

Big Data in the Legal Industry

Wang Peiyu

G21469074

CSCI 6444: Introduction to Big Data and Analytics

Semester: 2019 Summer

1. Introduction

Big Data is at its exaltation for it is applied to almost every aspect and field in our everyday life from self-driving cars to facial recognition applications. It has significantly eased and even automated people's life since its appearance. However, there are several fields that are slow to change. One of them is the legal industry. Legal professionals often argue that the law is a form of art. Therefore, they are suspicious about the possibility of integrating Big Data into legal field to get back accurate results that can actually help them. Some legal professionals are worried that, with the development of Big Data techniques, they will be replaced by Artificial Intelligence. In fear of being replaced, some of them are resisting the idea of even integrating Big Data into the legal industry. However, no matter legal professionals want it or not, the integration of Big Data Analytics techniques in the legal industry is inevitable over time. As someone who has been educated in the science and engineering field and is also determined to dive into the legal world later, I am very interested and enthusiastic in exploring the possibility of cross-field integration of Big Data and Law. I have read and reviewed multiple papers and books regarding existing applications of Big Data and Artificial Intelligence and potential future applications in the legal field and summarized them in this paper.

The remainder of this paper is organized as follows: Section 2 introduces the background of this paper, why is it an interesting topic to conduct research on and why is it important to Big Data. Section 3 talks about the current applications of Big Data and potential future applications in the legal world. Finally, Section 4 summarizes the research done in this paper and reach a conclusion while pointing out some directions for future work and exploration of Big Data in the legal world.

2. Background

Big Data has been applied to many fields and significantly eased and even automated people's life thus far. However, legal field is the one of the few fields left where you haven't seen a lot of applications of Big Data techniques. People have different opinions about the cause of this phenomenon. Legal professionals are either doubtful that Big Data can actually be of help or fearful that they will be replaced by Big

Data automation and Artificial Intelligence. Therefore, a lot of legal professionals resist the idea of even deploying Big Data in the legal field. As one of the few gold mines left that have not been heavily explored and integrated with Big Data, I think it is both fascinating and interesting to explore the possibility of combining those two fields together. It is also alluring to dive in and look for places where people can make breakthroughs happen. The fact that the legal field has not been heavily touched by Big Data, if at all, means that there is still plenty of room for cross-fields innovations to happen. However, the integration of Big Data into the legal field is inevitable. In this case, the question that should be asked is not “whether it can or will happen” but “how and where it will happen”. It is very unlikely for Big Data to completely replace legal professionals unless there is a complete breakthrough in Natural Language Process that can make machines understand the meaning of words and the logic behind all the legal documents. Besides the fact that Artificial Intelligence might not be able to overcome the interpretation limitation, it is also not ideal to replace legal professionals with Artificial Intelligent. In the legal world, it is not only about following the established legal rules but also the constant human re-evaluation and challenging of whether specific legal rules are still applicable to the contemporary situation. The re-evaluation and challenging of legal rules with social background being taken into consideration is essentially what is pushing the legal world to adapt and evolve over time.

In conclusion, the integration of Big Data can be very helpful when it is applied to the right place, and there can be many ways in which Big Data techniques can be applied to in the legal industry. Even though Big Data is not able to replace legal professionals at this moment, the applications of Big Data techniques can significantly lift the burden off attorney’s shoulders by automating or refining the repeated and tedious part of legal work. The applications of Big Data can also help cut down the lag in legal work, shorten court waiting time, and accelerate the processing of legal cases for all legal professionals.

3. Introduction to Big Data and the Application of Big Data in Legal Field

Regardless of the fact that Big Data and its potential applications may have just attracted the majority of public’s attention, data has been long-standing. With the

invention of computer and internet, everything is going online and paperless. People are more interconnected than before by sharing everything with each other through internet which creates the well-suited environment for Big Data techniques to work. In this era, getting enough data is not the concern anymore. The concern is that there are so many data that finding and extracting the meaningful and useful data is time-consuming, if not impossible. That is why Big Data techniques and tools are so important. Big Data, by definition, is “datasets whose size is beyond the ability of typical database software tools to capture, store, manager and analyze” [1] according to the McKinsey Global Institute. From the definition, we can see that the core of Big Data is the ability to clean up the large number of data that might be in different structures originally so that they can later be analyzed or integrated together with computational tools. Useful information can later be extracted systematically. Last but not least, analysis can be done on the extracted data or information and the results from the analysis will be used to guide decision making process. In conclusion, the power of Big Data is the systematic cleaning and extraction ability of large number of data regardless of what original structures they have. In this case, data can be significantly reorganized in form and shrunk in size .

There have been some epoch-making innovations, which would have not been made into reality without the help of Big Data Analytics techniques. Self-driving car is a very good example. Without the help of Big Data, computers would not have had the ability to script, clean, reorganize, analyze and prepare the data needed for Machine Learning. If that were the case, the Artificial Intelligence applications in the self-driving car would not have the ability to learn how to react under certain circumstances. Another good example is the financial industry. For a long time, financial industry has been an lucrative industry that is heavily depending on the empirical intuition and expertise of the professionals to decide what to do in a certain case. However, with the help of Big Data, financial professionals now have the means to look into historic data and build predictive models that could be trained and used for prediction purposes. The predictive models can significantly decrease the risks associated with human empirical intuition, increase the accuracy and correctness of the decision professionals make in theory, and lowers the bar of having a successful career in financial industry for many young professionals. Many young professionals in the financial industry can now have successful careers even if they

do not have many years of industry experience because their success does not need to depend solely on the number of years of industry experience they have anymore.

