A Machine Learning Approach to Cold Call Insurance Prediction

By: Sana Bahari

Under guidance of Professor Asil Oztekin

Abstract:
Increasing demands in Data-Ming approaches makes organizations to use data driven tactics in decision making. This paper studies customer behavior in buying car insurance with using data-mining approaches. Data-mining technique helps to discover unknown pattern in dataset and lead to superior results. The steps that are involved in Data-Ming includes Business Understanding, Data Understanding, Data preparation, Modeling, Evaluation and deployment. Moreover, knowledge discovery stemming from Data Mining techniques results in more stable and reliable results which, in turn, addresses the current issues in the organization and, ultimately, offers proper approach for managerial decision making. The studied dataset is from a bank who also sells insurance to customers. This articles studies 4 different classification models to compare and specify which model predicts results precisely. Random data splitting with 80/20 percent and 10-fold cross-validation are performed to select between models.

Keywords: Car insurance, Cold Calls Performance, Client Response, Supervised Methods, IBM SPSS Modeler, Neural Network

1. Introduction

1.1 Motivation

In the car insurance industry, the amount of customers a company has greatly determines its success. Even though car insurance is an essential for every car owners in the country, each car insurance company needs to reach out to those owners. The companies are not able to solely rely on the customers to look for their services due to the high amount of available supplies. Thus, in order to increase their customers base, they need to connect with as many people as possible. One way to achieve that is to have a designated department, Who is responsible for getting new customers. Their duties are to contact people through various platforms and convince them to purchase the services the company has to offer. One way to reach customers is to randomly call as many as potential clients as possible. This process is often called cold calling. The company's representatives do not know whether they will lend the client or not. Therefore, it is important to identify variables that has high influence on the outcome. This can help allocate company's resources to focus on the essential variables. Furthermore, it will save times and cut

costs for each gain of customers. With the ever growing market for predictive analytics, car insurance companies are sitting on goldmines of data at their disposal. Thus, by employing different mining techniques, they can stay ahead of the competitions. As emphasized by Asma and Alshamsi [1], "… since these databases are so quite big to be processed manually, the computation tools were needed to process such data and extract the important information".

1.2. Literature Review

Many researchers have studied the car insurance using predictive analysis. However, most of such studies try to predict and assess the risk of the car insurance policies and also to determine customers behaviors based on policies.

Kašćelan et al. conducted a study on risk prediction in car insurance using nonparametric data mining approach [2]. They used techniques such as clustering, support vector regression, and kernel logistic regression. By using these techniques, Kašćelan et al. [2] attempted to classify risk and predict claim size based on data, in order to help the insurer to accurately assess the risk and calculate actual premiums. In the article, they report "These huge quantities of data hide very important information, which could contribute to easier decision-making and risk assessment in car insurance". Data mining is capable of extracting this important information and it can also justify the investments of insurance companies in data. Kašćelan et al. [2] use nonparametric methods from modern statistical machine learning and data mining theory such as support vector regression and kernel logistic regression. Their main reason for using these instead of generalized linear models is because they believe their a restriction to these models. These generalized models are limited to only linear forms, which they believe is too rigid for this type of studies . From their studies, they concluded that the level of premium does not determined based on tariff cases. They were able to produce an 80% prediction accuracy.

Asma and Alshamsi [1] conducted a study on predicting car insurance policies using random forest. The aim of their studies is to predict the choice of a customer for each insurance policies option. In their study , they used Random Forest technique to predict customers' decisions regarding the choice of insurance policy options. Their main argument for using random Forest approach is because it is found to perform better than other classifiers having many advantages such as its robustness against over fitting, and reasonable computing times and it is very user-friendly. In the study, the data is divided into two sets, training and testing sets.

Within those two sets, the data is divided into seven groups based on the policies. Once the data has been processed, random tree uses a method called bagging to build the model. According to Asma and Alshamsi [1], the accuracy result using random forest was above 98% for all different sets of policies. Using his result, new insurance policies can be easily predicted. Furthermore, Asma and Alshamsi [1] considered to use clustering method to find out different patterns that clustered the policies.

In another study, D'Arcy and Stephen [3], set their project's goal to generate predictive models to enhance the insurance claim investigation practices. To achieve this goal, their research examine almost one-half million of data sets for patterns of claim behavior. The predictive model assists in identifying which claims will generate cost saving by investing the claim more extensively. By doing so, the insurer will be able deter fraudulent claims by customers. The method D'Arcy and Stephen [3] used to study and produce model is called Data to Knowledge (D2K) developed by NCSA. This method was able to fill in missing data using random distribution. The challenge of study by D'Arcy and Stephen [3] is that the model does not provide a clear result and analysis of the outcome.

Thakur and Sing [4], conducted a study on *Mining Customer's Data for Vehicle Insurance Prediction System using Decision Tree Classifier*. Their goal is to use decision tree classifier to predict the class label of unknown records. Thakur and Sing [4] in their evaluation process used methods such as holdout method, random sub sampling, cross-validation and bootstrap. Their model had an accuracy of 82% and error rate of 17.66.

2. Methodology

The steps that are followed on this project are based on CRISP-DM methodology. Such method is selected because it is well known and highly accepted in the Data-Mining realms. The method constitutes of six steps starting with Business Understanding and continues to the last step, i.e. Deployment. In each step, Data-Mining techniques improves the decision making and pave the way to next step. In general CRISP-DM methodology includes these steps:

First Step: Business Understanding, which helps to identify why the particular type of business needs Data-Mining approach for the decision making. In this Project we selected the Car Insurance dataset. In the context of car insurance industry, there is a substantial competition

between insurance companies and, hence, it is vital for the organizations to follow new methodology to remain alive and relevant in the market.

Second step: Data-Understanding; it is crucial to gain insight about nature of data-set before modeling and processing. Understanding each variable leads to increased efficiency of taking the correct approach in Data-Mining. This data-set is collected from Kaggle [5], and was a part of competition which was held in 2017.

Third step: Data Preparation, which is performed by different techniques to transform raw data to proper and usable dataset for constructing Data-Mining models. This step identifies the percentage of missing values, deleting or replacing such values, finding the outliers and dealing with them, and correlation between variables to identify only independent predictors to be used in Data-Mining Models. More detailed will be described in next sections [6].

Fourth step: Modeling. After cleaning and arranging original data set, is a time to choose proper Data-Mining models to make prediction. In this project, we study a supervised problem with classification response and, therefore, the Data-Mining models are chosen based on their popularity in these areas. As mentioned before, Data-Mining tries to suggest new patterns which are not discovered beforehand. Consequently, it is useful to try models which have not been deployed to find new pattern.

