

Developing Decision Tree based Models in Combination with Filter Feature Selection Methods for Direct Marketing

Ruba Obiedat

King Abdullah II School of Information Technology
The University of Jordan
Amman, Jordan

Abstract—Direct Marketing is a form of advertising strategies which aims to communicate directly with the most potential customers for a certain product using the most appropriate communication channel. Banks are spending a huge amount of money on their marketing campaigns, so they are increasingly interested in this topic in order to maximize the efficiency of their campaigns, especially with the existence of high competition in the market. All marketing campaigns are highly dependent on the huge amount of available data about customers. Thus special Data Mining techniques are needed in order to analyze these data, predict campaigns efficiency and give decision makers indications regarding the main marketing features affecting the marketing success. This paper focuses on four popular and common Decision Tree (DT) algorithms: SimpleCart, C4.5, RepTree and Random Tree. DT is chosen because the generated models are in the form of IF-THEN rules which are easy to understand by decision makers with poor technical background in banks and other financial institutions. Data was taken from a Portuguese bank direct marketing campaign. A filter-based Feature selection is applied in the study to improve the performance of the classification. Results show that SimpleCart has the best results in predicting the campaigns success. Another interesting finding that the five most significant features influencing the direct marketing campaign success to be focused on by decision makers are: Call duration, offered interest rate, number of employees making the contacts, customer confidence and changes in the prices levels.

Keywords—Direct marketing; data mining; decision tree; simpleCart; C4.5; reptree; random tree; weka; confusion matrix; class-imbalance

I. INTRODUCTION

Direct marketing has become a trend topic for academics and researchers over the past few years due to high competition between companies, increasing marketing campaigns costs and the changing demands of customers which make it hard to predict [29] [22]. Direct marketing is about finding the most potential customers for a certain product based on their characteristics, interests, behavior and needs, then trying to make customized marketing campaigns for these customers. All industries aim to increase their returns of marketing campaigns and their sales consequently through using the right marketing channels and techniques directed to the right customers at the right time [15]. Banks present one of the major sectors which have a great pressure to increase profits and reduce costs through using the right marketing strategies [17].

There are two approaches for promotions: mass marketing and direct marketing. Mass marketing uses the

traditional media for promotion such as television, radio, newspapers and broadcast messages to be distributed randomly without any customization [15], [12]. This type of marketing becomes less effective with time because of the great competition and the large number of available products these days along with its high cost. Usually the response rate which presents the percentage of customers who are influenced by the marketing and actually buy the promoted products does not exceed 1% which is considered a very low percentage. It is to be noted that, industries hope to increase this rate using direct marketing [13][29] [22].

Data mining techniques, machine learning and business intelligence present important models which can be used for direct marketing since there is a huge amount of available data about customers stored in the databases [4], [13], [29][32], which makes it impossible to analyze this data manually [15] [2], [20]. This data can be studied and analyzed in order to discover the customers' behavior, interests and pattern of buying. This information presents an important source of data for decision makers to help them predict the most potential customers to focus on with direct marketing and increase the respond rate consequently [13], [29], [2], [12]. This, ultimately leads to better management of the available resources to target these customers [19]. Direct marketing is used widely by many industries especially retailers, banks and insurance companies to promote their product and services such as loans and retirement insurance [13]. The reason why they use it is the massive amount of available data about their customers which is generated on a regular basis in an electronic format [2]. Most of the time, classification data mining approach is applied for this purpose in direct marketing to predict whether the customers are classified as buyers or non-buyers [19]. Nevertheless, the marketers' poor skills and knowledge of the data mining models makes it difficult for them to use these models[29].

This study aims to use a simple and comprehensive data mining model which is easy to be understood by users with little or no technical background, especially that decision makers in this case are usually sales persons and managers who are responsible for the direct marketing decisions and it is hard for them to use, understand and interpret more complex models even if these models have more predictive power. In one way or another, Decision Tree algorithms are the best choice here since the results they give are readable, comprehensive rules which can be translated easily to a natural language as a series of IF-Then statements for marketers instead of

black boxes models.

The main problem in using data mining with direct marketing is the high imbalance in the class distribution; as the response rate for these campaigns is less than 1% which presents the positive examples (buyers and respondent) and the rest 99% is identified as negative percentage. Most data mining algorithms do not behave well with this imbalance [13], [19]. Some studies such as [13] proposed using a learning algorithm which not only classifies examples but can also can compute probabilities and rank the example from most likely to least likely buyers. Hence lift analysis was used for evaluation.

This paper is structured as follows: Section II discusses the related works. Section III identifies the methodological approach followed in this research. Experiments and results are discussed in Section IV and finally conclusions are drawn in Section V.

II. RELATED WORK

This section reviews the main studies that discussed the usage of data mining techniques in direct marketing and highlights the main algorithms they applied along with their obtained results.

A two step approach was followed by [13] in order to discuss the data mining methods used for direct marketing. Firstly, Data mining was used to categorize the current customers into likely buyers and non-buyers in order to focus promotion on the likely buyers then apply the chosen data mining algorithms.

Three data sets taken from three different sources were used by the study for direct marketing. Only a small number of customers were identified as buyers. After that they tried to find the potential customers from the current non-buyers. The first data set was taken from a well known Canadian bank using their promotion for loan product. 90.000 records were studied and each customer has 55 attributes and after preprocessing, 62 attributes were used for data mining. The second data set was taken from a major life insurance company using a registered retirement saving campaign. The data set contains 80.000 customers with 7% identified as buyers and each customer has 10 attributes. The third data set belongs to a company that runs a bonus program for 100 sponsors. The data set contains 104.000 customers with 1% responders and each customer has 299 attributes [13].

