

IMPROVING THE INSURANCE INDUSTRY: A CONCEPTUAL FRAMEWORK FOR APPLYING MACHINE LEARNING BASED ON A SYSTEMATIC LITERATURE REVIEW

Nikola Medan¹, Dejana Kresović²

¹ Faculty of Organizational Sciences, University of Belgrade

² Securities Commission

Received: 4.7.2025; Accepted: 15.12.2025

Review Article

Correspondence: dejana.kresovic@sec.gov.rs

Abstract: The insurance industry is undergoing a significant transformation driven by advancements in technology, particularly machine learning. As insurers seek to enhance operational efficiency, risk assessment, and customer experience, machine learning offers promising applications across various domains, such as underwriting, claims processing, and fraud detection. Despite the potential of machine learning, its integration into traditional insurance practices faces numerous challenges, including data quality, regulatory concerns, and organizational readiness. The aim of this paper is to examine the possibilities and characteristics of the application of machine learning in insurance, in order to determine the machine learning approach that is most often used and that provides the best results. Drawing on insights from systematic literature reviews, the framework will provide a comprehensive understanding of how machine learning can reshape insurance practices. By exploring these aspects, this paper contributes to a more structured and informed approach to implementing machine learning in the insurance industry.

Keywords: machine learning, insurance, learning algorithms

Introduction

Today, the insurance sector is highly developed and complex, not only because it provides risk coverage and protection for various forms of property and life. In fact, insurance plays a crucial role in socio-economic development, as well-developed insurance markets contribute to financial stability, foster business innovation, and support social progress.¹ Insurance serves as a cornerstone of modern society and economic systems, allowing individuals and organizations worldwide to rely on healthcare programs to protect their lives and assets. For instance, car insurance is legally required in most countries. However, the insurance industry faces significant challenges that raise the cost of doing business. To offset financial pressure, insurers often increase premiums, which negatively impacts honest policyholders and undermines overall industry efficiency. Traditionally, insurance companies have relied on human experts to manage

these challenges. While this approach can be effective, it is inherently complex, time-consuming, and costly, requiring substantial human resources. Moreover, manual processes increase the risk of errors and often lead to delays in claim resolution. These delays reduce operational efficiency and may even result in business losses. Recent advances in digital technology - particularly in machine learning and big data - offer new opportunities to address these issues.² The insurance industry holds significant potential to harness algorithmic capabilities to enhance every stage of the value chain. In this context, machine learning and artificial intelligence offer numerous opportunities, with the potential to boost revenue generation, improve loss prediction accuracy, and develop more effective loss prevention strategies.³

"Machine learning is a process that involves learning from data and transforming it into relevant information, which can then be used to

¹ Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., and Castignani, G.: Explainable artificial intelligence (XAI) in insurance. *Risks*, Vol. 10, 1-50. (2022)

² Guo, Y.: Application of machine learning in insurance fraud detection: achievements and future prospects. *Proceedings of the 2024 International Conference*

on Artificial Intelligence and Communication (ICAIC 2024), 619-626. (2024)

³ Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., and Castignani, G.: Explainable artificial intelligence (XAI) in insurance. *Risks*, Vol. 10, 1-50. (2022)

generate knowledge. It is a multi-disciplinary field that draws on elements of statistics, philosophy, epistemology, psychology, and neuroscience. Machine learning approaches offer numerous application possibilities in the insurance sector, and their impact is widely recognized by industry professionals. Based on this, the aim of this paper is to explore the possibilities and characteristics of applying machine learning in insurance, with the goal of identifying the most used approaches and those that deliver the best results. In line with this objective, the following research question has been defined:

How can machine learning transform the insurance industry and what conceptual framework can be developed to enable its effective implementation in insurance companies?

Given the lack of research on this topic within the domestic academic community, the results are expected to provide valuable insight into the application of machine learning in the insurance industry and offer a practical framework for its adoption. By thoroughly categorizing how machine learning is applied in areas such as fraud detection, risk assessment, customer segmentation, claims processing, premium de-

termination, and similar fields, it will be possible to build a solid information base for better understanding the scope and depth of its integration in the insurance sector. Identifying the most used algorithms and evaluating their effectiveness in addressing specific insurance challenges will help highlight the strengths and weaknesses of each approach, enabling the selection of those that deliver optimal results. The study is also expected to uncover under-researched areas in the existing literature, laying the groundwork for future studies.

Literature review

Conceptual aspects of machine learning

Machine learning is a broad field that integrates information technology, statistics, probability, artificial intelligence, psychology, neurobiology, and several other disciplines. It facilitates problem-solving by building models that accurately represent selected data sets. Machine learning is a subfield of computer science focused on developing systems that can learn from experience and improve their performance over

time.⁴ As a research field focused on the theory, performance, and characteristics of systems and algorithms, machine learning has impacted nearly every scientific domain, significantly influencing both science and society. It is a branch of computer science aimed at enabling systems to learn without explicit programming. As a subfield of artificial intelligence, machine learning emphasizes practical applications such as prediction and optimization. Systems learn by improving their task performance through experience, which usually involves fitting models to data. Consequently, the line between machine learning and statistical methods is often blurred, with classification depending more on historical context than fundamental differences. Despite methodological similarities, machine learning prioritizes predictive accuracy over hypothesis-driven inference and typically handles large, high-dimensional datasets with numerous variables.⁵