Fifth step: Evaluation. This step includes the comparison of results from various model to come up with the best choice. Moreover, analyzing the variable importance based on their sensitivity is performed during this step. Information Fusion Sensitivity Analysis provides the best approach to choose the most significant predictors among all the variables.

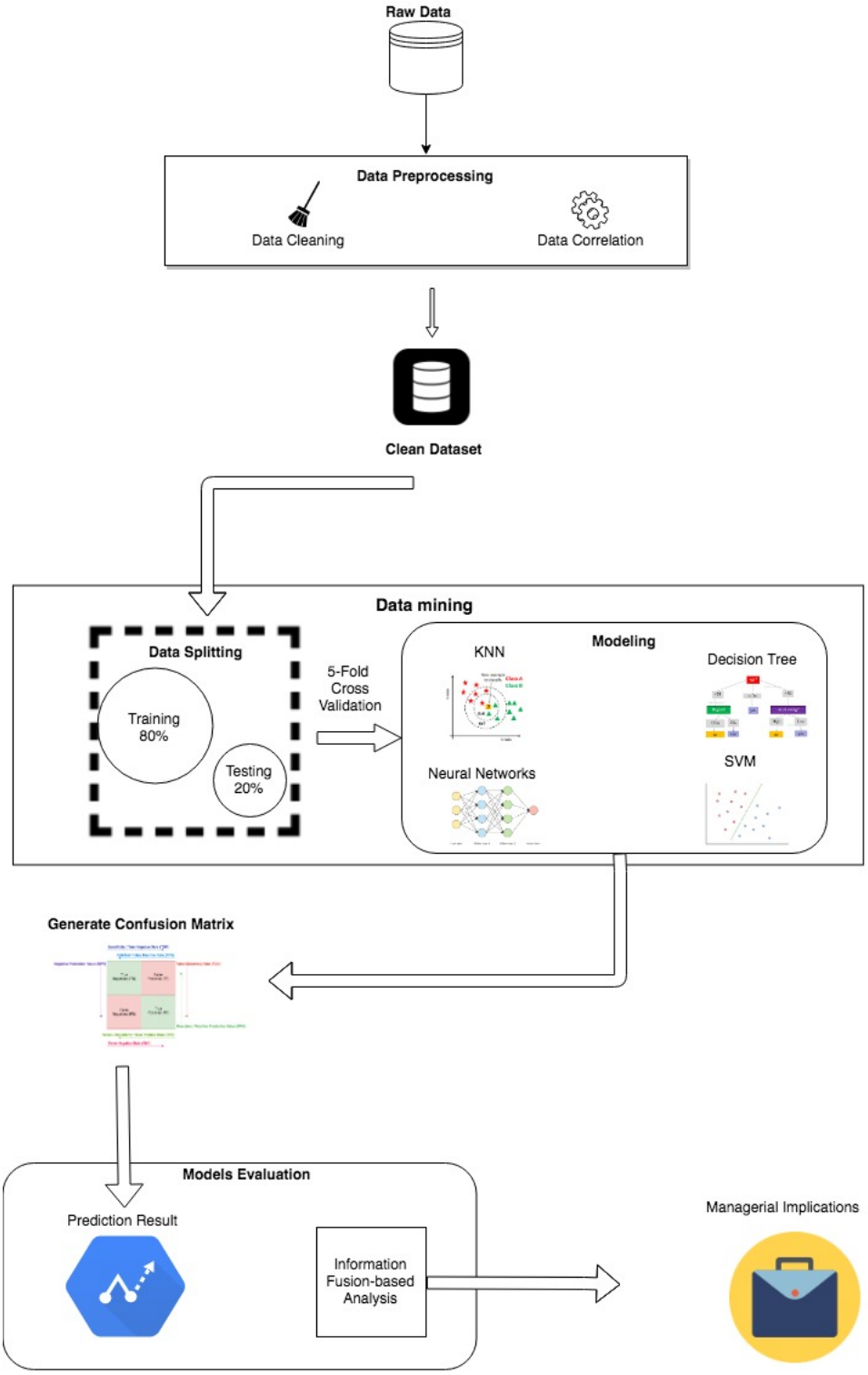

FIGURE 1. Flowchart of the methodology used in this project.

Sixth step: Deployment. This step includes reporting all the results from Data-Mining models to select the best decision and practice for managerial purpose. Therefore, user decides when and how deploy the Data-Mining results to the business organization to save time and money through Data-Mining approaches [7].

2.1. Data description

As mentioned before, the data set which is used for this study is obtained from Kaggle website [5]. It is originally collected from a bank in the United States, which also sells car insurance to the customers. The bank arranges different marketing campaigns to identify new clients. This data set are information from previous campaigns and includes client information as well as other variables which will be described in next step. Original dataset includes 1000 records and 19 columns, each column will be described briefly in the following table 1. Please note the types are based on IBM SPSS Modeler software.

Table 1. Dataset Variables with Type and Description

Variable	Type	Description
ID	Numeric	A unique number which is assigned to a client. This variable does not have any impact on our prediction models, so it is ignored
Age	Numeric	Customer age
Job	Nominal	Customer job
Marital	Nominal	Marital status of client
Education	Ordinal	client education level
Default	Numeric	Whether the customer has been default in credit before or not
Balance	Numeric	Average yearly balance in USD
HHInsurance	Numeric	Whether the household is insured or not
Carloan	Numeric	Whether the client has a car loan or not
Communication	Nominal	Contact communication type

LastContactMonth	Nominal	Month of the last contact
LastContactDay	Numeric	Day in month of the last contact
CallStart	Numeric	Start time of the last call (HH:MM:SS)
CallEnd	Numeric	End time of the last call (HH:MM:SS)
NoOfContacts	Numeric	Number of contacts performed during this campaign for this client
Daypassed	Numeric	Number of days passed by after the latest time the client has been contacted from a previous campaign
PrevAttempts	Numeric	Number of contacts performed before this campaign and for this client Outcome: Outcome of the previous marketing campaign
Outcome	Nominal	Outcome of the previous marketing campaign
CarInsurance	Numeric	Whether the client has subscribed to a car insurance

