ISSN (online) 2664-9535 (print) 2664-9527 https://atcr.kra.go.ke

Machine learning for intelligence-driven Customs management

Job Kavoya [1]

[1] Business Process & Information Management, Transformation Leadership Office, Kenya Revenue Authority, Nairobi, Kenya

E-mail: job.kavoya@kra.go.ke

Received 18 October 2019
Accepted for publication 25 October 2019
Published 15 January 2020

Abstract

The growth of big data is evident as organizations' application of information technology continue to improve and data storage costs continue to fall. The growth of big data presents an opportunity for organizations to better understand their customers, develop strategies that will generate additional revenue, and grounds for business model innovation. However, a very small portion of data collected by organizations gets analyzed. This scenario creates a loophole that may deny established business additional revenues, and threatens their long-term existence if new market entrants explore this weakness. Intelligence-driven organizations analyze data to generate actionable insights that guide decision making. Customs administrations generate huge amount of unstructured data, but what percentage gets analyzed? This paper presents two frameworks that can be customized by customs to develop strategies for intelligence-driven operations. First, is the SCALE framework that defines attributes of intelligence-drive organizations and secondly, the data-value framework that defines how organizes can transform data to value. These frameworks are enhanced by a review of three customs services in the world. In summary, two key lessons are reviewed. First, is the focus on enterprise-wide adoption of analytics and secondly, is the role data in becoming intelligence-driven. The paper concludes by highlighting use cases where customs can leverage machine learning capabilities to enhance operations.
Keywords: intelligence-driven, big data, analytics, customs, machine learning

ISSN (online) 2664-9535 (print) 2664-9527 https://atcr.kra.go.ke

1. Introduction

The amount of data generated globally is expected to reach 44 zettabytes by 2020 Desjardins (2019). This represents an exponential growth of data driven by interactions of people using mobile devices and internet Devakunchari (2014), Khan et al. (2014), Andersen et al. (2018). Some of this data is produced as a by-product of a process, usually referred to as digital exhaust data McKinsey (2011). In a single day, 500 million tweets are sent, 294 billion emails are sent, 4 petabytes of data content is created on Facebook, 5 billion searches are made, 65 billion messages are sent on WhatsApp Desjardins (2019). The demand for using evidence-based decision-making has led to growing trends in faster processing and data analysis Devakunchari (2014), Michael and Miller (2013). Faster processing of big data requires specific techniques, including machine learning, data mining, and visualization Khan et al. (2014). Machine learning is a subset of artificial intelligence. Machine learning applies algorithms that are able to learn from data without being explicitly programmed Pyle and Jose (2015). Organizations planning to create new value through data must be deliberate to invest in the right resources Batistič and Laken (2019). A combination of resources and capabilities, including information technology infrastructure, organizational culture, and skill set will create new value Fosso et al. (2015), Gupta and George (2016). This combination stimulates data-driven decision-making capabilities, increasing the precision of judgements McAfee and Brynjolfsson (2012).

The amount of data that is analyzed remains very low. Henke et al. (2016) show that only one (1) percent of data is analyzed, Gantz & Reinsel (2012) show only 0.5 percent of data is analyzed. Data analytics is required for organizations to become intelligence-driven Andersen et al. (2018), Dyckhoff et al. (2012). Intelligence-driven organizations have to deliberately stop making decisions based on common sense and adopt algorithms and systems that are able to learn based on big data Andersen et al. (2018). Intelligence-driven organizations have a capability to SCALE. This means that intelligence-driven organizations can sense (observe and record both external and internal environment), comprehend (use data to discern context, detect patterns and make inferences), act based on evidence-based insights, learn (acquire knowledge to adapt and improve behavior), and explain (show how something works, articulate purpose, set direction) their environment and their interactions with it Andersen et al. (2018).

Customs administrations generate huge amounts of raw data. This data majorly includes transactional data relating to imports, exports, transit, transshipment, and warehousing. In addition, customs agent and trader data including importers, exporters, goods description, values, quantity, and company ownership details including directors and staff details is available. Raw data does not offer any guidance on what actions should be taken Dyckhoff et al. (2012), Sclater, Webb and Danson (2016). Actionable insights should be generated to offer the necessary guidance Jørnø and Gynther (2018). The data generated by customs forms a very useful source of insights that could transform risk management, compliance management, enhance efficiency in operations, improve revenue, and ultimately become an intelligence-driven organization. This paper presents a review of two proposals on how organizations can transform to become intelligence-driven, lessons that Kenya Customs can learn from other customs, and finally, the use cases that customs can adopt in its quest for becoming intelligence driven.