As Lee et al point out, machine learning is usually classified into supervised and unsupervised machine learning.⁶ To this can be added semi-supervised and reinforced machine learning. *Supervised machine learning* involves teaching a system to learn a function that maps inputs to outputs using labeled training data, consisting of sample input-output pairs. The two primary supervised tasks are *classification*, which categorizes data, and *regression*, which fits data to predict continuous values. *Unsupervised learning* analyzes unlabeled data without human intervention, following a data-driven approach. It is primarily used to discover hidden patterns, trends, structures and groups within data. Typical unsupervised tasks include cluster analysis, density estimation, dimensionality reduction, anomaly detection, and association rule discovery. *Semi-supervised learning* combines both supervised and unsupervised methods, working with a mixture of labeled and unlabeled data. This approach

⁴ Surden, H.: Machine learning and law. *Washington Law Review*, Vol. 89, 87-115. (2014)

⁵ Bi, Q., Goodman, KE, Kaminsky, J., and Lessler, J.: What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*, Vol. 188, 2222-2239. (2019)

⁶ Li, T., Johansen, K., and McCabe, MF: A machine learning approach for identifying and delineating agricultural fields and their multi-temporal dynamics using three decades of Landsat data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 186, 83-101. (2022)

is useful in situations where labeled data is scarce but unlabeled data is abundant. The goal of semi-supervised learning is to achieve better predictive outcomes than using only labeled data. It is applied in areas such as machine translation, fraud detection, text classification and data annotation. *Reinforcement learning* is a type of machine learning that allows systems to determine optimal behavior in a given environment to maximize rewards or minimize penalties. This approach is determined by the environment and relies on feedback, where actions are taken either to increase reward or to decrease risk. It is widely used in applications such as robotics, autonomous driving, manufacturing and supply chain optimization, but is generally not suitable for simple problems.⁷

Multitask learning is a subfield of machine learning that aims to solve multiple tasks simultaneously by leveraging the similarities between them. This approach can enhance learning efficiency and act as a regularizer. In multitask learning, a single model is trained on several related tasks to improve perfor-

mance by sharing knowledge across all tasks. This differs from traditional deep learning methods, which typically focus on solving one task per model. Ensemble learning involves combining multiple models, such as classifiers or experts, to address a computational intelligence problem. Its primary goal is to improve overall model performance or reduce the risk of selecting a poor model. Ensemble methods are also used for assigning confidence to model decisions, feature selection, data fusion, incremental learning, non-stationary learning, and error correction. Bagging, or bootstrap aggregation, is a technique designed to increase the accuracy and stability of machine learning algorithms. It applies to both classification and regression tasks by reducing variance and preventing overfitting. Bagging creates multiple subsets of training data through random sampling and trains separate models on each subset, ultimately enhancing overall model performance.⁸

Machine learning algorithms are diverse, with artificial neural networks being one prominent type mo-

⁷ Sarker, IH: Machine learning: algorithms, real-world applications and research directions. SN Computer Science, Vol. 2, 1-21. (2021)

⁸ Mahesh, B: Machine learning algorithms - a review. International Journal of Science and Research, Vol. 9, 381-386. (2020)

deleted after the behavior of neurons in biological neural networks. Comprising interconnected neurons, these networks analyze complex relationships between measurable variables to predict outcomes. The networks consist of multiple layers of neurons connected by "axons," organized into three types: 1) an input layer, 2) one or more hidden layers, and 3) an output layer. Neurons in the input layer represent independent variables, while those in the output layer correspond to dependent variables.⁹ A decision tree is a graphical representation of choices and their outcomes arranged in a tree structure. Nodes represent decisions or events, while edges represent decision rules or conditions. Each tree consists of nodes and branches, where nodes correspond to attributes of the group being classified, and branches represent possible values for each attribute. The Naive Bayes classifier is based on Bayes' theorem and assumes independence among predictors, meaning the presence of one feature in a class is considered unrelated to the presence of others. Naive Bayes is commonly used in text classification and tasks involving clustering and

classification based on conditional probabilities. Support Vector Machines (SVM) are widely used supervised learning models for both classification and regression, capable of handling linear and nonlinear classification by mapping inputs into high-dimensional feature spaces using the "kernel trick." Linear Discriminant Analysis (LDA) is a classifier that creates a linear decision boundary by fitting class conditional densities to the data and applying Bayes' rule. It generalizes Fisher's linear discriminant by projecting data into a lower-dimensional space to reduce model complexity, typically assuming each class follows a Gaussian distribution with a shared covariance matrix. LDA is related to techniques such as ANOVA and regression analysis, aiming to express the dependent variable as a linear combination of features. Logistic regression is a probabilistic statistical model used for classification, employing the logistic (sigmoid) function to estimate probabilities. K-Nearest Neighbors (KNN) is an instance-based learning algorithm that does not build a general model but stores training instances in an n-dimensional

⁹ Bi, Q., Goodman, KE, Kaminsky, J., and Lessler, J.: What is Machine Learning? A Primer for the Epidemiologist.