The variable age indicates the age of clients and it ranges from 18 to 95. Job variable is categorical and nominal with 12 levels of management, blue-collar, student, technician, admin, self-employed, service, retired, entrepreneur, unemployed, housemaid and others. The marital variable is also categorical and nominal with 3 levels of married, single, divorced. The education variable is categorical and ordinal with 3 levels of primary, secondary and tertiary. Default variable is a numeric flag type with either 1or 0 as indicator of whether the client has been default or not based on available history. The balance is a numeric type showing yearly balance in USD in the bank. HHinsurance is numeric flag with either 1 or 0 showing whether the household of the client is insured or not. Carloan is another flag with 1 or 0 corresponding to, respectively, whether the client has or has not car loan based on the historic data. Communication is categorical and nominal variable with 2 levels of cellphone or telephone. LastContactMonth is a categorical and nominal variable with 12 levels of all months year-round. LastContactDay is numeric that ranges between 1 to 31 based on the day of the month the

contact has been made to the client. CallStart and CallEnd are variables to identify the time of start and ending of the phone call conversation and are both numeric type. NoOfContacts is numeric variable and indicates the number of contacts to the client. Daypassed is numeric variable showing the days after the last contact. PrevAttempts is a numeric variable and shows how many times during previous campaigns the client has been contacted. Outcome is categorical and nominal variable with success and failure levels and indicates the outcome of previous campaigns in terms of convincing the client to buy the insurance. Finally, CarInsurance is the response of the dataset and is a binary numeric with 1 or 0 corresponding to, respectively, the decision of client to buy the insurance it decline the offer. It should be noted that in IBM SPSS Modeler, the numeric type we used is either continuous or flag depending on the nature of the variable. For instance, the variable "default" that has only two levels, i.e. 0 or 1, so it is set as flag in data type in the Measurement tab. Conversely, the variable "daypassed" is considered as a continuous numeric in the software.

The column "Outcome" column had 73 percent missing values, and after analyzing available options to deal with this predictor, it is decided to keep the variable but refilling the missing parts with the "Failure" option due to its highly repeated frequency. Otherwise, the original data set is considered to be clean enough to be employed in the modeling.

Also in the preprocessing step, it is desired to assess whether all predictors are independent and there is no correlation between them. We used the correlation matrix for such analysis. The correlation matrix refers to the symmetric array of numbers [8]

$$\mathbf{R} = \begin{pmatrix} 1 & r_{21} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

where the elements r_{jk} are Pearson correlation coefficients between variables x_j and x_k

$$r_{jk} = \frac{s_{jk}}{s_j s_k} = \frac{\sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ik} - \overline{x_k})}{\sqrt{\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2}\sqrt{\sum_{i=1}^{n}(x_{ik} - \overline{x_k})^2}}$$

It can be shown that, using Cauchy-Schwarz inequality [9], that $|r_{jk}| \leq 1$.

To construct R, the categorical data should be converted into numeric ones. Therefore, dummy variables are introduced to transform the non-numerical data as the enabling method to

correlation analysis. The categorical values which are converted to dummy variables are “Job”,” Marital”, Education”, “Communication”, “LastContactMonth”, and “Outcome”. Once dummy numeric values are assigned to such variables, the new dataset is loaded through RStudio for run and to find the correlation between the variables.

As illustrated in Figure 2, the correlation matrix results show that the off-diagonal elements in R is much smaller than unity other than that of “outcome”. Therefore, in the modeling step we remove such variable. The small values of off-diagonal elements, other than “outcome”, implies that there is no significant correlation among other variables, and as a result, all variables minus “outcome” are used for the Data-Mining models.

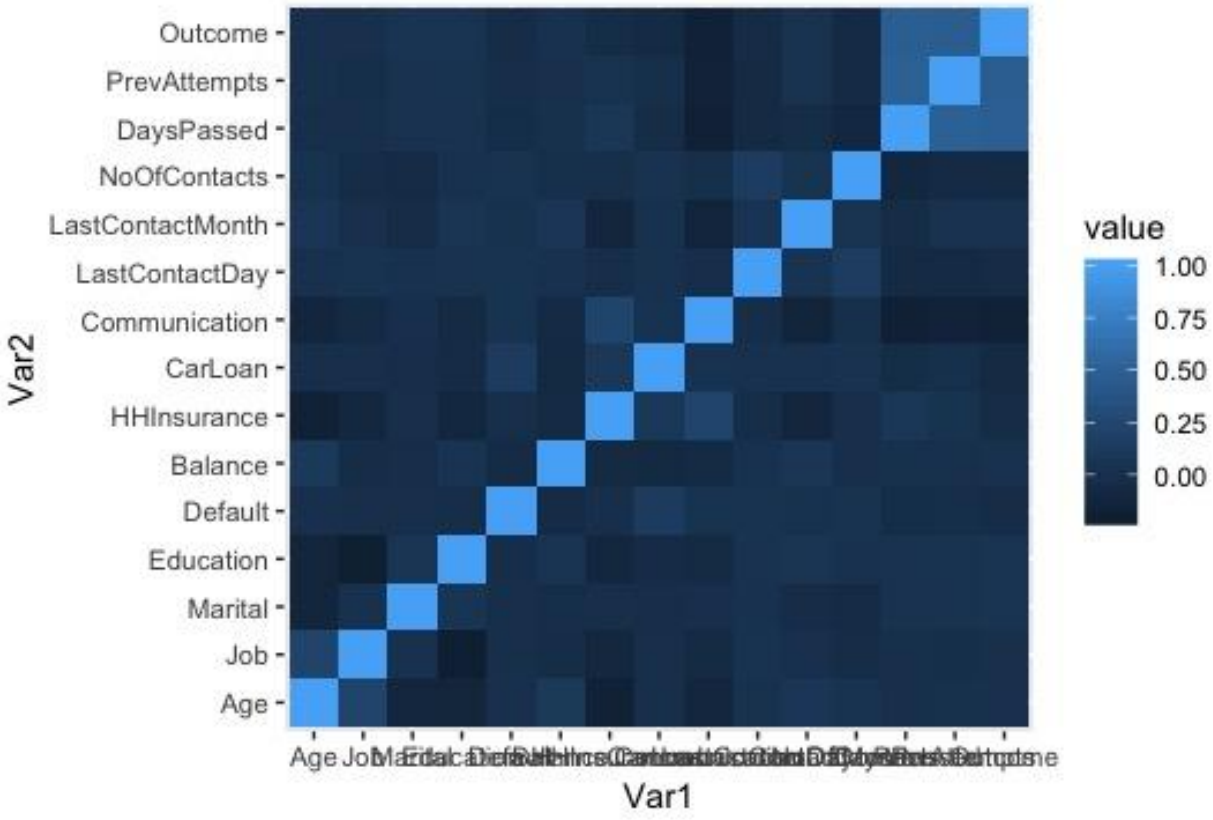

FIGURE 2. The correlation matrix for Car Insurance dataset. Language R is used to generate this figure.