2. Empirical review

This sections seeks to clarify two things: first, is to examine the characteristics of intelligence-driven organizations and secondly, how organizations can transform data to create new value.

2.1 Intelligence-driven organizations

Facebook took 3 years to reach 50 million users, WeChat took 1 year to reach the same number of users. Mobile phones took 12 years, internet 11 years, computers 14 years, credit cards 28 years. Interestingly, Pokémon Go took only 16 days to reach the same number of users Desjardins (2018). Technological advances create a capability that ensures businesses can attract huge number of users in a very short period of time due to presence of network effects, faster communication, and the nature of goods Desjardins (2018). To succeed in this fast changing environment, businesses have to be intelligence-driven Andersen et al. (2018) and deliberately invest in organizational innovation Fjeldstad et al. (2012), Galbraith (2010). To understand the characteristics of intelligence-driven organizations, this paper will use Andersen et al. (2018) SCALE framework. These organizations are intelligent systems that can Sense, Comprehend, Act, Learn, and Explain (SCALE). The framework is presented diagrammatically in the figure below:

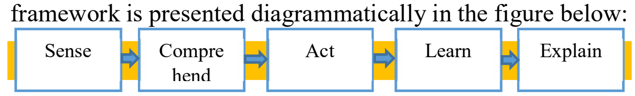


Figure 1: Intelligence-driven organizational characteristics: Source: Andersen et al. (2018)

Sense

Intelligence-driven organizations are able to sense: a capability to observe and register the external and internal environment Andersen et al. (2018). Organizations are able to sense the need to make necessary changes internally as well as externally Teece, Pisano, and Shuen (1997). This need is identified through data analytics. This data is generated from different sources: internet of thing devices, customer data available on digitally enabled devices and services, as well as primary data from core systems. For customs, this data will originate from sensors (electronic cargo tracking system), declarations data, licensing data, offence data, data from third parties, staff feedback, and social media data sets. The sense characteristic should enable customs know what is happening in their environment. This condition must be met for the other attributes to be defined.

Comprehend

Data collected from sense is raw data, and will not be useful without analysis to generate actionable insights. Comprehend involves analytics to discern context, detect patterns, and make inferences Andersen et al. (2018). To comprehend the environment, customs will need to deploy different models that can identify what is causing the observed behavior. For example, if customs revenue is declining while import volumes are growing, comprehend should help the organization to understand what is happening and propose ways of rectifying it. The proposed response could be modeled before it is deployed to ensure the actions will deliver the required results. Intelligence-driven organizations can scan the environment and derive meaning from the environment. This function should be automated or embedded into the core system. The outcome of comprehend is generation of actionable insights.

Act

This attributes refers to decision-making aspect based on insights generated at comprehend. Action maybe automated. For example, a number of organizations have deployed chatbots that are capable of responding to standard questions. In customs, customer queries including the clearance status of declaration, payment status, bond status, license renewal could be handled by a chatbot. Act could include performance measurement insights that could help organizations decide the nest step of action. In the United States, Joe & the Juice deployed a system to assess market potential in different locations Andersen et al. (2018). In customs, this could include decisions on whether a license applicant should be given license based on risk analysis of similar licenses. In post clearance audit, act could be used to develop a system that recommends which company should be audited based on past audit findings.

Learn

Learning is the ability to acquire knowledge, including experimentation, model improvement among others Andersen et al. (2018). This attribute includes rethinking on what data sets does the organization lack that, if available, can improve performance, service delivery, or risk management. For instance, if transit declarations data was shared by the country of destination with the customs where transit started, this will reduce the probability of dumping goods. Availability of this data will help customs in refining models on transit monitoring. Artificial intelligence models will create additional benefits.

Explain

Explain shows how something works, and machine learning models are limited in this aspect Andersen et al. (2018). These five attributes describe how intelligence-driven organizations work. For customs, these five attributes are necessary to become intelligence driven.

