Artificial Intelligence in the legal sector. 
A comparative analysis of expert and AI approaches to predicting court decisions

Nadezhda Kaliazina
MSc in Business Information Systems

2019
Thesis Supervisor: Zsolt Ződi
Table of contents

1. Introduction ............................................................................................................... 1
2. Literature Review ...................................................................................................... 4
3. Theoretical Background ............................................................................................ 6
   3.1 Background of Artificial Intelligence ................................................................. 6
   3.2 Natural Language Processing as a Key AI Technique for Decision Making ...... 9
   3.3 Corpus Linguistic ............................................................................................... 15
4. Methodology............................................................................................................ 16
5. Analysis of AI Legal Software ................................................................................ 18
   5.1 Legal Software Landscape ................................................................................ 18
      5.1.1 Legal Technology Classification Frameworks ............................................. 18
      5.1.2 Legal AI Software Landscape ................................................................... 20
      5.1.3 Evaluation of Legal AI Solutions ............................................................... 24
   5.2 Legal Analytics & Prediction Software Overview. Outline of Deployed Methods
      and Techniques ...................................................................................................... 27
6. Analysis of Expert and AI Approaches to Predicting Court Decisions ................. 34
   6.1 Expert Approach to Predicting Court Decisions ............................................... 34
   6.2 AI Approach to Predicting Court Decisions ...................................................... 40
7. Limitations of AI Approach to Predicting Court Decisions .................................... 47
8. Analysis and results ............................................................................................... 52
   8.1 Summary of AI Legal Software Analysis Results .............................................. 52
   8.2 Comparison of Expert and AI approaches to Court Decision Making ............ 54
9. Conclusion ............................................................................................................... 58
10. References ............................................................................................................ 60
1. Introduction

Artificial Intelligence (AI) is not a new theory, and first talks about its leverage date back in the 1950s (McGuire, 2006). Nowadays artificial intelligence is becoming a more and more popular field of study which is developing intensively and is being used in different areas; it could optimize data mining and processing, improve the quality of products and business processes.

Another possible area of AI application is workplace automation. According to McKinsey research (2017), “Near 50% of current work activities are technically automatable by adapting currently demonstrated technologies” and “6 of 10 current occupations have more than 30% of activities that are technically automatable”. Although those findings are more related to routine work and primarily low-skill and low-wage roles, even the highest-paid occupations have a significant amount of activity that can be automated, including preparing staff assignments, reviewing status reports, and even analysing data to perform decision-making. Indeed, banks and insurance companies use machine learning techniques to assess risks when granting loans or insurances. Another example is bookmakers using similar techniques to determine the probability of winning to find the most profitable bets.

However, there are still many areas where a decision is mainly made by humans. One of the best examples is the judicial system, the automation of which in the forthcoming years seems unlikely due to the importance of its verdicts. Nevertheless, the analysis of decisions taken by the court is quite an actual issue as it could help lawyers to make risk assessments even before the start of the trial.

The objective of this paper is to analyse if the artificial intelligence approach for predicting court decision-making has advantages over the expert approach with the initial hypothesis that AI approach is more beneficial as it is more objective and accurate. To achieve this goal, and to better understand the technology behind it, this paper will also explore what artificial intelligence (AI) and neuro-linguistic programming (NLP), as a part of AI, is, and how legal applications leveraging those technologies. Therefore, this thesis aims to achieve the following objectives:

1. To analyse the scope of currently available AI applications in the legal industry
2. To analyse AI legal analytics and prediction software’s methods
3. To evaluate possible limitations of using AI in the legal sector and, particularly, court decision-making

4. To appraise the AI approach for predicting court decision-making (algorithms, variables, accuracy) and to compare it with the expert approach

The research questions that this paper aims to answer with the findings of the empirical part are the following:

- **Main research question:**
  - In which way the artificial intelligence approach for predicting court decision making is different from the expert approach to it?

- **Sub-questions:**
  - What are the areas of application of modern AI software?
  - What are the most common techniques and methods used by legal analytics and prediction software?
  - What are the characteristics of the expert approach for predicting court decision-making, its benefits and limitations?
  - What are the characteristics of the AI approach for predicting court decision-making, its benefits and limitations?

This paper is divided into six sections. The first Section presents the literature review describing the three most well-known scientific papers about the development of AI models to predict the behaviour of the court. Additionally, the article describing the limitations of using predictive technologies is considered.

The second section is for the theoretical background which contains descriptions of key research concepts such as artificial intelligence, neuro-linguistic programming and corpus linguistics.

In the third section, the methodology is presented. It summarises the main research methods mentioned in the literature review articles.

*The main part* elaborates the central topic of the thesis work. It begins with an analysis of the scope of currently available applications in the legal industry and their classification. Furthermore, it outlines the software technological background: methods, models and algorithms needed. Deep understanding of how juridical predictive analytics applications work helps us to investigate how those technologies could be used for
predicting court decision-making and to evaluate if the artificial intelligence approach is more beneficial than a human one.

*The limitations part* describes the possible constraints on using AI in the legal sector.

All observations from the previous parts are aggregated in *the analysis and results* section. It contains a comparison of AI and expert approaches for predicting court decision-making.

*The conclusion* is a summary of the paper; it discusses whether the thesis’s objective is reached and options for future research.
2. Literature Review

The use of AI for court predictions has been widely investigated in recent years. One of the first significant researches belongs to Ruger, Kim, Martin, and Quinn – professors from the leading USA Universities. In their paper “Supreme Court Forecasting Project: Legal and Political Science Approaches to Supreme Court Decision-Making” (2004), the decision tree algorithm predicting the court’s affirm/reverse results with accuracy 75% was developed. However, being pioneers in the fields, they did not manage to build a predictive model which could be flexible and broadly generalizable. The main limitation was that the court was unusually stable during the reporting period (1994-2000), consisting of the same nine justices.

This has led authors such as Katz, Bommarito, and Blackman to create a new model by employing a relatively new machine learning approach known as random forests in 2017. Researches used the database of the Supreme Court of Washington University in St. Louis as the main source for analysis. Aside from the outcome (affirm, reverse, other), this database contained hundreds of variables about nearly every Supreme Court decision for the period of past 60 years which helped to build a quite robust and generalizable model. It was able to predict 69.7% of the decisions of the Supreme Court of the United States. Although the accuracy was lower compared to Ruger’s et. al’s model, it was still a significant improvement over prior works being a more useful and innovative approach.

Another widely known research on the topic was conducted in 2016 by Aletras, Tsarapatsanis, Preoțiuc-Pietro, and Lampos from University College London. They created an algorithm that used only the textual information from the European Court of Human Rights judgments and gave its verdict in binary terms (affirm, reverse). Overall, 584 English-language decisions relating to the 3rd, 6th and 8th articles of the European Convention on Human Rights were analysed. Scientists chose these articles for two main reasons: they contained the great volume of data and the number of them was large enough to test the final mathematical model; it reached a degree of accuracy of 79%. Authors investigated that the most reliable factors for decision prediction were words describing the actual circumstances of the case. The study also proved that the European Court of Human Rights judges consider the actual circumstances of the case more than the formal legal norms.

However, some experts have expressed doubts about the use of predictive analytics in law. Limitations of this approach were discussed by Devins, Felin, Kauffman, and
Koppl (2017). They noted that predictive models are mostly based on the measured dimensions neglecting the unmeasured ones without knowing whether they are meaningful or not; it is similar to the multitask principal-agent problem in economics. Additional weaknesses of the approach refer to the big data design paradigm and risk assessment model. Overall, scientists identify that there is a trend for overestimating the advantages of big data and related technologies, not considering that it is, primary, an algorithm that cannot be updated without human interaction and moreover, cannot go behind its boundaries. The law system, on the contrary, is abstract and value-oriented, it is constantly being developed and interpreted in different ways by different judges.
3. Theoretical Background

3.1 Background of Artificial Intelligence

Scientists have not yet agreed upon a unique definition of Artificial Intelligence (AI). In this paper, we will refer to Barr and Feigenbaum, researches in the field of computing theory, who proposed the following definition of artificial intelligence (AI) in the early 1980s:

Artificial Intelligence is the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behaviour (1981).

Later the concept of AI was broadened; at the present time, several algorithms and software systems are attributed to be a part of AI, the distinguishing characteristic of which is the ability to solve problems in the way human would do. The characteristics of AI also include autonomy and adaptivity (University of Helsinki, 2018). Autonomy is the ability to perform tasks in complex environments without constant guidance by a user, and adaptivity is the ability to improve performance by learning from experience. In the literature, there are several frameworks of how AI methods and techniques can be classified. One of the examples is AI Knowledge Map which was proposed by Francesco Corea (Forbes, 2018).

In the AI Knowledge Map, X-axis and Y-axis stand for the AI Paradigms and the AI Problem Domains, respectively (Fig. 1). The AI Paradigms represent the approaches used by AI researchers to solve specific AI-related problems. It includes logic-based tools (knowledge representation and problem-solving), knowledge-based tools (based on ontologies and huge databases of notions, information, and rules), probabilistic methods, machine learning, embodied intelligence, and search and optimization. These six paradigms are classified into three categories — symbolic, sub-symbolic and statistical. Symbolic AI is using human concepts expressed via strings of characters or symbols; for given elementary problems we have available symbolic processors, which accept symbolic input information on their input side and create symbolic output information on their output side. Subsymbolic AI systems do not manipulate a symbolic representation to find solutions to problems. Instead, subsymbolic models are based on a metaphor of a human brain, where cognitive activities of the brain are interpreted by theoretical concepts that have their origin in neuroscience while the statistical approach uses mathematical tools to solve specific subtasks.
Figure 1. AI Knowledge Map (Forbes, 2018)
The AI Problem Domains indicate the types of problems AI can solve. Additionally, it shows the potential capabilities of AI technologies. They are the following: reasoning, knowledge, planning, communication, and perception, where perception is the ability to interpret, reason about, and transform sensory data.

There is a big volume of technologies in the AI Knowledge map. However, in the paper, we mainly focus on expert and fuzzy systems, decision network, and natural language processing (NLP). NLP will be dealt with in more detail in the next section.
3.2 Natural Language Processing as a Key AI Technique for Decision Making

Different approaches and AI techniques could be used to optimize the decision-making process. Nevertheless, NLP is believed to be one of the most efficient current methods for court decision making. NLP stands for natural language processing, an actively developing scientific discipline, which is aimed to recognize and analyze textual data. NLP-based models are able to solve a vast number of tasks, from simple to complex ones such as spell checking, keyword search, information extraction from websites, documents classification, machine translation, spoken dialogue systems, and complex question answering.

It should be considered that the problem of determining the possible court decision is based on the motivation part of the decision. Thus, it is similar to the task of determining the connotation of a document on the basis of its content (sentiment analysis from NLP), because for both cases the solution lies in the binary text classification. Therefore, it is worth to discuss some methods for solving sentimental analysis tasks.

1. Supervised learning

During supervised learning or learning with a teacher, machine learning algorithms determine the dependencies in the labelled data set, where the inputs and the outputs are known. The raw data divided into two parts. The first part is then used to train one of the classification algorithms (for example, support vector machines (SVM) or Naive Bayes), and the other one is used to test the trained algorithm. The majority of scientific works based on this method focuses on choosing the most relevant features (the variables found in the input data set that can sufficiently improve the accuracy of the predictive model). For instance, the study of Kushal, Lawrence, and Pennock (2003) shows that the use of sequences from a given number of consecutive words (N-gram) as features allows achieving better results than traditional machine learning techniques (like SVMs) and common metrics (like mutual information). This approach was used in the previously mentioned work on the prediction of decisions of the European Court of Human Rights (2016).

2. Sentiment Lexicons

In order to solve the problem of sentiment analysis, many algorithms use the so-called sentiment lexicons. These lexicons, in fact, are tables mapping words with a certain
numerical value describing the sentiment of a word. Often these lexicons are created manually, but there are also ways to build them based on various meta-data. However, this approach is not quite applicable for solving the problems in the legal industry as it is not customary to use words that clearly reflect the position of the author in the formal texts.

