

# Application of Artificial Neural Network and Information Gain in Building Case-based Reasoning for Telemarketing Prediction

S.M.F.D Syed Mustapha<sup>1</sup>, Abdulmajeed Alsufyani<sup>2</sup>

Computer Science Department, College of Computer and Information Technology  
Taif University, Taif, Saudi Arabia

**Abstract**—Traditionally, case-based reasoning (CBR) has been used as advanced technique for representing expert knowledge and reasoning. However, for stochastic business data such as customers' behavior and users' preferences, the knowledge cannot be extracted directly from data to build the cases in reasoning in making prediction. Artificial Neural Network that is known to be able to build model for predicting unprecedented business data is used together with Shannon Entropy and Information Gain (IG) to identify the key features. 8 attributes have been identified as key features from the 17 attributes which are based on the telemarketing data. These attributes are used to select the key features in building CBR. The weightage for the key features in the cases is obtained from the IG values. The mechanism of creating the cases based on the input from the ANN is discussed and the integration process between ANN and CBR is given. The process of integrating the ANN and CBR shows that both techniques complement each other in building a model in predicting a customer who would subscribe one of the promoted new banking services called "term deposit".

**Keywords**—Artificial neural network; prediction model; telemarketing; Shannon Entropy; feature selection; case-based reasoning

## I. INTRODUCTION

Generally, marketing is an essential tool to increase sales and telemarketing is one of the marketing mechanisms since the advent of internet technology. There are two types of telemarketing, namely, the inbound and outbound [1]. The former describes the initiation made by the customer while the latter is managed by a group of telemarketing team that is designated to perform the systematic task in soliciting potential customers. In order to have an effective telemarketing communication, some background data are required to understand the potential of the customer in subscribing the proposed product prior to the teleconversation. The background data are related to the customer such as the age, profession, financial standing, prior marketing engagement with customers and even academic qualifications of the customers. These attributes are ought to be the ones that are significant to contribute to the final decision made by the customer whether to subscribe or otherwise on the proposed product.

In order to have a better prediction of the telemarketing situation, the attributes have to be selected based on its

relationship to the predictors (in this case the decision to be made by the customer on the subscription of the product). Each attribute will be measured based on the features of the attributes using Shannon Entropy and Information Gain (IG). The attributes are ranked based on IG values prior to feeding to the ANN for predicting the customer's response. We work on 45,212 data where about 10% is kept for testing and the remaining for training the neural network. The accuracy is reported based on the tested data against the actual result.

Our work is motivated by the past work in using machine learning techniques for making prediction mainly in sales, marketing and consumer behavior. According to Dilek et al. [2], marketing is an expensive operating cost for an organization that making a close to accurate forecasting on the potential market segment will be useful to reduce the cost of the production and materials, directly and indirectly. The ability to classify the consumer behavior using machine learning techniques such as regression, non-linear principal component analysis and classification was demonstrated by Richard et al. [3]. Modeling customer satisfaction has also resorted to ANN as its ability to determine the non-linear relationship between the consumer satisfaction and the factors that influence its behavior [4]. ANN is also used for volatile relationship like stock market on daily basis based on the index stock market indicator [5]. In market related application, ANN is able to predict housing prices by comparing ANN and hedonic economic model [6]. Similar work is done by Waheeb and Ghazali in predicting chaotic time series using high order neural network [15]. ANN was also proven to be able in predicting technical skills of potential soccer player [16].