There have been many other applications of Big Data in a lot of other industries as well, like healthcare, supply chain and et cetera. Big Data is changing and even revolutionizing the way we work and live for better.

In the remainder sections of this section of the paper, the focus will be on existing applications of Big Data in the legal industry and potential future applications in the legal industry respectively.

3.1 Current Big Data Applications in Legal Field

Even though Big Data has been applied to many industries and many innovations have thus been created, there are still some fields, in which you could rarely see the usage of Big Data tools or techniques, among them there lies the legal field. According to Bernard Marr from Forbes Magazine,

“the legal profession is traditionally one of the most conservative fields of professional or academic activity... The first big data tool to be made available to lawyers generally focused on billing, time management, marketing and customer relations functions, and not until recently, there has been little innovation in the way that legal profession uses Big Data... The current data-driven legal research is led by two entities- LexisNexis and Westlaw, which contain huge amount of cases and details but mainly function as search engines and offer little advanced analytics tools or insights.” [2]

In most cases, legal professionals still conduct their legal research or work in filing cabinets searching through mountains of case files even when all other professionals are going paperless with all kinds of databases. As for the cause of this deficiency, the co-founder of Ravel Law, who provides services designed to help legal professionals draw insights and connections using advanced analytics, thinks that it is because legal professionals all came from backgrounds with English Literature or Philosophy as their majors in the past. However, the new generations of attorneys who have entered the legal world in the past 5 or 10 years came from more diverse backgrounds. Some of them have degrees in Finance, some of them even have PhD degrees in biochemistry or other science and engineering fields. Due to their familiarity with quantitative analysis and data, they have high expectations about what technology should do for them [2]. He also thinks

that the new generations of attorneys are driving the evolution of how legal research should be conducted and maybe the revolution of the legal world as a whole.

Some of the revolutions in the legal world are already underway. Ravel law has taken on a Big Data Project along with Harvard Law School aiming to digitalize the faculty's entire US case library and make it available to anyone to access online for free. In order to achieve this goal, the legal documents will be scanned and then converted to computer text with optical character recognition technology [2]. From there, more analysis could be done through Text Analytics or Natural Language Processing and other Big Data Analytics techniques.

Another private analytics company, Lex Machina, aims to predict the cost and outcome of intellectual property litigations to guide decisions made by attorneys and their clients. "In the public sector, big data techniques have also been used in less than 10% jurisdictions to make bail decisions regarding the objective and scientific measure of risk a bail can lead to" [3] according to an Arnold Foundation report. One of the usages of Big Data in the legal field happened in Virginia jurisdictions when Virginia jurisdictions linked parole decisions to statistical data concerning rates of reoffending for people in different categories [4].

There are some other Big Data applications in the legal field as well. One of the most important rules in the legal world is that "decisions by judges and police comply with legal rules, and the norms of relevance are that decisions be based on legally relevant factors where that is required, that decisions do not discriminate on non-permissible grounds such as race, and that people are treated as innocent until proven guilty" [13]. Thus, the ability to prove or counterprove the guiltiness of opponents is the key to winning cases. Legal expert systems is a decision-support tool explicitly based on "Traditional" Reasoning with a computer program to model how a lawyer might answer certain questions to mimic legal expertise [13] is a perfect example of how Big Data can be integrated into the legal field. The quick and accurate sharing of information can also be crucial in legal industry when it comes to law enforcement, administrative law and criminal justice. Thus, there is also Police Information Systems which serves to enable better informed law enforcement decision-making by providing police with accurate and complete information [13]. By providing police with broad range of data from both

internal and external sources in different forms in a short time, Police Information Systems broadens the range of information available to the police and shortens the time needed for gathering the data. However, research suggested that this kind of Police Information Systems had been constrained by the traditional structure of policing and by the traditional role of the officer and has had limited impact on police practices [13].

Sentencing Database is another Big Data application in the legal industry. It is a data-oriented approach to judicial decision-making which is aiming to provide jurors and judges with historical sentencing patterns and other information to reduce the inconsistency and disparity in sentencing outcomes. Unlike the other two applications mentioned previously, Sentencing Database was more of a success. According to The Australian Law Reform Commission, “the establishment of a national database had overwhelming support from government and non-government organisations, Commonwealth prosecuting authorities, judicial officers, legal practitioners, federal offenders and academics” [13]. L. Moses et al. think that Big Data is different because it “speaks the language of probability, enhancing decision-making by estimating the likelihood that particular facts are or will be true” [13]. they also think the use of larger datasets makes it possible to detect correlations and patterns that might otherwise be missed, which could be used to “justify conclusions with reference to a sufficiently large dataset that would not have been statistically justified if based on smaller samples” [13]. The evaluations of applying Big Data Analytics to legal and policing decisions are based on effectiveness, acceptability and appropriateness (in legality, accountability and transparency) and the conclusion reached was that Big Data techniques can be deployed in the legal world to effectively identify correlations and enhance decision-making.

