

The importance of Big Data: US Internal Revenue Service

<https://doi.org/10.5281/zenodo.5039621>

In the Internet era, Big Data is term applied to datasets overflowing the boundaries of traditional database technologies. It brings together the power of computers and repositories with the largest inventories of ultra-individualized information to provide insight into every aspect of modern life and human behavior. This is a world that produces an estimated 2.5 quintillion bytes of data every day: representing the creation of ninety percent (90%) of the world's data in the last two years. It is also a world that creates huge challenges and gigantic opportunities for regulators.

Data scientists, policymakers, and tax experts are looking into ways of using big data mechanisms, tools, and solutions to advance the study and reform of taxation. The debate on plans for comprehensive tax reforms are currently underway by US leaders, while they gain increased access to big data analytics to inform their policies, priorities, and strategies.

The scale and detail of data gathered by services or government agencies like IRS needs to be meaningful and reflect the reality of their user base. IRS is ubiquitous for the USA and its citizens. It is present in our daily lives, impacting life decisions and even routine. IRS can gather and organized data about the majority of the population and all types of businesses. For tax data, nearly every American citizen and corporate entity is responsible for paying and reporting various tax information, disclosing a considerable amount about themselves. That means an unimaginable amount of data is available for gathering, organizing, managing, and analyzing. This is the very definition of big data.

IRS receives and processes more than 250 million tax returns every year. Budget cut and workforce attrition has negatively impacted IRS capacity as it fights an estimated tax gap of more than \$450 billion annually. Working smarter is the solution for more efficiency and more tools to battle tax fraud and tax evasion as identified by the IRS' Criminal Investigation Division.

Big data characteristics (for instance: volume, velocity, variety, veracity) also mean that big data employs significantly large storage space from diverse sources, stored in different formats, with different update intervals. Tax fraud analysis use of big data is a game changer as the industry releases methods, techniques, and technologies. Data mining through analytics is employed in the knowledge discovery in databases process, deploying predictive and descriptive tasks.

Through data mining, fraud investigation analyzes large volumes of data to discover unrecognized or unperceived patterns in data sets by leveraging statistical analysis and database technologies to find those patterns.

Predictive tasks work with machine learning and related technologies to make a prediction for each observation resulting from datamining. Prediction employs regression analysis to examine relationships between independent variables and dependent variables. Financial complexity demands the volume of variables provided by big data to make more accurate predictions. The statistical techniques for these include linear regression, multivariate linear regression, nonlinear regression, and multivariate nonlinear regression (as well as the more complex logistic regression, decision trees, and neural networks). Other, more complex predictive techniques of data mining appropriate to fraud detection or prevention include rule-based fuzzy reasoning, genetic algorithms, Bayesian belief networks and fuzzy neural networks.

Descriptive tasks, which include association rules and cluster analysis, describe the data under analysis. The use of these tasks to create models of behaviors (or transactions) that could fall under suspicious categories represents an important advancement. Descriptive tasks might be types of association rule analysis including multilevel association rules, multidimensional association rules, and quantitative association rules. Association rule algorithms generate rules describing potentially fraudulent situations. Cluster analysis collects data into related subsets patterns, identifying patterns used to discover or prevention financial fraud.

Complex and large-scale analytics such as what IRS fraud detection employs, requires big data, or the use of multiple data sources. An audit executed to discover fraud would integrate large internal and external datasets (demographics, taxpayer or corporate profiles, previous filings, call center data, and audit histories). The data analyzed could include a comprehensive backlog of years of historical data as well as external data. The volume and variety of data would be difficult to analyze without the analytics toolset of big data and the work of data scientists.

There are sources indicating the deployment of “spiders” by the IRS (automated computer programs) to review social media sites. Reports have also indicated the adoption of phone tracking technologies (for example, “Stingray,” a cell site simulator). Also, IRS keeps considerable volumes of data through utilizing more traditional technologies (for example, NRP and Individual Master File database). Independently from the accuracy of these sources and reports, a solid conclusion is that IRS has access to a large amount of data sets.