2.2 Data to intelligence

Customs data should provide guidance on which declaration is risky, which product has been undervalued, what product has been misclassified, which consignment should be targeted. This value is created purely through information exchange between different parties. Customs operations, therefore can be classified as an information-intensive service Lim and Kim (2014). To create value from data, organizations need to have a clear strategy starting from systems design to how this intelligence will be used and applied. Systems should be designed to capture data in the most appropriate way, this data is transmitted and stored in a place that it can be accessed for analytics. Customs can apply the data-value framework defined by Lim et al. (2018). This framework clarifies how data can be transformed into value. In the high level, this framework can be summarized as follows:

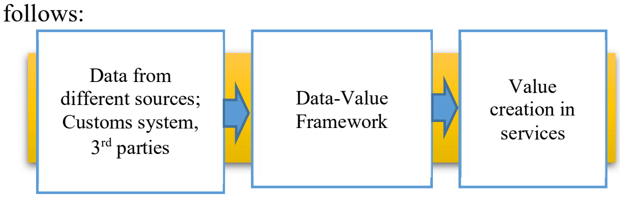


Figure 2: Data-Value framework high level: Source: Lim et al (2018).

Most organizations have data already within their servers and where third party data is required is well known. Additionally, organizations know very well the value they intend to create to serve their customers better, to improve service efficiency, to cut costs or to save time. The data-value framework, Lim et al (2018) acts as a bridge to connect the two sides. The framework has nine (9) steps to transform data to value. These are: data source, data collection, data, data analysis, information on the data source, information delivery, customer (user), value in information use, and provider network Lim et al (2018).

#	Steps	Description	Example, RECTS
	Data source	Identification of data source(s)	Cargo tracking device (hardware), customs core system.
	Data collection	How the information is collected	Through GPS system, sensors
	Data	Description of dataset	Device location, state of device, time taken, distance, speed, direction, driver details, truck details, date.
	Data analysis	What data analysis methods are to be used?	Automated signals on diversion of cargo (in-built in the system), requires human intervention.
	Information on the data source	What to be understood from the data source?	Understand effectiveness of ECTS in preventing diversion or dumping, percentage of false alarms.
	Information delivery	How information is to be communicated	Alarms, telephone calls, rapid response team.
	Customer (User)	Who will use the data	Drivers, KRA customs, Traders, Regional revenue authorities.
	Value in information use	How will the information benefit the user	Protect local industries by preventing diversion and dumping, apply intelligence-based targeting.
	Network of providers	Feedback to partners involved	RECTS hardware manufacturers, data transmission service providers, regional revenue authorities.

Figure 3: Data-Value framework: Adapted from: Lim et al (2018).

Developing such a step-by-step guide can help customs to become intelligence-driven. Customs knows what it intends to achieve in different operational areas. By working backwards, it possible to fill in the gaps and come up with strong strategies on how to achieve the required goals.

Lessons from other customs services

2.3 New Zealand customs

New Zealand customs is moving towards intelligence-driven customs operations. The New Zealand customs started transformation with the establishment of Joint Border Management System (JBMS); a Trade Single Window (TSW) system that links other agencies systems including passenger clearance and entries lodgment, and delivers tools for risk management and intelligence Okazaki (2017). JBMS was introduced in 2007 bringing together customs, department of labor, ministry of agriculture and forestry, and the ministry of transport, with a focus to deliver a practical way to improve border management in New Zealand customs (2015). The TSW provides an all-in-one model with a global data standard to achieve assurance in trade and travel New Zealand government (2012). The delivery of this solution has created a new capability that enables New Zealand customs to better predict threats, monitor trends, and target high risk transactions Okazaki (2017). In addition, it is now possible to monitor risks in the supply chain, achieve targeted interventions, and post entry assurance to ensure correct revenue is collected New Zealand government (2012). To support the JBMS, New Zealand implemented supporting legislations to ensure clear guidelines on data exchange and cooperation between government bodies.

A report by KPMG (2019) clarifies that for customs to increase compliance and efficiency, data and analytics should be incorporated in their strategy. New Zealand established a joint border analytics team comprising experts from different organizations, including customs, immigration, and partner government ministers New Zealand Customs Service (2014). The team comprise of data modelers, data scientists, data wranglers, business analysts, and subject matter experts. The source of data for analysis includes data on cargo, passenger data, data on parcels and open data sources. To ensure intelligence and insights generated from this data is put to the intended use, New Zealand has implemented guidelines on data sharing, privacy impact assessment, and continuous engagement with the joint border analytics governance group. Privacy impact assessment ensures personal information's privacy is not compromised during analytics New Zealand Customs Service (b).