American Journal of Epidemiology, Vol. 188, 2222-2239. (2019)

space. It classifies new data points based on similarity measures like Euclidean distance, with classification determined by majority vote among the nearest neighbors. The main challenge in KNN is selecting the optimal number of neighbors to consider.¹⁰

With the rapid advancements in the field, advanced learning methods have emerged as key trends in machine learning. Deep learning, in particular, has become dominant by utilizing deep architectures to automatically learn hierarchical representations that capture complex patterns in data. This allows deep learning models to outperform traditional shallow methods in areas such as speech recognition, computer vision, and natural language processing. Distributed learning addresses the challenge of processing large datasets that exceed the capacity of a single machine by distributing computations across multiple workstations, effectively scaling learning algorithms without centralizing data, thereby saving time and energy. Alongside distributed learning, parallel machine

learning techniques, supported by multi-core processors and cloud computing, are increasingly accessible for large-scale applications. Transfer learning enables the application of knowledge gained from one task or domain to different, often related, tasks or domains, helping overcome data scarcity in new tasks. Kernel-based learning techniques have become prominent for handling nonlinear problems by projecting input data into high-dimensional feature spaces using kernel functions, where linear methods can be applied efficiently. This approach empowers the solution of complex problems such as online classification and parameter estimation through implicit high-dimensional mappings.¹¹

Possibilities of applying machine learning in insurance

Machine learning has the ability to identify complex patterns within data sets, thereby enabling the discovery of relationships and structures that may not be immediately apparent. The potential of machine learning is reflected in its ability to

¹⁰ Sarker, IH: Machine learning: algorithms, real-world applications and research directions. SN Computer Science, Vol. 2, 1-21. (2021)

¹¹ Qiu, J., Wu, Q., Ding, G., Xu, Y., and Feng, S.: A survey of machine learning for big data processing. EURASIP Journal on Advances in Signal Processing, Vol. 1, 1-16. (2016)

Medan, N., Kresović, D. Improving the Insurance Industry: A Conceptual Framework for Applying Machine Learning Based on a Systematic Literature Review

push the boundaries of forecasting, offering new tools and methods that could transform forecasting and analysis.¹² Hence, the effective application of machine learning in insurance can be highlighted. For example, insurance companies have widely recognized the impact of artificial intelligence in the insurance industry, particularly in claims prediction. Machine learning has been applied to almost every aspect of the insurance process, including claims processing, fraud detection, decision making, loss prediction, risk management, and the like.¹³

Grise et al. (2020) highlight the positive impact of machine learning in non-life insurance, especially in risk assessment, which improves the long-term profitability of insurance companies.¹⁴ The insurance industry holds significant potential to leverage algorithmic capabilities that can enhance various stages of the value chain. According to Pavlović (2019), the European Insurance and

Occupational Pensions Authority (EIOPA) conducted a study on big data analytics, revealing that about one-third of European insurance companies use machine learning in their operations, particularly in health and car insurance. The study involved 222 insurance companies and intermediaries from 28 jurisdictions. Machine learning methods are actively employed by 31% of these companies, while another 24% are still evaluating their potential applications.¹⁵

Previous research analyzes various applications of artificial neural networks in the insurance industry, including insolvency management, fraud detection, revenue forecasting and customer segmentation. One study developed a three-layer neural network model to provide early warnings of insolvency for insurance companies using annual financial data, with the model achieving classification accuracy above 88%.¹⁶ Another study used artificial neural

¹² Paruchuri, H.: Conceptualization of Machine Learning in Economic Forecasting. *Asian Business Review*, Vol. 11, 51-58. (2021)

¹³ Poufinas, T., Gogas, P., Papadimitriou, T., and Zaganidis, E.: Machine learning in forecasting motor insurance claims. *Risks*, Vol. 11, 1-19. (2023)

¹⁴ Grise, YL, Wolfram, F., and Christian, L.: Machine learning applications in

nonlife insurance. *Applied Stochastic Models in Business and Industry*, Vol. 36, 523-537. (2020)

¹⁵ Pavlović, B.: Challenges in application of machine learning in insurance industry. *Insurance Flows*, Vol. 35, 7-34. (2019)

¹⁶ Tian, X., Todorović, J., and Todorović, Ž.: A machine-learning-based business analytical system for insurance

networks to predict insurance company revenue growth over a 41-year period, predicting a 120% increase based on historical premium data.¹⁷ In other research, machine learning algorithms have been used to predict insurance customer types, estimate premiums, manage risk, and the like.¹⁸ Regardless of the field or methodology, machine learning offers clear advantages to insurance companies. Traditional actuarial methods, which rely on statistical techniques and historical data, often produce less accurate premium estimates, especially for complex risk profiles. In contrast, machine learning models incorporate diverse data inputs, including real-time data streams, enabling deeper analyses that uncover patterns and correlations often overlooked by conventional approaches. This flexibility allows machine learning systems to rapidly adapt to

changing risk factors and personalize premiums for individual policyholders. For instance, decision trees and neural networks can adjust to new behavioral data and evolving risks, delivering more accurate and customized risk assessments and premium calculations compared to traditional methods.¹⁹

