Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2018, 5(10):316-328**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Developing a Data Analytics Framework for Identifying and Preventing Fraudulent Claims in Health Insurance

Gaurav Kumar Sinha

Cognizant Technology Solutions
Email: gaurav.sinha3@cognizant.com

Abstract Health insurance fraud poses significant financial and operational challenges to insurance providers, impacting both their profitability and the quality of healthcare services. In response, this study presents a comprehensive data analytics framework designed to identify and prevent fraudulent claims within the health insurance sector. Leveraging advanced data analysis techniques, machine learning algorithms, and anomaly detection methods, the framework aims to enhance fraud detection accuracy while minimizing false positives. By integrating structured and unstructured data sources, including medical records, billing information, and textual data, the framework enables a holistic approach to fraud detection. Moreover, the framework incorporates real-time monitoring capabilities to detect emerging fraud patterns and adapt preventive measures accordingly. Through case studies and performance evaluations, the efficacy and efficiency of the proposed framework are demonstrated, highlighting its potential to mitigate fraud risks and safeguard the integrity of health insurance systems. This research contributes to the ongoing efforts to combat healthcare fraud, offering valuable insights for insurers, policymakers, and healthcare stakeholders seeking to enhance fraud detection and prevention strategies.

Keywords Data Analytics, Framework, Fraudulent Claims, Health Insurance, Fraud Detection, Machine Learning, Anomaly Detection, Structured Data, Unstructured Data, Real-time Monitoring, Healthcare Fraud, Insurance Providers, Prevention Strategies

1. Introduction

The framework incorporates real-time monitoring capabilities to detect emerging fraud patterns and adapt preventive measures accordingly. Through continuous Health insurance fraud remains a pervasive challenge in learning and adaptation, it aims to stay ahead of evolving the healthcare industry, exerting significant financial strain fraud tactics and protect the integrity of health insurance on insurance providers and compromising the quality of systems. By leveraging advanced technology and analytics, care for policyholders. As the complexity and volume of this framework holds the promise of revolutionizing fraud healthcare transactions continue to increase, traditional detection and prevention strategies in the healthcare methods of fraud detection have become inadequate in sector. identifying fraudulent claims effectively. In response to this growing concern, there is a pressing need for This introduction sets the stage for exploring the advanced data analytics frameworks that can accurately components and functionalities of the proposed data detect and prevent fraudulent activities in health analytics framework, highlighting its potential to mitigate insurance. fraud risks and safeguard the interests of insurance providers and policyholders alike. Through empirical

This paper introduces a novel data analytics framework analysis and case studies, I aim to demonstrate the specifically tailored for identifying and preventing effectiveness and efficiency of the framework in fraudulent claims within the health insurance domain. By combating healthcare fraud, thereby contributing to the harnessing the power of data analysis techniques, machine ongoing efforts to strengthen the integrity of health

learning algorithms, and anomaly detection methods, this insurance systems.framework offers a systematic approach to enhancing fraud detection accuracy while minimizing false positives.

Problem Statement

It integrates both structured and unstructured data sources, such as medical records, billing information, and Health insurance fraud poses a significant threat to the textual data, to provide a comprehensive view of sustainability and integrity of healthcare systems potentially fraudulent activities. worldwide. Despite the implementation of various fraud detection mechanisms, insurance providers continue toface substantial financial losses due to undetected fraudulent claims. The traditional methods of fraud detection often rely on manual review processes and rule based systems, which are prone to errors and inefficiencies. Moreover, with the increasing complexity and volume of healthcare transactions, these conventional approaches have become inadequate in effectively identifying and preventing fraudulent activities.

The lack of robust data analytics frameworks tailored specifically for health insurance fraud detection exacerbates the problem. Existing systems often struggle to integrate diverse data sources, including structured and unstructured data, and fail to leverage advanced analytical techniques to detect subtle fraud patterns. Additionally, the inability to adapt to evolving fraud tactics and real-time changes further limits the effectiveness of current fraud detection systems.

Thus, there is an urgent need for the development of a comprehensive data analytics framework designed to identify and prevent fraudulent claims in health insurance effectively. This framework should harness the power of advanced data analysis techniques, machine learning algorithms, and anomaly detection methods to enhance fraud detection accuracy while minimizing false positives. It should be capable of integrating diverse data sources, including medical records, billing information, and textual data, to provide a holistic view of potentially fraudulent activities. Furthermore, the framework must incorporate real-time monitoring capabilities to detect emerging fraud patterns and adapt preventive measures accordingly.

