CHAPTER 16

Government Analytics Using Machine Learning

Sandeep Bhupatiraju, Daniel Chen, Slava Jankin, Galileu Kim, Maximilian Kupi, and Manuel Ramos Maqueda

SUMMARY

The use of machine learning offers new opportunities for improving the productivity of the public sector. The increasing availability of public sector data and algorithmic approaches provides a conducive environment for machine learning for government analytics. However, the successful deployment of machine-learning solutions requires first developing data infrastructure of the required quality to feed these algorithms, as well as building the human capital necessary to develop them. Ethical principles regarding the use of machine-learning technologies must be defined and respected, particularly for the justice system. This chapter provides an overview of potential applications of machine learning in the public sector and in the justice system specifically, as well as the necessary steps to develop them sustainably and ethically. It then analyzes the case of machine-learning deployment in India to illustrate this process in practice.

ANALYTICS IN PRACTICE

- Machine learning is fundamentally a methodological approach: it defines a performance indicator and uses collected data to train an algorithm to improve this indicator. Because of this relatively broad definition, machine learning includes different algorithms and may be applied in a variety of domains, from payroll fraud detection to court rulings. This flexibility requires practitioners to make key design decisions: what kind of performance indicator will be used? What training data and algorithm will be deployed? These decisions may substantially alter the machine-learning algorithm's results. Making these decisions thus requires close collaboration between machine-learning engineers, domain experts, and the agencies that will use the technology.

The authors' names are listed alphabetically. Sandeep Bhupatiraju is a consultant in the World Bank's Development Impact Evaluation (DIME) Department. Daniel Chen is a senior economist at DIME. Slava Jankin is a professor of data science and public policy at the Hertie School. Galileu Kim is a research analyst at DIME. Maximilian Kupi is a PhD candidate at the Hertie School. Manuel Ramos Maqueda is a research analyst at DIME.

- Machine learning can leverage large amounts of administrative data to improve the functioning of public administration, particularly in policy domains where the volume of tasks is large and data are abundant but human resources are constrained. Governments generate large amounts of administrative data on an almost-daily basis, but these data are seldom used to improve the production function of public administration. At the same time, civil servants are constrained in the amount of time they can dedicate to complete tasks—as well as in the amount of information they have readily available. Machine learning can process large amounts of administrative data, structuring them around performance indicators that algorithms are well suited to optimize. For example, machine-learning algorithms can be trained, using procurement data, to predict whether new, incoming contracts are irregular or not, at a scale and speed which far exceed human capacity.
- While machine learning can offer efficiency gains in public administration, governments need to be aware of their role in generating and using measurements in public administration, as well as their ethical responsibilities. Machine-learning algorithms require extensive data and measurements on both citizens and civil servants, but these data are often collected without their consent. As a result, governments should be transparent about how these data are being used to enhance public administration and how these technologies are being used to affect public administration. Care should be taken not to reproduce biases, such as racial or gender discrimination, in the machine-learning algorithms.
- To fully reap the benefits of machine learning, governments must undertake long-term investments in data infrastructure and human capital. Before machine learning is implemented, investments in data infrastructure must take place. Data quality needs to be improved and data pipelines must be developed to train the algorithm. Additionally, specialized machine-learning engineers must be hired and trained to implement the technology in the public sector. These investments are costly and require long-term planning: governments should not expect machine-learning technologies to be developed overnight.
- Machine-learning experts should collaborate with subject matter experts to guide how machine-learning technology will benefit the civil servants who will use it. Besides the technical knowledge to develop and operate machine-learning technologies, substantial levels of domain and political expertise, as well as awareness about potential ethical and legal pitfalls, are necessary to ensure the effective use of a machine-learning solution. For example, if the machine-learning technology is meant to assist judges in reducing racial bias, judicial experts and judges themselves should be consulted to ensure that relevant performance indicators and data are used. Including civil servants in the development of the solution also facilitates the adoption of the new technology.
- Machine learning is not a panacea, and practitioners should be aware of the limitations of the approach to fully leverage its benefits. Algorithms are limited to what is measurable by data, and performance indicators may reflect the bias of machine-learning engineers and the data themselves. Improving a particular performance indicator may not necessarily be the best way of achieving a policy goal. As a result, machine-learning applications should not be considered as a substitute for policy making but as a tool to complement and enhance decisions made by government agencies and their civil servants.

INTRODUCTION

Machine learning is a discipline that focuses on the development of computer systems (machines) that, through the analysis of training data, can improve their performance (learn) (Jordan and Mitchell 2015). Recent advances in data collection and processing power have expanded opportunities for machine-learning applications in a variety of fields. Advances in machine learning have brought tangible benefits in the worlds of business and medicine and in large-scale systems more generally (Brynjolfsson and McAfee 2017). However, this growth in opportunity has often led to excessive optimism about what machine learning can accomplish, as well as a tendency to downplay the potential steep costs of deploying machine-learning technologies (Chen and Asch 2017).

We start our discussion by offering a general definition of machine learning. What distinguishes machine learning from other methodological approaches is the definition of a learning problem under a statistical framework. Following Jordan and Mitchell (2015, 225), we define a learning problem as a "problem of improving some measure of performance when executing some task, through … training experience." The following example illustrates a machine-learning approach. Suppose a government is interested in reducing irregularities in its payroll system. One measure of performance would be the proportion of irregular paychecks correctly identified by the machine. The training experience—or data—would be a collection of paychecks manually classified by payroll analysts (see case study 2 in chapter 9). The learning problem would thus define a statistical model that learned how best to predict irregular paychecks by being trained on historical payroll data.[1]

In this chapter, we discuss applications of machine learning to public administration. We outline the data infrastructure and human capital requirements for developing machine-learning applications, as well as potential complementarities with public servant surveys. As noted in chapter 9 of *The Government Analytics Handbook*, the foundational step for any form of data analytics—including machine learning—is the development of a robust data infrastructure. We also highlight ethical concerns regarding the development and deployment of machine-learning applications, which relate directly to the discussion in chapter 6. Despite our focus on machine learning, we consider the broader shift to a data-driven and statistically informed culture—regardless of the implementation of algorithms—to be often already sufficient for bringing substantial benefits to public service delivery. These benefits include organizational changes, data literacy, and performance monitoring. With the transition to data-driven policy making, machine-learning applications are a natural next step in government analytics, automatically leveraging data to improve the performance of public administration through well-defined performance metrics.

Following this broader discussion of machine learning in public administration, we focus on machine-learning applications in justice systems. Within public administration, the justice system generates a large number of case rulings linking legal cases and actors (training data) to ideally fair rulings (performance). Thus, a potential machine-learning learning problem is: can we identify and reduce racial bias in court rulings by training an algorithm on a collection of case rulings? Instead of defining what a fair ruling is, we might define what an unfair ruling is. For instance, the decision of the judge should not be influenced by extraneous factors, such as the time of day or the race of the defendant. By identifying cases in which the decision has been influenced by such biases, a machine-learning model can potentially identify and ultimately prevent unfair rulings. This approach thus allows machine learning to improve the quality and fairness of judicial decisions (Ramos-Maqueda and Chen 2021). Despite this promise, limited data literacy and statistical training inhibit applications of data analytics in general—and machine learning specifically—in the judiciary.

Machine learning is not a panacea: it requires significant investments before any of its benefits come to fruition. As we detail throughout this chapter, machine-learning applications require investments in data infrastructure and the development of the human capital necessary to develop and deploy machine-learning algorithms. Undertaking these investments—and often long cycles of development—requires resources and long-term horizons before the benefits of the approach become apparent. Additionally, ethical concerns regarding embedded racial or gender biases in training data highlight how technologies can inadvertently reproduce the same human biases they were designed to eliminate. Thus, initial optimism regarding the revolutionary potential of machine-learning approaches should be balanced by a recognition of their limitations (Cross 2020).

Practitioners have much to gain from deploying machine learning in public administration. Defining performance metrics and automating the training of algorithms through large-scale data can improve the functioning of public administration—particularly when oriented toward well-defined tasks with an abundance of quantitative data. Performance improvement can come simply from embedding a data-driven approach to government functioning. There is not always a need for sophisticated approaches in machine learning to make progress. In fact, machine learning generally comes only after more basic steps have been taken on data management and analytics in public administration, as highlighted in other chapters of the *Handbook*.