The study chose Naïve Bayes algorithm and decision tree C4.5 algorithm with a slight modification to produce Certainty Factor (CF). Lift index was used for evaluation. Ada-boost methods of ensembling classifier were applied before applying the learning algorithms. Results shows that data mining can improve the efficiency of direct marketing in terms of the number of respondents and profit [13].

Other studies such as [11] applied data mining technology in the credit card marketing to help banks use the favorable strategy in finding the target clients based on real data taken from Chinese commercial banks. Firstly, they used K-mean clustering to divide the credit card holders, then built four classification models (C5.0, neural network, chi-squared, classification and regression tree). The result revealed that the decision tree is the best model to obtain the necessary features (e.g. monthly income,

family size and age) for successful credit card direct marketing.

Furthermore, [4] applied a Multi-Layer Perception Neural Network (MLPNN), Bayesian networks, Logistic Regression (LR) and (C5.0) decision tree in order to increase the efficiency of the marketing campaign. Real-world data of bank deposit was used. Results proved the effectiveness of these algorithm in predicting the best contact channel with the customers for subscribing deposits. Three statistical measures were used for evaluation, which are accuracy, sensitivity and specificity.

The same data set used by this study was collected and used by [17] who applied logistic regression, neural network, decision trees and support vector machine on the data set of the same bank with 22 selected features. Neural Network had the best results regarding the used metrics AUC of 0.8 and LIFT of 0.65. Moreover the results prove that 79% of successful contacts can be achieved by contacting only half of the better classified customers instead of calling all of them. Finally, sensitivity analysis and DT were applied and revealed that three months euribor rate followed by the call direction (inbound or outbound) was the most relevant feature. In addition, [19] also contained real data from a Portuguese bank concerning 17 phone marketing campaigns. Three CRISP-DM iterations were followed. The researchers applied many data mining algorithms such as Naïve Bayes (NB), Decision Trees (DT) and Support Vector Machines (SVM). The results showed that SVM has the highest prediction performance followed by NB and DT respectively. The most relevant feature was the call duration and the month of contact came next. In the same context using the same dataset [18] proposed a divide and conquer strategy using neural network data mining technique in order to divide the problem into a smaller manageable sub-problems. Each sub-problem is characterized by certain features. Experts evaluated the top influential features of the campaign and considered the call direction (inbound/outbound) as the most relevant one.

On the other hand [2] discussed a case study of a rural bank in Ghana. It applied J48 decision tree and Naïve Bayes. The data set contained 1000 instances with 10 features. The experiment found that the DT accuracy was better than NB with 92.5% and 91.6% respectively. Additionally, it identified the number of contacts as the most important attribute for the J48 DT.

Some studies followed two steps analysis starting by clustering the customer according to their characteristics and needs then made the classification models. For example [15] defined a set of users and tried to align them with the most appropriate communication channels and products. It followed two methods which are partitioning and model based prediction. First it clustered the products and channels then used these clusters in order to predict the customers' decision. The best results in term of accuracy and positive ration were obtained using 5 clusters. However, in terms of the classification methods C4.5 decision tree and Naive Bayes were the best. Finally the results showed that the partitioning method alone increased the accuracy, TP and TN values whereas combining the partitioning method with the classification model yielded to higher accuracy.

Other studies followed a comparative approach such as [33], which used a UCI repository data set with 16

attributes and 45,211 instances to compare between different classification techniques in bank direct marketing. The study chose four algorithms which are SVM, LAD-tree algorithm, J48 and Radial Basis Function Network (RBFN). SVM achieved the highest accuracy while RBFN was the worst one with percentages of 86.95% and 74.34%, respectively.

In general it can be noticed that most of the previous works focused on applying different data mining techniques and comparing between them in terms of efficiency. Nevertheless, not much attention has been given to complexity issues which present a serious concern here, since it is difficult for decision makers with little technical background to understand the complex relationships between the considered attributes. Therefore, this work attempts to cover this gap by focusing on applying a simple model which is easy to interpret since the decision makers in this case are managers and sales person who are not technical employees in the first place, which made DT the most appropriate option.

III. METHODOLOGY

There are many methodologies that can be adopted for constructing the data mining model. This paper follows a five stage methodology framework that aims to examine and modify the prediction model. This process of data mining is useful, simple and flexible to many people who have fair experience in the field of data mining . Fig. 1 below illustrates the proposed methodology of this research.

The main five ideas this experiment is interested in are shown through the following steps:

- **Feature Selection:** Also Known as Attribute selection. It is a useful method to reduce the number of attributes by illuminating the irrelevant attributes that do not highly affect the utility of data [10]. Using Feature selection techniques reduces the computation time, simplify the model and reduces the over-fitting. In Weka, there are three options for performing attribute selection which are using the attribute selection tab directly, using a meta-classifier and using the filter approach. This experiment used the meta-classifier option and the select attributes tab to obtain the numerical weight of each attribute.
- **Tree based Models Building:** This paper discusses four types of decision trees classification algorithms (SimpleCart, C4.5, RepTree and Random Tree). Decision trees are considered one of the most powerful and common tools for classification and prediction. Decision trees produce rules, which can be understood and interpreted easily by humans working in any domain.
- **Performance Evaluation:** This study has used the most common model evaluation metric such the accuracy, True Positive Rate (Recall), Precision, F-Measure and ROC area which are all derived from the confusion matrix without the need of any manual calculation. Moreover, this study reviews an additional evaluation metrics for the

evaluation of a model's efficiency in the presence of highly imbalanced data. It does so by applying the Geometry Mean (G-Mean).

