Implementation of Classifiers for Choosing Insurance Policy Using Decision Trees: A Case Study

CHIN-SHENG HUANG[1], YU-JU LIN[2], CHE-CHERN LIN[3]

1: Department and Graduate Institute of Finance
National Yunlin University of Science and Technology

2: Ph.D. Student, Department and Graduate Institute of Finance
National Yunlin University of Science and Technology &
Department of Finance, Fortune Institute of Technology

3: Department of Software Engineering
National Kaohsiung Normal University

TAIWAN

huangcs@yuntech.edu.tw[1]; kitty@center.fotech.edu.tw[2]; cclin@nknucc.nknu.edu.tw[3]

Abstract: - In this paper, we use decision trees to establish the decision models for insurance purchases. Five major types of insurances are involved in this study including life, annuity, health, accident, and investment-oriented insurances. Four decision tree methods were used to build the decision models including Chi-square Automatic Interaction Detector (CHAID), Exhaustive Chi-square Automatic Interaction Detector (ECHAID), Classification and Regression Tree (CRT), and Quick-Unbiased-Efficient Statistical Tree (QUEST). Six features were selected as the inputs of the decision trees including age, sex, annual income, educational level, occupation, and risk preference. Three hundred insurants from an insurance company in Taiwan were used as examples for establishing the decision models. Two experiments were conducted to evaluate the performance of the decision trees. The first one used the purchase records of primary insurances as examples. The second one used the purchase records of primary insurances and additional insurances. Each experiment contained four rounds according to different partitions of training sets and test sets. Discussion and concluding remarks are finally provided at the end of this paper.

Key-Words: - Insurance policy; Decision tree; Decision model; ECHAID; CRT; CHAID; QUEST; Classification tree; Decision support system.

1. Introduction

In insurance business, the possible insurance purchasers might have different backgrounds such as gender, education, age, income, risk preference, etc. For insurance consultants, how to select an appropriate insurance policy for their customers has become an important issue in recent insurance-related studies. Decision models for determining insurance policy have been established to help insurance consultants choose the best insurance policies for their customers. Mainly there are five types of insurances in the insurance business including life, annuity, health, accident, and investment-oriented insurances. The five insurances vary in different aspects such as purpose, coverage, and period of benefit return. In the past, insurance consultants recommended the insurance policies for their customers by experience. It might result in prejudice because no decision tools can help them to make the decisions.

The purpose of life insurance is to compensate the death of an insurant. Therefore in a life insurance, an insured amount should be specified when an insurant purchases a life insurance. The returns of the benefits depend on pre-specified conditions (live or dead situation of the insurant). In annuity insurance, an insurant periodically receives benefits from the insurer in his life time. The annuity insurance now becomes a popular insurance for those people who want to combine the functions of retirement program and insurance. The purpose of health insurance is to cover medical expenditures. The policyholder of a health insurance receives compensation to make up the loss due to his illness. The accident insurance pre-specifies a certain compensation amount to covert the loss of an insurant's injury due to an accident. The

investment-oriented is a new insurance product combining the functions of insurance and investment. The policyholder should take the investment risk and the insured amount is based on the earning of investment.

Recently, with the fast development of information technology, artificial intelligence techniques have been used to solve real world problems in many fields. Commonly utilized artificial intelligence techniques include fuzzy logic, data mining, neural networks, genetic algorithms, decision trees, etc. Fuzzy logic basically uses membership functions to describe gradual changes of belonging relationship between elements and a fuzzy set. It has been widely utilized to solve uncertainty problems. Data mining discovers deep knowledge from a data set using association rules. The apriori algorithm is one of popular data mining techniques used for basket analysis. Neural networks use neurons and weights to mimic the brain structure and the cognition processing of human beings. A neural network basically uses a training algorithm to optimize its weights. This procedure is called a training procedure. Many neural network models have been developed to solve real-world problems. The back-propagation algorithm is one of the most popular neural network methods. It is widely used in solving classification, estimation, and prediction problems. A genetic algorithm imitates the Darwin's theory of evolution to search the optimal solution using genes and chromosomes. A fitness function is used to measure the performance of genetic computing. Biologic evolution processes are involved in genetic computing such as mutation, crossover, and reproduction. A decision tree is a tree-structured decision model established by a set of production rules. We will discuss it in details later in this paper.

