
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202401.00281

A Review of “Data Mining in the Law Firm”

Authors: Liu Wenhao

Date: 2024-01-23T00:00:00+00:00

Abstract

Chapter 9 of this book is a relatively short paper authored by Kumar Jayasuriya, titled “Data Mining for Law Firms: Leveraging Internal Expertise to Drive Decision-Making.” It primarily discusses how law firms can utilize internal data to drive decision-making. The author contends that this data represents the most important yet underutilized resource for law firms; those with access to these data sources will have opportunities to create novel legal services. By harnessing unused data resources, law firms can transform into data-driven organizations and secure market advantages. The chapter is structured around seven sections: big data, adding value to legal practice, data-driven strategies, data teams, strategic target data, discovering internal expertise, and leveraging big data for innovation.

Full Text

Reflections on *Data Mining in the Law Firm*

This paper reflects on Chapter 9 of *Data Mining in the Law Firm*, a brief essay by Kumar Jayasuriya titled “Data Mining in the Law Firm: Leveraging Internal Expertise to Drive Decision-Making.” The chapter primarily addresses how law firms can utilize internal data to drive decision-making, arguing that this data represents the most important yet underutilized resource for law firms. Firms that can access these data sources will have opportunities to create new legal services, and by leveraging unused data resources, law firms can become data-driven organizations and gain market advantages. The essay discusses these issues through seven sections: big data, adding value to legal practice, data-driven strategies, data teams, strategic target data, discovering internal expertise, and leveraging big data for innovation. Rather than summarizing the chapter’s content, this paper explores questions raised during classroom presentations and in my own reflections, drawing on relevant literature.

1. Legal Big Data: Different from “Big Data” and “Judicial Big Data”

Data and big data are terms frequently mentioned in both academic and practical circles today. Unlike ordinary data, although big data has brought tremendous changes to our lives, few people can accurately define its concept. To understand something’s content, function, and utility, we must first understand its concept. In Chapter 9, the author cites the McKinsey Global Institute’s definition: a data collection whose scale in acquisition, storage, management, and analysis far exceeds the capacity of traditional database software tools, characterized by massive volume, rapid flow, diverse types, and low value density. However, the author considers this definition too narrow to fully explain big data’s connotation and proposes his own perspective. In the big data era, various industries have gradually adapted to development trends by formulating big data strategies, and the legal industry is no exception.

At the Fourth National Judicial Statistics Work Conference held by the Supreme People’s Court in August 2013, President Zhou Qiang explicitly articulated the judicial big data statistical concept of “big data, grand pattern, and great service.” China has gradually established judicial big data platforms such as the China Judicial Big Data Service Network. Current definitions of judicial big data mainly include connotative and denotative types. The former holds that judicial big data represents the combination of judicial activities and big data, a process described as: judiciary leveraging big data to complete established procedures, using publicly adjudicated results as the data foundation and source, employing big data’s complete index system to analyze all data, promoting case trials and enforcement, thereby transforming judicial management methods and enabling data to serve intelligent judicial construction. The latter considers judicial big data primarily as data oriented toward courts, comprising the sum and interrelationships of trial processes, enforcement information, legal documents, trial activity information, judicial administration, judicial personnel, and external assistance data formed by over 3,500 courts and more than 10,000 people’s tribunals nationwide in their judicial work. Regardless of connotative or denotative definitions, understandings of judicial big data remain basically consistent with only slight differences, indicating that academic concepts of judicial big data are relatively clear.

However, scholars’ definitions of legal big data differ substantially. Most scholars equate legal big data with judicial big data, a view holding that the data source for legal big data is publicly available court judgments. This perspective has certain limitations, as it overlooks the diversity of legal big data sources. From a legal concept perspective, “law” covers a far broader scope than “judiciary”; judicial big data is only the main data source for legal big data, and the two cannot be equated. Other scholars, from a “domain big data theory” perspective, consider law as one of the most important domain applications of big data. However, this view subordinates legal big data to the big data discipline, treating it merely as a component under big data, a position with which

I disagree. Legal big data is both a discipline and a research method. It is an emerging interdisciplinary field spanning law, statistics, computer science, and other disciplines—a typical cross-disciplinary subject. Simultaneously, legal big data is also a research method for studying legal issues, requiring both traditional research methods and big data research concepts and approaches. Data can provide intuitive, objective analysis results, thereby supporting problem analysis processes in digital form. As a broader data collection, big data yields more credible conclusions. As scholars have described, big data is called the “oil” of the future. People can collect, organize, and analyze massive data during work processes, abstract and mine valuable information behind the data, then apply it in practice to generate greater wealth. Big data can not only significantly impact practical work but also play a unique role in theoretical research.