Case-based reasoning (CBR) was introduced as a rapid approach in building expert's knowledge which is acquired from the past solvable cases [17]. Each case is presented with features and key features which are identified as discriminative factors to select suitable cases. Integrating CBR-ANN as complementing approach of both techniques has been shown by several works recently. Platon et al. deployed two machine learning techniques, ANN-CBR and a feature selection technique called PCA (Principal Component Analysis) to predict the power consumption on hourly basis [12]. The data that have readings with high variability is significant to be detected and to be used for prediction, it is captured using PCA. Electric consumption readings are random in nature, hence ANN would be the suitable method to model its behavior. Another proof of using a combined ANN-CBR is on

the design of green building based on the past successful building design. ANN is used to model the non-linear relationship of the key features in the cases and to predict the suitable case for a give case query [13]. Another suitable application of CBR-ANN integration is appraising the pricing for the domain name where the charges are based on arbitrary attributes such as length of domain name, words component, number of clicks, number of searches, etc. [14]. Biswas et al. took the advantage of using ANN to determine the feature weightage for the CBR cases. The ANN tree is pruned by taking into consideration of the four aspects in determining the feature weightages – sensitivity, activity, saliency and relevant [18].

Feature selection is one of the essential components prior to identifying key attributes that could enhance the accuracy in the prediction or classification using any machine learning techniques. Feature selection has been discussed and applied elsewhere in the literature. For example, feature selection using entropy has been successful for extracting the salient power consumption readings [7]. Another work reported on the usefulness of feature selection was on prediction of risks on hepatitis disease where feature selection had proven to be significant in enhancing the performance of the learning algorithm [8].

Our work in relation to the related work is on the application of feature selection on the telemarketing data using Shannon Entropy [9] and utilizing ANN and CBR for making prediction on the potential subscriber for the term deposit. In the following section, we describe the details in our research methodology.

This paper is organized in the following manner, Section 1, Introduction as discussed in this section, followed by Section 2, Materials and Methods, and finally Results and Discussion in the Section 3.

## II. MATERIAL AND METHODS

In this project, neural network takes the input values within the range of 0 and 1 and the data was transformed accordingly as described in subsection Data Preparation. In the subsequent subsection, the application of Shannon Entropy and Information Gain on the 16 attributes (as the 17th attribute is the decision) is illustrated. The theoretical ANN model and CBR concept are expounded and demonstrated in terms of how these two techniques can be integrated.

### A. Data Preparation

Data is obtained from UCI Machine Learning Website (<https://archive.ics.uci.edu/ml/datasets/Bank+Marketing>) which was made available since 14th Feb 2014. The data was first tested by Moro [10,11]. There were 17 attributes about a customer which are age, job type, marital status, education level, credit in default, housing loan, loan, personal loan, contact device, last contacted month and week as well as duration of the conversation by telemarketers, number of times the customer being contacted, the number of days lapse after the last contact, number of times the customer being contacted prior to the recent campaign, outcome of the previous campaign and finally the decision on whether the customer subscribes the term deposit product or not. The data provided

was in original format that it has to be converted to numerical scale between 0 – 1. For a qualitative categorical data (such as qualification – high school, tertiary, uneducated, etc.), the numerical is assigned in an even interval. For a binary category such as “Yes” or “No”, each is assigned at an extreme value, 0.9 and 0.1 respectively. For the attributes with numerical values, they are normalized and discretized between 0.1 and 0.9. This is important for data such as salaries or total saving amount which comes in infinite variations. The final outlook of the entire data is numerical values between 0 and 1. The decision whether to subscribe to the term deposit or not is valued as 0.9 and 0.1, respectively. There are 56,000 customer data in which 46,000 is separated as training data and the remaining is for testing.

### B. Shannon Entropy and Information Gain the Shannon

Entropy of  $Q$  is Given as

$$Q(A) = -\sum_{i=1}^n p_i(a) \log_2 p_i(a) \quad (1)$$

where  $p_i$  is the probability of the features within a given attribute; and  $n$  is the number of features of the attribute in relation to the decision type “Yes” or “No”, hence  $n = 2$ . For the data being used in this project, the number of features varies for each attribute and the number of unique features are determined automatically. For example, the attribute *Occupation* has  $M$  features ( $m_i \in M$  where  $i = 3$ , for example,  $m_1 =$  B-blue-collar,  $m_2 =$  W-white-collar and  $m_3 =$  U-unemployed). Hence,  $Q(A_m)$  is calculated as

