BIG DATA FRAMEWORKS

Medicare Fraud Detection

TEAM MEMBERS

Himanshu Mittal

20MIA1035

Kushagra Joshi

20MIA1054

Sanjay Kumar

20MIA1008

MOTIVATION

Medicare fraud is a growing problem that costs taxpayers billions of dollars each year and puts vulnerable populations at risk of receiving substandard care. Developing effective methods for detecting and preventing Medicare fraud is crucial to ensure that the program can continue to provide necessary healthcare services to those who need it most. By undertaking a project focused on Medicare fraud detection, students can not only gain valuable skills in data analysis and machine learning, but also contribute to the development of solutions that can help address this pressing issue.

ABSTRACT

In this project, we present in-depth discussions on Medicare Big Data processing and exploratory experiments to demonstrate the best learners and datasets for detecting Medicare provider claims fraud. We utilize Spark on top of a Hadoop YARN cluster capable of handling these large datasets. Our unique data processing steps include data imputation, deciding which variables to preserve, aggregating data from the procedure-level to the provider-level, and constructing the Combined dataset. The fraud labels are used to assess fraud by leveraging previous exclusion information as well as Medicare payments made to currently excluded providers. The resulting processed datasets are referred to as Big Data, and we utilize these datasets for our fraud detection experiments. Overall, this project aims to demonstrate the effectiveness of data mining and machine learning for detecting Medicare fraud, providing a cost-effective solution to help combat this growing problem.

INTRODUCTION

Medicare is a federal health insurance program in the United States, providing coverage for over 62 million beneficiaries. Unfortunately, this program is vulnerable to fraudulent activities, which cost taxpayers billions of dollars every year. Therefore, the detection and prevention of Medicare fraud have become crucial in ensuring the program's sustainability and protecting its beneficiaries. In recent years, advanced technologies such as machine learning and artificial intelligence have been used to develop fraud detection models, which can analyze vast amounts of data and identify potentially fraudulent claims. This project aims to explore the use of machine learning algorithms for Medicare fraud detection by analyzing historical data, identifying patterns, and developing models that can detect fraudulent activities with high accuracy. The project's success will have significant implications for the healthcare industry and the government, as it can lead to more efficient and effective fraud detection and prevention measures.

WHAT HAS BEEN DONE SO FAR?

Medicare fraud, various techniques and tools have been used, such as data analytics, machine learning algorithms, and anomaly detection. These techniques analyze patterns in healthcare data to identify unusual or suspicious activity that could indicate fraud. Additionally, collaborations between government agencies, healthcare providers, and private organizations have been established to enhance fraud detection and prevention efforts. Overall, the field of Medicare fraud detection is continuously evolving and advancing as new technologies and approaches are developed.

WHAT WE ARE DOING?

In this project, we are aiming to develop a machine learning model for the detection of Medicare fraud. We will be using publicly available datasets to train the model to identify fraudulent activities in Medicare claims. The model will be designed to identify patterns and anomalies in the data that indicate fraudulent behavior. The final goal is to develop an accurate and reliable system that can help prevent fraudulent activities, reduce wastage of resources and ultimately improve the overall efficiency of the Medicare system.

NOVALITY

The novelty of our approach lies in the utilization of machine learning algorithms to detect and classify Medicare fraud. Traditional methods of fraud detection rely heavily on manual reviews and rules-based systems that can be time-consuming and inefficient. Our project aims to leverage the power of machine learning to automate the detection of fraud, allowing for more efficient and accurate identification of suspicious claims. Additionally, we will explore the use of anomaly detection techniques to identify fraudulent behavior patterns that may not be immediately apparent through traditional detection methods.

RESULTS OBSERVED

We aim to evaluate the performance of the developed fraud detection model by using various evaluation metrics such as accuracy, precision, recall, and F1 score. The results will be compared with existing methods and benchmarks to demonstrate the effectiveness of our proposed model. Additionally, we aim to provide insights and recommendations for the development of a more robust fraud detection system to reduce the incidence of Medicare fraud.

LITERATURE SURVEY

Medicare fraud is a major concern in the healthcare industry and has become increasingly prevalent in recent years. Various studies have proposed different approaches to detect and prevent Medicare fraud, including data mining techniques, predictive modeling, and anomaly detection. These techniques leverage large volumes of healthcare data to identify fraudulent patterns and anomalies in healthcare claims. Additionally, some studies have explored the use of machine learning algorithms, such as neural networks and decision trees, to improve the accuracy of fraud detection. However, challenges remain in implementing effective Medicare fraud detection systems due to the complexity of healthcare data and the evolving nature of fraudulent schemes. Therefore, further research is needed to develop robust and adaptive fraud detection systems.

PROPOSED METHOD

The methodology used in this project involves collecting Medicare claims data and processing it using machine learning algorithms to detect potential fraud. The data is first preprocessed by cleaning and transforming it into a usable format. Feature selection and extraction techniques are then applied to the data before training a machine learning model on a labeled dataset. The trained model is then used to predict fraud in new data. To evaluate the performance of the model, metrics such as precision, recall, and F1-score are calculated. This methodology allows for efficient and accurate detection of potential Medicare fraud cases.

SYSTEM ARCHITECTURE

The system architecture for the Medicare fraud detection project comprises of several components. The data source for the project is Medicare claims data, which is obtained from the Centers for Medicare & Medicaid Services (CMS). The data is preprocessed using techniques such as data cleaning, data transformation, and feature extraction. The preprocessed data is then used to train machine learning models, including supervised and unsupervised learning algorithms. These models are trained on historical data to identify patterns and anomalies that indicate fraudulent activities. Once the models are trained, they are deployed to a cloud-based platform for real-time fraud detection. The platform consists of several components, including an API for data ingestion, a fraud detection engine, and a dashboard for visualizing the results. The system architecture is designed to be scalable, reliable, and secure, with a focus on minimizing false positives and false negatives. The system is expected to achieve high accuracy in detecting fraudulent activities, which will help reduce Medicare fraud and save taxpayer dollars.

THANK YOU FOR LISTENING!

TEAM MEMBERS

HIMANSHU MITTAL
20MIA1035
KUSHAGRA JOSHI
20MIA1054
SANJAY KUMAR
20MIA1008