Modern Economy, 2021, 12, 1213-1231
https://www.scirp.org/journal/me
ISSN Online: 2152-7261
ISSN Print: 2152-7245

An Insurtech for the Online Insurance: A Customer Repurchasing Behavior Study in Taiwan

Yun-Chiang Tai[1], Hsin-Neng Hsieh[2], Ching-Chih Wu[3*]

[1]Department of International Business, National Taiwan University, Taiwan
[2]Department of Statistics and Information Science, Fu Jen Catholic University, Taiwan
[3]Department of Finance, Providence University, Taiwan
Email: *ccwu886@gmail.com

How to cite this paper: Tai, Y.-C., Hsieh, H.-N., & Wu, C.-C. (2021). An Insurtech for the Online Insurance: A Customer Repurchasing Behavior Study in Taiwan. *Modern Economy, 12,* 1213-1231.
https://doi.org/10.4236/me.2021.128063

Received: June 29, 2021
Accepted: August 10, 2021
Published: August 13, 2021

Abstract

This study explores online insurance ventures by customers' behavioral profiles from data mining and visualization. First, we employ a simple decision-tree statistical learning method on unique hand-collected and processed data in Taiwan. This method considers first-time online insurance subscribers as new ones without a proper marketing connection after six months. Second, using advanced clustering technique and decision tree statistical learning on first-time purchasing customers, we find they repurchase online travel insurance for different purposes varying periods. Finally, we get robust results engaging different segmentations of customer data. In these ways, we enable marketing strategies to work with generating decision rules.

Keywords

Artificial Intelligence, Data Mining, Decision-Tree Statistical Machine Learning, Isurtech, Online Insurance, Unsupervised Clustering Learning

1. Introduction and Literature Review

Following Fintech's footsteps, Insurance Technology (henceforth Insurtech) actively develops along with Taiwan's online insurance business and technical collaboration. Online insurance was launched on July 31, 2015, in Taiwan with government authorization. At the outset, travel insurance is the first insurance product permitted selling on the internet. Insurance companies have strived to develop their online platforms to attract customers who are willing to purchase travel insurance in person. The government has not permitted insurance companies to sell health insurance online in Taiwan due to information asymmetry, even though online health insurance is popular in South Korea or Japan. However, after online travel insurance, online sales of savings insurance and annuity insurance were permitted to be launched in two years since 2015. Many business models have created new trading without agents to help in insurance sales. Online insurance is then a new chapter and a unique purchasing experience for the insurance buyer to have an insurance policy via an online platform for Taiwan's life insurance companies. Consumer purchasing behavior and online insurance shopping experience have become much more critical with artificial intelligence. Artificial intelligence tools such as text robots or website directives help facilitate the sales growth of online insurance.

This paper intends to portray online insurance purchasers' profile and know how to attract them to buy online insurance products with proper marketing strategies. Big data analytics, data and website platforms construction, and artificial intelligence are needed to accomplish these tasks. Graham (2018) mentions some connections between customer choices and platform construction and provides some insight into the issue of choice. Moreover, Yu and Chen (2018) use transaction cost theory to study customers' online travel insurance purchase intentions. Luo, Chen, Zhang and Xu (2019) investigate the three dimensions of trust belief on policyholders' purchase intentions in a third-party online insurance platform. Therefore, platform user experience and behavioral science of purchasing customers have become essential when starting a new online insurance business. Artificial intelligence tools embedded in webpages and statistical machine learning with customer servicing help E-insurance growing. Targeting customers is quite different from a variety of insurance policies. Various customers have their unique needs. Big data analytics, such as statistical machine learning methods, facilitate us to distinguish customer characteristics.

In business fields, including accounting, business management finance, and general insurance, machine learning methods have been applied for specific purposes. In accounting, machine learning helps to identify wrong booking records. As to business administration, statistical machine learning assists in decision-making. The business considers data mining a better strategy and first step to decision-making with visualization and classified tabulation during limited knowledge of potential online insurance customers. Kumar and Verma (2012) provide a concise introduction to data mining. Wahbeh, Al-Radaideh, Al-Kabi and Al-Shawakfa (2011) have suggested some data mining tools with artificial intelligence algorithms to help in decision-making. In finance, statistical machine learning helps customers' credit scoring (Hand & Henley, 1997; Henley & Hand, 1996) and financial product sales. Lee, Chiu, Chou and Lu (2006) use statistical learning methods to identify customers' credit profiles. Parodi (2012) indicates several computational intelligence applications on insurance issues. Cortis, Debattista, Debono and Farrell (2019) demonstrate what disruptive finance is in which Fintech and Insurtech have emerged from zero to a variety in a decade. Barry and Charpentier (2020) mention the risk associated with automobile insurance's aggregate data from sensor detectors and edge computing with big data analytics. Diana, Griffin, Oberoi and Yao (2019) provide some insurance applications with machine learning methods. Collaboration technologies in insurance are not new, but the one without the help of agents is. Scholars have found that machine-learning approaches could be pretty valuable in insurance analysis.

Statistical methods have once been considered a way to face unknown situations where we have no idea how to analyze those raw data. In our research, we depend on trailing online customers' footsteps to explore and extend our knowledge. How to deal with new compliance issues on online insurance becomes a big challenge when online customers do not show their faces and characteristics in-person. Moral hazard and adverse selection issues are resolved from the outset when customers visit online platforms. Online platform compliance and risk management are the priorities when constructing a platform under the regulator's requirement and supervision. New technology has been applied together with Google Analytics to grasp potential customers in the business. We provide the OTP (one-time-password) service when facing issues of stringent customer identification if needed. Furthermore, data integration with proper data mart construction is necessary, so are proper statistical machine learning methodologies for the classification of potential and actual customers.

The classification and regression tree (CART) (Breiman & Friedman, 1985; Breiman, Friedman, Olshen, & Stone, 1984; Friedman, 1977) is the first well-developed statistical learning method to help groups classify and identify new data from the status quo classification methodologies. From the beginning, scholars do not like applying tree analyses due to a lack of theoretical ground and with several drawbacks on computation. Until the 1980s, scholars have added theoretical methods to the decision-tree-based analysis. Many practical algorithms for pruning and cutting the tree make tree-based statistical helpful learning for analytics. Now, many researchers use the CART method on ordinal, cardinal, and numerical data. Wilkinson (1992) thinks statistical machine learning helps us to know the customer better. With more statistical measures developed with the decision-tree statistical machine learning, we have more valuable tools to identify customers and classify targeted customers for marketing. It brings a new point of view to the business developer (Kim & Loh, 2001). Brodley and Utgoff (1992) further make more incredible progress in introducing the multivariate decision tree with helpful software. Breiman (1995) even considers implementing this tree-structured data analysis with parallel computing in a workstation environment. With proper computation algorithms, parallel computing creates much faster and accurate classification results because of faster tree pruning and cutting. Breiman (1999) works with Cutler to solve missing value imputation, outlier detection, and cluster discovery issues. From 2000 to 2004, Breiman works on data visualization with random forests to have ensemble methods. The random forest has become the standard for data mining studies in the field of data science. Drossos, Papagelis, and Kalles (2000) report a library of decision tree algorithms in Java that shows widespread usage of tree-based learning. All these lead us to have the latest applicable decision-tree learning toolkits.

