Journal of Infrastructure, Policy and Development 2024, 8(16), 8803.
https://doi.org/10.24294/jipd8803

Article

Sales strategies with a probabilistic business model of an insurance company: The role of updating and targeting

Krisztián Koppány[1,*], István Á. Harmati[2], Dávid Fülep[2], Norbert Kovács[1]

[1] Department of International and Applied Economics, Széchenyi István University, H-9026 Egyetem tér 1 Győr, Hungary
[2] Széchenyi István University, Department of Mathematics and Computational Sciences, H-9026 Egyetem tér 1 Győr, Hungary
*** Corresponding author:** Krisztián Koppány, koppanyk@sze.hu

CITATION

Koppány K, Harmati IA, Fülep D, Kovács N. (2024). Sales strategies with a probabilistic business model of an insurance company: The role of updating and targeting. Journal of Infrastructure, Policy and Development. 8(16): 8803. https://doi.org/10.24294/jipd8803

ARTICLE INFO

Received: 26 August 2024
Accepted: 20 September 2024
Available online: 18 December 2024

Abstract: The major goal of decisions made by a business organization is to enhance business performance. These days, owners, managers and other stakeholders are seeking for opportunities of modelling and automating decisions by analysing the most recent data with the help of artificial intelligence (AI). This study outlines a simple theoretical model framework using internal and external information on current and potential clients and performing calculations followed by immediate updating of contracting probabilities after each sales attempt. This can help increase sales efficiency, revenues, and profits in an easily programmable way and serve as a basis for focusing on the most promising deals customising personal offers of best-selling products for each potential client. The search for new customers is supported by the continuous and systematic collection and analysis of external and internal statistical data, organising them into a unified database, and using a decision support model based on it. As an illustration, the paper presents a fictitious model setup and simulations for an insurance company considering different regions, age groups and genders of clients when analysing probabilities of contracting, average sales and profits per contract. The elements of the model, however, can be generalised or adjusted to any sector. Results show that dynamic targeting strategies based on model calculations and most current information outperform static or non-targeted actions. The process from data to decision-making to improve business performance and the decision itself can be easily algorithmised. The feedback of the results into the model carries the potential for automated self-learning and self-correction. The proposed framework can serve as a basis for a self-sustaining artificial business intelligence system.

Keywords: sales strategy; targeting; insurance; contracting probability; Bayesian updating; artificial intelligence

1. Introduction

The ideal customer is one that is cheap to acquire and generates a large amount of revenue with a high profit content. Customer acquisition is cheap if one can reach new clients easily and quickly, in a targeted way, through cost-effective sales channels, i.e., by reaching those who have a high willingness (probability) to buy the product. In addition to sales expenses, the profit that can be achieved should also be taken into consideration, and based on this, by targeting groups of potential clients for whom the expected value of profit is the highest, the business performance of the company can be enhanced by the proper recommendation of products (Vij and Preethi, 2021).

The expected behaviour of potential customers can be predicted by analysing the characteristics of existing or previously contacted potential customers. These characteristics (demographic data, place of residence, place and date of birth, age, income, marital status, etc. in the case of individual (B2C) clients; industry, company size, turnover size, headcount, etc. in the case of business (B2B) partners) are the criteria by which clients can be classified into homogeneous groups, a.k.a. clusters or segments (Qadadeh and Abdallah, 2018). The goal is to determine how likely it is to contract with a new customer with certain characteristics (belonging to a particular group). A company may have an advantage over its competitors if (a) it targets potential customers who are more likely than usual to do profitable business; (b) for each type of customer, it can select the sales channels through which customer acquisition is the most effective, or vice versa, it can select the customer groups to be targeted for a given sales channel; or, if (c) it contacts a potential customer through any channel, it offers a product that has the highest buying probability and profit, i.e. expected profit.

This study presents a simple theoretical business model serving the goals defined above. The proposed probabilistic microeconomic framework uses internal and external information on current and potential clients and performing calculations followed by immediate updating of contracting probabilities after each sales attempt. This can help increase sales efficiency, revenues, and profits in an easily programmable way and serve as a basis for focusing on the most promising deals customising personal offers of best-selling products for each potential client. The search for new customers is supported by the continuous and systematic collection and analysis of external and internal statistical data, organising them into a unified database, and using a decision support model based on it.

This model was developed by joint research with a Hungarian insurance company which has clients mainly from the domestic market. All company data provided by the company, as well as the results obtained using them are sensitive and confidential business information. The authors have not been given permission to publish them. For this, the study focuses on the theoretical framework. The illustrations present a fictitious model setup and simulations for a general insurance company considering different regions, age groups and genders of clients when analysing probabilities of contracting, average sales and profits per contract. Of course, there are other important classification variables that can and must be taken into consideration in a real-life situation. The fewer variables used in this presentation only serve simplicity.

Although the examples of the paper are about an insurance company, the elements of the model can be generalised or adjusted to any sector. Results show that dynamic focus strategies based on model calculations and most current information can outperform static or non-focusing ones in any industries. Moreover, the process from data to decision-making to improve business performance and the decision itself can be easily algorithmised. The feedback of the results into the model carries the potential for automated self-learning and self-correction. The proposed framework can serve as a basis for a self-sustaining artificial business intelligence system.

Insurance companies are increasingly leveraging advanced technologies such as artificial intelligence (AI), machine learning (ML), and Bayesian methods (BM) to enhance their business performance and sales efficiency. These technologies enable more accurate predictions, better customer targeting, and improved decision-making processes. Section 2 presents a literature review on how these techniques are being utilized in the insurance sector.

The theoretical framework proposed by this study will be discussed in Section 3. It is designed for a hypothetical insurance company, however, can be interpreted and applied to any sector. The industry is only important to justify the choice of customer characteristics used in the examples, as well as the internal and external information considered. Because of the limitations described above, the aim of the study is not to show concrete real-life results but present the theoretical framework. It avoids using numerical values wherever possible. Where it is inevitable, like in the simulations of Section 4, all the figures are fictitious. The purely theoretical presentation of the model with some fictive numerical illustrations still allows drawing some conclusions on the potential application of artificial intelligence in Section 5.

2. Literature review

Artificial intelligence (AI) is revolutionizing the insurance and finance sectors, transforming claims, underwriting, distribution, and pricing, while also impacting robo-advisory, fraud prediction, trading strategies, risk assessment, and chatbots (Fung et al., 2021).

AI in insurance companies improves claim clarity, fraud detection, and loss prevention, leading to better service and rates for clients while increasing competitiveness and employee productivity (Lupačov and Stankovic, 2022). Specifically, machine learning and deep learning, can effectively address various challenges in the insurance industry, including reducing uninsured individuals and improving risk management (Singh, 2020).

AI applications in the insurance industry are transforming marketing and business value creation, with a future focus on digital transformation and AI customer lifecycle models (Holland, 2022). According to Zarifis et al. (2019) AI-driven automation in the insurance sector can lead to four emerging business models: allowing others with superior AI and data to take a larger part of the value chain, improving effectiveness, adapting to fully utilize AI, or a technology-focused company offering insurance.

Some of the focus areas are the following.