This chapter is structured as follows. Section 1 describes applications of machine learning in public administration. Section 2 then outlines a road map to applying machine learning in the public sector, focusing on data infrastructure requirements and human capital needs. Section 3 shifts our focus from public administration in general to the justice system. In doing so, it highlights applications of machine learning in the justice system, as well as the data infrastructure and human capital required to implement them. Section 4 presents a case study from India illustrating machine-learning approaches to justice in practice. Section 5 moves beyond descriptive analysis to outline how machine learning can be used to assist causal inference. Finally, we conclude.

MACHINE LEARNING FOR PUBLIC ADMINISTRATION

The use of machine learning is spreading across many functional areas of public administration. While European Union (EU) governments focus on service delivery and public engagement, other areas, such as internal management and law enforcement, are progressively being targeted for the deployment of machine-learning solutions to increase their efficiency and effectiveness (Misuraca and van Noordt 2020). The applications are diverse, from detecting COVID-19 outbreaks to simulating the impact of changes in macroeconomic policy. Machine-learning applications thus provide novel ways for governments to use their data to improve public administration. Chapter 15 of the *Handbook* highlights how machine learning can be used to detect similarities between goods in public procurement.[2]

The use of machine learning provides a few advantages compared with more standard analytical approaches. Standard data analytics provides the analyst with tools bounded by the analyst's capacity to investigate connections between variables in the data—often the coefficients in a regression specification. However, in many public administration settings, the analyst is confronted with factors—individual or organizational—that may influence a policy outcome without the analyst's knowledge. Machine learning enables the exploration of relationships between variables in a principled, and often unsupervised, way. However, causality in machine learning is a relatively recent development, and it presents considerable challenges.[3] Potential applications, therefore, focus less on causal interventions or experiments and more on solutions that, based on given data, best perform an accurate prediction.

The primary focus of this section is on applications of machine learning for administrative data.[4] However, governments may leverage public servant surveys to complement this analysis, particularly for personnel data. For example, a government may be interested in better understanding job satisfaction and how it relates to staff turnover. While a machine-learning analysis could be useful in identifying potential patterns in civil service exit from the full population of interest as a function of demographics (sex, age, education, or race), it might not provide much information about the attitudes of the staff who are at risk of exiting. A public servant survey provides a complement for answering this kind of "why" question, but it may not be large enough to find general patterns in the first place—particularly if it is not linked to administrative data on exits, which is often difficult to do. Thus, machine-learning applications on human resources data outperform surveys at identifying certain kinds of patterns, but they need to be complemented by surveys explaining these patterns and highlighting potential interventions that might address problems.

Machine-Learning Applications

Applications of machine learning can be subsumed under the following three categories.

Detection and Prediction

Machine learning can help policy makers detect and predict destructive events, improving the design and implementation of adequate policy measures. This is the largest application of machine-learning approaches, addressing issues such as COVID-19 outbreaks, fake news, hate speech, tax fraud, military aggression, terrorist activity, cyberattacks, natural disasters, street crime, and traffic congestion—to mention only a few. While the detection and prediction of destructive events is only the first step toward effective government intervention, it is an important instrument for effective policy making.

For example, in Delhi, over 7,500 CCTV cameras, automatic traffic lights, and 1,000 LED signs are equipped with sensors and cameras that collect traffic data, which a machine-learning system processes into real-time insights. Local authorities can then decide how to balance traffic flow in real time and identify traffic patterns and congestion trends in order to plan for the long-term mitigation of traffic problems (Devanesan 2020). Besides these benefits, which are geared toward improving general traffic flow, these systems are also used by the Delhi Police to track and enforce traffic violations, such as speeding or illegal parking (Lal 2021).

More generally within public administration, machine-learning approaches have been used to improve the machinery of government itself. For example, chapter 15 of the *Handbook* highlights how machine-learning techniques have been applied to detect corruption in public procurement. One prominent example of this application is the use of decision tree models (random forest and gradient boosting machine) to detect the presence of Mafia activity in procurement contracts in Italy (Fazekas, Sberna, and Vannucci 2021). In Brazil, federal agencies have deployed machine learning to detect evidence of corruption in federal transfers to municipal governments, as well as in irregularities in paychecks issued to civil servants, as described in case study 2 in chapter 9.[5]

Simulation and Evaluation

Simulating and evaluating the impact of future policy measures is another widespread application area for machine learning. Simulating the potential costs of a policy measure against its expected benefits has become an increasingly relevant tool for governments. For example, in the United States, a simulation known as the National Planning Scenario 1 allows policy makers to simulate what might happen if Washington, DC, were subject to a nuclear attack (Waldrop 2018). Whether policies are designed to stimulate the economy or to contain the spread of a virus, simulation and evaluation provide valuable insight to policy makers before implementation, allowing them to choose which policies maximize intended effects.

Personalization and Automation

Machine learning can also be applied to the personalization and automation of government processes and services. For example, policy makers may customize digital government services for parents to every life stage of their newborn child or tailor the provision of health care services to each patient's particular needs. Additionally, the automation of repetitive tasks leaves more time for public servants to do other tasks. All in all, these novel technologies may help governments be more efficient in their use of time and increase their responsiveness to citizens' needs.

A medical example illustrates this approach. There has been growing interest within the US federal government in using machine learning to improve public health outcomes. A series of pilots to develop such machine-learning solutions have been rolled out. These include the prediction of potential adverse drug reactions using medical reports, the classification of whether a child is likely to have autism based on medical records, and the prediction of unplanned hospital admissions and adverse events (Engstrom et al. 2020, appendix). Another study has found, through the application of machine-learning techniques, that physicians overtest low-risk patients but simultaneously undertest high-risk patients (Mullainathan and Obermeyer 2022).

Practical Steps for Machine Learning in Public Administration

The implementation of machine learning in public administration comprises two key steps. The first is building a high-quality data infrastructure to feed the necessary training data to the machine-learning algorithm. Because public administration data infrastructures are often developed without machine-learning applications in mind, adaptation is often necessary. New data pipelines need to integrate public sector information systems that previously operated in isolation, such as public procurement and budget data. Data standardization practices, such as ensuring that variables in different data tables are named consistently, and other quality checks need to be in place to ensure that the data fed to the machine-learning system are accurate and comprehensive.

Another key step is developing the human capital necessary to deploy machine learning. Before fulfilling the promise of automated, self-learning algorithms, a team of human developers is necessary to set the system in place. In fact, the entire pipeline, from data infrastructure to the training of the algorithm to disseminating actionable insights for policy makers, has to be designed by humans. Having an in-house team capable of developing and maintaining machine-learning applications is crucial. Continuous collaboration between the machine-learning implementation team and policy colleagues who will use its insights ensures that applications are adapted for and stay relevant to public administration's needs.

In the following sections, we dig deeper into these steps. In so doing, we highlight examples of strategies to ensure that both the data infrastructure and the human capital requirements are in place to deploy machine learning in public administration.

Public Sector Data Infrastructure for Machine Learning

Machine learning requires large volumes of data. These data should be of high quality: they should be comprehensive, covering all measurements necessary for the algorithm, and complete, reducing to the extent possible any gaps in measurement that may arise. A robust data infrastructure ensures that these two principles are respected and is a prerequisite for any machine-learning application. The implementation process may require upgrading legacy information systems or integrating new systems into old ones to process the resulting, often large, data sets. Practitioners may benefit from referring to the *Handbook*'s wider discussion of how to reform data infrastructure for analytical insights; this discussion provides lessons that apply to machine-learning settings as well (chapter 9).

Some types of data structure may be more amenable to machine learning than others. Machine-learning applications often require structured data—with well-defined formats and measurements—so policy areas that traditionally deal with structured data, such as finance or budget data, lend themselves particularly well to it. (For an overview of using budgetary data for analytics, see chapter 11.) At the same time, governments produce unstructured data—which lack a predefined data format—such as written documents, meeting recordings, and satellite imagery. To take full advantage of this range of data, practitioners should develop a flexible storage solution that accommodates different types of data. This flexibility should be complemented by thorough documentation of data collection and standardization practices, as well as by measures to ensure compliance with data security regulations, such as the EU's General Data Protection Regulation (GDPR).