2. Data Mining in Law Firms: What to Mine and What Are the Obstacles?

(1) Internal Data Mining and Obstacles Law firm internal data has diverse characteristics. First is internal client data, which is key for law firms’ market competition. Through customer relationship management systems, firms can analyze internal clients and use big data to accurately predict clients’ practical needs in case representation, thereby improving legal service efficiency and quality. Second is personal information about lawyers, paralegals, and staff, training situations, business areas, inter-member cooperation, as well as financial and accounting information—all of which can become objects of big data analysis and are most easily overlooked by law firms.

Among these, knowledge management and experience exchange within law firms constitute a major domain for internal data sharing. The most effective method for lawyer knowledge management is sharing knowledge within the firm, digitizing and tagging cases according to specific facts and their legal departments. By relying on software template programming settings, subsequent case file abstracts, case progress, core disputes, and judgment or ruling conclusions can all be displayed on the firm’s big data software, leaving work traces on the platform. This enables case handling processes to be completed step-by-step under big data guidance while serving as an experience and knowledge exchange platform, even making firm management simpler and clearer.

However, this model also has problems. Whether lawyers are willing to share their professional knowledge and practical experience is key to the feasibility of the above model. The big data concept for law firms has been discussed in academic and practical circles for years, yet firms’ big data strategies have not made significant progress. The reasons for this status quo are diverse. On one hand, some law firms are unwilling to develop or purchase corresponding software equipment due to current technical limitations and cost considerations. On the other hand, law firms themselves face obstacles in sharing professional knowledge and practical experience between different teams and among differ-

ent lawyers within teams. Some teams or lawyers consider this knowledge as a tool to compete with other competitors for market share and are therefore unwilling or unwilling to proactively share all professional knowledge, sharing only some basic knowledge. The key to this issue lies in the fact that law firms are essentially partnership organizations, which differ from corporate enterprises, leading to team-based operations within firms. Each team has its own database and market clients, making teams relatively independent. However, this also provides direction for law firm big data development—software development companies or law firms can develop big data software for internal teams based on these characteristics to collect, organize, and analyze data, thereby promoting internal firm data strategies.

(2) External Data Mining and Obstacles Fee Prediction. As legal service providers, law firms charge certain fees for their services according to the *Lawyers Law*, but fee quotation has always been a headache for firms. Disputes over lawyer fees perennially rank first among various disputes. The key reason why lawyer fees easily lead to disputes is that legal services, as a type of professional service, create information asymmetry, making it difficult for clients to predict the reasonableness of quotations. The commercialization of lawyers' overall image also increases clients' distrust of quotations. Relying on big data, law firms can reasonably predict their fees. By establishing corresponding databases and analyzing factors such as lawyer level, region, time, and competitors' requirements in different legal services, data analysis can yield reasonable quotations. This can, on one hand, enhance client trust, effectively explain the basis for quotations, and thereby reduce future disputes over fees. On the other hand, it helps reduce lawyers' time and energy costs wasted on fee issues, improves lawyers' work efficiency and case volume, and enables lawyers to focus more on serving clients.

Of course, fee prediction also faces obstacles. The foundation of fee prediction requires large-scale pricing data. Due to different firms' protection of their own data and to avoid unfair competition among firms, it is currently difficult to ensure that all firms are willing to provide objective and authentic pricing data and bases. Lawyer Jiang Yong from Tian Tong Law Firm believes, "Only by establishing an integrity platform and completing it through crowdsourcing, allowing lawyers and clients nationwide to contribute their data to this platform, can this work proceed." I agree with this view and suggest that, relying on the current "two-combination" management system, administrative authorities or industry associations should take the lead in establishing big data platforms through bidding processes, thereby solving law firms' pricing issues, reducing legal service costs, and improving legal service quality and efficiency.