$$\sum_{i=1}^2 -p_i \log_2 p_i = \left(-\frac{C(Y)}{C(Y+N)}\right) \log_2 \left(\frac{C(Y)}{C(Y+N)}\right) + \left(-\frac{C(N)}{C(Y+N)}\right) \log_2 \left(\frac{C(N)}{C(Y+N)}\right) \quad (2)$$

where  $C$  is the cardinality and  $m$  is one of the feature in  $M$ . When value  $Q$  is close or equal to 1, the feature is impure and hence is significant to the attribute. For each  $m_i \in M$ ,  $Q(A_{m_i})$  is calculated and the sum is  $Q(A)$ .

Each attribute (total of 16 attributes) is measured using IG (Information Gain) to determine discriminative value in determining the final decision. IG for an attribute is calculated as the followings:

$$Q(TA) = \left(\sum_{i=1}^{C(M)} \left(-\frac{C(Y)_i}{C(Y+N)_i}\right) \log_2 \left(\frac{C(Y)_i}{C(Y+N)_i}\right) + \left(-\frac{C(N)_i}{C(Y+N)_i}\right) \log_2 \left(\frac{C(N)_i}{C(Y+N)_i}\right) - Q(A)\right) \quad (3)$$

The entropy value for an attribute is calculated based on the total number of “Yes” and “No” for all features under the attribute. Hence, IG is  $Q(TA) - Q(A)$ .

### C. Artificial Neural Network

We set our ANN with one hidden layer and sigmoid model as the transfer function. The output layer is set to be 0.9 to represent “Yes” or 0.1 as “No”. Sigmoid model is shown as follows:

$$y = \left( \frac{1}{1+e^{-x}} \right) \quad (4)$$

Every attribute that has been selected based on the Information Gain value will be coded as the input node. The number of hidden layer nodes is equal to the number of attributes. The entire network is depicted in Fig. 1. The initial values between 0 and 1 are fed into each input node which represents the attributes. The weights for the hidden layer are randomly computed as initial value. These values are propagated forward using sigmoid function. The final value is summed by totaling the values from all nodes from the hidden layer. The differences with the targeted value and total sum are computed in order to perform adjustment to the weightage. This is performed in many cycles until the sum error between the target and the computed weightage has reached to some threshold value. In our work, the stopping criteria is set to be error,  $\epsilon < 0.001$  or 1000 cycles.

#### D. Case-Based Reasoning (CBR)

Traditional CBR has four fundamental phases which are performed on each case such as retrieval, reuse, revise and retain [18]. In retrieval phase, cases are retrieved based on the similarities to the problems being matched. The matching is performed by measuring the similarities of the key features between the cases and the posed problem. Hence, one or more cases could be retrieved, and this depends on the threshold being set as the minimum value for a case to be retrieved. Reuse phase is when the past cases have similar problems to the posed problem, and hence the recommended solution from the past cases can be recommended to be reused. However, not all cases have good matching that the recommended solution could be reused as it is. In this regard, some adaptations on the recommended solutions need to be performed based on the discrepancies on the key features and this led to revision. Revise phase involves the process of adjusting the recommended solution in the partially matched cases. The new adapted solution is done based on few strategies such as reference to ontology, heuristics rules or semantic database. Beside adaptation, cases which are not useable or obsolete, the recommended solution has to be changed or the entire cases have to be discarded as the problems are no longer relevant to the current context. Retain phase is performed on the cases where adjustment on the solution is made and the cases can be treated as new cases being created and retained in the case library.

#### E. Integrating ANN and CBR

The purpose of integrating ANN and CBR is to take the advantage of both machine learning techniques. Traditionally, during the creation of CBR case, the key features are determined by the expert who will also advise on the weightage based on the importance of each key feature in order to discriminate effectively the cases. However, in many domains where the data are unprecedented, the expert could not give his/her intuitive idea. Information gain is deployed as the feature selection method to determine the key attributes. Information Gain (IG) is measured based on the ratio of an attribute's entropy value against the entire entropy. The attributes that is considered are those where the data could be

categorized with maximum of four categories. Hence, attribute such as National Security Number, Phone number are not categorizable as each data is unique. IG generates the values that the selection of the suitable attributes is determined manually. The quality of attributes selection is determined by applying in ANN model. The performance of the ANN will determine the right selection of the attributes which are significant to be the considered as key features in case creation. Fig. 2 shows the entire process of determining key features for the cases using feature selection and ANN for CBR.