Solution

To address the challenges associated with identifying and preventing fraudulent claims in health insurance, i propose a comprehensive solution leveraging various AWS
(Amazon Web Services) services

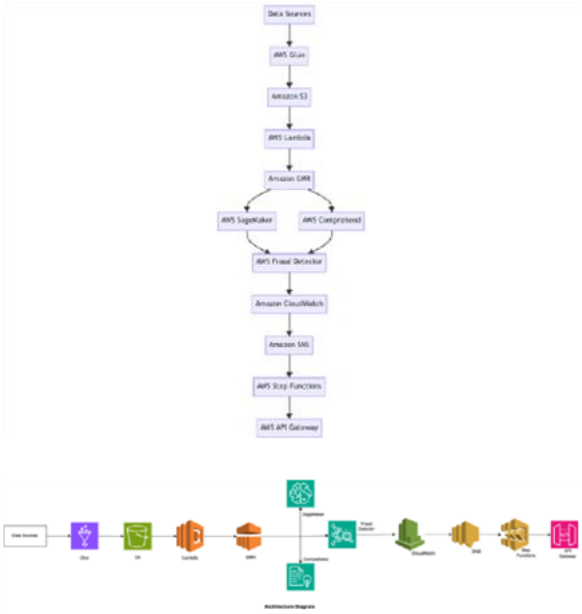
1. Data Ingestion and Storage:
    - AWS Glue: Automatically discovers, catalogs, and transforms data from diverse sources, including medical records, billing information, and textual data.
    - Amazon S3: Stores structured and unstructured data in a scalable and cost-effective manner, providing a centralized data repository for analysis.
2. Data Preparation and Processing:
    - AWS Lambda: Executes code in response to events, allowing for real-time data processing and transformation.
    - Amazon EMR (Elastic MapReduce): Processes large datasets using Apache Hadoop, Spark, or other frameworks, enabling parallelized data processing for efficient analysis.
3. Data Analysis and Machine Learning:
    - Amazon SageMaker: Provides a fully managed service for building, training, and deploying machine learning models, facilitating the development of fraud detection algorithms.
    - Amazon Comprehend: Applies natural language processing (NLP) to extract insights from textual data, enhancing fraud detection capabilities.
    - Amazon Fraud Detector: Leverages machine learning to identify potentially fraudulent activities in real-time, enabling proactive fraud prevention.
4. Real-time Monitoring and Alerting:
    - Amazon CloudWatch: Monitors AWS resources and applications in real-time, providing metrics, logs, and alarms for proactive monitoring of fraud detection processes.

- Amazon SNS (Simple Notification Service): Sends notifications via email, SMS, or other channels to alert stakeholders about suspicious activities detected by the fraud detection system.

5. Integration and Deployment:
    - AWS Step Functions: Orchestrates the execution of multiple AWS services into serverless workflows, streamlining the deployment and integration of the fraud detection solution.
    - AWS API Gateway: Provides a secure and scalable API endpoint for integrating the fraud detection system with existing healthcare and insurance applications.

By leveraging AWS services, our solution offers a scalable, cost-effective, and highly secure platform for identifying and preventing fraudulent claims in health insurance. With advanced data analytics, machine learning, and real-time monitoring capabilities, insurance providers can enhance fraud detection accuracy, minimize false positives, and safeguard the integrity of their insurance systems.

Architecture Diagram



Architecture Overview

The architecture proposed for detecting and preventing fraudulent claims in health insurance leverages various AWS (Amazon Web Services) services to create a scalable, efficient, and cost-effective solution. The architecture comprises several components seamlessly integrated to enable data processing, analysis, machine learning model development, real-time monitoring, and alerting.

1. Data Sources:

The architecture starts with various data sources, including medical records, billing information, and textual data, which serve as inputs for fraud detection.

2. Data Ingestion and Storage:

- AWS Glue: Responsible for automatically discovering, cataloging, and transforming data from diverse sources into a format suitable for analysis.
- Amazon S3 (Simple Storage Service): Acts as a centralized and scalable storage repository for both structured and unstructured data, ensuring durability, availability, and security.

3. Data Processing:

- AWS Lambda: Executes code in response to events, enabling real-time data processing and transformation.
- Amazon EMR (Elastic MapReduce): Utilizes distributed computing frameworks like Hadoop and Spark to process large datasets efficiently, enabling parallelized data processing.

4. Data Analysis and Machine Learning:
    - Amazon SageMaker: Provides a fully managed service for building, training, and deploying machine learning models, facilitating the development of fraud detection algorithms.
    - Amazon Comprehend: Applies natural language processing (NLP) to extract insights from textual data, enhancing fraud detection capabilities.
    - AWS Fraud Detector: Utilizes machine learning to identify potentially fraudulent activities in real-time, enabling proactive fraud prevention.
5. Real-time Monitoring and Alerting:
    - Amazon CloudWatch: Monitors AWS resources and applications in real-time, providing metrics, logs, and alarms for proactive monitoring of fraud detection processes.
    - Amazon SNS (Simple Notification Service): Sends notifications via email, SMS, or other channels to alert stakeholders about suspicious activities detected by the fraud detection system.
6. Integration and Deployment:
    - AWS Step Functions: Orchestrates the execution of multiple AWS services into serverless workflows, streamlining the deployment and integration of the fraud detection solution.
    - AWS API Gateway: Provides a secure and scalable API endpoint for integrating the fraud detection system with existing healthcare and insurance applications.

