A Methodology for Improving Complex Sales Success in CRM Systems

Doru Rotovei
West University of Timisoara
Computer Science Department
Timisoara, Romania
Email: doru.rotovei80@e-uvt.ro

Viorel Negru
West University of Timisoara
Computer Science Department
Timisoara, Romania
Email: vnegru@info.uvt.ro

***Abstract*—In this paper we propose a methodology for extracting complex sales expert rules by analyzing the data from the past lost/won deals stored in Customer Relationship Management Systems.**

We first used Multi-Adaptive Regression Splines model to identify the features importance, then we created a classification tree of lost/won sales using Random Forest and lastly we used the tree for extraction of the expert rules that gives insights into the rules of successful complex sales.

The proposed methodology was successfully validated using complex sales data from a CRM application and the results are presented and discussed in this paper.

***Index Terms*—Customer Relationship Management, Classification, Multi-Adaptive Regression Splines,Expert Systems, Random Forests, Decision Trees, Java Expert System Shell**

I. Introduction

A complex sale, such as buying an airplane, has the following properties [1]:

- high price tag
- high risk for the buyer
- multiple decision makers involved
- long life cycle (i.e. it takes a long time to close a deal)
- low volume (i.e. any sales representative sells fewer items in any given month compared to traditional non-complex sales)

Forecasting complex sales is a challenging operation mainly due to the multiple stages necessary to conclude a transaction successfully [8], [7], [9]. Consequently, a successful transaction depends on a series of factors, features and predictors that we studied in this paper using the Customer Relationship Management (CRM) Systems Data that captured the past lost/won complex deals.

Our study aimed to deal with the following research questions: 1) If we can classify won/lost complex deals, what is the feature importance? 2) Can we build a classification model for won/lost complex deals using data from past closed deals found in the CRM Systems? 3) What are the strategies the sales representatives can use to shift a potential losing deal towards a wining deal?

To address these research questions, we 1) used Multi-Adaptive Regression Splines (MARS) to extract feature importance 2) used Random Forest (RF) to create a classification tree of lost/won deals 3) used Java Expert System Shell (JESS) to extract expert rules using the RF model.

978-1-5090-5795-5/17/$31.00

The rest of the paper is structured as follows. In section II we present the current state of the art in CRM data mining and association rules extraction. In section III we introduce our methodology along with our data set used for validation. Here we also present the results obtained using the proposed methodology. In section IV we discuss the results and in section V we present our conclusion.

II. Literature Review

In this section we review previous CRM related research using data mining, forecasting and association rules extraction.

In paper [2], the authors introduced a 6 year long study on the use of Data Mining with CRM Systems. The paper identifies four dimensions of CRM systems: Customer Identification, Customer Attraction, Customer Retention and Customer Development and seven data mining functions, among them being Forecasting and Regression. Logistic regression is used in all four dimensions and Linear Regression was used in Customer Development. For forecasting only Markov Chain Model was used.

Independent Component Analysis (ICA) was proposed in [10] to screen features before creating an ANN model to predict sales in e-CRM. The screening of the feature was useful to improve the ANN model performance compared to other classifiers especially when there is a lot of noise in the data as in the case of data that comes from the internet. The ICA, which is similar to Principal Component Analysis (PCA), proved to be a successful preprocessing tool especially for CRM data.

Four classification algorithms, logistic regression, multilayer perceptron neural network, k-nearest neighbor and SVM were used in [11] to create response modeling for direct marketing i.e. model those customers more likely to purchase a marketing campaign product. The authors proposed an ensemble method based on clustering and under-sampling to improve the response models.

Much of the literature related to forecasting has been focused on creating models that identify the overall sales trend and value [7], [8], [9], [12], [13]. K-mean cluster and fuzzy neural networks (KFNN) where used in [7] to predict the

sales of a printed circuit board factory. The study compared the KFNN to the following models: Feed Forward Neural Network, Radial Basis Function Neural Network, Back Propagation Neural Network and Winter Exponential Smoothing and it shows that KFNN is a superior model in predicting sales.

One of the characteristics of the fashion retail is a very short product life cycle. A hybrid model using neural networks model coupled with a harmony search algorithm was proposed in [8] and compared to Auto-Regressive Integrated Moving Average (ARIMA) was superior in predicting fashion sales. Fashion retail forecasting was also studied in [9] were an evolutionary neural network was proposed that proved to be superior in predicting sales compared to the traditional Seasonal Auto-regressive Integrated Moving Average (SARIMA) models.