Business data may have physical meaning but the contribution to distinguish in decision making is not known. The attributes can only be signified by evaluating its contribution based on comparison of the information gain against all attributes. The result of the feature selection is discussed in Section III. The set of data with the selected attributes are performed on ANN to build the prediction model. A performance criterion of the ANN model is set (discussed in Section III) and if the results are unsatisfactory, the attributes will be reduced further. The adjustment of the feature selection is done iteratively until ANN model is able to meet the minimum required performance. Each feature selection has its IG value which is used to determine the weight for the key features in building the case.

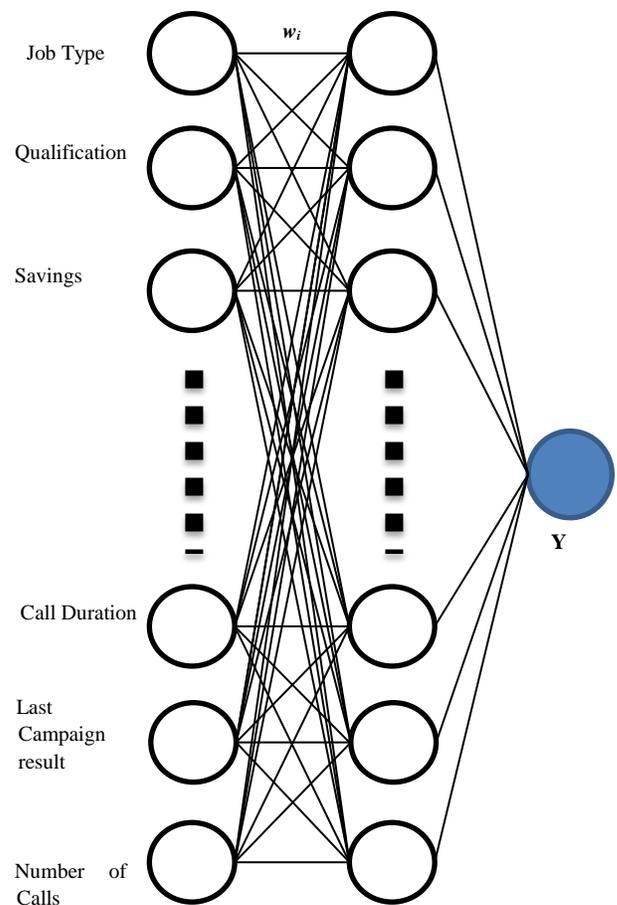


Fig. 1. Single Hidden Layer ANN.

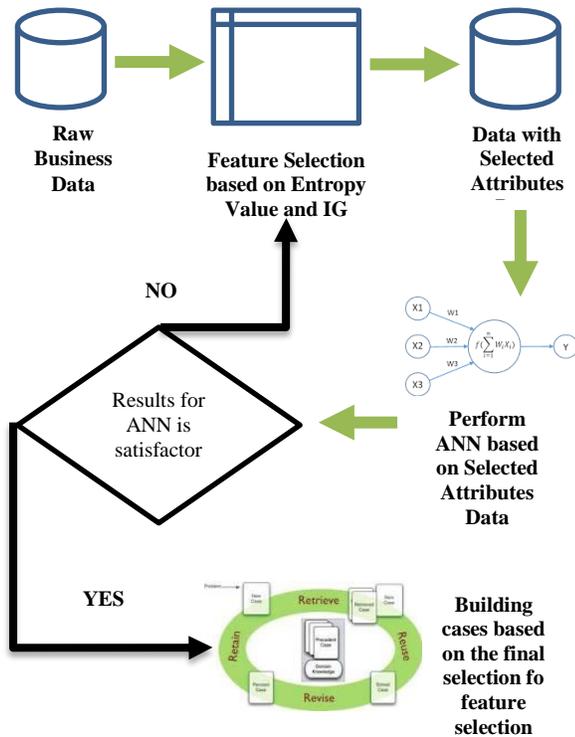


Fig. 2. Process of ANN-CBR Integration based on Feature Selection.