The deployment of this data infrastructure requires long-term, costly investment. Data engineers and information technology technicians should partner with the machine-learning implementation team to define data requirements, identify relevant variables, and connect the machine-learning applications to the data infrastructure. The development of a robust data infrastructure is foundational for the effective deployment of machine learning in public administration and should always precede it. Since the data infrastructure is embedded within public administration, its development requires careful coordination between machine-learning engineers, data engineers, and their institutional counterparts who own permissions to the data. Data should be integrated across government agencies to ensure that the largest pool of data is made available for training the application. Open communication between teams and agencies is therefore key.

Human Capital Requirements for Machine Learning in Public Administration

A sustainable machine-learning application is often best achieved by building on in-house human capital. This ensures that the developed solutions are in line with existing government regulations and that policy choices are encoded faithfully. Furthermore, in-house machine-learning experts will be more likely to possess the necessary subject and political expertise required to implement machine-learning solutions in a policy area. Finally, even if an agency decides to rely on external service providers, a certain level of embedded expertise is required to know what is technically possible and feasible, as well as to make informed judgments about the quality of contractor-provided solutions.

To build the necessary skills infrastructure in government organizations, it is first necessary to understand what competencies are needed for machine-learning developers. Naturally, knowledge about machine-learning and deep-learning algorithms is necessary. Beyond this basic knowledge, methods for dealing with large-scale data and databases in general and knowledge about distributed computing systems are also key. To successfully develop, deploy, and operate machine learning in government, familiarity with human-centered design and acquaintance with the legal and ethical frameworks in public administration are important. Finally, policy-area expertise and knowledge about governance and policy making in general enable machine-learning applications to be anchored in the operational needs of government.

Integrating the necessary skills infrastructure within government organizations often proves to be difficult. Hence, governments should follow one or more of the following best practices. Machine-learning talent does not usually follow the classical tenure path of public sector officials. Lateral entries or dedicated programs that allow entries for a limited amount of time can be effective methods for attracting these specialists into government offices. Furthermore, adapting job-classification schemes to include machine-learning-related job categories and increasing salaries and career prospects to better compete with comparable private sector job placements are advisable strategies. Ultimately, it is important to raise awareness among the target talent group about the motivating challenges (for example, social impact) and rewarding benefits (for example, job stability and work-life balance) of public sector work.[6]

Once in government, machine-learning experts' work can benefit greatly from exchange and knowledge sharing with colleagues. Establishing so-called communities of practice to cross knowledge boundaries within and across agencies can help gain legitimacy in relation to relevant stakeholders and foster collaboration among different agencies. Including nontech colleagues in these communities can also ensure that machine-learning applications are developed in a user-friendly manner and integrate well into the daily activities of public servants.[7] Another often-applied practice is the establishment of excellence centers that offer research, support, and training services and help agencies stay on the cutting edge of machine-learning technology. Finally, open communication, such as through blogs or dedicated events, can help other departments take notice and learn from each other's experiences.

Collaborations with external experts and research institutions can be another effective approach to bringing external expertise into a specific project while maintaining control and monitoring quality. Besides concrete project collaborations, establishing academic partnerships, like mobility or internship programs for the temporary assignment of personnel between government agencies and universities or research centers, can help institutionalize such collaborative efforts. Also, tailoring the machine-learning-related educational offerings of partner academic institutions to the particular needs of government organizations can be a viable way to ensure an inflow of machine-learning talent. Finally, building and sustaining intersectoral and interdisciplinary networking initiatives focused on the use of machine learning in government can help establish collaborations and foster learning and exchange.

Ethical Considerations for the Deployment of Machine Learning

Ethical considerations should be at the forefront of machine-learning deployment in public administration, and of analytical applications more broadly.[8] The social contract between governments and citizens differs substantially from the one that private sector companies have with their customers. Citizens or civil servants rarely have a choice about whether to share their data with the government. This makes data security and privacy particularly sensitive because most machine-learning applications require substantial amounts of data for appropriate training. On top of more general regulations, like the GDPR, ensuring the responsible usage and sharing of data, potentially by applying adequate anonymization techniques, should be a priority for governments to ensure citizens' trust (for a more extensive discussion, see chapter 28).

Another factor that can inhibit citizens' trust stems from the rare position of governments in relation to machine-learning technologies. Governments unify the roles of user and regulator in a single entity. This makes the public sector's use of machine learning a particularly delicate target of public scrutiny. Cases in which government machine-learning systems violate citizens' rights, like the recent case of the Dutch automated surveillance system for detecting welfare fraud, pose serious threats to citizens' trust (see box 16.1). It is therefore necessary to faithfully encode legal and political choices and ensure compliance with international regulatory frameworks to ensure ethical machine-learning applications in the public sector.

Applications of machine learning in government must consider that citizens often rely only on the government for public services like social security. This is a particular challenge to using machine learning in settings where the algorithm must make a choice. For instance, regarding social security systems, an algorithm might decide whether a citizen is eligible for a particular government benefit. In this situation, the algorithm would have to compare what would happen if the citizen were granted the benefit versus if the citizen were not. The algorithms that underlie this decision-making have to make assumptions about what would happen in each scenario, and the usefulness of the final decision depends on how appropriate these underlying assumptions were. If a citizen were denied a benefit due to an algorithm's decision, who would hold the algorithm accountable?

BOX 16.1 The Precedent Case of the Netherlands' Automated Surveillance System

On February 5, 2020, the District Court of The Hague ruled that SyRI (Systeem Risico Indicatie), a machine-learning application used by the government of the Netherlands to detect welfare fraud, violated Article 8 of the European Convention on Human Rights (ECHR)—that is, the right to respect for private and family life. This case is one of the first times a court has stopped a government's use of machine-learning technologies on human rights grounds and is thus considered to offer an important legal precedent for other courts to follow.

SyRI was designed to prevent and combat fraud in areas such as social benefits, allowances, and taxes by linking and analyzing data from various government and public agencies and generating personal risk profiles. It was deployed by the Minister of Social Affairs and Employment upon the request of various public agencies, including, among others, municipalities, the Social Insurance Bank, and the Employee Insurance Agency. The system mainly used a neighborhood-oriented approach, meaning it targeted specific neighborhoods where the linked data indicated an increased risk of welfare fraud.

Although the Court agreed with the government of the Netherlands that the fight against fraud is crucial and thus that employing novel technologies offering more possibilities to prevent and combat fraud generally serves a legitimate purpose, it ruled that the way SyRI was operated did not strike a "fair balance" between social interests and violation of the private lives of citizens, as required by the ECHR. In particular, the Court stated that due to the lack of insight into the risk indicators and the operation of the risk model, the system had violated the transparency principle and that discrimination or stigmatization of citizens in problem areas could not be ruled out.

The ruling, which led to the immediate halt of SyRI and caused public uproar far beyond the Netherlands, is a telling example of the potential negative consequences of applying machine-learning systems for government purposes without adequately addressing their potential ethically adverse side effects.

Often, modeling assumptions are not directly testable and hence require a substantial level of expertise over both what assumptions the algorithm is making and the suitability of those assumptions for a given public sector setting (Athey 2017). Public policy making through machine learning therefore raises important ethical questions. Choices may be made on behalf of government officials about citizen outcomes by machines they do not fully understand. For this reason, there is tension between the use of machine-learning technology to improve public administration and the oversight required to ensure that its use is in accordance with ethical principles. This tension becomes particularly salient when the use of previous administrative data for algorithm training introduces human biases into the system. Not uncovered, these biases can lead to "discrimination at scale" in sensitive areas such as racial profiling.

Finally, most applications of machine learning for government purposes are not static and should be adapted to evolving understandings of ethical principles. For example, algorithms for detecting fraud need to be constantly updated or retrained to address new forms of misconduct uncovered by agency employees and avoid an excessive focus on past forms of misconduct. Without such updating, algorithms may be biased toward past versions of criminal conduct. Constant updating by consulting domain experts and ethical advisors is necessary to ensure the effectiveness and ethical compliance of machine-learning technologies in government.

MACHINE LEARNING FOR JUSTICE

We now turn our focus to applications of machine learning in the justice system. The justice system is an institutional setting with high-frequency data, well-documented cases, extensive textual evidence, and a host of legal actors. As such, it is a useful setting within which to explore the use of machine learning for administration in the public service. An example of a core analytical question in justice is how the characteristics of judges impact judicial outcomes, such as rulings. This is a formulation of a wider question about how the individual characteristics of public officials impact the quality of public services provided by the government. It is a question that the analytics of public administration can investigate with the right measurement, data infrastructure, and skills for analysis.

