INSTITUT C.D. HOWE INSTITUTE

COMMENTARY

NO. 507

Regulatory Reform in Ontario: Machine Learning and Regulation

Regulators can better use big data and machine learning to save time and money for businesses, individuals and themselves, while addressing transparency and privacy concerns.

Anthony Niblett

The C.D. Howe Institute's Commitment to Quality, Independence and Nonpartisanship

About The Author

Anthony Niblett
is Associate Professor and Canada Research Chair in Law, Economics, & Innovation, University of Toronto.

The C.D. Howe Institute's reputation for quality, integrity and nonpartisanship is its chief asset.

Its books, Commentaries and E-Briefs undergo a rigorous two-stage review by internal staff, and by outside academics and independent experts. The Institute publishes only studies that meet its standards for analytical soundness, factual accuracy and policy relevance. It subjects its review and publication process to an annual audit by external experts.

As a registered Canadian charity, the C.D. Howe Institute accepts donations to further its mission from individuals, private and public organizations, and charitable foundations. It accepts no donation that stipulates a predetermined result or otherwise inhibits the independence of its staff and authors. The Institute requires that its authors publicly disclose any actual or potential conflicts of interest of which they are aware. Institute staff members are subject to a strict conflict of interest policy.

C.D. Howe Institute staff and authors provide policy research and commentary on a non-exclusive basis. No Institute publication or statement will endorse any political party, elected official or candidate for elected office. The Institute does not take corporate positions on policy matters.

Commentary No. 507
March 2018
Industry Regulation and Competition Policy

Daniel Schwanen
Vice President, Research

$12.00
isbn 978-1-987983-62-3
issn 0824-8001 (print);
issn 1703-0765 (online)

The Study In Brief

Government regulation of individual and business activity is part and parcel of modern society. But many businesses face difficulties in understanding and navigating the legal hurdles, rules, and uncertainty that come with modern regulation. Many governments in Canada have taken steps to reduce this burden by streamlining regulation and cutting unnecessary red tape.

In this *Commentary*, I explore how regulators can continue this trend toward more efficient and effective regulation: by embracing data analytics and machine-learning tools.

Big data, analytics and machine learning offer new and difficult challenges for regulators who oversee how many businesses make decisions. But regulators can also benefit from effective use of data science. Some of these benefits can be realized almost immediately by using data that the regulators already have.

First, regulators can better predict who should and should not be investigated. A regulator needs to make choices about how to allocate and prioritize scarce resources. With the right data and appropriate data analytics, predictions can be made about where to best place investigation resources.

Second, regulators must make choices over which cases to prosecute. Regulators should not waste resources litigating cases they are likely to lose. Instead, regulators should put resources only toward cases that they are likely to win. Regulators can turn to the data and use machine learning to predict how a court would resolve a particular problem.

Moving further into the future, big data and machine learning will change the way that laws and regulations will be consumed and produced. Lawmakers will have greater ability to provide relevant information *before* the individual or business acts, rather than waiting to adjudicate *after* they have acted. Businesses will seek prior authorization for many more regulated actions. Furthermore, the time and cost for regulators to respond to the queries will fall drastically. Instead of relying primarily on vague guidelines, regulators will be able to offer more expedient and personalized responses.

There are enormous benefits to regulators making decisions before individuals and business act. Advance rulings, given before investments are made, provide certain outcomes and reduce the likelihood of wasted investments. There are, of course, a number of potential barriers and issues that may arise. These include: the quality of the data, accountability and due process, the need for transparency, privacy and the reluctance to share data, the benefits of uncertainty, and the stability of social views and goals.

C.D. Howe Institute Commentary© is a periodic analysis of, and commentary on, current public policy issues. Michael Benedict and James Fleming edited the manuscript; Yang Zhao prepared it for publication. As with all Institute publications, the views expressed here are those of the author and do not necessarily reflect the opinions of the Institute's members or Board of Directors. Quotation with appropriate credit is permissible.

To order this publication please contact: the C.D. Howe Institute, 67 Yonge St., Suite 300, Toronto, Ontario M5E 1J8. The full text of this publication is also available on the Institute's website at www.cdhowe.org.

Government regulation of individual and business activity is part and parcel of modern society. But many businesses find puzzling the legal hurdles, inefficient rules, and uncertain and unpredictable principles that seem to pervade modern regulation.

They also face difficulties in interacting and communicating with government to better understand the regulatory environment. More efficient and effective government regulation would not only reduce the waste of public resources, but would also allow businesses and individuals to better understand their rights and responsibilities while minimizing uncertainty and the cost of interacting with the government.

In Canada and its provinces, governments have recently taken steps to reduce the regulatory burden placed upon businesses by seeking to eliminate duplication of regulatory steps, reducing compliance costs and shortening response times. Last year, for example, Ontario passed the *Cutting Unnecessary Red Tape Act*, 2017. The initiatives in this legislation allowed for greater alignment of regulations with national and international standards, reduced regulatory requirements for firms with good compliance records, and a commitment to electronic submission of any required documents.