In this paper, we use decision trees to establish decision models for the five insurances: life, annuity, health, accident, and investment-oriented insurances. Four decision tree methods were used to build the decision models including Chi-square Automatic Interaction Detector (CHAID), Exhaustive Chi-square Automatic Interaction Detector (ECHAID), Classification and Regression Tree (CRT), and Quick-Unbiased-Efficient Statistical Tree (QUEST). Six features were selected as the inputs of the decision trees including age, sex, annual income, educational level, occupation, and risk preference. A data set of three hundred insurants from an insurance company in Taiwan was used as examples for building the decision trees. Two experiments were conducted to evaluate the performance of the decision trees. The first one used the purchase records of primary insurances as examples. The second one used the purchase records of primary insurances and additional insurances. Each experiment contained four rounds according to different partitions of training sets and test sets. Discussion and concluding remarks are finally provided at the end of this paper.

2. Related Works and Preliminaries

Classification is a technique to categorize samples (instances) based on the features of the samples. Basically, to solve classification problems, we first establish a classification model based on well-known samples and then utilize the classification model to classify (categorize) unknown samples. Many classification methods have been developed to solve real world problems. The neural network method is one of commonly used techniques in solving the classification problems [1-3]. Basically, a neural network is established by using a hierarchically layered structure. Each of layers in a neural network consists of several nodes. Weights are utilized to link two nodes in adjacent layers. Training algorithms are used to update the weights in a neural network to get desired outputs. This is called a training procedure. After completing the training procedure, the unknown instances are classified by the well-trained neural network. Popular neural network models include feed forward neural network with back-propagation algorithm, self-organizing map, winner-take-all network, etc [4]. The computational details about these neural network models can also be found in [4].

The technique of decision tree (or classification tree) is another method to implement real world classification problems based on statistical approaches. Basically, a decision tree is a hierarchical tree-like structure with nodes and links. Tree-generating algorithms are used to build decision trees. Popular tree-generating algorithms include CHAID, ECHAID, CRT, and QUEST. Further explanation about the fundamental principle of tree generation and introduction of the four tree-generating algorithms will be discussed later in this section.

Shapiro indicated that neural networks, fuzzy logic, and genetic algorithms were three commonly used artificial intelligence techniques to solve insurance-related problems [5]. According to his paper, neural networks are widely utilized in classification problems (i.e., insurance claim frauds), prediction of financial crises or bankrupts, and medical care issues [5]. Fuzzy logic can be utilized in pricing strategies, asset evaluation, investment

policy, underwriting, and classification problems [5]. Genetic algorithms are employed in allocation of assets, insurance competitiveness optimization, and classification problems [5]. Huang et al. presented evaluation models for choosing insurance policy [6]. The evaluation models used a hybrid model consisting of Analytical Hierarchy Process (AHP), fuzzy logic, and the Delphi technique to determine the purchases of five major insurances: life, annuity, health, accident, and investment-oriented insurances [6]. Huang et al. further presented an empirical study using the hybrid model proposed in [6] where 300 purchase records were utilized to validate the hybrid model [7].

A decision tree is a hierarchically tree-structured decision model to classify patterns using nodes and links. Each of nodes in a decision tree represents an attribute of the decision model. Links derived from a node indicate the values or categorical items for this particular node. In general, a decision tree contains several input variables and one single output variable. Two types of nodes are used in a decision tree: terminal nodes and non-terminal nodes. Decision trees have been widely used to solve pattern classification problems in many fields [8-15].

A decision tree is established using a set of production rules. Basically, a production rule consists of two parts: an ***IF*** part and a ***THEN*** part. It is important to note that the ***IF*** part can contain several input variables while the ***THEN*** part contains a single output variable only. It is also important to mention that the inputs variables and the output variable are not reciprocal. It means the input variables can not appear in the ***THEN*** part and the output variable cannot appear in the ***IF*** part. Below, we present a simple example to explain how a decision tree is established by a set of production rules.