Those are some of the major applications of Big Data in the legal field that have already been made into reality. However, the number is still relatively small since the legal field has not been heavily touched by Big Data, which proves that there is still plenty of space left for innovations to happen.

3.2 Potential applications of Big Data techniques in legal field

3.2.1 Information Retrieval Systems in Legal Research

As mentioned in the previous section of this paper, the data-driven legal research is currently mainly led by two entities, one is LexisNexis and the other is Westlaw. Those two giants do have large databases containing huge number of case files and case details but they function more similar to search engines rather than Legal Analytics tools. Even as search engines, they sometimes do not function very well. When searching case files related to a specific judge, Westlaw gives back much fewer cases than LexisNexis by comparison, which is an indication of an incorrect or defective algorithm for information retrieval. In this case, applying better developed and more efficient algorithms to Legal Information Retrieval Systems can be very helpful when legal professionals want to transform into conducting data-driven legal research with databases from manually searching through paper-based legal case files.

“Many facts and insights that could help win an argument remain hidden because the information retrieval systems in the legal domain is not well suited... It is extremely difficult to search for a relevant case law by using Boolean queries or the references contained in the judgement, while a diverse result is much more intuitive and helpful to an attorney than a set of homogeneous results to win an argument” [7]. Legal Document Ranking with Diversification was proposed for queries to select a set of relevant and representative documents in a way that the diversity of the set is maximized. Document similarity and document distance (distance was used informally to refer to a dissimilarity measure derived from the characteristics describing the objects) were calculated in order to achieve this goal. Document similarity was queried first, then a set of documents based on their relevance scores was retrieved and re-ranked so that the top ranked documents were more diversified to cover more query subtopics. Because finding the optimum set of diversified documents is NP-hard, several algorithms were proposed to achieve this goal. The algorithms proposed include MMR, Max-sum, Max-min, Mono-objective, LexRank, Biased LexRank, DivRank and Grasshopper. By experimenting the proposed algorithms on legal documents and comparing the test results, the conclusion the authors reached is that, web search diversification techniques perform better than other approaches in legal diversification.

3.2.1 Big Data Applications in Law Firms

The legal field may be hesitant to changes, but law firms already started to look for ways to leverage their Data Analytics department and do it quickly enough to keep up with the business community [14]. Even though the legal industry is still fairly new to Data Analytics, Big Data, and Artificial Intelligence, law firms started to try deploying Big Data techniques through e-discovery by “developing and honing analytics” and predictive models and methods to help with e-discovery. Many law firms also started working on “client-facing Big Data projects centering on litigation risk assessment to predict the risks and results ahead of time before they materialize into very costly and time-consuming litigation” [14]. These projects are highly profitable because clients are willing to pay a lot for the ability to assess and predict the outcomes of cases ahead of time. They have also started using Big Data tools to look at the data that they have and figure out how and where to assign human resources more wisely rather than simply hiring more people when demand increases. Besides deploying Big Data techniques in e-discovery and human resource perspectives, “law firms are also trying making use of Big Data to leverage client data and serve it back at clients” [14]. Attorneys with Data Science backgrounds can advise clients on issues in their business analytics models regarding algorithmic bias and potential ethical concerns, i.e., potential litigation or merger and acquisition outcomes. Domain-specific Machine Learning and predictive models can then be developed to help navigate business decisions in legal context [14], which can only be done by Data Scientists who are also formally educated in law. Law firms are also teaming up with law schools and Legal Analytics labs on projects analyzing labor and employment cases for predictive analysis so that they can use the models built by Legal Analytics labs for assessment of what factors influence outcomes even when they don’t have the ability to do it in a law firm.

Besides all the applications within law firms above, lawyers are also leveraging data analytics tools mined from millions of court documents to predict how judges might rule. After carefully scrutinizing the data given back by the tools, attorneys could advise their clients on what to do next [8], according to S. Randazzo. Lawyers say the data can “help temper client expectations, influence courtroom decision-making and even save money by flagging strategies unlikely to succeed” [8]. E. Walters says that lawyers tend to answer the questions that are most important to their clients based on their professional