3. Text embedding

In the traditional NLP, words are considered as discrete symbols, which are further represented as one-hot vectors. A vector representation is a method of representing strings as vectors with values. A dense vector is constructed for each word so that words found in similar contexts have similar vectors. The vector representation is proved to be the starting point for most NLP tasks and makes deep learning effective on small datasets. Techniques of vector representations such as Word2vec and GloVe, created by Google (Mikolov, 2013) and Stanford (Pennington, Socher, Manning, 2014) respectively, are common tools used for NLP tasks.

a. Word2Vec

Word2Vec builds a semantic representation of the word. It is assumed that the context of words contains information about their meaning. A set of surrounding words in a given width (a model's hyperparameter) is selected around each word to train a model. There are two main ways to proceed to work with the context (Fig.2):

- **Skip-gram.** For this method, the context window containing k consecutive words is considered, where one word is missed. Next, the neural network, containing all words except the missing one, is learned to predict it. Therefore, if 2 words periodically share a similar context in a corpus, these words will have similar vectors.

- **Continuous Bag of Words (CBOW).** For this method, the context is represented by multiple words for a given target word. Each time the algorithm sees a word, the next word is taken. Next, the context words are used as the input of the neural network which predicting the word in the centre of the context. In the case of thousands of such context words and the central word, a dataset for the neural network is received. After, the neural network is trained and, finally, the output of the coded hidden layer is an embedding for a specific word.
The only drawback of Skip-Gram and CBOW is that they belong to the class of window-based models, which are characterized by low efficiency of utilizing the statistics of the corpus, which leads to non-optimal results.

b. **GloVe**

GloVe is believed to solve this problem by capturing the meaning of one word embedding with the structure of the entire visible corpus. Thus, it learns by constructing a co-occurrence matrix (word’s context) that basically count how frequently a word appears in a context using statistics and minimizing the standard deviation. This approach allows identifying the similarity of the word with the vector distance. Thus, text embedding creates some space for the semantic language description. At the same time, the given space has a certain internal structure; for example, words that are similar in meaning often are mapped to similar points in space. Moreover, in this space some metric relations are observed; the vector difference of the words “Russia” and “Moscow” is close to the vector difference of the words “Hungary” and “Budapest”.

Thereby, the text embedding model is allowing to construct a classifier that solves the problem of sentiment analysis using a simple bag of words, which is represented by the average value of its words. This representation can be used later for classical supervised learning.

4. **Deep Learning**

Deep learning is a part of machine learning, its subfield, that helps to achieve accurate results in several NLP tasks. It is the basis of today's outstanding solutions not only for
the problem of sentiment analysis but also for such tasks as machine translation and semantic text analysis. According to Zhang, Wang and Liu’s (2018) work “Deep learning uses a cascade of multiple layers of nonlinear processing units for feature extraction and transformation” (p.4). In the deep learning neural network, the layers which are closer to the inputs learn simple features, while more complex features, being the outputs of lower layers, are dedicated to higher ones (Fig. 3).

![Figure 3. Layers of Deep Learning Neural Network](image)

### a. Convolutional neural networks (CNN)

In the CNNs a multi-layer neural network with a convolutional layer is used. This technique is commonly applied for image classification and object detection models. It deploys a convolutional layer to the image with a set of filters and takes the new images it produces as inputs of the next layer. This is a benefit not only for object detection but for sentiment analysis as well. Collobert and Weston (2008) were among the first researchers to apply CNN-based frameworks to NLP tasks. The model builds word vectors and relevant features while the applied filters enable to highlight the intensely positive or intensely negative words. It will also learn particular words or n-grams that contain sentiment information. They then become inputs for a deep neural network which selects their best features to identify the sentiment of the sentence (Fig.4).

![Figure 4. Convolutional Neural Network for sentiment analysis](image)
CNN is an efficient AI tool as it allows mining semantic clues in contextual windows; however, it lacks preserving a sequential order and modelling long-distance contextual information. For this task, it is better to apply recurrent models.

b. Recurrent Neural Network (RNN)

RNNs can be effectively applied for analysis of sequential information as they process word one by one figuring out the sentiment after each step (Fig. 5). This approach allows memorizing the outputs of previous computations, afterwards, to use them for current ones. Nonetheless, as the distance in RNNs increases, they start to lose the ability to bind data. This issue was improved by the development of Long Short-Term Memory Recurrent Neural Network model (LSTM) by Hochreiter and Schmidhuber (1997). The key component of the LSTM is a cell state that passes through the whole neural chain representing the concept of long-term memory; information can easily flow through it without changing. LSTM can change information in the cell state by structures called gates. They determine which information will fall into the cell state, which will be erased from it, and which will affect the result of the next step. The drawback of this approach is the complexity of building a model capturing most of the useful long-term dependencies while avoiding the vanishing gradient problem.

RNNs and LSTMs are used to perform various NLP tasks such as word-level classification, language modelling, sentence-level classification, and semantic matching. It is relatively accurate for sentiment analysis tasks; however, it may slow down the evaluation process considerably.

![Figure 5. Recurrent Neural Network](image-url)
c. Recursive Neural Network (RNTN)

RNTN was firstly described by Socher (2013). It reflects language as a recursive structure – a hierarchy of words, sub-phrases, and other higher-level phrases. In such a structure, a non-terminal node is represented by all its children nodes. It looks similar to the tree model taking the output of the operation performed on a lower level (Fig. 6). One of the main benefits of this model is its interpretability as the result is seen for every node. They are also can be combined with LSTM techniques to solve the problem of gradient vanishing.

![Recursive Neural Network](image)

*Figure 6. Recursive Neural Network*
3.3 Corpus Linguistic

Corpus linguistics is seen as a research tool or a methodology, or, by some researches, as an academic sub-discipline that study the use of language based on large collections of its “real life” examples stored in corpora (or corpuses). Depending on the research question, corpora may contain newspapers, leaflets, transcriptions of spoken language, political speeches, text messages, fiction, videos, sign language, etc. Through the use of Corpus Linguistics, it is possible to investigate the language and the meaning of words and phrases. The basic tool which is used for this task is a concordance; a concordance is a type of a word display format which shows each occurrence of a word with a context. Concordances indicate the patterns and are often used to make dictionaries. Except for concordances, other tools adopted by modern corpus applications are:

- **Collocations**, statistical calculations of the words that most typically co-occur together;
- **Frequency lists**, a comprehensive list of words ordered either by frequency or alphabetically;
- **Keywords**, lists of words which are unusually frequent in the corpus in comparison to a reference corpus; like collocation, calculated with statistical tests.

Corpus linguistics can be used in several scientific fields, including lexicography, computational linguistics, cultural studies, psycholinguistics, NLP, and, many more. It also plays an important role in legal interpretation going forward. The main goal of Law and Corpus Linguistics (LCL) is to better understand the meaning of words in legal texts. A challenging area in this field is to discover the “ordinary meaning” of a legal text, which was examined in the paper of Lee and Mouritsen (2018). The ordinary meaning rule has both problems in the way it is theorized and the way that it is operationalized. First of all, judges still do not come to an agreement what “ordinary meaning” actually means. Secondly, they usually rely heavily on intuition and dictionaries in word interpretation. However, ordinary meaning cannot be based only on a single resource, many factors should be taken into account, including the context of an utterance, historical usage as well as the pragmatic aspects of the utterance, including the physical, spatial and social environment in which it occurs. Corpus linguistics allows measuring these factors and making identifying of ordinary meaning more transparent. Besides, corpus linguistics is believed to be an essential tool to apply for contract interpretation, intellectual property applications (patent claim construction or service mark’s genericness).
4. Methodology

Many authors have discussed the use of AI for court decision forecasting. We mentioned the most influential and relevant studies on the topic in the literature review part and now we consider the research methods they use. The summary is presented in table 1 where some additional articles were added which are used in the next chapters of the thesis. Category column is used to refer the study to its main topic.

Table 1.
Summary of the research methods in court decision forecasting

<table>
<thead>
<tr>
<th>Article’s Title</th>
<th>Author</th>
<th>Year</th>
<th>Methodology</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supreme Court Forecasting Project: Legal and Political Science Approaches to</td>
<td>Ruger, Kim, Martin &amp;</td>
<td>2004</td>
<td>Analytical research; survey</td>
<td>AI and expert approaches</td>
</tr>
<tr>
<td>Supreme Court Decision-Making</td>
<td>Quinn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A general approach for predicting the behavior of the Supreme Court of the</td>
<td>Katz, Bommarito, &amp;</td>
<td>2017</td>
<td>Analytical research</td>
<td>AI approach</td>
</tr>
<tr>
<td>United States</td>
<td>Blackman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predicting judicial decisions of the European Court of Human Rights: a Natural</td>
<td>Aletras, Tsarapatsanis,</td>
<td>2016</td>
<td>Analytical research</td>
<td>AI approach</td>
</tr>
<tr>
<td>Language Processing perspective.</td>
<td>Preoțiuc-Pietro &amp; Lampos</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Law and Big Data</td>
<td>Devins, Felin, Kauffman</td>
<td>2017</td>
<td>Secondary data analysis</td>
<td>AI approach</td>
</tr>
<tr>
<td></td>
<td>&amp; Koppl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Author(s)</td>
<td>Year</td>
<td>Research Type</td>
<td>Methodology</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------------------------------</td>
<td>------</td>
<td>-------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Lawyer as Soothsayer: Exploring the Important Role of Outcome Prediction in the Practice of Law</td>
<td>Osbeck</td>
<td>2018</td>
<td>Secondary data analysis</td>
<td>Expert approach</td>
</tr>
</tbody>
</table>

The aim of this research is to aggregate the results and finding from previous researches to complete a comparative analysis of AI and expert approaches to predicting court decisions. The research strategy is also to use online and printed secondary data sources to identify the core concepts and insights for the research.
5. Analysis of AI Legal Software

5.1 Legal Software Landscape

The practice of leveraging AI-related applications is affecting more and more business areas. This chapter of thesis work introduces an overview of AI software used by law firms and how they perceive these technologies and whether they are ready to implement them.

5.1.1 Legal Technology Classification Frameworks

The scope of solutions providing legal services is commonly known as Legal Technology or Legal Tech. The field is not a new one; first document management and artificial intelligence providers for law firms were set up more than 20 years ago (such as iManage\(^1\) and HighQ\(^2\)). Undoubtedly, Legal Tech now encompasses a wider range of technologies from machine learning algorithms for natural language processing, primarily, to enable to automate routine and day-to-day lawyers’ activities and to help with due diligence and disclosure exercises as well as performing quite specific tasks.

Many approaches exist with the purpose of classification legal software. One of them is a classification framework called “The legal innovation matrix” by Rackwitz and Corveleyn (2015). It is a matrix consisting of 4 quadrants: platform, network, software and know how (Fig.7).

Platform stands for a meeting point for supply and demand that provides access to legal services. Network includes people driven services; it is different from platforms as it offers services themselves in the form of managed services or on-demand staffing, especially during the period of a sharp increase in the amount of work, or specific expertise, or temporarily replacing a member of the team. Software quadrant is for solutions which directly performs legal tasks. The last dimension is know how, which focus on managing, creating, and delivering information workflow. All categories are driven either by people, process, resource or technology.

The Legal Innovation Matrix is a broad overview of the legal market which helps professionals to better understand the ways of legal service delivery. Many of law

---

1 Work product management solution for legal firms founded in 1995
2 SaaS file sharing, team collaboration and social network solution for legal and banking industries founded in 2001
companies are willing to be in the centre of the matrix, however, legal service providers are disintegrating and basically focussing on only one of the four quadrants.

![Legal Innovation Matrix by Rackwitz and Corveleyn](image)

*Figure 7. Legal Innovation Matrix by Rackwitz and Corveleyn*

In this paper, we will focus specifically on the software dimension of the matrix and on how to categorise solutions, it includes, based on functionality. To achieve it, another approach proposed by Praduroux, Paiva, and Caro (2016) should be discussed.

