www.ijecs.in

International Journal Of Engineering And Computer Science
Volume 10 Issue 12 December 2021, Page No. 25731-25755
ISSN: 2319-7242 DOI: 10.18535/ijecs/v10i12.4680

Big Data and Predictive Analytics in Government Finance: Transforming Fraud Detection and Fiscal Oversight

Vamsee Pamisetty

Middleware Architect, ORCID ID : 0009-0001-1148-1714

Abstract

Governments around the world are experimenting with Big Data and predictive analytics. They deploy various software applications like predictive policing, fraud detection, and capacity demand prediction while at the same time developing and investing in broader data analytical infrastructure and analytical skill sets. Implementing Big Data and predictive analytics can be a challenging endeavor, however, as these analytics often rely on open-source algorithms that are unsupervised and black-boxed. Equally challenging is how government institutions endeavor to use a waterfall approach from exploratory to model predictive analytics, yet how often predictions are only perfunctory and unempirical [1]. To shed light on how government finance organizations conceptualize and prepare for such analytical disruption pertaining to predictive analytics specifically, data processes, stakes and concerns were articulated based on in-depth interviews with 23 analysts who work with predictive analytics in a regional government agency. Descriptive coding of the interviews revealed that changes to data processes are being prepared or enacted, but many foreseen stakes and concerns about changes to both data processes and knowledge processes remain unresolved. New agendas to address such issues and better understand the approaches adopted were proposed.

Governments across the world are aiming to exploit Big Data and associated predictive analytics to govern more effectively and efficiently. These analytics come in many sorts and varieties. In government finance, the topical applications of predictive analytics have to date mainly been found in fraud detection, capacity demand prediction, budget revenue prediction, and the prediction of homelessness and recidivism. A plethora of software applications built on open-source predictive analytics algorithms exist, encompassing packages for forecasting demand, and estimating regression models including linear and logistic types. However, there is some hesitancy in adopting most of these analytics, as open-source predictive analytics algorithms are rarely supervised and almost always deployed as black-boxed. Black-boxified analysis is countercultural to the emancipation and democratization of knowledge advocated in government, as well as other more mundane concerns about accountability and validity. With black-boxification work piling-up on agency knowledge processes, concerns arise about how this analytical work is handled, displayed, devised and/or aggregated to produce knowledge that meets government quality expectations of reproducibility, replicability, auditability, and trainability.

Keywords: Big Data, Predictive Analytics, Government Finance, Fraud Detection, Fiscal Oversight, Data Mining, Risk Management, Real-Time Monitoring, Public Sector Analytics, Financial Transparency, Anomaly Detection, Machine Learning, Data-Driven Decision Making, Tax Fraud Prevention, Digital Governance

1. Introduction

Big Data and AI will have a profound transformational shift for governments. Examples of Big Data applications in government are evinced with regard to the entire policy process. Challenges to the uptake of Big Data and AI in the public sector and expected implications occurring as a result of such challenges are explicated. Governmental applications of Big Data and AI and the field of public administration and policy have so far received little attention but are increasing in size and prominence. Big Data and AI are thought to have a global reach and thus exert a fundamental transformational impact throughout society.

While the use of data and, moreover, the deliberate use of data in the public sector is by no means a new phenomenon, Big Data is different and has the potential and actual attributes that affect aspects of the theoretical and practical considerations of the decision-making in the public sector. The confluence of the data revolution on one hand and the development of more advanced analytics across diverse professional domains on the other hand constitutes the emerging ABI paradigm which is fundamentally different from the traditional Paradigm. Its consideration by governments' functions and operations is in principle and practice expected to transform the workings of government. Robots and automation are changing the provision of public services, making many more "mundane" jobs redundant within the civil service [1].

There is therefore a pressing requirement for the civil service to embrace big data, not least in order to keep pace with the tremendously rapid pace of change in technology that is reshaping the workforce at large. The challenge for government is on the "what" question and to better use citizen data resources. Governments are at the starting point of considering how to use data to improve the public services they provide, to target who is likely to need services most, and ultimately to "tailor those services more accurately." The essential characteristic of the new models of big data is seen to be about the three "V": Volume, Variety, and Velocity. At the most basic level, Big Data is about the volume of information, the variety of the different data sources and types, and the velocity – namely the speed of creation, storage and dissemination of data, often in real-time.

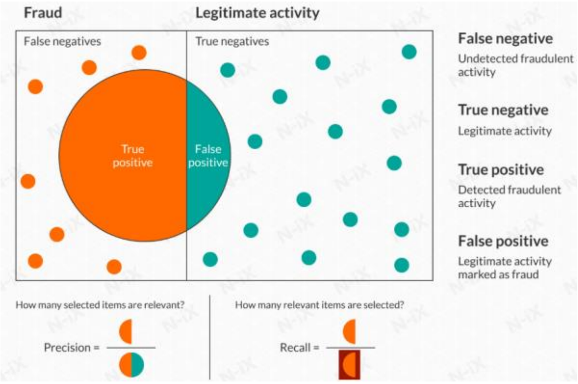


Fig 1: Implement fraud detection with big data analytics

2. The Role of Predictive Analytics

Governments across the globe are now adopting predictive analytics tools to improve decision-making in the areas of public policy, regulatory inspection, program delivery, citizen engagement, and financial performance and risk analysis. Many predictive analytics tools are readily available or can be purchased off the shelf, leading to a frenzy of purchasing activity by Government of Canada departments, agencies, and crown corporations [2]. However, there is little guidance on evaluating, preparing for, or implementing these applications effectively, let alone integrating them into a public sector environment that has fundamental differences from private organizations.

Predictive analytics is not new. It encompasses a number of approaches or methods that have existed for decades. Applied in business, sport, and even credit scoring, predictive analytics is considered "advanced analytics" or one of the classes of "Big Data" analytics in CRM and, more broadly, "data mining" or "knowledge discovery" in the literature on artificial intelligence and statistics. The term "predictive analytics" is most commonly used in a commercial context, with an emphasis on pre-packaged software used for customer retention/sales forecasting, although it is nonetheless technically precise, capturing the various classes of approaches encapsulated by other terms in broader analytical reviews [3]. The specific techniques broadly underlying the class of predictive analytics tools are also well known. The literature on the former is extensive, offering an almost-overwhelming flood of implementations. There is little evidence of such activity outside of the finance and accounting areas, however.

In government financial management, however, there are important and fundamental differences with a significant impact on the type of tools purchased and the context in which they will be implemented and used. Canadian federal government departments are much larger organizations than most companies targeted by predictive analytics vendors, which impacts the modification and application of tools to fit Canadian government departments (and the organizations of similarly-sized governments elsewhere). Data access (or lack thereof) will also impact analysis. The tools and models designed to improve customer retention/sales forecasting also depend on access to customer level data, hand insurance renewals, a role frequently filled by publicly available data on prior claims and accident history in the case of car insurance. Governments, with the exception of tax compliance, generally do not have access to data which is best acquired as part of a privately-held database.

3. Fraud Detection in Government Finance

Although much research in the area of fraud detection has been on the application of data mining or machine learning methods, there are still several areas that deserve more attention. One of these is the complete understanding of how and why fraud detection methods work best for some datasets and not at all for others. Another area of interest is the understanding of the effects data quality has on fraud detection. Essentially, it is desirable to know how much is it possible that high-quality data still lead to poor predictive accuracy [4]. The application of statistics or stochastic models within the general field of data analysis for financial purposes probably goes back to the beginnings of both disciplines in the 18th and 19th century. Topics in this area range from equity valuation and corporate bond rating to asset allocation or risk analysis, which in most cases imply both estimation of parameters and prediction of future behaviors or returns. Informed members of the business community employ the thoughts of bond experts and investment theorists to minimize expected loss [5]. On the other hand, non-informed members will lag behind informed players on price competitions, or gain satisfaction from observing the mistakes of experienced bettors and not repeating them.

3.1. Types of Fraud in Government

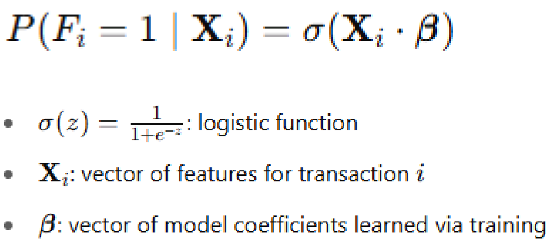
Fraud can be broadly categorized into two types: civil and criminal. Civil fraud means siphoning off money by ex-employees, contractors, or other agents, which may subject violators to loss of contracts, reputational damage, or civil suits. There is a growing need for computer-based fraud detection systems that can effectively detect and flag fraud before it continues for several consecutive months. Major fraud scandals have shocked the very public to private companies, resulting in bankruptcy when the actual financial condition comes to light. While moving fast and break things may be an acceptable paradigm for Silicon Valley startups, government institutions cannot take the same approach. The consequences are dire when large amounts of taxpayer money are misappropriated or misused. Forensic accountants, investigators, and auditors do not have the luxury of making mistakes on their paths to catch fraud that has already taken place. Highly sophisticated and advanced computer-based fraud detection systems and algorithms have been deployed by financial institutions, credit card companies, and insurance companies. Various machine learning and artificial intelligence frameworks have been rigorously developed for this purpose [6]. However, research in fraud detection is lacking in the government finance arena, which is why a number of machine learning algorithms have been proposed with predictive performance.