Consider a decision tree to classify the favor exercise (output variable) with three input variables: sex, age, and annual income. The input and output variables are described as follows:

Input variables

- Sex: categorical variable with two items: male (M) and female (F).
- Age: categorical variable with four items: less than 20 (<20), between 20 and 40 (20~40), between 41 and 60 (41~60), and higher than 60 (>60).
- Annual Income: categorical variable with five items: less than 25K (<25K), between 25K and 45K (25K~45K), between 46K and 65K (46K~65K), between 66K and 85K (66K~85K), and higher than 85K (>85K)

Output variable:

- Favor exercise: Categorical variable with 7 items: soccer, badminton, pingpong, yoga, bicycling, walking, and tennis.

Figure 1 demonstrates an exemplary decision tree where a gray node represents a terminal node and a white node indicates a non-terminal node. Each of terminal nodes is associated with a production rule.

The overall production rules related to the decision tree are shown as follows:

- ***IF*** "age > 60" THEN the favor exercise is **walking** (Node 4).
- ***IF*** "age < 20" AND sex = "M" ***THEN*** the favor exercise is **soccer** (Node 5).
- ***IF*** "age < 20" AND sex = "F" ***THEN*** the favor exercise is **badminton** (Node 6).
- ***IF*** age = "20~40" AND sex = "M" ***THEN*** the favor exercise is **tennis** (Node 7).
- ***IF*** age = "41~60" AND sex = "F" ***THEN*** the favor exercise is **bicycling** (Node 9).
- ***IF*** age = "20~40" AND sex = "F" AND income = "25K~45K" ***THEN*** the favor exercise is **pingpong** (Node 11).
- ***IF*** age = "20~40" AND sex = "F" AND income = "46K~65K" ***THEN*** the favor exercise is **yoga** (Node 12).
- ***IF*** age = "41~60" AND sex = "M" AND income ">85K" ***THEN*** the favor exercise is **golf** (Node 13).

Even thought there are many different types of decision trees, the fundamental principle of establishing a decision tree is the same. Below, we introduce a simple method of establishing a decision tree, called the C4.5 method [16].

The procedure of the C4.5 method [16]

Give a training set T.

1. Determine a parent node which is the most discriminative variable from the candidate list of variables in *T*.
2. Establish child links for the parent node. Each of the child links represents a value (or a categorical item) for the parent node. Divide T into subsets according to the values (items) of the child links.
3. For each of subsets produced in Step 2,

(i) If it fits the predefined conditions or if there is

no remaining input variables to be determined, keep this path as a qualified production rule.

(ii) Otherwise, go to step 1.

The C4.5 algorithm is a simple decision tree method. Below, we briefly introduce the four popular decision tree methods used in this paper: CHAID, ECHAID, CRT, and QUEST [16].

Based on the C4.5 method, CHAID employs the Chi-square test to build a decision tree starting from the most differential variable among the input variables. The input variables of the CHAID method are limited to categorical variables due to the Chi-square test. ECHAID is a modified version of CHAID, which exhaustively computes classification accuracies for all of possible combinations of tree architectures and then picks up the best one. In general, an ECHAID method obtains better classification result but spends more computational time than a CHAID method. CRT can deal with both categorical and numerical input variables. Basically, it recursively divides training data into two groups and hence continuously enlarges the scale of the decision tree. In CRT, goodness criteria are used to optimize the homogeneity of the output variable. A pruning process is also employed to decrease the complication of the tree architecture, preventing the tree from over-fitting problems. The QUEST method is a fast algorithm to generate a decision tree. It can remove the biases which are probably caused by other algorithms. It is disadvantaged with the limitation of using categorical input variables.