Praduroux et al. (2016) propose eight categories in which legal technology presence:

1. Lawyer-to-Lawyer Marketplace: online social networks connecting lawyers with clients and other lawyers;
2. Document Automation and Assembly: software enabling to optimize document workflow by populating automatically legal electronic documents;
3. Practice Management: software that allows managing operations, information, and schedule. Main functionality includes timekeeping, billing, and case management;
4. Legal Research: advanced technological software focusing on discovering useful information for building cases and drafting documents. Legal research applications often include analytics tools;
5. Predictive Analytics and Litigation Data Mining: software aiming to optimize the decision-making process. It analyses extensive volumes of data to find meaningful insights, and thereby predict future events;

6. Electronic discovery (also called e-discovery, ediscovery, or eDiscovery): “This is the electronic aspect of identifying, collecting and producing electronically stored information (ESI) in response to a request for production in a lawsuit or investigation” (Praduroux, Paiva & Caro, 2016). ESI includes various types of data such as text, images, calendar files, databases, spreadsheets, audio files, animation, web sites and computer programs. Many eDiscovery applications are integrated with databases or backup systems to streamline import and data collection processes;

7. Online Dispute Resolution (ODR): software helping to resolve disputes between parties online. Typical functionality enables secure upload of documents, proposal and discussion management, development of the settlement documents. ODR can be divided into two branches of solutions – technology-based systems (e.g. blind-bidding systems) and technology-assisted systems (alternative dispute resolution or ADR-related systems);

8. Data security technologies: software focusing on legal data protection and minimizing various security vulnerabilities.

Despite focusing on application areas, this approach may seem to be somewhat superficial, as, for example, eDiscovery has its own class while it can be considered as part of practice management or analytics and litigation data mining. Besides, it does not cover all possible application areas of legal software not addressing legal education and intellectual property. Nonetheless, it is a good basis to form a classification for legal software which leverages AI technology; we keep some categories, in which applications use AI, and add new ones not mentioned by Praduroux et al. We will identify them in the next section.

5.1.2 Legal AI Software Landscape

According to Forbes, investment in Legal Tech industry achieved more than 1.5 billion dollars in 2018. Interestingly, $362 million of all funding has been invested in legal solutions utilizing AI. This AI-focused funding alone in 2018 represents a bigger sum than the investment across all legal technology in 2017. The use of AI in the daily legal processes may transform almost every legal task. We adjusted the Praduroux et al.’s
framework and propose the following categories for categorisation of legal AI software: document automation, legal analytics, legal prediction, eDiscovery, legal research, legal due diligence, contract review, legal expert systems, eBilling, and intellectual property. Now, we will discuss them in a more detailed way with the examples and AI methods used in every category.

1. **Document automation**

   Document automation solutions allow managing the full contract life cycle and aiming to help businesses to reduce risks and increase sales. The typical functionality includes configuration and assembling of contracts to provide both contract parties with the tools to perform contract changes, store and access contract data, follow the workflow from inception to signing. The majority of solutions are usually integrated with customer relationship management (CRM) systems and use deep data analytics for reporting.

   This group also contains drafting process automation software which enables form filling and drafting of legal documents.

   *Main AI methods:* NLP, classification, information extraction, deep learning, optical character recognition (OCR)

   *Examples:* APTTUS, ContractPodAi, iManage, SURUKAM, LexCheck, Leverton

2. **Legal Analytics**

   Legal analytics solutions mine and aggregate legal data to search for valuable insights within it and to discover previously unknown patterns in the decisions of judges, lawyers, and other agents. They enable data-driven management, resource optimization, budgeting and pricing.

   *Main AI methods:* supervised and unsupervised machine learning, deep learning

   *Examples:* Clocktimize, IBM Analytics, Zero, Premonition, Bloomberg Law Litigation Analytics, Docket Alarm

3. **Prediction technology**

   Legal predictive software identifies relationships, patterns and trends through the legal data to statistically calculate a probability of an outcome or an opportunity. It primarily studies past juridical cases and judges’ behaviour to forecast how they might rule and in which timeframe. These solutions help lawyers to determine an optimal strategy to increase the rate of winning in litigation. The functionality of predictive
software is not restricted by case outcome prediction. It also includes early-warning systems that identify fraud red flags, for example, in the corporate emails. Some solutions can even leverage AI for crime prediction.

**Main AI methods:** unsupervised and supervised machine learning, NLP, deep learning

**Examples:** Case Crunch, Premonition, Gavelytics, Intraspexion, Lex Machina

4. eDiscovery

The definition of eDiscovery was identified in the 4.1.1 section of the paper. Generally, the eDiscovery process starts when a lawsuit is reasonably anticipated through post-trial. The main purpose of these solutions is to search across ESI (both structured and unstructured) for relevant data, at the same time, converting audio, video, and other content to the text format to enhance its understanding. In addition, eDiscovery software automates categorisation and pattern identification. It also supports businesses in reducing costs, meeting compliance requirements, and litigation readiness.

eDiscovery solutions got the lion’s share of investments in 2018 (Forbes, 2019).

**Main AI methods:** supervised machine learning, language detection, translation and transcription, object detection, face detection, identification, predictive coding

**Examples:** Veritone, OpenText, Text IQ, Catalyst, Relativity, Disco, Equivio

5. Legal Research

Legal research software is aimed to find out the solutions to legal questions by checking previous legal cases. It is typically more applicable to common law countries. Legal research applications revise case law, classify and encode it to provide relevant case examples needed to develop authoritative legal positions. Additional resources such as scientific and journal articles, reports are also reviewed to receive optimal results.

**Main AI methods:** NLP, machine learning, grammatical structure, word embeddings, cross-referencing

**Examples:** ROSS, WestLaw, Knomos, Bloomberg Law, Casemine, Casetext, Judicata

6. Legal Due Diligence

The term “due diligence” has been generally applied to the process of investigation taken before the conclusion of a treaty (primary mergers and acquisitions). Legal due diligence software reviews all the legal contracts and ancillary documents through the
company to determine their structure and obligations required in order to eliminate possible risks. Solutions of this type analyse both structural and linguistic anomalies and determinate outliers such as missing pages, additional clauses or abnormal wording. The use of AI methods for contract due diligence helps to automate document and clause clustering.

Main AI methods: supervised and unsupervised machine learning, NLP, pattern recognition

Examples: Luminance, Drooms, Kira, eBrevia, Diligen

7. Contract Review

Contract review or contract screening (pre-screening) software reviews and analyses a contract against an industry standard. In contrast to other categories, the end users of those solutions are both law professionals and ordinary audience. For lawyers, pre-screening applications highlight the key points from the large and sophisticated contracts; thus, lawyers are able to reduce contract review time by focusing only on the crucial contract parts. Another branch of screening solutions is usually presented in the form of chatbots that analyse user’s legal documents to summarise it to a brief overview which could be understood by non-professionals.

Main AI methods: NLP, deep learning, vector networks

Examples: ThoughtRiver, NDA Lynn, Legal Robot, Legal Sifter, LawGeex

8. Legal expert systems

Expertise automation solutions allow automation of expertise, reasoning and decision-making processes that used to be done by professionals in the field. They work on-demand delivering automated advice and answers to the legal questions; can be presented in the form of virtual assistants.

Main AI methods: machine learning, decision trees, rules-based algorithms

Examples: Neota Logic, Lisa, Autto

9. eBilling

Legal eBilling software’s goal is to automate invoice management by supporting tracking taxes, automated currency conversion, expense compliance and over-budget alerts. It allows minimizing manual paperwork to increase billing process efficiency.
These solutions also enable monitoring and checking the billing of lawyers’ fees and consulting services.

*Main AI methods*: supervised learning, NLP

*Examples*: Smokeball, SimpleLegal, Brightflag, InvoicePrep

### 10. Intellectual Property

Intellectual Property solutions allow tracking of trademarks, copyrights, patent utility models, domain names and other intellectual property. Their main goal is to optimize processes of docketing, invention disclosures, filing applications, valuing intellectual property portfolio, and budgeting. Moreover, intellectual property software provides law firms with a single centralised data repository that includes client information, license agreements, and opposition filings; it deploys data analytics for brand analysis and for proactively identifying upcoming issues. This software can often be integrated with case management applications to streamline the workflow of intellectual property procurement and litigation. However, some solutions in this group function solely as patent search engines or filing assistants.

*Main AI methods*: clustering and classification machine learning methods, deep learning


*Main AI methods*: clustering and classification machine learning methods, deep learning


This framework for classification AI landscape clearly has some shortfalls; the main weakness is that the usage of many solutions is not limited by only one group. Nevertheless, we believe it is still relevant as the examples of software for each group were chosen according to their core functionality.

### 5.1.3 Evaluation of Legal AI Solutions

The point that integrates all software mentioned above is that it is aimed to give law firms that will first implement these technologies a competitive advantage. Legal AI solutions allow optimizing the workflows, which reduces the costs of projects, and to demonstrate the ability to make data-driven decisions. Furthermore, they are able to reduce the time needed to complete tasks. In the experiment conducted by LawGeek in
2018, twenty top United States corporate lawyers were competing with AI. All participants had to read and analyse five Non-Disclosure Agreements (NDAs) in four hours and to identify more than 30 terms and problem areas in them. The average accuracy rate achieved by experts was 85% in comparison with 94% gained by the LawGeex AI algorithm. In addition, on average, the attorneys needed 92 minutes to analyse contracts, while the AI managed to do it in 26 seconds. In this area, AI has considerable potential to fully replace the human, however, in other subfields of the law industry, human intelligence remains an essential component creating teaching algorithms themselves and interpreting insights they provide.

While AI-based solutions have the capacity to transform a current legal workflow, it is still crucial to keep in mind that these technologies are not a panacea. For instance, in the contract management area, the use of AI is limited by access to the data it needs to process. Some organisations still keep using hard copies of the contracts or processing them in Excel and shared drives. The issue is not only about how to digitize legal contracts but how to create a unique structured knowledge base integrated them, so, AI can properly examine it. The data behind the AI algorithms is also a key factor of success for legal analytics and legal prediction as well as legal due diligence and contract review. Legal research sufficiency also deeply depends on available legal case databases. As for AI in intellectual property, AI-based software can analyse large volumes of data in the shortest possible time to control the counterfeit products. Meanwhile, it is difficult to create such a system due to the lack of a single database with images of all products for the analysis. For expertise automation, building a legal expert system requires deep legal knowledge in the phase of development and, then, system’s constant maintenance due to the nature of law which is constantly being changed and can be interpreted in many ways.

The systems available today, whilst advanced, are mostly focused on routine tasks. First, they are easier to automate as we can identify them as a certain algorithm which can be implemented into the program; secondly, they are easier to monetize as more potential customers are interested in them.

There are other obstacles such as an ethic and legal regulatory issue of using AI. New technologies are optimizing the work of lawyers, give them a room for professional development; however, artificial intelligence can never replace professions with moral and ethical aspects of work and in areas where creativity and communication skills are needed.
Therefore, the strengths and weaknesses of AI are not only about how accurate and sophisticated the algorithms are, but about the ability to work with them to reach the highest efficiency. Specialists are obligated to both program the AI systems, supervise them and review the results produced. As a further matter, legal professionals should be able to provide an understandable explanation of the gained insights and the decision behind them; the simple “Yes” or “No” labels are not enough for being evidence in the court. Thus, leveraging AI applications can be still a challenge for attorneys. In the eDiscovery field, for instance, one needs to understand mathematical statistics to use technology-assisted review (TAR) tools and to argue the findings during the trial.

Even though there are some constraints for AI in the legal industry, it does not mean that they will still exist in five or ten years. The ability of technology might change more radically and sooner than we expect; it is likely that we can wait a shift in the upcoming years.
Legal Tech market includes a wide range of solutions that have the potential to impact the practice of law and the lives of professionals. In the study, we concentrate on the software part of Legal Innovation matrix with an accent on AI-based software classification. This section of thesis work discusses its legal analytics and prediction technology category. Legal analytics area may be considered as a useful tool for the court decision prediction as it includes solutions that utilize historical legal data to identify patterns, trends and a correlation between variables. The input dataset usually contains a vast volume of unstructured data that later would be visualized via dashboards. Legal analytics is primarily descriptive.

Predictive analytics, on the contrary, processes historical data to make assumptions about future outcomes; in our case, to predict the outcomes of litigations. Solutions in both areas were selected by the funding raised and their mentions in the media to describe how they work and which technologies stand behind them.

