Streaming Network Analysis and Claim Velocity Scoring: Real-Time AI Frameworks for Insurance Fraud Detection and Prevention

Dr. Fei-Fei Li, Professor of Computer Science, Stanford University (Branch outside normal colleges)

1. Introduction

The insurance industry is plagued with an onslaught of fraudulent claims. In 2020 alone, U.S. insurance carriers collectively lost 7% to 10% of earned premiums due to fraud, costing an estimated $80 billion. As a result, one of the biggest priorities for many carriers is innovating their fraud detection and prevention methods. Instead of waiting for the claim to be paid and then seeking restitution, insurance carriers would greatly prefer to preemptively stop the payment from going out. The perfect time to catch insurance fraud is in real time. Real-time processing refers to the capacity of a system to process data immediately, without any delay. Processing the available data immediately leads to immediate advantages. In fact, when real-time data attains the capacity for analysis, it becomes an extremely valuable tool. Integrating AI and real-time processing increases net margin. Net margin expansion is the difference between fraud detection loss ratio improvements and the one-off cost of setting up the systems.

Given the enormous losses that can be attributed to insurance fraud and the resultant increase in cost for an individual policyholder and society as a whole, the selection of this essay topic is pertinent. The essay preoccupation is the swift detection and prevention of fraudulent insurance claims. Since preventing the payout is not just advantageous for the insurance carrier but also endorsed by relevant organizations, this essay will be directly answered by focusing on fraud detection. The inclusive objectives of the essay are to: • Discuss the key fraud detection challenges of the insurance carrier. • Examine the steady evolution and changing roles of real-time processing in fraud detection. • Analyze real-time AI's mechanics and its composition. • Scrutinize the potential use cases for real-time AI in fraud detection. • Contemplate real-time AI's

operational implications, particularly the degree to which it potentially offers the carrier a 'window of opportunity' to prevent loss.

1.1. Background and Significance of Insurance Fraud Detection

As insurance fraud has become a pressing issue, it has given rise to the development of numerous fraud detection methods. Some solutions use statistical, rule-based, or artificial intelligence to detect insurance fraud by creating fraud-related indicators. The currently used fraud detection systems have become very efficient in identifying known insurance fraud behavior through internal and external indicators. Unfortunately, a cut-off exists. Since these indicators are based on fraud signals, they are generally capable of identifying only the simplest chunks of fraud, mainly committed by inexperienced fraudsters.

In conclusion, insurance fraud detection is of paramount importance since it allows the insurance company to keep its fraud loss ratio low and minimize premiums. Policyholders will ultimately be the main victims in the event of insolvency of an insurance company because they will lose premiums pre-paid in advance. To tackle this issue, insurance companies have tried to use various new security measures and to innovate more advanced technologies based on different detection and handling methods to support it. To comply with legislation. Finally, complex fraud schemes can erode policyholder confidence in the financial system, since fraud is often associated with an insurance company and can promote reductions or outright cancellation of coverage by reinsurers, making retreat markets important. Since this is a threat to the availability and/or affordability of coverage, it is in the public interest to develop or pursue efforts to stop such behavior since society fills the coverage gap.

2. Fundamentals of Insurance Fraud

Fraud is present in many areas of business, and the insurance sector is no exception. It is a pressing concern for insurers, who have estimated the financial cost of insurance fraud at 15-20 percent of earned premiums. This problem is pushed down to policyholders in the form of increased premiums and limited choices for coverage, as insurers attempt to regain the funds lost due to fraud. In addition to the threat posed to policyholders, insurance fraud can also lead to the condemnation of well-intentioned individuals who have their claims for losses refuted. Insurers, in particular, run the risk of some insured individuals coming to believe they were defrauded by their insurer. The various means

by which insurance fraud occurs are wide-ranging. Health insurance fraud tends to revolve around situations where service providers make claims separate from services actually provided to policyholders. Dental offices, for example, could complete a discounted denture repair, later fabricating a second, more expensive repair while resubmitting the original request to the insurance company for reimbursement. Car insurance fraud could involve vehicle owners claiming rental costs or the cost of a new car while their depreciated car was repaired. One of the defining characteristics of insurance fraud is the real economic loss faced by the insurer, not just a reduction of account balances, from the deceit of policyholders, vendors, or insurers. Epidemiological studies on the prevalence of insurance fraud are scarce.