In this *Commentary*, I explore how regulators can continue this evolution toward more efficient and effective regulation through one particular channel: embracing data analytics and machine-learning tools. Academics and researchers have long realized how decisionmaking can be improved by the use of good data and effective use of algorithms (Grove and Meehl 1996). The commercial world has also begun to take notice and reap benefits. Recent advances in artificial intelligence, big data and machine learning have disrupted business models in healthcare, finance, law, and in many other industries and professions (Ford 2015; Brynjolfsson and McAfee 2014, 2017; Susskind 2017; Susskind and Susskind 2017; Alarie et. al. 2016a).

This new world of big data, analytics and machine learning presents new and interesting challenges for regulators who oversee how many businesses make decisions. But regulators can also benefit from effective use of data science. These technological advances can help resource-constrained regulators provide more efficient and streamlined regulation.

Some benefits of these data-analytic tools can be realized almost immediately. Regulators can use the vast amount of already-available data and machine-learning tools to predict where they should concentrate their regulatory efforts. These tools also can be used to better inform the regulator whom to investigate and inspect. Furthermore, machine learning can be used to better predict the outcome of likely litigation, ensuring greater cohesion between the views of the courts and those of the regulators.

The author thanks Benjamin Dachis, Tiff Macklem, Michael Trebilcock and anonymous reviewers for helpful comments on an earlier draft. He retains responsibility for any errors and the views expressed. The author is a co-founder of Blue J Legal, a start-up bringing machine learning to law. The company uses machine-learning technology to predict legal outcomes.

The use of such predictive tools enables regulators to streamline their operations, moving resources away from wasteful activities, such as investigating businesses that are almost certainly compliant with the law or proceeding with litigation that will likely be unsuccessful, toward activities that better achieve their regulatory goals.

Moving further into the future, the use of machine-learning tools may have a more dramatic impact upon regulation. Regulators will be able to use predictive tools that erode the tradeoffs between rules (clear, but inflexible) and principles (flexible, but unpredictable). The new predictive technologies discussed in this *Commentary* may permit a movement toward authorizing activities. This would allow businesses and individuals to better understand the legality of their intentions before they act. Instead of submitting forms electronically and waiting for feedback, regulated businesses may be able to receive instantaneous responses from regulators.

This *Commentary* proceeds as follows. First, I briefly explain what data science is and how machine-learning tools can help with prediction issues. Then, I explore two types of prediction problems that regulators currently face: which businesses to investigate and which cases to take to court. I argue that machine-learning tools can help with both, allowing the regulator to better allocate scarce resources. In the following section, I outline a vision of law and regulation in a world where machine-learning tools are pervasive. Essentially, regulation will become more ex ante in its approach and administration, allowing for greater certainty of actions. Potential roadblocks that stand in the way of this vision are also discussed.

Data Science and Machine Learning

Since 2010, there has been an explosion of activity in the field of artificial intelligence (White House Report 2016). This activity has been fueled by the availability of larger data sets, better algorithms, and bigger and faster computers. Recent advances in computing power have resulted in a drastic cost reduction in collecting, creating, storing, processing and analyzing data. It has been suggested that more data has been created in the past 12 months than in the entire history of humankind before 2017. This will likely be true again in 2018. Given the ever-diminishing costs of creating data and the enormous value that insights from these data can generate, the demand for and supply of data continues to grow.

Data science uses empirical observations of the actual behaviour of businesses and individuals. By analyzing these data, we can better predict how business and individuals will behave in the future. Machine learning provides tools for making such predictions. Machine learning refers to the subfield of computer science and statistics in which a machine is fed data and recognizes patterns among the data without being explicitly programmed to do so (Kaplan 2016). The term “machine learning” can be a little misleading. The machine is not learning in the same way that humans learn (Surden 2014). Rather, the machine seeks to optimize predictions within a subset of the data without being programmed by a human.

Of course, one could use traditional statistical or econometric techniques to make predictions. But these models commonly have a different focus: reducing estimator bias. Meanwhile, machine-learning algorithms seek to minimize the prediction error (Athey and Imbens 2017; Mullainathan and Spiess 2017). These objectives appear similar, but they differ in important ways. The traditional techniques are not optimized for prediction. The algorithm that best predicts using out-of-sample data is not necessarily an unbiased estimator.

Indeed, introducing some bias can improve predictability by reducing noise (Kleinberg et. al. 2015). Take the example of predicting real estate prices. Machine-learning algorithms have been shown to do a much better job than models based on ordinary least squares when it comes to predicting the sale price of real estate property

(Mullainathan and Spiess 2017). These machine-learning algorithms are designed specifically to solve "prediction problems."

What constitutes a prediction problem is broader than one might expect. Such problems go beyond the challenges of forecasting the weather or trying to work out the price of stocks in the future. Indeed, many real-life problems can be reframed as prediction problems. When we cross the road, we are predicting that we will not be struck by a vehicle. When we buy gifts for loved ones, we are predicting how much enjoyment they will receive.

