Machine Learning in Insurance: Enhancing Pricing, Claim Detection, and Index Insurance Innovation

Yuejiuhui Gao
School of Information Technology and Management, University of International Business and Economics, Hui Xin Dong Street, Chaoyang District, Beijing, China

Keywords: Machine Learning, Insurance Pricing, Claim Detection, Index Insurance, Data Science.

Abstract: The emergence of machine learning technology has brought revolutionary development to the insurance industry, providing new methods to solve traditional challenges. This article provides an in-depth analysis of how machine learning algorithms can be used to enhance insurance pricing, improve claim detection systems, and innovate in the field of index insurance. Machine learning models improve the accuracy of insurance pricing by processing large amounts of data and discovering complex patterns, predicting payout trends, and optimizing strategies. In addition, these models simplify the detection of fraudulent claims, thereby protecting the interests of insurance companies and policyholders. The combination of machine learning and index insurance has also been explored, emphasizing its potential in improving risk assessment models and developing personalized insurance products that adapt to dynamic market conditions. Through systematic literature review and qualitative analysis, this study emphasizes the transformative impact of machine learning on the insurance industry and outlines future research directions to fully utilize its potential and address related challenges.

1 INTRODUCTION

With the significant improvement of computing power and the rapid development of data science, machine learning technology has become a key driving force for innovation in the insurance industry. Machine learning algorithms can process and analyze massive amounts of data, revealing complex patterns and correlations that traditional analysis methods cannot capture, providing unprecedented insights for insurance companies.

Innovation in insurance pricing: In terms of insurance pricing, the application of machine learning technology not only improves the accuracy of pricing, but also enables insurance companies to provide personalized insurance products based on customers' specific risk situations. This personalized pricing strategy helps insurance companies better manage risks while providing customers with more attractive insurance solutions. Through in-depth analysis of historical claim data, machine learning models can predict compensation trends, optimize pricing strategies, and reduce information asymmetry issues. In addition, machine learning techniques have demonstrated significant advantages in handling data loss and structural bias, improving the robustness and transparency of models.

Intelligence of claim detection: In the field of claim detection, the application of machine learning technology has evolved from simple rule engines to complex predictive models that can monitor and analyze claim behavior in real time and detect abnormal patterns in a timely manner. This intelligent claim detection system not only improves the efficiency of fraud detection, but also helps protect the interests of honest customers and maintain fairness in the insurance market. Machine learning algorithms, especially deep learning structures, have shown great potential in identifying fraudulent claims. By analyzing patterns and anomalies in historical claim data, machine learning models can learn the characteristics of fraudulent behavior, thereby improving detection speed and accuracy.

Innovation of index insurance: As an emerging form of insurance, the combination of index insurance and machine learning has brought new growth points to the insurance industry. Machine learning technology plays an important role in processing climate data, optimizing risk assessment models, and developing personalized insurance

Gao and Y.
Machine Learning in Insurance: Enhancing Pricing, Claim Detection, and Index Insurance Innovation.
DOI: 10.5220/0013530000004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 585-593
ISBN: 978-989-758-754-2

products. This combination not only enhances the adaptability and flexibility of insurance products, but also provides new business opportunities for insurance companies. Machine learning models help insurance companies more accurately predict and quantify risks by analyzing historical claim data and market data, thereby optimizing insurance pricing strategies. The diverse uses of data mining techniques within the insurance sector, encompassing activities like evaluating risks, identifying fraudulent activities, and examining underwriting processes, all play crucial roles in the evolution of index-based insurance products.

Research Methods and Structure: This article systematically reviews relevant literature and uses qualitative analysis methods to comprehensively analyze the application of machine learning technology in insurance pricing, claim detection, and index insurance. The subsequent chapters of the paper will delve into these topics in depth, with the first chapter discussing the application of machine learning in insurance pricing; Chapter 2 will analyze the current situation and challenges of insurance claim detection; Chapter 3 will explore the combination of index insurance and machine learning. Each chapter will provide detailed case studies to demonstrate the effectiveness and potential of machine learning techniques in practical applications.

