2020].

Liu, Z. & Chen, H. 2017.A Predictive Performance Comparison of Machine Learning Models for Judicial Cases. *IEEE,* 978-1-5386-2726-6. [10 January 2020].

Lodder, A.R. & Oskamp, A. (eds) 2010. *Information Technology & Lawyers.* Available from: *Springer Netherlands*. [6 March 2020].

Marx, M. 1970. Citation Networks in the Law. HeinOnline. [6 March 2020].

McGinnis, J.O., & Pearce, R.G. 2014. The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services. *Fordham Law Review*, Volume 82 Issue 6. [4 October 2019].

Moens, M. & Angheluta, R. 2003. Concept Extraction from Legal Cases: The Use of a Statistic of Coincidence. *ICAIL 2003*, Edinburgh, Scotland, UK. [6 March 2020].

Nissan, E. 2015. Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement. AI and Soc (2017) 32:441-464. [22 May 2020].

Priddle-Higson, A. 2010. *Computational models of Ontology Evolution in Legal Reasoning*. Mphil Thesis, University of Edinburgh.

Pyrrho Investments Ltd v MWB Property Ltd & Ors [2016] EWHC 256 (Ch) (16 February 2016).

Raghupathi, V., Zhou, Y. & Raghupathi, W. 2018. Legal Decision Support: Exploring Big Data Analytics Approach to Modeling Pharma Patent Validity Cases. *IEEE* 2169-3536. [10 January 2020].

Remus, D. & Levy, F. 2017. Can Robots Be Lawyers? Computers, Lawyers, and the Practice of Law. *Georgetown Journal of Legal Ethics,* (Vol. 30, Issue 3). [3 October 2019].

Rissland, e.L., Ashley, K.D. & Branting, L.K. 2006. Case-based reasoning and law. *The Knowledge Engineering Review*, Vol. 20:3, 293-298. [23 May 2020].

Ross Intelligence. 2020. *Home*. Available from: https://www.rossintelligence.com/ [30 June 2020].

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,* 104-4. [12 June 2020].

Sangkeettrakarn, C., Haruechaiyasak, C. & Theeramunkong, T. 2019. Fuzziness Detection in Thai Law Texts Using Deep Learning. Confernce of Information and Communication Technology for Embedded Systems, IEEE 98-1-7281-1921-2. [10 January 2020].

Schweighofer, E. & Winiwarter, W. 1993. Legal Expert System KONTERM – Automatic Representation of Document Structure and Contents. *HeinOnline*. [11 March 2020].

Slowik, A. & Kwasnicka, H. 200. Evolutionary algorithms and their applications to engineering problems. Neural Computing and Applications. [22 May 2020].

Stranieri, A. & Zeleznikow, J. 2010. 'Knowledge Discovery from Legal Databases – using neural networks and data minng to build legal decision support systems.' In Lodder, A. & Oskamp, A. (eds) Information Technology & Lawyers. Available from: *Springer Netherlands.* [6 March 2020].
Susskind. 2020. *Home*. Available from: http://www.susskind.com/. [19 June 2020].

The Sunday Morning Herald. 2020. *Business*. Available from:
https://www.smh.com.au/business/world-first-as-law-firm-lists-on-asx-20070521-gdq6wz.html.
[30 June 2020].

Thomson Reuters. 2020. UK Home. Available from: https://uk.practicallaw.thomsonreuters.com/7-518-2435?transitionType=Default&contextData=(sc.Default)&firstPage=true&bhcp=1. [30 June 2020].

Tran, O.T., Le Nguyen, M. & Shimazu, A. 2013. Reference Resolution in Legal Texts. *ICAIL 2013*. [18 February 2020].

Wikipedia. 2020. Transfer learning. Available from: https://en.wikipedia.org/wiki/Transfer_learning [15 April 2020).

Zhang, S., Yan, G., Li, Y. & Liu, J. 2019. Evaluation of Judicial Imprisonment Term Prediction Model based on Text Mutation. IEEE 19th International Conference on Software Quality, Reliability and Security Companion. [10 January 2020].