MLE Student Capstone

1 Page Narrative

|  |  |
| --- | --- |
| Project Title | Healthcare Anomaly Detection |
| Industry Sponsor | -none- |
| Team Members | McKone, Sharma  Chavarria, Lederer |

Healthcare fraud is an expensive white-collar crime in the US and leads to an increase in healthcare premiums, and a reduction in quality and access to care. The National Health Care Anti-Fraud Association[1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9013219/#B1) conservatively estimates that about 3 percent of US healthcare spending is lost to fraud per year ($300 billion approximately).

Machine Learning techniques can identify current and evolving anomalies in claims data. As fraud becomes more sophisticated across an increasing number of annual transactions, an ML solution provides an opportunity to greatly reduce the effort, time and associated cost spent in identifying claims anomalies, and recouping any misappropriated funds.

To illustrate the capabilities of Machine Learning to identify claims anomalies, this capstone project team has developed two solutions:

* a supervised Logistic Regression Model to identify potential anomalies at the provider level
* an unsupervised KMeans Clustering Model to identify potential anomalies at the claim level

These solutions are hosted on Amazon EC2 utilizing a Streamlit front end for illustration. Fast API endpoints are in final development to expose the model for on-demand supervised and unsupervised predictions.