Machine learning excels at automatically capturing non-linear relationships in data, leading to more accurate models and providing the flexibility to adopt different functional forms.²⁰ Insurance companies in developed economies are increasingly adopting machine learning to leverage its advantages. For example, State Farm in the US uses machine learning to classify drivers based on their driving behavior, enabling them to offer tailored insurance products for different driver categories. Liberty Mutual, one of America's largest insurers, estab-

customer relationship management and cross-selling. *Journal of Applied Business and Economics*, Vol. 25, 256-272. (2023)

¹⁷ Bahia, ISH: Using artificial neural network modeling in forecasting revenue: Case study in National Insurance Company/Iraq. *International Journal of Intelligence Science*, Vol. 3, 136-143. (2013)

¹⁸ Tian, X., Todorović, J., and Todorović, Ž.: A machine-learning-based business analytical system for insurance custo-

mer relationship management and cross-selling. *Journal of Applied Business and Economics*, Vol. 25, 256-272. (2023)

¹⁹ Ejjami, R.: Machine learning approaches for insurance pricing: a case study of public liability coverage in Morocco. *International Journal For Multidisciplinary Research*, Vol. 6, 1-23. (2024)

²⁰ Casualty Actuarial Society. *Machine learning in insurance*. Casualty Actuarial Society, Arlington, USA. (2022).

lished Solaria Labs to drive innovation; in 2017, Solaria Labs created an open API portal to integrate IT projects with public data, aiming to develop a traffic safety application powered by machine learning. Allstate developed ABIE (Allstate Business Insurance Expert), a virtual assistant chatbot that uses machine learning to help agents sell complex property insurance products, significantly boosting sales. Progressive, another major US insurer, applied machine learning algorithms for predictive analytics to analyze driver data, better understand market trends, and improve motor insurance products. Their telematics program, Snapshot, collected 20 billion kilometers of driving data in 2016, enabling precise decision-making in motor insurance sales.²¹

However, the application of machine learning in insurance also presents several challenges. For decisions based on machine learning to be effective, they require a set of high-quality and objective data, which is often lacking. This can lead

to biased or discriminatory outcomes for certain categories of policyholders. Additionally, implementing machine learning and interpreting its results is highly complex and demands specialized expertise, which many insurance companies may not possess.²² In addition to these issues, there are ethical challenges associated with machine learning in insurance. Collecting and analyzing personal data, including demographics and behavioral patterns, requires robust data protection measures to prevent unauthorized access. Ethical concerns also focus on the fairness of machine learning algorithms, especially regarding sensitive data, as they risk reinforcing existing biases or creating new forms of discrimination, particularly when trained on unbalanced or biased datasets. Integrating machine learning into insurance systems poses technical challenges, often requiring major updates to legacy systems that may not support large-scale data analytics or real-time processing. Furthermore, data integration is complex, invol-

²¹ Pavlović, B.: Challenges in application of machine learning in insurance industry. *Insurance Flows*, Vol. 35, 7-34. (2019)

²² Jones, KI, and Sah, S.: The implementation of machine learning in the insu-

rance industry with big data analytics. *International Journal of Data Informatics and Intelligent Computing*, Vol. 2, 21-38. (2023)

ving the merging, normalization, and standardization of diverse data sources to ensure consistency for effective model training.²³

Research on the application of machine learning in insurance companies

Methodology research

To address the research question, a study employing a qualitative methodology was conducted, which included a systematic literature review and the development of a conceptual framework to explore the application of machine learning in the insurance industry. The methodology aimed to identify the most commonly used machine learning applications in insurance and their impacts. According to the author's knowledge, such studies remain limited in the academic community, particularly regarding the Republic of Serbia. The systematic review gathered and analyzed existing knowledge from peer-reviewed journals and conference proceedings to ensure data credibility and relevance. The review process followed a structured approach:

1. Search strategy: keywords such as machine learning in insurance, AI in insurance, and machine learning were used to search academic databases including PubMed, IEEE Xplore, Scopus, and Google Scholar.
2. Inclusion and exclusion criteria: priority was given to articles published in English over the last five years, focusing on studies with empirical findings. Publications lacking detailed methodology or relevance to the insurance sector were excluded.
3. Data extraction and analysis: selected articles were reviewed to extract information on key machine learning techniques, areas of implementation in insurance, benefits, limitations, and related aspects.

Based on insights from the literature review, a conceptual framework was developed to bridge theoretical knowledge and practical application of machine learning in insurance. This framework highlights the most common forms of machine learning applications across various insurance domains, along with their impacts. Analysis

²³ Ejjami, R.: Machine learning approaches for insurance pricing: a case study of public liability coverage in Morocco.

International Journal For Multidisciplinary Research, Vol. 6, 1-23. (2024)

Medan, N., Kresović, D. Improving the Insurance Industry: A Conceptual Framework for Applying Machine Learning Based on a Systematic Literature Review

of published case studies and empirical research, validated through secondary data, was used to demonstrate the feasibility and effectiveness of machine learning in the insurance sector. Drawing on these findings, conclusions were made regarding the effectiveness of different types of machine learning in insurance, based on comparative evaluation.

