A Machine-Learning-Based Business Analytical System for Insurance Customer Relationship Management and Cross-Selling

Xiaoguang Tian
Purdue University Fort Wayne

Jun Todorovic
Purdue University Fort Wayne

Zelimir Todorovic
Purdue University Fort Wayne

Effective cross-selling practices are integral to maintaining strong customer relationships and optimizing business processes within the insurance industry. This study presents a novel three-stage Machine Learning-Based System (MLBS) designed to enhance the identification of potential insurance customers and improve customer relationship management. This study proposes combining under sampling strategies and an ensemble approach to improve prediction performance. The proposed MLBS method involves selecting the best training sample using artificial neural networks and employing the stacking ensemble approach. It yields superior prediction results, exhibiting the highest recall, precision, and Area Under the Curve (AUC). These advancements substantially bolster the efficiency of cross-selling strategies. This research pioneers the application of stacking ensemble learning within the cross-selling domain, representing a novel contribution to the business field. The outcomes underscore the superiority of the MLBS system over baseline models across multiple performance metrics, thereby significantly enhancing support for cross-selling campaigns in various businesses.

Keywords: digital transformation, insurance cross-selling, customer relationship management, machine learning, artificial neural networks

INTRODUCTION

The innovative approach to marketing management has been associated with increased organizational performance and profitability. Marketing management is also a critical value-adding process in the business value chain of many organizations (Osarenkhoe & Bennani, 2007). As a subdomain of marketing management, cross-selling is commonly used by the financial service industry to increase sales and profitability. Although cross-selling strategy focuses on selling peripheral and more expensive services to existing customers (Vyas & Math 2006), it must be noted that not all cross-selling efforts are worth the time and monetary investments made in them (Rosen, 2004). For this reason, increasing the effectiveness of cross-selling activities while utilizing fewer resources is vital. As a result, many organizations today are attempting to be more proactive and innovative to gain competitive advantages through information and technology. The information available in current corporate databases enhances business opportunities and gains further insights. These insights will help "provide better customer service, make call centers more efficient, cross-sell products more effectively, help sales staff close deals faster, simplify marketing and sales processes, discover new customers, increase customer revenues, etc." (Osarenkhoe & Bennani, 2007, p. 156).

Some studies explored various methods of identifying potential customers as the target market for cross-selling efforts. For example, Kamakura et al. (1991) developed a probability model to identify prospects based on their acquisition of financial products. Building on their original work, Kamakura et al. (2003) improved their model by adding a "factor analyzer" to enhance their model efficiency. Harrison and Ansell (2002) used the statistical method known as the "survival approach" to identify prospects and determine which product will be bought and when the purchase will be made. Knott et al. (2002) developed a next-product-to-buy (NPTB) model to predict which product a customer would most likely buy next using four different statistical techniques focusing on logistic regression, multinomial logit, discriminant analysis, and neural nets. Knott et al. (2002) found that, although the NPTB model generated incremental cross-selling profits, its predictive power was not influenced by the statistical technique used during estimation. Due to the availability of data and the recent development of more sophisticated statistical programs, scholars can utilize available information to continue improving the predictive power of cross-selling models. Li et al. (2005) used a structural multivariate probit model to develop a product acquisition sequence based on customers' levels of demand maturity. Rather than identifying potential prospects, Li et al. (2005) focused on identifying the next product that existing customers would be more likely to buy.

Furthermore, Li et al. (2011) introduced a customer response model that can help companies recommend the right products to customers at the right time using the proper communication channels. As a subbranch of business process management (BPM), Customer Relationship Management (CRM) and its relationship with BPM have been studied widely. However, research in a specific area, like insurance cross-selling, is still limited. Current research results support using a novel machine learning system to explore further and maximize benefits from insurance cross-selling activities. The research aims to identify which customers are likely to buy a new caravan insurance policy and seeks to improve model accuracy by combining sampling strategies and an ensemble approach. The ensemble approach, a machine learning algorithm, is further discussed in this paper.

Previous studies have accumulated sufficient knowledge to help us understand consumers' loyalty behavior, which can significantly impact cross-selling efforts. Likewise, big data and artificial intelligence (AI) have increasingly led to the development of technologies that can help to mimic the human brain and thus predict human behavior more precisely, further facilitating more business activities. Analytical tools, like machine learning, can effectively help insurance companies harness the potential of big data, assist with gleaning meaningful information, and make informed decisions (Miklosik & Evans, 2020). It must be noted, however, that the insurance industry still faces challenges addressing the machine learning models' transparency and interpretability concerns (Wang & Siau, 2019; Tian & Han, 2022). To be deployed in real-world applications, the black-box machine learning approaches must be validated and use a rationale humans can understand. The lack of scientific and empirical evidence supporting the link between the model and real-world cases reduces our understanding of the machine learning model (Sullivan, 2022). Thus, the current study proposes a more transparent and novel machine learning approach to predict customer behavior and improve cross-selling efficiency. This study contributes to enhancing the transparency and interpretability of AI-based models. The clarity of machine learning will assist marketing practitioners in making better and more informed cross-selling decisions.

LITERATURE

The potential of machine learning has long been recognized. With the recent increase in computing power, machine learning is becoming a reality in an increasing number of industries and sectors. In this section, Machine learning is discussed from the perspective of the Fourth Industrial Revolution. Next, the application of machine learning is examined, followed by a review of machine learning, biological neural networks, and multilayer perception. Finally, a discussion of ensemble approaches completes this section.

Machine Learning in the Fourth Industrial Revolution

Klaus Schwab (2017), founder and executive chairman of the World Economic Forum, introduced the Fourth Industrial Revolution (FIR) concept. This concept is built upon the foundations of the first three industrial revolutions. During the first revolution, the advent of the steam engine allowed production to be mechanized for the first time. Electricity and other scientific innovations led to mass production in the second industrial revolution. The emergence of computers and digital technologies in the third industrial revolution increased the automation of the manufacturing industry. Artificial intelligence, 5G connectivity, and other technologies are identified as FIR technologies. These emerging technologies blur the boundaries between the physical, biological, and digital spheres (Schwab, 2017; Krafft et al., 2020). In examining how technologies can blur the boundaries between the physical, digital, and biological spheres, Krafft et al. (2020) introduced the "boundary object" concept to explore the diverse roles that various technologies assume. According to their study, three FIR phenomena—big data, machine learning, and artificial intelligence—can be used to process more information for enhanced learning and create a higher-level intelligence. Big data facilitates processing massive amounts of information, while machine learning enhances learning. The goal of these phenomena is to achieve a higher level of intelligence.