The predictions of machine-learning algorithms already pop up in everyday life. Take, for example, facial recognition software used by Facebook. The algorithm uses data in the pixels to predict whether one face is sufficiently similar to other faces that are characterized by other pixels. When Netflix makes a recommendation as to what movie you will enjoy, it is making a prediction based on what you and millions of other people like you have watched. Self-driving cars also use machine-learning technology when predicting how best to navigate complex systems of roads.

In recent years, academic researchers have used machine-learning techniques to explore the efficiency of government policies and business decisions. They have used machine learning to predict the value of elective surgeries on patients suffering from osteoarthritis (Kleinberg et. al. 2015); which restaurants are most likely to violate regulations on hygiene (Kang et. al. 2013); which candidates for jobs will be the best fit (Chalfin et. al. 2016); and, most impressively, predict which criminal defendants are most likely to either fail to turn up to court or commit a subsequent crime if they are released on bail (Kleinberg et. al. 2018). In this latter study, the authors show that machine-learning algorithms can better assess which defendants should be granted bail and which should be denied. Compared to human judges, the algorithm is better at identifying low-risk defendants. They show that the jail population of accused persons could be reduced by up to 41.9 percent without any increase in crime. Similarly, the algorithm identifies those defendants that are at high risk of violating the terms of their bail. Compared to human judges, the crime rate can be reduced by up to 24.7 percent without any increase in the jail population.[1]

The tools for machine learning are widely available at low cost. In the commercial sector, businesses can leverage the data they already have to generate insights that were previously unavailable. Many businesses are, no doubt, investing in creating and collecting more data relevant for prediction problems they face. With a good data scientist on staff, businesses are able to better tailor their products to consumers and better understand the competitive risks they face. There is probably not one large financial services firm in Canada that has not looked into using machine learning to predict which trades to make and which trades to avoid. Insurance companies are using the data to better understand the risk profiles of their clients. Energy companies use complex algorithms to predict usage on the grid to optimize energy storage. These companies are simply using the data they already have along with machine-learning technology to help them make better decisions.

This proliferation of big data and the use of machine learning will present new and unique challenges for imperfectly informed regulators. As more and more private companies use algorithms

1 To be clear, these numbers represent the upper bounds of the authors' estimates compared to the current situation of human decisionmakers. The 24.7 percent decrease in the crime rate is the upper bound of the estimated benefit from not releasing accused persons who are more likely to commit crimes while out on bail while releasing those who are less likely to commit crimes while out on bail.

to make their decisions, those citizens and other businesses that are adversely affected by the outcomes of these algorithms may seek recourse from lawmakers and regulators. For example, should a regulator intervene if insurance companies use predictive algorithms to deny coverage to consumers? Under what circumstances would reliance on the predictions of an algorithm fail to satisfy the principles of fair treatment that protect the rights of the consumer? What if a bank declines to lend funds to individuals based on the recommendation of an algorithm? (O'Neil 2016.) These are all questions that regulators must begin to address.

To oversee the use of these algorithms by the private sector, regulators will need to first understand how these algorithms work. They will need to learn where the data come from and what potential biases may be baked into them to create the algorithm. If particular biases of past human decisions lurk in the data, a decisionmaking algorithm may serve to entrench and replicate these biases. It will be important for regulators to understand what the objectives of the machine-learning algorithm were – what exactly is the algorithm predicting and what objective underpins this prediction. Regulators will have to dig under the hood to determine exactly what the recommendations or decisions of these algorithms actually mean. Put simply, regulators will need to come to grips with the underlying data science and advances in artificial intelligence in order to better regulate businesses that adopt these prediction tools.

But data science and machine learning do not just offer challenges and hurdles for regulators. These tools also offer incredible opportunities. In the next section, I will outline ways in which new tools can be used for more effective and efficient regulation.

How Regulators Can Use Machine Learning

Data science and machine learning are not just tools for regulators of the future. We don't have to wait. Many of the benefits can already be reaped using tools and data that are available today. Effective use of these tools is simply a matter of employing ready-made algorithms in order to extract maximum information from vast data and squeeze out useful insights. In many cases, regulators already collect data that are relevant to the prediction problems they face.

Many regulatory problems are simply prediction problems. Here, I discuss two such problems. First, consider how regulators can use machine learning to better determine who to investigate and who not to investigate. These tools can help regulators who inspect or audit businesses by helping them find the needles in the haystacks. Second, regulators must make choices over which cases to prosecute. But the decision to go to court is clouded with great uncertainty. Since the outcomes of adjudication can be difficult to predict, regulators can use data analytics to better determine which cases to take to trial and which cases to drop.

Targeting Who to Investigate and Who Not to Investigate

The question of where to place investigation resources is a common problem facing regulators. A regulator needs to make choices about how to allocate and prioritize scarce resources. Who should be investigated? Who should not? With the right data and appropriate data analytics, predictions can be made about where to best place investigation resources. Better-informed regulators can move resources away from investigating businesses that are at low risk of violating the law toward those that are high-risk violators.