We use the tree algorithms to have results-based decision-making. Decision tree learning contains several tree-growing algorithms, including CHAID (Chi-Squared Automatic Interaction Detection), Exhaustive CHAID, Classification and Regression Trees (CART) and Quick, Unbiased, Efficient, Statistical Tree (QUEST) (Loh & Shih, 1997) and all with the consideration of pre-pruning and post-pruning to prevent the growth of too many trees. As a result, avoiding overfitting makes the decision-making process to be more robust. With four algorithms, we could quickly handle missing values in data and unbalanced data clustering by ensuring that every tree-related algorithm undergoes rigorous testing before the decision.

For online insurance, statistical machine learning has two purposes: customer classification and the other is for customer behavior clustering. The former is supervised learning via generalized linear modeling, and the latter is unsupervised learning through multivariate classification methods.

From the sales of online insurance in Taiwan, we consider the online purchasing experience of customers. Because of that, a user-friendly website is necessary. It leads to some difficulties for the data analysis because we keep the online input items for customers as little as possible to complete the whole insurance process in seconds. It is a challenge to have data analytics with limited information from the online insurance purchasing process. To have some remedy, we collect some information with Google Analytics for potential customer verification. It is different from the marketing viewpoint because we develop an easy-to-use platform for potential customers to finish their insurance process. The platform only collects needed information for the insurance without considering customer behavior analytics.

Further, with coherent sales data, we provide some helpful data analysis. With the help of statistical machine learning, we have explored some interesting issues. This paper collected actual data for four years to explore the online insurance customer repurchasing behavior using decision-tree modeling, an unsupervised statistical machine learning method. We use travel insurance to analyze the feature and repurchasing behavior of customers. Customers may dissimilarly behave when facing different insurance products, such as savings insurance, annuity insurance, and travel insurance. In the paper, we explore some common drivers of customers' first purchasing experience in online travel insurance, impacting their repurchasing features. We utilize the decision-tree statistical learning model to analyze the common and diverse driving factors from different customer groups.

2. Research Methodology

2.1. Data Description

We build several datasets to have their further consolidation. One is the dataset from the online insurance website. The other is the dataset from the insurance company's database with the necessary items we need for statistical machine learning. Then the data provided from Google Analytics is used for cross verification. This study provides the items of coherent data in Table 1. We do not offer

Table 1. Data description (Coherent but without Classified Information).

Item	Item Description
Insurance Policy number	Internal Code Setting for Identification
Insured Date	Along with specific insurance policy
Insured Time	Along with specific insurance policy
Pay Date	Along with specific insurance policy
Pay Time	Along with specific insurance policy
Insured Processing Start Time	System record item
Insured Processing End Time	System record item
Member Number	The specific number of member to help identify
Days of the Insurance	System record item
Sex	Male, Female
Insured Age	Numerical item
Living Location	Taipei City, New Taipei City, Taoyuan, Hsinchu, Miaoli, Taichung, Changhua, Yunlin, Chiayi, Nantou, Kaohsiung, Tainan, Pingtung, Keelung, Yilan, Hualien, Taitung, Penghu, Kinmen, Mazu
Insured Sum-QADD	Numerical item
Insured Sum-QMR	Numerical item
Insured Sum-OHS	Numerical item
Insured Sum-OHS 1	Numerical item
Payment of Insured Sum-QADD	Numerical item
Payment of Insured Sum-QMR	Numerical item
Payment of Insured Sum-OHS	Numerical item
Payment of Insured Sum-OHS 1	Numerical item
Total Payment of Insurance Policy	Numerical item
Purpose of Insurance	Traveling, Business Trip, Family Visiting
Traveling Location	Domestic, Abroad
System Membership classification	Insured having member, Insured having/Trading member
Different Subscribing Members	New customer with higher insured sum, Old customer with higher insured sum, New Customer with lower insured sum
Insured Status	Unfinished, Finished
Travel Place	Taiwan, Mainland China, East Asia (Japan, South Korea and others), South Asia, America, Europe
Insurance Effective Date	Along with specific insurance policy
Insurance Effective Time	Along with specific insurance policy
Member Registered Date	System record item
Member Registered Time	System record item

any consolidated data items for privacy issues with personal data masking in tables. As we mentioned before, we keep the input items as little as possible to help the whole insurance process be quicker for analysis. Therefore, the items are necessary to support the insurance policy legally binding.

2.2. Data Mining with Visualization

Data mining is the first step with statistical machine learning on raw data for a better understanding when exploring a new business. We have domain knowledge about ordinary customers buying insurance from agents, but we have little knowledge about online insurance potential customers. With proper marketing to let all those potential customers know our products more, we collect and study their purchasing behavior from hand-collected real data. Visualization helps us to have some grasp of different purchasing behaviors with salient features of the customers. We could learn from the information obtained from statistical analytics, including tabulation, item classification, and visualization. Data mining is considered the first step for unsupervised statistical learning, especially for a new business venture, when we do not have enough domain knowledge to build an analytical model. That is why online insurance in Taiwan is suitable for data mining as the first step to exploring the unknown for modeling construction.

Visualization helps in data analytics. Let data speak for itself is the principle of grasping raw data. When the online insurance business started in Taiwan, there were three different categories of customers. The first one is the new customer who has not purchased any insurance policy from an insurance firm before. The second one has purchased insurance policies from insurance agents, not only travel insurance. The third one applies for their online insurance purchasing status from the servicing desks of an insurance firm. Taiwan's financial regulatory authority gives the three categories of customers different assured sums of insurance for the online travel insurance policy. The online travel insurance also provides two insurance packages: for domestic travel, and the other is for overseas travel. Customers could go abroad with a domestic travel package purchased with a limited sum assured of health insurance items. The younger generation commonly uses it because of limited budgets or shorter traveling days.

On travel insurance, we obtain real data from a life insurance company to make the learning reliable. We relate and refer to different data from other insurers to cross-reference the data while applying the analytic methods. Table 2 and Table 3 obtain new information that online travel insurance customers only buy the insurance once in a specific period. It is pretty standard since overseas traveling is usually once a year in Taiwan.

From Table 2 and Table 3, we find that the proportion by headcounts of first-time customers purchasing travel insurance increases gradually with time. The proportion also does not increase that much as expected we plan before. That is the reason for our study. When new insurers join the game and we adopt

Table 2. The primary information for online traveling insurance purchasing.

Time Period	Head counts of buying the online insurance policy only once	Total insured people	Total insured policies (including withdrawn ones)	Head counts of buying insurance more than once	Proportional head counts of buying insurance policy only once
2015/1/1 to 2015/12/31	6307	6863	7801	556	91.90%
2015/1/1 to 2016/12/31	16,974	19,737	24,322	2763	86%
2015/1/1 to 2017/08/29	28,424	33,937	43,300	5513	83.76%

Table 3. The primary information of online insurance purchasing.