**Intraspexion**

Intraspexion is the early warning system software with a focus on preventing litigation. It leverages deep learning to detect the risk of a potential lawsuit by checking the company’s emails for red signals in real-time; thus, corporate legal counsel can then check only a few flagged emails instead of thousands of them. Intraspexion uses Google's open-sourced deep learning TensorFlow algorithm. It is trained with past examples of litigation divided by specific type (so, later, the algorithm will be able to identify employment-related fraud emails assigned to a specific risk, for example, employment discrimination). After being deployed in the company, the algorithm is trained again. It indexes and extracts corporate emails with the embedded dtSearch text search engine, then uses a pattern matching mechanism to them against the fraud pattern. In the end, it identified two outputs — risk positive and risk negative. False positive and true positive results are then used to increase the algorithm’s accuracy. After several iterations, it is ready to score real-time corporate emails.

According to Intraspexion’s official website (2019), the software, on average, identifies near 100 red flag emails from every 2 million per month. It allows attorneys to significantly decrease the time of their check.
Lex Machina

Lex Machina is a software-as-a-service start-up with a focus on intellectual property litigation acquired by LexisNexis Legal & Professional in 2015 (Above the law, 2015). Its legal analytics platform provides insights on courts, judges, jurisdictions, parties and counsels to help attorneys to make data-driven decisions. It covers patent, trademark, copyright, antitrust, securities, commercial, employment, product liability, and bankruptcy litigation in federal courts.

Lex Machina captures raw data from public sources such as PACER (Public Access to Court Electronic Records), USPTO (United States Patent and Trademark Office), EDIS (Electronic Document Information System) and State Courts. As there is no single repository and court data is usually entered manually, the captured data is inconsistent and unreliable. OCR, NLP, text and data mining algorithms are used for cleaning and tagging data to create structured data sets. After, a proprietary language and tool-set classify legal text and extract key case events to label them as commercial, employment or intellectual property cases (fig. 8).

Figure 8. Lex Machina Technology Algorithm (Lex Machina official website, 2019)
Lex Machina deploys decision tree model to predict outcomes in both pre-litigation and litigation phases. The algorithm provides an output with such variables as prelim injunction, likely outcome, likely phase, damages costs, perm injunction, and time to termination. Besides, the solution gathers feedback from leading law firms and corporations to increase model accuracy.

Lex Machina customers are more than half of the American Law 100, and large corporates including Microsoft, Nike, eBay, Uber, and many others. It is free to use for academic institutions, federal courts, and other public-interest entities.

**Docket Alarm**

Docket Alarm is a legal analytics software automating checks of updates on a legal case’s docket. It also provides PTAB (Patent Trial and Appeal Board) analytics for case outcomes, timing and trends across subject matter, judge, law firm, and more. Its coverage covers 16 state courts and 250 million legal documents. The solution supports analytics workbench which allows users to set their own analytics based on customized variables. Docket Alarm data is open for integration through API.

Docket Alarm’s predictive module starts to work when an attorney enters legal facts into the system; those facts could be the assigned judge, law firm, party name or technology area. The system uses NLP semantic and pragmatic pre-processing to look for similar cases and text and data mining for a pattern search. Averaging litigation outcomes together, it generates a report with over 98% accuracy (Forbes, 2015).

**Premonition**

Premonition is a legal analytics solution which provides services in areas of legal services, insurance, health care and education. It analyses data on court and litigation, law firms and lawyers reports, panel selection, expert witnesses, arbitrators and others.

Premonition declares to have the world’s largest litigation database, it integrates publicly available data from most states and all federal courts in the USA as well as data on courts from the United Kingdom, Ireland, Australia, the Netherlands, Canada, India and the Virgin Islands. State level court data integration provides coverage of 97%. Premonition’s algorithm is able to analyse over 50,000 records per second by web-scraping operations via virtual private networks (VPNs). It facilitates Big Data technology and NLP to normalize cases found containing the data about attorneys, case types, judges and litigants. All output datasets are delivered via API and CSV to avoid
extra cutting and pasting. It also deploys machine learning, but most of the analytics does not rely on it.

The solution differs from other applications as for analytics it does not utilize valuables connected to the facts or the law. Its model found that “the judge-attorney pairing is worth 30.7 per cent of an outcome” (Above the Law, 2015) and that there is no correlation between a lawyer’s hourly rate and his performance. Premonition model heavily considers the context: it does not compare judges and attorneys from different courts and states, all cases are discussed within their specific types (108 case types by now), easy and hard cases are also taken into account for overall performance analysis.

**Ravel Law**

Ravel Law is a legal analytics solution delivering and visualizing insights on courts, judges, cases and over 400 United States law firms. The insights can be seen in three perspectives:

- **Case analytics.** It highlights how cases and their context are cited in other court filings. It is made with NLP instead of using keywords;
- **Judge analytics.** It shows the behaviour patterns of judges and relative cases, the percentage of grants and denies over motions. The insights are delivered by machine learning;
- **Court Analytics.** It analyses forums to assess possible outcomes showing patterns of each court with certain types of actions and motions.

Before providing its services to the customers, Ravel Law needed to solve one problem – access to a meaningful amount of legal data. The issue was overcome by partnering with Harvard Law Library; throughout a large project, approximately 10 million court decision were digitized to design an algorithm and be able to receive some insights. It was later trained with this input and retrained according to experts’ review. Currently, Ravel Law leverages NLP and data mining for understanding and extraction data from unstructured databases. It uses Scala, Spark, MongoDB, on SolrCloud on AWS for the law analysis and for deployment of the API. Then, machine learning is applied to mine and summarize legal cases.

**Bloomberg Law Litigation Analytics**

Bloomberg Law Litigation Analytics is a legal analytics software that proceeds legal data to recognize meaningful insights to help law firms with identifying their
litigation strategy and make a data-driven decision. The key Bloomberg Law Litigation Analytics capabilities include attorney, judges, law firms, parties and cases analytics. It allows seeing the whole judicial company’s history (lawsuits, law firms representing it, a company's federal litigation history), law firm’s federal litigation portfolio, judges’ portfolio, more cited opinions, appeal and motion outcomes, an average case length, attorneys’ portfolio and many more.

Litigation Analytics uses data of Bloomberg's company and Bloomberg Law's dockets information plus active and federal district court judge data. According to the solution’s official website, its data set includes “more than 3.5 million companies, 7000 law firms, 100 000 attorneys, and all active federal district court judges”. To extract, understand, and process data from these sources (normalization or digitization process) the software applies statistical text mining techniques. Its proprietary core consists of two levels; the robust real-time text mining fulfils low-level tasks such as tokenization, chunking and parsing, while, on the top-level, so-called entity extractors define firms and professionals in natural texts across Bloomberg’s news and social text databases for the following NLP sentiment analysis. In addition to that, the company leverages machine learning to accelerate case law research, to tag documents with normalized topics, and to uncover the underlying case law argumentation for a particular decision. For structured data analytics, Bloomberg Law deploys table detection and segmentation tools which help to increase the volume of ingested data.

To deal with data normalization issues that occur when law firms merge or change their names, machine learning and statistical algorithms are used. They validate data across Bloomberg's company data set. Unmatched companies are reviewed by the QA team.

**Casebook**

Casebook.ru is a Russian prediction and analytics solution that provides customers with information on litigations and legal entities participating in court proceedings especially bankruptcy proceedings (Casebook, 2019). It mines the company’s registration, management and shareholders data, and allows law firm evaluating the risks of concluding new deals and finding reliable partners.

The solution aggregates real-time data such as annual reports of legal entities, information about penalties and changes in pledge and non-tax liabilities, judicial activity of the company from public sources including database of decisions of arbitration courts.
and courts of general jurisdiction, the Federal Tax Service, the USRLE (Uniform State Register of Legal Entities) (also EGRUL).

Casebook has a focus on litigations of arbitration courts; its scale covers 99% of all registered in Russia companies and helps its clients to evaluate any firm with regard to its obligations to counterparties to reduce possible risks. The software deploys supervised machine learning to predict the outcome of a lawsuit based on similar cases and the history of decisions made by a designated judge. To train model, it uses the following features — plaintiff, defendant, claim amount, judge, court proceeding a case, and category of a case. The features are transformed into a text to be performed as TF-IDF (Term Frequency-Inverse Document Frequency) vector. After that, Naive Bayes classifier predicts to which category (claim returned, claim satisfied, claim partially satisfied, claim denied) a case belongs. As for now, the model has an accuracy of 82% (Yandex Academy, 2018). Casebook automatically reviews and recounts the probability of an outcome when there is a change in case data. A predicted result, as well as company’s information about claims, bankruptcy, litigation, outstanding loans, can be imported in the format of PDF, XLS, Doc.

Besides from its core functionality, now, Casebook is developing a new project which will predict a probability of execution of a judgment, as its average rate is near 50% for Russian Federation (Interfax, 2017).

Conclusion

Summarizing, the landscape of legal analytics and prediction software does include a few solutions, but their number is not compatible with the one from other industries. For instance, financial institutions, retail and FMCG firms have already actively used predictive analytics for classification and clustering their customer profiles for some years mining variety of data sources including CRM data sets, financial, sales and social media. The legal industry has the potential to do the same in a matter of judges and attorneys analysis as the field is still not mature. As for now, legal and predictive analytics is not separate core functionality of the solutions but mostly additional component helping the Legal Tech companies to gain a strategic advantage. Except for standing alone applications, the two main sectors of legal analytics implementation are case law research and eDiscovery (Disruptor, 2018).
One of the triggers of the legal analytics and prediction landscape evolving is the extensive volumes of available legal data. All the discussed above solutions are based on crawling, identifying and processing of various data sets from the public and their own databases. After that, they apply basic NLP and machine learning algorithm for analytics, predictions and visualisation in a form of dashboards. By and large, the majority of currently available solutions use the same AI techniques with a slight difference in the implemented methods and for a different final goal. The challenge for them is to create the fullest and the most large-scale data set with clean and meaningful data to avoid inaccurate results emerging during comparison a case against an incomplete data set with missing or erroneous case data. Thus, to become a leading solution in the industry, one should obtain a large volume of clean and accurate data as well as practice-specific tags that allow identifying the most relevant cases. This way, almost every aspect of law will become available for a detailed analysis.
6. Analysis of Expert and AI Approaches to Predicting Court Decisions

Predictions are the indefeasible part of everyday lawyers’ activities. Legal experts have to provide clients with information about possible prospects in litigation and associated costs. For decades, the human approach to the litigation outcome predictions was the only available option. The world has moved on, and nowadays, with the development of the technology field, statistical and algorithmic models are used on an equal basis with the human approach. In this chapter, we will consider the AI and experts processes of generating court decision predictions in order to determine general algorithms, factors or variables which are leveraged, and accuracy of both approaches.

6.1 Expert Approach to Predicting Court Decisions

Lawyers’ decision-making in everyday professional activities is inseparable from prediction. Accepting the case, they try to define the basic litigation metrics such as the likelihood of case duration, the court where the claim could be heard in, and the quality and attitude of its judges. They also consider how strong the adversary is, how likely it is to gather additional important evidence or whether the plaintiff will “fight up to the last-ditch” or change her mind and decide that the claim is no longer worth the time and effort to pursue — after the legal expert has invested her time, effort and talents. All these factors, and the likelihood of them to happen, exert an influence on the fees and expenses of the litigation. That’s why the correct prediction of case outcome is one of the greatest skills a lawyer gets with increasing professional experience.

Generally, legal experts can predict an outcome of a case by reading relevant law codes, relevant statutes, relevant court rules or relevant precedents which depends on the country’s legal system (civil law, common law, or statutory law). They analyse details of a case, court opinions, parties’ briefs and other legal materials. Needless to say, professionals consider a lot of factors behind the law boundaries to make their predictions.

The element-focused analysis, lawyerly experience, and the use of certain types of empirical information were named the tools attorneys rely on during case outcome forecasting (Osbeck, 2018). In his paper, Osbeck defines the element-focused analysis as an analytical process of breaking down the reasons of action or defence into its components, at that point deciding for every component whether it applies considering the known facts. Usually, this analysis takes place in the early stages of a dispute.
Lawyerly experience stands for professional intuition, knowledge and past legal experience; it helps not to skew the forecasting in favour of the customer. Empirical information that attorneys use includes past jury verdicts, confidential settlement data for cases when it is available, and jury behaviour for choosing an optimal strategy in the court.