2 THE APPLICATION OF MACHINE LEARNING IN INSURANCE PRICING

In the insurance industry, insurance pricing is the core link, which is directly related to the financial health and market competitiveness of insurance companies. The key to insurance pricing lies in accurately predicting future risks and payout probabilities. With the development of big data and machine learning technologies, insurance companies can now use these advanced tools to analyze historical claim data, thereby more accurately predicting payout trends and optimizing pricing strategies. This change not only improves the accuracy of pricing, but also provides an opportunity for insurance companies to maintain their advantage in fierce market competition.

Insurance is a financial mechanism aimed at managing economic uncertainty by diversifying risks (Kaffash, Azizi, Huang, & Zhu, 2019). Insurance companies traditionally rely on generalized linear models to handle data modeling of claim frequency and severity (Spekkers, Kok, Clemens, & Ten Veldhuis, 2014). However, with the successful application of machine learning in multiple fields, these techniques have begun to be widely adopted by insurance companies for improving risk assessment and pricing models. Machine learning models can learn from historical claim data and predict payout trends, providing insurance companies with a new tool to improve pricing accuracy and efficiency. Utilizing machine learning techniques allows insurance firms to forecast the likelihood of incurring losses with greater precision, which in turn mitigates the issue of information imbalance within the sector (Eling, Nuessle, & Staubli, 2022).

2.1 Flexible data processing methods

Traditional claim prediction methods typically involve fitting the frequency and severity of claims to a known probability distribution function and using it to predict future claims (Poufinas, Gogas, Papadimitriou, & Zaganidis, 2023). But this method relies on accurate fitting of claim data and may not capture all complex patterns in the data. Machine learning provides a more flexible approach that can handle more complex data patterns, including non-linear relationships and high-dimensional data. Machine learning models, including supervised learning, unsupervised learning, and reinforcement learning (Paruchuri, 2020), are being used to develop more advanced claim prediction systems. These models can automatically identify key patterns in data and be used to predict claim probabilities and amounts, thereby helping insurance companies optimize their funding preparation and pricing strategies. Spekkers et al. (2014) used regional aggregated claims data provided by Dutch private property insurance companies, including damages related to rainwater, such as damage caused by rainwater infiltration through roofs and ground flooding entering buildings. The study used decision tree analysis to investigate the influencing factors of rainfall related damage, which can handle nonlinear relationships, high-order interactions, and missing data. By constructing a decision tree model, researchers can identify the factors most relevant to claim frequency and average claim size. Research has found that the frequency of claims is most correlated with the maximum hourly rainfall intensity, followed by real estate value, ground floor area, household income, season (property data only), building age (property data only), ownership structure (content data only), and proportion of low rise buildings (content data only). Nonetheless, the construction of

a tree-based model that meets statistical standards for predicting the typical claim amount was unsuccessful, suggesting that the fluctuations in the typical claim amount could be associated with factors that are intractable to pinpoint on a regional level. Provided more imaginative space for data collection work (Spekkers et al., 2014).

2.2 Addressing Data Incompleteness

Missing data is a common issue when dealing with historical data of insurance applicants (Rusdah & Murfi, 2020). Traditional statistical methods typically require estimation or interpolation of missing values, but machine learning algorithms such as Extreme Gradient Boosting (XGBoost) can directly handle missing values without the need for interpolation preprocessing beforehand. XGBoost can accurately handle missing values through its sparse perception segmentation algorithm. Investigations reveal that the XGBoost algorithm, even without the application of interpolation for data preparation, achieves a similar level of precision to that of the model with interpolation. This finding substantiates the proficiency of XGBoost in managing data sets that have missing entries. This ability makes machine learning models more robust and capable of extracting valuable information from incomplete data. In addition, the application of machine learning technology in processing climate data is becoming increasingly widespread. Machine learning algorithms are used to process climate data such as precipitation, temperature, and soil moisture (Eltazarov, Bobojonov, Kuhn, & Glauben, 2023). Through spatial downscaling, the random forest algorithm can refine coarse resolution climate data to fine data with a resolution of about 5 kilometers, which is crucial for assessing risks related to weather, natural disasters, and other factors. The application of this technology can help insurance companies assess risks more accurately, especially when developing index insurance products.