2.1. Types of Insurance Fraud

There isn't a single universal definition for insurance fraud. In general, insurance fraud can be defined as an act committed intentionally by an insured or insurer to make money in an illicit way. Insurance fraud takes place across a wide variety of insurance sectors, such as household, automobile, and health insurance. This type of fraud can also occur because of a lack of understanding of the system or a lack of legitimacy in the insured's eyes. Herein, numerous types of fraud were identified, but these types are certainly non-exhaustive. We summarize some types of insurance fraud that take place using health fraud committed by policyholders.

Insurance fraud in terms of claimed expenses involves, for example, inflated claims. In this case, an individual policyholder, care provider, or representative may inflate the amount of the claim with or without the insured's involvement. By doing so, they obtain more or higher compensation than is allowed by law. This is a prevalent form of fraud since several combinations of insured and care provider representatives can be involved in the fraud. Examples of forged claims include altered invoices, overstated billings, inflating quantities, and finalizing a new product. Different branches of the insurance industry are exposed to distinct challenges in insurance fraud, as several distinguishable characteristics are revealed. Home and motor insurance naturally have more invasive fraud than many sectors and possess specific characteristics in terms of the perpetration methods fraudsters choose. For example, home insurance fraud will be about damaging a fridge or a television, whereas motor fraud can be about stealing a car or staging an accident. Insurance fraud is divided into contents of the claim and application, with

home insurance highlighted for having a 1 in 5 ratio of application to claim fraud, whereas it is then a ratio of 1 in 10 for motor insurance. Some of the aspects that distinguish the various categories from one another are driven by the methods the fraudsters use, the sums of money or emoluments at issue in particular lines of insurance, and differences in the frequency of fraud. For instance, relationship fraud is spread over a broad range of different types of insurance, but is particularly likely to occur when products are sold on the basis of the insured person and his financial history, as in card-protection insurance and income protection insurance. In others, such as controversial illness or death insurance, the motivations are usually financial. A statistical analysis of which types of fraud each investigator has had experience with, or which kinds of fraud make up the experience of numerous losses to this company, might shed light on their prevalence, but it is fair to say that most persons signing insurance proposals do so in good faith and that the best approach in detecting insurance fraud is to assume good faith unless it is detected to be otherwise. Generally, the impact of fraud on the activities of an MFI is aggravated by its potential to take money out of the business. Opportunities for premium increases can exert an ultimate cost on profitability, but can also be opportunities for management to consider fresh strategies and coping mechanisms such as natural risk sharing through greatly slowed bounds renewal. Finally, fraud is also very likely to cause serious stress for staff involved, particularly those communicating directly with the customer. Financial advice and services may be helpful. Experiences showing how policyholders might behave if they feel they have not been treated fairly also suggest the management of a successful investigation.

3. AI and Machine Learning in Fraud Detection

Introduction Artificial intelligence (AI) and machine learning are deemed the "next big thing" in combating insurance fraud, a problem that costs the industry billions of dollars a year. Whereas manual tactics to spot this type of crime can be time-consuming and inadequate, AI can automate and enhance the process, accomplishing it more quickly and often at a lower cost. At its heart, AI software applies advanced algorithms to straightforward data. As an illustration, machine learning agencies can instruct a computer application to recognize and address fraud by displaying it hundreds or thousands of distinct insurance claims or assertions that have previously been identified as false. The application then directly sifts through claims and situations in real time,

seeking to single out those that correspond to deceitful patterns. Manufactured fraud detection requests go through two types of examination that align with the divisions of in-sample and out-of-sample exposure. The first division, in-sample exposure, identifies system-centered or entitlement impropriety exhibited by claimants or employees who are pushing demands for reimbursements. The second type of examination, out-of-sample exposure, generates client assessments of prime prospects. These prospect ratings were decisive in agencies' decisions to offer loans, credit approvals, mortgages, employment, and insurance policies. An established global practice that utilizes AI and machine learning can benefit insurers because it distinguishes them from more traditional competitors, which is significantly vital in an extremely harsh economic environment. They manage to reduce fraud and directly digitize a large section of their operations. It is important to note that there is an ongoing conversation about the ethical implications of using AI to highlight fraudulent behavior in developed markets. Machines must impartially learn from unbiased data, representing only a truthful group to ensure precision. On the opposite side, AI is capable of expandable and rapid real-time detection. It is also proficient in recognizing the most intricate and singular acts of fraud that no existing system can currently identify. Finally, it can learn about the latest and innovative fraud trends in seconds.