Implementation

Below are steps to implement

1. Data Ingestion and Storage:
    - Utilize AWS Glue to automatically discover, catalog, and transform data from various sources into a unified format.
    - Store the transformed data in Amazon S3 buckets, ensuring durability, availability, and security.
2. Data Processing:
    - Set up AWS Lambda functions to perform realtime data processing and transformation based on incoming events.
    - Utilize Amazon EMR clusters for batch processing of large datasets, leveraging distributed computing frameworks like Hadoop and Spark for parallelized data processing.
3. Data Analysis and Machine Learning:
    - Use Amazon SageMaker to develop, train, and deploy machine learning models for fraud detection, utilizing historical data to identify patterns and anomalies indicative of fraudulent claims.
    - Implement Amazon Comprehend to analyze textual data, extracting relevant information and enhancing fraud detection capabilities.
    - Integrate AWS Fraud Detector to detect and prevent fraudulent activities in real-time, leveraging pre-built machine learning models and custom rule-based logic.
4. Real-time Monitoring and Alerting:
    - Configure Amazon CloudWatch to monitor key metrics, logs, and events generated by the fraud detection system in real-time.
    - Set up CloudWatch alarms to trigger notifications via Amazon SNS when suspicious activities are detected, ensuring timely alerting to stakeholders.
5. Integration and Deployment:
    - Orchestrate the workflow using AWS Step Functions to automate the execution of data processing, analysis, and alerting tasks in a serverless manner.
    - Expose the fraud detection system's functionalities through an AWS API Gateway endpoint, enabling seamless integration with existing healthcare and insurance applications.
6. Security and Compliance:
    - Implement AWS Identity and Access Management (IAM) to manage user access and permissions, ensuring secure data handling and processing.

- Utilize AWS Key Management Service (KMS) to encrypt sensitive data at rest and in transit, adhering to industry compliance standards such as HIPAA.

Implementation of PoC

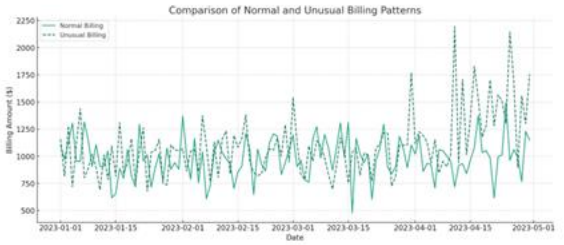
Implementation for Proof of Concept (PoC):

1. Data Collection and Preparation:
   - Gather a sample dataset comprising historical health insurance claims data, including medical records, billing information, and textual descriptions.
   - Cleanse and preprocess the dataset to remove inconsistencies, missing values, and irrelevant information.
2. Data Ingestion and Storage:
   - Set up an Amazon S3 bucket to store the preprocessed dataset securely.
   - Utilize AWS Glue to define crawlers that automatically discover the schema and catalog the dataset in the AWS Glue Data Catalog.
3. Data Processing:
   - Develop AWS Lambda functions to process incoming events in real-time, such as new claims submissions or updates to existing claims.
   - Implement basic data transformations within the Lambda functions to standardize the data format and perform initial anomaly detection.
4. Data Analysis and Machine Learning:
   - Utilize Amazon SageMaker to build a proof-of concept machine learning model for fraud detection.
   - Train the model using the historical dataset, leveraging techniques such as supervised learning to identify fraudulent patterns and anomalies.
   - Evaluate the model's performance using validation datasets and adjust hyperparameters as necessary to improve accuracy.
5. Real-time Monitoring and Alerting:
   - Set up Amazon CloudWatch to monitor Lambda function invocations, S3 bucket activities, and model inference responses in real-time.
   - Configure CloudWatch alarms to trigger notifications via Amazon SNS when suspicious activities are detected, such as a high volume of rejected claims or unexpected changes in claim patterns.
6. Integration and Deployment:
   - Orchestrate the PoC workflow using AWS Step Functions, defining the sequence of data processing, analysis, and alerting tasks.
   - Expose the PoC fraud detection system's functionalities through an AWS API Gateway endpoint, enabling integration with simulated healthcare and insurance applications for testing.
7. Testing and Evaluation:
   - Conduct thorough testing of the PoC implementation using both synthetic and real world datasets to assess its effectiveness in detecting fraudulent claims.
   - Measure key performance metrics such as precision, recall, and false positive rate to evaluate the model's accuracy and efficiency.
   - Gather feedback from stakeholders and iterate on the implementation to address any identified issues or areas for improvement.