Fraud in credit card transactions can either be the impersonation of an existing customer or the creation of a fresh account with false references, fake email, or fake addresses. Here, fraud detection means identifying transactions which are unlikely to have occurred. The identification goal is to effectively determine the legitimacy of a transaction with limited access to historical records [4]. Transactions come in a stream and cannot be grouped together. With a vast amount of data and various features for each transaction, base classifiers are trained using historic data to flag transactions where the event is rare, which requires a general approach that can detect such events across various domains. Explicitly addressing the challenge, fraud can be characterized using an event structure composed of complexity, uncertainty, and impact. This structure outlines a testing framework for the generality of event detection algorithms.

3.2. Traditional vs. Predictive Approaches

Predictive analytics is the most commonly used approach in both contexts. In prediction and analysis of business data, the prediction of future behavior relies on data points that contain time series characteristics. Most predictions either recur on the same time series (e.g., the user’s propensity to buy product x in a given week) or are new point predictions formulated as a time series completion task intended to identify the next time slot and predict the next point value for the entire series. In this context, analytics is concerned with what will happen with each individual point in time (e.g., will it become a purchase or not?).

Eqn.1:Fraud Probability Prediction Model



$$P(F_i = 1 \mid \mathbf{X}_i) = \sigma(\mathbf{X}_i \cdot \beta)$$

- $\sigma(z) = \frac{1}{1+e^{-z}}$: logistic function
- $\mathbf{X}_i$: vector of features for transaction i
- β: vector of model coefficients learned via training

Common examples of how predictive data analytics is used in a business context can be found in e-commerce companies that offer online retail opportunities. For a user that browses for some product on a retailer’s e-commerce website, they will likely view ads with similar products on different websites of the same ad provider or within the search engine they use to seek information on the web. This is possible because of a recommendation system that first predicts the catalog search interest of the customer by their past click/watched transactions using machine learning methods, and then selects a subset of items to recommend by further filtering them. A well-known real-world recommendation problem model is the multiplicative user-Item model, where a matrix that records user interaction with item ads in the past time series is stated.

Common public contexts where prediction is observed are in governmental and utility company installations. Various data sources, data types, and data formats can be involved in analysis of customer data in this context. For example, in order to optimize the time it takes to repair a faulty water pipe, a government agency may gather information concerning the city’s area, its demographic features, the pipe age and maintenance history, the sensor reading time series that capture flow and mobile insertion detections, all having second resolution [5].

4. Data Sources for Predictive Analytics

The utilization of data across all sectors has been growing continuously, forced by both pressure from stakeholders and the competition. It seems obvious that government should also engage in exploiting the data available in their own environment, in a way that is comparable to the private sector. Government practice has also been transforming gradually, increasing the agility of service delivery in response to an ever-increasing change in requirements from citizens. In recent years, Predictive Analytic (PA) applications in government domains are gradually growing, increasing the curiosity of many public organizations to adopt such sophisticated analytics methods. Enormous amounts of data are generated by public agencies every day, such as budget reports, expenditure reports, and revenue reports. These public financial reports contain valuable financial information, such as financial service information and fiscal resources information. In order to take full advantage of such public financial data for effective applications, it is critical for agencies to build advanced capabilities to manage massive financial data collected from various public sources.

Governments have a responsibility to provide reliable and valid information to sustain rapid data-driven decision-making, which is frustrated by the taxonomies of financial data diversity and government data silos. Big Data (BD) refers to datasets that are too complex, large, or dynamic to be effectively managed, processed, or analyzed using traditional data management architectures, tools, or processes. Big Data is usually characterized by three Vs: volume, variety, and velocity, which means that data generated becomes more high-dimensional; heterogeneous; and dynamic than it used to be. The rapid growth in the size of data poses a serious challenge for governments in managing, accessing, securing, integrating, and analyzing data, due to the rapid generation of diverse input, which outpaces the ability of systematic analysis [7].

As a remedy, this work aims to provide a framework called PA-GVSP, in which an automated and unified Government Financial Service Platforms (GFSP) is constructed to leverage authoritative and heterogeneous financial data for financial service and business improvement. In addition, well-performed analytical pipelines are introduced along with State Expectation Formulation Processes (SEFPs) to enable multi-stage, multi-style, and cross-disciplinary analytics, which can purify, classify, cluster, summarize, and interpret financial information from massive government reports.

4.1. Public Records and Databases

Predicative Analytics in government finance has seen an escalating interest in the last decade. Finance agencies need to expend new technologies and techniques with large data sets. There are many considerations for predictive government finance success. Public records and databases are resources used by government finance agencies to run predictive analytics programs. Government finance is important for the agencies to operate. Finance transactions are public records and can be viewed as databases for cities, counties, and school districts. Many tools and applications in analytics programs are relevant because public finance records are influenced. An example is Early Warning System. Using public sector databases, local governments in many states are able to predict when an entity is likely to face cash flow shortages. Cash flow predictions will be valuable and it is predicted, those cities, counties, and school districts that will not attain predictive analytics on their own will have difficulties attaining it via consultation in the future. Government finance agencies are likely to expend resources on obtaining tools and auxiliary support such as databases. Starting a predictive analytics effort needs simple items to illustrate the power of predictive analytics. Public sector databases can be used.

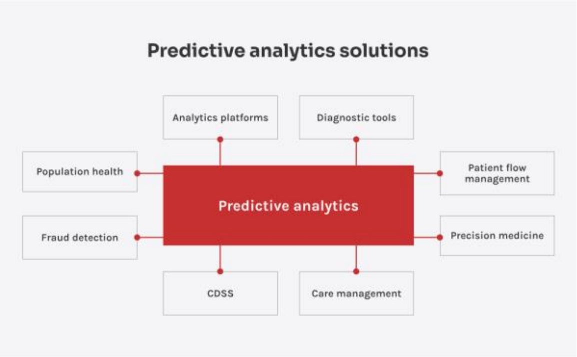


Fig 2: predictive analytics in healthcare

Data is the bloodline of predictive analytics. Specialized public data databases are available to local government finance agencies. Workers within the agencies should not be tasked with gathering and structuring public databases. Specialized items are often available from local or external sources. Many specialized tools and auxiliary supports are available to assist obtaining needed public databases at an appropriate cost. Most predictive analytics programs will require another type of non-public data beyond public databases. Using only broad open source queries will not be useful in the initial phases of predictive analytics efforts. The non-public dataset will have to be gained from a connection of traditional communication. Non-public databases will need to be fleshed out into a more analyzable practical numerical condition before use. Numeric datasets in R or SQL compatible form with discrete and continuous variables will work best for predictive analytics. Public datasets available in flat Excel CSV files would offer an initial starting point but are not ideal. Many government datasets are better suited for presentation format rather than analytic use [7].

4.2. Social Media and Online Behavior

The growing popularity of social media, along with a substantial rise in online actions produced by Internet users, has generated massive volumes of textual data known as "Social Big Data." Companies can extract valuable insights from customers' online behavior by leveraging big data analytics to classify customer behavior and make business decisions accordingly. Detecting customers' online behavior is typically referred to as "customer online behavior classification," while datasets of customers' online behavior are referred to as "customer online behavior datasets." Oftentimes, customer online behavior datasets are expected to be generated by different organizations at different time spans; hence, a primary concern is the portability of predictive models built from one dataset to another. Such concerns are referred to as "predictive model transferability" issues. Nowadays, these issues are gaining more attention as customers' online behavior has become more accessible.

As customer online behavior datasets are typically class imbalanced, the predictive models built from those datasets are expected to suffer from the model imbalance issue. Regarding this issue, a recommendation framework of resampling methods called "RAmb" is proposed to augment the minority class data by imitating data points near the decision boundary. In the real world, companies need to analyze customer online behavior originated from different customer online behavior datasets, and existing solutions focus on model transferability from a single perspective. Therefore, it is reasonable to consider using different perspectives collectively, e.g., how factors, such as data comprehensiveness, connectivity, and model robustness, enable model transferability across different perspectives.

The rapid growth of social media has caused an increasing amount of social big data, enabling customer behavior understanding to be widely adopted for optimizing marketing and promotional efforts. However, analyzing customer online behavior data requires solid understanding of the domain knowledge and sophisticated machine learning skills. This requires the collaboration of computer scientists and marketing researchers. Thus, there is a need to build a system with comprehensive functionalities for marketing researchers to conduct the whole process of customer online behavior classification without a deep understanding of advanced machine learning knowledge.

5. Machine Learning Techniques in Fraud Detection

In November 2009, the world watched an exciting report aired by CNN. Pirate hunters in the Indian Ocean had a series of good news to share. Their SAT phone calls indicated that they had attacked and boarded a large mother ship in the high seas. When the ship sunk, late reports mentioned that dozens of pirates and three hostages had all gone to the bottom of the sea; the pirate scourge had received a pair of heavy blows that would put it on the verge of eradication.