In the research of Goodman-Delahunty, Granhag, Hartwig, and Loftus 481 litigating attorneys, including new graduates and seasoned practitioners, from 44 states of the USA were asked to forecast litigation outcomes by answering the questions – “What would be a win situation in terms of your minimum goal for the outcome of this case?” and “From 0 to 100%, what is the probability that you will achieve this outcome or something better?” (Goodman-Delahunty, Granhag, Hartwig & Loftus, 2010). The experiment was based on the 481 civil and criminal. After the trials, the litigation outcomes were compared with the initial lawyers’ predictions. Results were classified into three groups: failure in achievement of minimum goals, achievement of goals, and goals exceeding. The outcomes of the cases were consistent with the goals set by the participants in 32% of the cases. Among the rest, lawyers were too modest only in 24% of the cases while being overconfident of the success of the case in 44%. This paper illustrates that a large portion of legal experts tends to think too positively during forecasting. However, nearly one-third of the participants were correct in their predictions.

Goodman-Delahunty et al.’s research has tended to focus on psychological aspects (confidence and calibration) of outcome prediction rather than case-related factors experts consider throughout this process. On the contrary, the study of Ruger, Kim, Martin and Quinn (2004), besides the development of a forecasting model, investigated the legal experts’ prediction making. 71 academics and 12 appellate attorneys were chosen among others based on their experience, training, researches and referrals from other colleagues to participate in the experiment. “Of this group, 38 clerked for a Supreme Court justice, 33 hold chaired professorships, and 5 are current or former law school deans” (Martin, Quinn, Ruger & Kim, 2004, p.3). Participants were inquired to predict a case outcome inside their professional areas without communication with one another. They were free to consider any information sources or factors that thought to be relevant to the prediction. Their forecasts took the frame of an “affirm or reverse” choice; this binary choice limitation was due to the need for the following comparison of the expert approach to the statistical model.
Even though it was incomprehensible to identify absolutely every aspect of expert forecasting, some data about the influencing factors were gathered. In one-two weeks after making a prediction, participants were asked to fill a survey to list these factors. The conducted results are presented in figure 9. As the figure illustrates, the experts considered traditional legal factors and materials such as precedent, statutory text, court decisions, and the briefs in the case as the most important. They also relied heavily on the previous Supreme Court opinions. Half of the variables used in the statistical model — a circuit of origin, an identity of the petitioner, and an identity of the respondent — were regarded as insignificant by most participants.

<table>
<thead>
<tr>
<th>Factor (Factors in bold were rated as important by a majority of respondents)</th>
<th>Mean Response</th>
<th>% Rating Factor As Not Important (1 or 2)</th>
<th>% Rating Factor As Important (4 or 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity of the court whose decision the Supreme Court is reviewing</td>
<td>2.1729</td>
<td>64.7</td>
<td>17.5</td>
</tr>
<tr>
<td>Existence of a divided court below</td>
<td>1.9792</td>
<td>66.7</td>
<td>10.4</td>
</tr>
<tr>
<td>Extent of disagreement in the circuits and/or state courts on the issue</td>
<td>2.5268</td>
<td>49.7</td>
<td>21.1</td>
</tr>
<tr>
<td>Identity of the petitioner</td>
<td>1.9470</td>
<td>71.9</td>
<td>13.7</td>
</tr>
<tr>
<td>Identity of the respondent</td>
<td>1.8788</td>
<td>75.8</td>
<td>12.1</td>
</tr>
<tr>
<td>Identity of counsel representing the parties</td>
<td>1.3806</td>
<td>92.6</td>
<td>4.5</td>
</tr>
<tr>
<td>Quality of the parties’ briefs</td>
<td>1.9776</td>
<td>66.4</td>
<td>10.4</td>
</tr>
<tr>
<td>Supreme Court precedent on point</td>
<td>3.8666</td>
<td>15.6</td>
<td>69.0</td>
</tr>
<tr>
<td>Supreme Court dicta on point</td>
<td>3.3947</td>
<td>23.6</td>
<td>54.4</td>
</tr>
<tr>
<td>Other statements by the Justices in prior opinions</td>
<td>3.2061</td>
<td>31.3</td>
<td>49.6</td>
</tr>
<tr>
<td>Text of relevant constitutional provision(s)</td>
<td>2.77/1</td>
<td>82.7</td>
<td>21.0</td>
</tr>
<tr>
<td>Text of relevant statute</td>
<td>3.5495</td>
<td>22.5</td>
<td>54.0</td>
</tr>
<tr>
<td>Text of relevant regulation</td>
<td>2.6250</td>
<td>52.1</td>
<td>35.4</td>
</tr>
<tr>
<td>Non-sexual evidence of meaning of constitutional, statutory, or administrative provision (e.g., legislative history, long-standing practice, etc.)</td>
<td>3.1327</td>
<td>50.1</td>
<td>44.3</td>
</tr>
<tr>
<td>Interpretive theories of the Justices</td>
<td>3.5564</td>
<td>19.0</td>
<td>62.9</td>
</tr>
<tr>
<td>Practical consequences of the decision</td>
<td>3.9254</td>
<td>9.7</td>
<td>73.8</td>
</tr>
<tr>
<td>Policy preferences of the Justices on the specific issue presented</td>
<td>3.6045</td>
<td>23.8</td>
<td>65.0</td>
</tr>
<tr>
<td>The conservative or liberal ideologies of the individual Justices</td>
<td>3.3282</td>
<td>29.0</td>
<td>54.2</td>
</tr>
<tr>
<td>Public opinion on the issue</td>
<td>1.7067</td>
<td>78.1</td>
<td>9.9</td>
</tr>
<tr>
<td>Composition and preferences of Congress</td>
<td>1.5388</td>
<td>90.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Composition and preferences of the Executive Branch</td>
<td>1.5429</td>
<td>84.8</td>
<td>4.6</td>
</tr>
<tr>
<td>The professional backgrounds of the Justices</td>
<td>1.5843</td>
<td>86.5</td>
<td>4.5</td>
</tr>
<tr>
<td>The personal backgrounds of the Justices</td>
<td>1.5970</td>
<td>82.1</td>
<td>5.2</td>
</tr>
<tr>
<td>Number of amici participating in the case</td>
<td>1.5984</td>
<td>87.8</td>
<td>2.4</td>
</tr>
<tr>
<td>Identity of amici participating in the case</td>
<td>1.7997</td>
<td>76.6</td>
<td>7.3</td>
</tr>
<tr>
<td>Position of the Solicitor General in an amicus filing</td>
<td>2.5904</td>
<td>69.0</td>
<td>97.5</td>
</tr>
</tbody>
</table>

Figure 9. Summary Statistics of Survey Responses (Martin, Quinn, Ruger & Kim, 2004)
The expert approach to forecasting had an accuracy of 59.1% counted as a summary of all available expert predictions. In the case of a unanimous or majority consensus result, this measure could be 65.6%. The rate of successfully predicted outcomes also varied for different case types and judges.

Among other factors influencing the outcome forecasting researches, Fox and Birke emphasized the specification level (2002). The study conducted with the help of 200 practising attorneys with an average 17 years’ experience revealed that legal professionals estimated the likelihood of an event to be higher when it is unloaded into an explicit disjunction (or separate evaluation) of constituent events. This hypothesis was proven for both categorical partition with qualitatively components (case settles, case is dismissed, case is dropped, etc.) and dimensional partition with quantitative continuum (award to plaintiff less than $25,000, $25,000–50,000, $50,000–100,000, over $100,000). Thus, Fox and Birke recommended lawyers to evaluate the probabilities of each part independently, at that point alter their values at the same time so that the probabilities sum was 100% in order to make more accurate forecasting.

Listed above prediction influencing factors are mainly connected with the case itself. However, attorneys not only analyse them but also rely significantly on the opinions of their colleagues. Jacobson, Dobbs-Marsh, Liberman, and Minson discussed this topic in their paper “Predicting Civil Jury Verdicts: How Attorneys Use (and Misuse) a Second Opinion” (2011). In their research, two groups of law student and practice attorneys were asked to estimate the full noneconomic award in personal injury cases. First, participants needed to make predictions by themselves, secondly, to discuss a case in with an option to adjust initial prediction, third, pairs were obligated to give one unified answer, and, in the end, they were free to give weigh for personal and partner’s prediction for overall estimation. Scientist revealed that for both groups the mean error decreased dramatically from round to round — from 0.806 to 0.292 for law students and from 0.387 to 0.130 for attorneys. The latter were more likely to fully ignore their partners’ estimations. This study has proved that the expert approach for court outcome forecasting could be more beneficial and accurate than the machine one as it includes the possibility of so-called “the wisdom of the crowd”.

Even considering all the aspects affecting legal expert prediction making process, it is impossible to see the whole pictures without understanding how judges actually judge. The answer to this apparently simple question turned out to be quite complicated. The response could be represented by two opposite models — the formalistic and realistic
(Guthrie, Rachlinski & Wistrich, 2007). The former is generally understood to mean that judges apply the law in a logical, mechanical and deliberative manner to the facts of the case. In contrast, the realistic theory proposes that judges apply intuition to make a decision and only after that they rationalize it with reasoning. The intuitive decision-making approach is believed to be more native and simpler than the deliberative one. Neither of these models exists in the legal world in their pure forms. Judges surely do not ignore facts of the cases and apply legal norms, meanwhile, sometimes, they trust their intuition.

Not only intuition, but other subjective elements have an impact on judicial decision making. In the study of 2016, Chen, Moskowitz, and Shue found that the judges are subject to the phenomenon of “gambler’s fallacy” when one starts to doubt the correctness of her decision if she has taken several of them in a row. For instance, after several following each other decisions to grant asylums for migrants, the judge tends to deny the next because of the belief that he becomes too indulgent.

Other anomalies that have a correlation with a court sentence were listed in the paper of Chen aggregating scientific works in the field (2019). Among them are such arbitrary and unfairly factors as:

- Political situation: U.S. federal appeals court judges are more politicized before upcoming elections while being more unified during the war;
- Discriminative factors:
  - Political views,
  - Race,
  - Masculinity;
- Others:
  - Birthday and name of the defendant,
  - Football game outcomes,
  - Weather,
  - Courtroom temperature,
  - Shared biographies.

These anomalies do take place due to a human factor. Judges, like other individuals, have two cognitive systems for making judgments — the intuitive and the
deliberative. Still, most judges use facts, evidence and highly restricted legal criteria to perform their everyday activities avoiding personal biases, attitudes, emotions and other subjective components. The intuitive approach might be a solution in some cases, but it can be a reason of unethical results in others.

Litigation, apart from being time-consuming and relatively expensive, always involves some risks. Before representing a client in the court, an attorney is often asked to estimate the probability of a successful outcome and to assign a value to a case. In determining how to approach the case, whether to settle or take its chances at trial, a client usually trusts the analysis of an attorney. Therefore, the consequences of a case outcome forecast error may result in the loss of credibility to an attorney and extra costs for a client. An attorney’s professional repute also depends dramatically on calculations of possible litigation’s results.

Legal schools prepare future lawyers to analyse cases from various perspectives; students learn how to discuss all aspects of a case in a form of a class discussion, an essay, an exam or a moot court event. Making first career steps, novice lawyers usually work with more experienced colleagues who encourage them to carefully analyse the strengths and weaknesses of each case. As lawyers climb the career ladder and are exposed to more cases, verdicts and statistics, their judgments are likely to become more accurate and consistent. However, in practice, such predictions are often considered to be an art, not a science. An assessment of the potential liability of a client or the likelihood of recovery may include many factors, ranging from examining the applicable legal principles or a case partition to considering the likely composition of the jury or policy preferences of the justices. Ultimately, a lawyer's assessment is usually subjective in assessing these different factors and the strength of each party's arguments. This subjectivity can be beneficial to lawyers providing them with the freedom to rely on their experience and to communicate with other experts to increase the rate of correct litigation outcome forecasts.