A new forecasting model using wavelet support vector machines (WN v-SVM) is proposed in [13]. Particle swarm optimization (PSO) was used to select the parameters of the WN v-SVM and the proposed model was used to forecast the sales of cars. A comparison with other traditional forecasting models was researched as well, proving the new model is superior to the transitional ones.

MARS and ANN were compared in [12] to forecast the sales of computer wholesalers in Taiwan. The study showed that MARS can be a good alternative in constructing forecasting models with the added benefit of identifying the importance of the feature used in building the model.

Bankruptcy forecasting using MARS was explored in [16] along with Fuzzy C-Means clustering to construct a hybrid model that outperformed ANN and Discriminant Analysis. The hybrid model was tested against data from the banking sector in Spain and it proved to be successful in helping lending decisions and profitability.

MARS and ANN were used in a a two-stage hybrid model to predict credit scores and credit risk in [20]. The hybrid model using MARS to identify the important variables and ANN to create the model based on those variables outperform logistic regression and discriminant analysis.

A new approach named Algorithm Learning Based Neural Network (ALBNN) is used in [25] to improve the learning and parameter adjusting of a normal ANN. ALBNN is using the knowledge gathered by different learning algorithms to extract the relevant initial points and accelerate the ALBNN model building. The new approach can be used successfully when high classification accuracy is needed.

Hybrid SVM techniques were used in in [21]. More precisely CART and SVM and MARS and SVM were used to predict credit scoring for a bank in China. The study found the hybrid model MARS+SVM for feature selection and model building outperformed the SVM, MARS and CART+SVM models.

Automated extraction of expert rules from databases was explored in [24] based on rough set theory and a survey of hybrid expert system is presented in [22]. The survey found a trend towards neuro-fuzzy and rough neural expert system.

The CRM transaction data from a small sized online shopping mall were used in [15] to create a clustering model of VIP and non-VIP customers based on frequency monetary value (RFM). The transaction were analyzed in the context of VIP and non-VIP customers and association rules were identified and proposed to improve future sales.

Association rules, as a pre-processing step, were research in [19] to predict customer churn along with back propagation neural networks and C5.0 decision trees. Furthermore, a new algorithm called Goal-oriented sequential pattern was proposed in [23] to discover patterns and behaviors of customers that might leave the business (customer-losing patterns). Furthermore, the authors propose retaining strategies based on the new knowledge gathered by analyzing network banking churn.

III. Methodology

In any CRM System that manages complex sales, actionable knowledge is vital for increased sales. Prior to our research, the company used for our study had limited knowledge about the patterns linked to successful closed deals.

To gather actionable knowledge associated to successful closed deals, our study answers three questions:

1) What are the features that most influence a won deal?
2) Are the past data collected during the previous deals enough to classify accurately lost/won deals?
3) What are the patterns or rules that can be implemented to increase the win rate of complex sales?

We used three different models to answer these questions in sequence order see Figure 1:

1) To understand what features play a more important role in successfully closing a deal, we created a MARS model to extract the feature importance.
2) To classify lost/won deals we used the feature importance found by MARS along with Random Forest to build a classification tree that we validated against a test data set.
3) Lastly we extracted expert rules associated with won deals. Having this knowledge can assist sales representatives shift a deal that could be lost into a winning deal. Here we used JESS and human experts to interpret the RF tree built at the previous step.

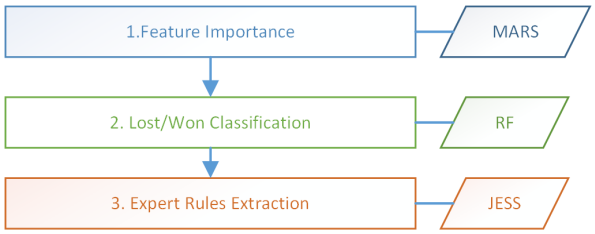

Fig. 1. Methodology Steps

Below we give an overview of the theoretical background used for the models along with the metrics used before introducing the data set and the experiment results using the proposed methodology.

A. MARS

Multi-Adaptive Regression Splines (MARS) model is a non-linear model that is formed by creating a piecewise of multiple linear models built using multiple regression equations that partition the input dataset [6]. The cut point or the hockey sticks are determined by using all the data points for each feature and the predictor that gives the smallest model error is chosen for the cut point. The process of using the next feature with its data points continues in a divide and conquer way until a stop point is reached that can be set by the user.

