

# Claims Adjudication: Application of Machine Learning

**Project/Initiative Name:** CAPE (Claims Adjudication Processing Engine)

**Project Manager:** Terri LaRoche

**Coaches/SMEs:** Sarah McGuire, Pamela Nash, Susheel Tiwari

**Team:** Matthew Bernstein, Jakob Herlitz, Nathan Hunt, Jagath Jai Kumar, Allan La, Aditya Parmar, Sarthak Patel, Samuel Sledzieski, Ashley White



## Background

In 2016, a team was tasked to explore Machine Learning in the claim domain through the Claims Adjudication Prediction Engine (CAPE). They successfully predicted whether a claim would be paid, closed, or denied. However some claims have to be manually adjudicated due to certain circumstances. Inspired by the success of CAPE, the goal of our project is to answer the following question. Can we automate the manual adjudication process?

## Example of Claim Being Predicted

Before

icn	prc_cd	rpi_id	seq_nbr	dg_cd
1	A		2	
2	B	XYZ	2	P

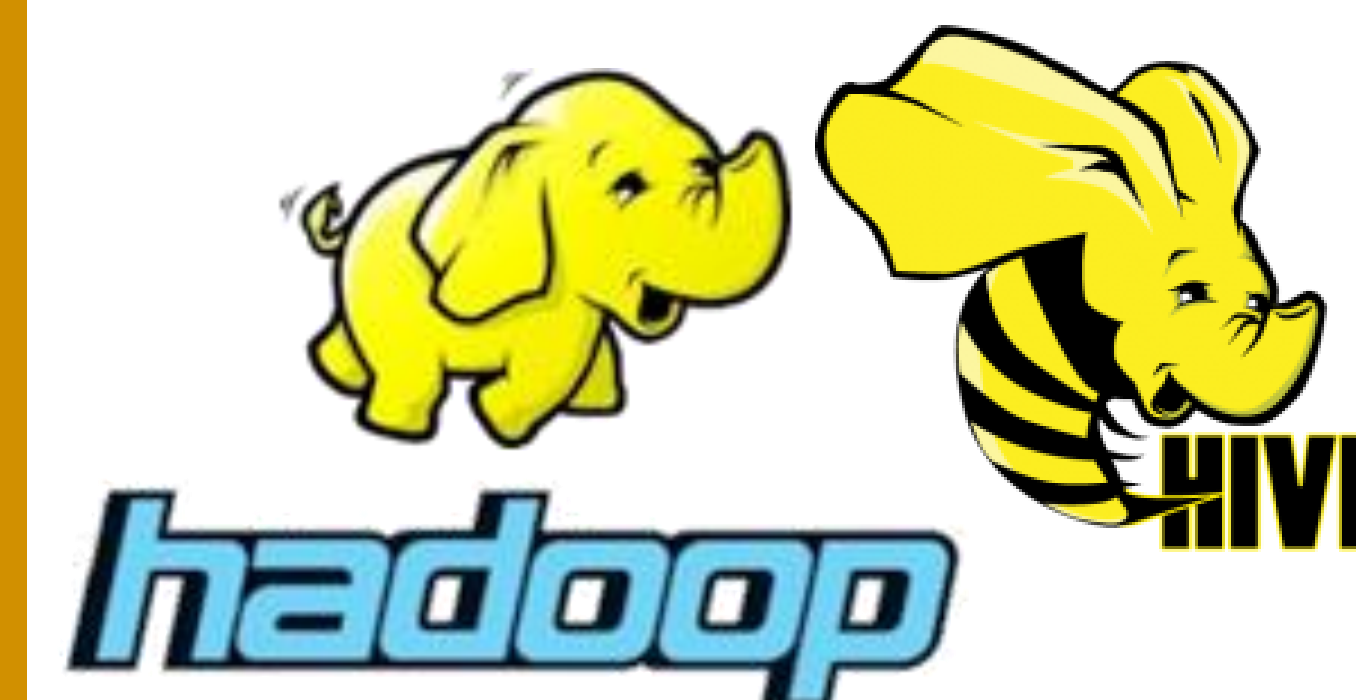
After

icn	prc_cd	rpi_id	seq_nbr	dg_cd
1	A	QWE	2	K
2	S	XYZ	2	P

We are using supervised machine learning to create a model to predict what a manually adjudicated claim should look like based on pre- and post-adjudicated data. By finding patterns in the “before” image, we can train the model to interpret and accurately predict the values that would be in an automatically adjudicated claim.

## Methodology

### Data Management



Hadoop and Hive are used for data acquisition. Examples include querying data where only certain conditions are met such as only the data where there are overlapping ICNs between multiple tables.

### Visualizations



We used the language R for visualizations. With libraries such as ggplot, tidy, dplyr, and more, R was the best choice to method to visualize our data. Visualizations include statistical information about numerical data, counts about categorical data, and other visualizations for data exploration.

### Machine Learning



Our machine learning objective is to predict each claim field with 85% accuracy. To accomplish this, we utilized the sci-kit learn package in Python to develop several models. These models include neural networks, decision trees, regression, and more.

### Documentation



By the end of the internship, we will have a comprehensive set of Jupyter Notebooks to facilitate a pipeline to recreate all the work done. It covers the following topics in depth, data acquisition, data cleaning, machine learning, and supporting background information.