- **Feature Analysis:** This step compares the achieved result of the feature selection and data reduction techniques of the most important 5,10 and 15 top attributes. This method selects the best 5 attributes from the total 21 attributes.
- **Rules Analysis:** This presents the last step in the methodology framework in which the most important rules are extracted as a series of IF-THEN statements relying on the tree with the best results. These rules highlight the most significant features to be focused on by decision makers.

A. Constructing the Prediction Model

Decision trees are one of the most commonly used models in machine learning and decision analysis; they help in determining the most successful strategy to reach the target. They are considered a predictive method which can be used for both classification and regression models. Decision trees are a supervised approach that seeks to find the relationship between input attributes and output attribute (class label) for optimal prediction [16].

The idea of the decision trees can be presented as the tree structure, where each node represents an attribute, each branch represents an outcome of the test, and each leaf node denotes a class label. The decision tree classifier traces the path from the root which is the main attribute of the set to the leaf node, which represents the class label [27]. The decision trees algorithm has a statement "if ... then ... else ..." construction which makes it easy to read and interpret. Moreover, Decision trees algorithms have different features which and this difference causes a difference in their results.

In this paper, different decision tree algorithms were used to predict the bank direct marketing campaigns.

which are:

- **C4.5:** This algorithm is developed by Ross Quinlan and is used to generate a decision tree [5]. It is an extension of ID3 algorithm that solves most of its problems, like dealing with noise and missing data and it is often used as a statistical classifier. C4.5 builds a decision tree based on the gained information. The attribute with the highest information gain is used as the splitting criteria. Moreover, C4.5 uses Gain Ratio for attribute selection criteria. This method contains two concepts which are Gain and Split Info. In other words, for continuous attributes this selection criteria gives the best result compared to ID3, which is only appropriate for discrete datasets [27]. Nevertheless, C4.5 has few disadvantages like the small variation in data, which causes different decision trees in addition to and the fact that it is not suitable for small training set [5]
- **RandomTree:** This is a supervised classifier developed by Leo Breiman and Adele Cutler. It can handle both classification and regression problems [8]. During The classification, each input feature is classified with all the trees in the forest,

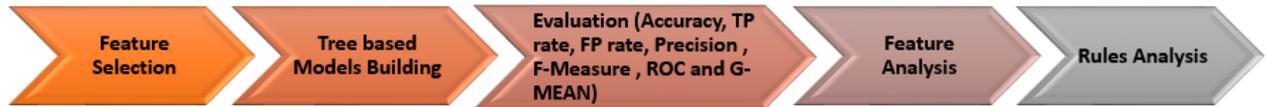


Fig. 1. Proposed Methodology.

and the class label will be the output of the majority. In regression problems, the classifier response is the average of the responses over all the trees in the forest [9].

- SimpleCart: CART is a prediction algorithm that was developed in the early 80s in Southern California by Leo Breiman [8]. It is considered as Classification and Regression Tree that uses historical data in order to generate a binary decision tree. It can operate with categorical or numeric attributes and this distinguishes it from other decision trees methods [26], [8]. One of the advantages of CART method is its strength to outliers. While splitting the algorithm it will isolate outliers in individual nodes. CART algorithm works as follow: Constructing the maximum tree which is the most time consuming part then choosing the right tree size and finally performing the classification of new data using a constructed tree [30]. The CART methodology includes automatic class balancing, handles missing values and allows for cost-sensitive learning, dynamic feature construction, and probability tree estimation [14].
- REPTree: Reduced Error Pruning Tree (REPT) is a fast decision tree algorithm. It applies regression tree logic and creates multiple trees in different iterations then finally selects the best one as the final tree. REPTree builds a decision tree based on the information gain and prunes it using reduced error pruning [8]. Pruning techniques have been used to minimize the complexity of tree structure without reducing the accuracy rate of classification. The basic of the REPTree is sorting values for numerical attribute once and handling the missing values by using C4.5's method of using fractional instances[9].

B. Data Description

The dataset is taken from direct marketing campaigns of Portuguese banking institution. It was collected and prepared by S.Moro,R.Laureano and P.Cortez [19], [17]. The dominant marketing campaigns were based on phone calls. The dataset contains (41188) instances and (20) attributes with one output attribute (target). All the available attributes in the dataset and their description are presented in Table 1 [19], [17].

As shown in Table 1, there are three kinds of attributes, which are Categorical, Numerical and Binary. The target attribute (Y) is binary with two classes which are "yes" which indicates that a deposit subscribed by clients and "no" which indicates that no deposit was subscribed by any clients. This dataset has 4640 clients with class label "yes" and 36545 clients with class label "no".

C. Evaluation Measures

A comparison between these algorithms is performed based on some standard performance metrics which are accuracy, precision, True Positive rate (TP) and F-measure based on the confusion matrix of each tree. The confusion matrix is a table that contains a summary of the prediction results of the classification system [31]. A confusion matrix for a binary classifier is shown in Table 2. It includes data about the actual and predicted values obtained by the classification model [24].