MARS model general equation is [6]:

$$f(x) = \beta_0 + \sum_{i=1}^{N} \beta_i h_i(X) \tag{5}$$

where $h_i(X)$ is a candidate function or a product of two or more functions from a set C of candidate functions.

$$C = \{(X_j - t)_+, (t - X_j)_+\} t \in \{x_{1j}, x_2 j, ..., x_N j\}$$
$$j = 1, 2, ..., p \tag{6}$$

where N is the number of observations and p is the number of predictors or features.

The piecewise linear basis functions are defined:

$$(x-t)_+ = \begin{cases} x - t & if x > t \\ 0 & otherwise \end{cases} \tag{7}$$

$$(t-x)_+ = \begin{cases} t - x & if x < t \\ 0 & otherwise \end{cases} \tag{8}$$

where $+$ means only the positive side of the function.

β_i are coefficients that represents the importance of the features and are calculated using Residual Sum of Squares (i.e. on average how far is the predicted value from the actual value).

In comparison to other feature selections algorithms [17], [18], MARS has few advantages [6]:

- MARS models are easier to interpret and understand which is important for the sales representatives using the CRM System
- MARS models do not require data preparation, therefore the algorithm can be applied directly on the raw data
- MARS models can handle categorical and continuous data
- MARS models select automatically the features that are important and exclude from the model the unimportant ones

B. Decision Trees and Random Forests

Classification and Regression Tree (CART) is a statistical model used to create solutions for classification or regression problems [28]. In building the tree, CART is using all the features and all the data points to find the split of the tree.

For the training sample:

$$S_n = \{(X_1, Y_1), ...(X_n, Y_n)\}, X \in \mathbb{R}^p, Y \in \mathbb{R} \tag{9}$$

where p are the number of features in the input set and Y is the predicted scalar value. The split is found statistically when the two partitions, S_1 and S_2, are found to minimize the Sum of Square Errors (SSE) [26]:

$$SSE = \sum_{i \in S_1} (y_i - \overline{y}_1)^2 + \sum_{i \in S_2} (y_i - \overline{y}_2)^2 \tag{10}$$

where $\overline{y}_1$ and $\overline{y}_2$ are the means within each group S_1 and S_2. The split continues recursively within each subclass of S_1 and S_2 until a threshold is reached.

Regression trees in general suffer from a few weaknesses[3]:

- instability - they are dependent on the data and data changes can affect the layout of the tree
- sub-optimal predictive performance due to strong partitioning of the data in clear rectangular spaces.
- selection biased - the feature with the most distinct values is selected
- over-fitting trees can over fit the data which leads to a lower predictive performance

To improve the predictability of regression trees, ensemble methods have been proposed [3]. One such ensemble method is Random Forest (RF).

RF is a bagging algorithm i.e. a sample is used to bootstrap the algorithm [27]. More concrete Random Forests trees are built by creating m models using k randomly selected predictors to bootstrap the algorithm with. Also in building the trees only a subset of the features are used called *mtry*. Subsequently the aggregation of the final tree is built by averaging the trees.

The Random Forests algorithm adds more smoothness in choosing the tree features and split points compared to a simple classification tree like CART[3].

The RF tree was used as as an intermediary step to extract the Exert Rules of successful complex deals.

C. Data Set and Performance Criteria

The data used to validate the proposed methodology represents complex sales results from a period of 7 years (2009-2016) with 598 sales where 424 deals were lost and 124 won. We extracted the data from the SQL database of the CRM System used to manage the complex deals.

In building our models we used R language along with the appropriate packages: "earth" for MARS and "randomForest" for the RF model. Also Java Expert System Shell (JESS) was used for rule extraction.To extract the accuracy and F1-measures we used 70% of the data for model building and 30% to collect the accuracy and F1-measures.

The predictors, or features used for our study are presented in Table I.