Recruitment Prediction. Recruitment is the main form of internal personnel renewal and supplementation for law firms, and recruitment quality affects firms' business development and long-term growth. Many large law firms, both domestic and foreign, have begun using big data to recruit young lawyers. By

analyzing employees' personal information (such as gender, age, educational institution, degree, major) and professional capabilities (previously drafted contracts, internship experience, etc.), big data can predict the fit between candidates and the firm, even predicting how long candidates might work at the firm. Therefore, big data can be widely applied in firms' human resources departments. During resume review and interview candidate screening processes, it can provide in-depth understanding of candidates' characteristics, thereby improving recruitment targeting and objectivity, enhancing recruitment efficiency, and avoiding the drawbacks of low efficiency and strong subjectivity in manual review.

Case Analysis and Win Rate Prediction. Analyzing case facts and making reasonable decisions accordingly are important standards reflecting legal professionals' expertise and the foundation for legal service providers' survival. In the big data era, through storing massive case facts and setting primary and secondary parameters for identification and analysis, reasonable analysis of case facts can be achieved, providing reasonable decisions for clients. For example, Tian Tong Law Firm's big data analysis process in business operations includes the preliminary demonstration stage, consultation and representation stage, case handling stage, and case closing and archiving stage, achieving precise grasp of case facts through full-process data analysis. (See Figure 1 below) Additionally, big data can achieve win rate prediction, which is clients' primary concern when retaining lawyers. In the big data era, through digitized analysis of massive cases, win rate prediction can be realized. As one data company states, "We help lawyers predict the future."

However, current case analysis and win rate prediction also face skepticism. This skepticism mainly stems from two aspects. First is the barrier to judicial data sharing. The degree of judicial data disclosure is seriously insufficient, with some judicial data becoming internal data for judicial administrative agencies, courts, and procuratorates. Additionally, some content is hidden in publicly available data. For example, current judgment documents hide judges' information, creating difficulties for lawyers conducting similar case searches. Second, some scholars believe current judicial big data analysis has not reached a mature state, and so-called big data analysis is merely a big data mindset or concept. Due to insufficient foundational data scale and immature corresponding technology, its predictive conclusions also have limitations.

[Figure 1: see original paper]

3. Data Mining Strategies for Small and Medium-Sized Law Firms: Why Important and How to Mine?

(1) Current Situation and Importance of Data Mining for Small and Medium-Sized Firms According to statistics from the Ministry of Justice website, by the end of 2020, China had 34,000 law firms. However, in terms of firm size, small and medium-sized law firms are numerous and widely dis-

tributed. Among them, there were over 22,700 firms with 10 or fewer lawyers, accounting for 66.02%; over 7,100 firms with 11-20 lawyers, accounting for 20.83%; over 3,400 firms with 21-50 lawyers, accounting for 10.08%; over 680 firms with 51-100 lawyers, accounting for 1.99%; and over 360 firms with 100 or more lawyers, accounting for 1.08%. (See Figure 2 below)

Law firms are important venues for lawyers' practice, and their strategic positioning affects industry competitiveness and development. Small and medium-sized firms are China's main form of law firms and the main force providing grassroots legal services. Research on big data's impact on law firms cannot ignore the status of small and medium-sized firms; corresponding research should be conducted to use big data to improve these firms' legal service efficiency and quality. However, a practical problem is that current leaders of small and medium-sized firms mostly lack big data concepts, let alone implement data-driven strategies, still operating in autonomous decision-making mode. Although these small and medium-sized firms cannot develop advanced big data application software, they can still use some basic data for statistical analysis to derive coping strategies and discover potential clients.