Unfortunately, regardless of how carefully and deeply a lawyer observe a case, there is still a room for an outcome or a verdict to vary from his initial assessment. In some cases, a lawyer may have overlooked a crucial legal factor or another component that may have affected the assessment; in other cases, a lawyer may have performed his job perfectly, but the result turns out to be unexpected due to individual anomalies in the judicial decision making. The use of analytical models can be a solution to solve this human subjectivity issue.
6.2 AI Approach to Predicting Court Decisions

A crucial transformation is taking place in the legal industry. The development of AI techniques can be used to examine, forecast, and respond to normative judgments and subsequently improving the law's effectiveness and fairness. Many scientists believed that algorithms for machine learning can significantly increase the accuracy of a litigation assessment. In this section, we highlight the main scientific works devoted to this topic with a close look at the models and variables they use.

Coming back to the study of Martin, Quinn, Ruger, and Kim, it is worth to mention their statistical model which predicts the outcomes of US Supreme Court with accuracy 75% (2004). The classification decision tree algorithm processed data of 628 cases (training set) utilizing several variables, only six of which were included in the final model. Those variables were “circuit of origin, issue area of the case, type of petitioner (e.g., the United States, an employer, etc.), type of respondent, ideological direction (liberal or conservative) of the lower court ruling, and whether the petitioner argued that a law or practice is unconstitutional” (Martin, Quinn, Ruger & Kim, 2004). The decision trees were built for each judge (overall 11 trees) adjusting different variables and in different ways. Despite the model was relatively simple, it allowed fairly accurately forecast how the particular judge would decide a case.

A much more sophisticated model was developed in the research of Bommarito, Katz and Blackman (2017). The input data set alone consisted of 240,000 justice votes and 28,000 cases outcomes collected for over two hundred years from the Supreme Court Database (SCDB). Each input case contained almost 150 variables of different types including categorical values. That’s why the selection of the model’s features was a challenge for researches. Some of them were taken from the initial data set (term, natural court, petitioner, respondent, manner court of origin and source of the case, issue area, etc.); others were additionally engineered. Extra features included such factors the Circuit Court of Appeals where the dispute arose, factors related to oral argument and case timing, and factors aggregating the behaviour patterns of a Justice, the Court, and the lower court. The model’s aim was to predict whether the Court would affirm/reverse a case as well as individual justice vote. The random forest classification algorithm was used which outperform support vector machines (LibLinear, LibSVM) and neural network with multi-layer perceptron. Describing the model in a simplified format, it learnt from a training dataset of the earliest years and then was tested whether it predicted case outcomes correctly. After that, it was adjusted for the following year and tested again.
This process continued through the data set until the most recent cases. As the approach did not rely on simple decision trees, but random forests, it was capable to aggregate all trees in the forest and average them.

The final model’s accuracy was 70.2% for the case outcome assessment and 71.9% for the justice vote assessment. Comparing this rate with ‘rule of thumb’ prediction rule (the baseline strategy is to always guess Reverse), the former is only slightly higher than the latter. For recent years if one predicts a reversal, the likelihood of it is 63%. On this basis, the difference between the machine learning approach and the rule of thumb approach is not significant. The model’s performance was consistent across time, judges, and nature of the cases; however, it is important to remember that the input was based on retrospectively data set generated a long time ago. The algorithm it provides may deteriorate in the future as the judges may decide new types of cases or in ways that somehow avoid the variables used to date.

This leads to the third study of Aletras, Tsarapatsanis, Preoțiuc-Pietro, and Lampos as it is different in the chosen approach, concentrating not on fact-related variables from a case but on the text associated with it (2016). The research was aimed to analyse the decisions of the European Court of Human Rights (ECHR) and to predict whether the court would decide if the Convention was violated or not. Text of the cases was used as an input data including cases attributed to the 3rd Article — prohibits torture and inhuman and degrading treatment — 250 cases, the 6th Article — protects the right to a fair trial — 80 cases, and the 8th Article — provides a right to respect for one’s private and family life, his home and his correspondence — 254 cases (The Human Rights Act, 1998). Typical case structure of ECHR contains procedure, facts, circumstances of the case, relevant law, the law, and operative provision parts. To achieve the goal, scientist developed a model leveraging NLP and machine learning techniques. Text representation methods were the top-2000 most frequent N-grams (contiguous word sequences) and clustering of related words (30 groups, e.g. prison, detainee, visit, situation, etc.). After that, each case was represented in a vector format with its elements representing a feature (n-gram or topic) with a weight based on the frequency of appearing in the case. Linear Support Vector Machine (SVM) was used for a case classification with data split 90/10 (training and testing sets). The received results were 78%, 84% and 78% of accuracy for 3rd, 6th and 8th Articles respectively with 79% accuracy rate on average. This rate was higher than Bommarito et al.’s model in numerical terms, but Bommarito et al.’s study considered a broader period of time and types of cases. The advantage of Aletras et al.’s
approach was that the researches randomly selected an equal number of cases with violated and non-violated outcome; thus, the rule of thumb in their case was only 50% — notably less than the overall model accuracy. Besides, their study reveals that circumstances of the case play a more vital role and have higher predictive accuracy than its law-related subsection, which supports legal realism theory.

However, there is something of a pitfall in the design of the study. The model learns from existing deviations and correlations in the data set without determining a causal relationship. In addition, the textual analysis of “facts” and “results” cannot be stored separately. It would be more effective if the algorithm captured a change in the judicial tone instead of determining whether certain patterns of facts could lead to certain outcomes. As the researchers suggested it would be beneficial to test the analysis of facts from additional documents such as the appeal pleadings and submissions aside from the judgements themselves. Another issue is how to interpret the model; clusters of words used to forecast judicial results are not very informative about what is happening and why (Fig.10). The problem of the transparency of algorithms can lead to questioning their legitimacy or raising the problem of regulation for future implementation.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Label</th>
<th>Words</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top-5 Violation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Positive State Obligations</td>
<td>injury, protection, ordered, damage, civil, caused, failed, claim, course, compensation, pecuniary, ukraine</td>
<td>13.50</td>
</tr>
<tr>
<td>10</td>
<td>Detention conditions</td>
<td>prison, detainee, visit, well, regard, cct, access, food, situation, problem, remained, living, support, visited, establishment, standard, admissibility, merit, overcrowding, contact, good</td>
<td>11.70</td>
</tr>
<tr>
<td>3</td>
<td>Treatment by state officials</td>
<td>police, officer, treatment, police officer, July, ill, force, evidence, ill treatment, arrest, allegation, police action, subjected, arrested, brought, subsequently, allegedly, ten, treated, beaten</td>
<td>10.20</td>
</tr>
<tr>
<td><strong>Top-5 No Violation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Prior Violation of Article 2</td>
<td>june, statement, three, dated, car, area, jurisdiction, gendarmerie, perpetrator, scene, june applicant, killing, prepared, bullet, wall, weapon, kidnapping, dated june, equal dated, slugged</td>
<td>−12.40</td>
</tr>
<tr>
<td>19</td>
<td>Issues of Proof</td>
<td>witness, asked, told, incident, brother, heard, submission, arrived, identity, hand, killed, called, involved, started, entered, find, policeman, returned, father, explained</td>
<td>−15.20</td>
</tr>
<tr>
<td>13</td>
<td>Sentencing</td>
<td>sentence, year, life, circumstance, imprisonment, release, set, president, administration, sentenced, term, constitutional, federal, appealed, twenty, convicted, continued, regime, subject, responsible</td>
<td>−17.40</td>
</tr>
</tbody>
</table>

Figure 10. The most predictive topics for Article 3 decisions represented by the 20 most frequent words (Aletras, Tsarapatsanis, Preoțiuc-Pietro, & Lampos, 2016)
Discussed above three articles, mentioned as well in the literature review section of this thesis, have the greatest coverage in the legal journals and other scientific works. Now we will review slightly less well-known works which are not less interesting in term of models and variables they use and the problem they try to solve.

Several studies have been carried out the litigation outcome predictions in a narrower way focusing on specific court process. Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan investigated a bail hearing forecasting (2017). A bail hearing is a court process after an arrest where a judge decides whether to release the defendant by setting a bail. Therefore, the research’s purpose was not to predict the guilty/innocent outcome but the release/detain result as well as to forecast a risk rate associated with a defendant. The input data contained almost 1.5 million New York cases dated 2008 – 2013. The dataset was divided in 80/20 proportion for training and testing of the model. Bernoulli loss function and Gradient Boosted Tree algorithm were deployed to train the model. Similar to the random forest method, it aggregated the outcomes from individual trees where each new tree helped to correct errors made by previously trained tree by adjusting the tree’s weights. The model used only variables available at the time of the bail hearing such as current offence, prior criminal history and it did not utilise factors of defendant demographics (except age) such as race, ethnicity or gender. Evaluation of the model based on accuracy as for the previous studies does not suit this one as the input data contains only crime outcomes for released defendants. The model’s accuracy metric was equal to 70.7%, but for a correct prediction defendant’s crime propensity had to be identified. For that, Kleinberg et al. introduced a payoff function that considers a defendant’s criminal propensity and the release decision. The research results demonstrated the inverse relationship between risk distribution and release late, which was naturally logical. Nevertheless, “the riskiest 1% of defendants have a predicted risk of 62.6% yet are released at a 48.5% rate” (Kleinberg, Lakkaraju, Leskovec, Ludwig & Mullainathan, 2017). Thus, the predictive model can be implemented in practice to reduce the crime rate.

Another issue raised by the study is whether judges take into account the social attributes of a case. It only revealed that some key observable variables (e.g. prior criminal record) were underweighted. However, decision-making correlation with sensitive case attributes including race, ethnicity, gender, etc. can be observed in future studies. Fairness should be the default in machine learning helping to understand human
error. Though the data we have were somewhat limited in this regard, algorithms applied to richer data might produce novel behavioural insights.

There are some works which do not only observe more specific cases but also unobvious factors as variables. Guimera and Sales-Pardo suggested a model that forecasted a justice’s vote based on the other justices’ votes in the same case (2011). As a baseline model, they considered a perfectly rational and free of any bias court, where judges have access to complete information about the case. In this idealised environment, all judges must make the same decision on a case; the key characteristic of this court is that justices’ votes are uncorrelated and do not depend on each other even in a slight way. If the court is not so ideal, the individual justice’s vote may be affected by some bias or another justice’s opinion. These hidden patterns may be forecasted based on historical data of previous cases. A computational model was built on the data of 150 cases from the Supreme Court Database and was able to correctly predict 83% of the individual justices’ decisions. The study also revealed that the average court predictability was lower during the time of democratic presidents. It highlights that real-world judicial proceedings may have some bias. In addition, this study varies from others as it does not consider the facts or text of the cases at all indicating that for prediction of case outcomes or judge’s votes there is a room for using methods developed for analysis of other tasks e.g. social network analysis.

Indeed, one area of AI and machine learning to be used is to warn judges about possible bias when the circumstances of the case in question coincide with similar ones, but a different decision is made. In this case, the judge might be asked to spend a few extra days to consider the case more carefully. Dunn, Sagun, Sirin and Chen examined the forecast of the asylum court decisions (2017). They had a focus not on case predictability (case is closed, all case data is available) but on early predictability (case is open, primary data is available). The research’s objective was to predict asylum granted/denied outcome based on the case common features such as nationality, language, notice to appear (NTA), base city, hearing location, case type, attorney, and judge – mostly categorical variables. For the input data set, the data from the Executive Office for Immigration Review (EOIR) was used, finalised to 602,500 records with 35% of the granted asylum cases. The Random forests algorithm was utilised as it showed better accuracy than support vector machine (SVMs) and neural networks with data set split as 80/20. Interestingly, the model was first trained with the smallest number of features to check how adding of additional fact influenced the case outcome. This approach revealed
that considering only two variables — judge ID and nationality — the model was able to correctly predict 76% of the cases. The final model accuracy with all available features reached 80%. It highlighted that an assigned judge plays a crucial role in prediction but also casts doubt on whether the judge considers all case facts during the decision-making. Thus, using AI models in this field may help both sides of the justice; asylum seekers will be able to evaluate the possibility of winning before a court proceeding (but at the moment when judge is assigned) and judges to be sure that they consider the individual factors of a case rather than predetermined judgments to make the adjudication process fairer.

