Network Behavioural Analysis and Claim Pattern Recognition: AI-Based Computational Strategies for Insurance Fraud Prevention

Dr. Aliaksandr Pinchuk, Professor of Computer Science, Belarusian State University of Informatics and Radioelectronics (BSUIR)

1. Introduction

Take the increasing number of claims made by the use of social media; paying benefits due to non-verifiable claims and the numbers used by organized crime to commit insurance fraud. All the processes described above represent surprise factors related to the underwriting result of an insurance company. Hence the importance that the insurance business gives to fraud prevention and detection strategies. In this context, this paper aims to understand how the process of prevention and detection of fraud is being discussed in literature. What challenges are being faced by actuaries in this process? It lists some of the AI and predictive techniques that can help actuaries in the control of fraud. In addition to understanding the current process through the analysis of the literature, this paper carries out documentary research and interviews with specialized personnel to address a critical analysis of how these people are preparing and what the challenges of actuaries are in detecting and preventing fraud.

2. Understanding Insurance Fraud

Insurance fraud involves any act committed based on a false representation of a fact, with the knowledge of this misrepresentation and the intent to receive a benefit. Insurance companies and consumers are affected by property and casualty insurance fraud, which occurs in several different forms including premium diversion, fee churning, asset diversion, and workers' compensation insurance fraud. U.S. insurance fraud exceeds $40 billion annually and can result in increased premiums and costs for all policyholders. Americans pay $100 billion in higher premiums due to fraud. To address this issue, insurers worldwide are investigating how risk management processes and risk assessment models can be expanded by utilizing AI.

Since insurance fraud behavior rarely provides direct evidence of the objective truth of a statement, directly identifying this behavior is nearly impossible. Thus, insurers often depend on claims handlers to recognize inconsistencies, unusual patterns, and red flags, and then track suspicious claims. For these claims, staff conduct investigations and make decisions. However, they process a large number of claims every day, which makes it difficult to pay extra attention to all suspicious claims. As a result, for some sophisticated fraud strategies such as committing multiple claims over a longer time, missing expensive crime, and inflating claims by lower-ranking staff, perpetrators take advantage of these human limitations to commit fraud. These human limitations have also induced the rise of commercial insurance fraud as well, as insurers are still using manual processes and technology that are unable to deliver effective pre-approval fraud detection. Furthermore, paper documentation can be kept in the absence of digital technologies, and consequences for fraud may not be serious, allowing companies to escape serious financial repercussions that help fraudsters evade detention and jail.

2.1. Types of Insurance Fraud

An insurance policy is a social way of sharing the risks. There are a lot of different types of insurance, such as life insurance, health insurance, private and commercial insurance, in-home insurance, etc. There are also a lot of types of insurance fraud. In general, type (A) is considered fraud in accordance with the policy conditions and the law in effect, and it involves false or misleading acts that are made knowingly. It has a number of aliases: 'hard fraud', 'material fraud', 'provable fraud', 'intentional fraud', 'criminal fraud', or fraud by deceit, etc. In general, type (B) – overstated claims – is fraud initiated by over-claiming. Usually, prior to the damage or loss occurring, the insured calculates the expected amount of payment and in response makes a claim. From the insurers' point of view, type (A) fraud is more interesting and much more challenging. This kind of fraud includes a number of different varieties and aliases, such as completely fraudulent policies and claims, claimant fraud, staged or fake accidents, multiple fake claims, and under-insuring, etc. Actually, the definition of fraudulent schemes in insurance is limited by the imagination of fraudsters. However, all these categories are based mainly on illegal financial gain due to the different types of insurance coverage. The task of the insurance company is to differentiate between them: to neglect the simple unintentional claims, to avoid conflicts between its clients, and, most of all, to protect itself and the other policyholders from those who commit hard fraud. In the

battle against insurance fraud, the leading role is played by the insurance companies. The administrative and legal framework also provides them with the flexibility of their actions.