2.2. k-Fold Cross-validation

In the absence of a very large designated test dataset, which is often the case in practical situations, estimating the test error rate needs special techniques. Such techniques exploit the

already available training data to estimate the error and accuracy or sensitivity of the model. One could employ a mathematical apparatus to adjust a model to training data and estimate the error (subset selection, model reduction, etc.) or, instead, estimate the error by holding out a subset of the training observations from the modeling, and then applying the model to those held out observations. We focus on the latter class of methods for error estimation and accuracy analysis. In this class methods such as validation set approach, Leave-one-out cross-validation (LOOCV), and k-Fold Cross-Validation, and the Bootstrap are suggested in literature [10]. We focus on k-Fold cross-validation method for a number of reasons; it computationally more economical, and it is often more accurate and has less variance and only slightly more bias compared to LOOCV [10]. Therefore, we employ the popular k-Fold Cross-Validation approach in this paper for model evaluation purpose. In this study, we set k=5.

k-Fold Cross-Validation approach involves splitting the set of observations into k groups, or folds, of approximately equal size. As illustrated in Figure 2, the first fold is treated as a validation set, and the model is constructed based on on the remaining k − 1 folds. We then calculate the accuracy, sensitivity, and specificity for such held-out observations, say, I_1. We continue such process for other folds, and for each process a separate I_i is calculated. Consequently, we generate k number of I_i at the end of the algorithm. The overall value of this method is computed by averaging these values,

$$I_{(n)} = \frac{1}{k}\sum_{i=1}^{k} I_i$$

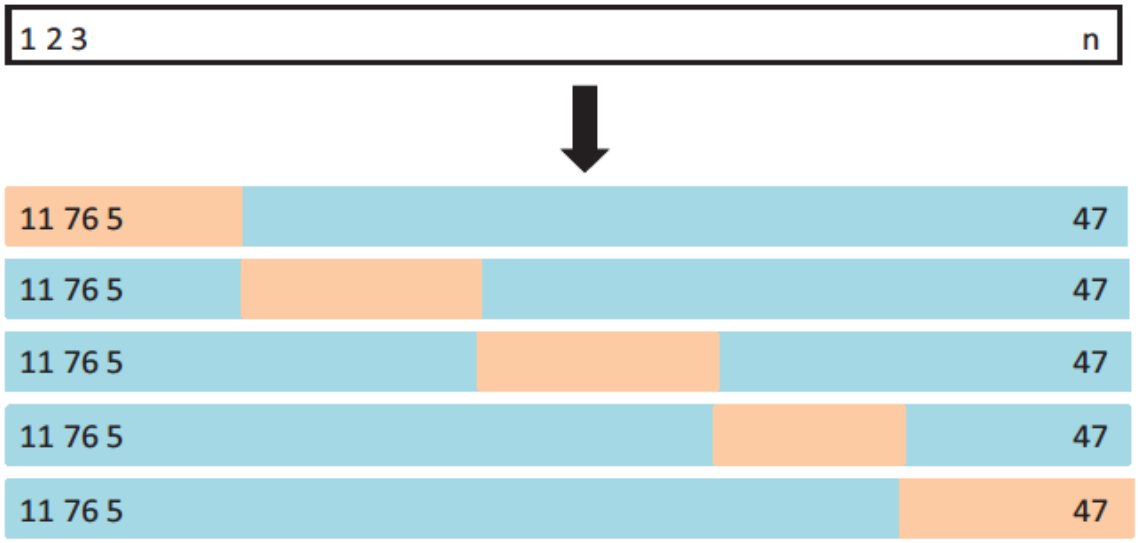

FIGURE 3. A schematic display of 5-fold cross-validation. Image Courtesy of [10]. N is the number of observations. The total set of observations is divided into k non-overlapping groups, namely folds. Each fold serves as a validation set (colored in beige), and the remainder as a training set (colored in blue). To Ensure accuracy, all k folds quantities are finally averaged.

2.3. Mining Methods Employed

Four Data-Mining methods are used in this project along with k-Fold Cross validation with k=5 to ensure the accuracy of the predicted values. The models are K nearest neighbors (KNN), decision tree, support vector machine (SVM), and neural networks. Such models are popular in prediction of supervised classification problems; however, we keep in mind that "there is no free lunch" in statistics that is no single method stands above all other method over all possible datasets [11].

The following definitions are used throughout this paper to define accuracy of classification models.

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

$$\text{True Negative Rate} = \frac{TN}{TN + FP},$$

$$\text{True Negative Rate} = \frac{TP}{TP + FN},$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

With TP, TN, FP, and FN are, respectively, true positive, true negative, false positive, and false negative values. IBM SPSS Modeler is used to implement each model and the details of such implementation is discussed in the following.

2.3.1. KNN Method

Nearest neighbor algorithms are based on the idea that similar (likely) cases are close to each other than dissimilar (less likely) cases. It is inspired by Bayes classifier, which proves the average test error rate for a classification problem is minimal if each observation is assigned to the most likely class, for a given set of variables. In practice, however, the distribution of response for the given input is not known. Many approaches instead try to estimate the conditional probabilistic distribution of output for a given input and the assign the observation to the most likely estimated class; K-nearest neighbors (KNN) classifier is among such approaches [12].

In KNN approach the conditional probability is estimated by using K, a positive integer, points in training data in the neighborhood of the observation for which response is a desired value over the total number of points in that neighborhood. It then applies Bayes rule, i.e. classifies the test observation to the class with the largest probability. It has been observed that KNN results are surprisingly close to that of Bayes classifier, which is the most optimal classifier. For instance, in [13] it has been proved that “For sufficiently large training set size n, the error rate of the 1NN classifier is less than twice the Bayes error rate.” As an example of such “closeness” of the two method, is demonstrated in Figure 4.