2.4 Her Majesty's Revenue and Customs (HMRC)

Organizations focused on becoming intelligence-driven should ensure they have assurance and integrity of systems and data. HMRC has implemented a Community System Providers (CSP) to address the issue of data quality and system assurance Okazaki (2017). As Okazaki explains, HMRC has symbiotic relationship with commercial operators enabling United Kingdom to have its intelligence tools embedded into the commercial systems. This delivers seamless data pipelines that ensure right data from the right source is available.

HMRC has deployed Connect system to detect and prevent fraud SAS (2014). The connect data warehousing and analysis project relies on over thirty (30) databases as a source of data Rigney (2016). A report by SAS (2014) highlights the benefit of using different datasets in tax administration: one data set may show a trader to be compliant but through other datasets, non-compliance will be identified. The Connect system has resulted in faster processing of documents, real-time detection of anomalies, and new opportunities for prevention and deterrence.

2.5 Canada Border Services Agency

Canada Border Services Agency (CBSA) has adopted enterprise data warehouse to manage big data, which it considers as an enabler Okazaki (2017). The data warehouse includes open source data to better manage risks related to cargo and people. According to Okazaki, the CBSA is exploring additional machine learning capabilities including facial recognition, lie detection and predictive modeling. The data warehouse and the advanced analytics team are being used to support CBSA's targeting and pre-arrival clearance. As Slowey (2017) puts it, border management has evolved and the ability of customs service to leverage data analytics and exploitation of data to guide decision making will result in better protection to society, improved border services, and growth of revenue. To fully leverage data analytics for decision making, CBSA is focusing on the following three areas: data governance, business intelligence, and advanced analytics Slowey (2017). Data governance focuses on data integrity, management of data, and automation of manual data entry tools. CBSA has moved from analysis based on siloes and has linked diverse datasets to have an enterprise-wide approach including culture shift to support data-driven organization.

2.6 Summary of lessons learnt

The experience from New Zealand, HMRC, and Canada Customs Service provides evidence that Customs can become intelligence-driven. This means that customs service can become proactive in risk management, efficient in trade facilitation, and facilitative in terms of secure and legitimate international trade. In this section, the paper highlights some pitfalls to avoid in the quest for becoming intelligence driven.

Focus on enterprise-wide adoption

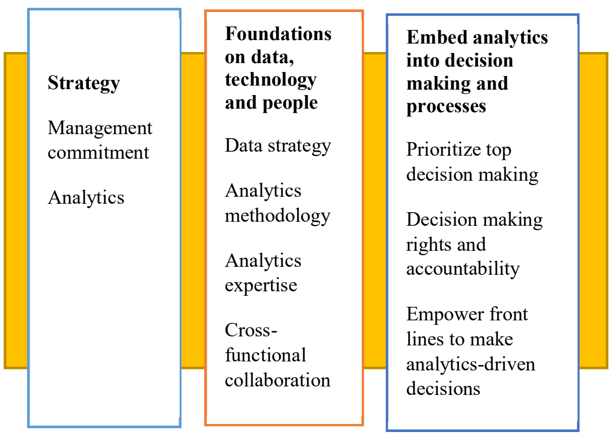
Organizations seeking to become intelligence-driven should develop strategies aimed at achieving enterprise-wide adoption of analytics for transformation to yield benefits. Focusing on some units of an organization becomes a hindrance from capturing the real value of AI Bisson at al. (2018). Customs may opt to start analytics in risk management with a small team of risk experts. After testing some algorithms and models, the analytics team may want to deploy the models to production environment. The model may work very well, flag risky declarations, save on time and enhance objectivity. However, if the customs officers at the borders are not familiar with the model workings, they would resort to using the manual risk analysis. Ultimately, the effort of analytics will not deliver the expected benefits. Organizations should, therefore, ensure enterprise-wide adoption of analytics and AI to realize benefits of transformation.