Recent academic studies have adapted Krafft et al.'s (2020) framework. For example, articles in a special issue (August 2020) of the Journal of Interactive Marketing discussed the performance of various tasks using different technologies and across different fields. Machine learning has already been used in CRM and other related areas as one of the technologies that can enhance learning. According to Forbes (Columbus, 2020), prevalent applications include consumer and market segmentation, computer-assisted diagnostics, call center virtual assistants, sentiment analysis/opinion mining, face detection/recognition, and human resources applications. Similarly, machine learning in the insurance industry is gaining momentum through evaluating risks, accelerating claims, and detecting fraud (Tian & Han, 2022; Hanafy & Ming, 2021; Tian, 2017). Machine learning offers more holistic, reliable, and representative data analysis and results to solve complex problems quickly. It could also produce a higher profit margin for the insurance business. On the other hand, only a few insurance companies have started exploiting machine learning, and knowledge of the technology is still relatively scarce among insurance professionals (Koster et al., 2021).

The Application of Machine Learning in Insurance Customer Relationship Management

Based on an analysis of related literature, CRM can be defined as the process of managing knowledge and customer interactions. It is the most critical process in BPM. Understanding and managing a company's customers involves people, processes, and technology (Lau et al., 2016). The definition suggests that CRM seeks to build and maintain a profit-maximizing relationship between companies and their customers. Previous predictive modeling insurance studies concentrated on customer segmentation, retention, profitability, and satisfaction.

Florez-Lopez and Ramon (2009) developed a three-stage customer segmentation approach that combines marketing feature selection, customer segmentation through the univariate and oblique decision-tree techniques, and global cost-benefit function to measure a program's success. The results show that decision-tree segmentation techniques can lead to higher overall performance while managers are still easily understood. Roodpishi and Nashtaei (2015) utilized k-means clustering to categorize 300 insurance customers based on their demographic variables such as gender, age, occupation, education level, marital status, place of residence, and clients' incomes. Subsequently, they employed the association rule method to discover hidden patterns in the insurance business, such as the characteristics of life insurance and engineering insurance customers. Thakur and Sing (2013) mined data from auto insurance customers and predicted their product demand using k-means clustering. Similarly, Jandaghi et al. (2015) utilized fuzzy clustering to group customers based on their demographic and behavioral data.

An emerging area is the dynamic segmentation of users based on a machine learning technique known as Latent Dirichlet Allocation (LDA). LDA is used to identify clusters of related entities in cases where the relationships are not clearly defined or even recognized. Through this method, characteristics emerge that allow for the extrapolation of behaviors to groups based on past behavior and other data sources, which lead to the subsequent creation of offers that appeal to the targets. LDA can help to identify clusters of users who purchase certain products and classify a particular user as belonging to a specific group (Earley, 2015).

Measuring customer retention involves identifying which customers would be most likely to leave and which would be most likely to stay with a company. Companies can increase their sales by focusing their marketing resources on high-quality customers who have repurchased intention or can bring high revenue and profit (Banasiewicz, 2004). Smith et al. (2000) presented a case study involving various machine learning techniques, such as logistic regression, neural networks, and decision trees, to understand the retention patterns of policyholders. Based on the possibility of policyholders canceling their policies, the insurer determined the cost of misclassification and the optimal pricing level required to improve market decisions.

Measuring customer profitability involves identifying patterns based on factors, such as the products used by a customer, customer satisfaction, and sell opportunities to predict the profitability of a customer. Fang et al. (2016) applied random forecast regression, a big data analytics method, to forecast insurance customer profitability. Their study data found that the region, age, insurance status, sex, and source of customers are the principal factors to consider when predicting insurance customer profitability. Analyzing customers' behaviors would increase the possibility of their retention and bring in more revenue to the company. Mehregan and Samizadeh (2012) employed a k-means algorithm to identify customers who purchased one or more insurance policies and the most functional attributes related to customers' purchase intention.

Customer satisfaction is a complex and crucial issue for insurance companies and is related to multiple business processes and various unstructured data. Feature selection is a critical task that must be completed before running a machine-learning model. Bockhorst et al. (2016) developed a machine-learning-based framework to predict customer claim satisfaction by extracting relevant information from claim losses, notes, calls, and activity log data. The part-of-speech (PoS) tagger tokenizes the claim handler's notes to generate the most frequent words in a term-frequency–inverse document-frequency (TI-IDF) matrix. This process reduces the dimensionality of the attributes in the principal component analysis and effectively extracts information from unstructured text data.

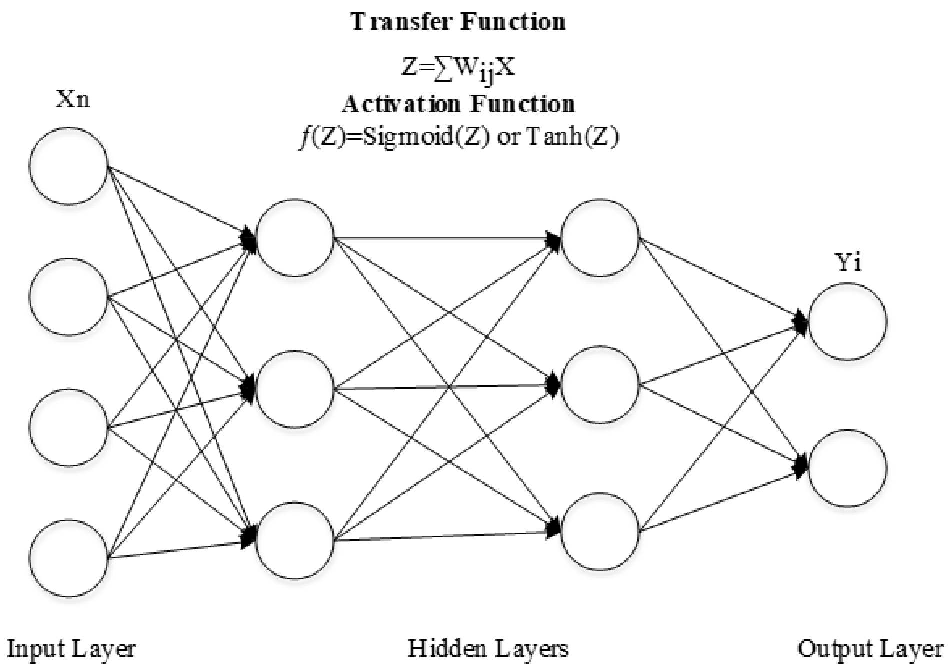
Cross-selling refers to the CRM practice of selling additional insurance policies to existing customers. For companies, the primary benefits of cross-selling include increased revenue and profitability. Cross-selling enables customers to purchase necessary products or services from trusted vendors and reduce costs. Despite the apparent importance of cross-selling, the topic has not attracted much attention from insurance marketers. Most strategic cross-selling decisions have relied on the intuition and experience of managers (Ansell et al., 2007). In practice, one of the most paramount important CRM decisions for an insurance company is determining which customers will buy a specific product or service. Insurance executives must become market- and data-driven (Miklosik & Evans, 2020). Machine learning approaches, especially artificial neural networks (ANNs), can be utilized to support this decision-making process.

