

## **Fraud/Abuse Detection for Insurance Companies**

### **Problem Statement**

In several countries fraudulent behavior in insurance claim is a major problem. Data mining tools and techniques can be used to detect fraud in large sets of insurance claim data. Fraud in insurance is done by intentional deception or misrepresentation for gaining shabby benefit in the form of showing false expenditures and claim. Data mining tools and techniques can be used to detect fraud in large sets of insurance claim data. Based on a few cases that are known or suspected to be fraudulent, the anomaly detection technique calculates the likelihood or probability of each record to be fraudulent by analyzing the past insurance claims.

The analysts can then have a closer investigation for the cases that have been marked by data mining software. The anomaly or outlier detection takes vital role in data mining. The outlier detection techniques again play an important role in insurance claim fraud detection and other web usage fraud detections. A huge exigency is there for effective fraud detection mechanism to improvise the insurance management system. This research gives an idea of machine learning based fraud detection technique for insurance companies. Basically this will try to detect the fraud in health insurance area, this technique can also be used for detecting fraud in other type of insurances. In the data mining technique and machine learning approach i.e., anomaly detection using Spectral Analysis is proposed to find fraudulent records in Health & all other insurance types.

### **Background**

There are different ways to classify fraud detection techniques. Frauds could be either internal or external. The efforts to stop this type of fraud can be applied on both either on fraud preventions or fraud detections. Machine learning based fraud detection methods can also be categorized into supervised learning or unsupervised learning. There are a number of fraud detection algorithms introduced in the past. One of the most popular algorithms for fraud detection is the rule-based algorithm, which is a supervised learning method that produces classifiers using IF-THEN forms. Decision tree algorithms also fall in this category. Such algorithms need pre-labeled data and suffer the drawbacks of misclassifications. They also do not take into account of the changing features of the fraud patterns. Neural network-based analyses are also frequently used in fraud detections and have yielded significantly good results from credit card fraud detections. These methods also need labeled datasets to train the neural network models in order to make prediction.

### **Methodology**

#### **Step 1: Data collection and dataset preparation**

This will involve collection of healthcare dataset that is real world dataset so this raw dataset needs preprocessing and feature extraction by noise reduction, feature extraction and data normalization so that this preprocessed data can be used on the proposed model.

#### **Step 2: Developing a Gap cut algorithm for Fraud/Abuse Detection for Insurance Companies**

In this step, a machine learning algorithm i.e., Gap cut algorithm is designed which first determine the minimum gap between values in order to separate different communities by using the spectral analysis. It is simple and efficient in the spectral analysis to divide the nodes into communities of unknown numbers. The goal is to detect the suspicious communities between primary care physician's (PCPs) and specialists in health care claim dataset. Using this proposed spectral analysis based community detection technique, aim is to find smaller communities and identify the most suspicious physician groups.

#### **Step 3: Training and experimentation on datasets**

The Gap cut algorithm model that is machine learning algorithm will be trained on the dataset to do the Fraud/Abuse Detection for Insurance Companies and notify them.

#### Step 4: Deployment and analysis on real life scenario

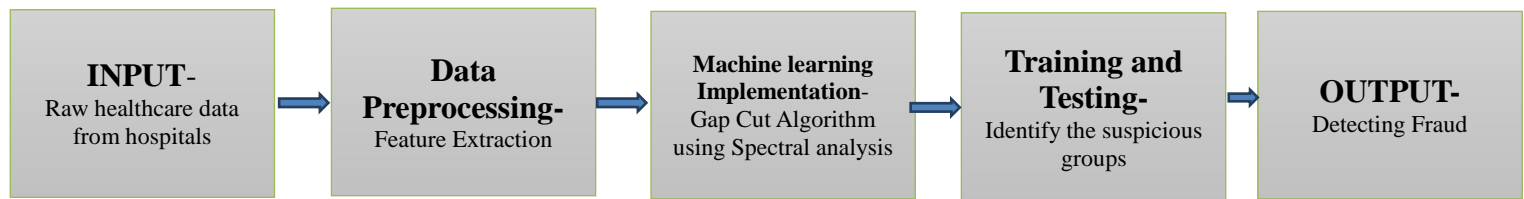


Figure.1 The proposed system architecture Fraud/Abuse Detection for Insurance Companies

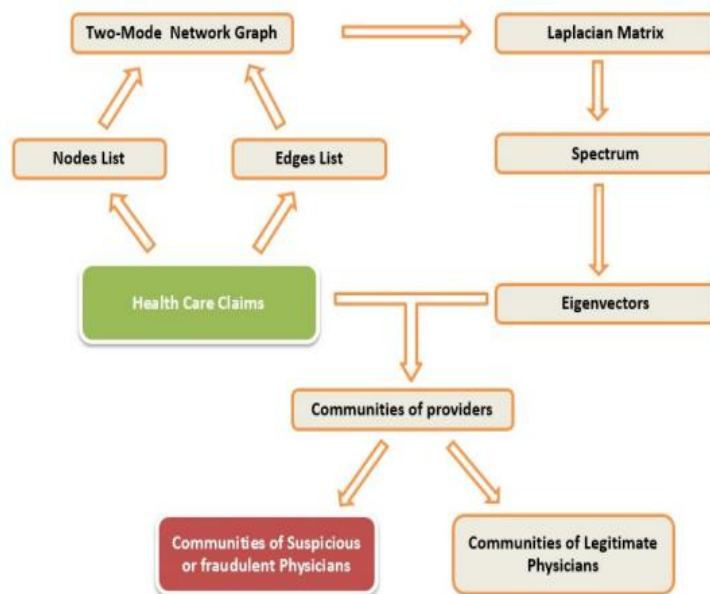


Figure.2 Overall architecture of proposed approach <sup>[1]</sup>

The trained and tested Fraud/Abuse Detection for Insurance Companies model will be deployed in a real-life scenario made by the human experts & will be leveraged for further improvement in the methodology and will follow the above architecture.

### Experimental Design

#### Dataset

Real world health care dataset is the dataset (available at <https://www.kaggle.com/c/hhp>), which contains US hospital data of patients and it is of very large size and contain tens of thousands of nodes, collected from patients admitted in the hospital will be used for experimentation and evaluation.

#### Evaluation Measures

Measures such as Accuracy, Modularity, Confusion matrix and comparison with other already present models also will be measured and evaluation is done for Fraud/Abuse Detection for Insurance Companies.

<sup>1</sup>C. Song, and A. Gangopadhyay. "A novel approach to uncover health care frauds through spectral analysis." IEEE International Conference on Healthcare Informatics (ICHI), 2013.

### **Software and Hardware Requirements**

Python based Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python and libraries such as Tensorflow will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the proposed system for Fraud/Abuse Detection for Insurance Companies.