Time Period	The increasing people of buying insurance only once	Total increasing people of buying online insurance	Total increasing number of insurance policies
2015/12/31 to 2016/12/31	10,667	12,874	16,521
2016/12/31 to 2017/08/29	11,450	14,200	18,978

new marketing strategies to against with newcomers, all brings a different game to the business. It is the motivation that leads us to explore online travel insurance's repurchasing behavior and period.

From Figure 1, we find that the purchasers of online travel insurance are mainly the younger generation aged between 26 to 40. Besides, from Figure 2 and Table 4, we discover that online travel insurance purchasers appear to be younger females and males. It is similar to purchasing behavior from other online sales where females constitute the leading purchasing group of daily necessities. Males are the primary buyers of fancy products. When we further classify customers according to their living areas, we reconfirm our findings with Figure 3(a), Figure 3(b), Figure 4(a), and Figure 4(b) that our main customers are younger females and males. A convenient digital servicing environment enhances online purchasing behavior. People living in a large city or more developed areas would use online insurance services much often. Due to sales' specific objectives, we find that most online insurance subscribers are mainly for traveling. Only a few are for business trips and family visiting. Online insurance subscribers use these services mainly for overseas traveling instead of domestic ones. It makes the insurance firm want to have marketing strategies for business trips or family visiting customers.

For the same reason, the analysis shows that if we do not take care of potential customers with their first logins on our websites, they would be considered new customers. The budget to catch their first intention is a waste. With statistical learning applied, we find that their repurchasing behavior becomes different after their first logins in a specific period.

2.3. The Exploration of Repurchasing Behaviour

As we mentioned before, an online insurance firm can make the purchaser easy to reach and have a swift online insurance process. It should make the information

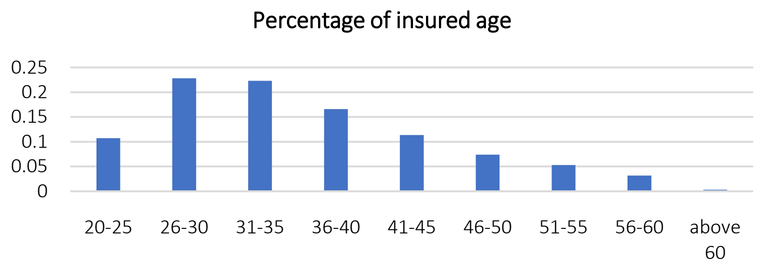


Figure 1. The percentage of insured age (in a form of age range).

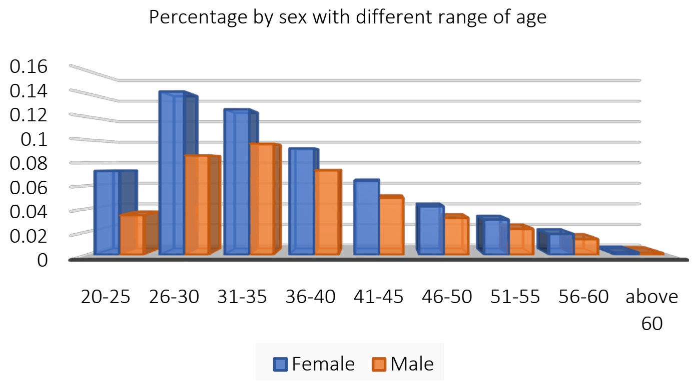
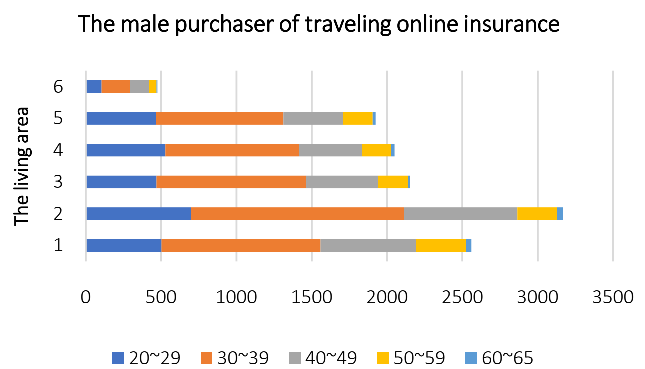
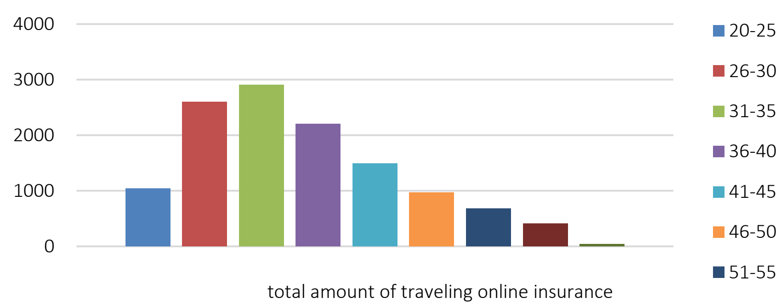


Figure 2. The percentage of insured age by sex (in a form of age range).



(a)

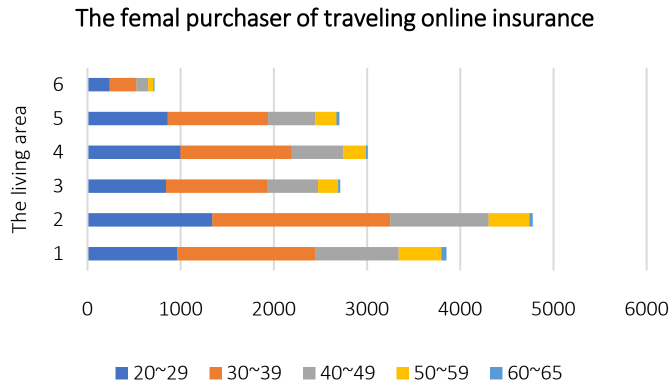


(b)

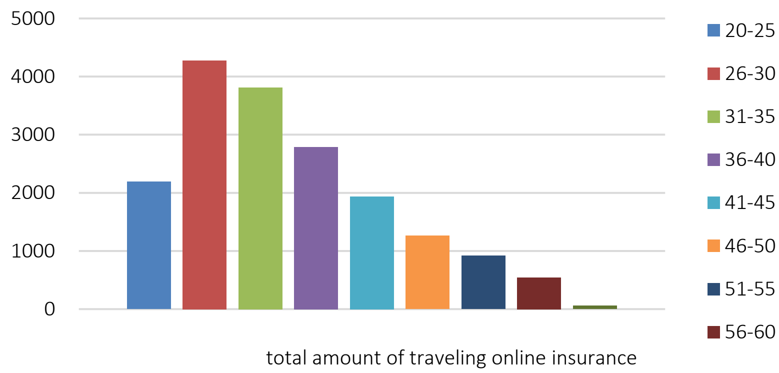
Figure 3. (a) The male purchaser of traveling online insurance (by living area). 1: Taipei City; 2: New Taipei City; 3: Taoyuan, Hsinchu, Miaoli; 4: Taichung, Chiayi, Changhua, Yunlin, Nantou; 5: Kaohsiung, Tainan, Pingtung; 6: Keelung, Yilan, Hualien, Taitung, Penghu, Kinmen, Mazu; (b) The male purchaser of traveling online insurance (by age classification).