#### F. Building CBR

Building case for CBR has the advantage that each case can be built as independent case solving unit. Hence, each data can be treated as a case. This is called a linear case library which may not be efficient for a fast case retrieval. The cases have to be grouped by using clustering method. Since the data sets are labelled, the cases are segregated by the two main decision making, “Yes” or “No” to subscribe the term deposit. Heuristics algorithm using k-means can be applied to cluster the data sets for both sides (“Yes” and “No”) to few more sub-clusters. Each sub-cluster is created to allow the grouping of different class of cases based on the similarity values. The procedure in generating sub-cluster is given below:

Procedure to generate sub-cluster:

- 1) Given set of data,  $N = \{n_1, n_2, \dots, n_k\}$
- 2) Choose a random centroid point,  $c$  where  $c \in N$
- 3) Let  $c_1$  and  $c_2$  are two sub-clusters
- 4) Determine  $d(c_1, c_2)$
- 5) Repeat
  - a) for each data  $n_i$ , determine  $d_{i,1} = (c_1, n_i)$  and  $d_{i,2} = (c_2, n_i)$
  - b) assign cluster to  $n_i$  based on the  $d_{i,1}$  value
  - c) reassign the new points for  $c_1$  or  $c_2$  depending on which cluster gets a new member
  - d) execute step a with new points of  $c_1$  or  $c_2$

Until the centroid  $c_1$  or  $c_2$  converged to the same position.

The same procedure could be applied to generate smaller cluster within the two clusters and the number of clusters depending on the complexity of the attributes. Since, using the feature selection has reduced to 8 attributes, two clusters are sufficient. For each  $n$ , the value  $d$  is calculated using the following Euclidean distance:

$$d_{i,j} = (\sum_1^k \delta(n - c)^2)^{1/2} \quad (5)$$

where  $\delta$  indicates the attributes value for each key feature. In this project, all attributes are normalized in order to allow numerical computation.

CBR is known as lazy learning in machine learning that its knowledge that are stored in the cases can grow through a simple comparison with the existing knowledge and stored when the new case does not match with the existing ones. Fig. 3 shows the process managing the case query. A case query could potentially be close to the centroid or far from any centroid. In the Cluster Assignment, the Case Query will be evaluated to determine its closest centroid. If it is found (new cluster is not needed), further search on case library is performed to find similar or partial match with the existing case. New cluster is recommended if no existing centroid have a close match. A close match value is determined if the Case Query falls within the boundary of the maximum and minimum distance of the case member within the same centroid. If none exist, then it is recommended that the Case Query needs a new cluster. The new Case Query will become the first centroid point. Cluster and sub-cluster avoid linear searching which can be  $O(N)$  for a large case library while clustered case library will reduce to  $O(\log N)$  depending on the number of clusters. For this purpose, other types of clustering algorithms such as AHC (Analytical Hierarchical Clustering [19]) is possibly used.