Results and discussion

Insurance claims

Several selected studies have applied machine learning techniques to predict motor vehicle insurance claims using claim history and telematics data. For example, some models forecast accidents by analyzing driving patterns, including annual distance traveled and the percentage of time spent in urban areas. Similar approaches use machine learning to predict zeroclaim occurrences by leveraging telematics data from auto liability insurance. Money laundering detection models have also been adapted for flood insurance claims by incorporating hydrological and socio-demographic data to enhance prediction

accuracy. Research involving Brazilian and Indian auto insurance datasets focused on predicting claim occurrence and amounts, identifying factors such as weather conditions and vehicle types as key predictors. Additionally, telematics-based studies have examined driving contexts like road type and traffic conditions to evaluate risk and anticipate motor insurance needs.²⁴

Alam and Prybutok (2024) conducted a study to predict health insurance demand in the US, employing six machine learning algorithms to forecast claims. The tested algorithms included support vector machines, decision trees, random forest, linear regression, extreme gradient boosting (XGBoost), and K-Nearest Neighbors (KNN). Their performance was evaluated using various metrics, and a feature importance analysis identified key variables influencing claim predictions. Results showed that XGBoost and random forest outperformed the others, achieving the highest coefficients of determination - 79% and 77%, respectively - with the lowest prediction errors. The analysis highlighted smoking habits, body mass index (BMI), and blood pressure le-

²⁴ Poufinas, T., Gogas, P., Papadimitriou, T., and Zaganidis, E.: Machine le-

arning in forecasting motor insurance claims. *Risks*, Vol. 11, 1-19. (2023)

vels as the most influential predictors. These findings emphasize the importance of incorporating these factors into insurance policy design and pricing strategies. The study demonstrates the transformative potential of artificial intelligence, particularly the XGBoost model, in enhancing the accuracy and efficiency of health insurance claims processing. By identifying critical variables and reducing forecasting errors, the approach offers significant cost-saving opportunities and underscores the value of machine learning for process optimization and data-driven decision-making in health insurance.²⁵

A study by Ejiyi et al. (2024) applied various machine learning algorithms to an insurance dataset from Africa to predict whether customers would file property claims. The data was preprocessed before implementation using Python. Different algorithms demonstrated specific strengths: Naive Bayes excelled in real-time predictions and multi-class problems, decision trees effec-

tively managed noisy data and prevented overfitting, while support vector machines were well-suited for sparse data. Performance was evaluated using the Gini index and SHAP values. The Gini index showed that the Kernel Support Vector Machine outperformed other models, with logistic regression closely following in prediction accuracy. SHAP analysis identified asset dimensions as the most influential factor, followed by asset type. Other algorithms designed for large or noisy datasets may perform better under specific conditions, making the choice of algorithm dependent on factors such as data-set size and quality.²⁶

The study by Pofuinas et al. (2023) aimed to predict the average cost of motor vehicle insurance claims using a dataset from Athens covering 2008 to 2020, structured quarterly due to data availability. Besides claims data, the dataset included variables such as the number of new and imported used cars in Athens and weather conditions (ma-

²⁵ Alam, A., and Prybutok, VR: Use of responsible artificial intelligence to predict health insurance claims in the USA using machine learning algorithms. *Exploration of Digital Health Technologies*, 30-45. (2024)

²⁶ Ejiyi, CJ, Qin, Z., Salako, AA, Happy, MN, Nneji, GU, Ukwuoma, CC, Chik-

wendu, IA, and Gen, J.: Comparative analysis of building insurance prediction using some machine learning algorithms. *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 7, 75-85. (2024)

ximum and minimum temperatures, days below zero, and rainy days) from three weather stations: Elefsina, Tatoi, and Spata. These stations were chosen based on data completeness. The research introduced two novel predictors for insurance claims: weather and car sales. Machine learning algorithms including Support Vector Machines, Decision Trees, Random Forest, and Boosting were applied to predict the average quarterly claim per insured vehicle. Key findings highlighted new car sales and minimum temperature in Elefsina as the most significant variables. Among models, a limited-depth Random Forest and XGBoost using the 15 most relevant variables showed the best predictive performance.²⁷

Customer segmentation

Machine learning plays an important role in the marketing function within the insurance sector. Owens et al. (2022) highlight that during the marketing phase of the insurance value chain, machine learning enhances the performance of various tasks. It enables more precise predictions of customer lifetime va-

lue and offers deeper insights into purchasing behavior, improving market and customer research. This allows for better identification of target segments and the creation of personalized premium strategies. Additionally, machine learning advances segmentation methods, ensuring communication and promotional efforts are effectively tailored to specific audiences.²⁸

Various studies have examined the use of machine learning and data analysis to enhance customer segmentation, retention, profitability, and satisfaction in insurance. Tian et al. (2023) review prior research, highlighting a three-level segmentation method that combines decision trees and cost-benefit analysis, with efficiency as a key outcome. Some approaches apply k-means clustering to categorize customers based on demographic attributes, followed by association rule mining to reveal hidden patterns. Dynamic segmentation using latent Dirichlet allocation is also used to identify behavioral clusters for personalized marketing strategies. For customer retention, models like logistic regre-