- Supporting underwriting decisions. AI-based risk intelligence models, such as the AI-Based Risk Intelligence Model (RIM), can improve insurance underwriting by analysing risk factors beyond traditional methods and enabling informed underwriting decisions (Area, 2023). AI-based underwriting can reduce claims by 9% and boost consumer trust, while enhancing underwriting accuracy and transparency in the life insurance industry (Maier et al., 2020).
- Purchase classification and recommendation systems. Graph Convolutional Networks (GCN) perform well for insurance purchase classification and recommendation using AI techniques, outperforming other AI techniques and addressing imbalanced data issues (Chu et al., 2023).
- Automated management of insurance agents' activities. The intelligent platform, integrating business intelligence, decision support system, and KPIs, effectively manages insurance agents' activities, improving decision-making efficiency and predicting future trends (Massaro et al., 2021).
- Interaction design of financial insurance products under the Era of AIoT. The aiCore intelligent dual-recording system significantly improves efficiency,

quality inspection pass rate, and user experience satisfaction in the insurance business by combining AI, data technology, and good design (An et al., 2022).

2.1. Machine learning (ML) in insurance modelling

Particularly, machine learning (ML) models has been widely adopted by insurance companies to predict sales, improve marketing strategies, manage claims, and increase profits more efficiently. Wang (2020) presents an insurance sales prediction model based on deep learning, specifically using an improved long short-term memory (LSTM) network. This model was designed to analyse historical sales data, identify both linear and nonlinear factors affecting sales, and remove redundant data through nondimensionalization and gray correlation calculation. The study concluded that the improved LSTM network outperformed other prediction models, demonstrating that data processing through multiple residual predictions can significantly enhance prediction accuracy.

Hanafy and Ming (2021) study focuses on the application of various ML methods to auto insurance big data. The researchers explored how automotive insurance providers incorporate ML models such as logistic regression, XGBoost, random forest, decision trees, naïve Bayes, and K-NN to predict claim occurrences. The study found that the random forest (RF) model outperformed other methods indicating its effectiveness in handling large datasets, predicting auto insurance claim occurrence, and improving customer service through better data interpretation and comprehension. Similarly, Kasuhal et al. (2023) finds that the Random Forest Regressor (RF) model provides better accuracy and captures more variability in retail sales data than Linear Regressor (LR) and Ridge Regressor (RR).

Velmurugan et al. (2023) confirms that data driven analysis of insurance claims using machine learning algorithms improves efficiency and accuracy of claims processing, while also providing a more convenient and personalized experience for policyholders.

2.2. Bayesian methods (BM) in insurance modelling

Bayesian methods, especially credibility theory, have a long history in actuarial science. Recent advances in computational methodologies have made fully Bayesian treatments more feasible and effective (Makov, 1996).

A number of studies suggest that Bayesian methods (BM) are useful in insurance modelling for handling complex data structures, providing complete predictive distributions, improving computational efficiency, and incorporating various sources of information.

Durante et al. (2015) presents a Bayesian hierarchical model which efficiently clusters agencies based on common monoproduct customer choices and co-subscription networks, informing targeted cross-sell strategies for insurance company customers.

In Qazi et al. (2017) Bayesian networks provide personalized insurance recommendations for customers, outperforming low-rank matrix factorization models in the insurance domain.

Bayesian methods have been increasingly applied in insurance modelling due to their flexibility and ability to incorporate various sources of uncertainty and prior knowledge. The key insights from the research papers on this topic are the following.

- Handling missing data and unsettled claims. Bayesian models can effectively handle missing values in covariates and partial information from open files, reducing sampling bias in insurance claims modelling (Côté et al., 2020).
- Predictive distribution for future claims. Bayesian approaches provide a complete predictive distribution for future claims, which is more informative than point estimates and helps in better decision-making (Hong and Martin, 2016a, 2016b).
- Scalability and efficiency. Distributed computing techniques have made Bayesian methods more efficient and scalable for large-scale insurance data, significantly improving computational performance. Bayesian analysis of big data in insurance predictive modelling using Apache Spark significantly boosts the speed of Bayesian computations, enabling their adoption in large-scaled insurance predictive modelling applications (Zhang, 2017).
- Flexibility in modelling complex distributions. Bayesian nonparametric models, such as Dirichlet process mixtures, are adept at capturing complex features like skewness, heavy tails, and multi-modality in insurance loss data, leading to more accurate predictions (Hong and Martin 2016b; Huang and Meng, 2020).
- Real-time risk monitoring. Bayesian methods facilitate real-time updating of risk measures, which is crucial for dynamic risk management and solvency risk assessment in the insurance industry (Hong and Martin, 2018).
- Incorporation of expert opinions. Bayesian models allow for the coherent incorporation of expert opinions and other sources of information, enhancing the evaluation of accident proneness and other risk factors (Insua et al., 1999).
- Addressing identification issues. Bayesian methods help alleviate identification issues in statistical modelling by incorporating prior knowledge, which is particularly useful in dealing with data limitations and misclassification problems.
- Asymptotic properties and posterior consistency. Bayesian methods have been shown to have desirable asymptotic properties, such as posterior consistency, which ensures the reliability of long-term predictions in insurance applications (Hong and Martin, 2016c).

In conclusion, Bayesian methods are highly useful in insurance modelling due to their ability to handle missing data, provide comprehensive predictive distributions, and incorporate expert opinions. They are particularly advantageous for modelling complex distributions, real-time risk monitoring, and addressing identification issues. Recent computational advancements have further enhanced their scalability and efficiency, making them a valuable tool in the insurance industry.

2.3. Bayesian updating (BU) supporting insurance sales activities

While the literature does not explicitly mention Bayesian updating used for supporting insurance sales activities, it is a statistical method that can be integrated with AI models to continuously improve predictions and decision-making processes. Bayesian updating allows insurance companies to update their beliefs about the probability of certain events as new data becomes available. This method can be particularly useful in dynamically adjusting sales strategies and marketing campaigns based on real-time data. It is proved in general or by many applications in other fields.

2.4. Bayesian updating (BU) in general and in other fields

Giovanis et al. (2017) shows that Artificial Neural Networks (ANNs) can increase computational efficiency in Bayesian updating with subset simulation (SuS) for numerical models by replacing full model evaluations with ANN estimates.

Wang and Shafieezadeh (2020) propose the adaptive Kriging-based method, BUAK, which significantly reduces computational demand and improves Bayesian updating accuracy in computationally intensive models.

In Heskes et al. (2003), JED, a neural-Bayesian system, can optimize single-copy newspaper sales by 1 to 3% with the same total number of deliveries, achieving consistent performance improvements across various newspapers and magazines.

Liu et al. (2019) presents an incremental Bayesian network structure learning algorithm, based on local updating strategy, achieves 98% average accuracy and low constant time in learning big data structures compared to mainstream algorithms.

2.5. Other AI methods in other industries supporting sales activities

Not only BU, but many other AI methods are used widely in other industries. AI-powered solutions in various business areas can improve efficiency, customer service, and decision-making, leading to smoother experiences and increased profits (Abousaber and Abdalla, 2023).

2.5.1. Deep learning models improving sales prediction accuracy

Sales prediction is a critical aspect of business operations, enabling companies to optimize inventory, manage supply chains, and make informed strategic decisions. Recently, deep learning models have been explored for their potential to improve the accuracy of sales forecasts compared to traditional and shallow machine learning techniques. The key insights are the following.