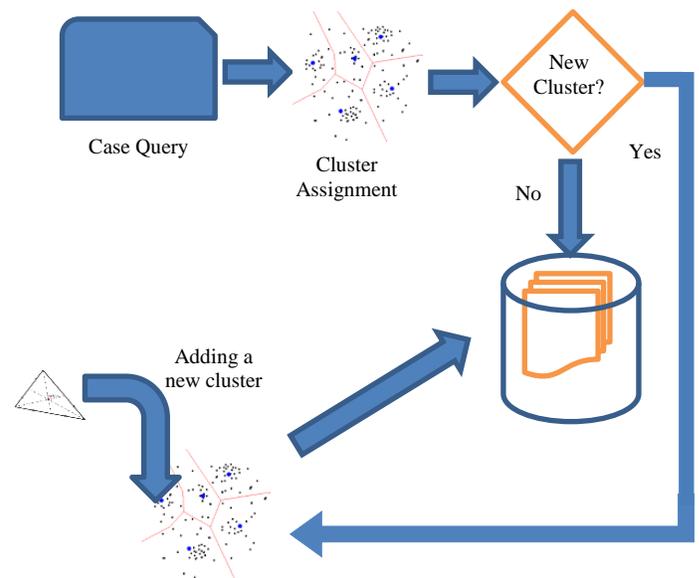


Fig. 3. Processing Case Query.

### III. RESULTS AND DISCUSSION

We represent the results in two manners, the outcome of the Shannon Entropy and Information Gain and the success rate of applying ANN in performing the prediction. Table I shows the physical meaning of the attributes as explained in the UCI Machine Learning website.

Some of the attributes use qualitative or descriptive values that the values have to be manually converted to some distinct numerical values to differentiate between the optional values. Table II shows the entropy and the information gain values for each attribute. Only those attributes with  $IG > 0.01$  will be considered as informative attributes (these attributes are bold).

The challenges that we are facing is that most of the attributes do not have high entropy value and they are not different from each other to make effective selection on the informative attributes. Hence, using information gain, we manage to reduce the attributes from 16 to 8.

The ANN program that we develop could take any number of attributes. For this experiment, we deploy only 8 attributes. There are 35,930 training data with the decision of “Yes” and 4760 data with “No”. The testing data for “Yes” and “No” are 3992 and 528 respectively. There are four experiments performed on the ANN, as shown in Table III.

TABLE I. ATTRIBUTE DESCRIPTIONS

Attribute No	Physical Meaning
Attribute 1	Age
Attribute 2	Employment Types (Student, Self-Employed, Retired, Management etc.)
Attribute 3	Marital Status
Attribute 4	Highest Education
Attribute 5	Credit in Default
Attribute 6	Balance in Saving
Attribute 7	Housing Loan
Attribute 8	Contact Mode (Mobile or Fixed Line)
Attribute 9	Last contact month
Attribute 10	Last contact day of the week
Attribute 11	Duration of the telemarketing conversation
Attribute 12	Number of times the customer being contacted
Attribute 13	Number of days lapse after the last telemarketing call
Attribute 14	Number of times being contact prior to the current campaign
Attribute 15	Number of days lapse since the last campaign
Attribute 16	Outcome of the previous marketing campaign
Attribute 17	Accept or Decline the term deposit

For each testing data, we apply the ANN model that has been built using the training data. Training data are separated based on the decision “Yes” or “No” and ANN is trained accordingly. The final weightage is computed to form the ANN model and the two sets of testing data (namely, with “Yes” and “No” decision) are applied for both ANN model with two different set of decision type. The targeted values of the decision for “Yes” and “No” are 0.9 and 0.1, respectively. Fig. 2 shows the sample of the results generated automatically in Excel using Visual Basic.

The “Computed Value” is the value generated from the ANN model. The “Difference” is the absolute value of the difference between the “Targeted Value” and the “Computed Value”. The “Difference” that is less than 0.001 and 0.01 are totaled on each of the testing data set on the two ANN model which is built on the two training data sets and the percentage shown in Table III represents the data with differences below 0.001 and 0.01. That means, if we set 0.01 as the threshold, we shall able to have at least 93% of testing data correctly predictable. If we set the threshold higher to 0.001 (smaller difference between the targeted and computed value), the percentages drop to 13% and 17% for the “Yes” and “No” cases. In the case of testing data that is executed on the ANN model which has opposite decision value, none of the testing data showed shows small differences, hence, it can be concluded that none of the testing data will be wrongly predicted. The testing data is not included as part of the data for training. Hence, the training data is considered as new.