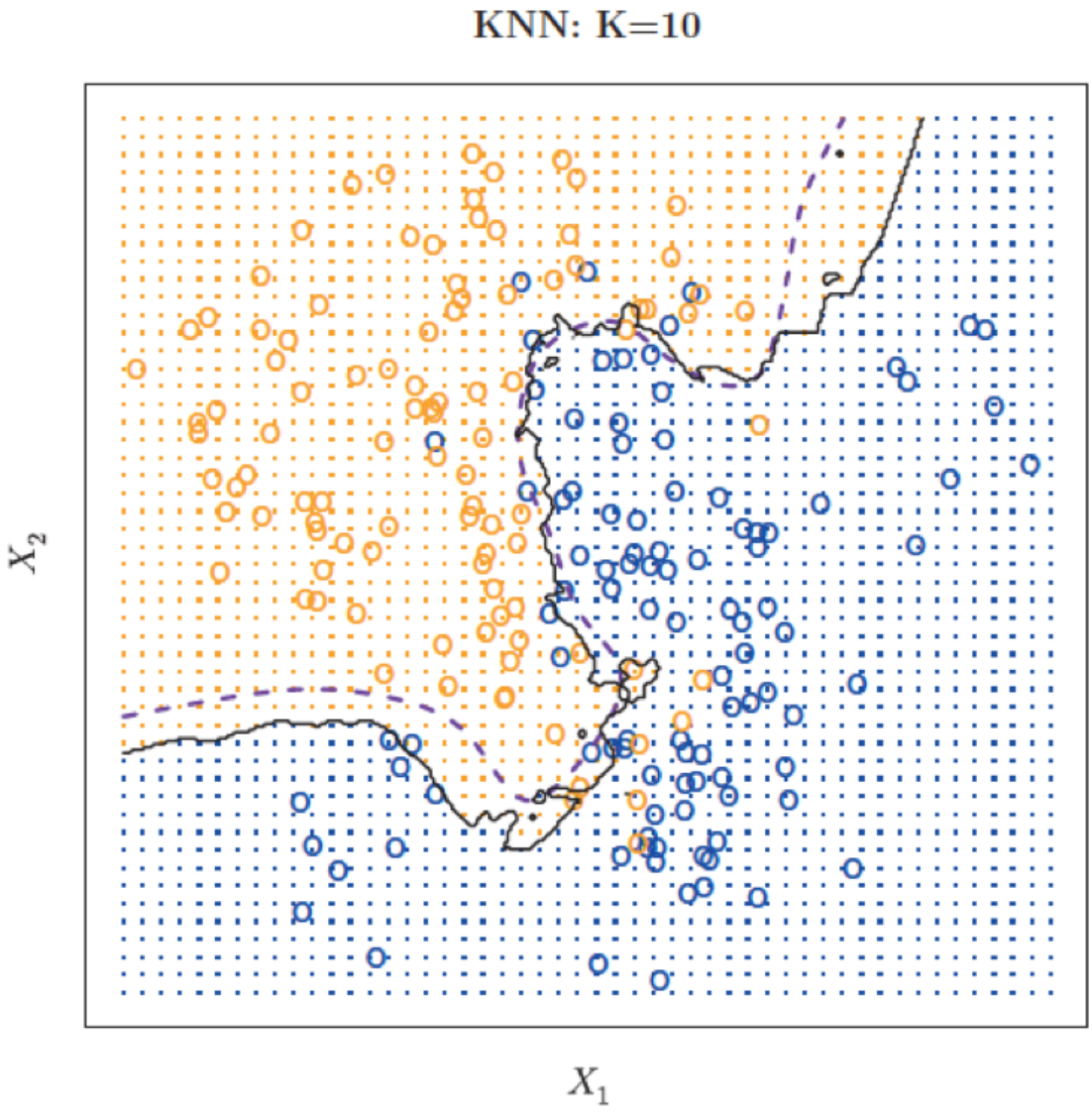

FIGURE 4. Training dataset consisting of 100 observations in each of two groups, indicated in blue and in orange. Image Courtesy of [10]. The purple dashed line and the black line, respectively, represents he Bayes decision boundary and the KNN decision boundary using K = 10.

The choice of K has a crucial impact on the result; the large is the K the lower is the variance but the higher is the bias. In this project, IBM SPSS Modeler chooses the value of k automatically and it was set as k=5.

2.3.2. Decision Trees

Decision Trees can be applied to both regression and classification problems. In this project, Decision Trees is another model which is used for the supervised classification problem. They have the advantage of being simple and therefore the interpretation of results is straightforward. On the other hand, they lack the predictions accuracy compared other methods used in this project, e.g. neural networks. This model consists of a series of splitting rules, starting at the top of the tree (see Figure 5).

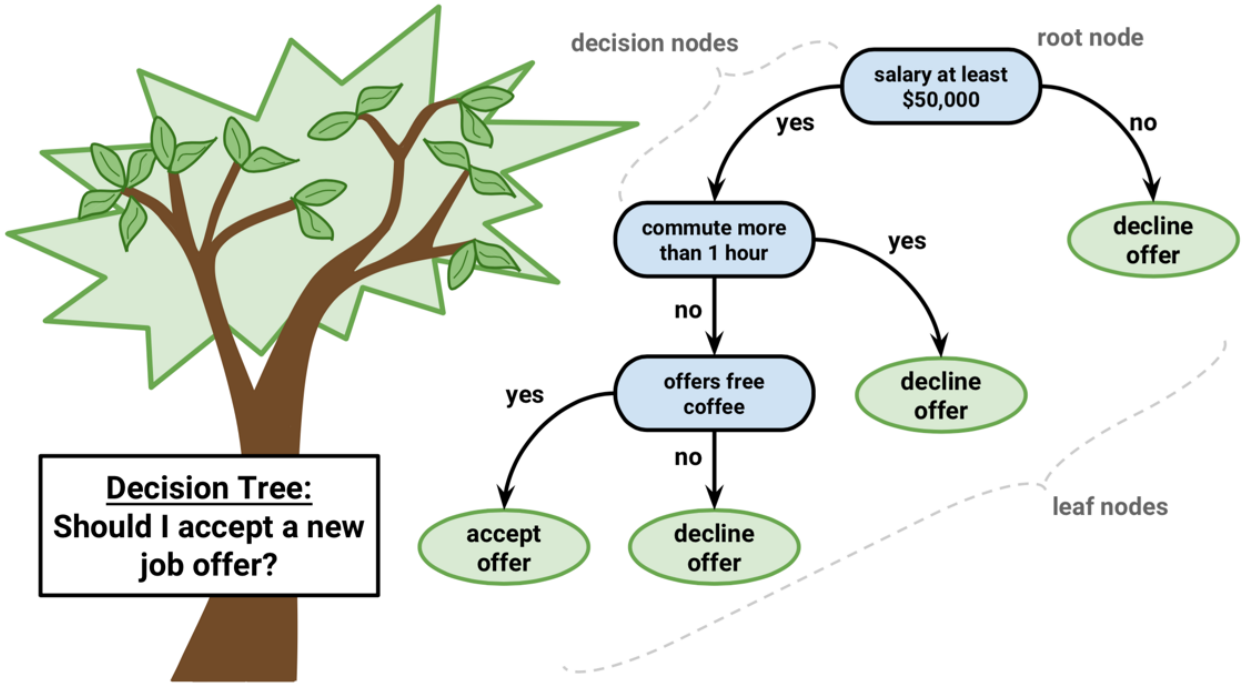

FIGURE 5. Image Courtesy of [18]. A decision tree structure for predicting whether a job offer should be accepted, based on the minimum salary, commute hours, and perks. At a given internal node, the label indicates the left/right-hand branch emanating from that split, and the right/left-hand branch corresponds to otherwise.