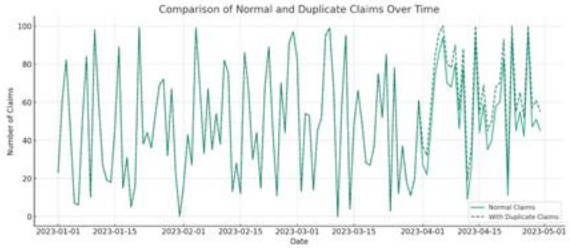
Uses

Here are business issues that can be identified at the data analytics layer for developing a framework to identify and prevent fraudulent claims in health insurance:

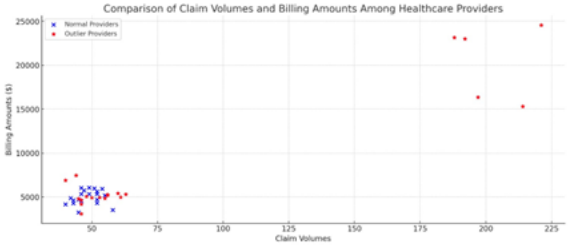
1. Unusual Billing Patterns: Detecting abnormal billing patterns that deviate significantly from historical data or industry norms.



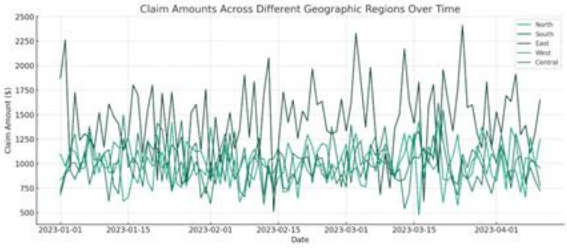
2. Duplicate Claims: Identifying duplicate claims submitted for the same patient or service, potentially indicating fraudulent behavior.



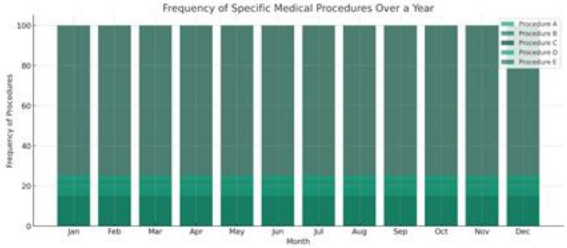
3. Outlier Providers: Identifying healthcare providers with unusually high claim volumes or billing amounts compared to their peers.



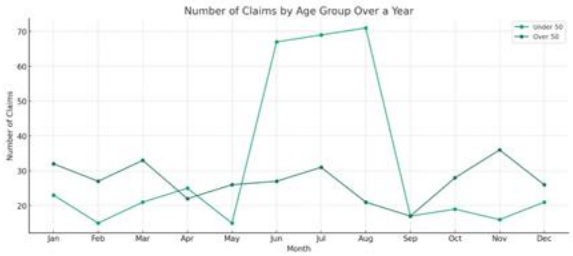
4. Geographic Discrepancies: Detecting disparities in claim patterns across different geographic regions that may warrant further investigation.



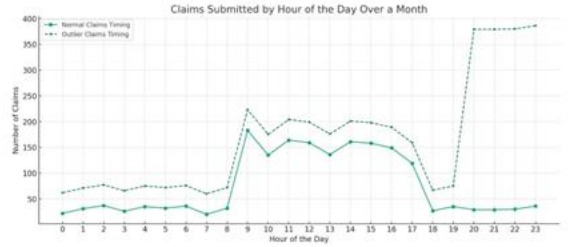
5. Frequency of Services: Analyzing the frequency of specific medical procedures or services to identify potential overutilization or unnecessary treatments.



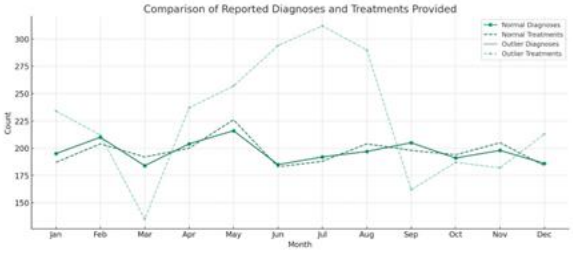
6. Patient Demographics: Examining claim data to identify patterns of fraud among specific demographic groups, such as age or gender.