Machine Learning and ANNs

An ANN is a machine learning technique inspired by the biological neural networks that constitute human brains. One type of ANN is called the multilayer perceptron (MLP). The most common method for training an MLP is backpropagation, which computes the weights of a multilayer network and employs gradient descent to determine the minimum squared error between the network's output values and the target values for these outputs. Typical learning approaches for ANNs are supervised learning and unsupervised learning. ANNs can be used in most machine-learning tasks, such as classification, regression, and clustering. Unlike traditional statistical techniques like discriminant analysis, the effectiveness of an ANN does not depend on various assumptions and conditions (Zhang, 2000).

An ANN is based on a collection of several interconnected units called neurons (nodes). Each connection can transmit information results between neurons located in different layers. The primary computing components of an ANN are weights, a summation transfer function, and an activation function. Figure 1 shows the four-layer structure and mechanism of an ANN.

FIGURE 1
MECHANISM OF AN ANN



The left layer of the network is called the input layer, while the right layer is called the output layer. The two middle layers are called the hidden layers; the number of hidden layers can range from 1 to ∞.

The most common choice among activation functions is the sigmoid function:

$$f(Z) = \frac{1}{1 + exp\,(-Z)}$$

The second most common choice is the hyperbolic tangent function:

$$f(Z) = \frac{e^{Z} - e^{-Z}}{e^{Z} + e^{-Z}}$$

The tanh(z) function is a rescaled version of the sigmoid, and its output range is [−1, 1] instead of [0, 1].

The third option for an activation function is the rectified linear unit (ReLU) function, which is defined as the positive part of its argument $f(Z) = Z^{+} = \max(0, Z)$, where x is the input to a neuron. This function will output the input directly if positive; otherwise, it will output zero.

These activation functions receive the output of the summation transfer function and produce a nonlinear decision boundary, which helps to solve several nonlinear classification problems. Two measures commonly used to measure the size of neural networks are the number of neurons and the network weights and biases.

Unlike other learning techniques, ANNs can identify complicated nonlinear relationships and interactions between independent and dependent variables. They can help to analyze noisy, incomplete, and less accurate data. Tuning parameters, such as hidden nodes within hidden layers, can improve the performance of an ANN.

Due to their powerful prediction ability and flexible applications, ANNs have been used in various business and finance research fields to solve regression and classification problems, such as bankruptcy prediction (Horak et al., 2020), market share forecasting, stock performance (Chhajer et al., 2022; Kurani et al., 2023), bond trades, and loan applications (Aslam et al., 2019). Most studies in these fields have suggested that ANNs perform as well as or better than other machine learning/statistical techniques (Wong & Selvi, 1998; Vellido et al., 1999). Current emerging areas in ANN research include image processing, text recognition (Kumar et al., 2016), disease detection (Kirmani & Ansarullah, 2016), and stock market analysis and prediction (Mahendran et al., 2020).

Previous studies have highlighted a few applications of ANNs in insurance, including insolvency management, claim fraud detection, revenue forecasting, and customer segmentation. Brockett et al. (1997) proposed a three-layer neural network model with three hidden units to provide the Texas Department of Insurance with early warnings of the insolvency of property/casualty insurance companies using annual financial data. With the appropriate selection of eight out of the 24 variables, the performance of this model was improved. In all three training sessions, the percentage of correctly classified data was above 88%, outperforming A.M. Best, a rating agency known for issuing insurance company insolvency ratings.

The first application of ANNs in insurance fraud detection began in 2000 (Phua et al., 2004; Viaene et al., 2002). Viaene (2005) explored the characteristics of neural network models using automatic relevance determination weight regularization, which provided domain experts with the option of finding the most informative predictors of fraudulent personal injury claims.

Bahia (2013) utilized an ANN to forecast an insurance company's next 41 years of revenue based on its actual premiums from 1970–2011. His research indicated that the insurance company would achieve revenue growth of 120%. The model measures, i.e., the mean squared error, showed that the best-fitting neural network structure would include one input layer, five hidden layers, and one output layer.

Sehgal et al. (2012) developed two ANNs to predict the types of insurance clients using tuning parameters such as the learning rate, hidden layers, number of neurons in each layer, stopping criteria, and activation function. They concluded that the ANNs are suitable for solving complex nonlinear problems with high accuracy and speed.

Yunos et al. (2016) discussed the factors influencing the performance of a backpropagation neural network (BPNN) in predicting the frequency and severity of expected claims and how to improve model performance by tuning the network structure, parameters, and error measurements. They found that the BPNN could tackle the problem using nonlinear claim data.

Recent studies have revealed that ANNs outperform other learning techniques in various cases (Paliwal & Kumar, 2009; Vassiljeva et al., 2017; Tolani et al., 2019). However, neural networks also have a disadvantage: the techniques are challenging to explain and interpret. The "black box" method limits individuals' understanding of the process and its conclusion. Furthermore, neural networks consume enormous amounts of computational resources during training. A simple network with fewer layers and neurons can mitigate these concerns.