3. Experiments and Discussions

The data were collected from an insurance company in Taiwan. Three hundred insurants were selected as the samples for the experiments. We used decision trees to establish decision models for the five insurances: life, annuity, health, accident, and investment-oriented insurances. The decision tree models used in this study are CRT, ECHAID, CHAID, and QUEST. Six attributes were used as the input variables for the decision trees including age, sex, annual income, educational level, occupation, and risk preference. They are described as follows:

- Age: encoded in year.
- Sex: categorical, 1 for male; 2 for female.
- Annual income: Encoded by a unit of 10,000 NTDs (New Taiwan Dollars).
- Educational level: categorical with 9 values:
  1: Elementary school;
  2: Junior high school;
  3: Senior high school;
  4: Vocational high school;
  5: Junior college or community college;
  6: Technical college;
  7: University;
  8: Master degree;
  9: Doctoral degree.
- Occupation: categorical, ranked in the ascending order of the occupational risks from 1 (lowest) to 6 (highest).
- Risk preference: categorical with values from 1 (lowest risk preference) to 10 (highest risk preference).

The insurants might simultaneously purchase multiple insurances including primary insurances and additional insurances. We conducted two experiments to establish the decision trees for determining the insurance policy for the five insurances. In Experiment 1, we used the transaction records of purchasing primary insurances as examples. In Experiment 2, we used the records of purchasing primary and additional insurances as examples. We divided the data into two sets: a training set and a test set. Each experiment contained four rounds according to different data partitions described as follows:

Round 1: Using all examples as a training set.
Round 2: Using 3/4 of examples as a training set and the rest of 1/4 as a test set.
Round 3: Using 2/3 of examples as a training set and the rest of 1/3 as a test set.
Round 4: Using 1/2 of examples as a training set and the rest of 1/2 as a test set.

Table 1 shows the sample sizes in Experiments 1 and 2. The sample sizes of life, annuity, and investment-oriented insurances in Experiment 1 are the same as those in Experiment 2.

Tables 2 and 3 show the results of Experiments 1 and 2, respectively. In Tables 2 and 3, symbol "—" denotes an un-appropriate decision tree in which all of samples are classified into a single class, i.e., either class 1 (purchasing insurance) or class 0 (not purchasing insurance).

Of the five insurances involved in this paper, the investment-oriented is the easiest one to establish the decision tree for classifying the purchase behaviors. The most difficult one to build the decision tree is the annuity insurance. The reason for that might be the data sizes buying an annuity insurance (N= 31) and not buying an annuity insurance (N= 269) of an annuity insurance are extremely un-equal. This will cause the decision built by the un-equal sizes of the training data set classifies all samples to be a single class (the class

having more sample which is the class not buying an annuity insurance). Observing the classification results of health and accident insurances in Table 2 gives the same conclusion.

As mentioned early, ECHAID is an enhanced version of CHAID by selecting the best decision tree from all possible tree architectures. Theoretically, ECHAID might get better classification results but spend more computational time than CHAID. In this study we applied statistical methods to analyze the classification performance between the CHAID and ECHAID methods using the classification accuracies shown in Tables 2 and 3. The statistical hypothesis is shown as follows:

H: The classification accuracies are different between the CHAID and ECHAID methods.

The statistical procedure for the hypothesis is described as follows

Step 1: Test the correlation between the classification accuracies between the CHAID and ECHAID methods using Pearson product-moment correlation method.

Step 2: Use the paired samples t-test.

Table 4 shows the Pearson correlation coefficient obtained by Step1. The correlation is significant at the 0.05 level with a two-tailed significance test. This result supports us to perform the test in Step 2. Table 5 shows the results of paired samples statistics. Table 6 displays the results of paired samples test on the pair of CHAID–ECHAID. From Table 6, we reject the hypothesis and conclude that the classification accuracies are not different between the CHAID and ECHAID methods.

Form the above discussion and the classification results in Tables 2 and 3, we conclude:

- The classification accuracies are not different between the CHAID and ECHAID methods.
- The decision tree model is a suitable technique to build the decision model for classifying investment-oriented insurance purchases.
- In establishing the decision trees for health and accident insurances, using the purchase records of primary insurances and additional insurances is better than using those the purchase records of primary insurances.

Figures 2-5 show the recommended decision trees for life, health, accident, and investment-oriented insurances, respectively.

4. Conclusions

We used decision trees to establish the decision models for purchasing insurances. Five major types of insurances were involved in this study including life, annuity, health, accident, and investment-oriented insurances. Six features were selected as the inputs of the decision trees including age, sex, annual income, educational level, occupation, and risk preference. Three hundred insurants from an insurance company in Taiwan were used as examples for establishing the decision models.