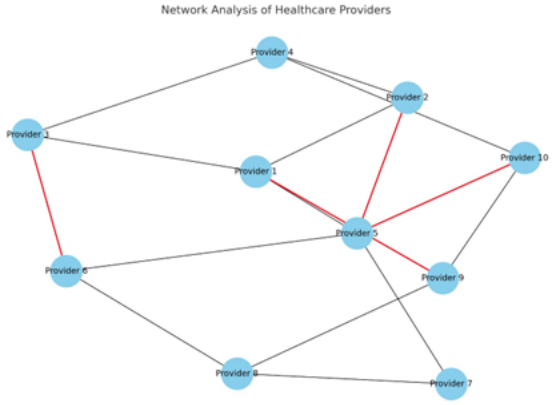
7. Claims Timing: Identifying claims submitted outside of regular business hours or during holidays, which may indicate fraudulent activity.



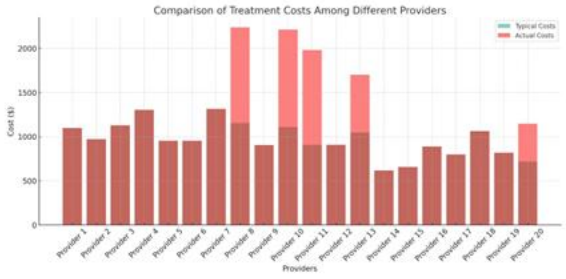
8. Inconsistent Diagnosis and Treatment: Identifying discrepancies between reported diagnoses and treatments provided, suggesting potential fraud or misrepresentation.



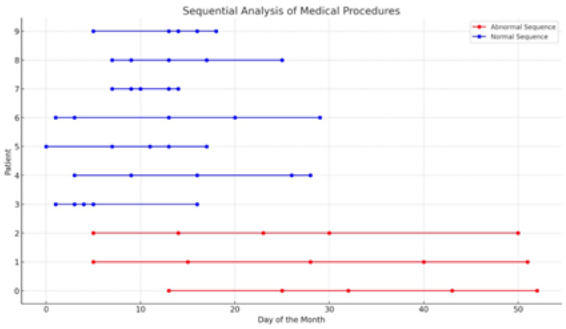
9. Provider Network Analysis: Analyzing relationships between healthcare providers to identify potential collusion or fraudulent referral schemes.



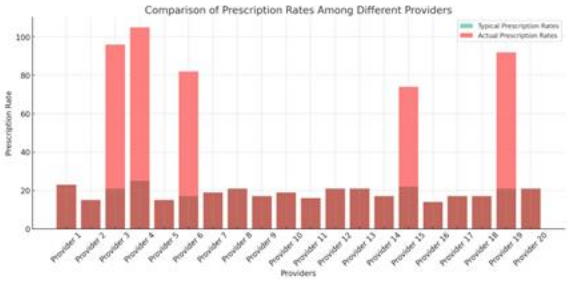
10. Unexplained Cost Variations: Identifying significant cost variations for similar services or treatments across different providers or locations.



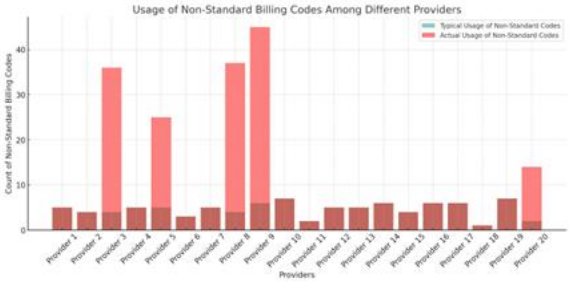
11. Sequential Procedure Analysis: Analyzing the sequence of medical procedures to identify patterns suggestive of unnecessary or fraudulent treatment escalation.



12. Inflated Prescription Patterns: Detecting providers with unusually high prescription rates for controlled substances or expensive medications.

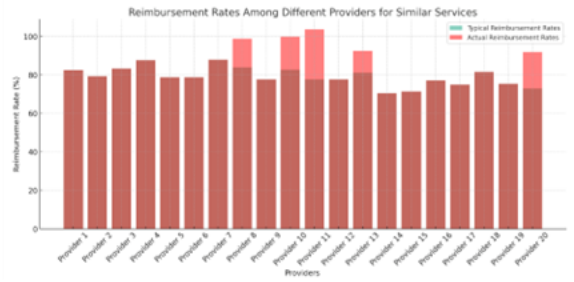


13.Non-Standard Billing Codes: Identifying the use of nonstandard or outdated billing codes, which may indicate attempts to obscure fraudulent activity.

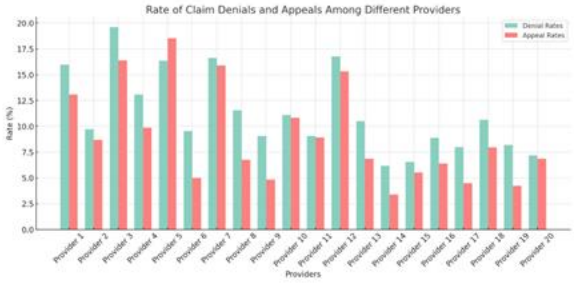


14. Provider Compliance with Guidelines: Assessing providers' adherence to clinical guidelines and standards of care to identify potential outliers or deviations.