For the most part, Canadian governments can take advantage of these machine-learning tools today because the data already exist. Ontario's Financial Services Commission, for example, has data on millions of pension plans, mortgages and insurance contracts. For its part, the Ontario Securities Commission collects data on millions of transactions every month. Using these data, regulators can place red flags on particular advertisements, products, deals or institutions. In many cases, it is simply a matter of using ready-made algorithms in order to extract maximum information from vast data and squeeze out useful insights. The data may need to be re-structured a little for optimal predictive power, but the data are there. In other cases, it may be that data need to be collected in order to generate the necessary insights.

Currently, governments around the world are using such tools to detect fraud. The UK Serious Fraud Office used machine intelligence to efficiently filter and search through 30 million documents during its 2017 investigation of Rolls Royce. The UK Financial Conduct Authority has also begun to develop machine-learning algorithms to help detect misrepresentations by financial advisers (Hunt 2017). Many tax authorities around the world have begun to realize the enormous potential of using data analytics to detect evasion and other suspicious patterns (Ernst & Young 2018) while competition regulators have begun looking into using such tools to detect the presence of cartels (OECD 2018).

One could easily imagine similar tools used to detect insider trading or fraudulent securities sales. The BC Securities Commission (BCSC) has done innovative work in this field, introducing a number of predictive risk models to improve targeting. Data-driven analyses have been developed to predict which financial dealers are more likely to experience significant deficiencies and to anticipate which issuers were more likely to have to restate their financial statements. While these tools were not based on machine learning, the BCSC recognized the challenges of having data that was not structured in a way that was conducive for prediction.

This data-driven process would work for targeting in cases where the regulator inspects or audits. Regulators of food hygiene, for example, could improve targeting by determining which restaurants and bars are most likely to fall short of the required standards. The process of improving targeting is already beginning in Ontario. The province is beginning to adopt simple practices that will reduce inspections for businesses with good compliance records. On the other hand, in a world where evidence-based decisionmaking is not used, compliant businesses would likely experience similar levels of inspections as those businesses that are non-compliant. This is an inefficient use of the regulator's resources.

Governments can better allocate resources by using data analytics that enable them to target non-compliant businesses. For their part, regulators collect data on a number of different performance measures. For example, a regulator of food hygiene would have data on how long restaurants and bars have been open, on the type of building the establishment operates in, how many other businesses are operated by the same operator, the type of food served, the prices of the products, the likely costs of food and labour, among many other variables. The regulator would also maintain detailed data as to whether these businesses have been compliant in the past. In our example, the food-hygiene regulator would know the past compliance grades of existing restaurants and bars.

Machine-learning techniques can be used to find patterns in these data. Regulators can develop predictive algorithms that allow them to better anticipate which companies are likely to be non-compliant. The regulator of food hygiene would be able to predict what "grade" new and existing restaurants will receive. They would also gain insight into when costly spot inspections are necessary. If regulators focus on those companies most likely to be non-compliant, resources are less likely to be wasted. It also means that compliant businesses will also save time and money, as they won't have to prepare for as many detailed inspections.

Of course, care must be taken to ensure that the existing data do not merely bake in the biases of previous regulators. For example, a concern with police using machine-learning algorithms to predict where crime will occur is that the data in the system may reflect inherent biases (O'Neil 2016). Using this data would serve only to reinforce these biases. Thus, algorithms need to be carefully developed and refined in order to maximize predictive value – and not merely recreate or entrench existing practices. But machine-learning tools can help here, too. In the context of bail decisions, for one, academics have shown that machine-learning algorithms can reduce errors and bias exhibited by human judges (Kleinberg et. al. 2018).

This discussion, however, presumes that regulators have collected and maintained detailed records of all relevant variables. If the relevant historical data do not already exist – or if they are not structured in such a way that is valuable – regulators should begin investing in collecting, creating and curating the necessary data that is "fit for purpose."

As noted above, the cost of data collection and processing is falling. For some regulators, though, generating these new data may appear to be prohibitively expensive, and traditional decision-making methods may prove optimal. However, this view is myopic, as the power of data-driven analysis will continue to improve in the near future.

Predicting When to Take Cases to Court

Machine-learning techniques can be used to predict how judges will decide cases (Katz et. al. 2017; Alarie et. al. 2016c). A resource-constrained regulator must learn to pick its fights wisely. Regulators should not waste resources litigating cases they are likely to lose. Instead, regulators should put resources only toward cases that they are likely – or predicted – to win. Regulators can turn to the data and use machine learning to predict how a court would resolve a particular problem.

This discussion, of course, presumes that the objective of litigation is to win cases. The regulator, however, may have other objectives. For example, the regulator may be challenging behaviour in order to achieve a deterrent effect. If this is the case, then the regulator can develop algorithms that predict the likely deterrent effect. Nonetheless, understanding the likelihood of victory is an important factor in deterrence.

Take the following example: imagine a pension-plan regulator who wishes to know whether a particular plan covers workers who believe they are employed in an industry, but whose firm claims they are independent contractors. The regulator may be unclear as to whether the courts will classify these workers as employees or independent contractors. This legal question can be quite complex. It turns on a vague legal standard that involves looking at a number of different aspects of the relationship between the worker and the party who hired the worker. This issue has been litigated more than 1,000 times under tax and employment law in Canada. The volume of litigation over this vague legal standard suggests that there is both uncertainty and disagreement as to how these cases should be decided (Alarie et. al. 2016c).