Determining the number of hidden layers and the nodes in each hidden layer is challenging. However, the number of variables can be selected as the number of units in the input layer and the number of classes as the number of units in the output layer. For hidden layers, the selection strategy varies. Researchers use single-layer neural networks to train models (Sehgal et al., 2012; Yunos et al., 2016). Such networks can be used for easy tasks and reduce the training time involved. However, the capabilities of such neural networks are limited.

Ensemble Approaches

Ensemble approaches use multiple machine learning algorithms to obtain better predictive outcomes than could be obtained from a single learning algorithm alone. In classification problems, ensemble learning models' generalization ability and prediction accuracy make them more robust than a single model. Based on their structures, ensemble learning approaches can be divided into homogeneous and heterogeneous categories (Aburomman & Reaz, 2017). The first is generated using the same models, while the second involves multiple models. Regarding the final model selection method, standard methods used to build accurate ensemble models include bagging, boosting (Opitz & Maclin, 1999), and stacking. The current study applies a stacking ensemble due to its proven performance in several previous studies (Džeroski & Ženko, 2004; Todorovski & Džeroski, 2003).

Leveraging big data analytics empower businesses to target customers effectively, provide personalized solution, reach and retain customers, and gain competitive advantages ((Miklosik & Evans, 2020). This study introduces a three-stage machine learning-based system (MLBS) to identify insurance customers and manage customer relationships accurately. The MLBS combines undersampling strategies and an ensemble approach to improve prediction performance. It involves selecting the best training sample using artificial neural networks and employing the stacking ensemble approach to design the prediction model. The system can assist non-expert users in analyzing typical insurance marketing problems, even a lack of insurance and cross-selling expertise.

DATASET AND METHODS

Dataset

The study dataset was obtained from the CoIL Challenge and is owned by Sentient Machine Research, a Dutch data mining company. The dataset is based on real-world business data and contains 86 variables, including product purchase data and socio-demographic data derived from zip codes. The variables beginning with M, P, and A refer to the demographic, product ownership, and insurance statistics, respectively, for each postal code. The research objective was to identify customers interested in purchasing a new caravan policy. A description of the variables is provided in the Appendix.

Methodology

Insurance cross-selling is a binary classification problem requiring that customers' insurance transactional information (training data) is combined with the known purchase history to predict purchase intention from unknown incoming customer records. Mathematically, it is equivalent to the optimization problem of building a decision function $f(x)\colon \mathcal{R}^p \to Y$, given a training dataset $X = \{x_i, y_i\}_i^n, x_i \in \mathcal{R}^p$, $y_i \in Y$ ($Y = \{1,2\}$, by minimizing the expected loss defined on $\mathcal{R}^p \times Y$: $L(f) = \int_{\mathcal{R}^p \times Y} l(f(x), y)dxdy$, where $l(f(x), y)\colon Y \times Y \to \mathcal{R}^+$ is a defined loss function (e.g., $l(f(x), y) = (f(x) - y)^2$).

The best way to determine an appropriate distribution for predictive modeling is by conducting multiple experiments and selecting the distribution that produces the best classifier. The distribution should balance time complexity and model performance (Hassan & Abraham, 2016; Chan et al., 1999; Chan & Stolfo, 1995). In this study, we used a fixed number of buying customers and merged them with the entire non-buying customers group. The modeling subsamples were formed by combining all buying observations (586 customers) with different non-buying observations. The distributions obtained were 50:50, 60:40, and 70:30, with the non-buying versus buying observation ratios of 586:586, 879:586, and 1367:586, respectively. We also applied sampling strategies with replacements and without replacements. Finally, we obtained six subsamples for predictive modeling.

Besides obtaining better learning results with more data, other techniques, such as the stacking ensemble (Wolpert, 1992) and cross-validation techniques, are appropriate for further improving machine learning performance. Stacking combines several classifiers to obtain better prediction results. In the stacking process, low-level classifiers are trained using the original data. Next, the stacked model is trained using the outcomes derived from the low-level classifiers instead of bagging and boosting (Dietterich, 2000). The stacking architecture integrates two or more based classifiers and a meta-classifier to obtain the predictions of the base classifiers. Machine learning can be used as the base classifier and further integrated into each stacking layer. Typically, stacking can generate better performance than other heterogeneous methods (Džeroski & Ženko, 2004; Todorovski & Džeroski, 2003).

The current study proposed a three-stage system for cross-selling predictions based on undersampling strategies, ANNs, and stacking ensembles. In the first stage, a novel undersampling strategy was proposed to process the imbalanced dataset. The dataset was then divided into the majority group and the minority group. All observations in the minority group were retained. Meanwhile, two sampling methods, replacement and without replacement, were applied to the majority group. After sampling a particular ratio of observations from the majority group, we combined it with the minority group. Finally, we obtained six different subsamples for modeling. These subsamples were 50:50 (not buy: buy) with/without replacement, 60:40 with/without replacement, and 70:30 with/without replacement.

In the second stage, the undersampling techniques were assessed using ANNs, which are efficient and effective in terms of predictions and comparing the different undersampling methods and the original dataset. The model designed using the subsamples was compared with the model developed using the original dataset. The undersampling method that produced the model with the best performance (highest recall) was used for the third stage of the experiment.

In the third stage**,** a stacking method using logistic regression and a support vector machine (SVM) as aggregators and ANNs as learners was employed to identify the best pipeline. The binary SVM model constructs an optimal hyperplane $y = w^T x + b$ to separate two groups of data points of the training data $X = \{x_i, y_i\}_{i=1}^{n}, x_i \in R^p, y_i \in \{-1, +1\}$, where y_i is the label of the observation x_i, by solving a quadratic programming problem:

$$\begin{aligned} &min_w \ \frac{1}{2} w^T w + \mathrm{C} \sum\nolimits_{i=1}^{n} \xi_i \ , w \in R^d \ , \xi_i \in R, b \in R \\ &s.t.\, y_i (w^T \varphi(x_i) + b) \geq 1 - \xi_i \ , \xi_i \geq 0, i = 1,2 \cdots n \end{aligned} \tag{1}$$