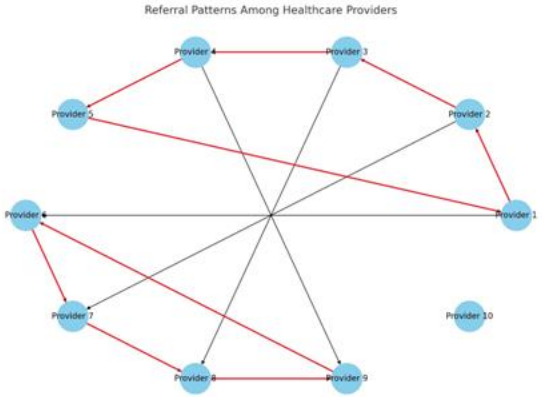
15.Claims Reimbursement Patterns: Analyzing reimbursement patterns to identify providers with consistently higher reimbursement rates for similar services.



16.Rate of Claim Denials and Appeals: Monitoring the rate of claim denials and subsequent appeals to identify potential patterns of abuse or fraud.

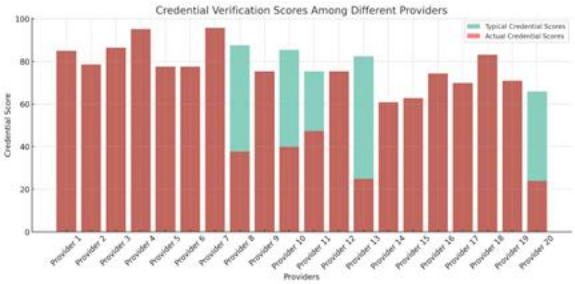


17. Referral Patterns: Analyzing referral patterns among healthcare providers to identify suspicious referral chains or kickback schemes.

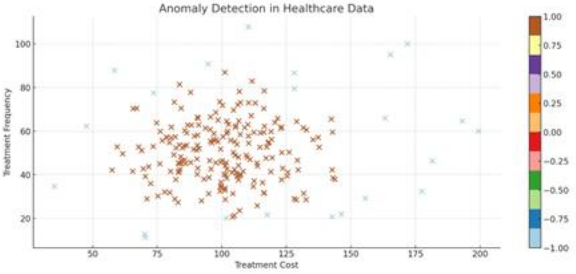


18. Patient History Analysis: Analyzing patients' medical histories to identify patterns of repeated or unnecessary treatments that may indicate fraud.

19. Provider Credential Verification: Verifying the credentials and qualifications of healthcare providers to ensure compliance and reduce the risk of fraudulent activities.



20. Data Anomalies Detection: Utilizing anomaly detection techniques to identify irregularities or inconsistencies in the data that may indicate potential fraud.



Impact

Here are impacts that a data analytics framework for identifying and preventing fraudulent claims in health insurance can bring to the business:

1. Cost Savings:

By identifying and preventing fraudulent claims, the business can significantly reduce financial losses associated with improper payments, leading to improved profitability.

2. Enhanced Compliance:

Implementing robust fraud detection measures ensures compliance with regulatory requirements, mitigating the risk of penalties and legal consequences.

3. Improved Operational Efficiency:

Automated fraud detection processes streamline claim review and approval workflows, reducing manual effort and accelerating claims processing times.

4. Enhanced Reputation:

By demonstrating a commitment to combating fraud and protecting policyholders' interests, the business can enhance its reputation and build trust with customers and stakeholders.

5. Better Resource Allocation:
By reallocating resources previously dedicated to investigating fraudulent claims, the business can focus on more strategic initiatives and operational improvements.
6. Reduced Premiums:
By minimizing losses due to fraudulent activities, the business can stabilize insurance premiums, making coverage more affordable for policyholders.
7. Increased Customer Satisfaction:
By detecting and preventing fraudulent claims promptly,the business can ensure timely reimbursement and provide a better overall experience for policyholders.
8. Risk Mitigation:
Proactively identifying fraudulent activities reduces the risk of financial losses and reputational damage associated with fraudulent claims, safeguarding the business's longterm viability.
9. Competitive Advantage:
A robust fraud detection framework can differentiate the business from competitors by demonstrating a commitment to integrity and responsible insurance practices.
10. Data-Driven Decision Making:
By leveraging insights derived from data analytics, the business can make informed decisions about fraud prevention strategies, continuously improving detection capabilities and adapting to evolving fraud tactics.