3.1. Overview of AI and Machine Learning Techniques

Fraud detection processes involve many areas such as deep learning, classification, feature selection, handling big data, clustering, similarity analysis, and unsupervised or supervised learning. This section aims to present an in-depth overview of different AI and machine learning techniques used in fraud detection.

Supervised learning techniques, mainly classification, use historical data to predict or flag transactions based on their fraud likelihood. This model learns on labeled data in order to make predictions on new and unseen data. Logistic regression, support vector machines, decision trees, and random forests are some popular techniques adopted. Unsupervised learning, by contrast, leaves the model to explore the data without knowing its internal structure or being trained by any desired outcome. In the case of fraud detection, these models can use a variety of clustering techniques, including k-means or hierarchical clustering methods, to segment similar records into groups. Neural networks, on the other hand, are a family of models built using layers and can

learn patterns from data by adjusting weights between layers. As a result, an optimal neural network architecture can be developed based on the study details and the structure of the dataset. Existing hybrid models combine two or more models, such as unsupervised, supervised, or a combination of both, to see whether they outperform individual models.

A great amount of data may be collected under normal conditions but not necessarily representative samples of the fraudulent cases to be detected. A core requirement for the development and deployment of fraud detection models is the availability of sufficiently large amounts of labeled training data. However, non-fraud cases considerably outnumber the fraudulent ones, which may lead to an imbalanced dataset. Techniques to address the imbalance consist of randomly undersampling the majority class, oversampling the minority class, or using techniques by generating synthetic fraud records. Although larger and higher quality data yield more accurate models, the biggest risk is that big data can contain a significant amount of noise or irrelevant instances. It is important that the data be highly accurate and relevant to its application scenario in order to improve the accuracy of the learning model. As a result, data quality in AI techniques and the algorithms forming AI-enhanced systems is key to assuring the effectiveness of AI models.

4. Real-Time Data Processing in Insurance Fraud Detection

As data has become the world's most valuable resource, and the analysis of that data has become more sophisticated and performed at a massive scale, real-time data is now key to understanding fraud loss trends. Speed is of the essence in insurance fraud. The increasing role of AI in the control of insurance fraud has led to real-time analysis becoming crucial to preventing fraudulent claim payouts. Every second that is saved in passing data and assessing it could prevent fraudulent payments. Insurers must process and analyze high volumes of data, generated continuously by applications to ensure that fraudulent activities are caught when they occur. Ingestion and processing may take hours, rendering them unsuitable for today's real-time world. New fraud detection mechanisms allow immediacy, relying on big, real-time data that utilizes specific big data tools including for storage and processing as parallel distributed processing platforms, and for speed, with low latency concerns for big data processing and storage functionality.

Today, the availability of big data on the cloud has made big data analytics an interesting pursuit. Utilizing cloud computing fits particularly well with big data processing due to the unparalleled computational resources available. However, a significant amount of resources needs to be dedicated to running the most effective real-time big data analytics applications. This makes it necessary to investigate real-time data processing that meets both the computational requirements and maximizes utilization. Delayed alerts or static hotlist feeds do not serve a purpose here. Real-time detection is the way to drive a fraud management program that can reduce fraud exposure and itself act as an embedded fraud prevention program. As there exists ample insurance data and resources, insurance companies today need to be equipped with granular big data and fast algorithms if they really want to embark on fraud detection. Once this is done, the adoption of run-time scorecards would eliminate entire fraud by prompting fraud changes. They enable insurers to identify and respond to anomalies and new patterns of fraud almost instantaneously, and can even issue an alert as they detect the fraud. Such integrated real-time detection and fraud management systems are beginning to be seen in the retail industry.