judgement, which may well be from experience [10]. But experience differs in person, which means that new graduates may have very little related industry experience or different views which could significantly affect their ability to give an accurate prediction for those questions back to their clients. Even for highly experienced attorneys, the answers to the same question could vary from person to person. However, these questions can be easily answered by scrutinizing historic data. While many other fields or industries are significantly more revolutionized with Big Data techniques, it might not be enough not to answer clients' questions with no number or evidence supporting their claims anymore. The everyday life of legal professionals' clients is now filled with Big Data and numbers. According to E. Walters, "the considered judgement of an experienced lawyer is unquestionably valuable. However, on balance, clients would rather have the considered judgement of an experienced lawyer informed by the most relevant information required to answer their question" [10]. Experience is important, but with the help of Big Data, even newly graduated attorneys can pursue and achieve the kind of success that cannot be achieved or even imagined ten years ago. E. Walters also proposed to solve the top three headaches that keep law firm managers awake at night, which are "1) acquiring new clients and new business, 2) clients demanding more services for less money, 3) wasting too much time on administrative tasks" [10] with data-driven solutions. For example, data-driven legal marketing, which deploys different systems like CRM to keep track of potential clients and lead sources they came from can be utilized to record and trace client data. From there, law firms can analyze client data and figure out where the majority of their clients came from so that they will be able to efficiently distribute marketing resources to the right places. They can also identify which clients, practice areas and lawyers bring in the most profits with the help of Big Data Analytics tools linking revenues and expenses for different clients, practice areas and lawyers. Some clients might bring in a lot in revenue to the law firm but after the expenses are deducted, the net profit is actually not great, which, without the help of Big Data, might be hard to find out. Law firms are also under the pressure to deliver results and control prices. Clients of law firms tend to bring the legal work in-house with the help of big data due to the uncertainty and risks if some of the legal work can be highly predictable and commoditized. E. Walters proposed a "fixed-fee legal work" solution for this line of legal

work to “shift the uncertainty from clients to law firms” [10] so that law firms are more likely to retain their clients’ business and stay competitive. Law firm can collect and standardize the information about time usage, expenses and billables for each type of legal work and decide the fixed-fee respectively for different lines of legal work. With the help of Big Data, law firms can even have access to information like seasonality billing data for certain practice groups and distribution of flat fee versus hourly billing for different practices [10], to help them stay competitive in the market. Law firms can also decrease the amount of time wasted on administrative work by automating administrative work with Big Data tools and use the time on bring in more new clients or maintaining their relationships with old clients instead. Litigation can be expensive and risky, but the risks can be quantified with Data Analytics. Lex Machina, Ravel Law and Bloomberg Law are offering granular information about judges which could be used by attorneys to decide which jurisdiction to file the lawsuits in order to decrease the waiting time and increase the probability of winning the case. With the granular information about judges, attorneys could be able to find out which jurisdictions are known for having shorter waiting time or favoring their clients. Ravel Law is even using Machine Learning to find the patterns in judges like the specific kind of language judges like to use. In addition, some attorneys would use the Ravel Law to find the specific analogies that a certain judge doesn’t like to hear and try avoiding that when in court. A lot of these Big Data tools or software are already invented by other companies, and key is for law firms is to think differently about their workflows.

It is also suggested in the paper written by C. Ma et al., that attorneys could use Patent Model Tree based on Neural Network to calculate patent similarity which could help them deal with patent infringement detection, patent search, enterprise competition analysis and patent layout [5]. They proposed using patent model tree to translate the patent text into a structured tree so that “the limitation posed by different authors’ word choice and writing style could be eliminated” [5]. Different similarity calculation formulas were proposed for structured data and for unstructured data respectively. In their research, patent claims were used as the textual experimental data because patent claims are “clear, complete and detailed descriptions which not only reflects the main technical content of the patent as a whole but also writes in accordance with a certain

format and structure requirement” [5]. The advantages and disadvantages of different Neural Network algorithms were compared and Siamese LSTM was chosen as the similarity calculation algorithm because it has obvious advantages and is by far the best method for calculating similarity of texts [5]. C. Ma et al. experimented with the 2016 patent documents from State Intellectual Property of the People’s Republic of China and had a group of students compare the patents manually. They reached the conclusion that the experimental results calculated with the proposed method were highly consistent with the manual comparison. In this case, applying Siamese LSTM algorithms to patent claims can significantly decrease the workloads of patent attorneys when looking for arguments for patent infringement.

Applied to the same practice area, but more specifically to Pharmaceutical Patent Validity cases, the paper written by V. Raghupathi et al. presented their research conducted within pharmaceutical patents utilizing Big Data and Distributed Computing platforms Hadoop MapReduce and applying a series of Machine Learning algorithms including clustering, word count and word co-occurrence to find out the top reasons why patents are ruled invalid by courts [16]. They aimed to help litigator decide their litigation strategies when representing clients, no matter if their clients are seeking help with filing the claims or counterclaims regarding invalidating an existing patent. V. Raghupathi et al. mentioned in the paper that, “computational intelligence applications in legal decision making... enables lawyers to devote less time gathering data and more time applying the law” [16]. V. Raghupathi et al. used clustering and text analytics techniques to help identify the keywords in the patent validity cases by customizing the stopwords corpus from NLTK library and applying TF-IDF for re-indexing the importance of the words presented in the cases. Their result is shown as Fig.1. They were able to elicit “obviousness” as one of the important reasons along with written descriptions, inequitable conduct and et cetera. Their results showed that their models are more accurate in predicting the outcomes of cases which are invalid due to the fact that courts are more likely to be more detail-oriented and list all the reasons why a patent is invalid than when the patents are decided valid [16], which could be overcome with a larger set of case data and details where the patents are decided valid. Even though this research is

only limited to pharmaceutical patent validity cases, the same idea and techniques could be easily replicated and applied to other practicing areas in the legal domain.