True Positive(TP) presents the number of the correct observation assigned to the positive class. True Negative (TN) presents the correct observation assigned to the negative class. The False Positive (FP) is the number of incorrect observations assigned to the positive class. Finally, False Negative (FN) presents the number of incorrect observations assigned by the model to the negative class [31] [3].

A classifier accuracy reflects it's overall prediction correctness and is defined as the number of the correct predictions to the total number of predictions. Accuracy is given by the Formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

True Positive Rate in machine learning refers to sensitivity or recall. It is used to measure the percentage of actual positives which are correctly predicted as positive. Recall is given by the Formula:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

Precision is a good measure to determine how precise the model is, and to tell the number of actual positive class among the predicted positive ones. The high precision indicates a small number of FP. Precision is given by the Formula:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

F measure represents both recall and precision with the formula:

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

Moreover, the Receiver Operating Characteristic (ROC) has been considered in the present study as one of the most commonly used metrics to evaluate the performance of classification models. ROC curve is presented by plotting the true positive rate (Y-axis) against the false positive rate (X-axis). An optimal model will have a ROC value of 1.0 [31].

The used dataset considered class-imbalanced data since class "yes" (minority) has very low proportions in

TABLE I. LIST OF ATTRIBUTES AND THEIR DESCRIPTION.

Attributes	Type	Attributes Descriptions
Age	Numeric	Clients age
Job	Categorical	("admin.", "blue-collar", "entrepreneur", "housemaid", "management", "retired", "self-employed", "services", "student", "technician", "unemployed", "unknown")
Marital	Categorical	("divorced", "married", "single", "unknown"; note: "divorced")
Education	Categorical	("basic.4y", "basic.6y", "basic.9y", "high.school", "illiterate", "professional.course", "university.degree", "unknown")
Housing	Categorical	("no", "yes", "unknown")
Loan	Categorical	("no", "yes", "unknown")
Contact	Categorical	("cellular", "telephone")
Month	Categorical	("jan", "feb", "mar", ..., "nov", "dec")
day_of_week	Categorical	("mon", "tue", "wed", "thu", "fri")
Duration	Numeric	last contact duration in seconds
Campaign	Numeric	number of contacts performed during this campaign and for this client
Pdays	Numeric	number of days that passed by after the client was last contacted from a previous campaign
Previous	Numeric	number of contacts performed before this campaign and for this client
Poutcome	Categorical	("failure", "nonexistent", "success")
emp.var.rate	Numeric	employment variation rate - quarterly indicator
cons.price.idx	Numeric	consumer price index - monthly indicator
cons.conf.idx	Numeric	consumer confidence index - monthly indicator
euribor3m	Numeric	euribor 3 month rate - daily indicator
nr.employed	Numeric	number of employees - quarterly indicator
Y	Binary(categorical)	has the client subscribed a term deposit?("yes", "no")

TABLE II. CONFUSION MATRIX

		Predicted Values	
		Positive(1)	Negative(0)
Actual Values	Positive(1)	TP	FN
	Negative(0)	FP	TN

the dataset compared to the class “no” (majority class). Hence, the classifiers are skewed towards the majority class and performed unfairly on the minority class. In such a case, other performance evaluation metrics should be applied in addition to the accuracy. One of the most popular techniques that solve this problem with good evaluating measures is Geometric Mean (G-Mean).

G-Mean is a metric that measures the balance between classification performances on both the majority and minority classes. A low value of G-mean indicates that the positive cases are weakly categorized even if negative cases are correctly classified [1]. G-Mean is given by the equation:

$$G - Mean = \sqrt{sensitivity * specificity} \quad (5)$$

Sensitivity (6) is also called true positive rate or recall. It measures the ratio of actual positives that are correctly classified as positive, while specificity (7) is also called true negative rate that measures the ratio of actual negatives that are correctly classified. .

$$Sensitivity = \frac{TP}{TP + FN} \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

IV. EXPERIMENTS AND RESULTS

This work has used a bank telemarketing dataset from UCI machine learning repository which consists of 41188 instances and 21 attributes collected by [19], [17] then applied four different decision tree algorithms (C4.5, REPTree, RandomTree and SimpleCart). Moreover, dataset is divided using K-Fold cross validation which is one of the most popular methods for evaluating the performance of classification algorithms, especially when the volume of the data set is large [7]. In the Cross validation technique, the data set is divided randomly into K of approximately equal parts(folds). The first fold is used as a testing set, and the remaining K-1 folds are used as training set. This process is repeated K times

until each fold has been used as the testing set. Then the model accuracy is calculated as the average of the obtained accuracy in each round [31]. The K value must be chosen wisely. It is usually set to 5 or 10 folds. As K increases, the overlap between training sets also increases. Choosing the value of K equal 10 is more likely and very common to be used because it makes predictions using 90% of the data [25]. Therefore, in this paper the data is split using a 10-fold cross validation to evaluate the predictive model performance.

For easy understanding of the learning process, there is a need to work with an algorithm which gives a maximum classification accuracy rate with simple structure in case of the existence of a huge set of data .

In this experiment, the algorithms are implemented on “Weka”, which is an open-source tool written in java used for data mining tasks. It was developed at the University of Waikato in New Zealand, and it can be executed on many platforms, like Windows, Linux and Macintosh operating systems [28].