Significant progress has been made in answering this question using machine learning (Chen 2019a, 2019b; Rigano 2018). In the United States, machine learning is already used in processing bail applications, DNA analysis of crimes, gunshot detection, and crime forecasting (Epps and Warren 2020; Rigano 2018). The large volume of data from surveillance systems, digital payment platforms, newly computerized bureaucratic systems, and even social media platforms can be analyzed to detect anomalous activity, investigate potential criminal activity, and improve systems of justice. For example, after the January 6, 2021, riots at the US Capitol, investigators used machine-learning-powered facial-detection technologies to identify participants and initiate prosecutions (Harwell and Timberg 2021). Machine-learning systems can also reduce the barriers to accessing courts by providing users with timely information directly, rather than through lawyers. Sadka, Seira, and Woodruff (2017, 50) find that providing information to litigants in mediation reduces the overconfidence of litigants and nearly doubles the overall settlement rate, but this only occurs when litigants are informed directly rather than through their lawyers.

The application of machine-learning systems to justice systems is useful because slight tendencies in human behavior can have significant impacts on judicial outcomes. A growing body of work demonstrates how small external factors, most of which participants are unaware of, can leave their mark on the outcomes of legal cases. Analysis of courts in the US, France, Israel, the United Kingdom, and Chile, for example, has found that in various settings, the tone of the words used in the first 3 minutes of a hearing, the incidence of birthdays, the outcomes of sporting events, and even the time of day of a hearing or a defendant's name can affect the outcome of a case (Chen 2019a). An analysis of 18,686 judicial rulings by the 12 US circuit courts (also known as courts of appeals or federal appellate courts), collected over 77 years, illustrates that judges demonstrate considerable bias before national elections (Berdejó and Chen 2017). Similarly, there is new evidence on sequencing matters in high-stakes decisions: decisions made on previous cases affect the outcomes of current cases even though the cases have nothing to do with each other. For instance, refugee asylum judges are two percentage points more likely to deny asylum to refugees if their previous decision granted asylum (Chen, Moskowitz, and Shue 2016).

Given the abundant evidence of how bias shapes decisions made by officials in the justice system, machine-learning methods can identify these sources of bias and signal when they shape judicial outcomes. The subtlety of different forms of bias requires an approach that searches through a very large number of relationships to detect their wider effects, an approach for which machine learning may be well suited. This can result in a more streamlined system and a reduction in backlog. Such tools can identify discrimination and bias even when these are not evident to the participants in the courts themselves, thereby strengthening the credibility of the judiciary (Bhushan 2009; Galanter 1984; Kannabiran 2009). Moreover, as large backlogs of cases are a significant problem for the efficiency of the judiciary in developing countries, interest is growing in performance metrics that will improve the functioning of the judiciary.

The adoption of machine-learning systems, however, is not an easy-to-implement solution, in particular for the justice system. Despite the growing availability of judicial data, it is first necessary to process these data in a way that is amenable to machine learning. This requires the integration of data from different sources, the processing of textual data into quantifiable metrics, and the definition of indicators for learning tasks that reflect either performance objectives or the operationalization of the concepts of fairness and impartiality. This is not an easy task: it requires substantial investments in data infrastructure and human capital, as well as building a conceptual framework. Therefore, to implement machine-learning algorithms successfully, justice systems need to acquire and train teams of machine-learning engineers, subject matter experts, and legal actors to develop machine-learning algorithms that are operationally relevant. These considerations are similar to the ones highlighted in the broader consideration of public administration.

Finally, there are ethical concerns regarding machine-learning applications for judicial outcomes. Practitioners and citizens may raise questions regarding the interpretability of algorithms because technological sophistication creates a "black box" problem: increases in technology's sophistication make its operation less interpretable (Pasquale 2015). The challenge of interpretability also raises concerns about accountability and oversight for these systems. Furthermore, the gap between those who can and those who cannot access and understand these technologies exacerbates existing social divisions and intensifies polarization. For all these reasons, machine-learning tools should be seen as complements to rather than substitutes for human decision-making, in particular for institutions that make life-altering decisions, such as the judiciary.

Judicial Data Infrastructure for Machine Learning

It is increasingly recognized that "the world's most valuable resource is no longer oil, but data" (*Economist* 2017). Like oil, raw data are not valuable in and of themselves; their value arises when they are cleaned, refined, processed, and connected to other databases that allow for the generation of insights that inform decision-making. This is particularly true in the field of machine learning, which requires large amounts of data to build accurate predictive models that provide information on the process, behaviors, and results of any indicators of interest.

Judiciaries collect vast amounts of data daily. Despite the availability of data, judicial data have rarely been analyzed quantitatively. In recent years, the transition from paper to e-filing and case management systems has facilitated the systematic analysis of massive amounts of data, generating performance metrics that can be used to evaluate courts and justice actors. Furthermore, with advances in machine learning, natural language processing (NLP), and processing power, these data create valuable opportunities to apply machine-learning models to evaluate and improve justice systems (Ramos-Maqueda and Chen 2021). Nonetheless, the extent to which countries can utilize novel approaches in machine learning and data analytics will depend on the available data infrastructure. The question, then, is which data do (and should) judiciaries collect?

In the justice domain, an integrated justice system brings together data on each case and connects these data with information on the actions and decisions at each case milestone. For instance, this includes information on the case filing, initial decisions, hearings, rulings, and sentences for each case. These data should also relate to potential appeals in order to help understand the evolution of the case. By implementing NLP on the text of case filings or judicial decisions, judiciaries can automate the revision of case filings or identify relevant jurisprudence for judges and court actors, for instance. Beyond information on the justice process itself, judiciaries will gain valuable insights from connecting these data with other information, such as human resources data, information on recruitment, or data from judicial training, to understand how best to select, train, and motivate judges depending on their background and experience.

To evaluate the impact of machine-learning interventions in justice, judiciaries should ideally collect information from other agencies involved in the justice process, as well as the economic outcomes of the parties involved. In criminal justice, an interoperable data ecosystem will connect judicial data with data from the prosecutor's office, the police, and the prisons, which will enable judiciaries to understand where the case has come from and the implications of judicial sentences. In civil cases, this may include economic data on citizens and firms who participate in the justice system, such as tax data, social insurance data, or procurement data. This way, judiciaries will be able to evaluate the impact of machine-learning applications not only on the judicial process but also on the lives of citizens and the financial status of firms that use the justice system.

In addition to the external and internal databases, it is also recommended that judiciaries carry out surveys that complement administrative information with the experience of the parties and employees involved in the justice system. Administrative data will not capture important elements of user or employee satisfaction, for instance, so survey data are a necessary complement for understanding the impacts of any new machine-learning models. We also recommend surveying those who are not necessarily part of the justice system—but who could be potential users in the future—through legal needs assessments.

Finally, there are additional complexities to developing data ecosystems for machine learning. Data must be of high quality, and large volumes need to be collected and stored to make them amenable to artificial intelligence (AI) algorithms, which, in general, presuppose big data. This requires data extraction, transformation, and loading (ETL) processes designed to support AI pipelines and, in most use cases, dedicated data engineers to maintain them.

Human Capital Requirements for Machine Learning in Justice

Justice systems often have limited in-house access to the necessary human capital to develop machine-learning applications. Judicial officers are rarely experts on data analysis—which is seldom part of their training—and machine-learning engineers generally lack the domain expertise necessary to understand the functioning of the law. In courts without sufficient human capital to take advantage of available data, training public servants in even simple data analysis skills may be a valuable long-term investment for improving the functioning of courts. Nevertheless, the development of machine-learning approaches may remain out of reach for public officials whose training does not include statistical modeling or data engineering.

An alternative approach is relying on nongovernmental organizations, international organizations, or even private companies to develop machine-learning applications. An example of this approach is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) tool in the United States, which is an algorithm that generates recidivism risk scores to aid judges in their rulings. However, judiciaries should consider the long-term sustainability of an outsourced solution as well as ethical concerns. COMPAS has been the target of controversy due to its proprietary algorithm and the inability of public officials and citizens to understand how it operates under the hood.[9] Additionally, a reliance on external contractors often substitutes for in-house development of the necessary human capital to develop machine-learning technologies, reproducing external reliance on nonjudicial actors for both the maintenance and expansion of machine-learning solutions.

