



# DETECTING FRAUDULENT HEALTHCARE CLAIMS

FOR THE U.S. DEPARTMENT OF LABOR

Elder Research partnered with the U.S. Department of Labor Office of Inspector General to develop a custom analytics solution to prioritize potentially fraudulent workers' compensation claims for investigation. Consolidating data sources into a centrally managed data repository and interactively visualizing information reduced case research times from days to hours.

## INDUSTRY

- » Healthcare

## BUSINESS NEED

- » Provide the Department of Labor OIG with an effective predictive model to detect fraud in Office of Workers' Compensation Program data and prioritize high-risk claims for investigation

## SOLUTION

- » Developed, validated, and deployed a predictive model ensemble to score fraud risk and present results in RADR, an intuitive, interactive visual tool which enables auditors to access data on high-risk claims

## BENEFIT

- » The RADR tool amplifies the productivity of DOL-OIG auditors
- » Reduced case investigation time from days to hours, combining client goals and expertise with objective measures of risk

## THE CHALLENGE

Department of Labor Office of Inspector General (DOL-OIG) contracted with Elder Research to create a predictive model to detect fraud in Office of Workers' Compensation Program (OWCP) data. The goals of this project were to highlight abnormalities in the claims data that could be used to form the basis of future audits and to create visualization tools to enable auditors to explore model results.

## THE SOLUTION

Elder Research implemented a healthcare case fraud predictive model and RADR data visualization tool for the Office of Audit (OA) staff. The supervised risk model was based on the attributes of known fraud cases from the Office of Labor Racketeering and Fraud Investigations (OLRFI) case management database.

The DOL-OIG provided five additional data sets to support this project. The data was comprised of over 900,000 FECA Workman's Compensation (OWCP) cases and their supporting data (Case Management, Bill Pay, Compensation, and Chargeback data) from the four previous fiscal years.

Catching fraud is very difficult and our modeling approach needed to be more sophisticated than normal predictive models due to there being very few positive fraud examples, no negative examples, and a very large number of variable categories. Another difficulty is that those investigating fraud cases often do not verify which cases are non-fraudulent, since only fraudulent cases advance through their workflow. In this project only 0.00034 percent of cases within the provided data were marked as fraudulent. Our analysis indicated that this was unlikely to be the true prevalence of fraud; that is, some of the cases marked as "unknown" were surely actually fraud cases, so not all those left unmarked could naively be used as negative (non-fraud) examples.

To minimize the chance of using unmarked fraudulent cases as examples of "non-fraud", an advanced two-step modeling approach was developed. The first step classified the unknown cases and ranked them based upon their likelihood of fraud. The second step used three different algorithms (logistic regression, random forest, and neural networks) to predict fraud. These models were then scored against the validation data to determine how effective they were at predicting

### Headquarters

300 W. Main Street, Suite 301  
Charlottesville, VA 22903  
(434) 973-7673  
[www.elderresearch.com](http://www.elderresearch.com)



**ELDER RESEARCH**  
DATA SCIENCE & PREDICTIVE ANALYTICS

### Office Locations

Arlington, VA  
Linthicum, MD  
Raleigh, NC

© 2017 Elder Research Inc.

fraud using an ROC curve. The ROC curve compares true positive rate to false positive rate and one may select the model with the highest area under the curve (AUC) as the best. The random forest model performed best, followed very closely by the neural network.

Once these three separate models were created, tested, and found to yield good results, we combined their strengths in an ensemble which often yields more accurate results and also avoids any one model being too sensitive to the specifics of the data used to train them. Two model ensembles were tested; a combination of random forest and neural network algorithms, and combination of all three. Based on AUC score, the random forest and neural network ensemble was selected as the production model.

Three of the OWCP Federal Employees' Compensation Act (FECA) databases were joined together at the case level allowing disparate data sources to be displayed together in a meaningful way. The model's results and supporting data are viewable in an easy-to-use visualization tool called RADR (Risk Assessment Data Repository). The tool enables auditors to access and visualize data on high-risk claims. Auditors can view claim risk scores in either a list view (Figure 1) or a map view.