Contract analytics can also be a big area for Big Data applications, according to Kingsley Martin's article "Deconstructing Contracts: Contract Analytics and Contract Standards". Deconstructing is essentially "seeing through the cloud of complexity and identifying the core and determining the way those pieces are assembled together" [11], which has been applied to all fields. When more standardized processes are introduced into the legal world and integrated through automation, contract technology can "orchestrate all phases of the contract life cycle". The current contract analytics is a set of technologies used for parsing contracts and looking for and identifying patterns across all agreements through hierarchical deconstruction. The procedure is shown as Figure 2. The contracts or agreements are first disassembled into component clauses, then each component clause is disassembled into sentences and the contract analytics software looks into the language of each provision. After that, sentences are again disassembled into words for identifying key contract variables. TF-IDF algorithm is applied afterwards to assign weights to the words. Words are then aggregated into a single, common outline. Contract Analytics software also calculates the frequency of clauses during the aggregation process for key statistics [11] to check if a specific provision appears in most agreements or only in some agreements for analytics and automation purposes, the simplified version is shown as Figure 3.

While deconstructing legal documents may be hard, supervised deconstruction could be introduced to monitor this process and highlight the core of the provision or agreements for machines to learn the patterns more accurately. Natural Language Processing techniques could also be applied to further automate the deconstruction process by creating checklists for all types of agreement. Big Data tools like regular expressions can also be used for metadata extraction for identifying patterns and automation. With contract modularization, standardization, simplification methods, deep learning and other Machine Learning techniques applied to the machine, the automation of contract analytics life cycle can be achieved.

S. Wolfram personally has worked on his own language-- knowledge based Wolfram Language for more than three decades, which is expected to work as the bridge

“between the precision of traditional computer languages and the ability to talk about real-world construct” [12], according to the paper “Law, Symbolic, Discourse and the Constitution”. Wolfram Language has “a general formalism for describing human activities... a definition of specific objects, broken down by all kinds of details” [12]. In Wolfram Language, the natural language form will be “immediately converted to the precise internal form to establish a symbolic discourse system and a lot of details about the world” [12]. With the symbolic discourse system and details about the world ready, Wolfram thinks that practical systems, or framework, could be set up for computational law and from there, ethical and principle rules will be defined for AIs to follow. New laws that are “born computational” can also be generated if AIs follow Wolfram Language. Human beings could communicate with AIs through symbolic discourse language because Wolfram Language will be able to be understood by both human beings and AIs in that case.

3.2.2 Big Data Applications in Legal Departments

When law firms are trying to leverage their Data Analytics department and Data Analytics tools, other non-law firms are also trying deploying Big Data Analytics to help their legal departments make decisions about if they want to insource or outsource specific types of legal work. Given the appearance of Big Data tools, a lot of the legal work which could only be done by law firms before could now be automated. In that case, more and more companies are trying to use internal personnel and in-house counsels to handle their legal work. The budget constraint is another reason why non-law firms are trying to take legal work back in their own hands. However, they could not possibly retain legal teams that have the abilities to handle all types of legal work in-house. Therefore, they will have to decide which lines of legal work to retain for in-house legal teams and which to outsource to law firms. According to J. Tiano, “In the context of the qualitative insourcing analysis, the goal is to insource legal work that has a moderate-to-high risk/value profile” [9]. With the help of Big Data Analytics tool, it is possible for firms to quantify the risks and values to help them make the decision. Besides, Big Data Analytics techniques can also be used to “pinpoint specific tasks, not just general matters to be insourced” [9]. With Big Data, firms can generate “the universe of tasks” handled by outside counsels and inside counsels respectively and check if the tasks outsourced

can be accomplished by in-house legal teams to potentially save outside counseling fee. Task-based analysis can also be generated with Big Data Analytics tools. Task-based analysis can help companies “organize the frequency of specific types of tasks and the personnel best equipped to handle a specific type of legal work and the cost along with it” [9]. Big Data Analytics techniques can also help companies look at historic legal work and identify the types of legal work that are “handled outside, substantively similar, recurring and multi-jurisdictional and insource them to internal law or legal department to capitalize this opportunity” [9]. Doing this could potentially save the firm hundreds of thousands from legal fee. Big Data Analytics tools can also be used to identify the departments that are “heavy consumers of outside counsels” for the firm to decide if insourcing those lines of legal work would be more efficient and cost-effective. In addition to using Big Data Analytics tools to analyze internal data, Big Data Analytics tools can also be deployed to analyze outside counsel’s primary practice areas to see if the firm want to hire attorneys to handle the legal work in the most common primary practice areas of all the outside counsels. By employing an expertise-centric approach, they can also see if they should switch to practitioners in different fields that could lead to more cost-effective results. Moreover, Big Data Analytics tools can also be used to conduct “across the board seniority analysis” for in-house counsels to help make human resources decisions. The identification and retention of top talents can be done with the help of Analytical processes. By scrutinizing employee data, an organization’s competitive advantage can be enhanced which would essentially lead to higher employee productivity [17].