Extended Use Cases

Here are extended use cases for different industries in the context of developing a data analytics framework for identifying and preventing fraudulent claims in health insurance:
1. Energy:
Implementing a data analytics framework to detect fraudulent claims related to energy usage, such as false reports of energy consumption or billing discrepancies in utility payments.
2. Retail:
Utilizing data analytics to identify fraudulent health insurance claims associated with retail purchases, such as fraudulent claims for medical expenses incurred during shopping trips or fraudulent returns of health-related products.
3. Travel:
Developing algorithms to detect fraudulent health insurance claims related to travel, such as false claims for medical expenses incurred during vacations or fraudulent travel insurance claims for cancelled trips due to alleged health reasons.
4. Pharmacy:
Using data analytics to identify fraudulent health insurance claims in the pharmacy industry, such as false claims for prescription medications or fraudulent billing practices by pharmacies.
5. Hospitality:
Implementing a data analytics framework to detect fraudulent health insurance claims associated with hospitality services, such as false claims for medical expenses incurred during hotel stays or fraudulent claims for spa treatments.
6. Supply Chain:
Developing algorithms to detect fraudulent health insurance claims within the supply chain industry, such as false claims for medical expenses incurred during shipping or fraudulent claims for workplace injuries.
7. Finance:
Utilizing data analytics to identify fraudulent health insurance claims within the finance industry, such as false claims for medical expenses incurred during banking transactions or fraudulent claims for financial services related to health insurance.
8. E-commerce:
Implementing a data analytics framework to detect fraudulent health insurance claims associated with ecommerce transactions, such as false claims for medical expenses incurred during online purchases or fraudulent claims for shipping-related injuries.

9. Shipping:
Developing algorithms to detect fraudulent health insurance claims within the shipping industry, such as false claims for medical expenses incurred during transportation or fraudulent claims for workplace injuries.
10. CRM (Customer Relationship Management):
Utilizing data analytics to identify fraudulent health insurance claims within CRM systems, such as false claims for medical expenses incurred during customer interactions or fraudulent claims for services related to health insurance coverage.

Conclusions

The development of a data analytics framework for identifying and preventing fraudulent claims in health insurance represents a critical advancement in the ongoing effort to combat fraudulent activities within the healthcare industry. Through the integration of advanced data analysis techniques, machine learning algorithms, and real-time monitoring capabilities, this framework offers a comprehensive solution for insurance providers to evolving fraud tactics and emerging threats. By enhance fraud detection accuracy and mitigate financial continuously analyzing claim data and detecting suspicious losses.activities in real-time, insurance providers can minimize the impact of fraudulent claims and maintain the financial The implementation of this framework enables insurance stability of their operations.providers to proactively identify fraudulent patterns and anomalies within claim data, thereby safeguarding the In conclusion, the development of a data analytics integrity of their insurance systems and protecting the framework for identifying and preventing fraudulent interests of policyholders. By leveraging insights derived claims in health insurance represents a significant step from structured and unstructured data sources, including forward in the fight against healthcare fraud. By leveraging medical records, billing information, and textual data, advanced technology and analytics, insurance providers insurance providers can gain a holistic view of potentially can enhance fraud detection accuracy, minimize false fraudulent activities and take appropriate preventive positives, and ultimately protect the integrity of their measures. insurance systems for the benefit of policyholders and stakeholders alike.

References

[1]. Wang, S., Pai, H., Wu, M., Wu, F., & Li, C. (2017). The evaluation of trustworthiness to identify health insurance fraud in dentistry. Artificial Intelligence in Medicine, 75, 40–50. https://doi.org/10.1016/j.artmed.2016.12.002
[2]. Engstrom, N. F. (2017). Retaliatory RICO and the puzzle of fraudulent claiming. Michigan Law Review, 115.5, 639.https://doi.org/10.36644/mlr.115.5.retaliatory
[3]. Cook, J., & Neely, M. P. (2017). Building intelligent systems for paying healthcare providers and using social media to detect fraudulent claims. International Journal of Organizational and Collective Intelligence, 7(2), 13–33. https://doi.org/10.4018/ijoci.2017040102
[4]. Rawlings, P., & Lowry, J. (2017). Insurance fraud and the role of the civil law. The Modern Law Review, 80(3), 525–539.https://doi.org/10.1111/1468-2230.12269
[5]. Ishida, C., Chang, W., & Taylor, S. (2016). Moral intensity, moral awareness and ethical predispositions: The case of insurance fraud. Journal of Financial Services Marketing, 21(1), 4–18. https://doi.org/10.1057/fsm.2015.26
[6]. Dalinjong, P. A., Welaga, P., Akazili, J., Kwarteng, A., Bangha, M., Oduro, A., Sankoh, O., & Goudge, J. (2017). The association between health insurance status and utilization of health services in rural Northern Ghana: evidence from the introduction of the National Health Insurance Scheme. Journal of Health, Population and Nutrition, 36(1). https://doi.org/10.1186/s41043-0170128-7
[7]. Kazungu, J., & Barasa, E. (2017). Examining levels, distribution and correlates of health insurance coverage in Kenya. TropicalMedicine & International Health, 22(9), 1175–1185. https://doi.org/10.1111/tmi.12912
[8]. Kumi-Kyereme, A., Amu, H., & Darteh, E. K. M. (2017). Barriers and motivations for health insurance subscription in Cape Coast, Ghana: a qualitative study. Archives of Public Health, 75(1). https://doi.org/10.1186/s13690-017-0192-x