2.3 Reduce structural deviation

The unstructured nature of machine learning algorithms helps reduce structural biases in traditional models (A et al., 2020). These biases typically stem from simplified assumptions and data requirements of the model. Machine learning models can adapt more flexibly to the true distribution of data, thereby providing more accurate risk assessments. This advantage is particularly important in insurance pricing, as it can help insurance companies quantify risk more accurately and set more reasonable premiums. In addition, when evaluating the performance of machine learning models, it is necessary to consider using multiple performance indicators to increase the transparency of the modeling process, such as (Hanafy & Ming, 2021), Accuracy, Error rate, Kappa, AUC (Area Under the Curve), Sensitivity, Specificity, Precision, Recall, F1. The application of machine learning technology in insurance pricing provides a new perspective, which optimizes insurance pricing by processing large amounts of historical data, predicting payout trends, and reducing information asymmetry. With the advancement of technology, insurance companies can expect more accurate and flexible pricing strategies to adapt to the constantly changing market environment. The development of these technologies not only improves the pricing accuracy of insurance products, but also provides insurance companies with stronger market adaptability and competitiveness. Emerging technologies such as federated learning have also shown potential in personalized insurance pricing, providing more accurate and personalized pricing strategies while ensuring data privacy. With the continuous exploration and application of these advanced technologies in the insurance industry, we have reason to believe that future insurance services will be more intelligent, efficient, and user-friendly.

3 MACHINE LEARNING IN CLAIM MANAGEMENT

Claim management is a core part of insurance company operations, which directly affects the financial stability and customer satisfaction of the insurance company. In the insurance industry, timely and accurate handling of claims is crucial for maintaining company reputation and customer trust. As machine learning technology progresses, insurance firms are equipped with sophisticated tools to enhance both the speed and precision of claims management. This section will delve into the ways machine learning assists insurers in making significant strides in the detection of fraudulent claims, the forecasting of claim amounts, and the optimization of the claims process. Models powered by machine learning, particularly ensemble algorithms and deep learning architectures, have shown remarkable efficacy in forecasting claims and uncovering fraud. These models are adept at handling extensive data sets and discerning pivotal elements that influence the likelihood of claims (A & B, 2021).

M. K. Severino and Y. Peng provided a macro profile of fraudsters based on real data in the property insurance field: approximately 60.14% of fraudsters are male; 48.16% of fraud cases are "premature claims", which refer to claims made shortly after the start of the insurance contract; 52.81% of fraudsters are unmarried; 79.95% of the insurance amount is used for electrical damage or theft claims; The average age of fraudsters is 41 years old; 72.61% of fraud cases involve new insurance policies rather than renewed ones; In fraud cases that have been detected but not yet confirmed in court, the average claim amount for fire/lightning/explosion insurance is the highest.

3.1 Detecting Fraudulent Claims with ML

Claim fraud detection is an important issue in the insurance industry, as it not only affects the financial health of insurance companies, but may also erode the interests of honest customers. Machine learning algorithms, especially deep learning structures, have shown great potential in identifying fraudulent claims (Spekkers et al., 2014). These algorithms can learn the characteristics of fraudulent behavior by analyzing patterns and anomalies in historical claim data. For example, algorithms such as logistic regression, XGBoost, C50, and random forest have been proven to be highly effective in predicting the probability of claims occurring in car insurance (Hanafy & Ming, 2021). In addition, big data technology can efficiently process and analyze massive medical insurance claim data, thereby identifying abnormal patterns and potential fraudulent behavior (Subrahmanya et al., 2022). Through in-depth analysis of claim data, it can be discovered that there are unconventional claim patterns, such as abnormally high claims, frequent claims by individuals or medical institutions, and diagnostic and treatment patterns that do not align with common medical practices. In fact, many insurance companies have increasingly adopted patterns and anomaly detection to automatically detect fraudulent claims (Eling, Nuessle, & Staubli, 2022), which not only improves detection speed but also enhances accuracy.