- Improved accuracy with deep learning. Deep learning models, such as LSTM and CNN, have shown superior performance in sales forecasting compared to traditional machine learning models like Random Forest and ARIMA, particularly for non-linear and complex data (Helmini et al., 2019; Kaneko and Yada, 2016; Messaoudi, 2023; Pavlyshenko, 2022; Yin and Tao, 2021). They can handle multi-attribute variables effectively, maintaining high accuracy even with a large number of product categories (Kaneko and Yada, 2016), and complex, non-linear relationships in data, which traditional models struggle with Helmini et al. (2019), Messaud et al. (2023).
- Comparison with shallow techniques. While deep learning models generally perform well, some studies indicate that the performance gains over sophisticated shallow techniques like Random Forest can be modest for certain evaluation metrics (Loureiro et al., 2018; Ma and Fildes, 2021).
- Handling high-dimensional data. Deep learning models are effective in managing and predicting sales with high-dimensional data, maintaining accuracy even when the number of product attributes increases significantly (Kaneko and Yada, 2016).

- Temporal and sequential data and patterns. Models incorporating temporal patterns, such as LSTM and bidirectional LSTM, are particularly effective in capturing the sequential nature of sales data, leading to more accurate forecasts (Helmini et al., 2019; Messaud et al., 2023). Sequential deep learning models, such as those incorporating LSTM and 1D convolution, excel in capturing temporal patterns and complex relationships within sales data, leading to precise sales predictions (Helmini et al., 2019; Messaoudi et al., 2023)
- Meta-learning and ensemble methods. Meta-learning frameworks using deep convolutional neural networks can provide superior forecasting performance by combining multiple base-forecasting methods, though the accuracy gains over some ensemble benchmarks are modest (Ma and Fildes, 2021)
- Trend correction. Incorporating time trend correction in deep learning models can significantly improve the accuracy of forecasts for non-stationary sales data (Pavlyshenko, 2022).
- Comparative performance. While deep learning models generally outperform traditional methods, the extent of improvement can vary. In some cases, the performance gains over advanced machine learning models like Random Forest are not substantial (Eglite et al., 2022; Loureiro et al., 2018).
- Adaptability and generalization. Deep learning models, particularly CNNs, demonstrate strong generalization abilities across different product categories and are adaptable to various types of online products (Yin and Tao, 2021).

Deep learning models generally improve sales prediction accuracy, especially for complex, non-linear, high-dimensional, multi-attribute, and sequential data. They outperform traditional and shallow machine learning models in many cases, though the performance gains can vary depending on the specific metrics and models used for comparison. The ability of deep learning models to handle non-stationary data, non-linear relationships and capture temporal patterns makes them a valuable tool for businesses aiming to optimize their sales forecasting and decision-making processes in diverse retail environments.

2.5.2. Human-AI collaboration in retail logistics

Loske and Klumpp (2021) presents an example of human-AI collaborations in retail logistics. AI-based route planning can lead to efficiency advantages, with sales managers' perspective being more effective than logistics managers' in achieving customer-oriented logistics systems.

2.5.3. Applying AI techniques to predict the success of bank telemarketing

AI techniques, specifically logistic regression, can effectively predict the success of bank telemarketing, offering lower costs, better customer interaction, and faster sales processes (Chen and Chiu, 2020).

2.5.4. AI-based sales forecasting model for digital marketing

AI-based neural networks can accurately predict future sales volumes of products, aiding digital marketers in extracting value from customer feedback (Biswas et al., 2023).

2.5.5. A fuzzy case-based reasoning model for sales forecasting in print circuit board industries

The fuzzy case-based reasoning (FCBR) model effectively forecasts future sales in printed circuit board industries, improving business strategy (Chang et al., 2008).

2.6. Ethical considerations, regulatory challenges, copyright, privacy, and data security concerns

Khaleel et al. (2023) explores ethical implications and concerns surrounding the use of AI and machine learning in the insurance industry, combining insights from professionals and an AI ethics specialist.

Generative AI can improve the insurance industry's marketing efficiency, economic performance, customer satisfaction, and service level, but must address copyright, privacy, data security concerns, ethical considerations, and regulatory challenges (Area, 2023; Lin and Ruan, 2023; Ravi and Vedapradha).

AI-enabled systems in life insurance underwriting increase the risk of unfair discrimination and ethical concerns, necessitating a framework to establish national standards, certification, and periodic audits to ensure compliance (Filabi and Duffy, 2021).

"Intelligent sales" enhances customer care through Big Data and AI, but ethical concerns arise in training data and the use of AI outputs in everyday work practices (Wolf, 2020).

AI ethics must be managed to ensure public trust and maximize the benefits of AI, but there are grounds for optimism due to public awareness, government engagement, and investment in ethical research (Stahl et al., 2023).

AI insurance solicitors and insurance companies specializing in insurance product sales require a revised Financial Consumer Protection Act to ensure consumer protection and the right to receive duty of disclosure (Cho, 2022).

The European Insurance and Occupational Pensions Authority (EIOPA) has developed an AI and ethical framework for the European insurance market, addressing complex ethical issues in AI applications across the value chain (Mullins et al., 2021).

The 'race to AI' has led to a 'race to AI regulation', with countries aiming to ensure trustworthy AI and boost legal certainty, while balancing competition and convergence (Smuha, 2021).

2.7. Literature review conclusion and the research gap

Data science and complex AI algorithms, such as Deep Learning, Adversarial Learning, Federated Learning, Transfer, and Meta Learning, are driving future changes in insurance markets (Śmietanka et al., 2020). Such as Bayesian updating (BU). The core method of the algorithm to be described in the paper resembles the latter both in terms of its aims and principles, even if it is simpler.

Calculation (continuous recalculation) of the current contracting probability per cohort (current client divided by total attempts) and multiplying this by the actual potential customer base (total population minus those attempted) are straightforward statistical operations. Using them to determine prioritisation for future attempts, however, can significantly improve business performance. As this paper will show, a simulation of automated and continuous updating based on endogenous internal and exogenous external information, and the use of this algorithm in advanced integrated AI and BU systems in the insurance sector (or other industries) can have promising results in improving sales efficiency.

AI models, such as the improved LSTM network and random forest, have demonstrated their capability to accurately predict sales and manage claims. Although its potential to enhance decision-making processes through continuous data integration is evident, applications of Bayesian updating has not been explicitly addressed in the reviewed insurance studies. This paper tries to fill this gap and outline some preliminaries to show the benefits of the method. Future research presenting results with real-world data and using complex models could further explore the combined use of AI and Bayesian updating to optimize sales strategies and customer targeting in the insurance sector.

3. Methodology and model framework

The presentation of our proposed probabilistic microeconomic model of the insurance business follows a matrix-algebraic scheme. **Table 1** shows the general structure of matrices to describe or calculate the elements (data and results) of the framework. Assuming life insurance (with different type of policies) as the product (with variants) under investigation, the most objective and the most important features of the clients include the gender (male/female), the age group (cohort 1, …, k), and the region (county, district, city, etc.) (region 1, …, m) where they live. These data are available for the entire population of Hungary (which can be assumed as the target market of the company) in the Dissemination Database of the Hungarian Central Statistical Office (HCSO, 2024). (The dimensions of the matrices can be increased or changed according to the number of essential features which should be taken into consideration in the industry and company under investigation.)

Table 1. The general structure of matrices used in the model.