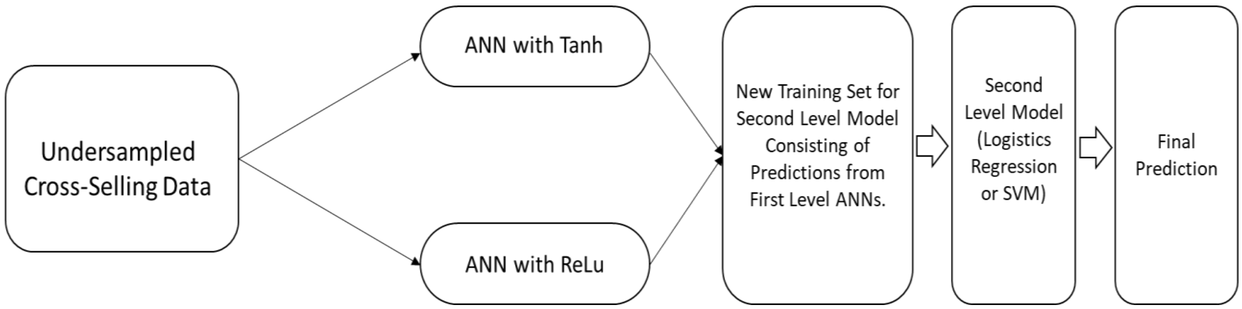
where $w \in \ R^p$ is the normal vector of the hyperplane and b is the offset, $\mathrm{C} \in R^+$ is the regularization parameter, and $\varphi(\cdot)$ is an implicit feature function mapping input data to the high-dimensional feature space for evaluation using kernel tricks. The logistic regression employs a conditional probability model $P(\mathrm{y} = \pm 1 | x, w) = \frac{1}{1 + e^{-y w^T x}}$, in which x is an observation, y is its label, and w is the weight vector, to conduct classification. Given the training data $X = \{x_i, y_i\}_{i=1}^{n}, x_i \in R^p, y_i \in \{-1, +1\}$, logistic regression optimizes the following the regularized negative log-likelihood by solving a quadratic programming problem:

$$min_w \ \tfrac{1}{2} w^T w + \mathrm{C} \textstyle\sum_{i=1}^{n} \log\left(1 + e^{-y_i w^T x_i}\right), w \in R^p \tag{2}$$

where C $w \in R^+$ is the penalty parameter.

Two stacking methods were applied during the experiments. The first method used an SVM as an aggregator, and the second used logistic regression. The aggregated learners were ANNs with two activation functions: Tanh and ReLU. The stacking procedure used in the study is shown in Figure 2.

FIGURE 2
STACKING PROCEDURE



A ten-fold cross-validation method was used in the study to obtain statistically significant results—nine subsamples were used for training, and one subsample was used for testing. Subsequently, the results of the ten tests were averaged.

The most common model performance criteria were accuracy, recall, specificity, precision, and area under the curve (AUC). Table 1 presents the calculation and descriptions of accuracy, recall, specificity, and precision.

TABLE 1
CONFUSION MATRIX AND MEASUREMENTS

Total population	Prediction positive	Prediction negative
True condition positive	True positive (TP)	False negative (FN)
True condition negative	False positive (FP)	True negative (TN)

$$\text{Accuracy} = \frac{\Sigma\,\text{TP} + \Sigma\,\text{TN}}{\Sigma\,\text{Total population}}$$

$$\text{Recall} = \frac{\Sigma\,\text{TP}}{\Sigma\,\text{True condition positive}}$$

$$\text{Specificity} = \frac{\Sigma\,\text{TN}}{\Sigma\,\text{True condition negative}}$$

$$\text{Precision} = \frac{\Sigma\,\text{TP}}{\Sigma\,\text{Prediction positive}}$$

We used recall as a selection measure to better evaluate cross-selling models with an imbalanced dataset. Unlike accuracy and precision, recall cannot be manipulated by the majority (negative) class. It also meets the goal of predictive modeling because it focuses on identifying as many potential buyers as possible. The AUC values, which indicate excellent test accuracy when they are close to 1, are also provided for reference.

RESULTS AND ANALYSIS

The proposed six undersampling techniques were evaluated by running a three-layer ANN with three neurons in the hidden layer. The performance scores for each technique are shown in Table 2.

TABLE 2
ANN RESULTS FOR THE SIX UNDERSAMPLING METHODS

Undersampling method	Recall	Precision	AUC
50:50 no replacement	0.637	0.65	0.701
60:40 no replacement	0.672	0.701	0.681
70:30 no replacement	0.762	0.783	0.691
50:50 replacement	0.672	0.659	0.711
60:40 replacement	0.695	0.727	0.713
70:30 replacement	0.795	0.791	0.719
Original dataset	0.055	0.172	0.669

The undersampling method 70:30 with replacement yielded the best recall (0.795).

Next, two stacking ensemble models were designed using logistic regression and an SVM as aggregators and ANNs as learners. There were two low-level learners, both of which were ANNs with three neurons in each network's hidden layer. However, one used the Tanh function, while the other used the ReLU function. Additionally, we ran the modeling using single classifiers, such as logistic regression, SVM, ANN, and random forests (RFs), to compare the model's performance.

The analysis results showed that the stacking method based on logistic regression significantly improved the model's performance. This method generated a recall of 0.904, which is higher than that of the non-stacking models. However, the stacking method with SVM was not appropriate for the study dataset since it had the lowest measurement values. Table 3 provides the comparison results for the stacking and traditional machine learning models.

TABLE 3
PERFORMANCE MEASUREMENTS FOR STACKING AND UNSTACKING MODELS

Classification method	Recall	Precision	AUC
Stacking with logistic regression	0.904	0.757	0.719
Stacking with SVM	0.544	0.677	0.536
Logistic regression	0.849	0.777	0.726
SVM	0.764	0.725	0.594
ANN with ReLU function	0.751	0.787	0.688
ANN with Tanh function	0.795	0.791	0.719
Random forest	0.866	0.753	0.712

CONCLUSION AND DISCUSSION

Insurance CRM involves the process of managing business and customer interactions effectively. The challenge of a complex, dynamic influx of incoming big data in various formats requires insurance companies to develop advanced methodologies for future revenue increases and profitability. Automatic tools and processes, especially machine learning and AI-based approaches help insurance companies achieve such goals by building continuing relationships and personal touchpoints (Behera et al., 2020). However, insurance practitioners and researchers need to make their machine-learning and AI-based systems explainable to foster trust and transparency in the black box of the machine-learning approaches (Samek & Müller, 2019). The machine learning algorithms, model structures and procedures, sampling strategies, and real-world insurance data discussed in the study provide marketing researchers and practitioners with a complete and verifiable path to design and test intelligent marketing information systems.