3.2 Predicting Claims Amounts Using ML Models

Accurate prediction of claim amount is crucial for insurance companies' fund management and customer satisfaction. Machine learning models can predict claim amounts by analyzing historical claim data, thereby helping insurance companies better prepare funds. Machine learning models can also dynamically adjust prediction strategies based on the complexity and variability of claim data to adapt to the constantly changing claims environment. Many health and car insurance companies have applied predictive analytics to data from connected devices and developed new innovative, personalized usage based insurance products (Eling, Nuessle, & Staubli, 2022).

In addition, machine learning methods have also provided new avenues for claim prediction (Poufinas, Gogas, Papadimitriou, & Zaganidis, 2023), which can help improve traditional claim processing processes. For example, using machine learning methods, particularly random forest regression and classification algorithms, to predict flood insurance claims in New York State. Combining flood insurance claim records from the National Flood Insurance Program (NFIP) with hydrological and demographic data to improve the accuracy of flood exposure maps, and taking into account socio-economic factors such as the proportion of minority residents, property values and ages, and political differences in electoral districts. The research results indicate that combining socio-economic data can improve flood exposure estimation, especially at the regional level of population census. These data significantly enhance the predictive ability of the model and provide a new perspective for understanding flood risk (A, J. K. et al.,2020) .

3.3 Automating and Optimizing Claims Processing

Streamlining and refining the claims process not only boosts operational efficiency but also contributes to an elevated level of customer satisfaction. The application of machine learning technology in this area includes automated claim classification, identification of abnormal claims and fraudulent behavior, as well as reducing human intervention. In specific operations, there are many innovations in feature selection methods. The Chi-squared test serves as a statistical technique for assessing the relationship between various features and the outcome variable, prioritizing features according to their pattern of occurrence. Recursive Feature Elimination (RFE) operates as an encompassing strategy that determines the most relevant subsets of features by iteratively eliminating them, often paired with classification algorithms like logistic regression. Meanwhile, tree-based feature selection functions as an intrinsic approach that leverages the inherent

capability of tree-based models, such as Extra Trees, to evaluate the significance of features. These feature selection techniques help reduce the dimensionality of the dataset, improve the predictive accuracy of the model, and reduce computation time. Using appropriate feature selection methods can not only streamline the feature space, but also improve algorithm performance. Research (Spekkers et al., 2014) has found that the random forest algorithm performs best after feature selection, especially when using tree based feature selection methods. In addition, certain specific customer characteristics (such as age, BMI, number of steps, number of children, smoking status, and region) are significantly associated with health insurance and travel insurance claim behavior, and these findings can help insurance companies better understand and predict claim patterns.

Another study on car insurance fraud detection used the Boruta algorithm for feature selection (Aslam, Hunjra, Ftiti, Louhichi, & Shams, 2022). Researchers were able to identify the most influential features for fraud detection - "faults," "basic policies," and "policyholder age," and applied three prediction models (logistic regression, support vector machine, and naive Bayes) to develop fraud detection mechanisms. For each feature, the Boruta algorithm calculates its importance score in the random forest and compares it with a random feature (i.e. a randomly selected unimportant feature in the feature set). If the importance score of a feature is significantly higher than that of a random feature, then that feature is considered important and retained in the model. If the score is lower than the random feature, then this feature is considered unimportant and is excluded in subsequent iterations. When all features have been evaluated and the importance score has stabilized, the iterative process terminates. Ultimately, the Boruta algorithm will output a list of features that are considered to be most influential in predicting the target variable (in this study, fraud detection). Boruta algorithm identified the 9 most important features for detecting car insurance fraud through the above steps, including "vehicle age", "vehicle category", "number of days of accident", "policyholder age", "gender", "marital status", "accident liability party", "insurance type", and "basic insurance policy". These features are then used to construct machine learning models to improve the accuracy of fraud detection. These features can also serve as a reference for subsequent car insurance fraud detection projects.