Weka provides an easy interface and implementations to different learning algorithms for regression, classification, clustering, association rule mining and attribute selection that can be applied to new datasets [6], [28]. All algorithms import the input file in the form of ARFF format. In this experiment, Windows 10 operating system with 8GB RAM was used to run Weka 3.9.3.

Table 3 presents the experimental results of all the proposed decision tree algorithms applied on the bank dataset. These values represent the rare class “yes”.

Based on accuracy, the following can be observed:

- When applying the SimpleCart algorithm, the correctly classified instances values are proposed directly by Weka software without the need for a manual calculation. 37664 instances from the total 41188 instances are correctly classified, which indicates an accuracy equal to 91.44%. The rate of the true positive values is equal to 0.552, and it has a G-mean with 0.728.
- C4.5 and REPTree algorithms show good accuracy results too with 91.19% and 91.07% values respectively. Additionally, the G-mean values for C4.5 is 0.719 and for REPTree is 0.702.
- When comparing the RandomTree algorithm to

TABLE III. PREDICTIVE METRICS FOR ALL DT ALGORITHMS

Model	Accuracy(%)	TP-Rate/Recall	FP- Rate	Precision	F-Measure	ROC Area	G-Mean
SimpleCart	91.44	0.552	0.040	0.639	0.593	0.903	0.728
C4.5	91.19	0.538	0.041	0.627	0.580	0.884	0.719
REPTree	91.07	0.517	0.039	0.626	0.566	0.903	0.702
RandomTree	88.62	0.475	0.062	0.495	0.484	0.726	0.677

the other used algorithms, it proposed The Correctly Classified Instances value equal to 36499 from the total 41188 instances with a percentage of 88.61%, so it can be seen that it had the least accuracy rate among all other values.

After evaluating the accuracy results, it has been found that SimpleCart and C4.5 have a competitive performance with the highest accuracy of classification compared to the other trees algorithms (REPTree and RandomTree). SimpleCart classified instances 0.25% more accurate than C4.5, which makes SimpleCart algorithm the best model with respect to accuracy.

TP rate and FP rate are also reviewed to compare the results of the different classifiers. The TP rate and FP rate values for SimpleCart are (0.552, 0.040), (0.538, 0.041) for C4.5, (0.517, 0.039) for REPTree and finally (0.475, 0.062) for RandomTree. This shows that SimpleCart has scored the highest TP rate while RandomTree has scored the lowest TP rate. By comparing between the results of the TP rate and FP rate of all the algorithms it is obvious that all these algorithms perform a better prediction for the positive cases. Examining other performance measures, such as the precision and F-Measure of all the algorithms, has showed very close differences in the results. The highest precision value is 0.639, and it is scored by SimpleCart, while RandomTree had the lowest precision value of 0.495. Also Simplecart has scored the highest F-Measure value which is 0.593, while C4.5, REPTree and RandomTree have scored (0.580, 0.566 and 0.484) respectively.

It is also to be noted that the measurement of the experimental result based on the Receiver Operating Characteristic (ROC) that are also presented in Table 3, shows that SimpleCart and REPTree have an equal values of 0.903. As observed, these values are the highest values among all other trees, followed by C4.5 with a value of 0.884 and RandomTree with a value of 0.726. This indicates that SimpleCart and REPTree predictive models can distinguish between the true positives and negatives with a good result that is nearest to the optimal classification point. Moreover, these models are compared based on G-Mean values that were calculated manually according to the equation (5). SimpleCart has also scored the highest G-Mean value of 0.728 compared to the other trees algorithms.

In addition, during the analysis of these algorithms two parameters are taken into consideration; which are the model construction time and the tree complexity. In terms of complexity, Table 4 presents a comparison between all the proposed trees. The tree complexity is clearly governed by the use of the stop criteria and the pruning process. However, the complexity of the tree is generally measured by the following measurements: the total number of nodes (tree size), the total leaf, the depth of tree, the number of attributes that are used in [16]. As shown in

Table 4, SimpleCart produces 47 total numbers of nodes while REPTree, C4.5 and RandomTree produce 992,1143 and 15505 nodes, respectively. Therefore, SimpleCart is better than all other trees in term of classification accuracy (i.e. the number of instances correctly classified) besides the tree size complexity, which presents an important factor affecting the algorithm efficiency, especially with decision tree classifiers.

Furthermore, the time needed to build the model has been taken into account, As shown in table 4, even though SimpleCart can classify the instances more accurately, it might crash on for larger datasets. Therefore, for large datasets, SimpleCart may be an ineffective algorithm. From the obtained results, the following conclusions are drawn:

- RandomTree is much faster than SimpleCart, as it needs much less running time.
- Although RandomTree does not accurately classify instances as SimpleCart does, it retains larger datasets while SimpleCart crashes.
- Due to RandomTree's ability to handle larger datasets, it can be used for processing unstructured data and for large-scale analysis.

A. Feature Selection and Importance Analysis

After applying all the classification models using all 21 attributes of the analyzed dataset, Weka provides a method for Attribute selection. Attribute selection is the process of removing the irrelevant attributes of the data mining task. Also, it aims to search for a main set of attributes that produce comparable classification results with all used attributes [23]. Even though the accuracy is high, the number of attributes used is relatively high. Hence, Weka is used to reduce the number of attributes to get a relatively better accuracy. Since SimpleCart is the best model according to the performance and tree size, three different selection attributes methods are applied on it, which are:

- InfoGainAttributeEval, which evaluates the relevance of an attribute by measuring the information gain of the attribute with respect to class label[21].
- ChiSquaredAttributeEval, which evaluates the relevance of an attribute by computing the value of the Chi-Squared statistic with respect to the class label[21].
- CorrelationAttributeEval, which evaluates the relevance of an attribute by measuring Pearson's correlation between it and the class label.