Whether in-house or externally sourced, the forms of human capital required for machine-learning applications are diverse and costly. The implementation team should include machine-learning engineers, software developers to code the user interface, data engineers to develop the data infrastructure, legal experts, and project managers to communicate the judiciary's needs to the implementation team. Because each component of the project relies on the others—there can be no user interface without a data infrastructure to feed it—the team should ensure that their timelines are aligned. A sufficient budget should be allocated to the project to cover the team's time for both the implementation and the monitoring of the technology after the first version of the application is developed.

Ethical Concerns

Developers of machine-learning applications should carefully consider the ethical implications of their use by judicial actors. Only technologies that aid human decision-making, rather than replacing it, should be adopted in the courts. There are multiple reasons for this recommendation. As noted earlier, algorithms have the "black-box" problem of interpretability—that is, it is not easy to trace the output of complex algorithms to the data inputs themselves. Additionally, biases in the decisions of judicial actors are reflected in the algorithm's training data and may be encoded into the algorithm itself. Thus, using machine-learning algorithms to inform judicial decisions without critical oversight raises the risk of replicating these biases elsewhere in the system. The inherent choice of performance metrics can also reinforce existing biases by decision-makers within the system. Addressing these issues requires a participatory and deliberative approach to the design, implementation, and evaluation of machine-learning technologies.

A reasonable demand to guarantee trust and fairness is that algorithms be interpretable. A judge may request the reason why a particular decision has been recommended by the algorithm. This transparency enhances judges' trust in the technology and allows for disagreement with its recommendations. Given the complexities of working with machine-learning algorithms, any rollout must be preceded by a phase of comprehensive study and rigorous testing of the systems themselves. Randomized controlled trials that carefully estimate the causal impacts of the adoption of these algorithms to properly evaluate their costs and benefits are essential. A carefully constructed trial can provide important benchmarks on cost, efficiency, user satisfaction, and impact on key performance metrics, all essential for a justice system to credibly serve citizens.

CASE STUDY: MACHINE LEARNING FOR JUSTICE SYSTEMS IN INDIA

This case study illustrates how machine learning was implemented in the national justice system of India by the Data and Evidence for Justice Reform (DE JURE) team. Due to India's large population and volume of cases, justice officials are often unable to effectively manage cases in a timely fashion. India has just 19 judges per million people and 27 million (2.7 crore) pending cases (Amirapu 2020; Chemin 2012; Kumar 2012). To address this, the Indian justice system has made considerable advances in the adoption of information technology and has released large volumes of data to court users and encouraged them to use electronic systems. Yet legislative, institutional, and resource constraints have limited the full impact of these advances (Amirapu 2020; Damle and Anand 2020).

In this section, we describe how the DE JURE team implemented machine-learning applications in India. We first highlight the data infrastructure requirements for implementing machine-learning applications, as well as how these applications could enhance the functioning of the justice system.

Judicial Data Infrastructure in India

In the past 15 years, considerable efforts have been made to adopt and deploy information technology systems in the courts of India. One of the most significant projects, the e-Courts project, was first launched in 2005 by the Supreme Court of India through the National Policy and Action Plan for Implementation of Information and Communication Technology (ICT) in the Indian Judiciary. The e-Courts initiative introduced technology into Indian courts in a variety of ways.

The most obvious feature of the system was the deployment of technology within courtrooms. Judges were provided with LCD touch screens, screens and projectors were connected via a local network to disseminate information to lawyers, and electronic boards at the courts displayed the queue of case numbers scheduled for hearing on a particular day. Outside the courtroom, e-filing procedures were established, and a data management architecture was created that ranged from the scanning of old cases into the electronic system to the creation of digital archives. The ICT plan also established direct electronic communication with litigants and an online case management system.

These investments eventually paved the way for the creation of the National Judicial Data Grid (NJDG), a database of 27 million cases that allows court users to view the status of pending cases and access information on past hearings. For the DE JURE team's goal to implement machine-learning tools, the most significant resources were the digital archives of cases. The team was able to scrape these publicly available digital archives to construct an e-Courts district court data set of 83 million cases from 3,289 court establishments.[10] They were able to curate details, like the act under which the case was filed, the case type (criminal or civil), the district where it originated, the parties to the case, and the history of case hearings, in a manner that made the data amenable to large-scale analysis.

A wider data ecosystem has been created by joining additional sources to the case data, including the following:

- **Data on judges:** To better understand the impact of specific judges—their identity, training, and experience—the team constructed a database of judges for the courts of India. They began this task by extracting data from editions of *The Handbook on Judges of the Supreme Court of India and the High Courts*, released by the Supreme Court of India, and appending to it information from various High Court websites. So far, the team has assembled detailed information for 2,239 judges from the handbooks for the years 2014–20. Most notably, 93.5 percent of these judges are men and 6.5 percent are women, and their range of experience covers a period spanning approximately 70 years.
- **Database of Central Acts:** This auxiliary data set is intended to give a definitive list of standardized act names. This could then be used to standardize the act names appearing in various cases. This would allow the team to analyze all cases filed under a given act. The team has, for example, examined all cases related to the Water Act of 1974 and found a total of 978 such cases at the Supreme Court and High Courts of India. The list of central (federal) acts can be viewed on the Legislative Department website of the Ministry of Law and Justice. There is currently no centralized source for all state legislation: this needs to be obtained from state websites separately.
- **Other administrative data:** Data on other institutions can be linked to the judicial data at the district as well as the state level. For example, data on Indian banks and their branches are available through the Reserve Bank of India. This database contains information on names, locations, license numbers, license dates, addresses, and other unique identifiers. The team has scraped, cleaned, and organized these data for further analysis. The database contains about 160,000 entries. The unique identifiers and location information allow the team to merge these data with banks appearing in litigation in courts that are present in the e-Courts databases. Merging these data with the legal data allows the team to examine a variety of interesting questions about the development of financial markets in an area, participation in the justice system, and the impacts of legal rulings.

The quality of these data varies significantly: there is no nationally standardized system for defining variables or reporting on them. For instance, in some states, the legal act name and section numbers are well delineated, but in other states, this is not the case. This makes it difficult to compare individual case types across courts and states (Damle and Anand 2020). There are no standardized identifiers within the data to follow a case through its potential sequence of appeals in higher courts. In a similar vein, there is no easy way to track a criminal case from its entry into the system as a first information report (FIR) to its exit as a judgment. There are inconsistencies in identifying information about participants, their attributes, and the types of laws or acts that cases relate to. There are also issues of incorrect reporting and spelling mistakes.

Machine-Learning Applications in the Courts of India

The quality of data in India's justice system is often compromised: case data are incomplete or litigants' identities are not registered. The DE JURE team has constructed a robust data pipeline to collect often-incomplete judicial data, as well as machine-learning tools to clean and prepare them for analysis. In this section, we contextualize the problem: how data quality issues in judicial data manifest themselves in India. In the following section, we describe the solution: how machine-learning tools have been designed to enhance the quality of judicial data for analysis.

Legal data released by the Indian judiciary are voluminous, messy, and complex (Damle and Anand 2020). The typical case has clear tags for some key dates (filing date), the key actors (petitioner, respondent, and judges) and the court name, but information about the type of case, the outcome of deliberations, and pertinent acts cited is often not clearly identifiable in the textual body of the judgment. Cleaning and preprocessing the data is critical for any form of analysis, especially for supervised algorithms trained on these data. Traditional empirical legal studies have typically addressed this issue by relying on small-scale data sets in which legal variables are manually coded and the scope of inference is related to a small body of legal cases pertinent to a single issue (Baxi 1986; Gauri 2009; Krishnaswamy, Sivakumar, and Bail 2014).

These traditional approaches are unable to keep up with the incoming volume of cases. In this context, machine-learning tools provide an alternate approach to detecting errors or gaps in the data and correcting them in an automated fashion. Using machine learning, it is possible to infer the identities of participants even when these data are not registered. Additionally, laws used as precedents for a ruling can be identified through text analysis. Beyond the data quality itself, machine-learning approaches can help identify biases and discrimination in judges' rulings.