But the volume of cases presents a strong opportunity for those seeking to understand how courts decide such cases with data analytics. Simple machine-learning techniques can be applied to make such predictions. The data, here, may need to be structured in order to be effectively analyzed and generate predictions. The decisions of judges and other adjudicators are commonly found in written opinions that, for the purposes of data science, are largely unstructured. Even if the data is not structured, recent advances in natural-language processing may be used in the near future to provide additional insights into legal decisions.

If the precedent cases are suitable for analysis, data analytics can be used to anticipate the outcomes of new cases. These new scenarios can be compared to all previous court decisions. Predictions can be made about the likelihood of any particular worker being classified as an employee. Cases where the regulator is objectively likely to lose based on the facts of the case can be dropped. Resources can, therefore, be allocated to cases with greater likelihood of success.

Legal uncertainty is costly not only for regulators. It is also enormously costly for the regulated businesses. Companies pouring resources into defending cases that the regulator is unlikely to win creates a lose-lose situation. But even without incurring large litigation costs, legal uncertainty is still costly for businesses. If they are unaware of how regulators and courts will resolve their case, this creates risk for any decision they make.

The magnitude of legal uncertainty is likely greatest when the laws that are being regulated establish vague standards. The outcomes of cases where the laws are clear bright-line rules are far more predictable for both regulators and the regulated parties. But how these vague standards play out in practice commonly depends on how courts view what is "reasonable" or "excessive" or "material." Careful data analytics can help resolve this legal uncertainty.

Indeed, the outcomes of litigation can be predicted in any field of law, provided there are sufficient data (i.e., previous decisions). One might imagine that such data analytics could provide great certainty for regulators of professionals to better determine whether a court will find a legal obligation has been breached. Machine-learning tools could also be used to predict the quantum of damages in insurance cases.

Using algorithms to help with decisionmaking also helps reduce inconsistency within a regulatory body. In doing so, the use of machine-learning algorithms may promote fairness. One particular investigator may be more aggressive than another working for the same regulator. One lawyer may be more aggressive than another. But regulated parties may not know whether the investigator or lawyer assigned to their case is aggressive or lenient. This generates further risk. The use of predictive algorithms, however, can help regulated parties better understand their obligations and monitor whether they receive fair treatment, irrespective of which investigator or lawyer is assigned to their case.[2]

Regulation in the Near Future

A Roadmap for Regulation

In this section, I will outline how big data and machine learning may disrupt the regulatory environment in the near future. Such changes are more far reaching than those discussed previously. In a series of recent papers with various co-authors, I have explored how big data and artificial intelligence will change the way that laws and regulations will be consumed and produced in 10, 20 or 50 years (Casey and Niblett 2016, 2017a, 2017b; Alarie et. al. 2017).

The predominant theme of our work has been that the law will, over time, provide greater certainty to individuals and businesses. Lawmakers will be able to create laws that indicate with greater precision what types of behaviour are illegal and which are not. Lawmakers will have greater ability to provide this information *before* the individual or business acts, rather than waiting to give a legal determination *after* they have acted. We suggest that the reliance on ex-post determination and adjudication will diminish somewhat over time.

We predict that law and regulation will move toward an ex-ante model. Instead of regulators or judges investigating, inspecting or auditing the behaviour of individuals and businesses who

2 The focus in this section of the *Commentary* has been on how machine learning and data analytics can help regulators to use existing the resources they have more effectively. The data they have can be used to determine who to target and improve compliance. But the benefits of effective data use of data by the public sector extend far beyond these objectives. For example, the government could make effective use of its existing data to create wealth for Canadians. Take the following example: with our single-payer healthcare system, the government has stored billions of data points, such as images. This health information could be anonymized and provided to start-ups to develop new AI tools that could lead to greater wealth creation and improved healthcare.

have already undertaken actions, we predict that regulators will be able to rubber stamp these actions beforehand. Future regulations thus will offer the predictability of ex-ante rules while still preserving the flexibility and tailored nature of legal principles (Casey and Niblett 2017b).

Two types of technology are paving the way for this change. First, the predictive technologies discussed above will enable lawmakers to envisage the likely impact of particular laws. As the cost of prediction falls, it will be easier for regulators to understand the consequences of the laws they enforce. Second, the costs of communicating are dropping dramatically. Regulators can communicate decisions to many businesses and individuals at the same time. This means that businesses and individuals can receive communications from regulators and other lawmakers instantaneously, before any action is taken.

There are enormous benefits to regulators making decisions before individuals and business act. Such laws greatly reduce uncertainty. Advance rulings, either authorizing or prohibiting particular behaviour before investments are made, provide certain outcomes and reduce the likelihood of wasted investments. Regulators currently use advance rulings and authorizations in some spheres. For example, the Canada Revenue Agency provides advance tax rulings on particular issues, while other regulators provide "no action letters" to notify regulated businesses that their intended behaviour will attract no further scrutiny.