Incorporating machine learning technology into the analysis of insurance claims not only enhances the speed and precision of the claims handling process but also aids insurance firms in more effectively managing risks and allocating resources optimally. With the continuous advancement of technology, future insurance companies can expect more intelligent claim processing systems to adapt to the constantly changing market environment. Accurate cost estimation can help health insurance companies and an increasing number of healthcare delivery organizations plan for the future and prioritize the allocation of limited nursing management resources (Ul Hassan et al., 2021). By utilizing machine learning technology, insurance companies can provide more personalized and efficient services while reducing operating costs and improving customer satisfaction.

4 INDEX INSURANCE INNOVATIONS WITH MACHINE LEARNING

Machine learning technology provides strong support for the design, pricing, and risk assessment of index insurance products. By utilizing machine learning models, insurance companies can more accurately predict and quantify risks, thereby providing customers with more reasonable insurance products. As technology continues to evolve, forthcoming index insurance offerings are poised to become increasingly smart and tailored, assisting both insurers and policyholders in more effectively managing the financial repercussions of severe weather occurrences.

4.1 ML-Driven Innovations in Index Insurance

Index insurance, as a risk management tool, focuses on providing compensation by linking it to indices such as weather and natural disasters, thereby avoiding adverse selection and moral hazard issues in traditional insurance (Nguyen, Mushtaq, Kath, Nguyen-Huy, & Reymondin, 2024). This form of insurance determines compensation through objective weather indices, which are highly correlated with actual losses and provide more timely and cost-effective compensation (Zhang et al, 2022). Machine learning methods can be used to fuse information from different sensors and data sources to improve the accuracy and robustness of the exponential insurance model. The application of machine learning technology, especially in video and image analysis,

provides an early warning system for the development of index insurance products, which is crucial for managing and predicting weather related risks (Eling, Nuessle, & Staubli, 2022).

4.2 ML in Risk Assessment and Product Development

Machine learning models assist insurance firms in forecasting and quantifying risks with greater precision by examining past claims and market information, which in turn allows for the refinement of insurance pricing strategies. The diverse uses of data mining within the insurance sector, including evaluating risks, identifying fraud, and analyzing underwriting, are all crucial elements in the development of index insurance products (Spekkers et al., 2014). Machine learning technology provides strong support for the design, pricing, and risk assessment of index insurance products. We used historical weather data and crop yield data from Illinois, USA, including 72 weather indices such as precipitation, temperature, dew point temperature, maximum and minimum temperatures, and water vapor pressure deficit. Innovatively applying neural network models to design weather index insurance contracts, learning from high-dimensional and nonlinear weather data through machine learning methods to improve the risk management capabilities of insurance products. Research has found that compared to insurance products based on traditional linear models, neural network models can significantly improve farmers' utility and determine equivalent wealth (CEW), reduce underlying risk, and demonstrate better performance in test samples (Chen, Lu, Zhang, & Zhu, 2024).