²⁷ Poufinas, T., Gogas, P., Papadimitriou, T., and Zaganidis, E.: Machine learning in forecasting motor insurance claims. *Risks*, Vol. 11, 1-19. (2023)

²⁸ Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., and Castignani, G.: Explainable artificial intelligence (XAI) in insurance. *Risks*, Vol. 10, 1-50. (2022)

ssion and neural networks predict policyholder behavior and help optimize pricing. Profitability forecasting often incorporates customer demographics and purchasing habits, with random forest models applied to predict client profitability.²⁹

Machine learning enables effective customer segmentation by analyzing behavior and preferences, aiding in new product development. Jones and Sah (2023) present a method using interpretable machine learning algorithms to analyze online insurance product reviews and assess the importance of specific features. The main challenge was uncovering and interpreting nonlinear relationships between feature satisfaction and overall customer satisfaction. To address this, the researchers applied interpretable techniques that balance strong predictive performance with clear, understandable results. Validated through a case study and compared to sentiment-based segmentation, their method showed superior clus-

tering performance and uncovered new opportunities for innovative insurance product development.³⁰

Qadadeha and Abdallah (2018) analyzed a dataset of 9,822 customers from 2,000 insurance companies, provided by Sentient Machine Research, a Dutch data mining firm. Each customer record included 86 attributes covering demographics, behavior, purchasing habits, and more. While the K-Means algorithm demonstrated effective clustering abilities, the Self-Organizing Map (SOM) outperformed it by achieving superior speed, higher clustering quality, and better data visualization. This highlights SOM's advantage in handling complex insurance customer segmentation tasks.³¹

Identification of insurance fraud

Machine learning techniques play a crucial role in detecting fraudulent claims across different insurance sectors. In auto insurance, money laundering methods are applied to classify suspicious claims and uncover potential fraud by analyzing

²⁹ Tian, X., Todorović, J., and Todorović, Ž.: A machine-learning-based business analytical system for insurance customer relationship management and cross-selling. *Journal of Applied Business and Economics*, Vol. 25, 256-272. (2023)

³⁰ Jones, KI, and Sah, S.: The implementation of machine learning in the

insurance industry with big data analytics. *International Journal of Data Informatics and Intelligent Computing*, Vol. 2, 21-38. (2023)

³¹ Qadadeha, W., and Abdallah, S.: Customers segmentation in the insurance company (TIC) dataset. *INNS Conference on Big Data and Deep Learning 2018*, 277-290. (2018)

accident data gathered from insurers. In health insurance, interactive frameworks incorporating money laundering techniques help identify fraudulent claims that involve multiple parties. Moreover, these methods can automatically categorize various types of motor insurance fraud, significantly reducing the reliance on manual investigation and enhancing efficiency.³²

Fraud prevention technologies are essential for reducing fraudulent insurance claims. Both statistical methods and machine learning approaches have demonstrated effectiveness in detecting various types of fraud, including money laundering, credit card fraud, telecom fraud, and cyber attacks. Studies on Medicare data show that supervised learning techniques typically outperform unsupervised methods. However, due to challenges in obtaining high-quality annotated fraud datasets in insurance, unsupervised learning methods are often favored. According to the Casualty Actuarial Society (2022), commonly used unsupervised techniques include K-means clustering, self-orga-

nizing maps (SOM), and principal component analysis (PCA).³³

Guo (2024) synthesizes various studies and highlights several key machine learning approaches for identifying and preventing insurance fraud. Supervised learning techniques, including ensemble learning, neural networks, and natural language processing, are widely applied by training models on labeled data to distinguish false claims from legitimate ones. When labeled data is limited, unsupervised methods such as clustering and anomaly detection are used to spot unusual patterns. Additionally, graph neural networks prove valuable in complex settings like health insurance by analyzing relationships and detecting collusion among multiple parties. Hybrid models that combine supervised and unsupervised techniques enhance fraud detection by both identifying anomalies and accurately classifying them. Continuous learning and frequent model updates are crucial to keep pace with evolving fraud schemes. Integration of blockchain technology further strengthens data security and ensures claim authenticity. Overall, combining these ad-

³² Poufinas, T., Gogas, P., Papadimitriou, T., and Zaganidis, E.: Machine learning in forecasting motor insurance claims. *Risks*, Vol. 11, 1-19. (2023)

³³ Casualty Actuarial Society. *Machine learning in insurance*. Casualty Actuarial Society, Arlington, USA. (2022).

vanced methods forms a robust strategy to improve fraud detection effectiveness and adapt to new fraud tactics in the insurance industry.³⁴

A landmark study developed an automated fraud detection framework designed to minimize employee intervention, enhance security, and reduce financial losses in insurance. This framework incorporates a blockchain-based system that enables secure transactions and data sharing among multiple agents within the insurance network. The study applied the extreme gradient boosting algorithm (XGBoost) for fraud detection and benchmarked its performance against other advanced algorithms. Results demonstrated that XGBoost outperformed competitors, including decision tree models, achieving up to 7% higher accuracy on an auto insurance dataset. Additionally, the research introduced an online learning solution that efficiently handles real-time data updates, surpassing other online algorithms in performance.³⁵