In practice, in Decision Trees method a top-down, greedy approach is favored, i.e. a recursive binary splitting is initiated with an initial region of all observation that is divided into two regions based on the best branching of that region instead of a global perspective. Based on such splitting of the variables at the nodes, various Decision Trees models are proposed. In summary, in developing Decision Trees model a recursive binary splitting grows based on optimization of, say Gini index as a measure of total variance, is used until a termination criteria, say number of observation in the region, is observed. Then a sequence of best subtrees is obtained using pruning techniques and K-Fold cross-validation.

In IBM SPSS, the Decision Trees add on can group individuals based on their characteristics, i.e. segmentation, it can identify the common features, i.e. profiling, and it can estimate a future event, i.e. prediction. Hence, in our project, Decision Trees can help to profile the customers who accept, in opposed to the ones who reject, the car insurance offer. While various number of algorithms for Decision Trees exist, in this project C5.0 method is used.

C5.0 is an improvement to C4.5 algorithm developed by Ross Quinlan [14], which itself is an extension to ID3 algorithm by the same author [15], and is based on entropy minimization of the attribute. C4.5 has gained lots of popularity, e.g. Witten et al. describe it as "a landmark decision tree program that is probably the machine learning workhorse most widely used in practice to date" [16]. C5.0 has even faster speed and better usage of memory and support of boosting as mentioned in [17]. Hence, in this project C5.0 model is used for prediction.

2.3.3. Neural Networks

Neural network is a data mining method that is based on the human brain to process information. This model is not an exact copy of how the brain functions but is biologically inspired by it. The main reasons why neural networks works well are because of its ability to study and learn the data, the nonparametric nature of the data (not rigid assumptions) and also their ability to generalize. Neural networks is often used for forecasting and business classification applications. In the paper *Effective Data Mining Using Neural Networks,* by Lu et al. [19] , a drawback with neural networks is " A neural network is usually a layered graph with

the output of one node feeding into one or many other nodes in the next layer. The classification process is buried in both the structure of the graph and the weights assigned to the links between the nodes. Articulating the classification rules becomes a difficult problem" [19]. One of the drawbacks of using neural networks for data mining is that it can takes longer execution time compare to other models. Thus, in order to efficiently use, a powerful computer is recommended. In this paper, the neural networks is expected to study the relations between the variables that contribute to the car insurance datasets and produce an output. In this case, the output is a single decision on whether the customer bought car insurance or not. During the process, neural networks will study hidden layers between the variables.

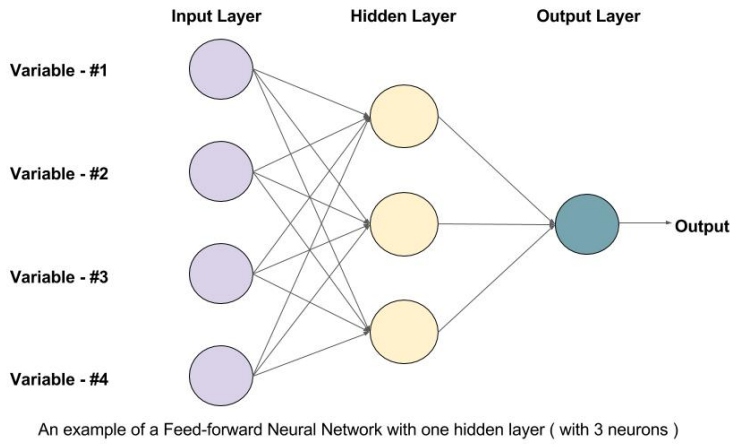

FIGURE 6. An example of feed-forward Neural network with one hidden layer

2.3.4. Support Vector Machines

Due to its superior predictive capability and theoretical foundation, support vector machines (SVM) are one of the most popular machine learning techniques being used. Support vector machines produce an input-output vectors that can be both classification function and also regression function. Due to their strong mathematical foundation, support vector machines are demonstrated in various real-world prediction problems, such as bioinformatics, medical diagnostics, demand forecasting and text mining. The justification for using support vector machines for this case are that it has simple geometric interpretation of the data and it gives sparse solution. Furthermore, in order to prevent an overfitting problem, support vector machines

can be count to solve this issue. However, support vector machines also has its drawbacks. For example, a limitation that SVM possesses is speed and size. This limitation applies for both the training and testing processes. Therefore, it requires time-demanding calculations and also extensive memory requirements.

3. Results and discussion

The next step for this project is evaluating performance of each data mining model. This is step 5 of CRISP-DM methodology described in section 2, i.e. evaluation. Here we focus on evaluating the results, reviewing the process and determining the next steps. Hence, the pros and cons of each model is discussed.

Table 2 provides a summary of confusion matrix, accuracy and sensitivity of all the four models based on IBM SPSS Modeler output. The confusion matrix results are the left two columns of each figure (ignoring the field No on the left most column); each row corresponds to a fold. As a remind a 5-Fold Cross Validation was used in this study. Confusion matrix allows visualization of the performance of each algorithm. The structure of confusion matrix is as follows: the upper left is true positive, the lower right is true negative, the upper right is false positive and the lower left is the false negative results for the classification problem in hand

KNN models performs the worst among the four models. It has 66.93% Accuracy, which is not desired in this classification model. Furthermore, the sensitivity 68.51% and specificity 36.68% values are not acceptable for this model as well. KNN performs a bad job in predicting false negative which has high impact on business performance. For instance, for 5th fold there are only 155 true negatives while 174 false negatives are observed. Therefore, the performance of KNN model for insurance call is not acceptable and thereby not recommended. This may be related to the high dimensionality of the problem.

Support Vector Machine (SVM) also performs poorly and only slightly better compared to KNN in terms of accuracy, but still is inferior to Decision Trees and Neural Network methods. For SVM the accuracy is 68.30%, which is almost identical to that of KNN, the sensitivity is 71.72% which is slightly better than that of KNN. However, it performs unacceptable results for specificity with 31.37%. For a similar comparison, for 5th fold there are 153 false negatives while

176 true negatives exist. Hence, just like KNN, SVM would not be a good choice for the application at hand.

Table 2. Data Mining Results: Confusion matrix, accuracy, sensitive, specificity and precision are given for each fold. The last row indicates the average values.

