FTA Awards Nomination/Entry Form

Person who led this effort or project

Name	Rahul Sarathy
Agency Name	Rhode Island Division of Taxation
Phone Number	(401) 574-9239
Email	Rahul.Sarathy@tax.ri.gov

About your program, idea, or project

Name your program, idea, or project: Rhode Island Fraud Analytics

What is the problem that you wanted to solve?

Note: the following describes sensitive methodology and as such, specific details concerning what was done are purposefully left out. Should reviewers want more details or insight, please let us know.

As Tax Administrators, we face an increasingly challenging set of problems with respect to combatting refund fraud. We have seen hardship across all agencies in hiring and retaining employees, an increasing amount of refund fraud in both quantity of filings and breadth of schemes, and find that “effective” refund fraud schemes are broadcast far and wide across multiple social media mediums. Coming out of the COVID-19 pandemic, we saw unemployment and identity theft fraud on a scale previously unobserved, and an increasing prevalence of schemes emerged from the dark web into the public internet.

The Personal Income section handles a volume of approximately 500,000 requests each year for refunds. Localized events such as external agency database breaches and dark web identity credential availability resulted in the enhanced need for vigilance, and as such we observed extreme growth in the percentage of those refunds that required review, equaling 163,000 (32.6%) in 2021 and 193,000 (38.6%) in 2022. It was apparent that the volume of refunds needing review based on pre-existing, traditional refund fraud business rules exceeded our capability, when normalized against prior refund issuance cadences.

The crux of the problem is how to keep pace with fraudsters, while handling increasing volume of issues with dwindling resources, and still allow for valid refunds to be issued in a timely basis without experiencing staff burnout.

Who was involved in addressing the problem?

The team that was involved with finding a novel solution was the Chief of Tax Analytics & Strategy, Chief of Personal Income, Personal Income refund fraud analysts and auditors, data analysts and the Front Office at the Division of Taxation. All members were State of Rhode Island employees.

How did they go about finding a solution?

To try to better identify a solution, we first built prototype logistic and linear regression models that took into consideration a few traditionally relevant fraud rules, and the model was trained on those refunds that were manually reviewed. To do so, we first allowed for refunds to be worked naturally by our auditors, in turn building up our training data, and then we attempted to model (logistic) if a refund was fraudulent or not, based on what had or had not been cleared before. We also attempted a linear regression model on how much fraud there was in a given refund. We ran into an issue here, in that there was an overwhelming lack of heteroskedasticity and output variance in the data, resulting in a nearly one-sided prediction. In other words, the data was too similar to itself to separate out the “noise” of the data from the statistically viable population.

Once we realized the group was too similar, we realized that we needed to split into smaller groups, and then within each group create a model for prediction. However, we found two paramount issues: 1) how to create the groups and 2) how to handle the fact that the treatment output was more complicated than “is it fraud or not.”

Conceptually, creating groups based on prior “intelligence” gleaned through refund fraud review was problematic as it only splits the data according to how it was proactively observed and disregards similarity of data that were not recognized. As such, to create groups, we turned to supervised machine learning (ML) and incorporated a Classification and Regression Tree (CART) model. A CART model systematically looks at all pieces of metadata among each refund, and then attempts to create rules to split it into “like” groups and estimate the result within that group by training on the “worked” data and predicted what will happen on the “unworked” data. It will repeat this idea across all possible groupings, until it finds a single rule that best divides the data as a whole. This division in theory creates the largest group where the result of the prediction is the highest measured intelligence gain. It then repeats the entire process for the remaining population, resulting in multiple non-overlapping groups.

To address the output issue, we looked at the common treatment scenarios, and eventually devised the following outputs from any given human review of a refund, with respect to the refund requested amount:

- The refund is issued unchanged;
- The refund is issued for a higher amount;
- The refund is issued for a lower amount;
- The return is amended or adjusted, resulting in a different refund requested amount;
- There is further processing in the period (the data is in-motion).

We set our parameters of the minimum group size and minimum predictive accuracy, and then allowed the CART model to dictate to us the business rules that result in the creation of each group as well as the output of that group, from the five options above.