Table 5 demonstrates a comparison between the three different methods to extract the most important attributes

TABLE IV. PERFORMANCE OF TREES IN TERM OF COMPLEXITY

Dataset	Instances	Attributes	Algorithm	Time to Build Model(sec.)	Size of Tree
Portuguese banking institution.	41188	21	SimpleCart	19.14	47
			C4.5	1.56	1143
			REPTree	0.66	992
			RandomTree	0.2	15505

and the obtained performance values that were derived from the confusion matrix except for G-Mean which was calculated manually.

By using the reported results in Table 5, it can be seen that the classification accuracy of the SimpleCart model achieved a highest percentage value of 91.4732% when reducing the number of attributes to 10 and by using ChiSquare selection attributes methods. This study will take into consideration G-Mean metric to evaluate the performance of the SimpleCart when reducing the number of attributes, since the dataset has imbalanced classes and G-mean is the best measurement to rely on when the class distribution is imbalanced. Table 5 shows the highest readings of the G-Mean when using Information Gain and ChiSquare for selecting the top 5 most relevant attributes with equal values of 0.736. Fig. 2 shows a performance comparison among the three different selection methods.

Table 6 presents the top 5 ranked features that were obtained directly from Weka by using *Select attributes* Tab. As can be noticed, Information Gain and ChiSquare selection methods have the same G-Mean value because their results provide the same features but with a different order. Moreover, it can also be observed that the attributes *duration*, *euribor3m* and *nr.employed* are common in the three selection methods, so they are considered the most important features for SimpleCart model. *duration* Indicates that a long contact with clients (in seconds) can increase the probability of successful deposit campaigns. Next comes *euribor3m*, which is short for Euro Interbank Offered Rate, and it is a very important reference for rates in the European markets. The offered euribor rate is for three months and is updated daily. Finally, an interesting outcome indicates that the number of employees (*nr.employed*) who make the calls and contact the clients has an influence on the probability of subscribing a successful deposit.

However, Information Gain and ChiSquare nominated *cons.price.idx* and *cons.conf.idx* (as a monthly average) attributes, meaning that economic indicators like changes in the price levels and the customer confidence in the current and future economy may lead them to save more than to spend.

It is also found that *Pdays* and *emp.var.rate* are influenced and controlled by the decisions of the bank managers. Hence, it can be seen that managers can increase the deposit rate when considering these variables (i.e the number of days after the last interaction with the customer from a previous campaign and employment variations rate).

B. Extracting Interesting Rules

In this part of the study, experiments extract the most important rules in the previously built tree based models. This step is very important to give an insight for decision makers and to assist them in taking efficient decisions

utilizing these extracted features. Features have been reduced from 21 features to 5 features which is almost a 75% reduction, and it has been found that the reduction in the number of attributes has achieved better results as presented in Table 5. Moreover, reducing attributes to 5 simplified the practical use of the Simple Cart model for marketers and managers and enabled them to use it in their marketing campaigns.

The most important extracted rules of the top 5 attributes for Simple Cart model are illustrated in algorithm 1. There are 18 *if... then* statements. Take the first statements for example; they can be explained as follows: "If the quarterly average of the total number of the employees is below 5087, bank managers should consider two important features for best response from clients, which are call duration and the euribor rate. If the call duration with the client is less than 172 seconds then the response of the client for depositing money in the bank will not succeed, while when increasing the call duration from 172 to less than 250 seconds and the euribor3m below 0.71649 then the model predict a successful campaign". But if the bank employee performed a long call with the client for a period of time longer than 250 seconds, then the model predicts a successful response. According to these statements the bank managers should pay attention to these three features to bring high profits.

The second part of the algorithm takes into consideration other important features which are constant confidence index *cons.conf.idx* and constant price index *cons.price.idx* in addition to the call duration and euribor rate features. This can be explained as follows: "Given that in the bank number of employees who makes calls more than 5087 (as a quarterly average of the total number of employees) and the call duration is less than 606.5 seconds and the constant confidence index of the client is above -46.65 then the model will predict unsuccessful response from the client. While a successful response prediction was obtained for a call duration more than 835.5 seconds and a constant price index less than 93.956. The analysis above indicates that using data mining technology in the direct marketing campaigns especially in bank sectors, is valuable and will lead to useful and great profits with high competition."