Table 4. The percentage in different age range of traveling insurance buyer.

insured age range	20 - 25	26 - 30	31 - 35	36 - 40	41 - 45	46 - 50	51 - 55	56 - 60	above 60
Female	7.28%	14.19%	12.64%	9.25%	6.43%	4.20%	3.06%	1.81%	0.20%
Male	3.44%	8.63%	9.65%	7.33%	4.95%	3.21%	2.25%	1.36%	0.14%



(a)



(b)

Figure 4. (a) The female purchaser of traveling online insurance (by living area). 1: Taipei City; 2: New Taipei City; 3: Taoyuan, Hsinchu, Miaoli; 4: Taichung, Chiayi, Changhua, Yunlin, Nantou; 5: Kaohsiung, Tainan, Pingtung; 6: Keelung, Yilan, Hualien, Taitung, Penghu, Kinmen, Mazu; (b) The female purchaser of traveling online insurance (by age classification).

provided by online customers as little as possible. We only obtain their traveling information and some demographic data to consider if they could get insured or not. Online insurance customers are the younger generation with less insured experience, meaning that we have no prior knowledge about potential customers. In this paper, we want to explore their repurchasing behavior after their first-time subscription. The consecutive period between the first-time and the second-time of subscription is our first issue. The job is to know how long the first-time online travel insurance buyer would come back to repurchase. The second issue is to know the demographics or other features of customers repurchasing online travel insurance. Therefore, we could target first-time purchasers to launch specific marketing activities to attract their repurchases. We conduct the repurchasing model with statistical learning tools on a much higher quality data set that combines valuable old customer database with new purchasing customer information. Table 5 provides the details.

From Table 5, the first issue to consider is to know the repurchasing period between the first-time and the second-time of subscription with possible factors. Furthermore, from the repurchasing model, the second one is to explore reasons behind customer repurchasing behavior with the decision tree statistical learning.

2.4. Data Processing with Analytics

We have several data sources to make them fit into one data mart. We have insurance data from traditional insurance policies and information from online insurance purchasers. We also deploy the Google analytic service to provide some delegated customers' guesses. To make different data sources coherent into one data set or data mart, we need to use more elaborated data processing methods

Table 5. The process of exploring repurchasing behavior.

Process	Repurchasing Behavior
Purpose	We use statistical learning methods, supervised and unsupervised classification analytics, to build our customer repurchasing model for the online insurance business.
Discuss issues	Issue 1. We want to know the repurchasing time period between the first-time and the second-time subscribing and possible factors. Issue 2. From the repurchasing model we build, we want to explore reasons behind customer repurchasing behavior.
Targeted customer	For the first-time subscriber of online traveling insurance, we use the repurchasing model to identify the possibility of repurchasing.
Expected achievement	Use the repurchasing model to identify the potential customers and all those identifications from the model are written as selective rules or conditions.
The procedures of execution	Step 1. To integrate different customer data sets, both new customers with old customer in the old database are included. We may need to perform the missing value replacement or data cleansing to have a higher quality coherent dataset. Step 2. With data mining results, we define new categories for further analytics. We have to design some new variables for the time period between two purchases to explore important factors on the customer repurchasing behavior. Step 3. Perform the decision-tree statistical learning method for specific repurchasing issues to build the repurchasing model with clear rules of classifications. Decision-tree statistical learning method is used on the integrated high quality data set to have classification rules we could follow with further marketing strategy. Step 4. With the combination of unsupervised clustering method and supervised decision tree learning, we rebuild a customer purchasing model on specific groups to explore the possibility of repurchasing behavior. The classification rules and conditions are provided from new repurchase modeling.

to have a much more reliable data set for further analytical usage. After data processing, we deal with the issue of missing values from different data sources. We also use the data masking method to protect customers' privacy, such as identification number, name, and personal health traits. We built coherent data sources from different data sources to meet the requirement that we could not identify customers' names and IDs. This criterion makes data scientists obtain the information easily without considering the breach of customers' privacy. Further, the statistical learning methods could apply the purpose-specific integrated data easily.

In short, data cleansing and integration, as a process, makes statistical learning to be a more powerful tool. We show these from the data processing steps in Figure 5 and Figure 6. In Figure 6, compared with the procedures in Figure 5, we perform the clustering method first to identify several larger clusters and

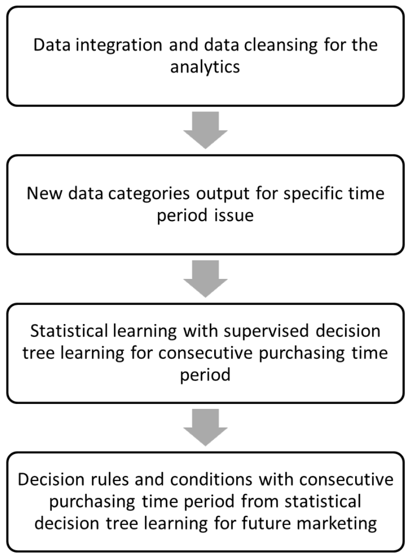


Figure 5. Flow chart of the first issue.

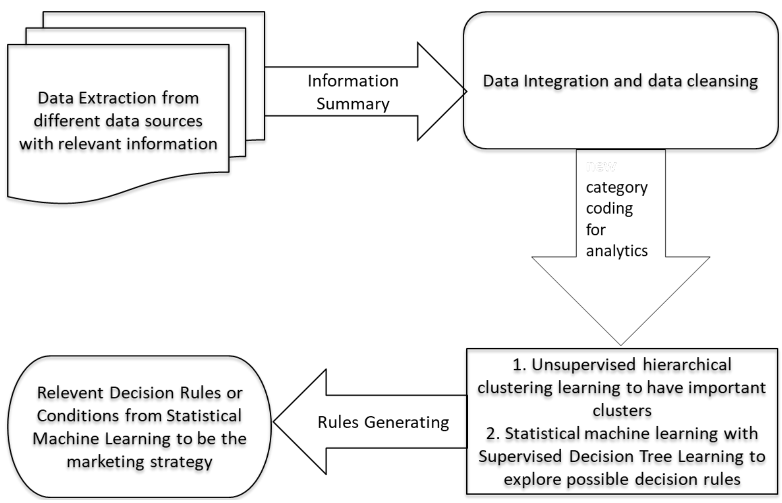


Figure 6. Flow chart of the second issue.

then to explore their features by the statistical decision-tree learning. As to the other issue, we consider the cost of marketing in which we decide to use an unsupervised clustering method first. In order to target the most suitable and profitable customers, we need to have more common features. Statistical decision-tree learning is the way out. Also, we consider the difference between the one-time online insurance buyer and the repurchasing customer to identify their standard features.

We provide some detailed procedures to deal with those two issues. Data integration is the first step to integrate different data sets we have, and data cleansing is the second step when we have many missing values from different data sources. We build a reliable data set or database for further analytical requirements. Further, from the data mining with the first exploratory information, we could use statistical machine learnings to have the periods or rules to explore our first-time purchaser's repurchasing online travel insurance behavior.