2.2. Impact of Fraud on Insurance Industry

The insurance sector is experiencing a drastic increase in fraud, and it is projected to cost non-life insurance between 3% and 5% of its claims. Health insurance fraud ranges from 3% to 10% of all health-related costs. Finally, the life insurance sector is estimated to lose $4 billion per annum because of fraudulent behavior. The scam appears to be growing with technological advancements. The ability to use online systems to supply incorrect, delayed, or non-provided services is the cause of this. These misinformed actions gather resource savings that participate in unethical behavior, while at the same time using the instruments that would help them sound like law-abiding citizens. This can be derived from the reliance of the fraudster and the societal context in which they live. The insurance industry is economically punishing fraudulent behavior. Normally, cases of fraud are maintained in secret. However, mass media and the public belief that the insurance industry is overcharging them are very high in paying for fraudulent behavior. The association represents an industry that benefits from fraud by developing products to minimize the risk of fraud and by establishing the expertise and jurisdiction required to examine allegations of possible fraud. In addition, the credibility of an insurance market reveals that insurers proactively attempt to minimize the financial weight of fraudulent practices.

3. Role of Technology in Fraud Detection

Artificial intelligence methods strive to utilize the computer age by producing intelligent behavior and implementing an intelligent system that follows an appropriate course of action in an intricate environment. In comparison to standard methods, which depend mostly on rigorous rules and experience-based knowledge, numerous new approaches based on machine intelligence are being developed and shown to produce autonomous learning, adaptation, and rational behavior. Examples incorporate various knowledge representation schemes, heuristic search for problem-solving, intelligent agents, and solid-state science, such as robotics and parallel processing technology, as networks. In such fields of technology, artificial intelligence methods are becoming essential to producing more secure, robust, and extensible system attributes.

In this area, recent AI developments have given rise to uncommon fraud detection solutions. Hackers have made an entire business around committing fraud, utilizing real identities, stealing, and counterfeiting documents. While digitizing a process makes it less susceptible to fraud, the process must be secured against electronic identification misappropriation. Additionally, artificial intelligence can be applied to best avoid several types of identity theft attacks, while also ensuring the privacy of the user and the data. With the age of data theft, new innovations in fraud detection throughout sectors are being introduced every day. Insurance industries are assured that these fraud prevention techniques will safeguard privacy and secure processes while swiftly tracking down fraudulent individuals within borrowed identities.

3.1. Traditional Methods vs. AI-Based Approaches

Insurance fraud can significantly affect various insurance operational tiers, cause substantial financial losses, and potentially expose insurance companies to insolvency, solvency, and market instability threats. Consequently, the development of effective countermeasures, such as advanced fraud detection and prevention techniques, constitutes a priority. Understanding the factors and motives of insurance fraud incidents, the characteristics of fraudulent schemes, and fraudster trends have led to the identification of various fraud prevention strategies and the utilization of IT platforms that support the specific activities. Traditional approaches use rule-based and statistical data analysis methods, such as predictive modeling and organizational rule engines, to examine insurance contracts and claims aiming to identify unusual patterns, from atypical variables to subjects engaging in abnormal transactions or behaviors. Unusual and exceptional materializations of contract-related and claim-specific parameters potentially signal fraud. An integral part of the traditional approaches is the process of negotiating and securing information exchanges with a variety of agent networks, public sources, and internal and external databases. Recently, machine learning models have been developed to identify unusualness indicated by traditional rule-based and statistical outlier-targeted variables and have reinforced the related prevention strategy. On the other hand, AI-based approaches, utilizing deep learning and big data analytics, treat the variables and information included in insurance contracts and the claims process as data that can be exploited for correlation analysis and vastly improved statistical detection precision. They can support the measurement of relationships among the detected risk items, provide prediction methods in order to estimate the loss

due to organized fraud, and extract high-dimensional data feature space characteristics that effectively inform risk management models. In turn, this can significantly enhance the fraud management functionality of insurance companies, since they will make superior decisions and actions implementing strategies adjusted to the actual attribute space restrictions and conditions encountered.

4. Machine Learning in Anomaly Detection

4.1 Introduction In-memory computing capabilities of modern business software enable the direct application of machine learning techniques on transactional data, which is the main source of fraud pattern formulation. Unlike traditional rule-based analytics, unsupervised machine learning approaches do not depend on past empirical rules and experience but are able to model historical data distributions accurately. An industry-adequate fraud measurement system can analyze and interpret vast amounts of firm-specific transactional data for regularly occurring pockets, large cumulative, and sophisticated fraudulent activities. Machine learning tailored for small differences in fraud identification, other than balanced and juror-assessor-guilty questions, are key components. However, group deterministic and non-deterministic and non-temporal transactional sequences can be found in explainability feature importance from many machine learning techniques to gain important interpretability from such analyses. We noticed that fairly standard logistic regression results in somewhat lower feature importance for positive events, including fraud. Because all transactional data is extensively cleaned, rules designated with high positive b-metric scores or values need to be screened for errors, omissions, and relevant fraud characteristics. The fraudulent adjuster claim case subset and the data imbalance reduced minority sample sizes available for the non-fraudulent caliper considered as input in the accepted sampling instance.