	Male				**Female**			
	Cohort 1	**Cohort 2**	**…**	**Cohort k**	**Cohort 1**	**Cohort 2**	**…**	**Cohort *k***
Region 1								
Region 2								
…								
Region *m*								

3.1. Current customers

The most obvious source of data is the records of the company's current customers who can be organised into a matrix described in **Table 1**. This matrix is denoted by **A**, whose general element, $a_{i,j,g}$ is the number of clients from the ith region, jth age group, and of gth gender, where $i = 1, 2, ..., m$, $j = 1, 2, ..., k$ and $g = 1, 2$ (male or female).

3.2. Sales efforts

Matrix **B**, with a general element of $b_{i,j,g}$, expresses the sales efforts to acquire customers in region i, age group j, and gender group g. The elements of **B** can be interpreted in several ways: (a) the monetary value of sales expenses (agency commissions, targeted sales campaigns, costs of telephone calls, online campaigns, other acquisition costs); (b) in the case of the traditional agent channel, they can be described by the number of customer inquiries (client-agent meetings): if a potential client is contacted several times, he or she will appear in the database and be counted several times; or (c) it refer to the number of clients contacted: in this case, a potential customer will appear and be counted once in the database irrespective of the number of contacts.

3.3. Sales efficiency or probability of a new contract

Assuming matrix **B** is available (or producible based on existing company data collections and records), the element-by-element (Hadamard) quotients of **A** and **B** ($\mathbf{A} \oslash \mathbf{B}$) can be defined as follows (according to the cases (a)–(c) in the previous subsection): (a) customer acquisition efficiency indicating how many customers were acquired with one unit of sales expenditures in each region, age group, and gender; (b) the probability of underwriting a policy (accept an offer and contract) per agent-client meeting; (c) the probability of contracting per client. Interesting and commercially useful model variants may be built for each approach, however, only the third (c) will be discussed in this paper. Applying probabilities of signing a contract on the matrix of potential customers, the expected number of new clients can be obtained.

3.4. Potential customers

Let **C** be the matrix of the number of potential or targeted customers. When the "targeted market" is the total population of Hungary, i.e., in the case of "no targeting", **C** can be produced using the estimated population data of the HCSO or the actual territorial data from the National Regional Development and Spatial Planning Information System (TeIR, 2024). For a more targeted sales campaign, however, **C** may also be the matrix of targeted customers who can be reached through a specific or a preferred (selected) sales channel. (In this case, however, matrices **A** and **B** must also refer and be restricted to this sales channel only.)

When analysing a specific insurance policy and assuming that existing clients cannot be new clients again (at least for the same product), one must define a reduced matrix of potential clients as the difference between matrices **C** and **A**, **C** – **A**. Alternatively, **C – B** reflects another interpretation when both existing and potential clients who have previously accepted or rejected an offer will no longer be revisited. Based on the above facts, the company can direct its future sales efforts to the elements of **C** – **A** or **C** – **B** matrices.

In the absence of **B**, only the **C – A** approach of the reduced matrix of potential clients can be followed together with the Hadamard quotient of **A** and **C** ($\mathbf{A} \oslash \mathbf{C}$) for the relative frequency of the case that a Hungarian person from a given region, age group, and gender is the client of the company.

3.5. Focused sales strategies

By considering the elements of the $\mathbf{A} \oslash \mathbf{B}$ (or $\mathbf{A} \oslash \mathbf{C}$) and $\mathbf{C} - \mathbf{B}$ (or $\mathbf{C} - \mathbf{A}$) matrices as horizontal and vertical coordinates, respectively, one can create scatter charts and portfolio segmentation diagrams like in **Figure 1**. To increase its business performance, the insurance company should focus on the regions, age groups, and gender where both the probability of contracting and the number of potential clients are high enough or higher than the average (the upper-right quadrant). This helps with choosing the target areas for future sales actions.

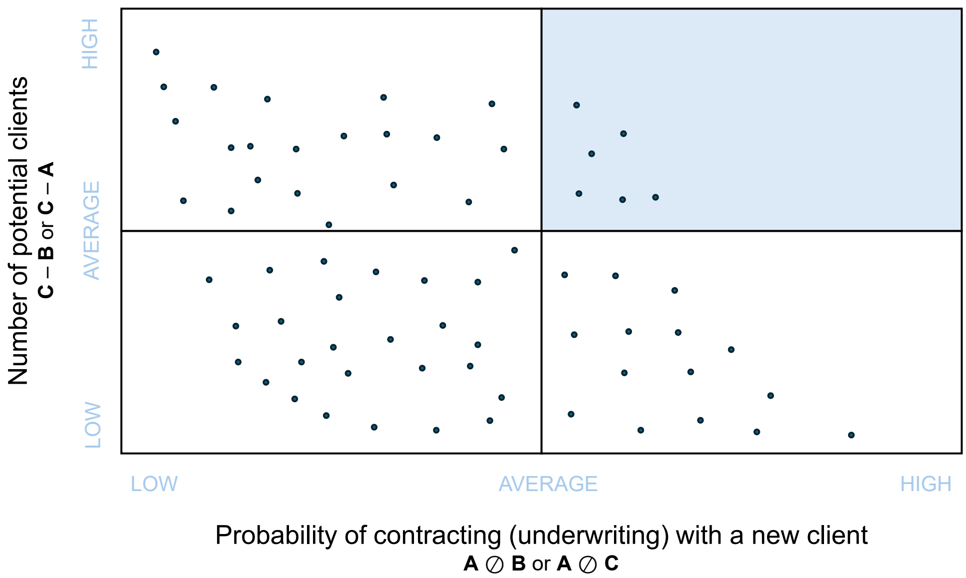

Figure 1. Portfolio diagram for focused sales strategies.

3.6. Expected number of new customers

The information represented by the two axes of the coordinate system in **Figure 1** can be compressed into a single indicator (reducing the number of the considered dimensions significantly).

By forming the matrix:

$$(\mathbf{C} - \mathbf{B}) \odot (\mathbf{A} \oslash \mathbf{B}) \text{ (or } (\mathbf{C} - \mathbf{A}) \odot (\mathbf{A} \oslash \mathbf{C})), \tag{1}$$

where $\odot$ denotes the element-by-element (Hadamard) multiplication, the expected number of new clients (new contracts) per region, gender, and age group can be determined. The elements of the matrix arranged in descending order can drive the sales strategy.

To compare different types of policies (or products) when developing and following efficient sales strategy, one can also calculate the matrices of the expected number of new customers by policy

$$\begin{gathered}(\mathbf{C}_1 - \mathbf{B}_1) \odot (\mathbf{A}_1 \oslash \mathbf{B}_1),\\ (\mathbf{C}_2 - \mathbf{B}_2) \odot (\mathbf{A}_2 \oslash \mathbf{B}_2),\\ \ldots,\\ (\mathbf{C}_p - \mathbf{B}_p) \odot (\mathrm{A}_p \oslash \mathbf{B}_p)\end{gathered} \tag{2}$$

$$\begin{gathered}(\text{or } (\mathbf{C}_1 - \mathbf{A}_1) \odot (\mathbf{A}_1 \oslash \mathbf{C}_1),\\ (\mathbf{C}_2 - \mathbf{A}_2) \odot (\mathbf{A}_2 \oslash \mathbf{C}_2),\\ \ldots,\\ (\mathbf{C}_p - \mathbf{A}_p) \odot (\mathbf{A}_p \oslash \mathbf{C}_p)),\end{gathered} \tag{3}$$

where p is the total number of compared policy types (or products). To maximise the number of new clients in a region, cohort, and gender, one should offer the product with the highest value in the related cell of the expected new customers matrix.