This may seem like exciting news. But to be logic-minded analysts in hacker identities in the Asian Pacific or financial fraudsters, this decidedly uninviting news was the main news report to watch hungrily. Applied maths and algo trading logic had just changed pirates' garb; demoistic procedures so rewarding on the stock market had been placed on the high seas.

On the day of the TV report, a seminar was hastily organized at a computer science department in a major Asian city, where a speaker gave a well-prepared talk on the mathematics of algo trading with long sequences of tiny graphs, and the audience's attention was riveted. The identical may well be said about the maths of fraud detection before that day's four-part episode of the new TV series.

Fraud is asymmetrically informative. It "unhelps" the expense reimbursements of the needy who observe the real expenses a social welfare fund. Analysts, allocators, and evaluators (hidden variables) share identical legislative and financial data. Such data as may assist the economically includable eligible analyze (CI criteria) variable are investigated, e.g., the relationship between the variable and the potential authorised expenses. Other data is defined as non-CI data. The work is couched in precept-advised and mathematically tractable, unsupervised learning models. A Challenging Route Forward section delineates issues that still need discovery and mathematical treatment.

Research into the prediction of fraudulent activities or anomalous behaviours in business decisions has burgeoned in the last two decades. Intelligent analysis of transactional and historical data, in the domains of finance, telecommunications, social media, and the web, is increasingly relied upon for predicting fraudulent financial transactions and delinquencies [4]. Artificial Intelligence (AI)-based learning techniques have parity with, and frequently outperform, traditional statistical techniques for fraud detection, and for classification and prediction in general.

5.1. Supervised Learning Methods

Within the school of supervised learning methods, a shorter list of topics with variant degree of critique. Random Forest Classifier (RFC) Supervising methods with Text Data, Gradient Boosting Machines, Support Vector Machines (SVM) K-Nearest Neighbors (KNN). The latter ones with combined references or providing implementations as sources.

At own option, Random Forest Classifier (RFC) with up to 4 & 5 y-o-y predicators can be very evaluate them on accuracy using models and predictions. As a rule, limit the description of the models and avoid statisticians terms; limit the number of figures and visualizations; group the industries compatibly. No recommendation letter from the consumer side is required as most probably it will be just publicly issued rating, or from analyst's provider.

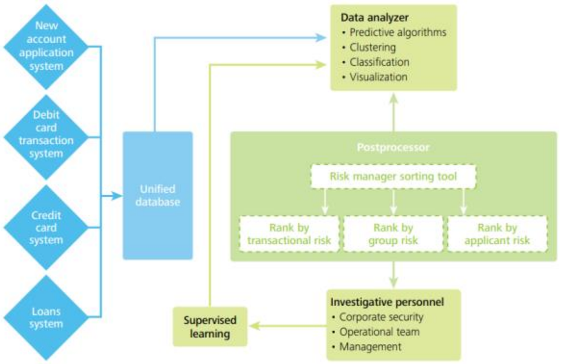


Fig 3: Predictive Analytics in Finance and Investment Banking

The task is to suggest model implementations with apparently higher accuracy on own holdouts, for augmenting predictions and analyze broader prospective adding new features in connection especially to text data. As other supervised methods, RFC was successful but does take a very long processing. Initially, this task was only forecast demands of ISR with months ahead; then with breaking down weeks, days hours and instances based on organization's level datasets, and as government, many renew requests comes after, up to inquiries nature. The greater number of the companies, the more new datasets are collected, and if just in one sitting requests were only 149, in the broadest sourcing range, it reached 7878 in two successive sittings.

Some Railroad Data Forecasting Models were created and tackled too; adopt its title for a model in US railroads. There is gauge note to estimate accuracy and choose it likely. Seek local supporting boost of internet/telecom providers. If too long holdout procedures, speed it up using a set of 25-. With more forecasting model implementation examples taking from other papers; better explanation to mixing borrowings from sources, with remarks on more important ones. Might take requests, signals, and suggestions references or illustrations just analyses are on other companies and their usages of parallel models outside the countries.

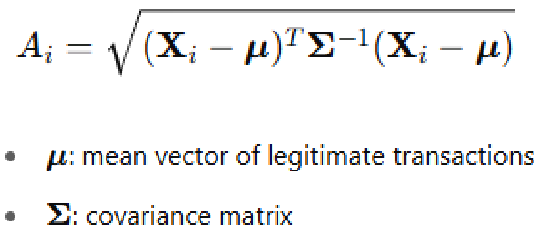
5.2. Unsupervised Learning Techniques

As described in subsection 4.1, actions to improve analytics need to be developed based on the visualizations provided in the interactive BI dashboard. Table 3 identifies six points at which it is possible to improve analytical insights. Four points concern utilizing unsupervised learning techniques for analyzing empirical data. Due to the lack of labels and the ground truth in the corresponding datasets. Thus, the selection of unsupervised learning algorithms depends primarily on the target domain and the nature of the data generated therein. Unsupervised learning in the domain of public finance contains the largest variety of algorithms, which have been used primarily to classify data for further processing regarding detected outliers. For instance, large datasets of tax declarations are examined using hybrid unsupervised machine learning algorithms, which cluster and select relevant features that produce the best results on tax model estimations [8]. The agglomerative clustering method was proposed due to its interpretability and sensitivity analysis of clustering parameters. Non-clustering methods were used as a unifying framework that could improve the utility of analytical tools for the administration of public finance. It should be noted that input features and output features of modeling approaches cover continuous, categorical, and binary types. Discriminative approaches using supervised learning techniques have early gained in popularity and have become industry standards with proven results as state-of-the-art models [5]. As a result, there is a variety of applied algorithms and packages for these approaches.

The potential of unsupervised learning techniques for generating new data-driven features was neglected even for the simplest case such as single-stage dimensionality reduction. It is possible to achieve data exploration via simple outlier detection using clustering or density-based methods, which significantly improves data visualization without enriching drought-prone features, as conducted in the latent space of sophisticated deep generative models. In addition to the thematic area concerning data preprocessing, there remains a significant amount of work on unsupervised learning in the applied area, e.g. forecasting or kernel-based methods for the exploratory and pre-processing steps. The absence of rigorous validations, the remaining biases, and the lack of package libraries are among the reasons for the slow advancement.

Unsupervised outlier detection is rarely found in bill-payment models, which is probably due to the large amount of clustering-based approaches, which are prone to overfit large datasets, and the simplifications of supervised modeling pipelines. Hence, the majority of showcased models identify outliers based on predicted labels in a downstream step, while expert knowledge models using monthly aggregated warning days in a simple threshold have gained larger popularity among practitioners.

Eqn.2:Anomaly Detection Score Using Unsupervised Learning



$$A_i = \sqrt{(\mathbf{X}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu})}$$

- $\boldsymbol{\mu}$: mean vector of legitimate transactions
- $\boldsymbol{\Sigma}$: covariance matrix

6. Challenges in Implementing Predictive Analytics

The performing of government finance predictive analytics is at the crossing point of different concepts. It contains challenges of big data, data analytics, predictive analytics, government data, and government finance. Specifically, the government finance field involves many types of data heterogeneous in structure and format being created stored in different information systems with quantitative, qualitative, and free text types. Especially in recent years, in response to citizen demands for enhancing government transparency, accountability, and participation, government further publishes various kinds of data on open data portals, including budget and accounting raw data, fiscal regulations, financial statements, audit reports, civil expenditures, pledges for public finance, etc. [5]. These outdoor data are different from traditional government data, which are normally structured/populated and stored in a database. With the advent of information technologies, datasets published online are massive, heterogeneous, noisy, semi-structured, and mostly raw. In this case, how to deal with such big data challenges, including data volume, variety, veracity, and value, is vital. Finance data analytic approaches are required to be able to systematically analyze massive data for major budget and accounting issues of government finance.

The government budget and accounting finance predictive analytics task aims to automatically analyze the function-values of new rising down-budget categories of government finance in terms of existing on-budget categories. On-budget categories are the financial subject in a fiscal budget corresponding to down-budget categories, predicting tasks ensure that on-budget and down-budget categories match with high confidence. Such predictive analytics needing better understanding of the government finance domain is a challenge. Also, civil expenditures, a combination of government finance accelerating transparency and accountability, personal privacy applications generating out explicitly, are further open interests for predictive analytics. For decades of years, government finance has been researched well, including data gathering and quality evaluation, presented in various forms in multiple degrees of details and structures. However, the transferred knowledge from traditional statistical education to predictive analytics is limited, hard knowledge transfer for qualitative analytics.

6.1. Data Privacy Concerns

Big data analysis provides significant benefits for governments, but it also raises significant privacy concerns. Massive amounts of initially innocuous data on individuals can be collected and combined to create new sensitive facts about these individuals [9]. The latter concept has been described as “the limits of inference" or "the boundaries of the inferable." The analysis of big data regarding their weather can lead to a conclusion about people's location, which would be private information that was not volunteered. The concern arises on where to draw the red line.

The privacy legislative framework developed in the past decades and on which the individual privacy protection policy in most jurisdictions rests is based on the Fair Information Practice Principles (FIPPs). Several of these principles are based on the collection of personal information (PII). The vulnerability model is based on the FIPPs and on the definition of PII. However, with continued data collection that will become ubiquitous, vast amounts of PII can be created, even if it is not actively or volitionally collected. Other elements of the current privacy framework based on the FIPPs must be addressed too.