3.2.3 Big Data Applications in Administrative law and Criminal Justice

In addition to sentencing databases that have already been put into use which aims to decrease sentencing outcome disparity, Big Data Analytics platforms are proposed for information sharing in the connection between administrative law and criminal justice. Apache Pig and Hadoop can be used to handle structured, semi-structured and unstructured data, then Cassandra and Zookeeper can be used to extract, transform, deliver and store data. Machine Learning and Neural Network algorithms can also be deployed to classify and mine potential knowledge and the results can be stored in MongoDB so that all staffs can have access to the results through Web App [6]. With Big

Data, “behavioral optimization and personal law in which legal decisions and rules can be optimized for better outcomes and where law is tailored to individual consumers based on past analysis of past data” [6] can be achieved. According to L. Moses et al., “The court can combine the evidence in the platforms to conduct cases based on related law and through the proposed system, cases can be automatically classified into different categories like administrative personnel crimes or serious criminal offences. The results given back by their experiment shows that the filing rate efficiency is improved by more than 20% and the time used for processing large number of cases is significantly decreased via this system” [13].

4. Conclusion and Future Work

In conclusion, one of the few fields where you still barely see applications of Big Data today is the legal field with only Big Data tools like time management, billing, marketing, client relation management, legal or sentencing databases available. Even when all other industries are going paperless you could still see legal professionals reading and searching through mountains of paper based legal documents in the filing cabinet to conduct their due diligence. The legal industry is one of the most conservative industries historically. Even when the era of Big Data is at its peak at this moment, there are still a lot of suspicions regarding either whether Big Data can be applied to the legal world or how accurately and how efficiently Big Data can work in the legal field. However, the conservatism existing in the legal industry is also the exact reason why the legal industry has barely been touched by Big Data which makes it one of the most lucrative fields that still have the most potential for Big Data innovations. The more diversified background of the new generations of legal professionals made the applications of Big Data techniques or tools in the legal world inevitable for they are more tech-savvy and they demand more from technology. Besides the different backgrounds of the new generations of legal professionals, the exposure that legal professionals’ clients have had to Big Data techniques and tools outside of the legal world is also pushing the legal world to evolve to keep up with the pace of the revolution brought by Big Data. The other major reason for Big Data innovations to inevitably

happen in the legal world is that, the invention of Big Data techniques made it possible for semi-structured data and unstructured data to be processed and analyzed which means that, legal documents can be processed and analyzed with Big Data techniques. There can be many ways where Big Data applications or techniques can be integrated into the legal industry. Machine Learning or Natural Language Processing algorithms can be utilized to digitize legal documents and facilitate further processing and analysis. After digitalization, legal information can be stored and shared across different entities through Big Data and distributed computing platforms. The two leading legal database entities LexisNexis and Westlaw act more as search engines in legal research which leaves plenty of space for better information retrieval systems or information retrieval algorithms to come out to help legal professionals find the most related and also diverse sets of cases; Through other Big Data Analytics tools, legal professionals could gain insights about the waiting time and the potential case outcomes in specific jurisdictions. They can also have insights about the judges they are assigned which could help them make legal decisions like whether it is better to take specific cases to court or settle. If they want to take it to court, which jurisdiction should they file the lawsuits to have the least waiting time and best outcome possible. The insights given back by those tools can also be used to temper or persuade clients.

Judge analytics tools can also be used to analyze the judges that attorneys are assigned which could help attorneys avoid certain analogies or terms that the judges they are assigned don't like. This may or may not help them in getting a better case outcome. Contracts and other legal documents automation could significantly save attorney's time which could be more efficiently used in finalizing legal documents, maintaining their relationship with their clients or making litigation strategies. In that case Law firms can more efficiently utilize the time they bill the clients, so that they can lower the legal fee or even commoditize some of the legal work without having to do much extra work. They can assign a fixed legal fee to the lines of legal work that can be highly automated and commoditized which could help mitigate the complaints of their clients regarding high legal fees. The data that law firms have could help law firms figure out the sources of their clients, which could potentially help them make better marketing decisions to distribute the resources to the best place to have better outcomes. Machine Learning and

predictive models can also be designed by legal professionals who have Data Science backgrounds to better advise their clients in legal context. For patent attorneys, Big Data Analytics techniques can be utilized to help determine the similarity between patterns which could be of great help in patent infringement litigation. Other companies can also use Big Data tools to analyze if they want to take specific types of legal work that is usually outsourced back in-house since legal work outsourced to law firms is significantly more expensive. This could potentially save them money and give back better outcomes especially if the kinds of legal work require a lot of domain knowledge and internal information. There are many more possible applications of Big Data in the legal field that could dramatically ease legal professionals' life which are not concluded or suggested here.

However, there are also a lot of concerns associated with the applications of Big Data tools and techniques in the legal area mainly because of the lack of related law and regulation. For example, some law firms may cooperate with law schools for special projects since law firms themselves want to use Big Data techniques in their line of work but usually do not have the ability or resources to develop those techniques or tools. In this case they may need to share their confidential information with law schools or labs in order to get back the specifically designed Big Data techniques or tools. The confidential information is protected by attorney-client privilege in general but when a third party is involved and there is no law regulating the sharing of confidential information from the third party which means that clients' information could be leaked through the third party. Also, if the opposing counsel knows about the sharing of the confidential information to the third party, they could potentially subpoena the third party to potentially work around the attorney-client privilege, which will be a big concern for law firms or attorneys and might even be the obstacle for wider utilization of Big Data Analytics tools and techniques in the legal world. Future work could focus on the security of Big Data tools or data privacy laws and regulations which could decrease or eliminate the probability of leaking privileged or confidential information.