Discussed above legal prediction papers are diversified, they have a focus on various issues and use different methods, algorithms and variables. Nonetheless, previous works have been limited in a similar way to the prediction of binary outcome such as affirm/reverse or satisfied/denied. The researches from Great Lakes Institute of Management (Chennai, India) developed a model predicting 12 possible outcomes of US Court of Appeals as a final project of their MBA course (Mohan, Hosurkar & Mishra, 2016). The input data consisted of public the Judicial Research Initiative (JuRI) data set at the University of South Carolina with case records dated 1997-2002. The possible outcomes were the following: stay petition or motion granted; affirmed; reversed; reversed and remanded; vacated and remanded; affirmed in part and reversed in part; affirmed in part; reversed in part and remanded; vacated; petition denied or appeal dismissed; certification to another court; not ascertained; affirmed, vacated and remanded. After cleaning the data, chi-square and boruta analysis were performed to identify key predictor variables; among them were facts about case origin, participants and nature of the case. Judges-related variables were not chosen by the algorithm but still were included in the final model. The model had to suit the requirements of multi-variable output, thus, random forest, neural network and XGBoost were selected. Random forest was over-fitted with accuracy rates as high as 99% due to the small size of the initial data set. Eventually, trained with neural network and XGBoost the model reached an accuracy of 82% and 91% respectively. Moreover, the study defined that which one of the 13 Appeals Courts in the USA hears the case vitally influences the outcome and that natural-born-citizens have higher chances during the litigation. Although the research was limited by the input data set, it still shows promising results with high accuracy. Besides, it gives hope that in future we can expect more works forecasting not binary cases outcomes.

Determination of possible bias is a big advantage which should be implemented in practice to increase the fairness of law and to make a step forward to the legal
formalism. It is essential not to forget that AI may only be as objective as the scientists creating it and as the data set it uses as input data. That’s why it would be optimal to first use AI techniques to discover bias pattern in historical data, and only after that to train model on clean data carefully considering the features. Apart from it, AI algorithms can be used to assess the effects of the judges’ decision-making. Many legal studies analyse the potential consequences of a change in the law, analogically, case outcome forecasting may be utilized for subsequent analysis of the causal assessment of the effects of court decisions. Predictions will be used as a baseline for the following evaluation of the law's effectiveness and fairness except just suggesting a correct decision.
7. Limitations of AI Approach to Predicting Court Decisions

While AI promises to significantly improve the life of legal professionals in many ways, there are some major concerns about how beneficial it really is in a matter of using AI for the legal forecasting. We observed five main shortfalls regardless of the topic; going from more practical to more philosophical levels, they may be classified as challenges related to data, model transparency, AI objectivity, the nature of law, and AI legitimacy.

**AI & Data**

Determining outcomes of legal cases is demanding, time-consuming, and tedious. Even though there are many publicly available databases with case records, it is not an easy task to aggregate all this data in a systematic way. The first matter that arises is not data insufficiency but that the data is locked and difficult to access. Court records must be found, read and classified according to the appropriate branch of law. After that, the data should be cleaned and transformed — categorical variables into easy-to-analyse and built into the model binary ones. During this process, some useful valuable insights may be missed. Moreover, the problem of missing data might also occur — historical data is possible to omit some values that may have a significant influence on the future model. Secondly, a case itself must be defined. Vital case-related information is often not part of the court record. For instance, official court documents rarely indicate whether a monetary settlement has been reached, and, if so, the terms of this settlement. Instead, the court records simply state that the court affirmed or dismissed the case. Why the case was dismissed or under what conditions the case is terminated is rarely revealed (Harris, Peeples & Metzloff, 2008).

However, if one is determined to find data that will support his initial hypothesis, the most probably, she will find it; it is another data-related issue. Many psychological and perceptual experiments emphasize how our expectations or initial impressions influence people to seek, perceive, and find evidence to support their initial ideas (Awh, Belopolsky & Theeuwes, 2012). The forth downside is the opposite of the previous one; we do not have any missing data but a large and all-encompassing input data set. With the increasing volume of data, the number of correlations expands dramatically too. Some of them uncover previously hidden insights, but others are just useless. This leads to
selecting a limited number of features, and, therefore, to the model simplification. Eventually, we face up to a data dilemma, on the one hand, the data is limited to observe all possible correlation, on the other hand, there is so much data that we need to remove the complexity while accepting the risk of missing important dependencies. How to find the golden mean? It is the question we still do not have an answer for.

**AI & Model transparency**

The use of AI in the legal industry raises questions about the fairness of decisions made by machines. Most AI-based programs are locked in the “black boxes”. The lack of transparency might cause a lack of trust in the models. Neural networks and other machine learning algorithms are designed to work like a human brain, thus, due to their very nature, cannot be transparent. This process is hidden and constantly changing; the models create their own connections and correlations rather than being explicitly programmed. It causes the risk of limiting the ability of the judge to make a data-driven decision that may be fully explained and the defender's ability to logically argue it while protecting his clients. While AI methods allowing tracking which input data features lead to a specific outcome is being gradually developed, they still only define how certain variables are combined and not why they are combined.

It is clear that AI is necessary to be explainable with regard to healthcare or self-driving cars of the automotive industry. But the legal field is not an exception. It is a question of “white box”, which is total transparency, versus “black box”, which reveals nothing about how the algorithm works. These expectations are aimed to provide predictive routing with a good level of transparency. This transparency explains which case factor, and to what extent, plays a role in making optimum decisions either by an attorney or by a lawyer. In this case, transparency allows to improve the training of legal professionals or to add restrictions to the system in order to avoid possible bias. One of the studies revealed that one tends to apply greater examination to information when expectations are violated (Harvard Business Review, 2018). In the experiment conducted by René Kizilec, a Stanford PhD student, three levels of transparency were tested. The provided students with a low, medium and high explanation of how the grading algorithm works. While medium transparency increased trust significantly, high, on the contrary to expectations, minimalised it to the level of trust equal or even lower than low transparency.
Another transparency issue is that it makes algorithms vulnerable. If the full code is realised it is possible for defendants to adjust case facts description to add more keywords to increase the chances of a favourable outcome.

Nowadays, AI technology allows the automation of various tasks using the most complex algorithms. Some of them are so sophisticated that it is impossible to explain them in a simplified manner. Thus, a white box approach that reveals all the steps behind the model is not feasible, but a black box is also usually not acceptable. The best approach may be to understand the basic information about the factors that determine algorithmic solutions and how this contribution is analysed or to integrate multiple AI paradigms into a hybrid solution combining them with more traditional solution techniques (Chowdhury & Sadek, 2012).

**AI as an objective approach**

It is commonly believed that AI will bring in our lives data-driven decisions bringing anomalies to light while omitting bias and human-based errors. It may encourage impartial court proceedings highlighting only those variables that should be crucial for a specific case or the article of law.

However, this common belief about the objectivity of AI has not been dealt with in depth. Even though AI relies on mathematical calculations, it is still developed and controlled by humans. Thus, there is a chance that AI may contain some bias inherited from its creator. Another fundamental issue is that the AI model is only as fair as the input data sets. Since machine learning is usually trained on real data, it is natural that real-world deviations are reflected in the final models. For example, the researchers from ProPublica found that the risk assessment rates are more favourable to white individuals rather than black (ProPublica, 2016). This phenomenon is known as “machine or algorithmic bias” — when regardless of the model used, any mathematical predictions relying on pattern recognition will exhibit the same kind of biases present in the training data. In addition, AI bias is not limited to the matter of data, it also includes model, learner and system prejudice (Massachusetts Institute of Technology, 2017).

One of the most commonly eliminated sources of AI bias is the human interpretation of algorithm results. Data always demands experts to assess the gained values and insights; findings can be made solely on the basis of correlations while the importance of causation decreases. Because of this interdependence, it may be quite a hard task to determine why the AI system worked in an anomaly way that hides bias that the person,
who built it, might have added earlier. Then, the attempt to explain the results as well as
the possible limitations of the solution would be difficult too. Thus, human-in-the-loop
model has some imperfectness. As the articles from section 6.1 shows, judges may also
be partial, their bias combined with bias in case outcome predictions could lead to even
more unfair results.

**Static AI & Dynamic Law**

Well-grounded definition of law as a metaphor was proposed by Devins, Felin,
Kauffman, and Koppl (2017). They saw law as a dynamic field that is constantly
evolving. With time passes, legal experts and judges can gradually interpret law in
different ways according to the political and historical situations as it has a set of
affordances, or possible uses and interpretations. Case law is also eventually changing
with appearing of new precedents and new social context. Moreover, every legislation
and regulation is based on language, and language has a dual nature too. Although it can
be used to build logical systems, its essence is vague, contextual and constantly
developing.

The same rule applies for law, on the one hand, it is systematic and logical, on the
other, it still semantic, undefined and built on compromise. As for AI, it is a set of precise
algorithms that are empirical and deterministic. It is not always capable to mine new
meanings in laws which provide new patterns of action and, therefore, new outcomes in
a form of risk or reward, which, in turn, stimulates new legal adaptations. Furthermore,
the AI model may not be complete enough to follow the process of judges finding
inconsistencies between legal principles or reacting to unforeseen events. One example
of such changes is the law of surrogacy. When surrogate motherhood became possible,
the justice could not agree who “the mother” of the child was (Devins, Felin, Kauffman,
& Koppl, 2017). To distinguish the “genetic mother” from the “native mother”, the law
had to split the idea of motherhood which was an unexpected and previously unneeded
distinction.

NLP promises to solve the problem of language ambiguity and sentiments. Though
NLP models also limited to specific words or indexing not focusing on more general
concepts. They perfectly fit for the automation of legal search allowing scanning,
retrieving, and ranking documents based on the wording. However, for the tasks of
understanding the dual meaning of law, it is still needed to be enhanced.
AI & Its legitimacy

Discussing possible drawbacks of the AI realization in the legal field, we will shift now to its big picture and a more abstract level. It is crucial to keep in mind that decisions made by AI-based applications will raise legal issues such as issues of liability for civil or even criminal wrongs. The process of AI implementation has already been started. In New Jersey, bail hearings are replaced with algorithmically informed risk assessments (The New York Times, 2017). Anyone eligible for release can eliminate cash bail if she meets certain criteria. To ensure objective, scientific decisions, judges rely on machine evaluations and predictions. Automated recommendation serves as guidance and does not replace judicial discretion. However, the program raises questions about the declared neutrality of machine thinking and the wisdom of reliance on mechanical judgment. Most of the models do not reach 99% accuracy; as it was previously discussed (sections 5.1.2 & 6.2), the average likelihood of correct forecast is about 80%, so, nearly every fifth case may be a false one. When the analytical model fails in its prediction, it may be an issue to allocate liability under current legal schemes (Barfield, 2018). Moreover, since these algorithms are proprietary, they are not subject to open government state or federal laws. This leads to defendants being unable to challenge results accuracy and infringing their right to due process.

Conclusion

We have discussed the main reasons why AI may be still a challenge for the judicial system. AI simply cannot consider all case and law circumstances and nuances, this factor violates its implementation in the process of justice. It is not yet possible to outsource human values and ethics to empirical and logical models, but AI methods should give the judges a set of guidelines to work with. It will create a more secure legal system where the AI keeps the judges informed and the judges consider based-on-data forecasts but do not unconditionally rely on them. This way two sides are going to help one another.
8. Analysis and results

8.1 Summary of AI Legal Software Analysis Results

The conducted in the 5th chapter research highlighted that legal software landscape is diverse and capable to support legal professionals in almost every daily activity. The design of the AI software classification was geared towards the frameworks of Rackwitz and Corveleyn, and Praduroux, Paiva, and Caro. The first framework helped us to see a variety of innovations in the legal field. From general to specific, we investigated the Software quadrant of the Legal Innovation Matrix (section 4.1.1., fig. 7) and discussed it through categories of Pradouroux et al.’s approach. However, this structure is proposed to be applied for every legal software, therefore, we made some modifications to introduce a classification for legal software leveraging AI technology. It included the following categories: document automation, legal analytics, legal prediction, eDiscovery, legal research, legal due diligence, contract review, legal expert systems, eBilling, and intellectual property. The main benefit of using them for law firms is the ability to reduce the time of routine tasks and to trawl through several static databases for needed legal data (e.g. precedents, law, patterns). For all solutions, machine learning is the fundamental AI technique; many also leverage NLP, and only some eDiscovery applications utilize more sophisticated methods including face and object detection.