By utilizing machine learning models, insurance companies can more accurately predict and quantify risks, thereby providing customers with more reasonable insurance products. With the continuous advancement of technology, future index insurance products will become more intelligent and personalized, helping insurance companies and policyholders better cope with the financial consequences of extreme weather events. Eltazarov et al. (2023) utilized optical bands and indices recorded in NOAA AVHRR climate data, as well as SRTM digital elevation model data, as input features for machine learning models to train models to predict climate parameters with finer spatial resolution. Research has found that in most cases (70%), weather index insurance crafted with climate data that has been spatially refined through machine learning techniques has shown enhanced effectiveness in risk mitigation, with these enhancements being statistically significant. Notably, insurance products tailored using downscaled temperature and rainfall data have demonstrated superior performance in diminishing inherent risks and amplifying the potential for risk reduction. In addition to common agricultural related indices such as weather and yield, parameter insurance has also launched products in disasters such as fires, floods, and typhoons. Myoung and Sunghai (2020) have proposed a new Korea Fire Risk Index (NKFRI) that covers all types of buildings and factories, particularly those designated by South Korean law as exceeding 3000 square meters and other specific buildings. It improves the accuracy of fire risk assessment by optimizing the weights of each component. NKFRI considers various variables (components) related to fire occurrence, which are divided into different modules and categories, such as basic hazards (such as building age, number of floors, structure, scale, fire load, etc.), ignition hazards (such as fire sources, gas facilities, hazardous material facilities, power facilities, etc.), and process hazards (only applicable to factories). The research results indicate that Deep Neural Network and NKFRI provide superior performance in fire risk prediction and management compared to traditional Korea Fire Risk Index (Choi & Jun, 2020).

4.3 Future of Health Index Insurance with ML

The development of index insurance products in the health industry is still in its infancy, but with the development of big data and artificial intelligence technology, this field has shown great potential and necessity. With the popularity of social media and smartphone applications, real-time monitoring of personal health data has become easier. Data obtained from electronic health records (EHR), electronic medical records (EMR), and electronic patient records (EPR), as well as data collected through social media and healthcare related smartphone applications, provide strong data support for the development of insurance products (Subrahmanya et al., 2022). These technologies can monitor individual health parameters, provide real-time data for index insurance products, help insurance companies assess risks related to specific events, and develop and price health index insurance products based on this. When developing index insurance products for the health industry, insurance companies can use big data and machine learning technologies to improve the accuracy of risk assessment and achieve personalized pricing strategies. For example, by analyzing

individual health data, insurance companies can customize insurance costs for each customer, thereby making premium rates more equitable (Hanafy & Ming, 2021). In addition, machine learning models, especially ensemble methods and deep neural networks, have demonstrated superior performance in predicting claims and fraud detection (A & B, 2021), which helps to improve the speed and efficiency of health insurance payouts.

Although the development of index insurance products in the health industry is feasible, there are also some challenges in their promotion. For example, the existence of underlying risk is the main reason for the consistently low demand for index insurance (Sun, 2022). To overcome these challenges, insurance companies can take the following measures:

- Personalized risk assessment: With the development of IoT (Internet of Things) technology, insurance companies are able to collect a large amount of personal health data. Combined with machine learning algorithms, this data can be used to create more refined personal health profiles. On the premise of respecting the privacy rights of the insured, insurance companies can analyze detailed personal health data, including data from wearable devices such as smartwatches and fitness trackers. In addition, the integration of genetic technology has opened up new possibilities for personalized risk assessment. Genetic information and genetic testing results can be used to predict the incidence rate of certain insured diseases, enabling insurance companies to provide more accurate pricing of insurance products.
- Product innovation: Machine learning models, especially deep learning techniques, are helping insurance companies develop new index insurance products. These products may be based on specific health conditions or treatment outcomes, for example, by analyzing an individual's genetic information to predict the risk of developing specific diseases and designing insurance products accordingly. In addition, the combination of wearable devices and health management services provides new directions for insurance product innovation. Insurance companies can provide insurance products related to health trackers, encouraging users to reduce premiums through healthy lifestyles. Such products can not only attract customers with strong health awareness, but also help reduce the insurance company's claims risk.RegTech: RegTech is the field of utilizing new technologies to meet regulatory requirements. In response to escalating regulatory demands, insurance companies must manage and scrutinize substantial volumes of data to remain compliant. Machine learning can help insurance companies automate and optimize these processes, improving compliance efficiency. For example, by using data analysis and machine learning techniques, insurance companies can identify and report potential fraudulent behavior faster, while also better understanding and predicting market trends, thus preparing and responding to potential risks in advance.

5 CONCLUSIONS

The application of machine learning technology in insurance pricing, claim detection, and index insurance has brought significant changes to the insurance industry. This article reviews how these technologies can improve the efficiency and accuracy of the insurance industry, and demonstrates the potential of machine learning in the insurance industry through theoretical analysis and literature review.