4.1. Supervised vs. Unsupervised Learning

In order to prevent fraud with artificial intelligence (AI)-based methods, fraud scenarios and indicators of fraudulent activity are used by machine learning methods to train a model that utilizes these scenarios or indicators to make decisions. The basis of this approach is that similar problems within the dataset are solved in a similar way. These scenarios are often derived from known fraud cases and databases. Therefore, good quality data is key to establishing good quality scenarios. Different machine learning

paradigms exist, and two broad categories should be distinguished: supervised learning and unsupervised learning. When the products sold by an insurer are such that few frauds occur, the insurer might have a problem using supervised learning as there are very few or no labeled fraud cases. This is called a low incidence model. In such cases, unsupervised learning models may be used. We will now explain the difference between these models in more detail.

Before training a model, the available data is split into a training set and a validation set. In the training set, the features or data that describe the problems need to be labeled: Is it gained in a fraudulent way or not? Fitting a model to separate fraudulent from non-fraudulent transactions is not directly the goal, but the model should fit fraud scenarios. We need to strike a balance in the trade-off between finding enough fraud but not too much against making sure that we do not waste resources in detecting as much fraud as possible. This applies to unsupervised models as well, but is less straightforward there. When a model is trained, its quality is tested out-of-sample in the validation set. Accuracy, positive predictive value, sensitivity, and specificity, among others, are examples of metrics to calculate this quality. Identifying fraud is important, but the ability to identify fraud without generating false alarms is also important. If a score is estimated, efficient thresholds can be chosen for company objectives. When satisfactory outcomes are yet to be achieved, additional labels can be added to the training data, and the model training and validation can be re-run.

4.2. Feature Engineering and Selection

Feature Engineering and Selection: Feature selection acts as a critical preprocessing step prior to developing fraud detection models. There are several practices within this step that can heighten the fitness and accuracy of the models. Automating the analysis will significantly cut down time and resource costs. Appropriate feature selection is very important due to the fact that the performance of AI algorithms is highly dependent on the quality of input features. Researchers usually spend a lot of time on feature extraction, preprocessing, and selection, working up to the algorithm configuration. However, good problem data, preprocessing, and smart feature data extraction call for effective application of data mining.

Studies often utilize conventional feature selection methods such as information gain, chi-squared tests, gain ratio, or feature selection algorithms such as genetic algorithms,

particle swarm optimization, and ACO. However, the process comes with its specific challenges. Some of the challenges include the large number of features compared to small and not sufficiently balanced data, the diverse nature of features, and missing data. For large feature sizes, the conventional feature selection techniques have significantly increased the time cost. These methods are also typically coupled with a heavy time requirement to compute the best pattern recognition model.

5. Challenges and Limitations of AI in Insurance Fraud Detection

Limitations must be acknowledged in the face of the claims made about the capabilities of AI-based solutions to detect insurance fraud. In fraud detection, too much focus on the use of AI technologies without acknowledging the restrictions and limitations of these technologies relative to human knowledge and creativity to commit insurance fraud itself is not encouraging. To recognize AI's limitations in fighting insurance fraud, insurance companies with digital champions in place are more likely to succeed with a holistic AI application strategy. These 'experienced adopters' of AI are harnessing the technology to improve the customer experience and to strengthen their business. They are also starting the journey from perimeter to embedding AI fraud intelligence capabilities across the enterprise.