Two experiments were conducted in this study. The first one used the purchase records of primary insurances as examples. The second one used the purchase records of primary insurances and additional insurances. Each experiment contained four rounds according to different partitions of training sets and test sets. Four decision tree methods were used in the experiments including CRT, ECHAI D, CHAID, and QUEST. The concluding remarks for the experiments are drawn as follows:

- The classification accuracies are not different between the CHAID and ECHAID methods.
- The decision tree model is a suitable technique to build the decision model for classifying investment-oriented insurance purchases.
- In establishing the decision trees for health and accident insurances, using the purchase records of primary insurances and additional insurances is better than using the purchase records of primary insurances.

As about the direction for future studies, it might be a good research topic to use some techniques to overcome the problem of un-equal sizes. Taking less samples from the class with a large sample size might be a possible approach to do that. This is called sub-sampling. How to get a suitable sub-sampled set to appropriately represent the original data might be the key point of a successful sub-sampling.

Reference:

[1] B. Watanapa, J.H. Chan, Neural network classification of extended control chart patterns, *WSEAS Transaction on Computers*, Issue 1, Vol. 6, 2007, pp. 160-166.

[2] G.. Munjal, S. Kaur, Comparative study of ANN for pattern classification, *WSEAS Transaction on Computers*, Issue 2, Vol. 6,

2007, pp. 236-241.
[3] P. Kraipeerapun, C.C. Fung, K.W. Wong, Uncertainty assessment using neural networks and interval neutrosophic sets for multiclass classification problems, *WSEAS Transaction on Computers*, Issue 3, Vol. 6, 2007, pp. 463-470.
[4] R.P. Lippmann, "An introduction to computing with neural nets," *IEEE Acoustic, Speech, and Signal Processing (ASSP) Magazine*, April, 1987, pp. 4-12.
[5] A. F. Shapiro, The merging of neural networks, fuzzy logic, and genetic algorithms, *Insurance: Mathematics and Economics*, Vol. 31, 2002, pp. 115-131.
[6] C. Huang, Y. Lin, C. Lin, evaluation models for choosing insurance policy using the AHP, fuzzy logic, and Delphi technique, *7th Int. Conf. in Applied Computer & Applied Computational Science*, Hangzhou, China, April 6-8, 2008, pp. 696-703.
[7] C. Huang, Y. Lin, C. Lin, Determination of insurance policy using a hybrid model of AHP, fuzzy logic, and Delphi technique: a case study", *WSEAS Transaction on Computers*, Issue 6, Vol. 7, 2008, pp. 463-470.
[8] B.-S. Yang, D-S, Lim, A.C.C. Tan, VIBEX: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table, *Expert Systems with Applications*, Vol. 28, 2005, pp. 725-742.
[9] J. Wang, S. Chan, Stock market trading rule discovery using two-layer bias decision tree, *Expert Systems with Applications*, Vol. 30, 2006, pp. 605-611.
[10] L.F. Mendonca, S.M. Vieira, J.M.C. Sousa, Decision tree search methods in fuzzy modeling and classification, *Intel. Journal of Approximate Reasoning*, Vol. 44, 2007, pp. 106-123.
[11] E.M. Mugambi, A. Hunter, G.. Oatley, L. Kennedy, Polynomial-fuzzy decision tree structures for classifying medical data, *Knowledge Based Systems*, Vol. 17, 2004, pp. 81-87.
[12] F.P. Sarasin, Decision analysis and its application in clinical medicine, *European Journal of Obstetrics & Gynecology and Reproductive Biology*, Vol. 94, 2001, pp. 172-179.
[13] N. Indurkhya, S.M. Weiss, Estimating performance gains for voted decision trees, *Intelligent Data Analysis*, Vol. 2, 1998, pp. 303-310.
[14] P. Deng, Using case-based reasoning approach to the support of ill-structured decisions, *European Journal of Operational Research*, Vol. 93, 1996, pp. 511-521.
[15] K. Tsujino, Implementation and refinement of decision trees using neural networks for hybrid knowledge acquisition, *Artificial Intelligence in Engineering*, Vol. 9, 1995, pp. 265-275.
[16] R.J. Roiger, M.W. Geatz, *Data Mining: a Tutorial-based Primer*, Addison-Wesley, New York, USA, 2003.