The interactive visualization provided by RADR allows the client to rank cases by risk, to filter cases by dollar value or other attributes, and to drill down into cases to quickly determine if they are worth pursuing, and if so, where to begin.

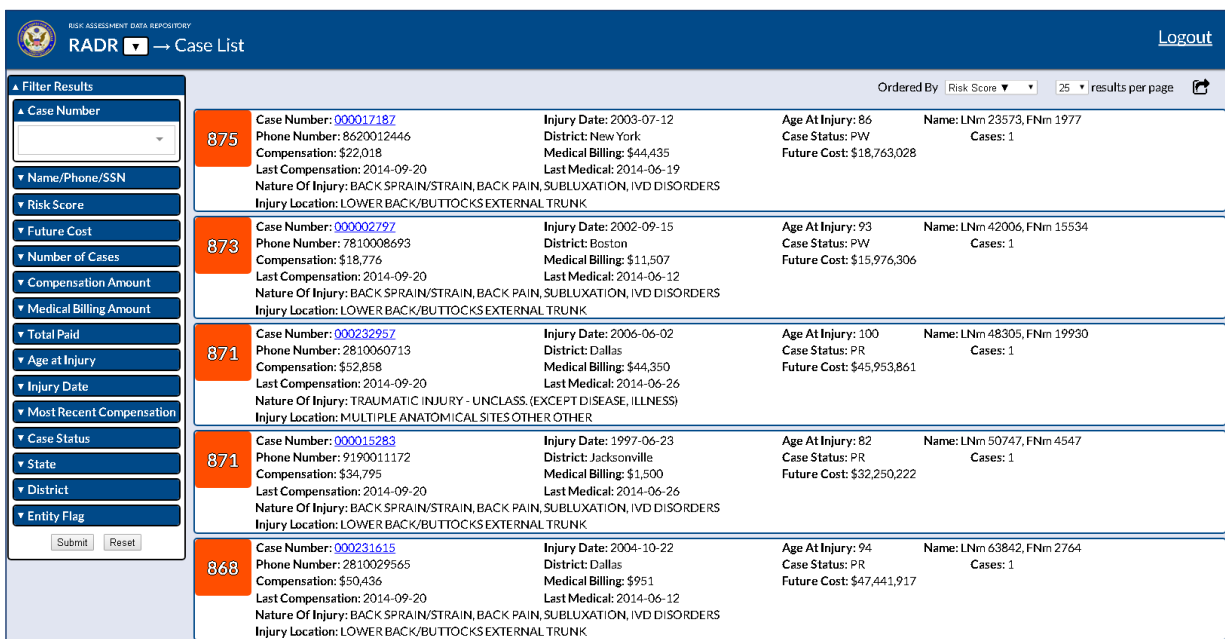


Figure 1. RADR visualization list view showing ranked list of cases and filters to help focus the investigation

## RESULTS

Once the ensemble model was put into production, a fraud risk score was generated for each of nearly one million cases in the observation data set. Since the known fraud cases were also scored through the model the team was able to determine that analyzing the top 21% of the cases with the highest risk scores would enable auditors to identify 90% of actual fraud cases. The RADR tool amplifies the productivity of DOL-OIG auditors by streamlining everyone's view into a centrally managed data repository

and analytic solution. Research that would have taken days is now done in hours, combining client goals and expertise with objective measures of risk.

The RADR fraud solution is being used by more than 150 DOL-OIG investigators and auditors across the country and RADR's effectiveness has led to it being mandated by the Special Agent in Charge to be the first step in every investigator's workflow.

### Headquarters

300 W. Main Street, Suite 301  
Charlottesville, VA 22903  
(434) 973-7673  
www.elderresearch.com



**ELDER RESEARCH**  
DATA SCIENCE & PREDICTIVE ANALYTICS

### Office Locations

Arlington, VA  
Linthicum, MD  
Raleigh, NC