[Figure 2: see original paper]

(2) Obstacles and Strategies for Data Mining in Small and Medium-Sized Firms Excessive cost is the biggest obstacle for individual practitioners and small and medium-sized law firms in utilizing big data. Software development costs and data utilization costs are the biggest challenges facing these firms. However, in fact, these small and medium-sized firms possess large amounts of current, past, and even potential client data, but this data lacks necessary management and utilization. For most small and medium-sized firms, effectively storing, maintaining, and utilizing this data is a feasible path. Regarding analysis software, these firms can use currently available market applications, even some free software, for data collection, storage, processing, and analysis.

Small and medium-sized law firms can utilize various data, including clients, internal personnel information, represented cases, financial information, and marketing strategies. These firms can collect and store this data through cloud-based electronic systems. Of course, how to collect this data is also a complex issue. Collecting client information requires dividing relevant client data into different fields according to various categories. Foreign research suggests this data is best placed in LPM (Legal Practice Management) systems or CRM (Customer Relationship Management) systems. These systems can achieve comprehensive management functions for practice, contacts, and finance. Additionally, some specialized financial management systems can also achieve comprehensive management of law firm financial data.

4. Ethical Regulation of Data Teams Under Data Mining: Who Regulates and Based on What?

Chapter 9 mentions that for law firms to become data-driven organizations, they need data teams composed of data scientists and business analysis experts for data management. It particularly emphasizes that team members must possess both legal practice knowledge and information technology training. This reflects that the legal big data context will 衍生 new profession types. The key question is whether these newly emerging professions are legal professions. If defined as internal legal professions within law firms, they would be required to comply with specific legal professional ethics norms—but what would these norms be? Another view holds that these professions do not belong to legal professions because this work can be performed by specialized data companies developing software for sale to law firms for their teams' use. Therefore, they are not legal professions and would be subject to professional ethics requirements but not legal professional ethics. However, based on the professionalism, complexity, and particularity of legal work, I support the first view.

1. American Bar Association Regulation of Non-Lawyer Staff

If these new professions are defined as legal professions, are there corresponding legal professional ethics norms to regulate these professionals? The American Bar Association's *Model Rules of Professional Conduct* Rule 5.3 addresses the responsibilities of non-lawyer assistants:

Rule 5.3 Responsibilities Regarding Nonlawyer Assistants

With respect to a nonlawyer employed or retained by or associated with a lawyer:

- (a) a partner, and a lawyer who individually or together with other lawyers possesses comparable managerial authority in a law firm, shall make reasonable efforts to ensure that the firm has in effect measures giving reasonable assurance that the person's conduct is compatible with the professional obligations of the lawyer;
- (b) a lawyer having direct supervisory authority over the nonlawyer shall make reasonable efforts to ensure that the person's conduct is compatible with the professional obligations of the lawyer; and
- (c) a lawyer shall be responsible for conduct of such a person that would be a violation of the Rules of Professional Conduct if engaged in by a lawyer if:
 - (1) the lawyer orders or, with the knowledge of the specific conduct, ratifies the conduct involved; or
 - (2) the lawyer is a partner or has comparable managerial authority in the law firm in which the person is employed, or has direct supervisory authority over the person, and knows of the conduct at a time when its consequences can be avoided or mitigated but fails to take reasonable remedial action.

Analysis of this provision shows that the U.S. constraints on non-lawyer assistants' behavior are achieved by imposing more obligations on lawyers—that is, lawyers should provide necessary guidance, management, and supervision in their conduct—without directly imposing professional ethics norms on these new professionals.

2. Chinese Bar Association Regulation of Non-Lawyer Staff

China's *Lawyers Law* lacks explicit provisions regulating these new professions, but Article 5 of the *Code of Conduct for Lawyers* stipulates: “This code applies to lawyers and law firms that are members of the All China Lawyers Association; other practitioners in law firms shall refer to this code for implementation.” This indicates that future new professions in law firms can be constrained by professional ethics using this code for reference.

However, whether American Bar Association provisions or Chinese bar association regulations, these norms for new professions are made indirectly and from the side, illustrating the lag of law itself. Non-lawyers can also become law firm partners. The UK's *Legal Services Act 2007* removed traditional restrictions on partner lawyer identity and nature in law firms. Article 18 of the *Hainan Special Economic Zone Lawyers Regulations* stipulates that “other professionals such as certified public accountants, certified tax agents, certified cost engineers, and patent agents can become partners in special general partnership law firms.” Due to the particularity and importance of data team professions, these professionals are likely to become law firm partners and therefore require specialized, direct professional ethics norms for constraint and limitation. Most academic research on lawyer professional ethics also includes other law firm personnel. Therefore, legislatively, for new professions emerging from data teams, certain rights and obligations should be granted to regulate their professional conduct and achieve consistency between rights and responsibilities.