3. Results from Statistical Machine Learnings

For the online travel insurance purchasers, we consider the customer who purchases the insurance more than once and studies their purchasing period between the first-time and the second-time purchases. It is the first issue we mentioned in Table 5. We have the process of Figure 5 implemented first, then perform the decision-tree statistical learning method to know the issue better. We show the result in Figure 7.

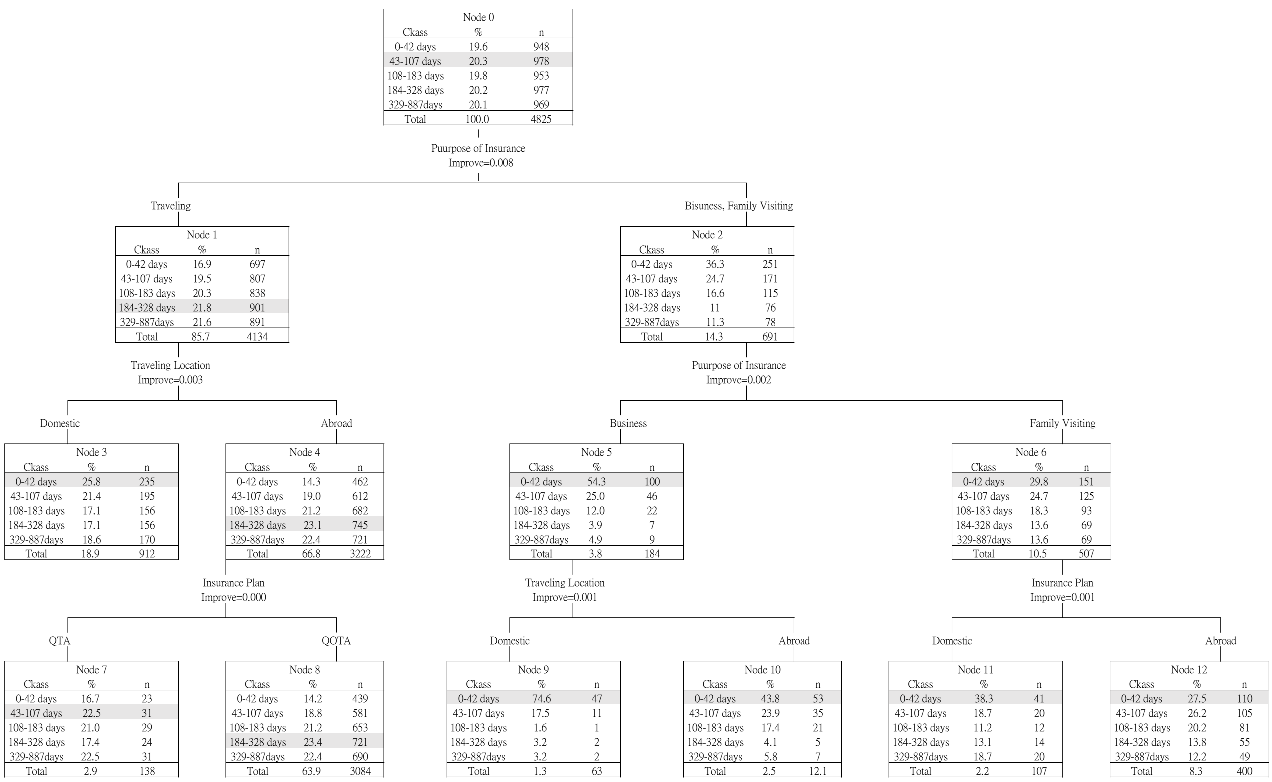


Figure 7. Decision tree for the first issue.

3.1. The Results of the First Issue

This paper forms a decision rule from the mother node to the end node in the decision-tree statistical learning. We have the discretion to judge which rule is more important than the others based on the conditional probability concerning its statistical learning tree-growing algorithm. From **Figure 7**, we know that the number of traveling activities is much greater than visiting and business trips. Further, with the traveling purpose, the consecutive traveling period between the first-time and the second-time for business trips and family visiting is around 0-42 days. People would buy a higher-priced insurance program for overseas traveling purposes, and the consecutive period is around 108 - 328 days.

Nevertheless, for the cheaper insurance package, the consecutive time is around 43 to 108 days shorter than the expensive package. It makes the online insurer have some different marketing strategies with customers of different insurance premiums. Some confounding effects maybe when the first year's marketing strategy is not mature enough to attract login customers. We have adjusted the marketing strategy several times in the first year. However, as we use data with four years, the impact is significantly reduced when there are many insured policies.

With decision tree statistical learning, we have some decision rules with overseas traveling wherein those customers are purchasing online travel insurance once a year. We find that for customers in their business trips in **Figure 7**, the consecutive period of domestic traveling with a cheaper-priced insurance program is around 0 - 42 days. However, the consecutive period of abroad traveling for business trips with a higher-priced program is around 0 - 42 days or 43 - 107 days. It shows that people with a business purpose to have online insurance may visit more than three or four times in a year, meaning that the online insurer should have some marketing programs with business trip customers to increase profit. The online insurer should recognize the customer to classify the trip to be a family trip or a single trip for the customer traveling once a year. We further need to classify if the trip happens around the time of summer vacation or not for a family trip. Therefore, specific marketing programs could launch to increase profit.

3.2. Robustness of First Issue Learning

We perform the decision tree statistical learning with tree algorithms to study the consecutive time between the first-time and second-time online insurance purchasing behavior. The learning is robust, with the algorithm enhancing the pre-pruning and post-pruning in the learning to prevent the overfitting or under-fitting issue. In **Figure 8**, the decision-tree learning with different periods confirms our findings from **Figure 7**. The tree structures are similar in those two figures with equal-numbers of mother nodes as the starting point but with only three categories in **Figure 8**.

Comparing **Figure 7** and **Figure 8**, we have a slightly different result. The