Table 1: Sample sizes of Experiments 1 and 2.

	Life	Annuity	Health	Accident	Investment -oriented
Exp. 1	143	31	60	22	156
Exp. 2	143	31	182	169	156

Table 2: Classification accuracies of Experiment 1 (primary insurances only)

Method	Round		Life	Annuity	Health	Accident	Investment -oriented
CRT	1	Overall samples	—	—	—	—	65.7%
	2	Training (3/4 of samples)	—	—	—	—	68.0%
		Test (1/4 of samples)	—	—	—	—	57.7%
	3	Training (2/3 of samples)	—	—	—	—	65.0%
		Test (1/3 of samples)	—	—	—	—	67.5%
	4	Training (1/ 2 of samples)	—	—	—	—	63.8%
		Test (1/ 2 of samples)	—	—	—	—	61.5%
ECHAID	1	Overall samples	—	—	—	—	65.7%
	2	Training (3/4 of samples)	—		—	—	64.5%
		Test (1/4 of samples)	—	—	—	—	69.4%
	3	Training (2/3 of samples)	65.2%	—	—	—	58.2%
		Test (1/3 of samples)	48.0%	—	—	—	50.9%
	4	Training (1/ 2 of samples)	66.2%	—	—	—	63.8%
		Test (1/ 2 of samples)	53.3%	—	—	—	60.0%
CHAID	1	Overall samples	—	—	—	—	65.7%
	2	Training (3/4 of samples)	—	—	—	—	67.8%
		Test (1/4 of samples)	—	—	—	—	52.8%
	3	Training (2/3 of samples)	59.5%	—	—	—	60.3%
		Test (1/3 of samples)	48.0%	—	—	—	65.6%
	4	Training (1/ 2 of samples)	58.2%	—	—	—	64.1%
		Test (1/ 2 of samples)	43.5%	—	—	—	58.7%
QUEST	1	Overall samples	—	—	—	—	66.0%
	2	Training (3/4 of samples)	—	—	—	—	68.8%
		Test (1/4 of samples)	—	—	—	—	53.7%
	3	Training (2/3 of samples)	—	—	—	—	68.8%
		Test (1/3 of samples)	—	—	—	—	58.4%
	4	Training (1/ 2 of samples)	—	—	—	—	70.0%
		Test (1/ 2 of samples)	—	—	—	—	61.3%

Remark: "—" denotes an un-appropriate decision tree in which all of samples are classified to a single class.

Table 3: Classification accuracies of Experiment 2 (primary insurances and additional insurances)

Method	Round		Health	Accident
CRT	1	Overall samples	—	—
	2	Training (3/4 of samples)	—	—
		Test (1/4 of samples)	—	—
	3	Training (2/3 of samples)	—	—
		Test (1/3 of samples)	—	—
	4	Training (1/ 2 of samples)	—	—
		Test (1/ 2 of samples)	—	—
ECHAID	1	Overall samples	—	65.0%
	2	Training (3/4 of samples)	—	62.9%
		Test (1/4 of samples)	—	53.5%
	3	Training (2/3 of samples)	65.2%	66.7%
		Test (1/3 of samples)	54.9%	47.9%
	4	Training (1/ 2 of samples)	—	65.8%
		Test (1/ 2 of samples)	—	54.7%
CHAID	1	Overall samples	—	61.3%
	2	Training (3/4 of samples)	—	62.5%
		Test (1/4 of samples)	—	57.4%
	3	Training (2/3 of samples)	59.1%	66.7%
		Test (1/3 of samples)	52.9%	47.9%
	4	Training (1/ 2 of samples)	—	62.8%
		Test (1/ 2 of samples)	—	61.3%
QUEST	1	Overall samples	—	—
	2	Training (3/4 of samples)	—	—
		Test (1/4 of samples)	—	—
	3	Training (2/3 of samples)	—	—
		Test (1/3 of samples)	—	—
	4	Training (1/ 2 of samples)	—	—
		Test (1/ 2 of samples)	—	—

Remark: "—" denotes an un-appropriate decision tree in which all of samples are classified to a single class.