Having addressed whether regulation is necessary, we must also respond to the question of who should regulate. I believe that for new profession types formed by law firm data teams, the regulatory subjects for their professional ethics should also be diversified. On one hand, internal regulation should be conducted by law firms, formulating rules and regulations for data teams, with firms and lawyers responsible for supervision and inspection. On the other hand, external regulation should comply with the *Lawyers Law*, establishing a “two-combination” supervision and management system of judicial administrative organs and bar associations—that is, judicial administrative organs formulate professional ethics norms at the macro level, while bar associations formulate specific implementation rules or procedural regulations to guide, supervise, and undertake professional ethics matters for law firm internal data team members.

5. Training Young Lawyers Under Law Firm Data Mining Strategies

Training young lawyers is an unavoidable topic for law firms. The particularity of the legal profession requires lawyers to possess not only professional theoretic-

cal knowledge and strong professional ethics but also long-term practice to truly improve their professional capabilities. As Justice Holmes stated, the life of the law lies not in logic but in experience. In traditional training systems, young lawyers in law firms mainly engage in basic legal work such as legal research, drafting basic legal documents, and simple legal consultations. However, in the big data era, these basic legal tasks that young lawyers use for practice are gradually being replaced by technology. Through big data retrieval and analysis, laws and regulations can be precisely located, legal documents automatically generated, and clients' legal questions promptly answered. Young lawyers will face tremendous challenges from big data. Moreover, with the continuous development of technology and big data, machine replacement of human labor will become a universal phenomenon in society. For the legal industry, large law firms can fully resist technological transformation and "survive" due to their abundant capital and advanced strategic positioning. Small and medium-sized law firms, however, will be at a disadvantage in fierce market competition and may even be eliminated due to deficiencies in technology, capital, and concepts. This will further lead to reduced internship positions and practice opportunities for young lawyers, profoundly demonstrating the challenge big data poses to young lawyer training models.

Facing this development trend, law firms must innovate young lawyer training models. On one hand, they must emphasize legal big data concepts, integrating legal big data throughout young lawyer training, and continuously improving young lawyers' professional levels and competitiveness through flexible use of big data tools. Combining emerging professions such as data teams and leveraging law students' professional advantages, young lawyers should be encouraged to engage in these emerging professions to enhance their competitiveness. Additionally, young lawyers must recognize the shortcomings of technology and big data. While legal big data can greatly improve the quality and efficiency of legal products, it is not omnipotent. "Currently, the vast majority of legal big data research primarily sources from publicly available judgment documents, but factors that significantly influence case handling may not all be written into judgments. Even when finding a needle in a haystack of massive cases, some effective information in judgment documents remains underutilized, and some legal big data research still faces the danger of small-sample thinking." Therefore, young lawyers can utilize their professional knowledge to engage in areas where legal big data cannot yet provide services or provides poor service quality, relying on professionalism to enhance competitiveness.

Moreover, this conceptual transformation must also permeate legal education. China is the world's largest legal education country, with the highest number of law schools and law students. This is both an achievement of China's higher education development and a challenge. Law schools should transform traditional education models, introducing new concepts such as technology, Internet Plus, and big data as key teaching content. Additionally, they can attempt to establish interdisciplinary subjects combining law and big data, encouraging statistics and computer science students to pursue law degrees to adapt to

current trends. Some institutions have already made attempts in this regard. For example, Tsinghua University established the Institute for Intelligent Rule of Law, actively conducting research on legal issues in information technology, empirical analysis of legal big data, and legal technology application research under the guidance of the rule of law and cyberpower strategies. Southwest University of Political Science and Law established the Artificial Intelligence and Law School. China University of Political Science and Law and Beijing Institute of Technology have also established data law and artificial intelligence law disciplines. Of course, it must be acknowledged that these current reforms still face issues such as insufficient teaching staff and mismatched training effects with training objectives, requiring further improvement.