Ex-ante authorization thus reduces the costs of uncertainty. For example, currently, two firms seeking to merge can ask the Competition Bureau for an ex-ante ruling. If ex-ante authorization were not available, companies seeking to merge would do so amid great uncertainty. Companies may waste resources structuring deals that the Competition Bureau disapproves of after the companies have invested in undertaking the merger. Furthermore, companies that could lawfully merge under the *Competition Act* may be dissuaded from doing so because of legal uncertainty over the merger.

In recent papers, co-authors and I argue that, in the near future, parties will seek prior authorization for many more regulated actions. Furthermore, the time and cost for regulators to respond to the queries will fall drastically. Instead of relying primarily on vague guidelines, regulators will be able to offer more expedient personalized and (eventually) legally binding responses.

Consider the evolutionary path of how algorithms will be adopted and used by regulators. Initially, these types of algorithms will be used merely to inform decisionmakers. For example, an algorithm could be used to help a regulator detect the presence of cartels. With more data, the algorithm will become better at identifying violations and non-violations of the law. Where there is very high confidence in the algorithm, its predictions may become the binding law. Where there is low confidence, then human regulators are still needed to exercise judgment about the algorithm's prediction. But consider that, over time, predictive algorithms in all forms of regulatory activities will continue to improve. They will incrementally move from the category of informing regulators to suggesting or recommending actions. And then, after another period of time, algorithms will move from recommendations to providing binding decisions that provide ex-ante authorization (or non-authorization).

To illustrate how this might work, an analogy and some examples may be helpful. First, consider how big data has changed how drivers consume directions while driving. Twenty years ago, drivers used paper road maps to determine their route from one place to another. However, the driver might not know where they were and could get lost.

Today, drivers simply use a smart phone or GPS for directions. They enter their destination and the device provides precisely tailored directions on how to get there. These directions are optimized using vast amounts of data and complex predictive algorithms. This information is provided instantly. The smart phone greatly reduces uncertainty. Drivers don't have to worry if the paper road map is out of date or if road conditions have changed. The device tells the driver how to navigate the complex network of roads in the most efficient way.

This analogy demonstrates how my co-authors and I see the future of regulation. Instead of a climate where regulators issue vague principles (the road maps in our analogy), we should be thinking about how to create algorithms and programs that provide tailored and personalized ex-ante regulations (the device directions). Regulators currently use guidelines that are vague and create uncertainty. But in the near future, we should be able to enter our query and ask: how can my business legally navigate the complex network of regulations and laws? (Alarie et. al. 2016a, 2016c; Casey and Niblett 2016).

Now, consider how this might work in practice. In the above example of whether workers are employees or independent contractors, businesses may be able to use an app operated by the regulator. Businesses enter the facts about the nature of the relationship with their workers. They could receive an instantaneous answer to their problem from the regulator. This app would be based on the machine-learning technology described above that compares the facts of the case at hand to every single precedent. No longer would the business need to wait until after they act for a regulator to determine whether the workers they have hired were eligible for pensions.

Also, consider the regulation of misleading advertisements. Ex-post adjudication of whether an advertisement is misleading is costly because consumers may have already suffered as a result of the misrepresentation. In the near future, regulators should be able to develop predictive tools that classify which advertisements are misleading under the law and which are permissible. If such a tool were to exist, businesses should be able seek ex-ante authorization for their advertisements. Those deemed to be misleading might never be seen by a single consumer.

Potential Roadblocks

Unpredictability and uncertainty in regulation imposes large costs on businesses in particular. Predictive analytics can help overcome some of these problems. A trend toward algorithm-assisted regulation and algorithmic regulation will mean less ex-post determination and more ex-ante authorization. This will lead to greater certainty for all parties. But there are a number of risks and considerations that need to be addressed before this vision is realized.

Quality of Data

Regulators need to ensure that the data are the right data and are of sufficient quality in order to develop these algorithms. Governments need to ensure that the right data are being collected and structured in such a way that is helpful when addressing each prediction problem. Of course, if the algorithms rely on poor data, or the objective is mis-specified, regulatory decisions will suffer.

Accountability and Due Process

What happens if an algorithm makes an error? Is it sufficient to say that algorithmic errors are fewer and smaller than if we left the decision-making to humans? Algorithms will make errors. How do we structure the law such that these algorithms are tuned correctly to minimize errors? While individual employees of the regulator may not be held liable for their actions, they may face repercussions in the form of demotions or loss of employment. When should we hold the creator of an algorithm responsible?

In a world where algorithms form the bases for regulatory responses, the questions of due process will be of even greater importance. Indeed, human regulators will need to understand how to best work with machine intelligence to provide efficient and effective regulation. When should the algorithm's answer be allowed to stand? When should humans intervene and override? One response is to never let the algorithm make binding decisions and always have some element of human responsibility. The algorithm merely provides suggestions or recommendations, at best. Of course, this adds elements of uncertainty back into the equation.