Table 4: The results of paired samples correlations

Pair	N	Correlation	Sig.
ECHAID - CHAID	20	.543	.013

Table 5: The results of paired samples statistics

Pair	Mean	N	Std. Deviation	Std. Error Mean
ECHAID	60.090	20	6.796	1.520
CHAID	58.805	20	6.701	1.498

Table 6: The results of paired samples test

Paired Differences (CHAID-ECHAID)			t	Sig. (2-tailed)
Mean	Std. Devi.	Std. Error Mean		
1.285	6.453	1.443	.891	.384

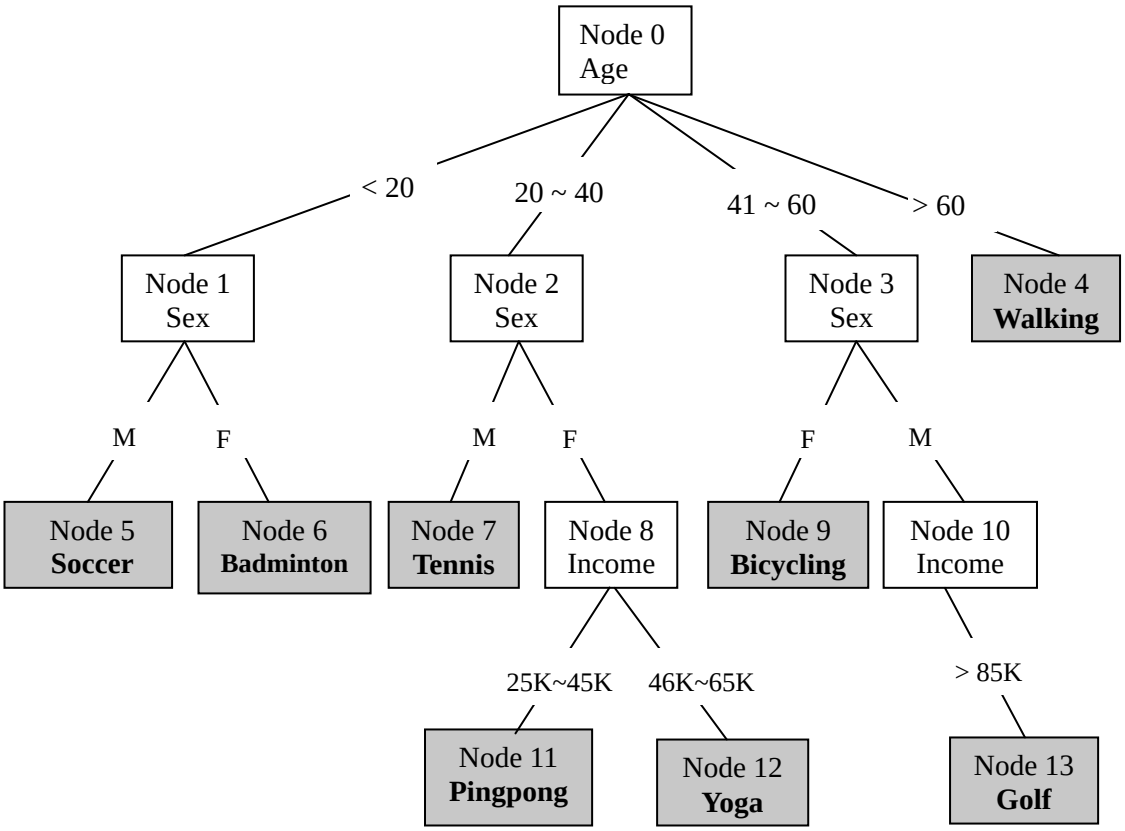
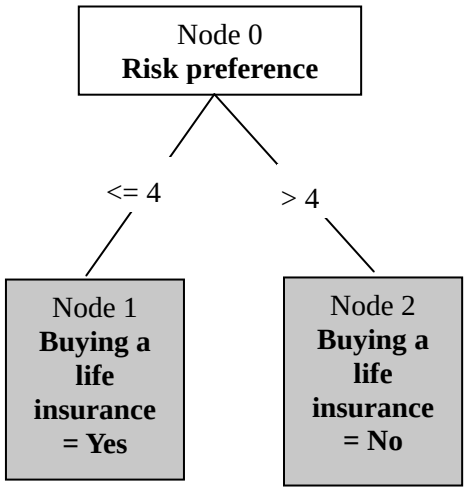


Figure 1: An example of decision tree.