Machine learning has become a vital tool for detecting fraud across industries, especially in auto insurance. By analyzing vast amounts of historical data, it identifies patterns, anomalies, and inconsistencies within insurance claims, strengthening fraud prevention efforts. Algorithms like neural networks, Bayesian learning, artificial immune systems, and support vector machines examine claimant details, vehicle information, accident reports, and claim histories, showing high accuracy in spotting fraudulent claims. Unsupervised techniques such as cluster analysis group similar claims to uncover suspicious patterns, allowing insurers to focus investigations effectively. Additionally, supervised models like logistic regression and decision trees are widely used for risk assessment and fraud detection. Commonly applied models in auto insurance fraud include Naive Bayes, random forest, and XGBoost.³⁶

³⁴ Guo, Y.: Application of machine learning in insurance fraud detection: achievements and future prospects. Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024), 619-626. (2024)

³⁵ Jones, KI, and Sah, S.: The implementation of machine learning in the

insurance industry with big data analytics. International Journal of Data Informatics and Intelligent Computing, Vol. 2, 21-38. (2023)

³⁶ Mouna, SA, and Ilham, K.: Auto insurance fraud detection using machine learning contrasting US and Moroccan companies. Proceedings of the International

Determination of insurance premiums

Machine learning techniques are increasingly enhancing insurance premium pricing. Linear regression remains important for its simplicity and ability to model relationships between variables like age, gender, and claims history. Decision trees and ensemble methods, such as random forests, handle complex datasets where variable interactions influence premium calculations, creating decision branches for more precise risk assessments. Gradient boosting regression refines accuracy by iteratively correcting previous errors, while neural networks excel at detecting intricate patterns in high-dimensional data that simpler models might miss. A study on Moroccan vehicle insurance data compared polynomial regression, decision tree regression, random forest regression, and gradient boosting regression, evaluating their performance with metrics like mean square error (MSE), root mean square error (RMSE), and R^2 , demonstrating their effectiveness in predicting premiums and adapting to emer-

ging risks³⁷ Table 1 shows summarized results on the most commonly used machine learning methods in insurance.

Table 1. Machine learning methods in insurance

Insurance area	Machine learning
Receivables	Decision trees, random forest, neural networks, machines with support vectors, XGBoost.
Customer segmentation	K-means clusters, decision trees, Latent Dirichlet allocation
Fraud identification	Logistic regression, decision trees, XGBoost, graph neural networks.
Determining premiums	Decision trees, random forest, polynomial regression, neural networks, regression with increasing gradient.

Table 1 shows that machine learning methods such as decision trees,

of public liability coverage in Morocco. International Journal For Multidisciplinary Research, Vol. 6, 1-23. (2024)

Conference on Industrial Engineering and Operations Management, 226-235. (2024)

³⁷Ejjami, R.: Machine learning approaches for insurance pricing: a case study

random forest and neural networks are common in various aspects, which shows their flexibility in handling data of different complexity.

Conclusion

Machine learning has significantly transformed the insurance industry by enhancing efficiency and accuracy in key areas such as fraud detection, customer segmentation, and premium prediction. Its integration improves decision-making, boosts predictive precision, and enables more personalized services.

For fraud detection, supervised learning models like logistic regression, decision trees, and support vector machines are commonly employed to classify claims as fraudulent or legitimate, while unsupervised methods like clustering and anomaly detection help uncover fraud when labeled data is limited. Advanced techniques such as graph neural networks excel at detecting complex fraudulent schemes involving multiple entities. In customer segmentation, clustering algorithms like k-means and hierarchical clustering group clients based on shared traits, enabling targeted marketing and product development, with dynamic methods like Latent Dirichlet Allocation offering behavior-based segmentation. Claims management leverages decision

trees and random forests for classification, alongside neural networks and ensemble models like gradient boosting to enhance forecasting accuracy. Finally, premium forecasting still relies on linear regression for its clarity but increasingly benefits from advanced methods such as gradient boosting and random forests that better capture complex data patterns.

This study contributes significantly to both theoretical understanding and practical implementation of machine learning in the insurance sector. Theoretically, it enriches existing literature through a systematic review of cutting-edge machine learning applications in insurance, emphasizing their transformative effects on conventional insurance processes. Given the limited current research, the findings establish a solid foundation for future investigations in this field. Practically, the study offers valuable guidance for stakeholders by presenting a framework that matches specific machine learning techniques to various insurance functions. However, the research has limitations, primarily its qualitative approach and reliance on secondary data. To refine the understanding of machine learning's potential in insurance, future studies should incorporate real-world data from insurance companies and con-

Medan, N., Kresović, D. *Improving the Insurance Industry: A Conceptual Framework for Applying Machine Learning Based on a Systematic Literature Review*

duct case studies to explore how these organizations implement machine learning in practice.