Inference about the Identities of Participants

Some databases of judgments provide no identifying information about participants in the cases themselves. To better understand who participates in the courts, the team first extracts litigant names from the raw text of the judgments and then uses matching algorithms to identify the type of litigant (individual, company, or state institution). Classifying participants can be challenging. If the identification exercise involves government agencies, it is first necessary to compile all the different agencies of the state and national governments. Manually tagging entities is prohibitively time-consuming, but the existence of latent patterns in the names makes this fertile ground for machine-learning applications.

The machine-learning application relies on similarity across names for participants that belong to the same "group" to classify a particular name as belonging to that group. Using prelabeled data—individual name A belongs to group B—the machine-learning algorithm can extrapolate to unlabeled data, where an individual's name is available but not their group. Some obvious groups of interest are gender, caste, and religion, which are not recorded in judicial data but are available in other data sources. Another group of interest may be whether a participant is a government agency or not. We focus here on people's first and last names, for illustration.

The team first formats individuals' names to ensure that each individual can be identified by an honorific title, a first name, and a last name. Honorifics, such as Shri, Sri, Smt., Mr., Mrs., and Ms., enable the algorithm to directly identify an individual's gender. To extend this classification to names without an honorific, the team trains an algorithm on a publicly available corpus of labeled common Indian first names. Training this algorithm, often referred to as training a classifier, is the process by which the algorithm learns patterns within the data related to a group. Here, these patterns are the statistics of co-occurrence of letters in names, the lengths of names, and other features that allow the algorithm to determine whether a name indeed belongs to a particular group: in this case, a gender.[11]

These algorithms formalize intuitive notions of why a name belongs to a given group by identifying frequently occurring patterns within names associated with that group. Caste assignment is more complicated because the same last name could be associated with multiple caste groups. The name "Kumar," for example, could belong to a person belonging to the Scheduled Castes, the Scheduled Tribes, or the category "Other."

In the case of such names, the team generates distributions of the last name across the different caste categories. They then use this distribution to generate a prediction and combine this with the predictions of other models to ensure a robust prediction. They assign a caste to each household based on a simple majority vote among these models.

Identification of Laws and Acts

Legal texts in India's justice system do not currently employ a standardized citation style for referring to acts or laws. For example, the Hindu Marriage Act may be referred to in a variety of ways, such as "u/s 13 clause 2 of Hindu Marriage Act," "u/s 13(b) Hindu Marriage Act," or "u/s 13 of Hindu Marriage Act 1995." Again, machine-learning tools can be used to address this issue.

In this project, the DE JURE team uses a set of tools that create mathematical representations of the text in the form of vectors. Term frequency–inverse document frequency (TF–IDF) is one popular method for representing a string of words as a numerical score that reflects how frequently a word is used within a given text and how infrequently it appears in the corpus. Applying this to act names, the team uses different clustering algorithms to group particular act citations based on how similar they are numerically. This approach groups the underlying act-name data in a manner that best preserves the coherence within groups (a particular act name) and the distance across groups to make the classification.

The identification of specific acts and how often they are cited opens new opportunities for legal analysis. The team can, for example, compare the different types of acts that are cited in the different courts within India's justice system. It can allow researchers and practitioners to identify the real-time evolution of legal citation—and legal thought—as judges refer to these acts.

Using Descriptive Analysis and Machine Learning to Identify Discrimination and Bias

Bias and discrimination can occur in different areas of policy making, particularly when civil servants exercise discretion, such as in judicial rulings. Judges may favor or deny due process to plaintiffs depending on their ethnic or gender identity, undermining the rule of law and the right to impartial judgment. This challenge is, of course, not unique to judiciaries. A broad academic literature has demonstrated that bias in a human decision-maker can have conscious as well as unconscious drivers and may manifest in complex ways and in a variety of contexts that can be difficult to prove (Banerjee et al. 2009; Bertrand and Mullainathan 2004; Ewens, Tomlin, and Wang 2014; Kleinberg et al. 2018). In other settings, such as labor markets and educational institutions, algorithms—rules that take "inputs" (like the characteristics of a job applicant) and predict some outcome (like a person's salary)—have been shown to create new forms of transparency and to serve as methods for detecting discrimination (Kleinberg et al. 2020; LeCun, Bengio, and Hinton 2015).

Machine-learning algorithms can identify these biases and forms of discrimination, enabling governments to measure their prevalence and design policy changes to reduce them. In the courts of India, algorithms could help judges make critical decisions about cases (for example, dismissals or bail applications) and reduce bias in their rulings. Building such algorithms requires a rich data set that includes litigant characteristics (caste and gender), lawyer characteristics, court characteristics, case details (filing details and evidence provided), and case outcomes (such as the granting of bail or the dismissal of a case). A machine-learning engineer would develop a "learning procedure" that would aim to provide a predicted outcome from a broad range of inputs and modeling approaches, such as a neural network (Dayhoff 1990). These models differ from traditional statistical methods, such as linear regression, which are more deductive (presuming a linear fit between a few sets of variables) than inductive (allowing the data to report the best fit between a large set of variables).

These insights could be invaluable not only within the courtroom itself but also in judicial education. Experiments are currently underway, in the Judicial Academy of Peru, for example, to assess methods to improve case-based teaching by using the history of a judge's past decisions, which can reveal potential bias or error. The data are also suitable for creating personalized dashboards and interfaces that provide judges, mediators, and other decision-makers with real-time information about their own performance relative to their own previous decisions and those of others who are comparable (Kling 2006). This information can be used to augment the capabilities of judges and lawyers, increase their productivity, and reduce potential bias in their decisions.

BEYOND DESCRIPTIVE ANALYSIS: IMPACT EVALUATION IN THE JUSTICE SYSTEM

Moving beyond descriptive analysis and more correlational analysis of data, an underexplored field in the justice system is policy experiments for impact evaluation. Legal scholars and judges have long debated the merits of implementing various laws and regulations and have justified their arguments with theories about the effects of these legal rules. This situation resembles the field of medicine a century ago: before clinical trials, medical research focused on theoretical debates rather than rigorous causal evidence.

A growing body of empirical research now demonstrates that causal inference is possible in judicial studies. For example, in situations where cases are randomly assigned to judges, the random assignment itself can be used as an exogenous source of variation to evaluate the impact of judicial decisions. Since judges do not choose their cases, observed rulings reflect the judge's personal characteristics (ideological preferences or biases) and features of the case rather than the judicial process as a whole. Additionally, informational treatments can have an impact on the behavior of judges, improving their performance (box 16.2).

BOX 16.2 Leveraging Data and Technology to Improve Judicial Efficiency in Kenya

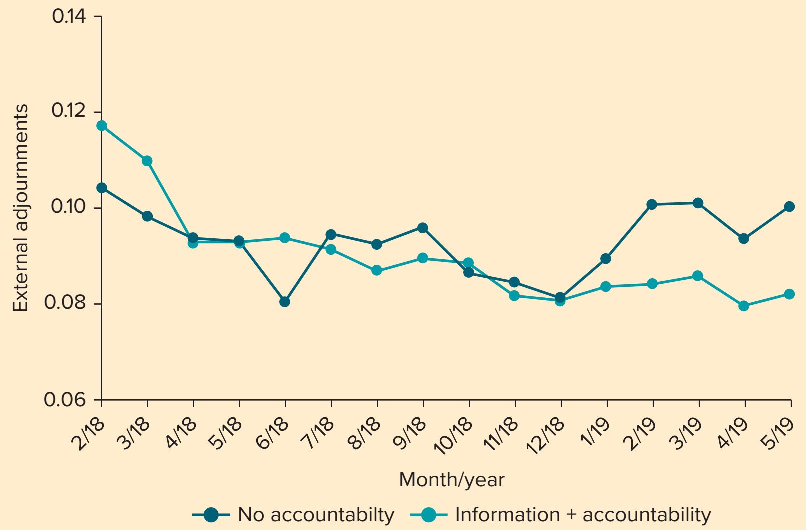
In partnership with the Judiciary of Kenya and McGill University, the World Bank's Development Impact Evaluation (DIME) Data and Evidence for Justice Reform (DE JURE) team has been leveraging underutilized administrative data to improve judicial efficiency. Through its case management system, the Judiciary of Kenya collects large amounts of administrative data on the characteristics of cases, the dates of hearings and reasons for adjournments, and other important metrics of court performance. These data are readily available for understanding and designing interventions to address challenges to the efficient delivery of justice, such as adjournments of hearings, which cause large delays in court proceedings. Despite the richness of these data, they were not being used for decision-making. DIME and the Judiciary of Kenya decided to leverage these data systems to design an algorithm identifying the greatest sources of delay for each court and presenting recommended actions. The team included performance information in a one-page feedback report. Then it studied whether this simplified, action-oriented information could reduce adjournments and improve judicial performance.