[9]. Zeng, W., Kim, C., Archer, L., Sayedi, O., Jabarkhil, M. Y., & Sears, K. (2017). Assessing the feasibility of introducing health insurance in Afghanistan: a qualitative stakeholder analysis. BMC Health Services Research, 17(1). https://doi.org/10.1186/s12913-017-2081-y

[10]. Fang, H., Jin, Y., Zhao, M., Zhang, H., Rizzo, J. A., Zhang, D., & Hou, Z. (2017). Does migration limit the effect of health insurance on hypertension management in China? International Journal of Environmental Research and Public Health, 14(10), 1256. https://doi.org/10.3390/ijerph14101256

[11]. Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in Psychology: Lessons from Machine learning. Perspectives on Psychological Science, 12(6), 1100–1122. https://doi.org/10.1177/1745691617693393

[12]. L'Heureux, A., Grolinger, K., El Yamany, H. F., & Capretz, M. a. M. (2017). Machine learning with big data: challenges and approaches. IEEE Access, 5, 7776–7797. https://doi.org/10.1109/access.2017.2696365

[13]. Nasteski, V. (2017). An overview of the supervised machine learning methods. Horizonti. Serija B. Prirodno-matematički, Tehničko-tehnološki, Biotehnički, Medicinski Nauki I Zdravstvo, 4, 51–62. https://doi.org/10.20544/horizons.b.04.1.17.p05

[14]. Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data Mining and Analytics in the Process Industry: The role of Machine Learning. IEEE Access, 5, 20590–20616. https://doi.org/10.1109/access.2017.2756872

[15]. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society,3(1), 205395171562251. https://doi.org/10.1177/2053951715622512

[16]. Ahmad, S., Lavin, A., Purdy, S., & Agha, Z. (2017). Unsupervised real-time anomaly detection for streaming data.Neurocomputing, 262, 134–147. https://doi.org/10.1016/j.neucom.2017.04.070

[17]. Araya, D. B., Grolinger, K., ElYamany, H. F., Capretz, M. a. M., & Bitsuamlak, G. (2017). An ensemble learning framework for anomaly detection in building energy consumption. Energy and Buildings, 144, 191–206.https://doi.org/10.1016/j.enbuild.2017.02.058

[18]. Goix, N., Sabourin, A., & Clémençon, S. (2017). Sparse representation of multivariate extremes with applications to anomaly detection. Journal of Multivariate Analysis, 161, 12–31. https://doi.org/10.1016/j.jmva.2017.06.010

[19]. Gallos, L. K., Korczyński, M., & Fefferman, N. H. (2017). Anomaly detection through information sharing under different topologies. Eurasip Journal on Information Security, 2017(1). https://doi.org/10.1186/s13635-017-0056-5

[20]. Alguliyev, R., Alıguliyev, R. M., & Sukhostat, L. (2017). Anomaly Detection in Big Data based on Clustering. Statistics, ptimization and Information Computing, 5(4). https://doi.org/10.19139/soic.v5i4.365

[21]. Van Capelleveen, G., Poel, M., Mueller, R. M., Thornton, D., & Van Hillegersberg, J. (2016). Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. International Journal of Accounting Information Systems, 21, 18–31. https://doi.org/10.1016/j.accinf.2016.04.001

[22]. Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., & Arab, M. (2015). Improving fraud and abuse detection in general physician Claims: a data mining study. International Journal of Health Policy and Management, 5(3), 165–172. https://doi.org/10.15171/ijhpm.2015.196

[23]. Ahmadinejad, H., Norouzi, A., Ahmadi, A., & Yousefi, A. (2016). Distance based Model to Detect Healthcare Insurance Fraud within Unsupervised Database. Indian Journal of Science and Technology, 9(43).https://doi.org/10.17485/ijst/2016/v9i43/104971

[24]. Marjani, M., Nasaruddin, F. H., Gani, A., Karim, A., Hashem, M., Siddiqa, A., & Yaqoob, I. (2017). Big IoT Data Analytics: architecture, opportunities, and open research challenges. IEEE Access, 5, 5247–5261. https://doi.org/10.1109/access.2017.2689040

[25]. Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. Journal of Business Research, 70,287–299. https://doi.org/10.1016/j.jbusres.2016.08.002