Need for Transparency

It is often the case that, given sufficient data, the more complex the machine-learning algorithm, the greater the level of predictability. But this comes at a cost. The more complex the algorithm, the more difficult it is to explain a prediction. This may not be a problem if the objective is to predict what transactions might be fraudulent and which should be investigated. However, where binding legal decisions are being made or recommended by algorithms, citizens and businesses will request reasons why the decision was made.

For example, if a proposed merger were rejected on competition grounds, it would be considered sufficient for the regulator to say the algorithm predicted this to be anti-competitive. There are simple machine-learning algorithms where coefficients can be used to explain the weights of different variables. Even with the more complex algorithms, if the regulated actors trust the data and understand the algorithm's objective, the question of transparency can be satisfied (Selbst and Boconas 2018). Currently, we rely on human decision-makers to provide reasons for their decisions. But human brains are even more of a black box than machine-learning algorithms (Pande 2018).

Privacy and the Reluctance to Share Data

The degree to which firms and individuals are willing to share data and information with regulators will affect the degree to which regulation moves from ex post to ex ante. In the same way that smart phones would not provide the benefits discussed above if drivers were unwilling to enter their destination, if regulated actors value privacy or their own data more than ex-ante regulation, guidance and ex-post adjudication will still be a pervasive form of regulation. Furthermore, regulators would need to share data with other regulators in order to harness the opportunities offered by machine learning – as well as providing enormous economies of scale.

Desirability of Uncertainty

Providing legal certainty may have downsides. If the algorithm fully discloses how to navigate the law, loopholes may be fully exposed. Currently, we allow for ex-post determinations to hold bad actors liable if they have abused the spirit of the law. For example, taxpayers may fall afoul of the general anti-avoidance rule if transactions are designed primarily to avoid tax even if the transaction doesn't contravene specific sections of the *Income Tax Act*. These types of standards allow lawmakers the space to acknowledge that the law does not fully capture all situations and contingencies. Thus, there may be particular laws where we may not wish to reveal the outcomes ex ante.

New Laws and Changing Laws

Predictive machine-learning algorithms are based on past behaviour and past precedents. But if new laws are instituted, it may be difficult to train an algorithm to predict likely outcomes. Similarly, social trends change over time. How can algorithms best account for changing social tastes and desires? One solution is to change the objective of the algorithm and change what it is that the algorithm is maximizing. But that may be insufficient where the data have been created to meet a particular objective that is no longer desirable.

Conclusion

As businesses make more and more decisions through algorithms, lawmakers will face new and difficult questions as to how best regulate such decisionmaking by the private sector. This *Commentary* has focused, however, on a different connection between big data and regulation: how can data analytics and machine learning assist regulators in cutting through information problems, streamline administrative burdens and pave the way for more efficient regulation.

The effective use of big data in regulation is one of the most important steps regulators can take in the near future. Implementation of the tools discussed in this *Commentary* is simply a continuation of a trend toward better regulation. But regulators who elect not to use the data they have in the most effective ways will be wasting resources. These regulators will further burden regulated businesses with inefficient targeting and litigation. Those regulators who fail to even collect data for policy decisions will fall further behind the ball.

Finally, a further ancillary benefit of regulators using machine-learning tools is they would learn a great deal about the technology and what it can do. This knowledge would be vital for allowing regulators to better design appropriate future regulations. The lack of knowledge and experience with machine learning among regulators means they are ill-equipped to address a slew of regulatory issues that will arise from the use of big data. If Canada seeks to be a leader in innovation, we need a regulatory apparatus that is innovation friendly. And we need regulators that understand how to best harness the power of machine learning.

REFERENCES

Alarie, Benjamin, Anthony Niblett, and Albert H. Yoon. 2016a. "Law in the Future." *University of Toronto Law Journal*, 66(4): 423.

———————. 2016b. "Regulation by Machine." *Journal of Machine Learning Research: Workshop and Conference Papers*. Barcelona: Neural Information Processing Systems.

———————. 2016c. "Using Machine Learning to Predict Outcomes in Tax Law." *Canadian Business Law Journal*, 58(3): 231.

Athey, Susan, and Guido Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives*, 31(2): 3.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brillliant Technologies*. New York: W.W. Norton & Company.

———————. 2017. *Machine, Platform, Crowd*. New York: W. W. Norton & Company.

Casey, Anthony J., and Anthony Niblett. 2016. "Self-Driving Laws." *University of Toronto Law Journal*, 66(4): 429.

———————. 2017a. "Self-Driving Contracts." *Journal of Corporation Law*, 43(1):100.

———————. 2017b. "The Death of Rules and Standards." *Indiana Law Review*, 92(4): 1100.

Chalfin, Danieli, et al. 2016. "Productivity and Selection of Human Capital with Machine Learning." *American Economic Review: Paper and Proceedings*, 106:124.

Ernst & Young. 2018. "The Tax Authority of the Future: How Tax Authorities Are Using Analytics to Deliver New Levels of Value." Available at: http://www.ey.com/Publication/vwLUAssets/ey-the-tax-authority-of-the-future/$File/ey-the-tax-authority-of-the-future.pdf.