Organizations can customize the framework developed by Bisson at al. (2018) to achieve an enterprise-wide adoption of AI and analytics, see figure 4 below. The framework is categorized into three areas: strategy, capabilities, and embedding analytics into decision making. Strategy requires organizations to incorporate analytics into the critical functional areas. Customs' critical functional areas include cargo clearance, warehousing, valuation and tariff, border management, risk management, and revenue collection. Customs should therefore develop a strategy that incorporates analytics into these functional areas. Employees in each of these areas should then be equipped with the right tools and possess required skills to interpret results from analytics. At all levels of management, there should be commitment to support the new model with top management setting a clear vision of integrating analytics across all operational areas. Organizations should focus on capabilities including capabilities on data, processes, technologies and people. There should be a clear determination to ensure data literacy at all levels of management. This will ensure customs officers have same understanding and data analysis results are interpreted in the same manner. This cohesive approach will result in faster adoption of intelligence consumption and decision making. Capabilities in data strategy including data ownership, data governance, data quality, data maintenance, and technical requirements for the data should be clearly defined Bisson at al. (2018).

Organizations should invest in capacity building including acquisition of required technologies and talent as a starter to becoming intelligence-drive. However, organizations have to ensure the structure of the organization, culture, and work activities are aligned to support enterprise-wide adoption of AI, Fountaine et al (2019). To drive corporate wide adoption of AI technologies, Fountaine et al. (2019) proposes that organizations should focus on the following three transformations: First, organizations have to adopt cross-functional collaboration. Business teams ought to work hand-in-hand with analytics team to achieve broad AI adoption within an organization. Collaboration between diverse teams enables organizations to identify where new changes are likely to be required and plan in advance. Secondly, decisions have to be based on data. Wider adoption will require employees at all levels to deliberately trust machine-generated insights as a backbone for decision making. Human intuition and suspicions should be used to generate hypotheses that, once tested through algorithms, can be adopted in decision-making. Lastly, organizations should adopt agile, experimental, and adaptive model of operations.

The insights generated from analytics should then be translated into outcomes. Organizations should work on making analytics user-friendly and customized for different users Bisson at al. (2018). This requires deployment of correct tools including APIs, dashboards, recommendation engines and mobile Apps. There is a huge gap between organizations AI dream (what the organization intends to achieve with AI) and actual adoption of AI. Moldoveanu (2019) posits that this gap is not technical but organizational and cultural. Hall, Phan, and Whitson (2016) explain that for organizations to gain value from machine learning, there has to be a fundamental shift in culture to support data-driven decision making. There also exist a shortage of skilled personnel Hall et al. (2016).

Figure 4: Achieving enterprise-wide transformation: Source: Bisson et al. (2018)



Data

Data, organizational culture and skills set a key pillars in becoming intelligence-driven. Organizations should keenly monitor the data they collect and store. The cost of dirty data diminishes the returns expected from data analytics. An estimate by Redman (2016) puts the cost of dirty data at USD 3.1 trillion is US alone in 2016. Redman (2017) puts the cost of dirty data to be 15% to 25% of most companies' revenues. Lee (2018) report show that data cleaning can take up to 80% of analyst's time. A survey of 500 companies by Gartner (2018) show that data wrangling teams are growing and data scientists are taking 48% of time in data visualizations and 46% in data preparation. This evidence should challenge customs to rethink data quality as it seeks to become intelligence driven. Customs data is submitted through third parties (clearing agents) in Kenya. Most of the critical data is descriptive: describing the attributes of a product including manufacturer, brand name, size, composition material, packaging type, intended end-use among others. These attributes play a very critical role in establishing the correct classification of a product and determining the correct value of the product. Submitting data with quality issues can lead to misclassification or undervaluation. To benefit from intelligence and proactive decision-making, customs should emphasize on data quality for any document submitted to customs. The scope should cover declaration data, import declaration form, additional goods description data, manifest data, transit declaration data, and any other information submitted to customs. Third party data sources should be verified for data quality issues including systems that integrate with customs system. Redman (2016) confirms that improving data quality will yield the best data opportunity. Organizations seeking to become intelligence driven should strive to have data issues addressed as soon as they are identified and if possible, at the point where data is captured or sourced from.