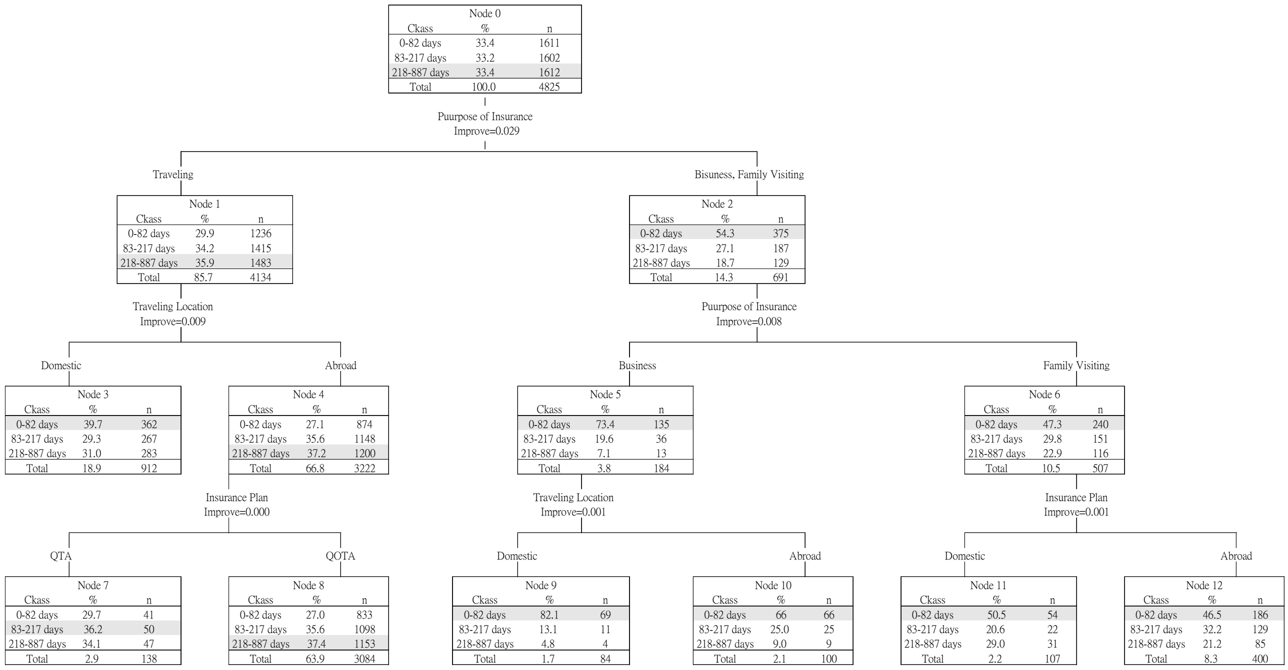


Figure 8. Rubust of decision tree for the first issue.

consecutive time with domestic traveling purposes is around 0 - 82 days. The cross-reference helps bring out a much narrower consecutive period to have a better marketing strategy.

3.3. The Implication of the Marketing Strategy

The first issue studies remind us that if we hope that our marketing strategy succeeds, the contact method with customers is essential. The hot contacting time with the first-time online insurance purchaser is within six months. After that, we should consider the customer as a new customer without receiving our marketing material before. How to help a customer know and build some emotional relations with other customers becomes crucial for the online insurer.

3.4. The Results of the Second Issue

For the second issue, we perform several tailor-made analytics to ensure the following marketing strategy is cost-effective. Therefore, we first consider the different features between a one-time purchaser and repurchasing customer of online travel insurance. From **Figure 9**, customers with family visiting purposes with higher insured sum would repurchase the online travel insurance. Old customers with higher insured sums love to repurchase the online insurance policy without any age distinctions. For the new customer with a higher insured sum, customers with business trip purposes would repurchase. On the other hand, customers with traveling purposes give slight improvement from the decision-tree learning. Only customers living in Taipei City, Chiayi, Hualien, Taitung, Penghu, Kinmen, and Mazu have a higher possibility to repurchase.

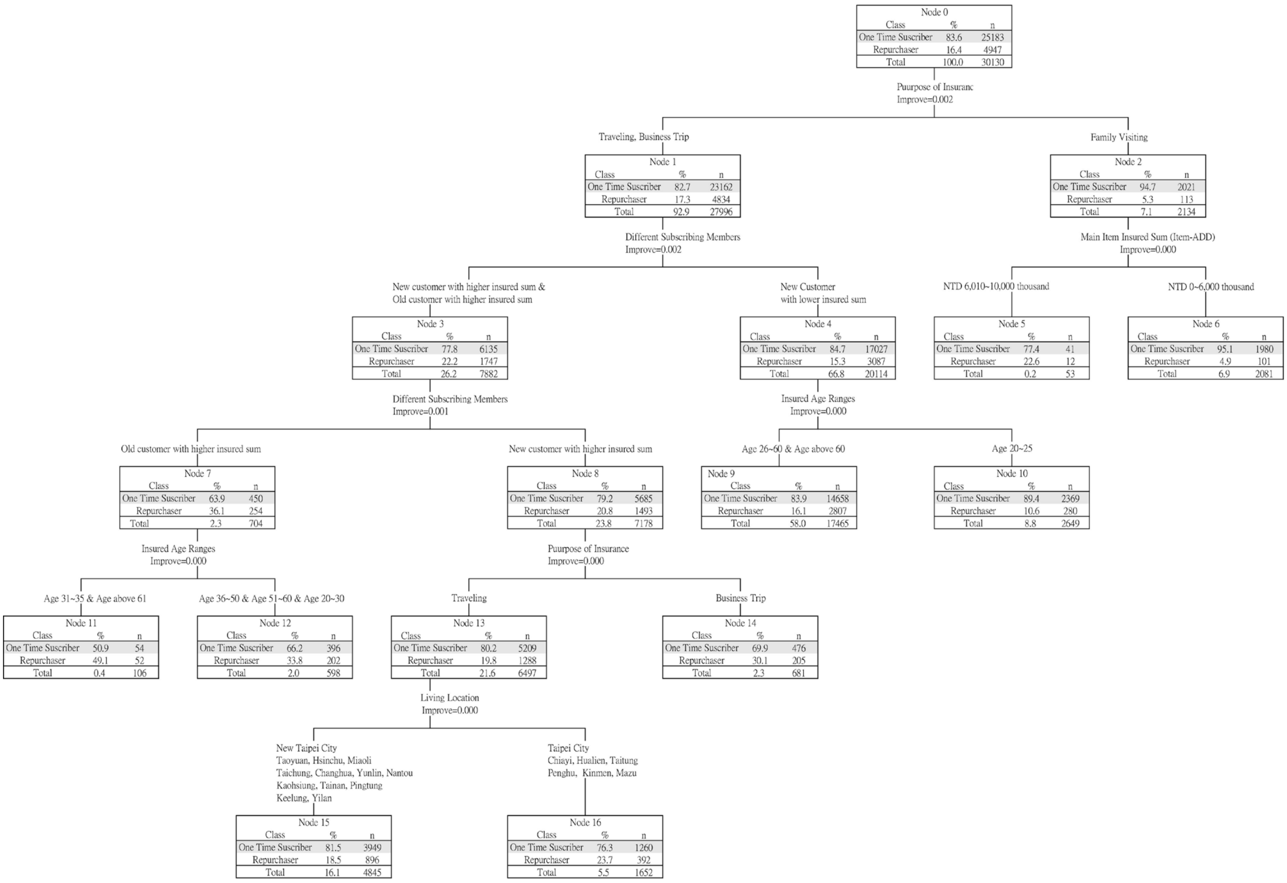


Figure 9. Result of decision tree for the second issue (1).

From Figure 9, we find the decision-tree statistical learning leads to an overfitting problem, and the decision rules are too many to help issues. It allows us to use a clustering method for the whole customers and consider only the two largest cluster groups of customers for further decision-tree statistical machine learning. The goal is to have helpful decision rules and focus on many profitable customers to help the online insurance business grow. Another reason behind this is for a cost-effective marketing strategy because of a limited marketing budget. Therefore, we perform a hierarchical clustering method, unsupervised statistical learning, to help have several classified groups. We choose the two most significant clusters to have the decision-tree statistical learning to know factors affecting customer repurchasing behavior. We provide the most prominent group's decision-tree learning results in Figure 10 and the second largest group in Figure 11.