As soon as these steps were carried out, we examined the group of legal analytics and prediction technology solutions to understand the details of their functionality and to see how the concept of AI for forecasting court decisions works in the real world in form of business applications. It appeared that the market is still not mature; the predictive software is only entering the industry with the majority of solutions based in the United States.

Solutions core functions are represented by data analytics (behaviour patterns for judges, attorneys, cases) and forecasting (typically case outcome predictions). For both capabilities, the main AI technique is machine learning. Data analytics software utilises supervised machine learning for data labelling and pattern recognition and unsupervised for clustering and searching for previously unknown correlations. Forecasting software mostly employs text mining for unstructured input pre-processing (tokenizing, stemming, parsing, etc.) and then processes it with data mining (pattern recognition, statistical models). Even though official solutions’ websites often refer to these steps as to “NLP”,

52
it does not stand for it. Only a few solutions leverage NLP technologies (vector representation, sentimental analysis). As for the data mining model training for case prediction, a decision tree algorithm is the most commonly used algorithm.

According to introduced in the theoretical background part AI Knowledge Map, legal analytics and prediction solutions fall into the statistical paradigm, primarily using mathematic methods. For domains, they cover perception (raw data transformation), reasoning (problem-solving) and knowledge (input representation and understanding). Contrary to expectations, we did not find a significant deployment of complex and deep AI. A recent Resulticks’s study found that nowadays AI is the most over-hyped term which is used for marketing and promotion (Forbes, 2018); it could be the reason why companies try to use the power of naming without a full understanding of the technology behind it. Although the implemented AI techniques are not outstanding, we still believe that with the evolution of the legal tech industry they will evolve too.
8.2 Comparison of Expert and AI approaches to Court Decision Making

In the 6th chapter, we discussed in detail the expert and AI approaches to court forecasting and how this topic was examined in the literature. In this section, we will compare them in order to answer the main research question and test the initial hypothesis.

One of the best examples of the competition between humans and AI is the FantasySCOTUS (FantasySCOTUS, 2019). The online platform, which was found by Josh Blackman and now is a part of LexPredict, gathers users to forecast the outcomes of the US Supreme Court. It is possible to predict either the binary outcome (whether the court affirms or reverses the lower court) or the split (9-0, 8-1, 7-2, 6-3, 5-4, 4-1-4, or fragmented). From a legal perspective, FantasyScotus helps to see which perceptions legal experts have with respect to the Supreme Court and how they identify judges’ legal and ideological inclinations. It is also essential for judges to keep their ears on „the wisdom of the crowd” representing liberal voting and the rule of law and objective, independent standards of judging. From a practical perspective, it makes possible to compare the aggregated vision of 25,000 attorneys, students and just people interested in law with AI. The average human accuracy is nearly 70% — almost the same as the rate of Blackman et al.’s model (Katz, Bommarito & Blackman, 2017). Taking into account that not all members of the platform are equally wise, it is crucial to notice that top 10 experts have been right about case outcomes on average in 77% of the time, which is higher than algorithm’s indicator. In any case, FantasySCOTUS could help researches to reveal which cases are more predictable by humans and which are better understood by AI. Combining this data, it is possible to create a forecasting algorithm using the best of both approaches.

Indeed, neither of approaches in their pure representations could claim to be a perfect prediction solution. Human decision-making and forecasting appeared with the beginning of law and legal system. With the development of statistical methods, scientists start to use quantitative methods for legal prediction (QLP-based technologies) to supply the shortcomings of human reasoning. Nowadays, AI technologies are trying to compete for the leading position in the field. First of all, humans are limited in the volume of their observations. The more experience an expert has, the more data he observes; while an experienced lawyer may know hundreds, if not thousands of previous cases, he has hardly seen thousands, hundreds of thousands, or millions of them. The same is true for the speed of documents analysis, humans lost this battle to the algorithms. Thus, giving a response
to the “What would be the outcome of the case?” question, experts think about the probability of it from personal observations which is particularly challenging for rare cases. AI, on the contrary, can look through millions of records in the seconds to search for similar examples.

Comparing the basic concepts behind both sides, we examined that legal professionals depend on both rational and intuitive predictions while AI has a focus only on algorithms developed over the initial data set. The term “data” refers to structured (legal databases) and unstructured legal data (text, audio and video files, etc.) that is relevant to use for the case outcome forecasting. Analogical reasoning underlies how lawyers make decisions; at law schools, they often learn from casebooks, therefore, their future reasoning can be presented as the following process described by Katz — “people who cite Case X also cite Case Y; lawyers who argue Principle X also typically argue Principle Y” to build parallels to the case (2012). We can find a similarity in the AI method; its models are developed from correlations and hidden patterns in the historical data. The most common models are utilized for training is random forest, decision tree, naive Bayes. As algorithms rely on previous case data, it basically uses its factors as variables, however, sometimes AI may reveal previously unknown correlations. Experts also mainly rely on the case description but pay attention to other elements as well (Martin, Quinn, Ruger & Kim, 2004).

The discussion about which approach is better may go on endlessly. Nevertheless, researches always look for numerical support of their argument, that’s why we cannot proceed without weighing up the accuracy rates of both sides. In broad terms, the average aggregated rate is higher for AI³. However, we must be very careful with accepting the statement that machines are better than experts in predicting the case outcomes. If we look separately at the forecasts of appeal attorneys in Ruger’s study, their prediction accuracy was more than 90% (Martin, Quinn, Ruger & Kim, 2004). The best FantasySCOTUS’s player correctly predicts cases more than 80 per cent of the time (FiveThirtyEight, 2014). This benchmark needs to be interpreted with caution, but we certainly need to keep the rates in mind as best human predictors overtake AI.

As for the characteristics of humans versus machines, the former is believed to be more subjective while the latter is more objective. Our previous findings revealed that AI objectivity is not fully reliable, but it still may be beneficial and helpful in avoiding

---

³ Based on the aggregated results from the papers mentioned in the 6th chapter
individual’s errors and bias. Sometimes subjectivity plays an essential role in case some empathy is needed. All conscious emotions are usually complex and difficult to learn by machine. The same applies to law in general, legal systems are complex adaptive systems with sophisticated levels of interconnections and feedback loops between their institutions and agents, laws and regulations. The exact level of complexity varies in different subfields, but this sometimes unstructured and dynamic environment is hard to be fully translated into a static model. Models have not understood yet the meaning behind the language, and they have not yet been able to constantly and automatically monitor changes in data and litigations. They are also restricted by determining the binary outcome while experts may predict case outcome details such as the amount of claim, terms of imprisonment and others.

Despite all the work describing the problems of prediction and uncertainty, there has been a limited number of case outcome predictive solutions and articles. This topic is becoming a more popular area of interest for scientific works and the next decade is likely to see more complex and interesting AI models or even new fields of their applications. Contrary to expectations, we did not find an AI superiority over the expert approach to predicting court decisions. As for now, AI is a perfect tool to assist legal professionals which cannot exist separately from them.

The summary of the described above results is presented in Table 2.
Table 2.
Comparative analysis of expert and AI approaches to court decision making

<table>
<thead>
<tr>
<th></th>
<th><strong>Expert approach</strong></th>
<th><strong>AI approach</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basis</strong></td>
<td>• Data</td>
<td>• Data</td>
</tr>
<tr>
<td></td>
<td>• Experience</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Intuition</td>
<td></td>
</tr>
<tr>
<td><strong>Model</strong></td>
<td>• Analogical models</td>
<td>• AI methods: text and data mining, NLP’s models</td>
</tr>
<tr>
<td><strong>Volume</strong></td>
<td>• Limited by experience and human ability</td>
<td>• Only limited by data sources; millions of cases</td>
</tr>
<tr>
<td><strong>Speed</strong></td>
<td>• Experts are 200 times slower than AI&lt;sup&gt;4&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td><strong>Input Data</strong></td>
<td>• From the case: precedent, statutory text, etc.</td>
<td>• From the case: petitioner, respondent, judge, case area, etc.</td>
</tr>
<tr>
<td></td>
<td>• Case-related: oral arguments, decision consequences, judge’s policy preferences</td>
<td>• Case-related: oral arguments, judges’ behaviour, other hidden patterns</td>
</tr>
<tr>
<td></td>
<td>• Additional: other experts’ opinions</td>
<td></td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>• 60-70%</td>
<td>• 70-80%</td>
</tr>
<tr>
<td><strong>Characteristics</strong></td>
<td>• Subjectivity</td>
<td>• Objectivity</td>
</tr>
<tr>
<td></td>
<td>• Unstructured Environment</td>
<td>• Structured Environment</td>
</tr>
<tr>
<td></td>
<td>• Diversified outcomes</td>
<td>• Primarily binary outcome</td>
</tr>
<tr>
<td><strong>Level of prevalence</strong></td>
<td>• Throughout the legal industry</td>
<td>• Additional functionality of legal AI software</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scientific works</td>
</tr>
</tbody>
</table>

<sup>4</sup> LawGeek’s Experiment (LawGeek, 2018)
9. Conclusion

The first discussions about using computer technologies for case outcome forecasting go back into the 1960s when a pioneer researcher in the field, Lawlor, identified the status of studies in the prediction of judicial decisions and applied his prediction method to the United States Supreme Court (1963). Today, we still debate how technologies, and AI particularly, can be used in the legal industry and for court decision forecasting. Researches have learned how to use the increased volume of observational data to build a model that may assist people in their daily routine, or even more sophisticated tasks. These tools allow cases identification, patterns extraction, estimation of winning litigation. They help to understand judicial decision-making and to identify areas for its improvement. Despite it, AI legal solutions are still only assisting tools that cannot exist sufficiently without human interaction.

In this paper, we have described numerous studies conducted in the field. Our research evaluated the landscape of legal AI software while focusing on legal analytics and prediction solutions. There is a consuming interest from investors in the AI software, but it is still not clear what exactly the legal tech’s company mean by this abbreviation. Some of them just use it as a marketing tool, another utilizes statistical models and call them “AI”, and only the rest is eager to develop sophisticated neural networks or similar models. Even among progressive and innovative start-ups, there is no single approach to court decision forecasting. They all use different initial data sets, algorithms, variables and features. The same is true for academic researches in the field. As for now, there is still room for further works. In the thesis, we described examples of prediction models for a bail hearing, bankruptcy litigation, the Supreme Court, the Court of Appeals and others. The judicial system is a very complex structure where different kinds of cases require different models. We hope that the future will bring the standard or framework which will be considered as a benchmark for the litigation outcome forecasting that would be scalable and accurate.

Besides, the thesis underlined the comparative analysis of the artificial intelligence approach to predicting court decisions and the expert approach. The initial hypothesis that AI approach is more beneficial than the expert one proved to be wrong. AI has some advantages including limited objectivity, greater speed and ability to process a large volume of data. However, its limitations related to data, transparency, legitimacy and statical nature do not allow it to surpass humans.
We are aware that our research may have two limitations. The first is that we looked for information about AI legal software available in English. Hence, there was a possibility that there might be other outstanding legal applications that were not commonly known or they were just available within their countries. The second is that we did not conduct any experiments by ourselves and did not develop an AI algorithm to check all our observations in practice. Therefore, for future research, it would be essential to utilize results from this study to develop a model which will use complex AI approaches including NLP and deep learning to predict a diverse (not binary) outcome. And, before that, the input data set should be analysed to identify possible historical bias. Only after cleaning the data we will be able to build a trustworthy model that can assist the legal professional in such a difficult matter as predictions.
10. References


FantasySCOTUS [official website]. Retrieved April 5, 2019, from https://fantasyscotus.lexpredict.com/


Harris, C.T., Peeples, R. & Metzloff, T. B. (2008) Does Being a Repeat Player Make a Difference? The Impact of Attorney Experience and Case-Picking on the Outcome of Medical Malpractice Lawsuits. Yale Journal of Health Policy, Law, and Ethics, 8, 2 (1)


Mohan, K., Hosurkar, N. & Mishra P. (2016). Indian Institute of Science (IISc), Bangalore India [pdf]