The three customs services reviewed above have focused on data quality. Key data challenges including data quality, data privacy, data security and governance, data integration and preparation should be addressed at the source. A significant percentage of customs data is captured by merchants (clearing agents, importers or exporters) who may not have the same interest as customs. This agency issue introduces loopholes in data quality and data integrity. The merchants may not capture a complete dataset to customs, which customs require for classification, valuation, post clearance audit, and risk management. Missing some critical information may mean misclassification, or risky declaration is not detected, or full revenue cannot be collected. Customs in Kenya should develop strategies to improve data quality. This includes developing partnerships that can provide information direct from source, or third-party data providers who can validate what has been captured by a merchant.

Machine learning use case in Kenya customs

Google's DeepMind achieved a 15% cooling efficiency above what human could do, PayPal is using machine learning to detect and prevent money laundering, a Singapore insurance company is using IBM's technology to automate claims processing Brynjolfsson and Mcafee (2017). Machine learning creates a new capability for businesses to grow and optimize processes to serve customers better Wellers et al. (2017). Machine learning is not a new concept. It first appeared in computer science research in 1950s Hall, Phan, and Whitson (2016) but it is gaining momentum now due to growth and affordability of data storage and improvements on data processing capabilities. In business context, machine learning is a tool that can be applied in fields where data is to be applied and acted upon Awad and Khanna (2015). Are there machine learning use cases that customs can apply in its operations? This paper presents two possible use cases that customs in Kenya can apply.

2.7 Machine learning and risk management

Machine learning algorithms create a capability for businesses to develop models that can sift through huge data sets to identify patterns and trends that can be useful in risk management. Machine learning has potential to grow business revenues. In the baking industry, traditional risk management techniques are not yielding the expected benefits Aziz and Dowling (2019). Babel et al. (2019) estimates that machine learning improves risk management and could generate a value of over USD 250 billion in the banking industry. An analysis of machine learning models on credit markets have shown that deep learning models outperformed traditional benchmark models Son, Byun, and Lee (2016). Chorzempa (2018) explain that Chinese citizens have been lending informally to each other, making it hard for most of them to have formal credit history that could be used for formal lending. ZestFinance, a start up in the Chinese market is using customer data available on Baidu including buying history, search engine data, and other relevant data and applies machine learning techniques to determine the credit score of an individual Aziz and Dowling (2019). This evidence clearly show that machine learning has the potential to create new value.

Customs operations involve examination of data captured by merchants to determine the risk level and the course of action. This calls for a balance between compliance and trade facilitation. Customs administrations should develop intelligence-based risk controls that can facilitate faster clearance without a compromise on compliance. The data submitted by merchants has patterns that can be identified through machine learning. Classification algorithms can be deployed to separate or classify declarations based on their risk scores. The scores can be divided into different classes as maybe necessary and the actions to be taken defined in the system. Risk scores could be mapped to the risk channels used in customs: Red for high risk, Amber for medium risk, and green for low risk. Classification algorithms include random forest, decision trees, neural networks, and K-nearest neighbor. Risk management in machine learning has the potential to deliver new value including faster clearance of cargo, intelligence-based targeting, proactive identification and prevention of risks, and growth in government revenues.

2.8 Machine learning and recommender systems

Recommender systems have grown in importance with e-commerce and social media. On YouTube, the recommender algorithm gives a list of videos that are similar to what a user has viewed. In amazon, recommender algorithms show items that other buyers checked after viewing the item on display. On Facebook, recommender algorithms show a list of friends that you might know but you have not connected with. Recommender systems create value by helping customers discover new and relevant content based on their identified behavior. In customs, recommender system could be used in licensing, post clearance audit, and enforcement. In licensing, a recommender system can be trained to identify applicants who are compliant and using data captured in application, the algorithm will generate a recommendation score for all applicants. In post clearance and enforcement, recommender algorithms will identify taxpayers who, once audited, revenue yields will be high.

3. Recommendation and Conclusion

This paper has presented two frameworks that could be used by customs to develop a strategy to become intelligence-driven. In addition, lessons that customs in Kenya can learn from other customs have been presented and a summary presented. It is important to ensure data collected by customs is of the required quality to avoid the costs and time-wastage associated with dirty and incomplete data. This paper recommends that customs should identify the future desired state and work backwards to the current state to identify what is required and what is missing to actualize the future state. This gap analysis will then be modeled into a strategy document that will define the operations of customs, where changes are required, where new skills and capacity is required. All officers within customs will need to be sensitized on the future state, trained, and their capacity enhanced to achieve the desired future state. In conclusion, intelligence-driven customs will result into safe and protected society, facilitation of legitimate trade, collection of government revenue in an efficient manner that ensures optimal allocation of resources.