V. CONCLUSIONS

This paper investigates experimentally four types of tree based classification algorithms for predicting the bank direct marketing campaign performance. The classifiers are: SimpleCart, C4.5, RepTree and Random Tree. This type of classifiers was chosen because of its interpretability, flexibility and prediction power. The results show that the best results were achieved using SimpleCart model with an accuracy of 91.44% a precision of 0.639 % and a recall of 0.552%. Furthermore, a feature analysis study is conducted based on different feature selection methods to gain an insight on which variables have more influence

TABLE V. COMPARISON BETWEEN THE NUMBER OF THE SELECTED ATTRIBUTES AND THE PERFORMANCE

Attributes Selection Methods	Accuracy(%)	Recall	FP-Rate	Precision	F-Measure	ROC Area	G-Mean
Best Top 5 Attributes							
Information Gain	91.211	0.567	0.044	0.620	0.592	0.902	0.736
ChiSquare	91.211	0.567	0.044	0.620	0.592	0.902	0.736
Correlation	91.4611	0.519	0.035	0.652	0.578	0.929	0.708
Best Top 10 Attributes							
Information Gain	91.4708	0.546	0.038	0.643	0.590	0.908	0.725
ChiSquare	91.4732	0.546	0.038	0.643	0.591	0.906	0.725
Correlation	91.4465	0.511	0.034	0.654	0.574	0.922	0.703
Best Top 15 Attributes							
Information Gain	91.3761	0.542	0.039	0.638	0.586	0.900	0.722
ChiSquare	91.3761	0.542	0.039	0.638	0.586	0.900	0.722
Correlation	91.4198	0.545	0.039	0.640	0.589	0.903	0.724

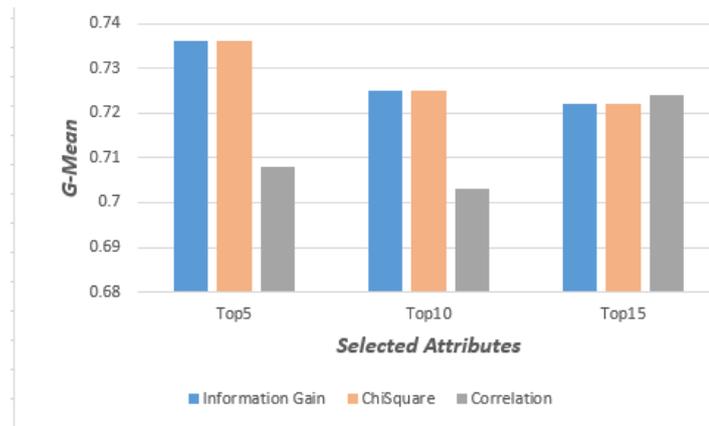


Fig. 2. Comparison

TABLE VI. TOP 5 RANKED ATTRIBUTES

Attributes Selection Methods	Top 5 Ranked Attributes	
Information Gain	0.1094127	duration
	0.1025698	euribor3m
	0.0980036	cons.price.idx
	0.0976254	cons.conf.idx
	0.0896296	nr.employed
ChiSquare	7896.1793	euribor3m
	7618.4243	cons.price.idx
	7570.0707	cons.conf.idx
	7518.7686	duration
	6955.9686	nr.employed
Correlation	0.405	duration
	0.355	nr.employed
	0.325	pdays
	0.308	euribor3m
	0.298	emp.var.rate

in the investigated problem. Best results were gained using top 5 selected features. This analysis showed that the most influencing features are call duration, offered interest rate, number of employees, changes in the prices levels and customer confidence. Such information can be very useful to decision makers, as it can enhance direct marketing campaign, increase the number of clients who subscribe the deposit and lead to a better management of the available resources by focusing on these most influential features. As future work other session features that had not been discussed in the study and may affect the Direct Marketing success can be addressed. Furthermore,

this study's results can be evaluated against other sectors. In addition, future work can discuss the effect of these features on different customer segments or investigate different marketing channels rather than phone calls.

REFERENCES

- [1] Josephine Akosa. Predictive accuracy: a misleading performance measure for highly imbalanced data. In *Proceedings of the SAS Global Forum*, pages 2–5, 2017.
- [2] Anokye Acheampong Amponsah and Kwaku Agyepong Pabbi. Enhancing direct marketing using data mining: A case of yaa

Algorithm 1: The extracted decision rules

```

if nr.employed < 5087.65 then
  if duration < 172.5 then
    | no
  else
    if duration < 250.5 then
      if euribor3m < 0.71649 then
        | yes
      else
        if duration < 202.5 then
          | no
        else
          if euribor3m < 1.0470 then
            if duration < 234.5 then
              | yes
            else
              | no
            end
          else
            | no
          end
        end
      end
    else
      | yes
    end
  end
else
  if duration < 606.5 then
    if cons.conf.idx < -46.6500 then
      if euribor3m < 1.5255 then
        if euribor3m < 1.3685 then
          if duration < 223.0 then
            | no
          else
            | yes
          end
        else
          | no
        end
      else
        if duration 173.5 < then
          | no
        else
          | yes
        end
      end
    else
      | no
    end
  else
    if duration < 835.5 then
      if euribor3m < 1.402499 then
        | yes
      else
        | no
      end
    else
      if cons.price.idx < 93.956 then
        | yes
      else
        if duration < 1614 then
          if euribor3m < 4.862 then
            | no
          else
            | yes
          end
        else
          | yes
        end
      end
    end
  end
end