3.7. Extensions: Expected sales and profits on new customers

The models discussed here are primarily sales-focused, supporting an approach to customer acquisition activities based on maximising the expected number of customers. Matrix operations $(\mathbf{C} - \mathbf{B}) \odot (\mathbf{A} \oslash \mathbf{B})$ (or $(\mathbf{C} - \mathbf{A}) \odot (\mathbf{A} \oslash \mathbf{C})$), however, can be augmented in several ways to support decisions that have a broader view of the business. To do this, the following matrices can be defined:

D is the matrix of annual premiums, a general element of which shows the average annual premiums written on a contract for a given region, age group, and gender on a contract; and

E is the matrix of net profits, a general element of which shows the average net insurance profits on a contract for an insured person of a given region, age group, and gender.

With these additions the company can take another significant step towards an effective decision-making tool that provides support for the development of a market strategy to increase annual premium income and the resulting market share using the matrices $(\mathbf{C} - \mathbf{B}) \odot (\mathbf{A} \oslash \mathbf{B}) \odot \mathbf{D}$ (or $(\mathbf{C} - \mathbf{A}) \odot (\mathbf{A} \oslash \mathbf{C}) \odot \mathbf{D}$); or to increase insurance profits by using matrices $(\mathbf{C} - \mathbf{B}) \odot (\mathbf{A} \oslash \mathbf{B}) \odot \mathbf{F}$ (or $(\mathbf{C} - \mathbf{A}) \odot (\mathbf{A} \oslash \mathbf{C}) \odot \mathbf{F}$) when prioritising between offered products and potential new clients from different regions, age groups, genders, and other key features.

3.8. Expressing uncertainty with interval estimates

The above matrix equations are probability models; therefore, in addition to the expected (average) values, it is worth paying attention to uncertainty. For example, the elements of matrix $\mathbf{A} \oslash \mathbf{B}$, on the one hand, are not exact probabilities, but relative frequencies, i.e., estimates of probabilities. On the other hand, values determined on a probability basis naturally carry uncertainty (to put it more precisely: not only the expected value, but also the standard deviation can be important).

The element-by-element quotient **A** ⊘ **B** approximates the willingness to contract or the probability of contracting. The generally accepted underlying assumption is that this probability can be approximated by relative frequency. The relative frequency is the ratio 'number of cases/total number of cases'. The number of cases in the insurance example is the number of current clients or contracts, these are the elements of matrix **A**. To determine the elements of matrix **B**, the actual number of requests and sales attempts are needed. Based on these, an interval estimate of the contracting probability can be given, in which one may even consider the size of the population (the number of inhabitants of a given settlement, district, county, region, etc., or the number of inhabitants meeting certain conditions within it). The interval estimate for the probability (with a given confidence level) is

$$\hat{p} \pm \beta \sqrt{\frac{\hat{p}(1-\hat{p})}{n} \frac{N-n}{N-1}} \tag{4}$$

where $\hat{p}$ denotes the calculated relative frequency, n is the sample size, N is the population size, and β is the factor resulting from the percentage size of the confidence interval. Using the matrix algebraic notations described above, the elements of the **A** ⊘ **B** matrix are the relative frequencies $\hat{p}$; the elements of matrix **B** give the number of experiments they correspond to n in the formula; and the elements of matrix **C** give the population size, N.

For example, if in a city with N = 15,000 inhabitants n = 1000 different agent-client meetings took place, and a = 180 contracts were concluded then the probability estimate (point estimate, single-value estimate) is $\hat{p}$ = 180/1000 = 0.18. As a result, 95% confidence interval estimate is obtained for the probability of contracting, which includes the uncertainty inherent in the probability model (and reality). This can be expressed as 0.18 ± 0.023 or with the [0.157; 0.203] interval.

Using the interval estimations for the contracting probabilities, interval estimations (from-to) can also be obtained for the values calculated from it (e.g., expected number of new clients), which characterises the uncertainty much better than a single number.

4. Applications and simulations

In the following, we provide the step-by-step calculations and results of the model on simulated data. Of course, it would be better to use real-life data sets from an existing insurance company (e.g., our research partner), but due to their privacy policy and data protection rules, real-life data are considered confidential. Thus, we were allowed to use them only for experimenting, but without publication.

The consistent application of the model implements a combination of human and machine learning. It focuses our attention and resources on customers with the highest business potential and identifies them based on client features used for group formation. The results of previous decisions can be easily verified, and the information set for subsequent decisions can be updated after each sales action. Based on more up-to-date information, the system is constantly learning. By immediately adding the results of any offer (client-agent contact) to matrices **A** and **B** (in the case of success), or to only **B** (in the case of rejection) (as shown in **Figure 2**), the correction will be realised instantly, and the contracting probabilities that form the basis of future decisions will change accordingly. This kind of immediate or slightly delayed updating process is widely used in probability-based forecasting, usually called Bayesian updating (Bai et al., 2022; Mosallam et al., 2013; Vairo et al., 2022).

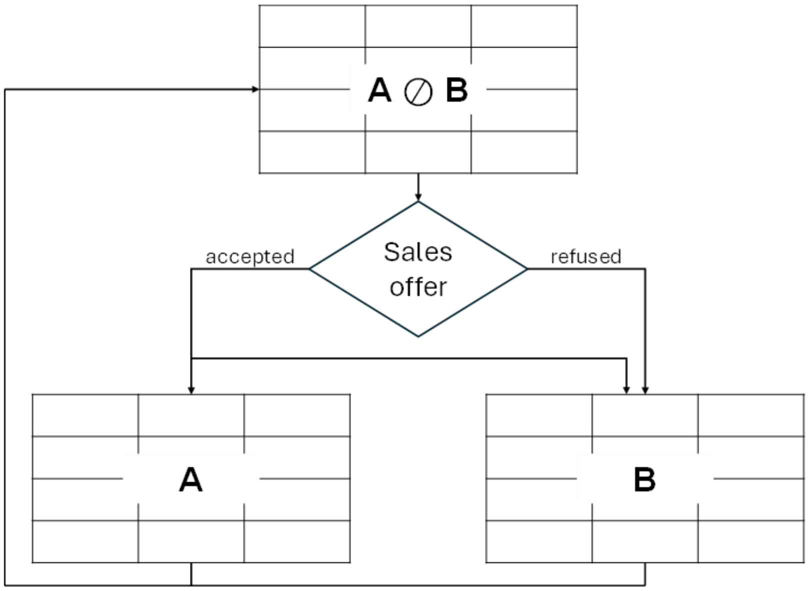

Figure 2. Learning with the evolution of matrices A and B during sales activities (client-agent contacts).

Model simulations highlight that dynamic focusing strategies based on the most current information always outperform static or non-focusing ones. To illustrate this, consider the following fictive example with initial matrices of **Table 2**. For the sake of simplicity, the gender dimension was omitted. Green, yellow, and red backgrounds indicate the first, second, and third highest (top3) element in each matrix, respectively.

Table 2. Initial matrices used in the simulation exercise.