References

[1] Awad, M., and R. Khanna. (2015). Machine Learning in Action: Examples. *Efficient Learning Machines*.

[2] Aziz S., M. Dowling (2019) Machine Learning and AI for Risk Management. In: Lynn T., Mooney J., Rosati P., Cummins M. (eds) Disrupting Finance. *Palgrave Studies in Digital Business & Enabling Technologies*. Palgrave Pivot, Cham

[3] Babel, B., K. Buehler, A. Pivonka, B. Richardson, and D. Waldron (2019). Derisking machine learning and artificial intelligence. *McKinsey & Company*, February 2019.

[4] Batistič S and P. Laken (2019). History, Evolution and Future of Big Data and Analytics: A Bibliometric Analysis of Its Relationship to Performance in Organizations. *British journal of management.* https://doi.org/10.1111/1467-8551.12340

[5] Bisson, P., B. Hall, B. McCarthy, and K. Rifai (2018). Breaking away: The secrets to scaling analytics. *McKinsey Analytics*. May, 2018.

[6] Brynjolfsson, E. and A. Mcafee (2017). The business of artificial intelligence: What it can — and cannot — do for your organization. *Harvard Business Review*, July, 2017

[7] Chorzempa, M. (2018). China Needs Better Credit Data to Help Consumers. Peterson Institute for International Economics. accessed from https://www.piie.com/system/files/documents/pb18-1.pdf

[8] Desjardens, J. (2018). How Long Does It Take to Hit 50 Million Users? *Visual capitalist.* Available at https://www.visualcapitalist.com/how-long-does-it-take-to-hit-50-million-users/

[9] Desjardins, J. (2019). How much data is generated each day? *World Economic Forum*. Available on: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

[10] Devakunchari, R. (2014). Analysis on big data over the years. *International Journal of Scientific and Research Publications,* Volume 4, Issue 1, January 2014 1. ISSN 2250-3153

[11] Dyckhoff, A. L., D. Zielke, M. Bültmann, M.A. Chatti, & U. Schroeder. (2012). Design and implementation of a learning analytics toolkit for teachers. *Educational Technology & Society*, 15(3), 58–76.

[12] Fjeldstad, Ø. D., C. Snow, R. Miles, and C. Lettl (2012). The architecture of collaboration. *Strategic Management Journal,* 33(6), 734–750. doi:10.1002/smj.1968

[13] Fosso S., S. Akter, A. Edwards, G. Chopin and D. Gnanzou (2015). How big data can make big impact: findings from a systematic review and a longitudinal case study, *International Journal of Production Economics*, 165, pp. 234–246.

[14] Fountaine, T., B. McCarthy, T. Saleh (2019). Building the AI-Powered Organization. *Harvard Business Review*, July–August 2019

[15] Galbraith, J. R. (2010). The multi-dimensional and reconfigurable organization. *Organizational Dynamics*, 39. doi: 10.1016/j.orgdyn.2010.02.001

[16] Gantz, J and D. Reinsel (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in Far East. *IDC Iview*

[17] Gupta, M. and J. George (2016). Toward the development of a big data analytics capability, *Information & Management,* 53, pp. 1049–1064.

[18] Hall, P., W. Phan, and K. Whitson (2016). The Evolution of Analytics: Opportunities and Challenges for Machine Learning in Business. *O'Reilly Media*, Inc., CA.

[19] Henke N., A. Libarikian, and B. Wiseman (2016). Straight talk about big data. *McKinsey Quarterly*.

[20] Jørnø, R. L., & K, Gynther. (2018). What Constitutes an 'Actionable Insight' in Learning Analytics? *Journal of Learning Analytics*, 5(3), 198–221. https://doi.org/10.18608/jla.2018.53.13

[21] Khan N., I. Yaqoob, I. Hashem, Z. Inayat, W. Ali, M. Alam, M. Shiraz, A. Gani. Big Data: Survey, Technologies, Opportunities, and Challenges. *The Scientific World Journal,* Hindawi Publishing Corporation. Volume 2014, Article ID 712826, 18 pages. http://dx.doi.org/10.1155/2014/712826

[22] Kolbjørnsrud, Vegard & Andersen, Espen & Johnson, John & Ragnvald, Sannes. (2018). The data-driven organization: *Intelligence at SCALE*. 10.18261/9788215031583-2018-03.