The following metrics were used to evaluate the models:

TABLE I
LIST OF FEATURES USED TO BUILD THE MODELS

Feature	Explanation
NoUsers	Number of users (1, 2-4, 5-10, 11-50, +50)
SalesCycle	Number of days it took to close the deal
Revenue	Potential revenue the deal can generate
LeadType	Can be: web lead, email lead, referral
LeadVendor	A number indicating the lead generator engine used if any
Timezone	The time zone of the potential customer like EST, PST, Australia etc.
ProbabilityToClose	The probability the deal will be won
PrimaryUser	The primary sales representative assign to the individual opportunity
NoOfActivities	The number of activities performed to close the deal like email, fax, phone call etc.
OpportunityStatus	The opportunity status **Won=1** or **Lost=0** to competition. This is the variable we create the model for

1) Accuracy:

$$A = \frac{N_c}{N_t} \tag{11}$$

where N_c is the number of correctly classified examples and N_t is the number of the total test samples. Accuracy ranges from 0-1 with 1 being the perfect classifier.

2) Precision:

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

Where TP is the True Positives and FP is the False Positives. Precision is a measure of the classifiers exactness, also called Positive Predictive Value (PPV). For a good classifier we are looking for a high precision that corresponds to a large number of True Positives.

3) Recall:

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

Where TP is the True Positives and FN is the False Negatives. Recall is a measure of the classifier completeness also called Sensitiviy or True Positive Rate (TPR). Recall measures how many items are predicted compared to how many should have been predicted.

4) F1 Measure:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{14}$$

F1 measure captures the balance between precision and recall in one metric, the harmonic mean of the two.

5) Kappa:

$$kappa = \frac{Accuracy - RandomAccuracy}{1 - RandomAccuracy} \tag{15}$$

$$RandomAccuracy = \frac{(TN + FP) \times (TN + FN)}{Total^2} + \frac{(FN + TP) \times (FP + TP)}{Total^2} \tag{16}$$

Kappa compares the accuracy of a system to the accuracy of a random system. Kappa is a normalized measure with values between 0-1 that can be used to compare different models[5].

D. Experiment Results

1) Feature Importance Extraction Using MARS: The first step in our methodology is to discover what is the order of the feature importance. The MARS model we created revealed the feature importance presented in table II.

We notice that some features are more important than others, for example, Sales Cycle or Revenue whereas Lead Vendor or Lead Type were not used to build the MARS model therefore we removed them from the training data set on the next step.

TABLE II
MARS FEATURE IMPORTANCE

Feature	Overall
NoOfActivities	100.000
PrimaryUser	33.306
SalesCycle	28.924
Revenue	9.153
TimeZone	0.000
LeadType	0.000
NoOfUsers	0.000
LeadVendor	0.000

2) Random Forest Classification Model: Knowing the feature importance helps sales representatives focus on the important features that affect the success of a complex deal. The next step is to create a classification model of lost/won deals. Based on the MARS feature importance we excluded from the input data the features that are not important and used only the important ones.

To build the RF model we used 500 trees with 3 variables tried at each split that converged with the tree shown in Figure 2. The two confusion matrices extracted during training and using the test data set are presented in Table III. In this context 0 means a lost deal and 1 means a won deal. Furthermore the evaluation metrics results against the test data set are shown in Table IV.

TABLE III
RF MODEL CONFUSION MATRIX

During Training			On Test Data Set		
	0	**1**		**0**	**1**
0	331	2	**0**	142	2
1	5	82	**1**	0	34

TABLE IV
RF MODEL CLASSIFICATION RESULTS ON THE TEST SET

Measure	Value
Accuracy	0.9888
Kappa	0.9644
Precision	1.0000
Recall	0.9444
F1-Measure	0.9930

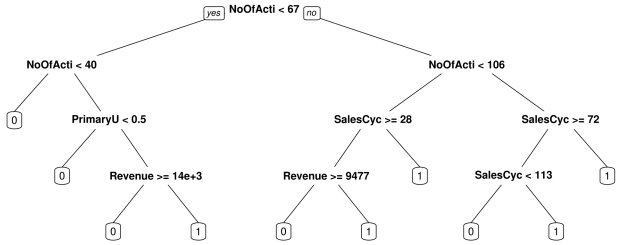

Fig. 2. Random Forests Output Tree

The RF model is used to extract expert rules that could help shift a potential losing deal towards a winning deal.

3) JESS: Java Expert System Shell (JESS) is a Rule Based Engine for Java.[1] As a last step in our methodology we used JESS to create rules that can improve the win rate of the complex deals. We extracted the rules by consulting sales experts based on the RF model. One such rule is present in listing 1.

The sales experts helped validate the rules of the RF model and also curate the rules based on the relevance to the sales representative.

