IMPROVING UNEMPLOYMENT INSURANCE CLAIM FRAUD DETECTION

Elder Research designed and deployed an automated fraud detection solution for the New York Department of Labor Unemployment Insurance Integrity Center of Excellence. The tool was estimated to have identified 1,200 claims annually before the current investigative process with annual projected savings of $972,000 in recoverable and nearly $392,000 in non-recoverable overpayments.

INDUSTRY

» Insurance

BUSINESS NEED

» Rapidly detect fraud and improper payments in UI claims

» Make prompt and accurate eligibility determinations to reduce payment errors, fraud, waste, and abuse

SOLUTION

» Developed a flexible and scalable UI fraud detection tool that enables the client to deploy increasingly sophisticated UI program strategies over time

BENEFIT

» Savings estimated at $972K in recoverable and $392K in non-recoverable overpayments annually

» Easy for states to adopt and integrate the solution into their existing systems

» Fused disparate data sets into a unified view to simplify investigations

The Challenge

The U.S. unemployment insurance (UI) system is run and funded primarily by the individual states with oversight and support from the U.S. Department of Labor. Individual state UI programs are entrusted with ensuring benefits are paid promptly and accurately to eligible claimants, preventing improper payments (both over- and under-payments), and ensuring that employers properly classify their workers and pay their contributions promptly and accurately.

The Unemployment Insurance Integrity Center of Excellence is a collaborative hub for all Unemployment Insurance state agencies to elevate the expertise of the UI community. The Integrity Center funded the development of a Data Analytics and Predictive Modeling (DAPM) tool to support predictive analysis for Unemployment Insurance Integrity across 53 state and U.S. territory constituents. The goal for the development was to use data analytics to improve detection of fraud, overpayments, and underpayments in unemployment insurance. The solution needed to identify risky unemployment insurance claims, claimants, and employers while also helping states balance the competing requirements of promptness and accuracy of payments.

Key criteria and plan for the DAPM tool:

- Be generalizable and open to reduce adoption barriers across UI programs.

- Be flexible to allow easy modification and to accommodate the statutory variety of constituent laws and operating environments of the states.

- Implement tools in pilot states, evaluate their performance, and use the results to refine the development approach for future implementations.

The Solution

Elder Research and our subcontractors, most notably NTELX, Inc., joined forces to develop an innovative, flexible, and robust predictive analytics tool that allows states to dig deeper into their data to uncover improper unemployment insurance payments. Three state workforce agencies with diverse characteristics (New York, Idaho, and Kansas) were selected to pilot the program. Claim volume for these agencies totaled about 350,000 per week. The goal is that these pilot states will

ELDER RESEARCH
— DATA SCIENCE · AI · MACHINE LEARNING —

be a catalyst for other states and territories in the United States to adopt similar techniques.

Elder Research worked with the client and subject matter experts from the pilot states to establish the business needs of each state's UI program, uncover pain points and improvement opportunities, inventory relevant data sources, and identify all technical, legal, and regulatory constraints for the project. This ensures that our efforts are closely aligned with stakeholders' vision for the project and helps determine the features in each state's data that are relevant indicators of fraud and abuse.

The solution architecture included three main components: Model Framework, a Normalized Risk Database (NRDB), and end-used Business Intelligence (BI) tools (Figure 1). The NRDB is a standardized data model for storing UI data for the scoring modules to ingest. State data is extracted, transformed and loaded (ETL) into the NRDB database using customized scripts for each state. The model framework fuses internal and external data sources and stores the fused data and model outputs (risk scores, supporting data, etc.) into the NRDB. The data sources include common data elements (shared across state UI programs), state-specific data elements (i.e., derived from any unique UI program legal and regulatory requirements), and any additional external data (such as social media data) to generate risk scores for UI claimants that are stored in the NRDB.
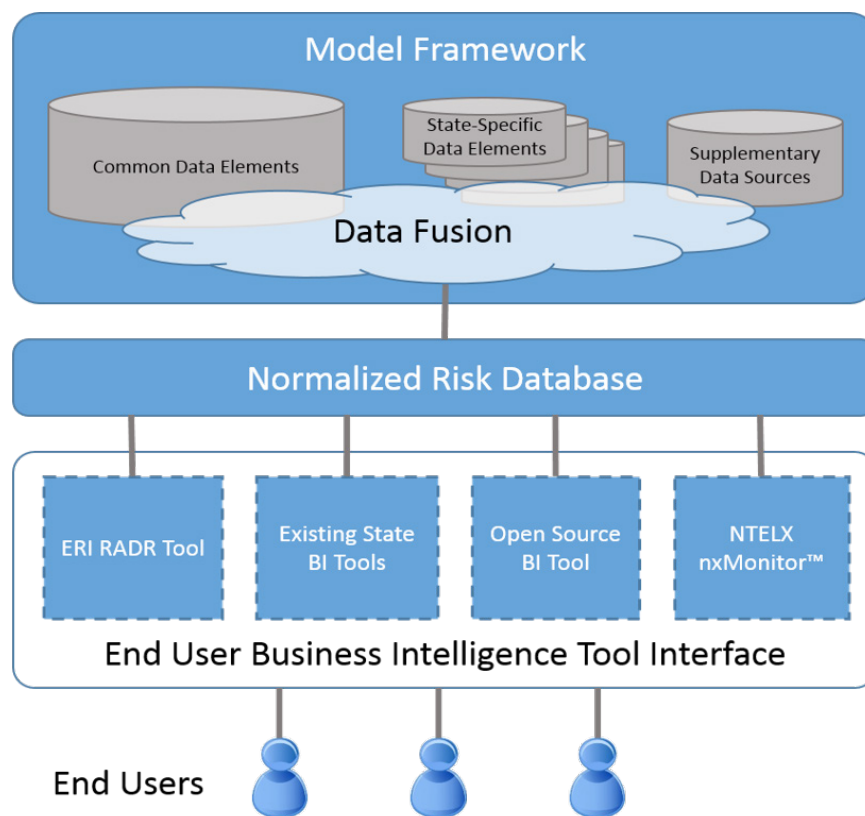