[23] KPMG (2019). Tax Intelligence Solution. New Zealand Customs and Trade. Available on https://home.kpmg/content/dam/kpmg/nz/pdf/Aug/tax-intelligence-solution.pdf

[24] Lee, C. (2018) MIT Sloan Conference Rundown - The Challenges of Big Data Analytics. *Discovered intelligence*. March 29, 2018

[25] Lim, C., & K. Kim (2014). Information service blueprint: A service blueprinting framework for information-intensive services. *Service Science*, 6(4), 296–312. https://doi.org/10.1287/serv.2014.0086

[26] Lim, C., K. Kim, M. Kim, J. Heo, K. Kim, and P. Maglio (2018). From data to value: A nine-factor framework for data-based value creation in information-intensive services. *International Journal of Information Management.* 39 (2018) 121–135.

[27] McAfee, A. and E. Brynjolfsson (2012). Big data: the management revolution, *Harvard Business Review*, 90, pp. 60–68.

[28] McKinsey (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute.*

[29] Michael K.,and K. W. Miller. (2013). Big data: new opportunities and new challenges, Editorial: *IEEE Computer*, vol. 46, no. 6, pp. 22–24, 2013.

[30] Moldoveanu, M. (2019). Why AI Underperforms and What Companies Can Do About It. *Harvard Business Review*. March 12, 2019

[31] New Zealand Customs Service (2014). Joint Border Analytics. Available on: https://www.customs.govt.nz/about-us/your-information/customs-information/joint-border-analytics/

[32] New Zealand Customs Service (2015). Accessed from: https://www.customs.govt.nz/about-us/about-customs/our-history/

[33] New Zealand Customs Service (b). Privacy by Design. Available on https://www.customs.govt.nz/about-us/your-information/your-personal-information/privacy-by-design/

[34] New Zealand Government (2012). Formal Review of the New Zealand Customs Service. State Services Commission, the Treasury and the Department of the Prime Minister and Cabinet. Available on https://ssc.govt.nz/assets/Legacy/resources/pif-customs-review-oct12.pdf

[35] Okazaki, Y. (2017). Implications of Big Data for Customs: How It Can Support Risk Management Capabilities. WCO Research Paper No. 39.

[36] Pemberton C. (2018). Key Findings from Gartner Marketing Analytics Survey 2018. Gartner, May 16, 2018

[37] Pyle D., and C. San Jose (2015). An executive's guide to machine learning. *McKinsey Quarterly*.

[38] Redman T. C. (2017). Seizing Opportunity in Data Quality. *MIT Sloan management review*. November, 2017

[39] Redman, T.C. (2016). Bad Data Costs the U.S. $3 Trillion Per Year. *Harvard Business Review*, September, 2016

[40] Rigney P. (2016). The all seeing eye - an HMRC success story? Available at https://www.ifa.org.uk/media/653935/Tax-HMRC-Connect-system.pdf

[41] SAS (2014), Recovering £7 Billion in Additional Tax Revenues, HM Revenue & Customs. Available athttps://www.sas.com/content/dam/SAS/en_gb/doc/CustomerStories/hmrc.pdf

[42] Sclater, N., M. Webb, & M. Danson. (2016). The future of data-driven decision-making. *JISC*, Bristol

[43] Slowey, C. (2017). Data analysis for effective border management – the Canadian experience. WCO News 82, Dossier

[44] Son, Y., Byun, H., & Lee, J. (2016). Nonparametric machine learning models for predicting the credit default swaps: An empirical study. *Expert Systems with Applications, 58,* 210–220.

[45] Teece, D. J., G. Pisano, and A. Shuen. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal,* 18(7), 509–533. doi:10.1002/(SICI)1097-0266(199708)18:7 <509::AID-SMJ882>3.0.CO;2-Z

[46] Wellers, D., T. Elliott and M. Noga (2017). 8 Ways Machine Learning Is Improving Companies' Work Processes, *Harvard Business Review*, May 31, 2017