	A: matrix of current clients			**B: matrix of sales attempts**			**A / B: matrix of contracting probabilities**		
	Cohort 1	**Cohort 2**	**Total**	**Cohort 1**	**Cohort 2**	**Total**	**Cohort 1**	**Cohort 2**	**Total**
Region 1	260	190	450	2550	2200	4750	10.20%	8.64%	9.5%
Region 2	128	164	292	1500	1330	2830	8.53%	12.33%	10.3%
Region 3	120	138	258	1300	1120	2420	9.23%	12.32%	10.7%
Total	508	492	1000	5350	4650	10,000	9.5%	10.6%	10.0%
	C: matrix of total population (target market)			**C - B: matrix of potential customers**			**(C - B) x (A / B): expected new customers**		
	Cohort 1	**Cohort 2**	**Total**	**Cohort 1**	**Cohort 2**	**Total**	**Cohort 1**	**Cohort 2**	**Total**
Region 1	23,000	27,000	50,000	20,450	24,800	45,250	2085	2142	4227
Region 2	29,000	24,600	53,600	27,500	23,270	50,770	2347	2869	5216
Region 3	22,000	24,400	46,400	20,700	23,280	43,980	1911	2868	4779
Total	74,000	76,000	150,000	68,650	71,350	140,000	6343	7880	14,222

Based on $\mathbf{A} \oslash \mathbf{B}$ the probabilities of contracting, the insurance company should focus its sales efforts to R(egion)2-C(ohort)2 firstly, and to R3-C2 and R1-C1, secondly and thirdly. In the example, $(\mathbf{C} - \mathbf{B}) \odot (\mathbf{A} \oslash \mathbf{B})$ matrix for the expected number of new customers gives almost the same order. Only in the third place there is a change. A smaller group's (see R1-C1, the green element of $\mathbf{C} - \mathbf{B}$, with 20,450 potential customers) higher success rate (10.2) is overshadowed by a larger group's (R2-C1 with 27,500 potential customers) lower rate (8.53%), which is also reflected in aggregated data (the contracting probabilities in total row and column). This is an example of the Simpson's paradox (Blyth, 1972; Good and Mittal, 1987; Simpson, 1951) which highlights that calculating probabilities across different regions and age groups might lead to uneven data representation due to population grouping and aggregation. This reminds us of the importance of careful preparatory statistical analysis, consideration of stratification, grouping, confounding variables and casual relationships, and the potential pitfalls of simplistic interpretations of data. Nevertheless, in our example, decisions are made not on the basis of the aggregates (row and column totals), but on the basis of R-C pairs (inner cells of the table) reflecting superior information. Although the group sizes are different, the variations are not too large. Thus, Simpson's effect on the comparison of percentages has a moderate influence. Additionally, when the targeting decision is based on the expected number of new clients, not only the proportions, but the group sizes are also matter.

The initial data also reveals that there is only a slight difference between R2-C2 and R3-C2 in the first and second place, respectively, which can result in a number of shifts (flip-flops) in the first please (top1).

The left diagram in **Figure 3** shows the results of model simulations for 100 consecutive client-agent meetings under different assumptions. No focusing means a random choice from the current potential customers. Dynamic focusing on Top3 and Top1 chooses the potential client for the next meeting from the actual top3 and top1 region(s) and cohort(s) based on current contracting probabilities. Due to the stochastic evolutionary process and the closeness of R2-C2 and R3-C2 contracting probabilities, top1 changes several times during the process. This is not the case in static focusing on Top1 which insists on the initial top1 region and cohort (R2-C2 in our example, see the green element of the top right $\mathbf{A} \oslash \mathbf{B}$ matrix in Table group 2. Note that the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A} \oslash \mathbf{B}$, $\mathbf{C} - \mathbf{B}$, Top3, and Top1 cells evolve endogenously during the sales operation sequence (except initial Top1).

Figure 3. Results of one-series simulations for 100 consecutive sales activities (client-agent contacts) (left) and 50,000-series Monte-Carlo simulations (right).

This series of experiments with a small number of attempts clearly shows the differences between different strategies. According to the single-series simulation results in the left diagram of **Figure 3**, without any focusing only 9 new clients could be gained, whereas by focusing on the varying top3 and top1, the number of new clients reached 14 and 17, respectively. By insisting on the initially best region-cohort, the number of new clients decreased again (to 15).

For a set of results more robust than single-series simulations on the left of **Figure 3**, 50,000-series Monte-Carlo simulations have been performed. (The simulations for each different strategies have been executed 50,000 times.) The average number of new clients for the four cases above (10.23, 11.69, 12.40, and 12.33, respectively) reflects that focusing strategies are significantly better than no focusing and the expected value of new customers follow the same order: Dynamic focusing on Top1 outperforms all others (see the box-plots on the right side of **Figure 3**).

5. Conclusions and further research

Despite the study's focus on new customers, with a consecutive application of the model framework and collection of data according to the matrices defined and used above, one can ensure the gradual transformation of the existing client portfolio in the directions that suits the company's goals and enhances its business performance. By continuous updating and regular monitoring of these matrices, the firm can shape its client portfolio not only by chasing the most profitable options but also reducing less beneficial businesses or repricing contracts. It may help set directions for product development, as well.

By updating not only **A** and **B**, but also **D** and **E** matrices with data from contracted new customers, the average revenue and profit figures for the related customer group will be modified, which again influence and correct future decisions: if a group's revenue and profitability figures improve or deteriorate compared to the level on which the decision was based, this will be reflected in **D** and **E** matrices and expressed in the following decisions based on them. Of course, not only **A**, **B**, **D**, and **E**, but also **C** changes and should be updated based on new external data.

Consequently, the presented business model on acquiring new customers, and the process of learning at the heart of it, requires a continuous and systematic collection of external and internal statistical and business data on current and potential clients, and the immediate recording of the results of each sales attempt into a unified database. Unfortunately, the lack of such database still prevents many companies from adopting the model. However, by overcoming this obstacle, the benefits can pay off the related investment costs. Fortunately, the digitalisation of the insurance industry makes the data-related tasks easier (Eckert and Osterrieder, 2020).

The simulation results showed that focused sales strategies can enhance business performance significantly. Although the example presented in the paper concerned an insurance company, the structure and the elements of the model are universal and can be generalised or rephrased for any sector.

The model simulations illustrated that the process from data to decision-making and even the decision itself can be easily algorithmised, and the feedback of the results into the model carries the potential for automated self-learning and self-correction.

Thus, the proposed framework can also serve as a basis for a self-sustaining artificial business intelligence system (AI). Developing the ways of practical implementation of such an AI system may be a subject of future research.

Author contributions: Conceptualization, KK; methodology, KK; validation, KK and IÁH; formal analysis, KK; investigation, KK; data curation, KK; writing—original draft preparation, KK; writing Subsection 3.8, IÁH; writing—review and editing, KK; supervision, KK, IÁH, DF and NK; project administration, NK; funding acquisition, NK. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This research was supported by the project "Development of customisable insurance products with the help of artificial intelligence 2020-1.1.2-PIACI-KFI-2021-00267". The authors thank to Katalin Kiss for proofreading the manuscript.

Conflict of interest: The authors declare no conflict of interest.