KNN

Fold No	Confusion Matrix		Accuracy	Sensitivity	Specicifity	Precision
1	990	228				
	464	389	0.6658619	0.68088033	0.36952998	0.81280788
2	979	232				
	401	391	0.68397404	0.70942029	0.37239165	0.80842279
3	1,000	215				
	459	377	0.67137981	0.68540096	0.36317568	0.82304527
4	354	80				
	172	149	0.66622517	0.6730038	0.34934498	0.8156682
5	365	95				
	174	155	0.6590621	0.67717996	0.38	0.79347826
Mean			0.669300605	0.685177069	0.366888458	0.68063522

SVM

Confusion Matrix		Accuracy	Sensitivity	Specifiifty	Precision
392	66				
158	161	0.7078507079	0.7127272727	0.2907488987	0.8558951965
379	75				
140	163	0.6856010568	0.7302504817	0.3151260504	0.8348017621
393	90				
143	176	0.6683291771	0.7332089552	0.3383458647	0.8136645963
353	81				
155	166	0.6728476821	0.6948818898	0.3279352227	0.8133640553
384	76				
153	176	0.680608365	0.7150837989	0.3015873016	0.8347826087
Mean		0.6830473978	0.7172304797	0.3147486676	0.8305016438

C5-Decision tree

Fold No	Confusion Matrix		Accuracy	Sensitivity	Specicifity	Precision
	398	60				
1	148	171	0.732303732	0.728937729	0.74025974	0.86899563
	391	63				
2	128	175	0.747688243	0.753371869	0.73529412	0.86123348
	421	62				
3	132	187	0.758104738	0.761301989	0.75100402	0.87163561
	374	60				
4	142	179	0.732450331	0.724806202	0.74895397	0.86175115
	399	61				
5	153	176	0.728770596	0.722826087	0.74261603	0.8673913
Mean			0.7399	0.7382	0.7436	0.8662

Neural Network

Confusion Matrix		Accuracy	Sensitivity	Specicifity	Precision
395	63				
95	224	0.7966537967	0.806122449	0.7804878049	0.8624454148
391	63				
88	215	0.8005284016	0.8162839248	0.773381295	0.8612334802
399	84				
98	221	0.7730673317	0.8028169014	0.7245901639	0.8260869565
376	58				
94	227	0.7986754967	0.8	0.7964912281	0.866359447
393	67				
94	235	0.7959442332	0.8069815195	0.7781456954	0.8543478261
Mean		0.793	0.806	0.771	0.854

The C5-Decision Tree shows 73.99% Accuracy. The Sensitivity level which is True positive is 73.82% and Specificity for this model which is true negative is 74.36%. This shows such method is acceptable for car insurance success calling application, with a room to improvement. As remark, it should be highlighted in the competition for which this dataset was used C5 algorithm was the winner. While, we also confirm the successful performance of such Data Mining method, we observe Neural Network outperforms C5 as explained below.

The last and the best model is Neural Network, which performs best among all the four data-mining models. It obtains 79.3% accuracy, which is the highest in terms of prediction. It also provides good results for sensitivity and specificity with 80.6% and 77.1%, respectively.

Thus, the best model that we choose is Neural network. To summarize, Table 3 below is given for all the methods used in this study.

Table 3. Data Mining Results Summary

Model	Accuracy	Sensitivity	Specificity	Precision
C5-Decision Tree	73.99%	73.82%	74.36%	86.62%
KNN	66.93%	68.51%	36.68%	68.06%
SVM	68.30%	71.72%	31.47%	83.05%
Neural Network	79.3%	80.6%	77.1%	85.4%

3.1. Information Fusion-based Sensitivity Analysis (IFBSA)

Information fusion is method which mixes information together and results in new evidence. This method is used to realize which variable plays an important role in making the prediction models. For this purpose, accuracy of all models should be standardized first and then the variable importance is calculated based on the standardized accuracy. For this project, IFBSA is performed and following results are obtained.

Table 4. Information Fusion-based Sensitivity Analysis

variables	importance for Decision Tree	importance for SVM	importance for Neural Network	IFBSA
Default	0.0243	0.034	0.004	0.019884707
DaysPassed	0.0267	0.032	0.054	0.038478533
CarLoan	0.0272	0.044	0.04	0.037280228
Age	0.0548	0.038	0.006	0.03141284
NoOfContacts	0.0746	0.032	0.116	0.076339755

LastContactDay	0.093	0.034	0.016	0.045753215
Communication	0.1078	0.112	0.142	0.121778804
HHInsurance	0.1111	0.156	0.132	0.133120916
PrevAttempts	0.1548	0.038	0.08	0.089963742
LastContactMonth	0.2668	0.252	0.164	0.224089213
Marital	0.014725	0.038	0.034	0.029260172
Education	0.014725	0.072	0.03	0.038606727
Job	0.014725	0.088	0.074	0.05996492
Balance	0.014725	0.03	0.108	0.054066228

Table 4 results indicate that "LastContactMonth" variable is the most important predictor in making prediction models. Figure 7 reaffirms such finding, when the bar chart is included for Information Fusion-based Sensitivity Analysis.