6. The Impact of Big Data on “Legal Professionalism”: Shock, Challenge, or Promotion?

“Legal professionalism” is the mainstream concept of Western lawyer professional positioning, recognizing and pursuing industry professionalism, public service, and autonomy. Legal professionalism is an outstanding achievement in Western lawyer system research and an important theoretical foundation in professional sociology and lawyer studies. However, this theory’s research on lawyer systems is based on traditional practice models. Big data and technology have brought tremendous changes to the legal industry. What changes will big data bring to the three cornerstones of legal professionalism?

1. Big Data and Professionalism

Both promotion and challenge. Legal professionalism emphasizes lawyers’ professionalism, requiring common knowledge backgrounds, value concepts, and emotional identification. The core of professionalism lies in receiving professional law school education and obtaining industry admission qualifications. Abel believes that the professionalization process of the legal profession is in fact a “professional project” aimed at monopolizing the market, with core characteristics being control over industry admission and control over legal service output processes. Lawyers can achieve “monopoly” of the legal market precisely because of their professionalism. Under big data conditions, establishing massive case databases can greatly improve the accuracy and efficiency of lawyers’ decisions, making lawyers more specialized—this is also the key to law firms’ big data strategies. However, we must also note that under the big data background, the boundaries of legal profession “monopoly” are gradually fading. On one hand, with the application of technology and big data, some traditional legal services such as case consultation, contract document generation, case analysis, and win rate prediction can be replaced by emerging software. The public without legal education backgrounds can also obtain corresponding legal services by paying certain fees or even for free, no longer seeking lawyers’ help. On the other hand, the big data background will create some new professions, such as data scientists or business analysis experts, who will become important components within law firms, the core auxiliary to lawyers in providing legal

services, and key to improving firms' competitiveness. However, these people may not possess legal knowledge, illustrating that the boundaries of legal profession "monopoly" are weakening, with more and more non-law graduates able to engage in law-related professions.

2. Big Data and Public Service

Promotion of public service. Public service represents the opposition between professionalism and commercialism. As Chief Justice Rehnquist of the U.S. Supreme Court metaphorically described: "A lawyer is a subtle mixture of merchant and priest—the former aims to maximize profits, while the latter completely disregards economic demands." It must be noted that the requirement of public service under professionalism is always primary, and commercialism can never surpass professionalism. Under the big data background, on one hand, since some basic legal services have been replaced, the public can obtain them at low cost or even for free, which will inevitably impact the profitability of law firms' basic legal services. However, reducing legal service costs also reflects public service to a certain extent. On the other hand, big data provides new platforms for law firms and lawyers, enabling lawyers to conduct legal aid and public service, demonstrating lawyers' social responsibility and further manifesting public service.

3. Big Data and Autonomy

Promotion of autonomy. Autonomy is the greatest characteristic of legal professionalism and the mark distinguishing it from other professions. Western bar associations are highly autonomous organizations that can control and manage their internal members, thereby maintaining industry reputation and achieving professional monopoly. A major characteristic of autonomy is supervision and regulation of internal members. Big data conditions create convenient conditions for industry association supervision. For example, as law firm sizes continue expanding, lawyer conflict-of-interest review is a very complex issue. By inputting lawyer information, represented cases, family relationships, and same-firm lawyer information into big data software systems, effective supervision of lawyers' violation of conflict-of-interest rules in representation activities can be achieved. Similarly, big data can also be used to manage former judges and prosecutors engaged in legal practice. In summary, big data has a coexisting relationship of challenge and promotion with legal professionalism. We must fully utilize big data concepts to achieve legal professionalism's professionalism, public service, and autonomy.

Wang Jiayun, "Judicial Big Data and Realization of Judicial Justice," *Journal of Jishou University* (Social Sciences Edition) No. 2 (2020): 137.

Sun Xiaoyong, "Application and Prospects of Judicial Big Data in Chinese Courts," *China Legal Science* No. 4 (2021): 124.