In terms of insurance pricing, machine learning technology analyzes historical claim data to enable insurance companies to predict payout probabilities more accurately, thereby achieving more refined pricing strategies. The application of this technology reduces the problem of information asymmetry and improves the accuracy and efficiency of pricing. Machine learning models are capable of handling complex data patterns, including non-linear relationships and high-dimensional data, providing insurance companies with a new tool to improve pricing accuracy.

The progress in claims detection is particularly significant. The application of machine learning technology has significantly improved the intelligence level of claim detection. By analyzing patterns and anomalies in claim data, insurance companies can more effectively identify fraudulent claims. The application of this technology not only improves detection speed, but also enhances accuracy, thereby protecting the interests of insurance companies and honest customers. Machine learning models, including supervised learning, unsupervised learning, and reinforcement learning, are being used to develop more advanced claim prediction systems.

In the field of index insurance, the application of machine learning technology has demonstrated tremendous innovation potential. By analyzing climate data and optimizing risk assessment models, machine learning provides insurance companies with the opportunity to develop new insurance products. These products can better adapt to market changes and provide customers with more flexible insurance solutions. Index insurance provides compensation by linking it to indices such as weather and natural disasters, thus avoiding the issues of adverse selection and moral hazard in traditional insurance.

Although machine learning technology has brought many benefits to the insurance industry, there are also some challenges. Issues such as data privacy, model transparency, and regulatory compliance require joint efforts from insurance companies and regulatory agencies to address. In addition, with the development of technology, ensuring the fairness and ethics of machine learning models is also an important direction for future research. Insurance companies need to ensure the security and privacy of customer data while utilizing these technologies.

This review is based on existing literature and theoretical analysis, and future research can further explore the application effects of machine learning technology in practical insurance business. Empirical research can provide deeper insights and help insurance companies better understand and apply these technologies. In addition, interdisciplinary research methods such as combining economics, statistics, and computer science may bring new perspectives and solutions to the insurance industry. Future research should focus on how to integrate machine learning techniques with the specific needs of the insurance industry, as well as how to evaluate and improve the effectiveness of these technologies in practical applications.

The application of machine learning technology in the insurance industry is a constantly evolving field, providing opportunities for insurance companies to improve efficiency, optimize risk management, and innovate products. With the continuous advancement of technology, insurance companies need to constantly adapt and innovate to fully utilize the potential brought by these technologies. Future insurance services will be more intelligent, efficient, and user-friendly, but at the same time, attention needs to be paid to the challenges and ethical issues brought by technology. Insurance companies should actively explore how to integrate machine learning technology into their business processes, while ensuring that the implementation of these technologies does not harm customer interests or violate regulatory regulations.

REFERENCES

Afshar, M. H., Foster, T., Higginbottom, T. P., Parkes, B., Hufkens, K., Mansabdar, S., et al. (2021). Improving the performance of index insurance using crop models and phenological monitoring. *Remote Sensing*, 13(5), 924. https://doi.org/10.3390/rs13050924

A, J. K., B, B. B., C, C. G., D, R. E., E, E. W., & F, B. R. (Year). Predicting flood insurance claims with hydrologic and socioeconomic demographics via machine learning: exploring the roles of topography, minority populations, and political dissimilarity. *Journal of Environmental Management*, 272. https://doi.org/10.1016/j.jenvman.2021.112421

A, M. K. S., & B, Y. P. (2021). Machine learning algorithms for fraud prediction in property insurance: empirical evidence using real-world microdata. *Machine Learning with Applications*. https://doi.org/10.1016/j.mla.2022.12.003

Aslam, F., Hunjra, A. I., Ftiti, Z., Louhichi, W., & Shams, T. (2022). Insurance fraud detection: Evidence from artificial intelligence and machine learning. *Research in International Business and Finance*, 62, 101744. https://doi.org/10.1016/j.ribaf.2022.101744

