cormac
HEALTH DATA ANALYTICS

Identifying Improper Medicare Payments

Unsupervised Machine Learning Prototype Using Medicare Skilled Nursing Facility (SNF) Claims Data

$3.3 billion in federal funds are lost each year due to Improper Payment at Skilled Nursing Facilities. CORMAC's Unsupervised Machine Learning Model provides a highly scalable machine driven way to identify claims that have a high potential of being improper claims.

This approach provides a path to curtail improper payments before they are paid.

Machine Learning Approach to Identifying Improper Payments

Background

Every year billions of dollars are wasted due to improper payments. The majority of these improper payments is due to insufficient or no documentation to substantiate the claim. CMS calculates the Medicare Fee-for-Service (FFS) improper payment rate through the Comprehensive Error Rate Testing (CERT) program. Each year, CERT evaluates a statistically valid stratified random sample of claims to determine if they were paid properly under Medicare coverage, coding, and billing rules.

Approach

CORMAC's approach to identifying improper payments has been through Machine Learning, particularly using an Unsupervised Learning Algorithm on SNF data. The available SNF data lacked labels which prompted the use of unsupervised machine learning algorithms to estimate the multi-dimensional probability density functions of the claims data.

Benefits

Analyzing each claim and assigning individual risk scores identifies a set of claims at high risk of being improper. If CMS reviewers limit their scope of evaluation to only this set of claims there would be substantial savings in labor, and time. In addition to potentially discovering (and rectifying) a larger number of actual improper payments, the reviewers would not be spending as much time looking at records that were highly likely to have been paid correctly. Aside from the cost savings, the time savings would be substantial as well.

Conclusion

The results produced from the Machine Learning Algorithm clearly shows that there are many potential improper payments can be identified before the payment is made to the provider. The machine learning model would train itself by continuously feeding back the improper claims. If these claims are analysed before the payments are made, numerous improper payments can be avoided thus saving a substantial amount of money and time for CMS.

- $3.3 billion lost annually in Skilled Nursing Facility Improper Payments
- SNFs are responsible for the 2nd highest amount of Improper Payments

• • •

- Unsupervised Machine Learning used
- Data: 2015 & 2016 SNF data from ResDAC
- Purely data driven

• • •

- Analyze every claim
- Identify in real time or pre-payment
- Model continuously adapts & learns with more data and feedback