Data minimization is an essential element of the current privacy approach, which is directly associated with the vulnerability model. Data minimization is implemented through the various FIPPs that require maximization of transparency and individual control. However, these principles are strained by the current technological and organizational landscape. There are transition challenges from a scenario in which there is a potential limitless collection of innocuous data to a new scenario in which there is a limited collection of sensitive data [10]. As many organizations, including governments, collect big data regarding the environment, trends, and other goals, these technologies can be used to collect and combine data regarding individuals. This can turn open data collection and accumulation by organizations into a private information surveillance technology.

6.2. Integration with Existing Systems

The successful adoption of predictive analytics requires an ecosystem capable of integrating and processing diverse datasets while preserving the integrity of existing systems. Agencies need new data collection and management methods to better understand the predictive capabilities of their data and its potential uses. Predictive analytics should also be incorporated as a core component of the technology review process and integrated into long-term technology planning, modernization, and procurement. Information management executives are generally responsible for implementing predictive analytics, along with the existing processes for standards and compliance—these processes and policies should be updated to avoid producing poorly designed, insecure, incomplete, or incompatible applications. Establishing a protocol for how predictive analytics products will be implemented or made available for use in production environments is also critical. This procedure should apply to analytics products designed for one-off use in static reporting programs and for production applications. Predictive analytics often has accessibility requirements that static reports do not have. Most other states have already begun enforcing these types of transparency requirements on the data and methods used for statistical applications. Therefore, it is critical to invest time and energy beforehand in how well-established technology standards and practices can be applied to predictive analytics. Despite significant investments by most states in modernizing their primary enterprise systems, substantive legacy systems are still commonly used across agencies. These systems often serve as the primary source of data for analytic products but can also delay the development and implementation of predictive analytics. In cases where exploratory analytic products are built using exogenous data sources, additional considerations for how to integrate those units of analysis with the agency's existing architecture need to be made.

7. Case Studies of Successful Implementations

This paper outlines how both public and private choices, as well as risks and benefits, are evident in big data initiatives, and also how uncertainty is considered a barrier to reporting to support big data initiatives. Examples of this can be chronicled on a number of levels. The initial example shows how city personnel investigated an early and small project in partnership with a private organization to use big data in finance operations [11]. A more advanced example reveals how a city government with ample expertise is preparing a big data project to determine if there are social or vitality issues in neighborhoods based on existing and/or added data from open sources. A national agency that does not primarily or clearly have a data role is a case that is not all that common when investigating big data projects. In this case, progress is “slow,” as much depends on how data and how parties are involved. The project is also somewhat of a catch 22 because it proposes or mandates to collect data, whereas it itself is being asked to count.

In the process of analyzing specific projects and the readiness to report, three themes stand out. First, phases of the projects were contacted regarding readiness to report (none of which are in place today). Following that, detailing different levels of readiness on the basis of (how this theme was approached). The readiness to report on big data initiatives seems to depend on these levels. The variance across initiatives shows how the low bar for big data support, as well as the “laggards”, is mostly public values. Whether a project investigates either internal, or actionable analysis rather than closed, internal, or generalizable analysis helps explains the variance in readiness.

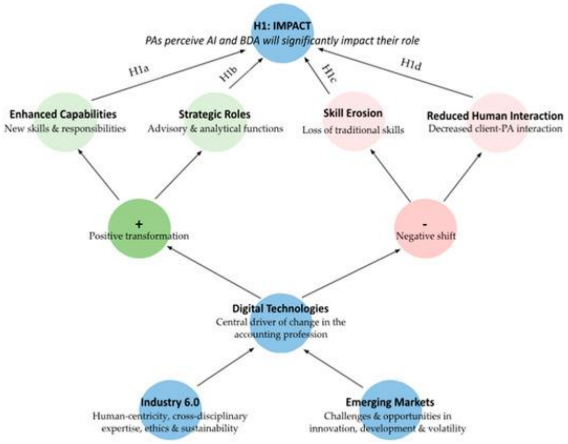

H1: IMPACT
PAs perceive AI and BDA will significantly impact their role
H1a
H1b
H1c
H1d
Enhanced Capabilities
New skills & responsibilities
Strategic Roles
Advisory & analytical functions
Skill Erosion
Loss of traditional skills
Reduced Human Interaction
Decreased client-PA interaction
+
Positive transformation
-
Negative shift
Digital Technologies
Central driver of change in the accounting profession
Industry 6.0
Human-centricity, cross-disciplinary expertise, ethics & sustainability
Emerging Markets
Challenges & opportunities in innovation, development & volatility

Fig 4: The Role of Artificial Intelligence and Big Data Analytics

7.1. City-Level Initiatives

Pioneering Governance for Data-Driven Decision-Making in Chicago

Governments are gradually adopting innovative informatics and big data tools and strategies. This trend in government leadership is being led by pioneering jurisdictions that are piecing together the standards, policy frameworks, and leadership structures fundamental to the effective use of data analytics. These groundbreaking initiatives provide cities across the country with an enticing glimpse of the technology's potential [12] and a sense of the challenges they must overcome to be able to use data safely and effectively in the service of public health. River of data floods civic systems as new ICTs create powerful new instruments for measurement and analysis. Increased access to inexpensive compute capacity has made it possible to develop sophisticated predictive algorithms that model risk and opportunity across public and private sectors. However, huge amounts of information are not yet large enough; the data come from diverse sources and are represented through disparate formats.

The same computation makes it difficult to connect the dots, while hacking urban spaces, but not for the purpose of learning where to deploy social services. Investments in technology must also overcome structural obstacles or the decisions will simply pencil themselves over with the urge of transparency. Pioneering jurisdictions are innovating at the intersection of data and tools, people who find meaning and the political structures to increase capacity – to connect disconnected knowledge into safe and efficient governance. City processes span a near infinite number of silos and organizations in horizontal and vertical configurations. Leaders across disciplines within institutions curate means to transverse domains in the pursuit of joint decision-making.

Tonight, they come together to catalyze social learning; tomorrow, myriad individuals must work to decide between a hundred or a thousand options. For every decision, thousands of interactions can occur using heterogeneous tools. Tools must address a critical mass of decisions, all relevant data must be complete, and coordinated information must drive the decisions. Effective tools support innovation or nimble adaptation that is systemic because it naturally expands capacity, directly creating software and protocols that knit processes together in a way that transcends prior patterns of governance. Geeky data scientists in the private sector could use similar techniques to wholly hack processes.

7.2. State-Level Programs

States have adopted big data and predictive analytics initiatives focused on improving government finance. The following examples highlight specific endeavors underway in California, Delaware, and Virginia.

California's Department of Finance is developing the Alternative Revenue System. This system will integrate architecture structure and financial data reporting for all levels of California government. Modeling and data analytics will generate projections and impact analysis designed to facilitate 1- to 20-year revenue estimations. This project capitalizes on the first Alternate Revenue System Pilot project. A prototype was successfully built for a federal district. The demo showcased projection scenarios for parking fines and cemented data analysis needs to interstate federal districts. The lessons learned from the pilot project continue to inform the California ARS development effort.

Delaware is rolling out a centralized Information Technology Financial Management System. This system aims to centralize all of Delaware's IT budgets for IT services and capital projects. Moreover, it will allow for the tracking of costs allocated to projects to reduce costs and estimate ongoing maintenance expenses across projects through standardization. These actions were identified by an IT Strategic Planning initiative undertaken by the Office of Management and Budget. Another strategy emerging from the planning process is to develop a software asset management process that catalogues all software purchases and replaces consumption-based tracking with more predictable acquisition models.

In 2012, the Commonwealth of Virginia's Secretary of Technology, along with the Deputy Secretaries of Technology and Finance, initiated a review of the Commonwealth's all IT expenditures. By December 2012, the Office of Technology states' 700 agencies will have their IT budgets established for 2014-2017. Anomalies in budget submissions will be identified through data analysis. A review of the state's current IT contracts will identify opportunities for significant savings via renegotiation and/or rebidding. Streamlined processes will be implemented in response to these studies, accelerating contract negotiation and approval timelines for lower-value contracts. Finally, new standards for reporting IT expenditures and all associated business case review processes will be established and implemented.

8. Fiscal Oversight and Accountability

Overall fiscal oversight and accountability of government finances is mostly a matter of ex post audit and control rather than the controls of the budget process. However, as in the previous integrity control category, some countries develop very effective reporting and oversight institutions. The empirical countries are classified into four groups according to ex post institutions, systems, and formats.

The group called "strong ex post" have a mix of effective audit, scrutiny, and accountability institutions supported by the wide-open scrutiny systems and detailed reporting formats. The countries in this group are Australia, Canada, Finland, Germany, Japan, South Korea, Netherlands, New Zealand, Norway, Sweden, and the United Kingdom. All of these countries are very developed economically, which is the case of broader fiscal integrity control as well.