With the availability of cheap and powerful computing resources and data storage, predictive modeling methods, such as the machine learning approach discussed in this study, can help insurers identify potential customers for future services quickly and automatically. It is crucial in the insurance value chain (Paruchuri, 2020). The iterative and automatic aspects allow for quick and continuous learning from big data and support fast, reliable decision-making; however, this study only examines the machine learning approach using small-scale data. The proposed system should be tested using a larger cross-selling dataset when available. Furthermore, machine learning-based marketing approaches can do repetitive and tedious tasks. To make the maximal benefits of machine learning, speed, accuracy, consistency, and transparency should be focused on and achieved. The automatic black-box machine learning should be explainable and understandable by average professionals. Companies must also embrace collaborative intelligence (Wilson & Daugherty, 2018) and foster an environment where humans and Artificial Intelligence systems can work together.

The results showed that the ANN model performed better when using the undersampling method than the original imbalanced data. Meanwhile, the ANN model demonstrated its powerful prediction ability for the insurance cross-selling problem. Choosing the hidden layer and activation functions are effective strategies for enhancing the performance of ANN models. Based on the stacking of neural networks, multiple ANNs can be stacked using another machine learning classifier as the aggregator. This could allow us to derive more information from insurance cross-selling data and identify more target consumers within an organization's database. The ensemble model leverages the performance of neural networks and enables more robust predictions and classifications. Thus, this study found that using logistic regression as an aggregator can significantly improve the model's performance in terms of cross-selling. The simple and interpretable models can be used for high-stakes decisions without losing accuracy (Semenova et al., 2022). Compared to opaque, complex, and hyper-realistic models, the simple approach provided more understanding and was less misguided. From a business strategy perspective, the transformative MLBS for CRM and cross-selling will help insurance companies exploit their business data better and create new business models to achieve their data-driven strategic objectives and goals. From the value creation perspective, the efficient, effective, and informative decisions from the outcome of the MLBS would create new value by improving internal and external customer satisfaction. From the BPM perspective, MLBS can help insurance companies simplify and streamline their business processes in CRM while lowering costs and saving time – all without damaging customers' experience.

The accuracy of the prediction in insurance marketing can have a significant effect on its business. Insurers should put adequate resources into training their marketing professionals to adapt and apply machine learning appropriately and ensure better data quality. Insurers' ability to monitor machine learning and make fast data-driven decisions based on machine learning will maximize its advantages. We also suggest evaluating model performance using multiple metrics, such as recall, AUC, and other measures. The various metrics can increase modeling process transparency and prevent discrimination and harm to customers and businesses (Hanafy & Ming, 2021).

Responding to severe market competition, companies, especially insurance companies, must strive to efficiently and effectively manage customer relationships to increase business performance. Best-performing companies must obtain specific knowledge of customers and technology (Lau et al., 2016). Applying machine learning methods assists the insurance industry in enhancing performance in marketing strategy development, innovation inspiration, revenue accumulation, and cost control. It not only allows insurers to identify customers they want to serve but also facilitates a fundamental shift from managing product portfolios to portfolios of customers (Chen & Popovich, 2003). The proposed system can be widely used to optimize insurance marketing processes and solve various classification and prediction problems in insurance functions, such as fraudulent claims, risk severity analysis, agent classification, and employee recruitment.

LIMITATIONS AND FUTURE DIRECTIONS

This study only focused on strategies for predicting cross-selling opportunities. Future research can apply other approaches, such as oversampling, cost-sensitive training, boosting and bagging ensembles, neural network parameter tuning, and other necessary procedures, to address class imbalance issues (Wang et al., 2019).

According to the posterior probabilities derived from the proposed model, potential customers can be divided into multiple groups for future cost analysis. When considering each customer's different soliciting costs and potential revenue, insurers should avoid investing too much in false-positive or false-negative targets. Future research could discuss this issue by exploring the gain scores method.

Ensemble methods integrate the benefits of several methods. They usually generate better and more accurate predictive results. The generalization ability and prediction accuracy of ensemble learning models are more robust than that of a single model. Future research can apply different ensemble approaches, such as boosting and bagging, to improve the model's performance.

Furthermore, future research can test the various parameters associated with different classifiers. For instance, experiments with RF and SVM classifiers involved varying the number of trees and other kernel functions. Similarly, the results from "L1" and "L2" can be used for model comparison for the logistic regression classifier.

REFERENCES

Aburomman, A.A., & Reaz, M.B.I. (2017). A survey of intrusion detection systems based on ensemble and hybrid classifiers. *Computers & Security*, *65*, 135–152.

Ansell, J., Harrison, T., & Archibald, T. (2007). Identifying cross-selling opportunities, using lifestyle segmentation and survival analysis. *Marketing Intelligence & Planning*, *25*(4), 394–410.

Aslam, U., Tariq Aziz, H.I., Sohail, A., & Batcha, N.K. (2019). An empirical study on loan default prediction models. *Journal of Computational and Theoretical Nanoscience*, *16*(8), 3483–3488.

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

Banasiewicz, A. (2004). Acquiring high value, retainable customers. *Journal of Database Marketing & Customer Strategy Management*, *12*(1), 21–31.

Behera, R.K., Gunasekaran, A., Gupta, S., Kamboj, S., & Bala, P.K. (2020). Personalized digital marketing recommender engine. *Journal of Retailing and Consumer Services*, *53*, 101799.

Bockhorst, J., Wang, Y., Gupta, S., Qazi, M., Sun, M., & Fung, G. (2016). Using temporal discovery and data-driven journey-maps to predict customer satisfaction. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 846–852). IEEE.

Brockett, P.L., Cooper, W.W., Golden, L.L., & Xia, X. (1997). A case study in applying neural networks to predicting insolvency for property and casualty insurers. *Journal of the Operational Research Society*, pp. 1153–1162.

Chan, P.K., & Stolfo, S.J. (1995). A comparative evaluation of voting and meta-learning on partitioned data. In *Machine Learning Proceedings 1995* (pp. 90–98). Morgan Kaufmann.