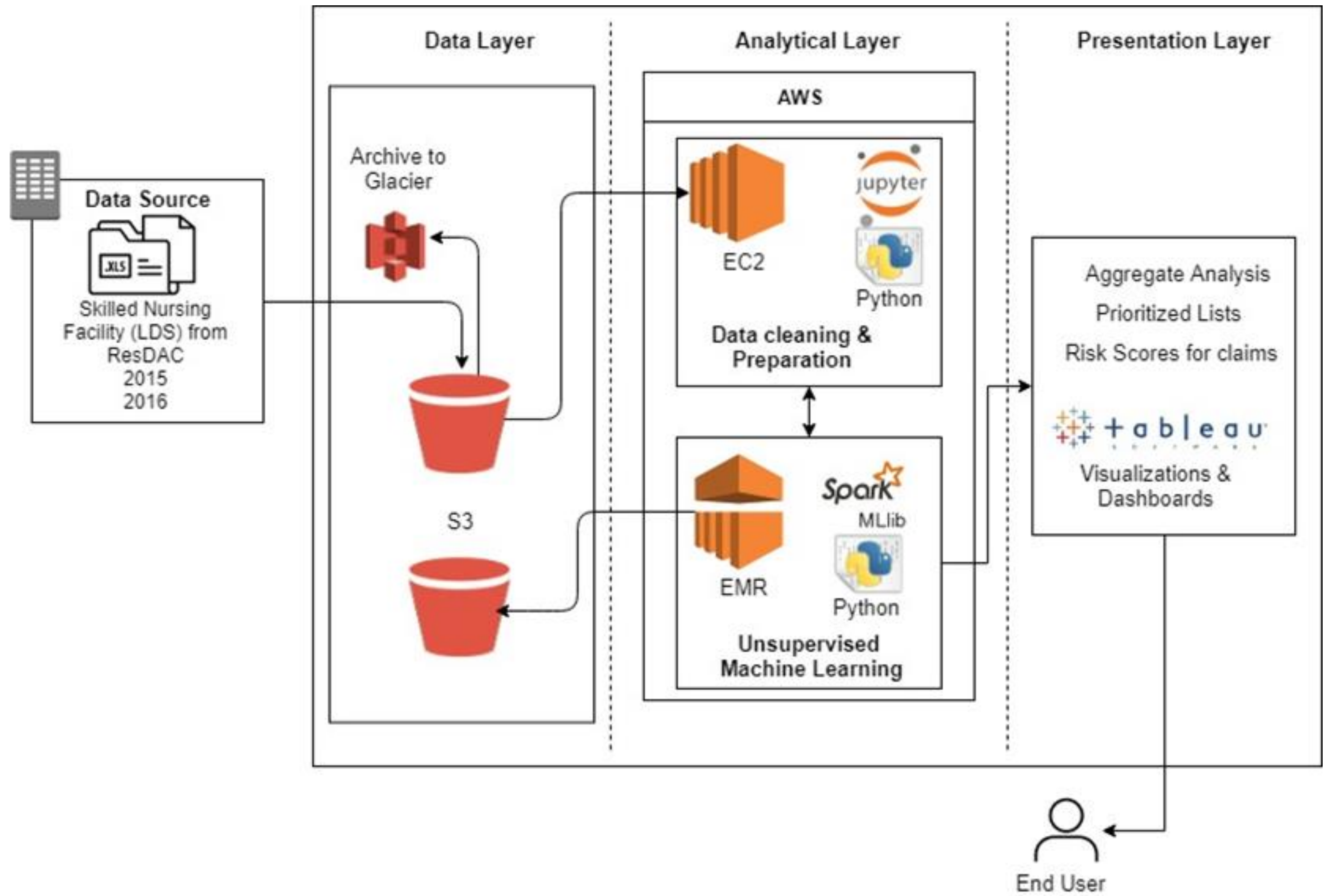
• • •

- Substantial savings in time and effort
- Introduce semi-supervised learning with feedback

• • •

Architecture

We leveraged our Innovation Lab residing in Amazon Web Services (AWS). The architecture diagram below depicts the process that SNF claims data underwent to produce the results. The results generated are displayed using the Tableau BI tool to depict visualizations (Graphs are shown in the oncoming pages).



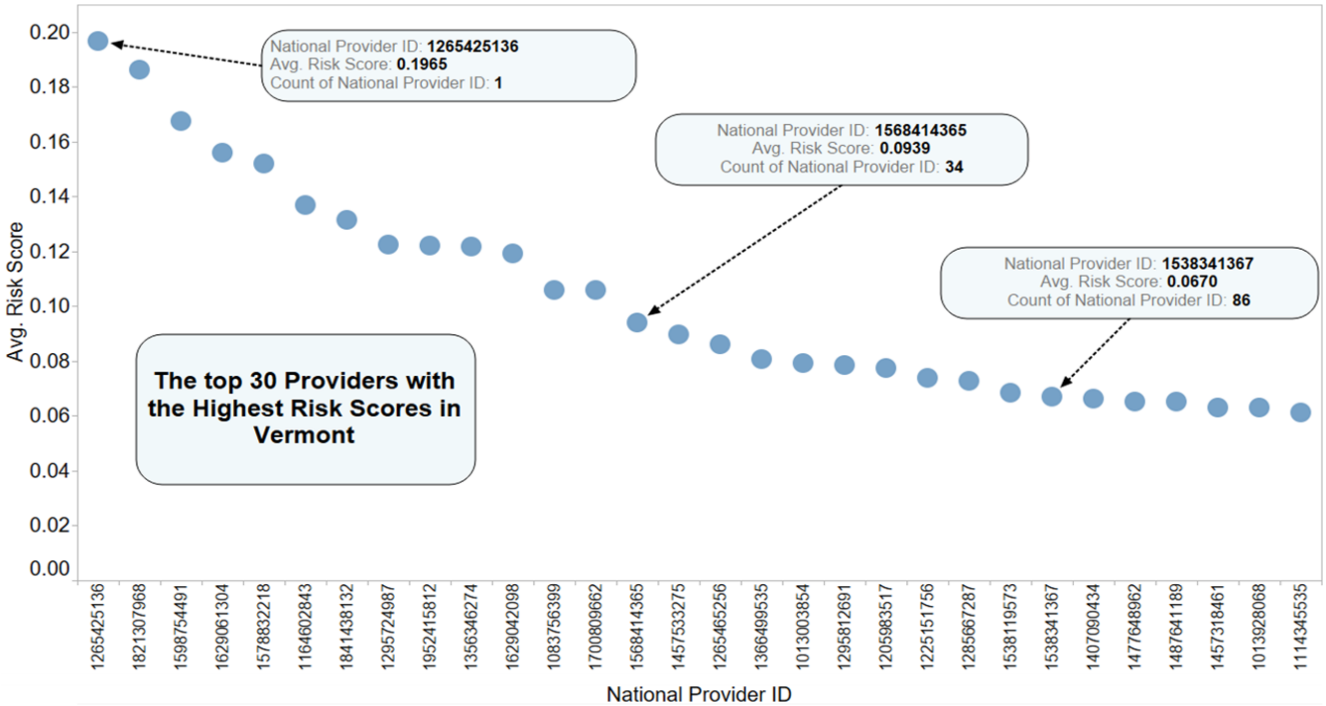
In preparation for analysis, the data from ResDAC was stored in S3 buckets and the data not in current use was archived in Glacier. The data was cleansed as the first step in the process of analyzing, identifying and correcting raw data. We used Python in a Jupyter notebook hosted in an EC2 instance to cleanse and prepare the data. The prepared data was then stored in S3 before use by the machine learning algorithms.