Figure 1. Open and reusable model design framework used for the DAPM tool.

The NRDB was integrated into the state's existing IT systems and built on an industry standard platform (such as Oracle or SQL Server). For data security the NRDB messaging services and programming interfaces are highly encrypted with hashes and keys to prevent unauthorized access and data interception. Three machine learning algorithms were trained using state UI data to help identify improper payments. Each method uses a different mathematical technique to study the same problem and complements the other modules. The benefits for each of the chosen techniques is shown in Table 1.

Supervised Learning Model	Benefits for Insurance Fraud Detection
K Nearest Neighbor	» Can add new instances without retraining » Performs well using only a few key features » Robust to unidentified fraud in training set and missing data
Support Vector Machines	» Can model complex relationships between features » Robust to noise and unidentified fraud in the training set » Can work with many features
Random Forest	» Robust to noise and missing values » Generalizes well to new data » Works well with a large number of features » Able to capture complex interactions between features

The outputs of several modules were fused into a Network Contagion module to predict the likelihood that an individual would commit fraud based on his or her network of associates. High-risk claimants are generally tightly connected to known bad actors as shown in the example link graph in Figure 2.

The NRDB contains the risk score history and necessary information to support the score and expedite claim determination. To access and analyze the claim risk scores and supporting data each state has the flexibility to select a BI tool that best integrates into their existing work processes.

States can choose to use their existing BI and workflow tools, open source tools, or tools developed and licensable by the Elder Research Team to manage UI claims data. The selected BI tool can display the claim, its risk score, the metrics that contributed to this score, and other supporting information such as graphs and hyperlinks that allow investigation of claims by drilling in to specific details or following cross-references to linked cases.

Pilot states used an easy-to-use BI tool developed by Elder Research called RADR (Risk Assessment Data Repository) — a custom solution that combines risk modeling with
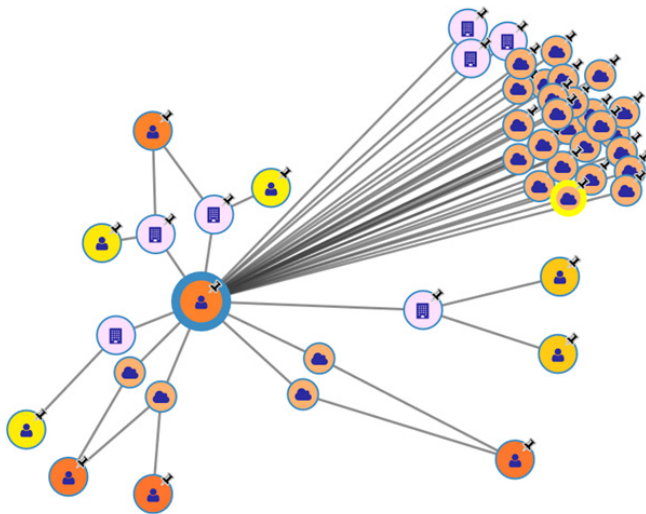

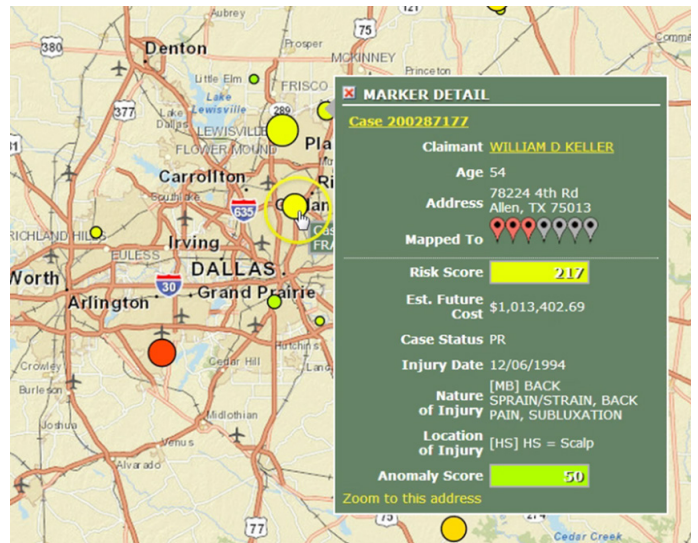Figure 2. Sample link graph showing network associations..