4.1. Challenges and Solutions

Real-time processing in the insurance industry brings with it numerous and exacting challenges. These may include data privacy concerns, a lack of integration between existing technology platforms and applications, and the sheer scale of the potential data in operation. All of these need to be addressed to ensure that real-time fraud detection methods are adopted correctly and efficiently. Data privacy is a serious constraint on the sharing of data; ensuring that the right data gets to the right place at the right time is fraught with potential dangers concerning misuse or misappropriation. Additionally, information compliance, either through the suitability of local laws or the management of global data protection laws and guidelines, is a genuine concern for organizations worldwide.

The lack of integration between the so-called 'silos' of technology platforms in existence, as well as the many software applications used by insurance companies and their customers, is also a reason to be cautious in adopting any real-time processing methods for data, including fraud detection methods. Although cloud integration and APIs may go some way toward encouraging universal data standardized storage solutions, it is

less likely that this issue can be resolved. Similarly, the data growth issue is reaching a critical mass; already, few companies are able to handle the data volumes they hold, never mind larger streams from real-time processing. The challenges surrounding real-time detection are serious indeed. However, while the challenges may be considerable, the potential—and probable—impact on the industry if these systems are adopted makes them well worth overcoming. Many are already working on solutions: in the case of data privacy, new and secure methods of sharing data are being developed. Similarly, limitations on data governance and oversight responsibilities have resulted in the development of a new, cloud-based data governance model well-suited for the real-time age. As well as investments in new technology, however, it is also necessary to train staff and stakeholders to use these new data processing systems. Exciting and potentially industry-changing, the solution to insurance fraud detection is no exception. A payment method that is protected by the transparency of blockchain technology seems a preventative measure against the misuse of data, providing insurance companies with peace of mind regarding data privacy.

5. Case Studies and Applications

We carried out the following case studies to explore how AI solutions can be used to detect and prevent insurance fraud. The case studies highlight instances where the application of AI technology has led to the identification of fraudulent activities and describe how these pilots or rollouts have changed the operational processes within the organizations. We end the section by drawing lessons from the case studies. The case studies provide evidence of the practical applications of AI technologies and their real-life impact on the operational efficiency and effectiveness of detecting and reducing fraud in insurance. The case studies described in this section cover a variety of types of insurance, purposes of AI deployment, and fraud types. This diversity of applications and operational environments illustrates how a variety of organizations have implemented real-time applications of AI in their claims handling.

Case Study 2: Cat Progressive. The implementation of any technology solution is a complex task that can be met with resistance on many fronts, including technical, political, and operational. The implementation of AI technologies in case studies 1 and 2 was successful, while it fell short of success in case study 3. This paper seeks to evaluate why AI yielded different results across different industries while detecting and

preventing the occurrence of various types of insurance fraud, including motor fraud, general liability, and health care. However, it is equally important to communicate AI deployment successes as it is to communicate when it has been unsuccessful. The insurance and AI industries can only advance through the development of proof of concept studies, pilots, best practices, and the facilitation of knowledge transfer in order for adopters to replicate success. The implementation of an AI solution in response to rising motor insurance fraud is particularly difficult as it involves radical reform of a business operation in a very price-competitive business environment. The success of the implementation of the AI solution in both of these insurance sectors lays out different perspectives within the organization, and these are described along with some of the processes of change management that were incorporated into the implementation phase. Requested outcomes of the evaluation included the determination of the level of success that the respective implementers had using the automated anti-fraud solutions, the lessons learned from each AI deployment, and the organization's recommendations for future AI deployments.

5.1. Successful Implementations in the Insurance Industry

Successful Implementations in the Insurance Industry

The following subsection is structured around successful instances, discussing the implementation, outcomes, and factors contributing to success.