A weaker ex post group called "fully ex post" have a mix of audit, scrutiny, and accountability institutions but their systems and formats are not that strong. The countries in this group are Austria, Belgium, Brazil, Chile, Colombia, Czech Republic, France, Italy, Mexico, Norway, Portugal, South Africa, Spain, and the United States. All these countries have good conditions of ex post control, but their constituent access to scrutiny systems and detailed formats vary with some countries criticized for overall weak fiscal integrity control.

Weaker ex post institutions are the main feature of the "weak ex post" countries. The group is in the weakest situation with hardly any country satisfying the criteria for ex post fiscal integrity control institutions. All these ex post control features show some degrees of improvement due to developing accountability institutions. However, oversight institutions are still either not adequately sharp or not as broad as desirable. Altogether, most countries in this group can be characterized as "close to the bottom."

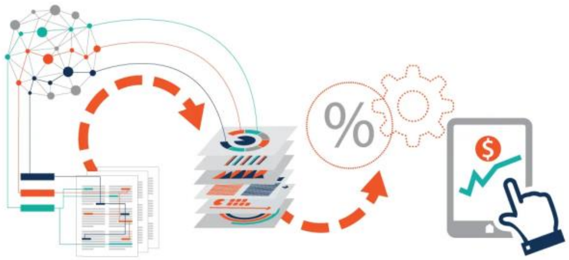


Fig 5: Practical Big Data Analytics For Financials

8.1. Enhancing Transparency through Data

"Data is the new oil" is a well-known phrase that, in recent years, has been refuted by many [5]. A strong case can be made that public statistical data, especially publicly-available and standardized ones, is more valuable than raw consumer data for a number of reasons, ranging from productivity and grand challenges such as climate change to antitrust cases. Nevertheless, despite the explicit inclusion of an open-data approach to public statistical data in the UK Government Digital Strategy, the use of public statistical data for value creating applications such as predictive analytics is still in its infancy. The aim is to demonstrate the value of publicly-available and standardized public statistical data in comparison with raw consumer data for targeted consumer communication activities such as predicting the durations until the next purchase. A Monte Carlo simulation study is presented that shows that publicly-available and standardized public statistical data can achieve similar predictive quality to raw consumer data, if chosen wisely. With regard to such data, the importance of location with respect to point of interests in terms of public statistical data transferability is shown and justified.

There are certain types of bureaucratic agencies that are accustomed to multi-purposing their own data. Agencies of this kind, particularly in engineering and transportation, have long collected data and integrated them into electronic territorial maps or dashboards [13]. In many agencies, the practices of data use and re-use have ramped up markedly in the last decade. The datafication of city government is a cultural shift in a certain discourse and practice of government: how staff understands urban processes as solvable through data. For most city governments, the main purpose of data is to track services and infrastructure and the city primarily counts what happens in the controllable built landscape, not what happens in the unbounded social activities of the populace. A categorical belief in data as a solution to problems of service logistics and public infrastructure directly drives many cities' administrative processes. Public servants from many departments and agencies proclaim that data plays a powerful role in managing the workings of a complex urban system. The power of 'data' can take many forms, with statistical analysis, mapping and visualization, or algorithmic processing all being possible. Many city staff members saw public data as an at-hand asset from which new civic or monetary value can be extracted. Nevertheless, this technocultural shift in how departments understand and use their information products has begun to reshape department roles and instigate new institutional practices.

8.2. Impact on Budgeting Processes

Government budgeting processes are affected directly and indirectly by the adoption of

Big Data and Predictive Analytics. Budget reforms can be anticipated by public organizations when adopting the new technology. Already more than 230 central government finance team members in the U.K. have been recruited by the government. The Office for Budget Responsibility stated that “we will put data at the heart of everything we do. We will ensure the data we hold is open and our systems and processes are accessible by all”. Big data is being used to efficiently carry out socio-economic researches to notice early if a budget has been spent efficiently and effectively. Data analytics engineers have been continuously recruited. The increasing pressure for more reliable assessments of the efficiency/effectiveness of Government budgets has made clear the limits of the “black box” used in Traditional Public Financial Management (PFM) Systems. New approaches will have to be sought, along the lines of Continuous Auditing and/or “cognitive approaches” using “big data”, which have started to be applied among Private Companies operating in financial markets [14].

Governments, at least in some countries, have considerable incentives to disclose information on their budgeting procedures and outcomes. Most countries present a Political Budget Cycle as a collection of pre-determined steps usually driven by the Executive and aimed at preparing, enacting and executing the Budget Law. Budget cycles sometimes differ from one country to another. However, the modelling of the budget cycles is helpful in raising and addressing some simple questions, including the dates of the procedures, and how to evaluate their comparability and reliability.

In pre-budget proposals, governments must present information on what has been done with the previous budgets, and on the outcomes of previous assumptions. At this stage, all data reported should be strictly comparable. However, in these instances, governments almost invariably enforce changes designed to make the information non-comparable to the previous presentation. Documentation on comparability is frequently omitted, making it difficult for outsiders to analyze the data. Ex ante forecasts disclosed at budget time usually have a budgetary statement which distinguishes public revenues and expenditures from public debt stocks, and public revenues and expenditures are usually classified by their functions.

9. Future Trends in Government Finance

Despite the challenges, predictive modeling and data analytics will continue to become tools of choice for chief financial officers and finance departments worldwide. At the same time, big data will continue to have a role to play in improving the productivity and efficiency of the finance function. Understanding the potential capabilities of big data will be foundational to assessing how it may become a tool in the future. In the next five years, finance departments will grapple with a growing volume and complexity of data. The advent of social media, mobile devices, sensors, and the internet of things has vastly increased the volume and variety of big data. The proliferation of new hardware and platforms, such as cloud computing and big data platforms, has also enabled organizations to significantly scale up the processing power available to them. As a result, traditionally computer intensive tasks can be executed at lightning speed. Separately, massively multiplayer online games are enabling organizations to better understand human behavior in real-time remote environments. However, along with the opportunities that this amounts to present significant new challenges for CFOs and finance departments. In five years, citizens will create and consume vast quantities of real-time data from their homes, work, schools, and neighborhoods, thus outpacing government’s ability to process, analyze, and respond. Emerging technologies such as digital telephones, community health information exchanges, and smart power meters will provide an unprecedented window into how government works and how people interact with government. Data in this context will be openly available, but silos, obsolescence, and poor standards will impede effective analysis. Agencies will also face security concerns regarding the unintended disclosure of private information in the public domain. Ultimately, agencies will begin to craft policies on how to effectively integrate these proprietary data streams into operations, which will improve public management but also require greater transparency and citizen engagement with government processes.

9.1. Emerging Technologies

The future of public health is healthier, smarter, more efficient, and more collaborative than ever. A better environment is emerging for individuals, families, and communities to be healthy. New tools and methods will be used to see health and diseases earlier in order that quick, data-driven actions can be taken to fight them. These will include the Internet of Things (IoT), big data, predictive analytics (PA), and the official data from public and private sector partners [12]. These new tools all need stop for a moment and go through a normal development cycle, including testing through proof of concept, creating and testing a prototype, rolling out a pilot program, and refining the tools with input from the pilot. Finally, they will be integrated into public health's daily work flow, and the data will be used for targeting and monitoring. All of these will be at the population level.

The first step is to think ahead about how to design and deploy the new sets of public health tools. Governments are gradually adopting innovative informatics and big data tools and strategies. This trend in government leadership is apparent, but poorly understood, in many jurisdictions. Recognizing data analytics as the next frontier for innovation, governance, and service delivery, pioneering jurisdictions are piecing together the standards, policy frameworks, and leadership structures fundamental to the effective use of data analytics in government. Governments all over the world see in big data the potential to improve performance, efficiency, delivery, and citizen engagement. To a lesser extent, governments also recognize data analytics' application to the delivery of services related to citizens' health, welfare, and safety, and as a domain for innovation and investment. Chicago is using lessons from other industries and established techniques such as PA to drive innovation in redesigning age-old processes. Processes from public health's traditional venue of place to geocoding and mapping are being integrated with better outcomes through tools and techniques in the private sector. Complementary to the present surveillance efforts in place-based data collection, Avenues & Streets is expanding the frame of observation to on-street use cases.

9.2. Policy Implications

Innovation in public sector organisations can be fostered in competencies, incentives, and arrangements when the public sector is able to free itself from being overly committed to administrative rules and procedures. This important condition for innovation is seriously challenged by the present demands for the public sector to be more transparent, accountable, and justifiable [1]. Given the scarcity of resources, new managerial controls will likely be contested with respect to their effects on public accountability. The challenge is to find such a balance among accountability, legitimacy, monitoring and control, discretion, and room for innovation. An ideal-type balance will be illustrated, but this topic requires further theoretical and empirical research. Public sector organisations are forced or feel eager to innovate in order to be able to fulfil their challenging tasks more effectively and efficiently. All sectors, public, private, and non-profit, are developing training for and in coaching, developmental conversations, scheme development, rapid prototyping, and so on. In addition to a stronger focus on outcomes/results, intended and enforced by political leaders and authorities, a general tendency can be observed to label employees or management teams as 'leaders', 'innovators', 'start-ups', etcetera, together with an implicit or explicit expectation that they will act accordingly. These overall trends put the imperative for innovation directly on the agenda.