Figure 3. RADR map view showing geolocation of claims and associated claim risk.

data visualization to prioritize caseloads, uncover new leads, and accelerate investigation.

RADR is a highly flexible application capable of running on the three major operating systems — Windows, OSX, and Linux — and is adaptable to many client deployment environments including Oracle and Microsoft SQL server, as well as open source database engines such as MySQL and H2. Supported source datasets include flat files and open or closed source databases such as Microsoft, Oracle, H2, mSQL, MySQL. The risk scores range from 0 (lowest risk, colored green) to 100 (highest risk, colored red).
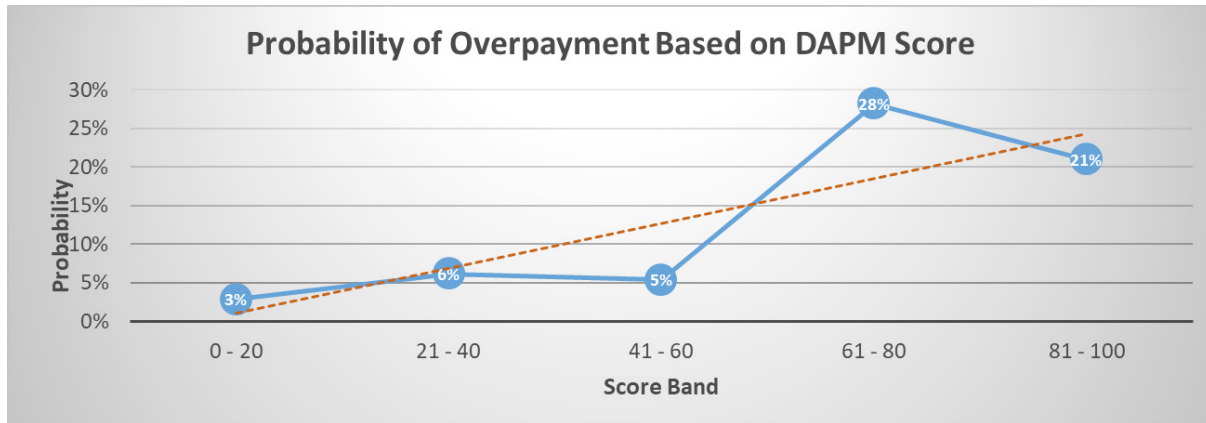

Figure 4. Overpayment probability bands for Idaho pilot test using the DAPM risk score

Results

The DAPM tool was highly acclaimed by the National Association of State Workforce Agencies (NASWA) and the New York Unemployment Insurance Center of Excellence. This innovative UI tool is scalable and deployable across all states, and provides a framework that enables the Center and the states to deploy increasingly sophisticated UI program strategies over time. Code for the DAPM tool is available from the Integrity Center to all states at their request.

The tool enables investigators to identify known fraud schemes in their data, uncover new fraud schemes as they emerge, provide data-driven answers to complex questions, and achieve significant return on investment by selecting cases that are most likely to uncover payment error, fraud, waste, or abuse. For example, Figure 4 shows the relation between fraud and DAPM score on the Idaho pilot test data. Note that the score separates problem claims very well, as new claims in the two highest score bands have a 28% and 21% probability of overpayment, whereas those in the lower three bands have 6% or less.

The state of New York saw similar success in their implementation of the DAPM tool. New York and Elder Research estimated the savings of implementing the tool by calculating how much sooner the tool identified an overpayment before the typical state process. The tool was estimated to have identified almost 1200 claims annually before the current investigative process (approximately 100 per month). If we were to assume New York stops payments when the tool identifies a claim, the tool could have saved $972,000 annually in recoverable overpayments and nearly $392,000 in non-recoverable overpayments.

The tool offers a single secure point of entry for identifying, analyzing, and tracking questionable UI claims across multiple data sources. End-users can work together to analyze real-time (or near real-time) data to uncover UI claims that have the highest risk for fraud or improper payments. Staff have comprehensive information on all UI claims and high-risk claimants are identified before and after payment and referred for further review and verification. The automated scoring algorithms directly improved the speed, quality, and consistency of claim determinations. The plan is to implement the DAPM tool nationwide over the next few years.