References

Abousaber, I., & Abdalla, H. (2023). Review of Using Technologies of Artificial Intelligence in Companies. International Journal of Communication Networks and Information Security (IJCNIS). https://doi.org/10.17762/ijcnis.v15i1.5743.

An, W., Wang, S., Chen, Y., Wu, Q., Wu, C., Wang, K., Wang, S., Chen, H., & Chen, Z. (2022). Interaction design of financial insurance products under the Era of AIoT. Applied Mathematics and Nonlinear Sciences, 7, 745 - 756. https://doi.org/10.2478/amns.2021.2.00162.

Area, D. (2023). Impact of AI in the General Insurance underwriting factors. Central European Management Journal. https://doi.org/10.57030/23364890.cemj.31.2.72.

Bai, Y., Xing, Y., Wu, J. (2022). Integrating knowledge graph, complex network and Bayesian network for data-driven risk assessment. Chemical Engineering Transactions, 90, 31-36.

Biswas, B., Sanyal, M., & Mukherjee, T. (2023). AI-Based Sales Forecasting Model for Digital Marketing. Int. J. E Bus. Res., 19, 1-14. https://doi.org/10.4018/ijebr.317888.

Blyth, C. R. (1972). On Simpson's Paradox and the Sure-Thing Principle. Journal of the American Statistical Association. 67 (338): 364–366. doi:10.2307/2284382. JSTOR 2284382.

Chang, P., Liu, C., & Lai, R. (2008). A fuzzy case-based reasoning model for sales forecasting in print circuit board industries. Expert Syst. Appl., 34, 2049-2058. https://doi.org/10.1016/j.eswa.2007.02.011.

Chen, K., & Chiu, H. (2020). Applying AI Techniques to Predict the Success of Bank Telemarketing. Proceedings of the 2020 4th International Conference on Deep Learning Technologies. https://doi.org/10.1145/3417188.3417198.

Cho, K. (2022). A Study on new Insurance Distribution Channel's Right to Receive the Duty of Disclosure and Legal Issues: Focusing on AI (Artificial Intelligence) Insurance Solicitors and Insurance Companies Specializing in Insurance Product Sales. Korean Insurance Law Association. https://doi.org/10.36248/kdps.2022.16.2.003.

Chu, J., Than, J., Thon, P., & Jo, H. (2023). Investigation on Insurance Purchase Classification for Insurance Recommendation Using Deep Learning and Class Propagation. 2023 IEEE 8th International Conference on Software Engineering and Computer Systems (ICSECS), 1-6. https://doi.org/10.1109/ICSECS58457.2023.10256401.

Côté, M., Genest, C., & Stephens, D. (2020). A Bayesian Approach to Modeling Multivariate Multilevel Insurance Claims in the Presence of Unsettled Claims. Bayesian Analysis. https://doi.org/10.1214/20-ba1243.

Durante, D., Paganin, S., Scarpa, B., & Dunson, D. (2015). Bayesian modelling of networks in complex business intelligence problems. Journal of the Royal Statistical Society: Series C (Applied Statistics), 66. https://doi.org/10.1111/rssc.12168.

Eckert, C., Osterrieder, K. How digitalization affects insurance companies: overview and use cases of digital technologies. ZVersWiss 109, 333–360 (2020). https://doi.org/10.1007/s12297-020-00475-9

Eglite, L., & Birzniece, I. (2022). Retail Sales Forecasting Using Deep Learning: Systematic Literature Review. Complex Syst. Informatics Model. Q., 30, 53-62. https://doi.org/10.7250/csimq.2022-30.03.

Filabi, A., & Duffy, S. (2021). AI-Enabled Underwriting Brings New Challenges for Life Insurance: Policy and Regulatory Considerations. Journal of Insurance Regulation. https://doi.org/10.52227/25114.2021.

Fung, G., Polanía, L., Choi, S., Wu, V., & Ma, L. (2021). Editorial: Artificial Intelligence in Insurance and Finance, Front. Appl. Math. Stat., Sec. Mathematical Finance Volume 7 - 2021 https://doi.org/10.3389/fams.2021.795207.

Giovanis, D., Papaioannou, I., Štraub, D., & Papadopoulos, V. (2017). Bayesian updating with subset simulation using artificial neural networks. Computer Methods in Applied Mechanics and Engineering, 319, 124-145. https://doi.org/10.1016/J.CMA.2017.02.025.

Good, I. J., Mittal, Y. (1987). The Amalgamation and Geometry of Two-by-Two Contingency Tables. The Annals of Statistics. 15 (2): 694–711. doi:10.1214/aos/1176350369. ISSN 0090-5364. JSTOR 2241334.

Hanafy, M., & Ming, R. (2021). Machine Learning Approaches for Auto Insurance Big Data. Risks. https://doi.org/10.3390/RISKS9020042.

Helmini, S., Jihan, N., Jayasinghe, M., & Perera, S. (2019). Sales forecasting using multivariate long short term memory network models. PeerJ Prepr., 7, e27712. https://doi.org/10.7287/peerj.preprints.27712v1.

Heskes, T., Spanjers, J., Bakker, B., & Wiegerinck, W. (2003). Optimising newspaper sales using neural-Bayesian technology. Neural Computing & Applications, 12, 212-219. https://doi.org/10.1007/s00521-003-0384-x.

Holland, C. (2022). Artificial Intelligence (AI) and Digital Transformation in the Insurance Market: A Case Study Analysis of BGL Group. Proceedings of the 55th Hawaii International Conference on System Sciences, 1-10. https://doi.org/10.24251/hicss.2022.553.

Hong, L., & Martin, R. (2016a). A Flexible Bayesian Nonparametric Model for Predicting Future Insurance Claims. North American Actuarial Journal, 21, 228 - 241. https://doi.org/10.1080/10920277.2016.1247720.

Hong, L., & Martin, R. (2016b). A Flexible Bayesian Nonparametric Model for Predicting Future Insurance Claims. ERN: Bayesian Analysis (Topic). https://doi.org/10.2139/ssrn.2690843.

Hong, L., & Martin, R. (2016c). A review of Bayesian asymptotics in general insurance applications. European Actuarial Journal, 7, 231-255. https://doi.org/10.2139/ssrn.2741074.

Hong, L., & Martin, R. (2018). Real-time Bayesian non-parametric prediction of solvency risk. Annals of Actuarial Science, 13, 67 - 79. https://doi.org/10.1017/S1748499518000039.

Huang, Y., & Meng, S. (2020). A Bayesian nonparametric model and its application in insurance loss prediction. Insurance Mathematics & Economics, 93, 84-94. https://doi.org/10.1016/j.insmatheco.2020.04.010.

Hungarian Central Statistical Office (HCSO) (2024). Dissemination database. Population and vital events. https://statinfo.ksh.hu/, accessed 01.03.2024.

Insua, S., Martín, J., Insua, D., & Ruggeri, F. (1999). Bayesian Forecasting for Accident Proneness Evaluation. Scandinavian Actuarial Journal, 1999, 134-156. https://doi.org/10.1080/03461239950132624.

Kaneko, Y., & Yada, K. (2016). A Deep Learning Approach for the Prediction of Retail Store Sales. 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), 531-537. https://doi.org/10.1109/ICDMW.2016.0082.

Kaushal, R., Giri, K., Raina, K., Choudhary, D., Naikade, K., & Kartikeya, M. (2023). AI-Based Approach for Retail Sale Forecasting. 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), 111-116. https://doi.org/10.1109/ICSSAS57918.2023.10331651.