We document customers repurchasing behavior in Figure 10. The first cluster is as follows 1) customers' age is between 31 and 35; 2) a higher total insurance payment is above NTD 552; 3) they are living in New Taipei City, Hsinchu city and Hsinchu county. Next, the second group is as follows 1) customers live in Taichung, Hsinchu, Nantou, Changhua, Yunlin, and Chiayi City; 2) their age is above 51; 3) A higher total insurance payment is above NTD 555. Then, the

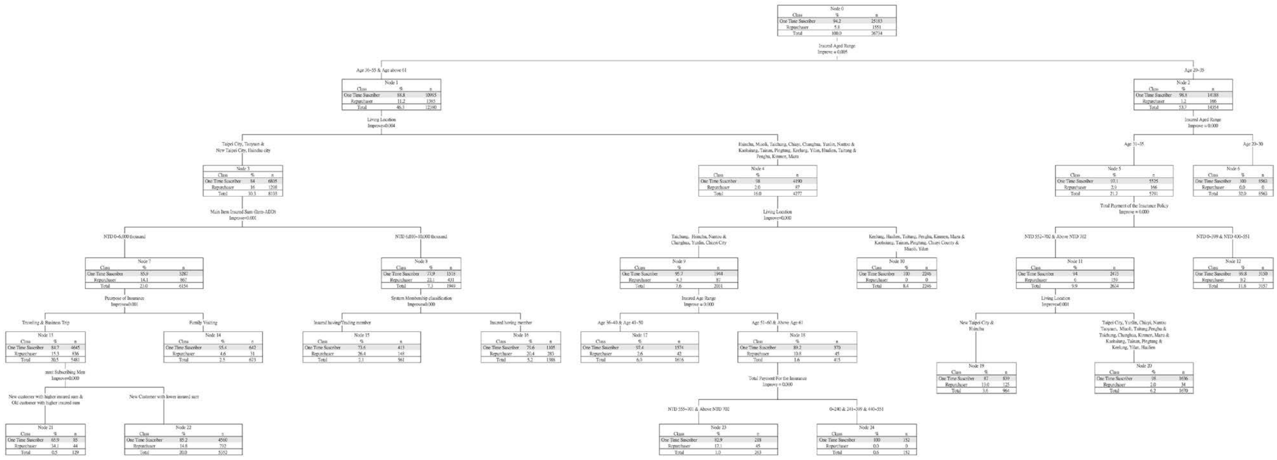


Figure 10. Result of decision tree for the second issue (2).

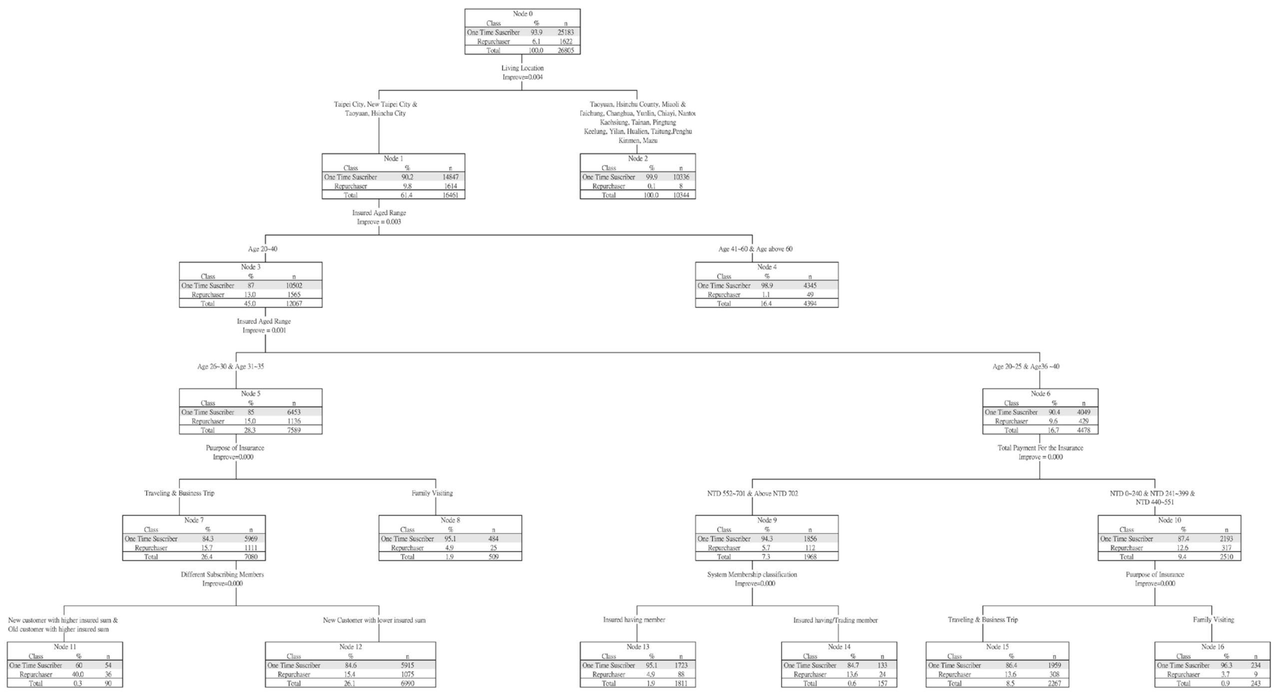


Figure 11. Result of decision tree for the second issue (3).

third group is as follows 1) customers' age is above 36; 2) they live in Taipei City, Taoyuan, New Taipei City, and Hsinchu City; 3) the insured sum is above NTD 6 million. Finally, the fourth group is as follows 1) customers' age is above 36; 2) they are living in Taipei City, Taoyuan, New Taipei City, and Hsinchu City; 3) the insured sum is between NTD 3 to 6 million; 4) customers both the purposes of traveling and business trip; 5) the higher insured sum, the higher repurchasing possibility.

From Figure 11, in the second cluster, the most salient group with repurchasing behavior is 1) customers aged between 26 and 35; 2) customers are to travel and business trip; 3) customer with a higher insured sum.

From Figure 10 and Figure 11, we could target our potential customers with salient features to launch specific marketing strategies to attract them to repurchase. Since we have only limited information about online insurance buyers' knowledge based on a few online subscribing items. We work with Google Analytics to explore potential customers to know their online behavior. Every start node to the end node could be a decision rule with the tree algorithm for marketing, but only a few works well from the conditional probability increasing substantially.

3.5. The Implication of the Marketing Strategy

From the second issue studies, decision-tree-based learning is not enough to understand travel insurance customers' repurchasing behavior. Therefore, we count on unsupervised clustering learning to comprehend the feature of customer repurchasing behavior. The decision rule is with a more meaningful conditional probability increasing for the repurchase group. It is also a rule for advance marketing strategies to be applied.

4. Conclusion

Taiwanese start another online insurance around six years. A new business model is needed to nurture this new business. With artificial intelligence toolkits, online insurance websites could use automated robots to answer customers' questions. On the other hand, the website could be equipped with artificial intelligence to accumulate more customer information when subscribing to online insurance. It is a new emerging business worth more effort to have a new business model with a better solution to customer needs.