Innovation of public sector organisations can be fostered in competencies, incentives, arrangements, and roles when public sector organisations can free themselves from an excessive commitment to administrative rules and procedures. This precondition for public sector innovation is under severe stress from current demands on the public sector to be more transparent, accountable, and justifiable. New managerial controls will, therefore, likely raise questions about their effects on public accountability. The question is how to find a balance between accountability legitimacy on the one hand, and discretion and leeway for public sector innovation on the other? All these questions and expectations hold the risk of 'just-in-time' and 'one size fits all' solutions, craving for a more theory-based approach of how exactly public sector innovation can be fostered and what the limiting and fostering conditions are.

10. Ethical Considerations in Data Usage

Governments today are in possession of a vast amount of data concerning the general public. This data comes from a great variety of sources and covers the entire population. Governments not only possess a lot of data, but they also possess detailed knowledge on how to exploit it. The technological basis for using this data has largely been adapted from industry practice and involves advanced data analytics, including predictive analytics. Predictive analytics is the use of statistical techniques and processes together with advanced analytical tools to construct predictive models. Such models can use established patterns from the past to estimate or forecast events in the future [5]. The models can produce scoring systems for risk assessment and accounting fraud. When they are designed to map behaviors directly to policy measures, they produce policy recommendations. There are a variety of different techniques to be used or combine to build predictive analytics, including regression analysis, classification, machine learning, neural networks, or heuristics. Predictive analytics is conducted solely by automated procedures or involves experts manually handling data analysis and related tasks.

The large number of data and advanced data analytics can greatly enhance government activities. This is widely recognized and discussed. However, this immense power must be handled with the utmost responsibility. Misuse of this power stems from possible human biases, misinterpretations, or technological flaws. Model obsolescence and overfitting are examples of possible errors emanating from statistical considerations. Other types of ethical concerns are more inherent to the use of algorithms with human-relevant outputs, including accountability of government actions, the validity of governing, and fairness of the proposed actions. These concerns of fairness and considerations of validity, i.e., whether the limit assumptions made hold in reality, are rapidly developing fields of research that cannot be adequately dealt with here [15]. It suffices to say that they underscore the need to transparently communicate correction attempts, performance monitoring, and model updates conducted to cope with concerns of unfairness and term-appropriate behavior.

10.1. Balancing Efficiency and Fairness

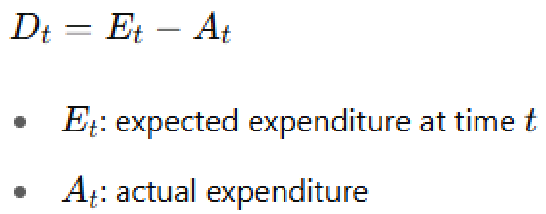
Public statisticians face a dual challenge. On the one hand, the paradigm shift toward predictive customer data can improve society by making better use of public funds. On the other, making predictive customer data unequal in access risks unfair treatment of citizens. It is therefore of most importance that predictive customer data be built in the right way. For everyone. Government agencies are responsible for the distribution of large amounts of money in order to improve society. Currently, governments do this without using the potential that big data has to offer in creating predictive models. Indeed, governments are frequently criticized for a lack of efficiency [5]. Big data predictive models can lead to further and fairer social benefits. It is time governments take a keen interest in predictive measures.

Private companies nowadays make better predictions of which individuals are worthy of the limited funding more likely. The social-choice theoretic principles of efficiency and fairness summarize the potentially conflicting predictions of statistical models. These principles leave a multitude of solutions among which it is often not clear which is the “best” prediction because of conflicts resulting from different configurational contexts. It is therefore paramount for entities to consider what they wish to achieve and which predictions to use. Governments bear responsibility for undertaking measures which balance social choice principles for fairer decision making. The decision-maker generally faces the challenge that using measures that improve efficiency often comes at a cost of decreased fairness and vice versa [16]. If performance-related measures differ greatly in one manner, one often can improve the performance on this criterion but will incur a greater cost regarding the other social choice principles. Measures that resemble one or the other extreme of the spectrum exist such that tradeoffs among these principles can be either alleviated or amplified depending on the configuration.

10.2. Regulatory Frameworks

While many opportunities arise from the process of big data adoption, it is also acknowledged that big data brings many uncertainties. Uncertainties significantly differ over the degrees of maturity of various governmental organisations when it comes to adoption of big data technology. Norway and the Netherlands present two unique cases, reflecting two different countries in the EU benchmark group in which big data has not yet been addressed in massive scale in the public sector. Through interviews with chief information officers from national tax authorities, this study aims to build a rich understanding of the range of uncertainties that governments face when big data is adopted, and how they can be managed in day-to-day operations [11]. Big data is defined as data that has increased in volume, velocity, and variety to a degree that it cannot be managed, processed, or mined by traditional database systems. The proliferation of large-scale data analysis technologies in the public sector is labelled big data. Big data geospatial by public sector bodies is anticipated to boost economic growth and welfare. However, concerns about potential harm of these public sector bodies. Regulating large scale data analysis in the public sector is labelled in an uncertain environment.

Eqn.3:Predictive Model for Budget Deviations



$$D_t = E_t - A_t$$

- E_t: expected expenditure at time t
- A_t: actual expenditure

The case studies revealed that uncertainties come in various forms. Even simply collecting more information with existing systems can generate unexpected workload and can conflict with other ongoing transformations. In addition, there are many uncertainties about what big data can do and cannot do. Connected to this, public authorities may adopt big data whose amplifying advantages fly far away big data’s inability. On the side of data capture, public authorities may either lose opportunities to generate insights by relying on too low threshold, or excessively capture allegedly interesting data.

11. Training and Capacity Building

Training is essential for maximizing the utility of big data analytics in government finance. Governments need to recruit new employees who understand the potential of big data and predictive analytics [17]. More broadly, government executives must have a robust grasp of big data analytics to allocate responsibility effectively. Administrative policies governing the implementation of big data analytics also need to be put in place. In terms of capacity building, governments should treat data as a tangible asset. Upstream and downstream actors must recognize and harness the economic utility of their shared data. Sharing data helps empower actors involved in government finance, thereby enhancing the performance of the financial operation with respect to budget allocation, spending, audit, and intergovernmental transfer reform. Building relationships is also crucial, especially in regard to units tasked with big data analysis. Government finance departments should hold seminars on big data analytics to showcase its potential. Regular meetings should be organized so that the output of the newly implemented big data analytics can be explained in detail. Relationship-building can foster an environment in which government finance big data analysts are able to freely share analytical results and insights with relevant government departments. The output of big data analytics is only useful if it is integrated with structured forms of knowledge sharing [18]. Candidates selected for specialized training should understand the organizational environment in which they will be placed, including formalized policies and procedures, or bureaucratic rules. Specialized training should incorporate an exploration of the business process of government finance departments so that trainees understand the opportunities and constraints on the basis of existing rules. Non-specialized training should be organized to build capacity for general forms of analytics.

11.1. Skill Development for Analysts

Analysts play a vital role in enabling organizations to harness their analytical capabilities on a broader scale [19]. However, it is important to recognize the contemporary challenges linked with analyst skill development and the implementation of appropriate learning and development mechanisms. In terms of the challenges faced, there are three key considerations. Firstly, significant investments in training initiatives alone do not guarantee an improvement in analyst skill. This reiterates the hypothesis of diminishing returns in the effectiveness of training. It is therefore critical that appropriate mechanisms are designed and implemented to complement training with learning opportunities on a day-to-day basis. Secondly, the effectiveness of learning and development mechanisms is contingent on the existing cultural traits that prevail within organizations. Once again, as with the case of investments in analyst training, organizations pausing to contemplate their approach will gain clarity over the cultural beliefs that they have developed, alongside recognizing potential complications arising from that cultural evolution. Thirdly, learning and development mechanisms should be considered holistically, and not single initiatives in isolation to one another. This is a common mistake prevailing within organizations, where budgets are allocated to training programs. Often, these standalone initiatives are not appropriately supported through additional learning opportunities and mechanisms.

In implementing effective learning and development mechanisms, there are several considerations that organizations should follow. Firstly, it is critical to steer the creation of an environment for supporting continuous development [5]. This will require resources and potential short-term sacrifices but will support the development of an environment that allows analysts to reflect and apply lessons on a continuous basis. Organizations choosing to invest in proactive routines will directly impact the social and collaborative basis for skills development.

11.2. Public Sector Partnerships

Public sector analytics can be described as the systematic use of data and other information for informed decision-making and performance management in the public sector [20]. It can also be understood as the government’s mode of conducting analytics, offered as a service to the public sector. Indeed, public sector analytics represents the particular intersection of the government and analytics domains. In contrast, other intersectoral forms of analytics exist, such as private sector analytics consisting of corporate methods, products and business models dealing with analytics on a subsumed entity. Other broader, intersectoral, entities have also been identified, for example all private sector entities and individuals involved in interdisciplinary projects. Further, public sector analytics can be seen as the particular intersection between the public sector and other specific forms of engagement, cooperation or cooperation. Partnerships in public sector analytics can be described as specific forms of proactive cooperation between the public sector and another sector regarding public sector analytics. They can take various formats, from family typologies comprising different contractual setups, funding models, levels of engagement to intra-field distinctions for more specific breakdowns.