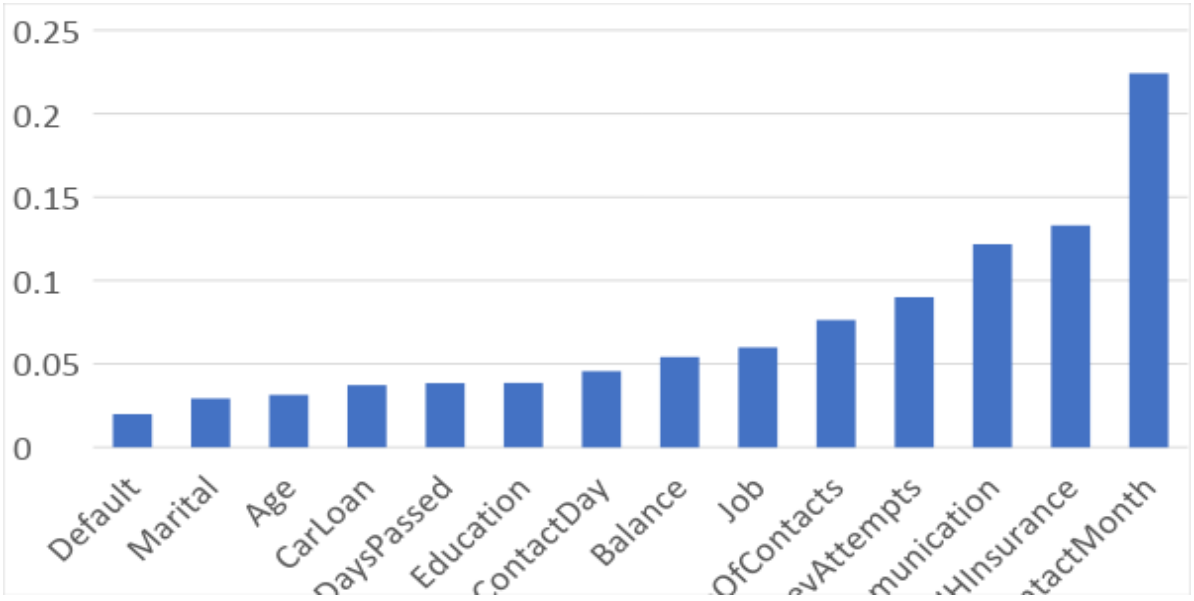

FIGURE 7. IFBSA bar charts

4. Managerial implications

There are several potential managerial implications from the result of this study. Managers of car insurance companies should be able to accurately predict of their insurance sale through cold calls. By using various variables that affect whether a cold call was successful, this study is able to determine and identify which variables are more important than others. The importance of this study is that it can greatly narrow a manager's focus when it comes to allocating resources. Instead of having to spend time and money on different factors that could or could not impact the sale of car insurance, managers can instead focus solely on few important factors instead. For example, based on the information fusion-based analysis result, it is evident that the top 4 factors that influence the outcome of cold call are previous attempts, communication, household insured or not, and last contact month. In a logical and business, such results are meaningful since the amount of attempts made previously can impact whether an individual will purchase the current car insurance policy or not. By persisting with a customer, he or she will be more encouraged to make the purchase due to the familiarity and abundance of information provided. Furthermore, the form of communication is also an important factor. An individual will be more accepting of cold calls depending on what type of communication devices they have been contacted from. A mobile phone can suggest that the individual was not contacted at a time period where they are open to hear about new information. Last but not least, the last time the individual was contacted regarding the car insurance has the highest influence on the result. This could be because the longer since a customer was contacted, the more likely they would have forgotten about the information.

Managers should utilize this information and capitalize on it. By investing in the right factors, their companies can benefit of gaining new customer quickly and effectively. Furthermore, it is recommended that managers continue to gather new data cases and add in more variables to foster the classification models.

5. Future Studies and Concluding Remarks

In order to produce an even more concrete and conclusive predictive models, there are a few changes and additions that can be made. One addition that can be added to the studies is the cost of insurance. By combining the original data with new datasets that contain the cost of each policies presented and sold to customers, it can gives a clearer understanding of customers behaviors and decision making process. It is evident that in the consumer market, cost has a great influence on customer's behaviors. Thus, car insurance companies can store data regarding this factor, it will gives future studies an opportunity to produce an even more accurate model to predict whether or not customers will buy car insurance or not.

Another addition that can be made to the study is the name of the policies. It has been extensively studied and implemented that name of certain products can produce a positive or negative influence on customer's purchases. Thus, if a company were able to provide data sets regarding the sales of insurance policies including the name of each policies, it would show whether the name can have a great impact on the sale or not.

Further addition could be information regarding the client demographic. If the cold calls were only made to potential customers that are located in a low income areas or areas that have low new vehicles purchases, then it can have a significant impact on the amount of insurance purchases. However, if the cold calls were made to customers in areas with high amount of new vehicle purchases, it is more inclined to receive a higher receptions from people due to higher demands and needs in the area.

As for drawbacks, there were a few methods and processes made during this study that could have been avoided. For example, originally, this study were conducted without removing outcome from one of the input variables. After careful analysis, it was found that since outcome had such a large impact on the result of the result, it would be a better model to not include outcome as one of the input variables.

To conduct this study, many processes were involved. It started by selecting a data set that is suitable with enough meaningful variables and cases. Furthermore, a long process of data preprocessing was needed in order to prepare the data to be usable. And by using four different modes KNN, Neural Networks, Support vector machines and decision tree, a conclusive model was produced to accurately predict which variables highly influence the result of insurance sales.

Predictive modeling is a powerful and important tools for car insurance companies and also many other businesses in a variety of industries. It can be used for many purposes, such as cold call variables predictor, risk assessment, policies assessment etc. There are still a lot more to complete before it could be fully comprehended and coherent. This study is only scratching the surfaces of what could be done with this one particular data set. The potentials is enormous.

References

[1] Alshamsi, Asma S. "Predicting car insurance policies using random forest." Innovations in Information Technology (INNOVATIONS), 2014 10th International Conference on. IEEE, 2014.

[2] Kašćelan, Vladimir, Ljiljana Kašćelan, and Milijana Novović Burić. "A nonparametric data mining approach for risk prediction in car insurance: a case study from the Montenegrin market." Economic research-Ekonomska istraživanja 29.1 (2016): 545-558.

[3] D'Arcy, Stephen P. "Predictive modeling in automobile insurance: a preliminary analysis." World risk and insurance economics congress. 2005.

[4] Thakur, S. S., and J. K. Sing. "Mining Customer's Data for Vehicle Insurance Prediction System using Decision Tree Classifier." International Journal on Recent Trends in Engineering & Technology 9.1 (2013): 121.

[5] https://www.kaggle.com/kondla/carinsurance#carInsurance_train.csv

[6] https://ieeexplore-ieee-org.umasslowell.idm.oclc.org/document/4631695

[7] https://ieeexplore-ieee-org.umasslowell.idm.oclc.org/document/4631695

[8] https://libguides.library.kent.edu/SPSS/PearsonCorr

[9] http://users.stat.umn.edu/~helwig/notes/datamat-Notes.pdf

[10] James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer, 2013.

[11] Wolpert, D.H., Macready, W.G. (1997), "No Free Lunch Theorems for Optimization", IEEE Transactions on Evolutionary Computation 1, 67.

[12] Devroye, L.; Gyorfi, L. & Lugosi, G. (1996). A probabilistic theory of pattern recognition. Springer. ISBN 0-3879-4618-7.

[13] http://cseweb.ucsd.edu/~elkan/151/nearestn.pdf

[14] Quinlan, R. C. "4.5: Programs for machine learning morgan kaufmann publishers inc." San Francisco, USA (1993).
[15] Quinlan, J. R. 1986. Induction of Decision Trees. Mach. Learn. 1, 1 (Mar. 1986),
[16] Ian H. Witten; Eibe Frank; Mark A. Hall (2011). "Data Mining: Practical machine learning tools and techniques, 3rd Edition". Morgan Kaufmann, San Francisco. p. 191.
[17] M. Kuhn and K. Johnson, Applied Predictive Modeling, Springer 2013
[18] https://hub.packtpub.com/divide-and-conquer-classification-using-decision-trees-and-rules/
[19] Lu, Hongjun, Rudy Setiono, and Huan Liu. "Effective data mining using neural networks." IEEE transactions on knowledge and data engineering 8.6 (1996): 957-961.