There are several fallacies and misperceptions about the robustness, accuracy, and effective use of AI in these disciplines that need to be considered when deploying AI in anti-fraud settings. Foremost, the word 'intelligence' is misleading when used with artificial as in 'artificial intelligence', since it never equated the two in pioneering work on AI research. Intelligence can be said to depend on knowledge, experience, and creativity, while machines have other capabilities that are unmatched by humans, such as memory, speed, and accuracy. Intelligence, knowledge, and creative thought are much more ambiguous and difficult to formalize than an AI approach. Intelligence should be a defining feature of economic crime prevention, personally. It is my belief that it requires creative thought to devise a solution to challenging economic crime problems, not just applied technology. Conventional AI-based fraud detection has the appearance of intelligence or smartness, but it is neither intelligent nor clever, particularly when viewed in the burgeoning context of digitization, machine learning, and reinforcement within the insurance industry to gain a more competitive advantage against a large number of entrenched financial technologies and fraud solutions.

5.1. Data Quality and Bias

In implementing AI systems and strategies that will be used for insurance fraud prevention and detection, one of the most important aspects that organizations need to consider is the integrity and bias of the data that will be used to train, test, and evaluate their AI system. AIs are particularly hungry for large quantities of data, but the importance of ensuring that data is relevant, not damaging, biased, competitive, or stolen, while ensuring the privacy and security of AI models must not be forgotten. Data privacy and security must also be considered to ensure the privacy of policyholders' claims and other personal data if services such as snapshot-based insurance products, which base premiums on policyholder behavior, are used. Before deploying an AI-based solution, insurers should assess the quality of the data. If data comparability or personal malfunctions are already established, the models will perpetuate these errors or beliefs over time when comparing similar models to other models that claim to improve AI and machine learning business models.

5.2. Interpretability and Explainability

While NLP has performed wonders in terms of advanced analytics such as text classification, understanding and automating this process can help insurance firms provide explanations to their queries and discussions. In very simple terms, the interpretability of a decision could help analysts evaluate machine learning models more effectively, improve data quality, or identify better features for building classifiers. Explainability, on the other hand, is more about making the ML systems function like a 'black box' – that is, understandable and clear to anticipate decisions, actions, and to take on better reasoning abilities. It is also about understanding the processes related to transparency, accountability, and social ethics. This also helps promote transparency and ethics when being used in real-life applications – for example, in insurance risk assessment tasks.

The ability to understand reports from NLP and NLU-based models is critical to a range of professions, including insurance practitioners and regulators. Not only parametric and logistic/linear regression models, but also deep learning models are seen as a 'knot of fathomless mysteries' wherein changes to the peculiar regularities of making decisions may not be accurately explained. Freely swapped from data, such models can be very misleading and harmful. It is a necessity for business applications to make these

highly complex models interpretable. It is always important to unbox the learning process accompanying the decision boundary. As such, it is essential to define outlines for the parameters of the model in a way that quantitative and qualitative explanations about classifications could be given as per the applications of the models.

6. Case Studies and Best Practices

In this chapter, we discuss various case studies and best practices around AI fraud detection. These case studies represent only a small fraction of the work of fraud detection at various insurance companies. Still, they provide impactful insights on implementing AI-based solutions to detect insurance fraud and advanced fraud prevention strategies being put in place by these insurance companies. Through these examples, we illustrate how insurance companies can detect and combat insurance fraud at different stages of the customer journey with holistic, AI-based fraud detection solutions. These insurance companies are leveraging AI to unlock greater efficiencies and improve loss ratios without compromising the customer experience. Additionally, we outline an advanced fraud strategy that targets bad claims before they are filed, aiming to shift the focus in fraud detection from detecting bad behavior to preventing it. In doing so, we completed sixteen case studies and provided quantitative insights into the value provided by AI to detect insurance fraud at various stages of the customer journey.

6.1. Real-world Implementations

Current use cases or real-world implementations of AI presented in this work are mainly based on research papers. These AI-based insurance fraud prevention strategies mostly need integration with insurers' existing fraud management systems in order to operate in the real world. There are several startups and a very limited number of established companies that offer anti-fraud solutions. Most of these companies provide solutions using machine learning and mainly deal with form-based insurance industries. Therefore, it is beneficial to have a review of industrial events and organizations that are investing in fraud prevention to provide some practical insight into the problem. Aiming to discuss the practical insights into insurance fraud prevention, we provide an inaugural discussion of commercially available anti-fraud solutions in the global insurance industry. Currently, many of these commercially available solutions still refer to standard pattern recognition techniques as AI. Particularly, they are product

applications in specific areas like health insurance or workers' compensation based on conventional machine learning algorithms that are exploited mainly for decision-making support in fraud identification. Therefore, in addition to the existing research and future research directions, we also reveal practical anti-fraud applications available or some work in progress that efforts to develop new practical applications.