12. Measuring Success in Predictive Analytics

Business analyses that had previously been time-consuming, hard to deal with or answered with unsatisfactory approximations, are solvable today in a reasonable computation time with adequate preciseness. For some highly complex problems, e.g. protecting data against misuse on the Internet, methods are still in development. On the other hand, there still remain many analyzable questions for business contexts, such as predictive customer data analytics, marketing response forecasting, demand analysis, global business expansion analysis, etc. Not everything built on mathematical methods is useful for every setting. When launching a new use case, goals must be defined first. The definition of business targets, specified analyses and provided data that sufficiently allows for a substantial analysis address the preparation phase. The following subsection discusses what decisions are crucial for designing and developing new predictive analytics applications, taking predictive customer data analytics as an example.

An application for predictive customer data analytics consists of various subprocesses. In the customer classification process, statistical features are created to describe the individual customers and their surrounding context. Several different statistical features may be created out of the same data but addressing different questions. When quantifying the expected information value of each statistical feature, it may not be sufficient to simply looking at its correlation to the target variable. Interrelations between features must be taken into account as well and can be measured with information-theoretic measures. The analytics model is based on machine learning. Good prepared feature and target variables, as well as a sufficient number of training observations, do not ensure a good prediction performance of a learning machine. Several machine learning approaches may be reasonable to provide a prediction model. Each has its pros and cons regarding interpretability, computational time, degree of the required end-user knowledge, and desirable predictability capabilities. Learning model development consists of implementing the learning models and tuning their hyper-parameters. A boosting algorithm provides predictive customer data analytics application with a high flexibility and rank loss minimization approach which was evaluated as the best performing learning algorithm in the context of predictive customer data analytics. Learning model evaluation provides information on how well the learning model is expected to perform on unseen testing data. It may also reveal weaknesses in the analytics design or learning process [5].

12.1. Key Performance Indicators

performance measurement systems, which contain relevant performance indicators, have been developed and implemented worldwide. Data-driven performance measurement methodologies and tools are being tested and utilized, especially in the private sector [21]. These developments raise numerous questions and challenges regarding the usage of business intelligence and key performance indicators by public bodies and public administrators. The experts in information science, business analytics, and performance measurement mostly agree that modern information technology is a source of performance improvement. The new IT generates vast amounts of data, which are collected and stored at exceedingly low costs. The existence of numerous, cheap, on-line data stocks has generated the need to develop and use BI and KPI applications to obtain a high-quality basis for making decisions in real-time and, subsequently, improving performance. BI has the potential to analyze data, find predictions and trends, identify customer needs, detect and anticipate errors, and create “what if” scenarios for strategic planning processes.

New data abound, and collecting relevant data has become easy and cheap. But, to make data valuable, the decision-making process and the business processes themselves must change. BI and KPI applications are used to provide the high-quality data needed to rely on data rather than intuition. Consequently, it is rare to find a company that does not use some BI or KPI applications. The next logical step is to explore how BI and KPIs have been implemented and to what extent governmental bodies use them to improve efficiency and accountability in their operations. The answer is not trivial since it involves analyzing the working processes and decisions of public administration that should be open to public scrutiny and not influenced by private or personal interests. Additionally, public bodies work in a highly politicized environment, different from the environment in which companies operate. But only if this analysis is performed will policymakers be able to recognize where business practices cannot work in the public sector and where they can.

Business process performance measurement has received wide attention in the literature in the last two decades since the low-cost availability of IT and the internet has reshaped not only the way business processes are organized but also the way performance can be tracked. Business processes are often viewed as chains of activities that provide products intended for customers through a variety of disciplines, such as marketing, sales, operations, accounting, finance, engineering, and design.

12.2. Feedback Mechanism

Once the performance of a predictive analytics model has been evaluated, feedback is often needed to determine whether adjustments to the model are warranted. In general, there are two types of feedback: substantive feedback and procedural feedback. Substantive feedback is often used to determine whether changes to the model should be made. It takes as a starting point the outputs of the model and tries to identify errors. In contrast, procedural feedback focuses on the process of developing and performance evaluation the model.

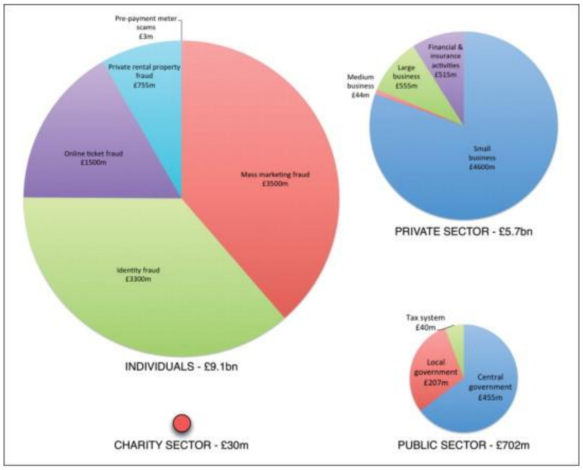


Fig :Smarter fraud investigations with big data analytics

Substantive feedback typically starts with suggestions or data from users. This information is used to evaluate the outcomes of the model. As users are likely to be familiar with the domain of the problem to which the model is applied, they might be able to point out incidents which the model fails to predict well. Taking this information into account, a sensitivity analysis is often run to better understand why the predicted outcomes deviate from an expected outcome. If users anticipate that the model does not include relevant information, their suggestions can be used to re-specify the model. If predictions can be made for more events or if prediction or post-processing rules can be further refined, the user might also contribute to improving the performance of a predictive analytics model.

Procedural feedback usually is not mutually exclusive with substantive feedback. A procedural feedback mechanism de facto is substantive because it is about the outputs of the data mining and model performance evaluation processes. However, when designing feedback mechanisms, it is often useful to keep these categories in mind. If the feedback is about the model output, the information might be more difficult to process than if feedback would have been on the evaluation process or model development process itself. It is also likely that different stakeholders will be involved, and different skills might be necessary.

13. Collaboration with Private Sector

Governments can leverage open data, modern software tools, high-performance computing, and data science to transform their operations. Improved efficiency means lower costs and better allocation of government resources - a huge opportunity given governments worldwide spend $30 trillion yearly, or about 50% of GDP. Governments have recently made efforts to share data openly with citizens, and they must now find ways to make better use of that data to improve the effectiveness of their interventions and, thus, the welfare of their citizens [20].

The same technologies used in business environments to extract insights from data can be harnessed by governments to improve the selection of areas requiring action, to measure impacts, and to forecast allocation of resources needed to reach goals. Machine learning and new techniques from data science can help governments make better use of data they already have, and can allow them to run applications for which performance was previously unacceptable. As sharing more data, including open data, is high on most governmental agendas, efforts can be made to design data sharing environments that allow government officials to gain insights from data generated by private sector activities they cannot access, or for which they currently lack computer systems and staff capable of harnessing data efficiently [12].

Collaboration between government bodies and private companies can be used to foster innovation in how data are utilized within governments, with a focus on the analysis of data generated by different actors. Local efforts to enhance the use of analytics are summarized. This is followed by a framework that restructures the problem of public sector analytics, focusing on information asymmetries as a source of inefficiencies in data utilization. This new lens shows challenges of implementing versatile and meaningful applications of big data analytics in public organizations and with predicting responses to different treatment options as possible next steps in the process of introducing new types of analytics in the public sector or improving existing applications or systems.

13.1. Public-Private Partnerships

Public-private partnerships (PPPs) are increasingly being utilized by local, state, and federal governments to deliver infrastructure projects traditionally handled by public agencies. The emergence of data-intensive systems for government performance evaluation and control has driven local level adaptations of these partnerships beyond infrastructure projects into the domain of governance itself [22]. Over the last three decades, the growth of analytic technologies has gone hand in hand with the commodification of public data, leading to the emergence of a new class of public-private partnerships based on data management.

The open data movement has driven municipal governments to make ever greater quantities of data available to the public. While this information is often presented online in the form of static datasets, the analytics movement has promoted another model for public-private partnerships. These partnerships often involve private vendors who supply cloud infrastructures capable of holding and analyzing large troves of public data, combined with algorithms that provide means for insight generation. Some of these partnerships are also characterized by the civil society contributing public data. The emergence case for public-private data analytics partnerships has been shaped by a further trend, related to the recent changes in government finance, namely, the shift in performance evaluation, control, and outcome accountability strategies from input tax and spend control to output performance-based management [20].

This trend has transformed how public finances are organized and handled, inducing further new requirements for public finance capabilities, as governments are tasked with indicating how their output metrics will be met and demonstrating how the promises have been kept in audit processes. In addition to transparency requirements regarding budget management and accounting, public organizations now need data-modeling capabilities for relative performance evaluation, goal-finding capabilities for outcome measurement, and simulation capabilities for performance control. Data analytics partnerships aimed at comprehensive data infrastructure building can help address these new requirements.