Chen, Z., Lu, Y., Zhang, J., & Zhu, W. (2024). Managing weather risk with a neural network-based index insurance. *Management Science*. Advance online publication. https://doi.org/10.1287/mnsc.2023.4149

Choi, M. Y., & Jun, S. (2020). Fire risk assessment models using statistical machine learning and optimized risk indexing. *Applied Sciences*, 10(12), 4199. https://doi.org/10.3390/app10124199

Eling, M., Nuessle, D., & Staubli, J. (2022). The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 47. https://doi.org/10.1057/s41288-022-00207-8

Eltazarov, S., Bobojonov, I., Kuhn, L., & Glauben, T. (2023). Improving risk reduction potential of weather index insurance by spatially downscaling gridded climate data-a machine learning approach. *Big Earth Data*, 7(4), 937-960. https://doi.org/10.1080/20964471.2023.2179475

Hanafy, M., & Ming, R. (2021). Machine learning approaches for auto insurance big data. *Risks*, 9(2), 42. https://doi.org/10.3390/risks9020042

Kaffash, S., Azizi, R., Huang, Y., & Zhu, J. (2019). A survey of data envelopment analysis applications in the insurance industry 1993-2018. *European Journal of Operational Research*, 284(3), 834-849. https://doi.org/10.1016/j.ejor.2019.07.034

Lyubchich, V., Newlands, N., Ghahari, A., Azar, M. T., & Gel, Y. R. (2019). Insurance risk assessment in the face of climate change: integrating data science and statistics. *Wiley Interdisciplinary Reviews:*

Computational Statistics, 11(4), e1457. https://doi.org/10.1002/wics.1457

Nguyen, T. T., Mushtaq, S., Kath, J., Nguyen-Huy, T., & Reymondin, L. (2024). Satellite-based data for agricultural index insurance: a systematic quantitative literature review. *EGUsphere*, 2024, 1-23. https://doi.org/10.5194/egusphere-egu2024-1

Paruchuri, H. (2020). The impact of machine learning on the future of insurance industry. *ABC Journals*, 3. https://doi.org/10.29029/abcj.2020.3.1.7

Poufinas, T., Gogas, P., Papadimitriou, T., & Zaganidis, E. (2023). Machine learning in forecasting motor insurance claims. *Risks*, 11(9), 164. https://doi.org/10.3390/risks11090164

Rawat, S., Rawat, A., Kumar, D., & Sabitha, A. S. (2021). Application of machine learning and data visualization techniques for decision support in the insurance sector. *International Journal of Information Management Data Insights*. https://doi.org/10.1108/IJIMDI-09-2021-0102

Rusdah, D. A., & Murfi, H. (2020). Xgboost in handling missing values for life insurance risk prediction. *SN Applied Sciences*, 2(8), 1-10. https://doi.org/10.1007/s42452-020-03426-1

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., & Ten Veldhuis, J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Natural Hazards & Earth System Science*, 14(12), 3345-3355. https://doi.org/10.5194/nhess-14-3345-2014

Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. M. Z., Paul, R., Smriti, K., et al. (2022). The role of data science in healthcare advancements: applications, benefits, and future prospects. *Irish Journal of Medical Science*, 191(4), 1473-1483. https://doi.org/10.1007/s11845-022-02439-4

Sun, Y. (2022). Enhanced Weather-Based Index Insurance Design for Hedging Crop Yield Risk. *Frontiers in Plant Science*, 13, 895183. https://doi.org/10.3389/fpls.2022.895183

Ul Hassan, C. A., Iqbal, J., Hussain, S., Alsalman, H., Mosleh, M. A. A., & Sajid Ullah, S. (2021). A computational intelligence approach for predicting medical insurance cost. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2021/6347849

Zhang, J., Zhang, Z., Wang, C. Z., Wang, X., Zhang, L. L., Ma, X., et al. (2022). Weather index insurance can offset heat‐induced rice losses under global warming. *Earth's Future*, 10. https://doi.org/10.1029/2022EF002531