Ford, Martin. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books.

Grove, William, and Paul Meehl. 1996. "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy." *Psychology, Public Policy, and Law*, 2(2): 293.

Hunt, Stefan. 2017. "From Maps to Apps: The Power of Machine Learning and Artificial Intelligence for Regulators." Beesley Lecture. Oct. 17, 2017. Available at: https://www.fca.org.uk/news/speeches/maps-apps-power-machine-learning-and-artificial-intelligence-regulators.

Kang, Jung Seok, et al. "Where Not to Eat? Improving Public Policy By Predicting Hygiene Inspections Using Online Reviews." 2013. In Proceedings of the conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 1443.

Kaplan, Jerry. 2016. *Artificial Intelligence: What Everyone Needs to Know*. New York: Oxford University Press.

Katz, Daniel Martin, Michael Bommarito, and Josh Blackman. 2017. "A General Approach for Predicting the Behavior of the Supreme Court of the United States." *PLoS One*, 12:4 e0174698.

Kleinberg, Jon, et al. 2015. "Prediction Policy Problems." *American Economic Review: Paper and Proceedings*, 105:491.

Kleinberg, Jon, et al. 2018. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics*, 133(1): 237.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives*, 31:87.

OECD. 2018. "Workshop on cartel screening in the digital area." Conference presentations. Available at: http://www.oecd.org/daf/competition/workshop-on-cartel-screening-in-the-digital-era.htm.

O'Neil, Cathy. 2016. *Weapons of Math Destruction*. New York: Crown.

Pande, Vijay. 2018. "Artificial Intelligence's Black Box Is Nothing to Fear." *New York Times*. January 25. Available at: https://mobile.nytimes.com/2018/01/25/opinion/artificial-intelligence-black-box.html?

Selbst, Andrew D,. and Solon Barocas. 2018. "The Intuitive Appeal of Explainable Machines." Unpublished manuscript. Available at:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3126971

Surden, Harry. 2014. "Machine Learning and Law." *University of Washington Law Review*, 89:87.

Susskind, Richard. 2017. *Tomorrow's Lawyers: An Introduction to Your Future*. 2nd ed. Oxford: Oxford University Press.

Susskind, Richard, and Daniel Susskind. 2015). *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. Oxford: Oxford University Press.

White House Report. Executive Office of the President. National Science and Technology Council. 2016. "Preparing for the Future of Artificial Intelligence." Available at: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

Notes:

Notes:

Recent C.D. Howe Institute Publications

March 2018	Ambler, Steve, and Jeremy Kronick. *Faulty Transmissions: How Demographics Affect Monetary Policy in Canada.* C.D. Howe Institute Commentary 506.
March 2018	Dachis, Benjamin. *Fiscal Soundness and Economic Growth: An Economic Program for Ontario.* C.D. Howe Institute Commentary 505.
March 2018	Robson, William B.P., and Parisa Mahboubi. "Inflated Expectations: More Immigrants Can't Solve Canada's Aging Problem on Their Own." C.D. Howe Institute E-Brief.
March 2018	Feldman, Michael K. *The Case for Longer Mortgages: Addressing the Mismatch between Term and Amortization.* C.D. Howe Institute Commentary 504.
February 2018	Schwanen, Daniel, and Rosalie Wyonch. "Canada's 2018 Innovation Policy Report Card." C.D. Howe Institute E-Brief.
February 2018	Robson, William B. P., Alexandre Laurin, and Rosalie Wyonch. *Righting the Course: A Shadow Federal Budget for 2018.* C.D. Howe Institute Commentary 503.
February 2018	Richards, John, and Parisa Mahboubi. "Measuring Student Outcomes: The Case for Identifying Indigenous Students in Canada's PISA Sample." C.D. Howe Institute E-Brief.
February 2018	Eichenbaum, Martin, Benjamin K. Johannsen, and Sergio Rebelo. *Understanding the Volatility of the Canadian Exchange Rate.* C.D. Howe Institute Commentary 502.
February 2018	Dachis, Benjamin. *Death by a Thousand Cuts?Western Canada's Oil and Natural Gas Policy Competitiveness Scorecard.* C.D. Howe Institute Commentary 501.
January 2018	Busby, Colin, Ramya Muthukumaran, and Aaron Jacobs. "Reality Bites: How Canada's Healthcare System Compares to its International Peers." C.D. Howe Institute E-Brief.
January 2018	Macdonald, Bonnie-Jeanne. *Headed for the Poorhouse: How to Ensure Seniors Don't Run Out of Cash before they Run Out of Time.* C.D. Howe Institute Commentary 500.
January 2018	Wyonch, Rosalie. *Risk and Readiness: The Impact of Automation on Provincial Labour Markets.* C.D. Howe Institute Commentary 499.

C.D. HOWE
INSTITUTE

67 Yonge Street, Suite 300,
Toronto, Ontario
M5E 1J8

Canadian Publication Mail Sales
Product Agreement #40008848