Lin, X., & Ruan, W. (2023). Research on the Marketing Transformation of Insurance Industry Under Generative Artificial Intelligence Technology. Proceedings of the 2nd International Conference on Public Management, Digital Economy and Internet Technology, ICPDI 2023, September 1–3, 2023, Chongqing, China. https://doi.org/10.4108/eai.1-9-2023.2338781.

Liu, D., Wu, J., Liang, L., Li, X., Luo, C., & Xu, G. (2019). An incremental Bayesian network structure learning algorithm based on local updating strategy. 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 1, 673-677. https://doi.org/10.1109/IAEAC47372.2019.8997724.

Loske, D., & Klumpp, M. (2021). Human-AI collaboration in route planning: An empirical efficiency-based analysis in retail logistics. International Journal of Production Economics, 241, 108236. https://doi.org/10.1016/J.IJPE.2021.108236.

Loureiro, A., Miguéis, V., & Silva, L. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. Decis. Support Syst., 114, 81-93. https://doi.org/10.1016/j.dss.2018.08.010.

Lupačov, J., & Stankovic, Z. (2022). Artificial intelligence in insurance companies. JTTTP - journal of traffic and transport theory and practice. https://doi.org/10.7251/jtttp2202077l.

Ma, S., & Fildes, R. (2021). Retail sales forecasting with meta-learning. Eur. J. Oper. Res., 288, 111-128. https://doi.org/10.1016/j.ejor.2020.05.038.

Maier, M., Carlotto, H., Saperstein, S., Sanchez, F., Balogun, S., & Merritt, S. (2020). Improving the Accuracy and Transparency of Underwriting with AI to Transform the Life Insurance Industry. AI Mag., 41, 78-93. https://doi.org/10.1609/aimag.v41i3.5320.

Makov, U., Smith, A., & Liu, Y. (1996). Bayesian methods in actuarial science. The Statistician, 45, 503-515. https://doi.org/10.2307/2988548.

Massaro, A., Panarese, A., Gargaro, M., Vitale, C., & Galiano, A. (2021). Implementation of a Decision Support System and Business Intelligence Algorithms for the Automated Management of Insurance Agents Activities. International Journal of Artificial Intelligence & Applications, 12, 01-13. https://doi.org/10.5121/IJAIA.2021.12301.

Messaoudi, F., Loukili, M., & Ghazi, M. (2023). Demand Prediction Using Sequential Deep Learning Model. 2023 International Conference on Information Technology (ICIT), 577-582. https://doi.org/10.1109/ICIT58056.2023.10225930.

Mosallam, A., Medjaher, K., Zerhouni, N. (2013). Bayesian approach for remaining useful life prediction. Chemical Engineering Transactions, 33, 139-144.

Mullins, M., Holland, C., & Cunneen, M. (2021). Creating ethics guidelines for artificial intelligence and big data analytics customers: The case of the consumer European insurance market. Patterns, Volume 2, ISSUE 10. https://doi.org/10.1016/j.patter.2021.100362.

National Regional Development and Spatial Planning Information System (TeIR) (2024). Population (31 December). https://www.oeny.hu/oeny/teir/#/tablo/5, accessed 01.03.2024.

Pavlyshenko, B. (2022). Forecasting of Non-Stationary Sales Time Series Using Deep Learning. ArXiv, abs/2205.11636. https://doi.org/10.48550/arXiv.2205.11636.

Qadadeh W., Abdallah S. (2018). Customers segmentation in the insurance company (TIC) dataset. Procedia computer science, 144, 277-290. https://doi.org/10.1016/j.procs.2018.10.529

Qazi, M., Fung, G., Meissner, K., & Fontes, E. (2017). An Insurance Recommendation System Using Bayesian Networks. Proceedings of the Eleventh ACM Conference on Recommender Systems. https://doi.org/10.1145/3109859.3109907.

Ravi H., Vedapradha R. (2023). Artificial intelligence service agents: a silver lining in rural India. Kybernetes. https://doi.org/10.1108/k-09-2022-1239.

Simpson, E. H. (1951). The Interpretation of Interaction in Contingency Tables. Journal of the Royal Statistical Society, Series B. 13 (2): 238–241. doi:10.1111/j.2517-6161.1951.tb00088.x.

Singh, S. (2020). A Commentary on the Application of Artificial Intelligence in the Insurance Industry. Trends in Artificial Intelligence. Volume 4 Issue 1 https://doi.org/10.36959/643/305.

Śmietanka, M., Koshiyama, A., & Treleaven, P. (2020). Algorithms in future insurance markets. International Journal of Data Science and Big Data Analytics. https://doi.org/10.2139/ssrn.3641518.

Smuha, N. (2021). From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence. Law, Innovation and Technology, 13, 57 - 84. https://doi.org/10.1080/17579961.2021.1898300.

Stahl, B., Schroeder, D., & Rodrigues, R. (2023). Ethics of Artificial Intelligence. Case Studies and Options for Addressing Ethical Challenges. SpringerBriefs in Research and Innovation Governance. https://doi.org/10.1007/978-3-031-17040-9.

Vairo, T., Bragatto, P., Milazzo, M. F., Pettinato, M., & Fabiano, B. (2022). DYN-RISK–Design and Development of a Dynamic Risk Assessment Tool. Chemical Engineering Transactions, 90, 325-330.

Velmurugan, K., Pazhanivel, K. Divyasree, R., Gowtham, E., & Guruharan, S. (2023). Data Driven Analysis of Insurance Claims Using Machine Learning Algorithm. International Journal of Advanced Research in Science, Communication and Technology. Vol. 3, Issue 1, May 2023 https://doi.org/10.48175/ijarsct-9689.

Vij A., Preethi N. (2021). Approaches Towards a Recommendation Engine for Life Insurance Products. 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2021, pp. 1-5, doi: 10.1109/ICMNWC52512.2021.9688436.

Wang, H. (2020). An Insurance Sales Prediction Model Based on Deep Learning. Rev. d'Intelligence Artif., 34, 315-321. https://doi.org/10.18280/ria.340309.

Wang, Z., & Shafieezadeh, A. (2020). Highly efficient Bayesian updating using metamodels: An adaptive Kriging-based approach. Structural Safety, 84, 101915. https://doi.org/10.1016/j.strusafe.2019.101915.

Wolf, C. (2020). AI Ethics and Customer Care: Some Considerations from the Case of "Intelligent Sales". Proceedings of 18th European Conference on Computer-Supported Cooperative Work. https://doi.org/10.18420/ecscw2020_n02.

Yin, X., & Tao, X. (2021). Prediction of Merchandise Sales on E-Commerce Platforms Based on Data Mining and Deep Learning. Sci. Program., 2021, 2179692:1-2179692:9. https://doi.org/10.1155/2021/2179692.

Zarifis, A., Holland, C., & Milne, A. (2019). Evaluating the impact of AI on insurance: The four emerging AI- and data-driven business models. Emerald Open Research. Volume 1 Issue 1 https://doi.org/10.35241/emeraldopenres.13249.1.

Zhang, Y. (2017). Bayesian Analysis of Big Data in Insurance Predictive Modeling Using Distributed Computing. ERN: Bayesian Analysis (Topic). https://doi.org/10.1017/ASB.2017.15