We utilized Spark deployed in EMR clusters to run unsupervised machine learning algorithms from Spark MLlib libraries and Python libraries. The output of this process was the assignment of individual risk scores to each claim. These risk scores were then prioritized, used for aggregate analysis and population analysis to identify potential improper payments. We also employed Tableau for visualizations and building dashboards.

Machine Learning Model Findings – Sample Visualizations

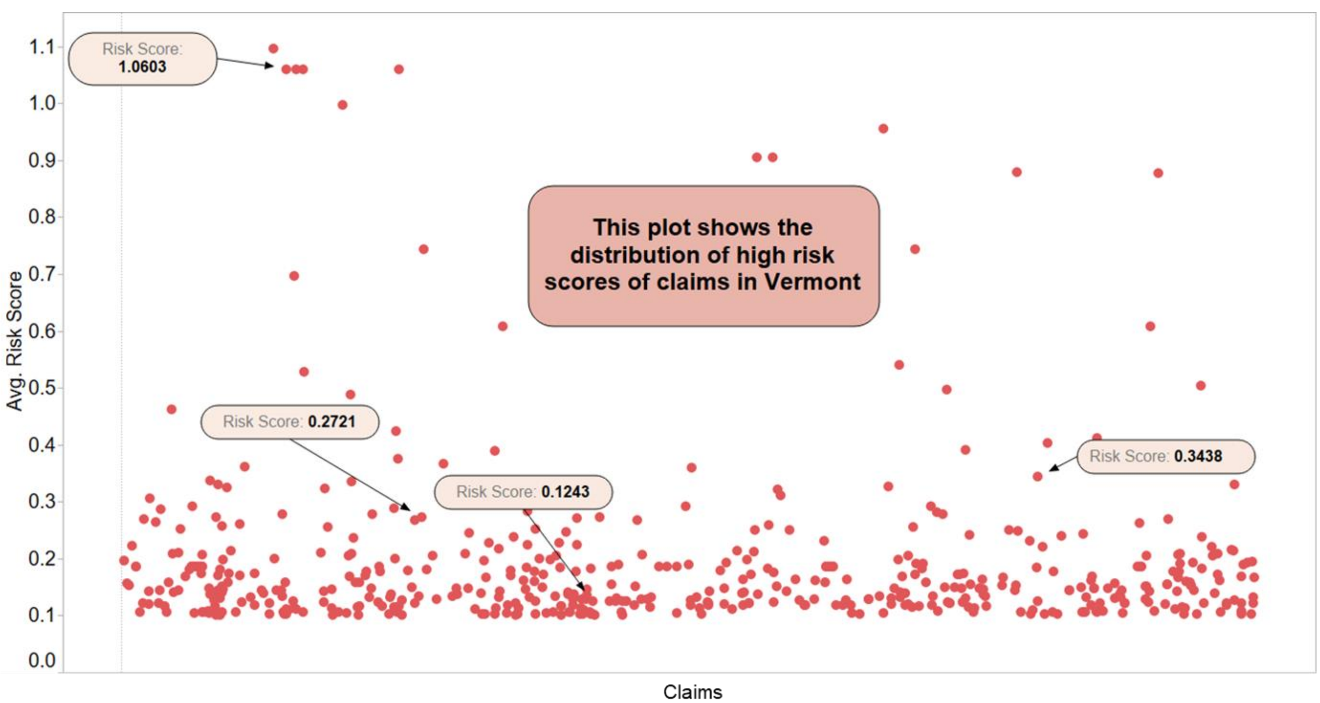
1. TOP 30 PROVIDERS WITH THE HIGHEST RISK SCORES IN VERMONT:

The following graph shows the providers with highest average risk scores. The dots at the top show the providers with higher risk claims that are worthy of further investigation.