In a randomized controlled trial across all 124 court stations in Kenya, the team compared the impact of sharing the one-page feedback reports only with judges and supervisors to the impact of sharing them with Court User Committees as well, the latter acting as an additional accountability mechanism (figure B16.2.1). The team found that the one-page feedback report with the accountability mechanism reduced the number of adjournments by 20 percent over a four-month period and increased the number of cases resolved (Chemin et al. 2022). The conclusion was that the report was more effective when both tribunals and Court User Committees received it. Thus, sharing performance information with courts may be effective to improve efficiency, but it is particularly effective when this information is also shared with civil society and court stakeholders. This study served as proof of concept that utilizing data to provide information to judicial actors can reduce adjournments and increase the speed of justice, which have a downstream impact on the economic outcomes of citizens and firms.

(continues on next page)

BOX 16.2 Leveraging Data and Technology to Improve Judicial Efficiency in Kenya *(continued)*

FIGURE B16.2.1 Impact of One-Page Feedback Reports on Case Delays in Kenya



Source: DE JURE, World Bank.

Randomly assigning cases to judges predicted to be harsh or lenient allows researchers to identify the long-run causal impacts of the length of sentences (Dobbie and Song 2015). To identify the causal effect of a sentence length of eight months or eight years, a randomized controlled trial would need to randomize the sentence, which is impossible. However, assigning a defendant to a judge predicted to assign an eight-month sentence or to another judge predicted to assign an eight-year sentence allows researchers to identify the causal impact of sentence length on subsequent life outcomes.

The same conceptual framework can examine the causal effects of debt relief on individuals' earnings, employment, and mortality (Sampat and Williams 2019). This causal approach sheds light on the impact these judicial decisions can have on individuals' welfare outcomes. By applying machine learning to infer the bias, lenience, and ideological preference of a judge, researchers can identify the causal effect of these variables on the judicial system and the life outcomes of those affected by judges' decisions.

CONCLUSION

In this chapter, we argue that machine learning is a powerful tool for improving public administration in general and the justice system in particular. Machine learning, at its core, emphasizes a methodological approach centered around a learning problem: defining indicators and using evidence to improve them. Under the umbrella of this methodological approach, multiple applications are available to tackle key issues in public administration. Algorithms can be written to draw inferences about the identities of participants and study the deliberative processes they employ within courtrooms. Machine-learning tools can also convert a high volume of textual data to numerical estimates that can be used for understanding the processes and outcomes of different types of case data, including public procurement, taxes, and the systems of justice themselves.

These tools, however, have several limitations and requirements that need to be addressed before they can be effectively deployed in the courts. At the very outset, there are significant issues related to the privacy of personally identifiable information, security, and the control of legal data. Next, the algorithms require data preprocessing, training on large, high-frequency data sets, and iterative refinement with respect to the actual use cases in which they are deployed. This requires strong pilot programs that are studied as part of randomized controlled trials. Insights on data privacy and costs as well as outcomes require that these pilots be constructed on a reasonable scale.

Public administration officials often execute a range of tasks, from the ordinary to the complex, such as the adjudication of a trial. The smart application of machine learning can enhance levels of automation, productivity, and the level of information extracted from data generated in the public sector. If done right, it can help reduce noise—and this can be one step toward aiding the impersonal execution of tasks, reducing bias, enhancing predictability, and making decision rules more transparent. But none of these outcomes can be presupposed from the machine-learning approach: they depend on the ethical framework and operational relevance underlying its implementation. Machine-learning practitioners are therefore advised to take necessary precautions and develop solutions that are accountable to the public and useful for government officials.

NOTES

1. Machine learning is therefore a methodological approach anchored in a learning framework. It is not a radical departure from classical approaches to statistics and, in fact, often builds on canonical models (such as linear or logistic regressions), nor is it exempt from well-known challenges, such as bias and model misspecification.
2. The machine-learning application is part of a broader study on bureaucratic allocation available in Bandiera et al. (2021).
3. For a discussion of causality in machine learning, see Schölkopf (2022).
4. Public servant surveys, due to their smaller sample frames and limited applications in prediction, are rarely used directly for machine-learning applications.
5. Federal transfers have been analyzed using machine learning since 2018 (CGU 2018).
6. The Inter-American Development Bank has written extensively on the topic. See chapter 3 of Porrúa et al. (2021).
7. For a concrete example of this approach, see case study 2 in chapter 9 of the *Handbook* on the Brazilian federal government's experience with the development of machine learning for payroll analysis).
8. For an overview, see chapter 6 of the *Handbook*.
9. For a discussion, see Yong (2018).
10. The e-Courts data are public and can be accessed via the district court websites, the e-Courts Android/iOS app, or the district court services webpage at https://services.ecourts.gov.in/ecourtindia_v6/.
11. To reduce model overfitting, we use the majority vote from multiple trained classifiers, including a logistic regression model and a random forest classifier to make predictions on gender. A logistic regression models the probability of a binary outcome or event. A random forest classifier will use decision trees (nested if-then statements) on features of the data to make the prediction. We have also made predictions of religion and caste using similar approaches. Muslims can be recognized in the data through the distinctiveness of Muslim names: common names such as Khan and Ahmed can easily be assigned and coded, but for others, we utilize the occurrence of specific letters (such as *q* and *z*) through appropriate classifiers to identify additional names.

REFERENCES

Amirapu, Amrit. 2020. "Justice Delayed Is Growth Denied: The Effect of Slow Courts on Relationship-Specific Industries in India." *Economic Development and Cultural Change* 70 (1): 415–51. https://doi.org/10.1086/711171.

Athey, Susan. 2017. "Beyond Prediction: Using Big Data for Policy Problems." *Science* 355 (6324): 483–85. https://doi.org/10.1126/science.aal4321.

Bandiera, Oriana, Michael Carlos Best, Adnan Qadir Khan, and Andrea Prat. 2021. "The Allocation of Authority in Organizations: A Field Experiment with Bureaucrats." *The Quarterly Journal of Economics* 136 (4): 2195–242. https://doi.org/10.1093/qje/qjab029.

Banerjee, Abhijit, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan. 2009. "Labor Market Discrimination in Delhi: Evidence from a Field Experiment." *Journal of Comparative Economics* 37 (1): 14–27. https://doi.org/10.1016/j.jce.2008.09.002.

Baxi, Upendra. 1986. *Towards a Sociology of Indian Law*. New Delhi: Satvahan.

Berdejó, Carlos, and Daniel L. Chen. 2017. "Electoral Cycles among US Courts of Appeals Judges." *The Journal of Law and Economics* 60 (3): 479–96. https://doi.org/10.1086/696237.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013. https://doi.org/10.1257/0002828042002561.

Bhushan, Prashant. 2009. "Misplaced Priorities and Class Bias of the Judiciary." *Economic and Political Weekly* 44 (14): 32–37. https://www.jstor.org/stable/40278698.

Brynjolfsson, Erik, and Andrew McAfee. 2017. "The Business of Artificial Intelligence." *Harvard Business Review*, July 18, 2017. https://hbr.org/2017/07/the-business-of-artificial-intelligence.

CGU (Controladoria-Geral da União). 2018. "Inteligência Artificial Analisará Prestação de contas em Transferências da União." Comptroller General of Brazil, October 23, 2018. https://www.gov.br/cgu/pt-br/assuntos/noticias/2018/10/inteligencia-artificial-analisara-prestacao-de-contas-em-transferencias-da-uniao.

Chemin, Matthieu. 2012. "Does Court Speed Shape Economic Activity? Evidence from a Court Reform in India." *The Journal of Law, Economics, & Organization* 28 (3): 460–85. https://doi.org/10.1093/jleo/ewq014.