Case Study: Implementing AI to Achieve Best-in-Class Results

A Europe-based insurer, realizing the limitations of its classical rule-based fraud detection system, decided to modernize its entire approach to fraud management using AI. The insurer had learned that rule-based and statistical scoring systems excel at pushing a strategic fraudster toward other insurers but fail to halt the new breed of youthful fraudsters who are gaming the system with non-conventional banking and identity theft fraud. The solution was to implement AI throughout the claims journey in this line of business, using out-of-the-box models specifically designed for first-notice-of-loss or claims setup fraudsters to overcome the initial skepticism that would be expected from decade-old claims decision-makers. They started by running a Proof of Value using real-life claims data from the line of business, which was labeled for fraud. The Proof of Value was successful, yielding positive results with a relatively low false discovery rate.

With the results in hand, the insurer then assembled a cross-functional project team consisting of data scientists, data analysts, business analysts, and IT.

The claims solution was split into four sprints for delivery, spread out over 20 weeks. Week 1 to Week 10 was mainly focused on building the Proof of Value in a clearer and scalable manner. These models were mainly out-of-the-box AI models that didn't require significantly large processing capacity due to its few Home Office hours in production. The second half was spent building the remaining models that were fully extended "real-time AI" models. A streamlined cut-and-paste process methodology was selected via deep-dive sessions with all stakeholders and data scientists. The idea was to run models on in-scope data with the business early, before the solution went to production, as a functioning tuning step in fraud strategy. The cross-functional team had daily calls and would agree on what the new current undetected fraud was, who was doing it, and strategies to catch future fraud. The most arduous task was not the building of the models nor productionizing them, but getting the operation skilled and trained to approach claim making from a pure new mindset across all roles, including the front-line call center staff. The results of the implementation were very successful, and the insurer eliminated tens of millions in claim fraud, reducing the long tail by a third. The next phase is to scale and replicate across other lines of business.

6. Future Direction

While the future of AI in the insurance industry will be greatly determined by the continuing evolution of technology, increased user acceptance, and regulatory norms, we anticipate expansion and convergence of its range in:

● With Interactive Technologies: AI will likely converge with other technologies that are commonly used in significant portions of insurance and should become better aligned with AI efforts. This includes how AI and/or blockchain might connect.

● With Real-Time and Predictive AI: The future of AI could become more predictive. AI will have the potential to not only immediately detect potential fraud in progress but also grievances that may have fraudulent potential in the future.

● Across the Financial Industry: AI does not typically exist only in the insurance industry. For this purpose, the inspection of insurance companies will also involve the examination of their regulatory preparations. AI processes its exchanges.

● With AI as Part of Various Controls: AI can lead to new technologies that evaluate the fraud risk of surveillance systems and other manual activities requiring rating models to change. New fraud tactics will be considered in the industry, and a multi-faceted solution can be investigated to target and share any tactics currently being undone if desired.

● With a New Ethics: As consumer interest in privacy and confidence in computing and innovative processes increases, AI designers are challenged to make the insurance charging procedure as clear as possible. A solution-based design should be implemented to ensure transparency and advocacy while also ensuring that ethical considerations are a concern for the future.

7. Conclusion

The insurance industry has been and still is one of the industries that have experienced a significant incidence of fraud. Fraudsters are becoming increasingly sophisticated and employ a variety of techniques to deceive insurance companies; however, insurance companies are developing increasingly sophisticated mechanisms to tackle them. This text surveyed the breadth of fraud detection techniques for real-time and real-time-like solutions. Speed is a critical parameter of any detection mechanism; thus, information should be processed as soon as it becomes available. These functionalities and techniques were classified as, for example, feature engineering, ensemble techniques, and post-processing based on the underlying algorithm and main characteristics of the methods, which were described in detail before providing an overview of selected methods. The survey results showed that there is still significant scope for improvement in real-time insurance fraud detection. The study concluded with a discussion of emerging trends and possible future challenges in big data and real-time fraud detection in AI-related fields.

Whether it is a part of the claim process or a standalone service, a real-time AI solution is expected to pave the way for better and more robust fraud prevention mechanisms in the insurance industry. The current version can give a hint at the magnitude of the possible benefits from AI solutions in fraud detection and prevention. Additionally, we expect this to lay a foundation for studying the minimum threshold of the level of risk the stakeholders could be willing to accept given resource constraints and strategic technological investment.