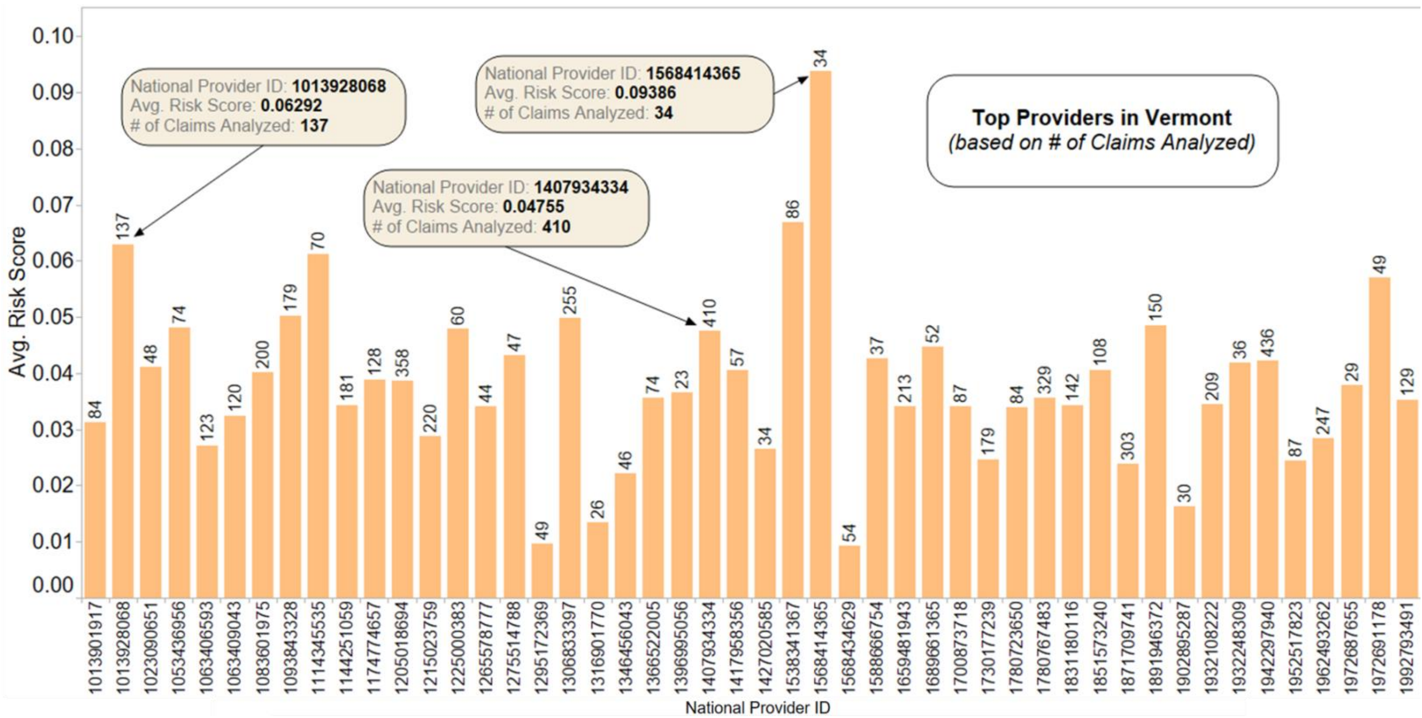
2. HIGH RISK CLAIMS IN VERMONT:

The following graph points out the high-risk claims in Vermont. The high-risk scores imply a greater probability of a claim being an improper payment through incorrect coding or lack of sufficient documentation



3. TOP PROVIDERS WITH RISK SCORES IN VERMONT

This bar graph shows a comparison of Risk Scores based on the claims analyzed for each provider. In particular, the higher bars indicate providers with fewer claims and higher risk scores; a combination which warrants further investigation.



4. MACHINE ANALYSIS OF A CLAIM:

An analysis of a single high-risk claim (MEDPAR ID 31307) spotted by the machine learning model.

- **Overall Statistics:** State of Arizona, Average DRG_PRICE: $74,418, DRG_PRICE of this claim: $193,460, Length of Stay in this claim: 25 days
- **Summary of Analysis:**
    - Long Stay, low-cost procedures and very high charge
    - Many diagnoses with unspecified conditions
    - More likely to be a case of insufficient documentation or incorrect coding

Diagnosis
Hypertension NOS (Unspecified essential hypertension)
2nd degree burn NOS (Blisters, epidermal loss [second degree], unspecified site)
DMII wo cmp uncntrld (Diabetes mellitus without mention of complication, type II or unspecified type, uncontrolled)
Chest pain NOS (Chest pain, unspecified)

Diagnosis
Acq coagul factor defic (Acquired coagulation factor deficiency)
Thrombocytopenia NOS (Thrombocytopenia, unspecified)
Liver disorder NOS (Unspecified disorder of liver)
Alcoh dep NEC/NOS-unspec (Other and unspecified alcohol dependence, unspecified)
Liver disorder NOS (Unspecified disorder of liver)

Procedures
Wound irrigation NEC (Other irrigation of wound)
Electrocardiogram
Venous puncture NEC (Other puncture of vein)
Skeletal x-ray of shoulder and upper arm
Injection of insulin