Chan, P.K., Fan, W., Prodromidis, A.L., & Stolfo, S.J. (1999). Distributed data mining in credit card fraud detection. *IEEE Intelligent Systems and Their Applications*, *14*(6), 67–74.

Chen, I.J., & Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. *Business Process Management Journal*, *9*(5), 672–688.

Chhajer, P., Shah, M., & Kshirsagar, A. (2022). The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. *Decision Analytics Journal*, *2*, 100015.

Columbus, L (2020). Roundup of Machine Learning Forecasts and Market Estimates, 2020. *Forbes Contrib*. Retrieved December 31, 2022, from https://www.forbes.com/sites/louiscolumbus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/?sh=46e9f2285c02

Dietterich, T.G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, *40*(2), 139–157.

Džeroski, S., & Ženko, B. (2004). Is combining classifiers with stacking better than selecting the best one? *Machine Learning*, *54*(3), 255–273.

Earley, S. (2015). Cognitive computing, analytics, and personalization. *IT Professional*, *17*(4), 12–18.

Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using big data analytics: A case study of the insurance industry. *Computers & Industrial Engineering*, *101*, 554–564.

Florez-Lopez, R., & Ramon-Jeronimo, J.M. (2009). Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting. *Social Science Computer Review*, *27*(1), 96–117.

Hanafy, M., & Ming, R. (2021). Machine learning approaches for auto insurance big data. *Risks*, *9*(2), 42.

Harrison, T., & Ansell, J. (2002). Customer retention in the insurance industry: Using survival analysis to predict cross-selling opportunities. *Journal of Financial Services Marketing*, *6*(3), 229–239.

Hassan, A.K.I., & Abraham, A. (2016). Modeling insurance fraud detection using imbalanced data classification. In *Advances in Nature and Biologically Inspired Computing* (pp. 117–127). Cham: Springer.

Jandaghi, G., Moazzez, H., & Moradpour, Z. (2015). Life insurance customers segmentation using fuzzy clustering. *World Scientific News*, *21*, 38–49.

Kamakura, W.A., Ramaswami, S.N., & Srivastava, R.K. (1991). Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing*, *8*(4), 329–349.

Kamakura, W.A., Wedel, M., De Rosa, F., & Mazzon, J.A. (2003). Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *International Journal of Research in Marketing*, *20*(1), 45–65.

Kirmani, M.M., & Ansarullah, S.I. (2016). Classification models on cardiovascular disease detection using neural networks, naïve bayes and J48 data mining techniques. *International Journal of Advanced Research in Computer Science*, *7*(5), 52–62.

Knott, A., Hayes, A., & Neslin, S.A. (2002). Next-product-to-buy models for cross-selling applications. *Journal of Interactive Marketing*, *16*(3), 59–75.

Koster, O., Kosman, R., & Visser, J. (2021, September). A checklist for explainable AI in the insurance domain. In *International Conference on the Quality of Information and Communications Technology* (pp. 446–456). Springer, Cham.

Krafft, M., Sajtos, L., & Haenlein, M. (2020). Challenges and opportunities for marketing scholars in times of the fourth industrial revolution. *Journal of Interactive Marketing*, *51*, 1–8.

Kumar, S., Kumar, K., & Mishra, R.K. (2016). Scene text recognition using artificial neural network: A survey. *International Journal of Computer Applications*, *137*(6), 40–50.

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, *10*(1), 183–208.

Lau, H., Nakandala, D., Samaranayake, P., & Shum, P.K. (2016). BPM for supporting customer relationship and profit decision. *Business Process Management Journal*, *22*(1), 231–255.

Li, S., Sun, B., & Montgomery, A.L. (2011). Cross-selling the right product to the right customer at the right time. *Journal of Marketing Research*, *48*(4), 683–700.

Li, S., Sun, B., & Wilcox, R.T. (2005). Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research*, *42*(2), 233–239.

Mahendran, A., Vasavada, K.A., Tuteja, R., Sharma, Y., & Vijayarajan, V. (2020). Stock market analysis and prediction using artificial neural network toolbox. In *Embedded Systems and Artificial Intelligence* (pp. 559–569). Singapore: Springer.

Mehregan, S., & Samizadeh, R. (2012). Customer retention based on the number of purchase: A data mining approach. *International Journal of Management and Business Research*, *2*(1), 41–50.

Miklosik, A., & Evans, N. (2020). Impact of big data and machine learning on digital transformation in marketing: A literature review. *IEEE Access*, *8*, 101284–101292.

Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligent Research*, *11*, 169–198.

Osarenkhoe, A., & Bennani, A.E. (2007). An exploratory study of implementation of customer relationship management strategy. *Business Process Management Journal*, *13*(1), 139–164.

Paliwal, M., & Kumar, U.A. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems With Applications*, *36*(1), 2–17.

Paruchuri, H. (2020). The Impact of Machine Learning on the Future of Insurance Industry. *American Journal of Trade and Policy*, *7*(3), 85–90.

Phua, C., Damminda, A., & Vincent, L. (2004, January). Minority report in fraud detection: Classification of skewed data. *Acm Sigkdd Explorations Newsletter*, *6*, 50–59.

Roodpishi, M.V., & Nashtaei, R.A. (2015). Market basket analysis in insurance industry. *Management Science Letters*, *5*(4), 393–400.

Rosen, T.A. (2004). Why your bank will fail at cross-selling. *Commerical Lending Review*, *19*(2), 41–44.

Samek, W., & Müller, K.R. (2019). Towards explainable artificial intelligence. *Explainable AI: Interpreting, explaining and visualizing deep learning*, pp. 5–22.

Schwab, K. (2017). *The fourth industrial revolution.* London, England: Portfolio Penguin.

Sehgal, P., Gupta, S., & Kumar, D. (2012). Application of neural networks in predictive data mining for insurance. *International Journal of Latest Trends in Engineering and Technology*, *1*(1), 1–4.

Semenova, L., Rudin, C., & Parr, R. (2022). On the existence of simpler machine learning models. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1827–1858).

Smith, K.A., Willis, R.J., & Brooks, M. (2000). An analysis of customer retention and insurance claim patterns using data mining: A case study. *Journal of the Operational Research Society*, *51*(5), 532–541.