There can also be downsides or misuses of Big Data tools in the legal world because of it is less intuitive and transparent in how it works. Also, it could give back false results if chosen wrongly. Even if chosen correctly, the models can still be intentionally or

unintentionally misused to produce wrong predictions or conclusions, which could be used to misguide jurors and judges. Thus, it is very important to ensure accuracy and guard against illegitimacy. It is always important to remember that a tool exists does not mean that it is appropriate to use it regardless the context in future research and applications of Big Data in the legal industry.

The immaturity of Natural Language Processing is also limiting the usage of Big Data Analytics in the legal industry. Computer Scientists and Data scientists nowadays are applying Neural Network to Natural Language Processing and Text Analytics in the hope of finding a way around the need for actual interpretation of the language. The legal field is a field where the accurate usage of words can be crucial and the logic can be very complex. This complexity makes the correct and accurate understanding of specific words in specific contexts essential to understanding the arguments and reasonings of legal professionals and rulings of judges. It is also essential for straightening the logic behind them, which is also why, legal professionals tend to think that legal arguments or reasonings are more of a form of art. Legal rules are often vague and complex. Legal rules are also subject to change if the judges of the highest court can be convinced that those rules are wrong or outdated so that they are not applicable anymore. This means that simply applying legal rules can be fatal when it comes to legal analysis. Without the actual understanding of the words which is key to the understanding of the actual logic hidden behind the arguments, it is not very likely that the logic or reasoning can be actually interpreted, or at least interpreted correctly and accurately. That is why a lot of legal professionals still doubt the ability and effectiveness of Big Data tools when it comes to mirroring the legal arguments or legal expertise. However, this all depends on the future developments or breakthroughs in Natural Language Processing. If researchers in Natural Language Processing field could find a way to get around the need of actual interpretation of words for accurate interpretations of the legal logic, or a way for machines to actually and accurately understand the meaning of words which could lead to the accurate interpretation of the logic behind the reasonings and rulings, legal expertise might really be mimicked or mirrored, which would lead to a complete revolution in the legal world. In this case, future work can focus on Natural Language Processing for more breakthroughs.

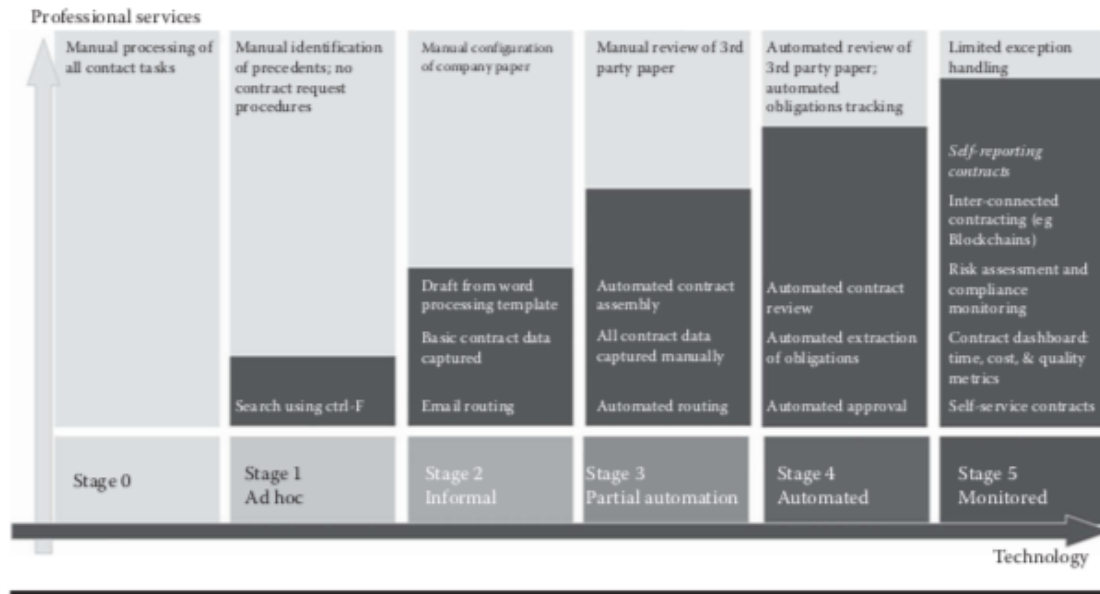


Fig.2 Contract Technology maturity model [15]