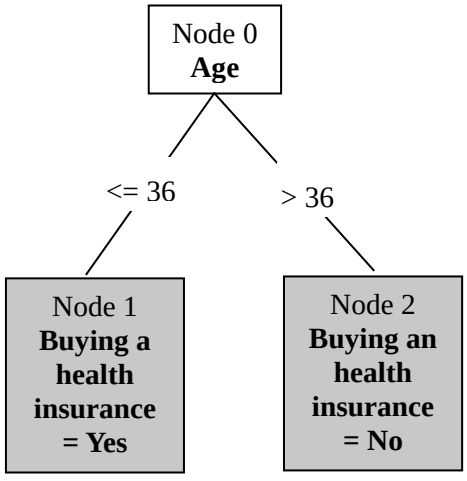



```
/* Node 1 */
IF (Risk preference <= 4)
THEN
  Buying a life insurance = Yes

/* Node 2 */
IF (Risk preference > 4)
THEN
     Buying a life insurance = No
```

(a) The tree (b) The decision rules

Figure 2: The decision tree and associated decision rules for buying a life insurance

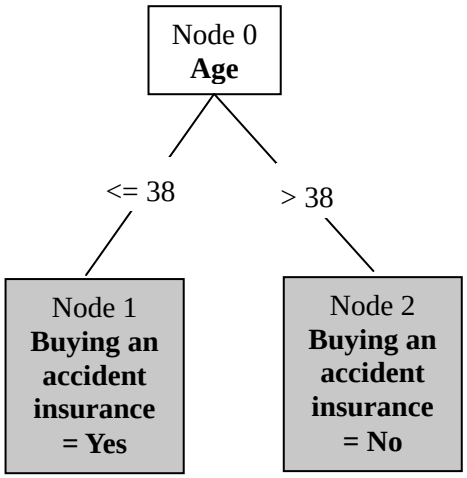



```
/* Node 1 */
IF (age <= 36))
THEN
  Buying an health insurance = Yes

/* Node 2 */
IF (age > 36)
THEN
       Buying an health insurance = No
```

(a) The tree (b) The decision rules

Figure 3: The decision tree and associated decision rules for buying a health insurance



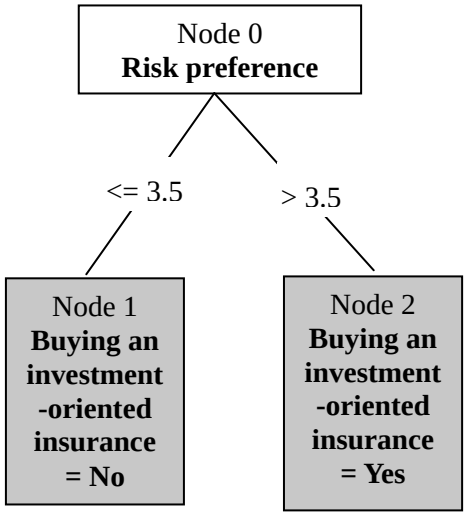
(a) The tree

```
/* Node 1 */
IF (age <= 38))
THEN
  Buying an accident insurance = Yes

/* Node 2 */
IF (age > 38)
THEN
      Buying an accident insurance = No
```

(b) The decision rules

Figure 4: The decision tree and associated decision rules for buying an accident insurance



(a) The tree

```
/* Node 1 */
IF (Risk preference <= 3.5)
THEN
  Buying an investment-oriented
     insurance = No

/* Node 2 */
IF (Risk preference > 3.5)
THEN
     Buying an investment-oriented
     insurance = Yes
```

(b) The decision rules

Figure 5: The decision tree and associated decision rules for buying an investment-oriented insurance