Chemin, Matthieu, Daniel L. Chen, Vincenzo Di Maro, Paul Kimalu, Momanyi Mokaya, and Manuel Ramos-Maqueda. 2022. "Data Science for Justice: The Short-Term Effects of a Randomized Judicial Reform in Kenya." TSE Working Paper 22-1391, Toulouse School of Economics, Toulouse.

Chen, Daniel L. 2019a. "Judicial Analytics and the Great Transformation of American Law." *Artificial Intelligence and Law* 27: 15–42. https://doi.org/10.1007/s10506-018-9237-x.

Chen, Daniel L. 2019b. "Machine Learning and the Rule of Law." In *Law as Data: Computation, Text, and the Future of Legal Analysis*, edited by Michael A. Livermore and Daniel N. Rockmore, 433–41. Santa Fe, NM: Santa Fe Institute Press. https://doi.org/10.37911/9781947864085.16.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. "Decision Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires." *The Quarterly Journal of Economics* 131 (3): 1181–242. https://doi.org/10.1093/qje/qjw017.

Chen, Jonathan H., and Steven M. Asch. 2017. "Machine Learning and Prediction in Medicine—Beyond the Peak of Inflated Expectations." *New England Journal of Medicine* 376: 2507–09. https://doi.org/10.1056/NEJMp1702071.

Cross, Tim. 2020. "Artificial Intelligence and Its Limits: An Understanding of AI's Limitations Is Starting to Sink In." *Economist*, Technology Quarterly, June 13, 2020. https://www.economist.com/technology-quarterly/2020/06/11/an-understanding-of-ais-limitations-is-starting-to-sink-in.

Damle, Devendra, and Tushar Anand. 2020. "Problems with the e-Courts Data." National Institute of Public Finance and Policy Working Paper 314, National Institute of Public Finance and Policy, New Delhi, India. https://www.nipfp.org.in/media/medialibrary/2020/07/WP_314__2020.pdf.

Dayhoff, Judith E. 1990. *Neural Network Architectures: An Introduction*. New York: Van Nostrand Reinhold.

Devanesan, Joe. 2020. "AI-Powered Traffic Management Is Slashing Asia's Congestion Problem." *Techwire Asia*, August 28, 2020. https://techwireasia.com/2020/08/ai-powered-traffic-management-is-slashing-asias-congestion-problem.

Dobbie, Will, and Jae Song. 2015. "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection." *American Economic Review* 105 (3): 1272–311. https://doi.org/10.1257/aer.20130612.

Economist. 2017. "The World's Most Valuable Resource Is No Longer Oil, but Data." May 6, 2017. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.

Engstrom, David Freeman, Daniel E. Ho, Catherine M. Sharkey, and Mariano-Florentino Cuéllar. 2020. *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*. Washington, DC: Administrative Conference of the United States. https://www.acus.gov/research-projects/agency-use-artificial-intelligence.

Epps, Willie J. Jr., and Jonathan M. Warren. 2020. "Artificial Intelligence: Now Being Deployed in the Field of Law." *The Judges' Journal* 59 (1): 16–39. https://www.americanbar.org/groups/judicial/publications/judges_journal/2020/winter/artificial-intelligence-now-being-deployed-the-field-law/.

Ewens, Michael, Bryan Tomlin, and Liang Choon Wang. 2014. "Statistical Discrimination or Prejudice? A Large Sample Field Experiment." *The Review of Economics and Statistics* 96 (1): 119–34. http://www.jstor.org/stable/43554917.

Fazekas, Mihály, Salvatore Sberna, and Alberto Vannucci. 2021. "The Extra-Legal Governance of Corruption: Tracing the Organization of Corruption in Public Procurement." *Governance: An International Journal of Policy, Administration, and Institutions* 35 (4): 1139–61. https://doi.org/10.1111/gove.12648.

Galanter, Marc. 1984. *Competing Equalities: Law and the Backward Classes in India*. Oxford: Oxford University Press.

Gauri, Varun. 2009. "Public Interest Litigation in India: Overreaching or Underachieving?" Policy Research Working Paper 5109, World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-5109.

Harwell, Drew, and Craig Timberg. 2021 "How America's Surveillance Networks Helped the FBI Catch the Capitol Mob." *Washington Post*, April 2, 2021. https://www.washingtonpost.com/technology/2021/04/02/capitol-siege-arrests-technology-fbi-privacy/.

Jordan, M. I., and T. M. Mitchell. 2015. "Machine Learning: Trends, Perspectives, and Prospects." *Science* 349 (6245): 255–60. https://doi.org/10.1126/science.aaa8415.

Kannabiran, Kalpana. 2009. "Judicial Meanderings in Patriarchal Thickets: Litigating Sex Discrimination in India." *Economic and Political Weekly* 44 (44): 88–98. https://www.jstor.org/stable/25663738.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein. 2018. "Discrimination in the Age of Algorithms." *Journal of Legal Analysis* 10: 113–74. https://doi.org/10.1093/jla/laz001.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Cass R. Sunstein. 2020. "Algorithms as Discrimination Detectors." *Proceedings of the National Academy of Sciences* 117 (48): 30096–100. https://doi.org/10.1073/pnas.191279011.

Kling, Jeffrey R. 2006. "Incarceration Length, Employment, and Earnings." *American Economic Review* 96 (3): 863–76. https://www.jstor.org/stable/30034076.

Krishnaswamy, Sudhir, Sindhu K. Sivakumar, and Shishir Bail. 2014. "Legal and Judicial Reform in India: A Call for Systemic and Empirical Approaches." *Journal of National Law University Delhi* 2 (1): 1–25. https://doi.org/10.1177/22774017 20140101.

Kumar, Vandana Ajay. 2012. "Judicial Delays in India: Causes & Remedies." *Journal of Law, Policy & Globalization* 4: 16–21. https://www.iiste.org/Journals/index.php/JLPG/article/view/2069.

Lal, Niharika. 2021. "How Traffic Cameras Issue E-Challans." *Times of India*, April 17, 2021. https://timesofindia.indiatimes .com/city/delhi/how-traffic-cameras-issue-e-challans/articleshow/82103731.cms.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521: 436–44. https://doi.org/10.1038 /nature14539.

Misuraca, Gianluca, and Colin van Noordt. 2020. *AI Watch: Artificial Intelligence in Public Services*. EUR 30255 EN. Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/039619.

Mullainathan, Sendhil, and Ziad Obermeyer. 2022. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care." *The Quarterly Journal of Economics* 137.2 (May): 679–727. https://doi.org/10.1093/qje/qjab046.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Porrúa, Miguel, Mariano Lafuente, Edgardo Mosqueira, Benjamin Roseth, and Angela María Reyes, eds. 2021. *Digital Transformation and Public Employment: The Future of Government Work*. Washington, DC: Inter-American Development Bank. https://publications.iadb.org/publications/english/document/Digital-Transformation-and-Public-Employment-The -Future-of-Government-Work.pdf.

Ramos-Maqueda, Manuel, and Daniel L. Chen. 2021. "The Role of Justice in Development: The Data Revolution." Policy Research Working Paper 9720, World Bank, Washington, DC. https://openknowledge.worldbank.org/handle/10986/35891.

Rigano, Christopher. 2018. "Using Artificial Intelligence to Address Criminal Justice Needs." National Institute of Justice, October 8, 2018. https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs#citation--0.

Sadka, Joyce, Enrique Seira, and Christopher Woodruff. 2017. "Overconfidence and Settlement: Evidence from Mexican Labor Courts." Unpublished manuscript. http://www.enriqueseira.com/uploads/3/1/5/9/31599787/overconfidence_and _settlement_preliminary.pdf.

Sampat, Bhaven, and Heidi L. Williams. 2019. "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome." *American Economic Review* 109 (1): 203–36. https://doi.org/10.1257/aer.20151398.

Schölkopf, Bernhard. 2022. "Causality for Machine Learning." In *Probabilistic and Causal Inference: The Works of Judea Pearl*, edited by Hector Geffner, Rina Dechter, and Joseph Y. Halpern, 765–804. New York: Association for Computing Machinery. https://doi.org/10.1145/3501714.3501755.

Waldrop, M. Mitchell. 2018. "Free Agents." *Science* 360 (6385): 144–47. https://doi.org/10.1126/science.360.6385.144.

Yong, ed. 2018. "A Popular Algorithm Is No Better at Predicting Crimes Than Random People." *Atlantic*, January 17, 2018. https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/.