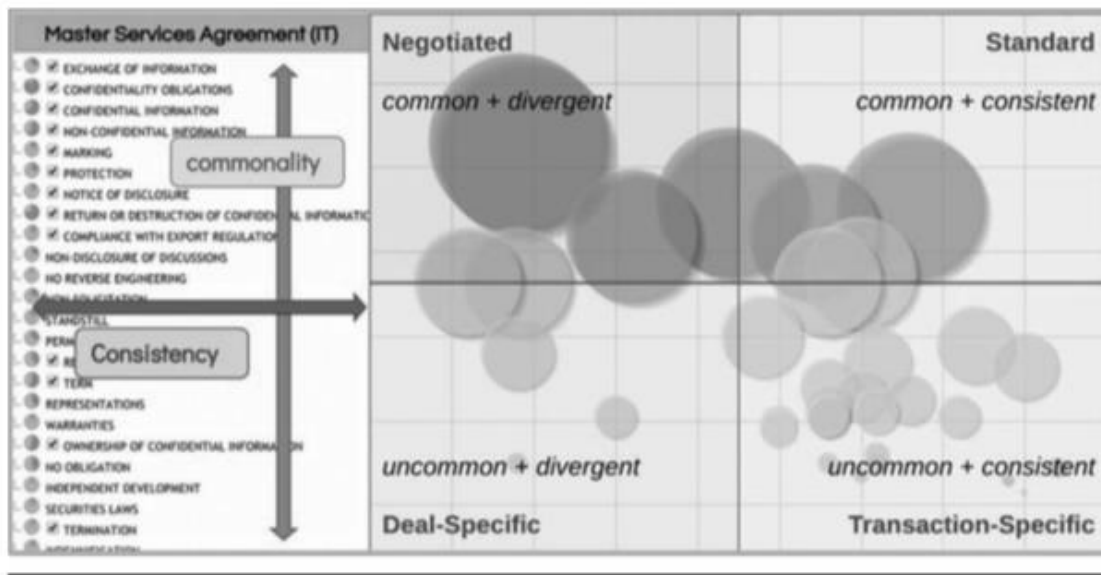


Fig.3 Contract Analysis Visualization [11]

5. Bibliography

- [1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinset Global Institute , 2011.
- [2] B. Marr, "How Big Data Is Disrupting Law Firms And The Legal Profession,"

- Forbes, 20 Jan 2016. [Online]. Available: <https://www.forbes.com/sites/bernardmarr/2016/01/20/how-big-data-is-disrupting-law-firms-and-the-legal-profession/#d78f0eb7c237>. [Accessed 20 July 2019].
- [3] Laura and John Arnold Foundation, "Developing a National Model for Pre-trial Risk Assessment," Laura and John Arnold Foundation, 2013.
 - [4] B. E. Harcourt, "Against Prediction: Sentencing , Policing , and Punishing in an Actuarial Age," *Public Law and Legal Theory Working Papers*, vol. 94, no. 13, p. 14, May 2005.
 - [5] C. Ma, T. Zhao and H. Li, "A Method for Calculating Patent Similarity Using Patent Model Tree Based on Neural Network," in *Advances in Brain Inspired Cognitive Systems*, Xi'an, China, 2018.
 - [6] N. Li, J. Zheng and M. Feng, "A Big Data Analytics Platform for Information Sharing in the Connection Between Administrative Law and Criminal Justice," in *Advances in Brain Inspired Cognitive Systems*, Xi'an, China, 2018.
 - [7] M. Koniaris, I. Anagnostopoulos and Y. Vassiliou, "Evaluation of Diversification Techniques for Legal Information Retrieval," *Algorithms*, 2017.
 - [8] S. Randazzo, "Data Tools Offer Hints at How Judges Might Rule; Services offer lawyers statistics on how likely a given case is to be dismissed," *Wall Street Journal*, 2016.
 - [9] J. J. Tiano, "Data analytics tools can inform a law department's initial insourcing decision," *Breaking News*, 2015.
 - [10] E. Walters, "Introduction: Data Analytics for Law Firms: Using Data for Smarter Legal Services," in *Data-Driven Law: Data Analytics and the New Legal Services*, CRC Press, 2019, pp. 1-11.
 - [11] K. Martin, "Deconstructing Contracts: Contract Analytics and Contract Standards," in *Data-Driven Law: Data Analytics and the New Legal Services*, CRC Press, 2019, pp. 33-58.
 - [12] S. Wolfram, "Computational Law, Symbolic Discourse, and the AI Constitution," in *Data-Driven Law: Data Analytics and the New Legal Services*, CRC Press, 2019, pp. 103-126.
 - [13] L. B. Moses and J. Chan, "2014 Thematic: Using Big Data for Legal and Law Enforcement Decisions 643 USING BIG DATA FOR LEGAL AND LAW ENFORCEMENT DECISIONS: TESTING THE NEW TOOLS," *University of New South Wales Law Journal*, vol. 37, no. 2, p. 643, 2014.
 - [14] G. O. Hernández , "Data (Gold) Mining: Law firms are beginning to figure out how to make money off their data analytics departments--if they can find the talent," *American Lawyer*, p. 48, 1 March 2018.
 - [15] K. Martin, "Legal Executive Institute," 16 June 2016. [Online]. Available: <http://legalexecutiveinstitute.com/contract-maturity-technology-assembly-line>. [Accessed 22 July 2019].
 - [16] V. RAGHUPATHI, Y. ZHOU and W. RAGHUPATHI, "Legal Decision Support: Exploring Big Data Analytics Approach to Modeling Pharma Patent Validity Cases," *IEEE*, vol. 6, 2018.

- [17] M. Burdon and P. Harpur, "Re-Conceptualising Privacy and Discrimination in an Age of Talent Analytics," *University of New South Wales Law Journal*, vol. 37, no. 2, p. 679, 2014.