References

Alam, A., and Prybutok, VR: Use of responsible artificial intelligence to predict health insurance claims in the USA using machine learning algorithms. *Exploration of Digital Health Technologies*, 30-45. (2024)

Bi, Q., Goodman, KE, Kaminsky, J., and Lessler, J.: What is Machine Learning? A Primer for the Epidemiologist. *American Journal of Epidemiology*, Vol. 188, 2222-2239. (2019)

Bahia, ISH: Using artificial neural network modeling in forecasting revenue: Case study in National Insurance Company/Iraq. *International Journal of Intelligence Science*, Vol. 3, 136-143. (2013)

Casualty Actuarial Society. *Machine learning in insurance*. Casualty Actuarial Society, Arlington, USA. (2022).

Ejjami, R.: Machine learning approaches for insurance pricing: a case study of public liability coverage in Morocco. *International Journal For Multidisciplinary Research*, Vol. 6, 1-23. (2024)

Ejiyi, CJ, Qin, Z., Salako, AA, Happy, MN, Nneji, GU, Ukwu-

oma, CC, Chikwendu, IA, and Gen, J.: Comparative analysis of building insurance prediction using some machine learning algorithms. *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 7, 75-85. (2024)

Grize, YL, Wolfram, F., and Christian, L.: Machine learning applications in nonlife insurance. *Applied Stochastic Models in Business and Industry*, Vol. 36, 523-537. (2020)

Guo, Y.: Application of machine learning in insurance fraud detection: achievements and future prospects. *Proceedings of the 2024 International Conference on Artificial Intelligence and Communication (ICAIC 2024)*, 619-626. (2024)

Jones, KI, and Sah, S.: The implementation of machine learning in the insurance industry with big data analytics. *International Journal of Data Informatics and Intelligent Computing*, Vol. 2, 21-38. (2023)

Li, T., Johansen, K., and McCabe, MF: A machine learning approach for identifying and delineating agricultural fields and their multi-temporal dynamics using three decades of Landsat data. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 186, 83-101. (2022)

- Mahesh, B.: Machine learning algorithms - a review. *International Journal of Science and Research*, Vol. 9, 381-386. (2020)
- Mouna, SA, and Ilham, K.: Auto insurance fraud detection using machine learning contrasting US and Moroccan companies. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 226-235. (2024)
- Nasteski, V.: An overview of the supervised machine learning methods. *Horizons*, 4, 51-62. (2017)
- Owens, E., Sheehan, B., Mullins, M., Cunneen, M., Ressel, J., and Castignani, G.: Explainable artificial intelligence (XAI) in insurance. *Risks*, Vol. 10, 1-50. (2022)
- Pavlović, B.: Challenges in application of machine learning in insurance industry. *Insurance Flows*, Vol. 35, 7-34. (2019)
- Paruchuri, H.: Conceptualization of Machine Learning in Economic Forecasting. *Asian Business Review*, Vol. 11, 51-58. (2021)
- Poufinas, T., Gogas, P., Papadimitriou, T., and Zaganidis, E.: Machine learning in forecasting motor insurance claims. *Risks*, Vol. 11, 1-19. (2023)
- Qadadeha, W., and Abdallah, S.: Customers segmentation in the insurance company (TIC) dataset. *INNS Conference on Big Data and Deep Learning 2018*, 277-290. (2018)
- Qiu, J., Wu, Q., Ding, G., Xu, Y., and Feng, S.: A survey of machine learning for big data processing. *EURASIP Journal on Advances in Signal Processing*, Vol. 1, 1-16. (2016)
- Sarker, IH: Machine learning: algorithms, real-world applications and research directions. *SN Computer Science*, Vol. 2, 1-21. (2021)
- Surden, H.: Machine learning and law. *Washington Law Review*, Vol. 89, 87-115. (2014)
- Tian, X., Todorović, J., and Todorović, Ž.: A machine-learning-based business analytical system for insurance customer relationship management and cross-selling. *Journal of Applied Business and Economics*, Vol. 25, 256-272. (2023)

UNAPREĐENJE INDUSTRIJE OSIGURANJA: KONCEPTUALNI OKVIR ZA PRIMENU MAŠINSKOG UČENJA ZASNOVANOG NA SISTEMATSKOM PREGLEDU LITERATURE

Sažetak: Industrija osiguranja prolazi kroz značajnu transformaciju vođenu napretkom u tehnologiji, posebno mašinsko učenje. Kako osiguravači nastoje da poboljšaju operativnu efikasnost, procenu rizika i korisničko iskustvo, mašinsko učenje nudi obećavajuće aplikacije u različitim domenima, kao što su preuzimanje, obrada potraživanja i otkrivanje prevare. Uprkos potencijalu mašinskog učenja, njegova integracija u tradicionalne prakse osiguranja suočava se sa brojnim izazovima, uključujući kvalitet podataka, regulatorne probleme i organizacionu spremnost. Cilj ovog rada je da se ispituju mogućnosti i karakteristike primene mašinskog učenja u osiguranju, kako bi se utvrdio pristup mašinskog učenja koji se najčešće koristi i koji daje najbolje rezultate. Oslanjajući se na uvide iz sistematskih pregleda literature, okvir će pružiti sveobuhvatno razumevanje kako mašinsko učenje može preoblikovati prakse osiguranja. Istražujući ove aspekte, ovaj rad doprinosi strukturiranim i informisanom pristupu implementaciji mašinskog učenja u industriji osiguranja.

Ključne reči: mašinsko učenje, osiguranje, algoritmi učenja