```

- asantewaa rural bank ltd. in ghana. *International Journal of Computer Applications*, 153(7), 2016.
- [3] Mohamed Bekkar, Hassiba Khelouane Djemaa, and Taklit Akrouf Alitouche. Evaluation measures for models assessment over imbalanced data sets. *J Inf Eng Appl*, 3(10), 2013.
 - [4] Hany A Elsalamony. Bank direct marketing analysis of data mining techniques. *International Journal of Computer Applications*, 85(7):12–22, 2014.
 - [5] Bhumika Gupta, Aditya Rawat, Akshay Jain, Arpit Arora, and Naresh Dhama. Analysis of various decision tree algorithms for classification in data mining. *International Journal of Computer Applications*, 163(8):15–19, 2017.
 - [6] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.
 - [7] Yoonsuh Jung and Jianhua Hu. Ak-fold averaging cross-validation procedure. *Journal of nonparametric statistics*, 27(2):167–179, 2015.
 - [8] Sushilkumar Kalmegh. Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news. *International Journal of Innovative Science, Engineering & Technology*, 2(2):438–446, 2015.
 - [9] Sushilkumar Rameshpant Kalmegh. Comparative analysis of weka data mining algorithm randomforest, randomtree and ladtree for classification of indigenous news data. *International Journal of Emerging Technology and Advanced Engineering*, 5(1):507–517, 2015.
 - [10] D Lavanya and Dr K Usha Rani. Analysis of feature selection with classification: Breast cancer datasets. *Indian Journal of Computer Science and Engineering (IJCSE)*, 2(5):756–763, 2011.
 - [11] Wei Li, Xuemei Wu, Yayun Sun, and Quanju Zhang. Credit card customer segmentation and target marketing based on data mining. In *2010 International Conference on Computational Intelligence and Security*, pages 73–76. IEEE, 2010.
 - [12] Ying Li, Pavankumar Murali, Nan Shao, and Anshul Sheopuri. Applying data mining techniques to direct marketing: Challenges and solutions. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 319–327. IEEE, 2015.
 - [13] Charles X Ling and Chenghui Li. Data mining for direct marketing: Problems and solutions. In *Kdd*, volume 98, pages 73–79, 1998.
 - [14] Venkatadri Marriboyina. A comparative study on decision tree classification algorithms in data mining. 04 2018.
 - [15] Merve Mitik, Ozan Korkmaz, Pinar Karagoz, Ismail Hakki Toroslu, and Ferhat Yucel. Data mining based product marketing technique for banking products. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 552–559. IEEE, 2016.
 - [16] W Nor Haizan W Mohamed, Mohd Najib Mohd Salleh, and Abdul Halim Omar. A comparative study of reduced error pruning method in decision tree algorithms. In *2012 IEEE International conference on control system, computing and engineering*, pages 392–397. IEEE, 2012.
 - [17] Sérgio Moro, Paulo Cortez, and Paulo Rita. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62:22–31, 2014.
 - [18] Sérgio Moro, Paulo Cortez, and Paulo Rita. A divide-and-conquer strategy using feature relevance and expert knowledge for enhancing a data mining approach to bank telemarketing. *Expert Systems*, 35(3):e12253, 2018.
 - [19] Sergio Moro, Raul Laureano, and Paulo Cortez. Using data mining for bank direct marketing: An application of the crispdm methodology. In *Proceedings of European Simulation and Modelling Conference-ESM'2011*, pages 117–121. EUROSIS-ETI, 2011.
 - [20] Sérgio Moro, Raul Laureano, and Paulo Cortez. Enhancing bank direct marketing through data mining. In *Proceedings of the 41th European Marketing Academy Conference (EMAC)*. European Marketing Academy, 2012.
 - [21] Jasmina Novaković. Toward optimal feature selection using ranking methods and classification algorithms. *Yugoslav Journal of Operations Research*, 21(1), 2016.

- [22] Youqin Pan and Zaiyong Tang. Ensemble methods in bank direct marketing. In *2014 11th International Conference on Service Systems and Service Management (ICSSSM)*, pages 1–5. IEEE, 2014.
- [23] Nikita Patel and Saurabh Upadhyay. Study of various decision tree pruning methods with their empirical comparison in weka. *International journal of computer applications*, 60(12), 2012.
- [24] Tina R Patil, SS Sherekar, et al. Performance analysis of naive bayes and j48 classification algorithm for data classification. *International journal of computer science and applications*, 6(2):256–261, 2013.
- [25] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. *Encyclopedia of database systems*, pages 532–538, 2009.
- [26] Philip J Sallis, William Claster, and Sergio Hernández. A machine-learning algorithm for wind gust prediction. *Computers & geosciences*, 37(9):1337–1344, 2011.
- [27] Shiju Sathyadevan and Remya R Nair. Comparative analysis of decision tree algorithms: Id3, c4. 5 and random forest. In *Computational Intelligence in Data Mining-Volume 1*, pages 549–562. Springer, 2015.
- [28] Trilok Chand Sharma and Manoj Jain. Weka approach for comparative study of classification algorithm. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(4):1925–1931, 2013.
- [29] Lilian Sing'oei and Jiayang Wang. Data mining framework for direct marketing: A case study of bank marketing. *International Journal of Computer Science Issues (IJCSI)*, 10(2 Part 2):198, 2013.
- [30] Roman Timofeev. Classification and regression trees (cart) theory and applications. *Humboldt University, Berlin*, 2004.
- [31] Efraim Turban, Ramesh Sharda, and Dursun Delen. *Business intelligence and analytics: systems for decision support*. Pearson Higher Ed, 2014.
- [32] Chakaran Vajiramedhin and Anirut Suebsing. Feature selection with data balancing for prediction of bank telemarketing. *Applied Mathematical Sciences*, 8(114):5667–5672, 2014.
- [33] K Wisaeng. A comparison of different classification techniques for bank direct marketing. *International Journal of Soft Computing and Engineering (IJSCE)*, 3(4):116–119, 2013.