IRS is cross-referencing and mining these data sets to execute run pattern recognition algorithms so that it identifies trends enabling the understanding of the relationships in the data. IRS has employed varied advanced techniques and tools in these efforts (including anomaly detection, advanced clustering, and neural networks), with the objective of improving case selection and coordination among IRS divisions. Data analytics and predictive policing will help the IRS identify tax-reporting anomalies and identify tax evasion on a larger scale.

Within the accounting and tax law profession, big data and analytics are associated with automation. Offloading data management and processing power to computers translates to less manual labor to dissect numbers, construct models, and conduct independent analysis. This does not at all the end of opportunities for those tax professionals. It represents instead new beginnings, fresh opportunities, new knowledge, and a renewed importance of the experts working with these machines.

The partnership between data scientists, programmers, and law professionals with master’s in law in taxation (LLM) degrees is necessary to ensure asking the right questions about the wealth of information found through big data, to interpret the feedback derived from algorithms and data queries, and to provide guidance for the development of future policies.

The accountant profession offers Master of Taxation for those professionals seeking advancement in the field. Like previous example, this expertise is important to hoard the collected data, managed, and organized, to ensure that the analytic potential of available technology tools is leveraged optimally. As big data enables policymakers to be rely more on scientific methods while working on tax reform, the need for experts in tax law and accounting continues to make the new robust information accessible, meaningful, and useful.

Big data umbrella with the analytic capabilities it provides, has already demonstrated disruptive for different industries. The internet age isolates very few entities (people, businesses, governments) living in the 21st century from the impact of big data. Tax processes, policies and

practices are no exception. Taxpayers are looking for solutions and opportunities to make tax planning and compliance more seamless and automatic. Government, likewise, is investing in new ways to collect, organize, and utilize big data to enforce and reform taxes in the United States. Human creativity together with with analytic capabilities of modern technology, represent a brand-new era for taxation.

Over the last ten-year period, IRS investments in big data analytics will result in good return, in areas such as international tax enforcement through the collaboration with international tax enforcement efforts (country funded programs, document leaks, among others). Information reporting and information-sharing agreements have led to important structural changes in the global collaboration of tax-related information. These results will only further strengthen the new initiatives (highlight to the Joint Chiefs of Global Tax Enforcement, known as the “J5”).

Positive results produced by those investments will pave the path to enable new efforts to focus on areas that are ripe for development. Certainly, the one example at the top of anyone’s list is Cryptocurrency. Cryptocurrency-related tax compliance is unknown, most likely enormous. According to IRS reports, less than 1,000 taxpayers reported gains with cryptocurrencies like Bitcoin during the 2013-2015 period. The IRS is actively mining newly received data from actions on different exchanges servicing the USA.

For the IRS, the future of fighting tax fraud has arrived. Among the different processes, tools and efforts, IRS has embraced big data analytics, only seen the beginning has surfaced. With a reported year-over-year 400 percent increase in tax fraud detection and more than 1,000 percent increase in the identification of proceeds from other financial crimes, IRS is likely to increase its skates on its bet on big data, big data technologies, and tools.

References

- Freeman, J. B. (2019, January). The IRS and Big Data: The Future of Fighting Tax Fraud. *Today's CPA*, pp. 5-6.
- Klasing, D. (2019, September 2). *How the IRS Uses Big Data Analytics to Catch (and Punish) Tax Evaders*. Retrieved from Klasing Associates: <https://klasing-associates.com/irs-uses-big-data-analytics-catch-punish-tax-evaders/>
- Malaszczyk, K., & Purcell, B. M. (2018, June). Big data analytics in tax fraud detection. *Journal of Finance and Accountancy*, 1-10.
- Villanova University. (2020). *Big Data and Tax Reform*. Retrieved from Villanova University Tax and Business Online: <https://taxandbusinessonline.villanova.edu/blog/big-data-and-tax-reform/>