To create the rules, we had to feed the model with metadata (variables) concerning the refund, which included traditional fraud “error” items (denoted by type), comparisons of demographic data, comparisons to prior filing and return history, inter-line item statistics from the submitted return, among many other categories. In this process, we grew from a dozen variables in our prototype to 93 unique variables in 2021 and 102 unique variables in 2022. To further assist in nuanced analysis, we allowed for each unique combination of the values of variables within categories to create interaction terms, thereby creating additional unique variables of logically similar material, selected based on values that had significant predictive power.

Describe the outcome. What is the new idea, approach, program, or activity?

The outcome was a massive savings in time coupled with a new, iterative, Machine Learning driven approach to finding and stopping refund fraud.

Each week, we would re-run the modeling methodology from scratch, adding or removing variables on occasion, obtaining fully realized sets of rules to find groups each week with a 95%+ prediction accuracy. With the addition of better data over time, we expanded from an average of 24 groups per week in 2021 to 51 groups per week in 2022.

Finally, we applied statistical sampling to the viable groups each week, allowing for the refund fraud analysts and auditors to manually review a percentage of the approved group's unworked population to confirm that the predicted result, when tested, were found to be statistically accurate. Once accuracy was confirmed, we would identify the individual refunds within each of the approved groups and allow for those refunds to be issued.

In 2021, we were able to predictively model, review, and release 42,000 "clean" refunds without human intervention. As the model was iterated based on previous findings, in 2022 we were able to release 90,000 "clean" refunds without human intervention, while finding a 400% increase in prevented fraud (in dollars). This results in a cost savings (time-based) of 46.63% of the effort it would have taken to review all our refunds using a traditional methodology. Assuming that each refund and return filing can be reviewed and cleared of fraud in 8-10 minutes, this process allowed for 8 additional full-time employees to spend time on other tax administration tasks.

This methodology of application of ML and predictive modeling to tax data at an enterprise scale is truly unprecedented. A formal Literature Review in 2022 found that there were very few predictive modeling studies in the published record concerning tax data. Those that were published dealt with subpopulations of datasets, largely reliant on aggregations (as published by the IRS) or done as a pilot on a small sample size. The rationale posed in the literature was that tax data is so highly protected that researchers are not able to use it in an academic setting, and that the agencies with data access lack the research and data analytics abilities to successfully execute academic-quality studies.

What has changed since this was implemented? How have your operations improved? Include any data, analytics or metrics that would show the value of your program. Don't forget management advantages such as improved morale.

As described above, this is the iterative result over multiple years. We have been able to keep up with an increase in volume from 163,000 (32.6%) refunds to be reviewed in 2021 to 193,000 (38.6%) in 2022, while also achieving cost savings of 46.63% despite identifying 400% more fraud (in dollars). The ability to free up resources to work on other tasks allows for more variety in assignment, increasing both the morale of the employees (avoiding doing the same task repeatedly) and their skill set (through experience of a more diverse set of tasks). The increased fraud detection is likely in part due to lessened pressure in having an individual review so many items, allowing the reviewer to have “fresh eyes” when reviewing a refund marked as potentially fraudulent. Lastly, the success of this project has internally invigorated our management team as to the possible uses of analytics, resulting in various section Chiefs embracing analytics and metrics, applying concepts towards everything from day-to-day management of resources, to identifying and pursuing more nuanced audit leads, to driving other novel efforts by virtue of being able to measure success.

Is there any component of your program that makes it workable only in your state or city?

Everybody ought to be doing this.

Is this an in-house project, or did you partner with an outside vendor or service-provider?

100% in-house

Additional information or comments about your usage of outside vendors or service providers.

The application of this methodology can be used across tax types and treatment scenarios, and even across government agencies’ approach to fraud detection. And we plan to do just that.

What comes next — will you be adding to your program, rolling it out more widely, trying additional approaches?

Next, we will be working to expand the fraudulent review beyond Personal Income tax into Corporate and Miscellaneous Business Income taxes, as well as to add complementary predictive models on suspicious filings resulting in underpaid and in balance period (which in turn helps find fraud on the federal returns).

Additional Optional Materials