In the future, we expect more general insurance (life or non-life) products to be launched online with the help of statistical machine learning algorithms. People could purchase their insurance online in a much quicker and better way without an agent's help. Statistical machine learning will help humans to have a more comfortable life. Scholars (Cevolini & Esposito, 2020; Tanninen, 2020) suggest that insurance policies' pricing issues should be algorithm-driven or behavioral-based. Life insurance companies desire to have, a tailor-made insurance policy for specific customer profiles with different insurance premiums. The algorithmic prediction would become a future research issue when having more diverse customers in which the insurance companies could have tailor-made insurance policies to sell. Taiwanese Health insurance policies with wearable computing devices launched in 2017. Life insurance companies will give a deep premium discount if insurance holders walk ten thousand feet, recorded from a wearable device, each month. It will be some breakthroughs when we have more data from those insurance customers.

The tools of artificial intelligence are not only for storytelling of actual situations but having predictive power. The first research limitation of this research is the predictive ability when we perform artificial intelligence learning with hand-collected actual data—the information structure changes when new regulation or deregulation comes. Lately, the platform of third-party, such as MOMO[1], could sell similar online traveling insurances from different insurers. The third-party platform has a more substantial marketing power when insurers provide several kinds of products or services to customers in time. The second research limitation of this research is the pandemic impact of COVID-19, which makes online traveling insurance extinct during the period. When the environment changing makes a lower predictive power, the modeling method we apply should adapt accordingly.

Acknowledgements

We would like to thank the Editor and the anonymous referees for insightful and constructive comments. We also gratefully acknowledge the financial support of the Ministry of Science and Technology (MOST) of Taiwan (MOST 109-2410-H-126-003).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Barry, L., & Charpentier, A. (2020). Personalization as a Promise: Can Big Data Change the Practice of Insurance? *Big Data & Society, 7,* 1-12. https://doi.org/10.1177%2F2053951720935143

Breiman, L. (1995). Current Research in the Mathematics of Generalization. *Proceedings of the Santa Fe Institute CNLS Workshop on Formal Approaches to Supervised Learning*, Addison-Wesley, 361-368.

Breiman, L. (1999). Pasting Small Votes for Classification in Large Databases and On-Line. *Machine Learning, 36,* 85-103. https://doi.org/10.1023/A:1007563306331

Breiman, L., & Friedman, J. H. (1985). Estimating Optimal Transformations for Multiple Regression and Correlation. *Journal of the American Statistical Association, 80,* 580-598. https://www.jstor.org/stable/2288473 https://doi.org/10.1080/01621459.1985.10478157

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees* (1st ed.). Routledge.

Brodley, C. E., & Utgoff, P. E. (1992). *Multivariate versus Univariate Decision Trees*. Citeseer.

Cevolini, A., & Esposito, E. (2020). From Pool to Profile: Social Consequences of Algorithmic Prediction in Insurance. *Big Data & Society, 7,* 1-11. https://doi.org/10.1177%2F2053951720939228

Cortis, D., Debattista, J., Debono, J., & Farrell, M. (2019). InsurTech. In T. Lynn, J. G. Mooney, P. Rosati, & M. Cummins (Eds.), *Disrupting Finance: FinTech and Strategy in the 21st Century* (pp. 71-84). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-02330-0_5

Diana, A., Griffin, J. E., Oberoi, J. S., & Yao, J. (2019). *Machine-Learning Methods for Insurance Applications—A Survey* (27 p). Society of Actuaries.

[1]A business-to-customer shopping platform in Taiwan.

https://kar.kent.ac.uk/id/eprint/71090

Drossos, N., Papagelis, A., & Kalles, D. (2000). Decision Tree Toolkit: A Component-Based Library of Decision Tree Algorithms. *Paper Presented at the European Conference on Principles of Data Mining and Knowledge Discovery*, Ghent, 14-18 September, 381-387. https://doi.org/10.1007/3-540-45372-5_40

Friedman, J. H. (1977). A Recursive Partitioning Decision Rule for Nonparametric Classification. *IEEE Transactions on Computers, C-26*, 404-408. https://doi.org/10.1109/TC.1977.1674849

Graham, T. (2018). Platforms and Hyper-Choice on the World Wide Web. *Big Data & Society, 5,* 1-12. https://journals.sagepub.com/doi/10.1177/2053951718765878 https://doi.org/10.1177%2F2053951718765878

Hand, D. J., & Henley, W. E. (1997). Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 160,* 523-541. https://www.jstor.org/stable/2983268 https://doi.org/10.1111/j.1467-985X.1997.00078.x

Henley, W., & Hand, D. J. (1996). AK-Nearest-Neighbour Classifier for Assessing Consumer Credit Risk. *Journal of the Royal Statistical Society: Series D (The Statistician), 45,* 77-95. https://www.jstor.org/stable/2348414 https://doi.org/10.2307/2348414

Kim, H., & Loh, W.-Y. (2001). Classification Trees with Unbiased Multiway Splits. *Journal of the American Statistical Association, 96,* 589-604. https://doi.org/10.1198/016214501753168271

Kumar, R., & Verma, R. (2012). Classification Algorithms for Data Mining: A Survey. *International Journal of Innovations in Engineering and Technology (IJIET), 1,* 7-14. https://ijiet.com/wp-content/uploads/2012/09/2-1.pdf

Lee, T.-S., Chiu, C.-C., Chou, Y.-C., & Lu, C.-J. (2006). Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines. *Computational Statistics & Data Analysis, 50,* 1113-1130. https://doi.org/10.1016/j.csda.2004.11.006

Loh, W.-Y., & Shih, Y.-S. (1997). Split Selection Methods for Classification Trees. *Statistica Sinica, 7,* 815-840.

Luo, C., Chen, Q., Zhang, Y., & Xu, Y. (2019). The Effects of Trust on Policyholders' Purchase Intentions in an Online Insurance Platform. *Emerging Markets Finance and Trade,* 1-18. https://doi.org/10.1080/1540496X.2019.1695122

Parodi, P. (2012). Computational Intelligence with Applications to General Insurance: A Review. I—The Role of Statistical Learning. *Annals of Actuarial Science, 6,* 307-343. https://doi.org/10.1017/S1748499512000036

Tanninen, M. (2020). Contested Technology: Social Scientific Perspectives of Behaviour-Based Insurance. *Big Data & Society, 7,* 1-14. https://doi.org/10.1177%2F2053951720942536

Wahbeh, A. H., Al-Radaideh, Q. A., Al-Kabi, M. N., & Al-Shawakfa, E. M. (2011). A Comparison Study between Data Mining Tools over Some Classification Methods. *International Journal of Advanced Computer Science and Applications, 8,* 18-26.

Wilkinson, L. (1992). Tree Structured Data Analysis: AID, CHAID and CART. *Sawtooth/SYSTAT Joint Software Conference*, Su Valley, July 1992. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.165.3625

Yu, T.-W., & Chen, T.-J. (2018). Online Travel Insurance Purchase Intention: A Transaction Cost Perspective. *Journal of Travel & Tourism Marketing, 35,* 1175-1186. https://doi.org/10.1080/10548408.2018.1486781