6.2. Key Success Factors

First of all, the new requirements for solutions ensuring efficient preventive control are put forward in the insurance field. Secondly, such solutions can be implemented with adequate use of the existing AI technologies that have been recently developed. In particular, the following key success factors for building and functioning of the AI solution for efficient fraud prevention can be observed: The IVC system can be assembled and launched in an operational mode only in very close collaboration of actuaries, fraud prevention units, and their IT staff. One of the most important parts of such cooperation is to develop a technical specification that satisfies not only actuaries but also IT staff. The practical implementation and operational phases are iterative and require constant communication and feedback from other departments involved to improve a single iteration.

The efficiency of the IVC system depends on the prevalence of the process of adjusting insurance rates. It means that it is better not to have an eternal underestimation of insurance rate accuracy without any chances of changing the accrual principles. If there is a prevalence of the 'evolutionary' mode of rate development, it is highly probable that the company is half-ready for the introduction of the IVC system. Moreover, it is the obligation of the IVC system developers to estimate the possible radical actions taken by rival companies.

6. Future Direction

In this chapter, we discussed in detail the motivations behind fraud prevention for big data, IoT, and AI-related applications. In addition to discussing the importance of fraud prevention earlier in the part, we also explained why the current set of solutions offered by the current generation of IoT platforms and recommendation engines are inadequate. We then evaluated the main challenges involved in insurance fraud prevention before offering a security-first vision. We underlined the need to deploy a tailored security protocol to minimize the trust required by a given AI. This is especially relevant for the

insurance economy where every stakeholder can have skin in the game. It is also relevant in the context of existing regulation because the smaller changes in the current approach could encourage a class war based on how many put their skin in the game. In the context of this protocol, we offered a preliminary standard in the form of a roadmap to transparency for fairness and fraud prevention in the explainable AI era. The first step of the protocol is a report.

A classical obstacle encountered in many related topics is the AI explained. If the insurance company's algorithm has not prevented fraud, then denial requires explaining and refusal requires further exploration. While giving material to a judge may help this conversation, it is much more useful to embed common sense information about risks directly into the algorithm used to make such decisions. Note that embedding information is a subtle design choice. In the case of machine learning reasoning, such information acts like a regularizer. Less obviously, feature engineering is in practice the process that takes more time, care, and effort in any machine learning project. It is therefore natural to treat skill at feature engineering as something that is almost as much a part of real-life intelligence as efficient optimization. Our second step to our reference client is an audit of the associated algorithm. Future directions include generalized algorithms and audits, recurring towards real-time and implications for AI-based insurance pricing.

7. Conclusion

The protection of insurance companies from loss is a pressing issue for insurance companies. They use information technology intensively to address the issue. Advanced technologies such as AI are implemented because they are significantly helpful. AI implementation is important not only from a particularly competitive service, but also from the perspective of loss management. The purpose of this research was to develop AI implementation strategies of insurance companies in dealing with fraud and to evaluate the effectiveness of the strategies. This paper suggested AI application strategies through in-depth interviews with relevant experts and conducted a survey of the AI application strategies as opposed to existing insurance companies. The findings are expected to support the effective practical applications of insurance companies, and it is certain that the findings should bring technological and literature contributions.

In this study, comprehensive AI-driven insurance company business models, that apply AI to business processes such as insurance, customers, channels, products or services were developed through five key applications. Current technological influences of AI on insurance-related processes, products, and services were considered. AI application strategies that have been identified for insurance businesses are expected to significantly contribute to enhanced relations with customers by building applied modules focused on customer needs. AI which is accurately adapted to insurance business models comprises business processes, customer relations, and business structures. Moreover, AI provides accurate customer evaluations, customer services, markets and product development competencies, and business frameworks specific to the insurance industry. Ideas for structuring insurance businesses in light of these facts were discussed. Insurance company leaders should note the influence of these ideas on business initiatives; efficient and sophisticated AI can enhance the functions of insurance companies, while also enabling customer value creation. To achieve these aims, insurance companies need to identify and specialize in AI modules that they could easily apply, and to adjust their business models to reflect the insights gained from this process. Furthermore, they should refine the practical application of AI modules in support of the companies' objectives and strategies.