TABLE II. ENTROPY AND IG VALUES

Attributes	Entropy values	Information Gain (IG)
<b>Attribute 1</b>	<b>0.50725</b>	<b>0.013375</b>
<b>Attribute 2</b>	<b>0.50870</b>	<b>0.011922</b>
Attribute 3	0.51783	0.002791
Attribute 4	0.51688	0.003748
Attribute 5	0.52020	0.000424
Attribute 6	0.51715	0.003475
<b>Attribute 7</b>	<b>0.50670</b>	<b>0.013927</b>
Attribute 8	0.51683	0.003794
<b>Attribute 9</b>	<b>0.50097</b>	<b>0.019659</b>
Attribute 10	0.52006	0.000563
Attribute 11	0.51406	0.006563
<b>Attribute 12</b>	<b>0.45886</b>	<b>0.061762</b>
Attribute 13	0.51743	0.003193
<b>Attribute 14</b>	<b>0.50334</b>	<b>0.017283</b>
<b>Attribute 15</b>	<b>0.50334</b>	<b>0.017283</b>
<b>Attribute 16</b>	<b>0.04241</b>	<b>0.478219</b>

TABLE III. RESULTS FOR THE ANN

Experimental Set Up		<0.001	<0.01
Training Data with "Yes"	Testing Data with "Yes"	13%	98%
Training Data with "Yes"	Testing Data with "No"	0%	0%
Training Data with "No"	Testing Data with "Yes"	0%	0%
Training Data with "No"	Testing Data with "No"	17%	93%

H	I	J	K
Attribute	Targeted value	Computed Value	Difference
16	0.2	0.9019777	0.0019777
	0.6	0.9056676	0.0056676
	0.4	0.9045856	0.0045856
	0.2	0.8956658	0.0043342
	0.6	0.9037097	0.0037097
	0.4	0.9040949	0.0040949
	0.2	0.8991377	0.0008623
	0.2	0.9011563	0.0011563
	0.6	0.9058895	0.0058895
	0.2	0.8967431	0.0032569
	0.6	0.9060655	0.0060655
	0.6	0.9050942	0.0050942

Fig. 4. Results generated using Excel Format.

#### IV. CONCLUSIONS

Business data like telemarketing is always large and the values of the attributes are not discernible as whether they should be included or excluded in the analysis. One may have attributes that may be seemed to be essential in terms of the functional aspects, but they do not carry any meaning in the data analysis. For example, the duration of the call in persuading the customer is an essential attribute but it may be meaningless if all of the values under the same attributes are the same. Hence, using Shannon Entropy and Information Gain will be useful in selecting attributes that could enhance in identifying the distinctive data. ANN model is proven to be useful to be used to model stochastic data such as telemarketing data as it has no standard model that can be used to model for making prediction. However, ANN is always be considered as black box where the decision being made is not transparent to the decision makers. Translating the black box model into white box model in the form of business rules and case presentation offer few benefits. The cases can be maintained and through periodical review of its decision without the need to rerun the ANN. The solution in the cases can be adapted to new situations. Practitioners could learn from the cases to understand the pattern in consumer behavior and preferences. Our work here shows that such model can be built using ANN and the prediction is reliable for certain threshold value.

The present work has few potential works for improvements. Adaptation is the powerful aspect in CBR as it allows the sustainability of the case library to be able to handle unprecedented situation. At present, adaptation which is done through adjustment to the parameters do not guarantee that the case is able to adapt to a more challenging situation. Deploying solid knowledge representation technique such as ontology shall allow reasoning and adaptation to new solution. Another aspect of improvement is the clustering in which the number of clusters is pre-fixed considering the number of attributes is small. For a larger set of attributes, the number of clusters and sub-clusters can be automatically determined to avoid as the behavior of the clusters cannot be determined manually.