Sullivan, E. (2022). Understanding from machine learning models. *The British Journal for the Philosophy of Science*, *73*(1), 109–133.

Thakur, S.S., & Sing, J.K. (2013). Mining customer's data for vehicle insurance prediction system using k-means clustering: An application. *International Journal of Computer Applications in Engineering Sciences*, *3*(4), 148–153.

Tian, X. (2017). Insurance fraud detection: An exploratory data mining approach. *Proceeding of Southwest Decision Sciences Institute 48th Annual Meeting*, pp. 121–129.

Tian, X., & Han, H. (2022). Deep convolutional neural networks with transfer learning for automobile damage image classification. *Journal of Database Management*, *33*(3), 1–16.

Todorovski, L., & Džeroski, S. (2003). Combining classifiers with meta decision trees. *Machine Learning*, *50*(3), 223–249.

Tolani, S., Rao, A., Worku, G.B., & Osman, M. (2019). System and neural network analysis of intent to buy and willingness to pay insurance premium. *Managerial Finance*, *45*(1), 147–168.

Vassiljeva, K., Tepljakov, A., Petlenkov, E., & Netšajev, E. (2017). Computational intelligence approach for estimation of vehicle insurance risk level. In 2*017 International Joint Conference on Neural Networks (IJCNN)* (pp. 4073–4078). IEEE.

Vellido, A., Lisboa, P.J., & Vaughan, J. (1999). Neural networks in business: A survey of applications (1992–1998). *Expert Systems with Applications*, *17*(1), 51–70.

Viaene, S., Dedene, G., & Derrig, R.A. (2005). Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications*, *29*(3), 653–666.

Viaene, S., Derrig, R.A., Baesens, B., & Dedene, G. (2002). A comparison of state-of-the-art classification techniques for expert automobile insurance claim fraud detection. *Journal of Risk and Insurance*, *69*(3), 373–421.
Vyas, R.S., & Math, N.R.B. (2006). A comparative study of cross-selling practices in public and private sector banks in India. *Journal of Financial Services Marketing*, *10*(4), 123–134.
Wang, J., Feng, K., & Wu, J. (2019). SVM-based deep stacking networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*(1), 5273–5280.
Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda. *Journal of Database Management*, *30*(1), 61–79.
Wilson, H.J., & Daugherty, P.R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, *96*(4), 114–123.
Wolpert, D.H. (1992). Stacked generalization. *Neural Networks*, *5*(2), 241–259.
Wong, B.K., & Selvi, Y. (1998). Neural network applications in finance: A review and analysis of literature (1990–1996). *Information & Management*, *34*(3), 129–139.
Yunos, Z.M., Ali, A., Shamsyuddin, S.M., Ismail, N., & Sallehuddin, R.S. (2016). Predictive modelling for motor insurance claims using artificial neural networks. *International Journal of Advances in Soft Computing and Its Applications*, *8*(3), 160–172.
Zhang, G.P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *30*(4), 451–462.

APPENDIX: DATASET VARIABLES AND DESCRIPTIONS

MOSTYPE: Customer Subtype
MAANTHUI: Number of houses 1–10
MGEMOMV: Average household size 1–6
MGEMLEEF: Average age
MOSHOOFD: Customer main type
MGODRK: Roman Catholic
MGODPR: Protestant
MGODOV: Other religion
MGODGE: No religion
MRELGE: Married
MRELSA: Living together
MRELOV: Other relation
MFALLEEN: Single
MFGEKIND: Household without children
MFWEKIND: Household with children
MOPLHOOG: High-level education
MOPLMIDD: Medium-level education
MOPLLAAG: Low-level education
MBERHOOG: High status
MBERZELF: Entrepreneur
MBERBOER: Farmer
MBERMIDD: Middle management
MBERARBG: Skilled laborers
MBERARBO: Unskilled laborers
MSKA: Social class A
MSKB1: Social class B1
MSKB2: Social class B2
MSKC: Social class C

MSKD: Social class D
MHHUUR: Rented house
MHKOOP: Homeowners
MAUT1: 1 car
MAUT2: 2 cars
MAUT0: No car
MZFONDS: National Health Service
MZPART: Private health insurance
MINKM30: Income < 30,000
MINK3045: Income 30,000–45,000
MINK4575: Income 45,000–75,000
MINK7512: Income 75,000–122,000
MINK123M: Income >123,000
MINKGEM: Average income
MKOOPKLA: Purchasing power class
PWAPART: Contribution private third-party insurance
PWABEDR: Contribution third-party insurance (firms)
PWALAND: Contribution third-party insurance (agriculture)
PPERSAUT: Contribution car policies
PBESAUT: Contribution delivery-van policies
PMOTSCO: Contribution motorcycle/scooter policies
PVRAAUT: Contribution lorry policies
PAANHANG: Contribution trailer policies
PTRACTOR: Contribution tractor policies
PWERKT: Contribution agricultural machines policies
PBROM: Contribution moped policies
PLEVEN: Contribution life insurances
PPERSONG: Contribution private accident insurance policies
PGEZONG: Contribution family accident insurance policies
PWAOREG: Contribution disability insurance policies
PBRAND: Contribution fire policies
PZEILPL: Contribution surfboard policies
PPLEZIER: Contribution boat policies
PFIETS: Contribution bicycle policies
PINBOED: Contribution property insurance policies
PBYSTAND: Contribution social-security insurance policies
AWAPART: Number of private third-party insurance 1–12
AWABEDR: Number of third-party insurance (firms)
AWALAND: Number of third-party insurance (agriculture)
APERSAUT: Number of car policies
ABESAUT: Number of delivery-van policies
AMOTSCO: Number of motorcycle/scooter policies
AVRAAUT: Number of lorry policies
AAANHANG: Number of trailer policies
ATRACTOR: Number of tractor policies
AWERKT: Number of agricultural machines policies
ABROM: Number of moped policies
ALEVEN: Number of life insurance policies
APERSONG: Number of private accident insurance policies
AGEZONG: Number of family accident insurance policies
AWAOREG: Number of disability insurance policies

ABRAND: Number of fire policies
AZEILPL: Number of surfboard policies
APLEZIER: Number of boat policies
AFIETS: Number of bicycle policies
AINBOED: Number of property insurance policies
ABYSTAND: Number of social security insurance policies
CARAVAN: Number of mobile home policies 0–1