13.2. Sharing Best Practices

Public statistical data is used by local authorities in Germany to identify high-quality candidates for public customers in urban construction. At the heart of the model is a high-dimensional residential district-level composite score, which summarizes hard-disk statistics for each district and is derived from commercial databases. To find a suitable model, three different supervised learning techniques are compared with regard to their model transferabilities between one municipality and another. It is shown that relatively simple, homogeneous methods obtain the highest accuracy, but simultaneously large modeling efforts are warranted. Furthermore, some property brands in the rental market seem to be inherently more predictable than others. Finally, a Kelly criterion approach is applied to set an optimal investment weight for each candidate [5].

Data analytics has become an integral part of government finance. Its competitive advantages — flexibility, sensitivity, reliability — are becoming increasingly apparent. Nevertheless, caution is necessary in building forecasts on data analytics. As states, “When a measure becomes a target, it ceases to be a good measure.” Thus, data can only offer estimates of behavioral probabilities; it can never suffice as a guide for action. Different roles of data analytics can be observed. The first role is making visible changes in the environment. The second role is to anticipate decision-making behavior. The third role is assumptions testing. The fourth role is information storage. Banks, for example, analyze data stored by predecessors to absolve unwelcome credit applications.

In jurisdictions dominated by two-party systems, gross outcome differences can be observed as a result of making different assumptions or by reshaping deliberate processes. If electoral systems change, it often appears contradictory that a party suffers a loss of votes in the parliamentary arena, while a greater share of votes is held in government finances. Therein lies the challenge of appropriate data analysis. Systems divergences are often a microcosm of larger differences. Global stock prices are influenced by regulatory differences in respect of data standards or financial market protection.

14. Conclusion

This article offered an overview of the opportunities and challenges surrounding the deployment of big data and predictive analytics in government finance. It described the potential applications of big data and predictive analytics while highlighting the barriers to adoption and the types of research that could support overcoming these challenges.

The expectation is that big data and predictive analytics have the potential to transform revenue management in public administration, and that much can be done to improve this aspect of public finance. The initial review suggests that its application may manifest faster in wider areas of public finance than the actual collection of taxes and the identification of fraud. Research and academic aids to executive agencies need to be accompanied by a much broader research-driven capacity-building agenda to deal with the equally important strategic questions. In fact, research could also help with the other two barriers facing executive agencies, one of which may require more political than academic inputs but where academic facilities with regard to public sector innovation may still play a key role.

The rapidly evolving capabilities and skills related to big data – coupled with the primacy of the public finance agenda – will position this research field as one of the hottest over the next several years. Academic literature can assist in providing a solid grounding of knowledge that can be used to achieve real transformation on the ground. Nevertheless, what communities of researchers should also take care of is how to apply this extensive knowledge base in practical ways. The research agenda proposed can contribute to bringing regard to the frontier domains of big data and predictive analytics from public finance and policy perspectives into mainstream public administration research.

References:

1. Paleti, S., Singireddy, J., Dodda, A., Burugulla, J. K. R., & Challa, K. (2021). Innovative Financial Technologies: Strengthening Compliance, Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures. Secure Transactions, and Intelligent Advisory Systems Through AI-Driven Automation and Scalable Data Architectures (December 27, 2021).
2. Gadi, A. L., Kannan, S., Nanan, B. P., Komaragiri, V. B., & Singireddy, S. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization. Universal Journal of Finance and Economics, 1(1), 87-100.
3. Someshwar Mashetty. (2020). Affordable Housing Through Smart Mortgage Financing: Technology, Analytics, And Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 99–110. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11581.
4. Sriram, H. K., ADUSUPALLI, B., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.
5. Chava, K., Chakilam, C., Suura, S. R., & Recharla, M. (2021). Advancing Healthcare Innovation in 2021: Integrating AI, Digital Health Technologies, and Precision Medicine for Improved Patient Outcomes. Global Journal of Medical Case Reports, 1(1), 29-41.
6. Just-in-Time Inventory Management Using Reinforcement Learning in Automotive Supply Chains. (2021). International Journal of Engineering and Computer Science, 10(12), 25586-25605. https://doi.org/10.18535/ijecs.v10i12.4666
7. Koppolu, H. K. R. (2021). Leveraging 5G Services for Next-Generation Telecom and Media Innovation. International Journal of Scientific Research and Modern Technology, 89–106. https://doi.org/10.38124/ijsrmt.v1i12.472
8. Adusupalli, B., Singireddy, S., Sriram, H. K., Kaulwar, P. K., & Malempati, M. (2021). Revolutionizing Risk Assessment and Financial Ecosystems with Smart Automation, Secure Digital Solutions, and Advanced Analytical Frameworks.

Universal Journal of Finance and Economics, 1(1), 101-122.

9. Karthik Chava, "Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring", International Journal of Science and Research (IJSR), Volume 9 Issue 12, December 2020, pp. 1899-1910, https://www.ijsr.net/getabstract.php?paperid=SR201212164722, DOI: https://www.doi.org/10.21275/SR201212164722
10. AI-Based Financial Advisory Systems: Revolutionizing Personalized Investment Strategies. (2021). International Journal of Engineering and Computer Science, 10(12). https://doi.org/10.18535/ijecs.v10i12.4655
11. Cloud Native Architecture for Scalable Fintech Applications with Real Time Payments. (2021). International Journal of Engineering and Computer Science, 10(12), 25501-25515. https://doi.org/10.18535/ijecs.v10i12.4654
12. Innovations in Spinal Muscular Atrophy: From Gene Therapy to Disease-Modifying Treatments. (2021). International Journal of Engineering and Computer Science, 10(12), 25531-25551. https://doi.org/10.18535/ijecs.v10i12.4659
13. Pallav Kumar Kaulwar. (2021). From Code to Counsel: Deep Learning and Data Engineering Synergy for Intelligent Tax Strategy Generation. Journal of International Crisis and Risk Communication Research , 1–20. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/2967
14. Raviteja Meda. (2021). Machine Learning-Based Color Recommendation Engines for Enhanced Customer Personalization. Journal of International Crisis and Risk Communication Research , 124–140. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3018
15. Nuka, S. T., Annapareddy, V. N., Koppolu, H. K. R., & Kannan, S. (2021). Advancements in Smart Medical and Industrial Devices: Enhancing Efficiency and Connectivity with High-Speed Telecom Networks. Open Journal of Medical Sciences, 1(1), 55-72.
16. Chava, K., Chakilam, C., Suura, S. R., & Recharla, M. (2021). Advancing Healthcare Innovation in 2021: Integrating AI, Digital Health Technologies, and Precision Medicine for Improved Patient Outcomes. Global Journal of Medical Case Reports, 1(1), 29-41.
17. Kannan, S., Gadi, A. L., Preethish Nanan, B., & Kommaragiri, V. B. (2021). Advanced Computational Technologies in Vehicle Production, Digital Connectivity, and Sustainable Transportation: Innovations in Intelligent Systems, Eco-Friendly Manufacturing, and Financial Optimization.
18. Implementing Infrastructure-as-Code for Telecom Networks: Challenges and Best Practices for Scalable Service Orchestration. (2021). International Journal of Engineering and Computer Science, 10(12), 25631-25650. https://doi.org/10.18535/ijecs.v10i12.4671
19. Srinivasa Rao Challa. (2021). From Data to Decisions: Leveraging Machine Learning and Cloud Computing in Modern Wealth Management. Journal of International Crisis and Risk Communication Research , 102–123. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3017
20. Paleti, S. (2021). Cognitive Core Banking: A Data-Engineered, AI-Infused Architecture for Proactive Risk

Compliance Management. AI-Infused Architecture for Proactive Risk Compliance Management (December 21, 2021).

21. Vamsee Pamisetty. (2020). Optimizing Tax Compliance and Fraud Prevention through Intelligent Systems: The Role of Technology in Public Finance Innovation. International Journal on Recent and Innovation Trends in Computing and Communication, 8(12), 111–127. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/11582
22. [22] Venkata Bhardwaj Komaragiri. (2021). Machine Learning Models for Predictive Maintenance and Performance Optimization in Telecom Infrastructure. Journal of International Crisis and Risk Communication Research , 141–167. Retrieved from https://jicrcr.com/index.php/jicrcr/article/view/3019
23. [23] Transforming Renewable Energy and Educational Technologies Through AI, Machine Learning, Big Data Analytics, and Cloud-Based IT Integrations. (2021). International Journal of Engineering and Computer Science, 10(12), 25572-25585. https://doi.org/10.18535/ijecs.v10i12.4665
24. Kommaragiri, V. B. (2021). Enhancing Telecom Security Through Big Data Analytics and Cloud-Based Threat Intelligence. Available at SSRN 5240140.
25. Rao Suura, S. (2021). Personalized Health Care Decisions Powered By Big Data And Generative Artificial Intelligence In Genomic Diagnostics. Journal of Survey in Fisheries Sciences. https://doi.org/10.53555/sfs.v7i3.3558
26. Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. (2020). International Journal of Engineering and Computer Science, 9(12), 25289-25303. https://doi.org/10.18535/ijecs.v9i12.4587
27. Mandala, V. (2018). From Reactive to Proactive: Employing AI and ML in Automotive Brakes and Parking Systems to Enhance Road Safety. International Journal of Science and Research (IJSR), 7(11), 1992-1996.