#### REFERENCES

- [1] Kotler, P. and Armstrong, G. (2004). Principles of Marketing, 10Ed. Pearson Education Inc. New Jersey
- [2] Dilek Penpece and Orhan Emre Elma, (2014). Predicting Sales Revenue by Using Artificial Neural Network in Grocery Retailing Industry: A Case Study in Turkey. International Journal of Trade, Economics and Finance, Vol 5, No 5. October 2014, pp 435 – 440.
- [3] Richard Briesch and Priyali Rajagopal, (2010). Neural Network Applications in Consumer Behavior. Volume 20, Issue 3, July 2010, pp 381-389.
- [4] Nooshin Zeinalizadeh, Amir Abbas Shojaie, Mohammad Shariatmadar, (2015). Modeling and analysis of bank customer satisfaction using neural networks approach. International Journal of Bank Marketing. Vol 33, Issue 6, pp 717 – 732
- [5] Qiu M, Song Y (2016) Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model, PLoS ONE 11(5): e0155133. doi:10.1371/journal.pone.0155133
- [6] Limsonbunchai V, Gan C and Lee M. (2004). House price prediction: hedonic price model vs. artificial neural network. American Journal of Applied Sciences,1(3), pp 193-201
- [7] Sergio J., Àngela N., Fransisco M., Narcís A. (2015) Hybrid methodologies for electricity load forecasting: Entropy-based feature selection with machine learning and soft computing techniques, Energy, Vol 86, 15 June 2015, pp 276 – 291.
- [8] Pinar Yildirim (2015). Filter Based Feature Selection Methods for Prediction of Risks in Hepatitis Disease. International Journal of Machine Learning and Computing, Vol. 5, No. 4, August 2015, pp 258 – 263.
- [9] Shannon, Claude E. (July–October 1948). "A Mathematical Theory of Communication". Bell System Technical Journal. 27 (3): 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
- [10] Moro, S., Cortez, P., and Rita, P. (2014). A Data-driven Approach to Predict the Success of Bank Telemarketing, Decision Support Systems, Elsevier, 62: 22 – 31, June 2014.
- [11] Moro, S. Laureano, R. and Cortez, P (2011). Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds), Proceedings of the European Simulation and Modelling Conference- ESIM'2011, pp. 117 – 121, Guimaraes, Portugal, October, 2011d. EUROSIS.
- [12] Platon, R, Dehkordi, V.R. and Martel, J. (2015). Hourly predictions of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis. Energy and Buildings, Vol 92, pp 10 – 18.
- [13] Cheng, J.C.P and Ma, L.J. (2015). A non-linear case-based reasoning approach for retrieval of similar case and selection of target credits in LEED projects. Building and Environment. Vol 93, pp 349 – 361.
- [14] Dieterle, S and Bergmann, R. (2014). A Hybrid CBR-ANN Approach to the appraisal of Internet Domain Names. In L. Lamontagne and E. Plaza (Eds.): ICCBR 2014, LNCS 8765, pp. 95–109.
- [15] Waheeb, W and Ghazali, R. (2016). Chaotic Time Series Forecasting Using Higher Order Neural Networks. International Journal on Advanced Science Engineering Information Technology, Vol 6, No.5, pp 624 – 629.

- [16] Nawi, N.M., Hussein, A.S., Samsudin, N.A., Hamid, N.A., Yunus, A.M., Aziz, M.F.A. (2017). The Effect of Pre-Processing Techniques and Optimal Parameters selection on Back Propagation Neural Networks. *International Journal on Advanced Science Engineering Information Technology*, Vol 7, No.3, pp 770 – 777.
- [17] Richter, M.M and Weber, R.O. (2013). *Case-based Reasoning*. Springer-Verlag.
- [18] Biswas, S.K., Chakraborty, M., Singh, H.R., Devi, D., Purkayashta, B. and Das, A.K. (2017). Hybrid case-based reasoning system by cost-sensitive neural network classification. *Soft Computing*, 21, pp 7579 – 7596.
- [19] Murtagh, F. (2014) Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, 31, pp 274- 295.