For example the Number of Activities is in the control of the sales representative and therefore by controlling the Number of Activities the sales representative can influence the win rate whereas the potential Revenue is in general fixed during the potential sale. For that reason selecting a JESS rule based on the Number of Activities is more relevant.

Listing 1. Jess Rule for increase activities

```
(defrule increase_activities_to_win
  "Increase the chance to win
the deal by increasing the number
  of activities"
   (< ?NoOfActivites 106)
   (< ?SalesCycle 72))
=>
   (printout t "Recommendation to win:
      Increase number of activities
        to win the deal.")
)
```

Additional relevant rules extracted are presented in Table V.

IV. Discussion

MARS model revealed that not all features are equally important and we discovered that, for example, the Number of Activities has a bigger impact on the success of an individual sale compared to the Lead Vendor or the Time Zone of the potential customer. Although activities such as phone calls, email, fax etc. have a high importance in predicting the success of a deal, there is a tipping point where too many activities can actually be detrimental to win the deal. In our study if the number of activities bypass 106 and the sales cycle is under 113 the deal could be lost.

[1]http://www.jessrules.com

TABLE V
Rules Extracted Based on the RF Model

Rule	Recommendation
If NoActivities $>$ 67	Close the deal in under 70 days.
If Revenue $<$ 14000 and NoActivities $<$ 40	Increase the number of activities to win the deal.
If NoActivities $<$ 106 and Revenue $<$ 9477	Keep control of the sales cycle and close in under 28 days.

The Random Forest algorithm helped build a decision tree from which we extracted Expert Rules. The rules extracted were manually constructed using Sales Experts. However a future direction of research is the automation of the rule extraction.

We note that there are features that are not directly modifiable and therefore not in the control of the sales representative like for example the potential Revenue. On the other hand there are features the sales representative can influence like for example the Number of Activities or setting an accurate Probability to Close. The automatic rule engine needs to differentiate between those, something the human Sales Experts intuitively took into account when building the knowledge in the Expert Rules shown in Table V.

We notice that the RF model was able to predict with high accuracy the lost/won deals when verified against the test data set. We can infer that, in this particular case, the sales representatives and the deals they are closing have a predictable pattern. However as the economies change, new products are brought to the market or sales team develops, the model will decay. Therefore a periodic model rebuilding is necessary, especially when new people are added to the sales team or new products are introduced.

Few factors influence the winning rate of the complex deals in the CRM System used for our research: Number of Activities, the length of the Sales Cycle and how big the deal is (Number of Users or Revenue). Our recommendation was to pay close attention to these features, have tools and reports in place to monitor them for each individual deal and, in addition, to present to each sales representative the expert rules found.

V. Conclusion

In this paper we proposed a methodology for expert rule extraction using complex sales data recorded by CRM Systems. In particular we proved that a sequence order of MARS to discover features importance followed by Random Forest classification of lost/won deals along with the extraction of Expert Rules using JESS can provide insights into how to improve the success rate of winning complex deals.

The future direction of research will be to automate the translation of the Random Forest trees into meaningful Expert

Rules. We would also like to study the impact on rule extraction of each individual activity (phone call, email, fax) as time series taking into account the time the activity was performed in relationship to the sales cycle.

ACKNOWLEDGMENT

This work was partially supported by InnoHPC - Interreg, Danube Transnational Programme - grant and VI-SEEM H2020-EINFRA 675121 grant. The views expressed in this paper do not necessarily reflect those of the corresponding projects consortium members.

REFERENCES

[1] Thull, J., 2010. Mastering the complex sale: how to compete and win when the stakes are high!. John Wiley and Sons.

[2] E. W. T. Ngai, L. Xiu and D. C. K. Chau. Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications 36(2), pp. 2592-2602. 2009

[3] Kuhn, Max, and Kjell Johnson. Applied predictive modeling. New York: Springer, 2013.

[4] Cherkassky, Vladimir, and Yunqian Ma. "Practical selection of SVM parameters and noise estimation for SVM regression." Neural networks 17, no. 1 (2004): 113-126.

[5] Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. biometrics, pp.159-174.

[6] Friedman, Jerome H. "Multivariate adaptive regression splines." The annals of statistics (1991): 1-67.

[7] Chang, P.C., Liu, C.H. and Fan, C.Y., 2009. Data clustering and fuzzy neural network for sales forecasting: A case study in printed circuit board industry. Knowledge-Based Systems, 22(5), pp.344-355.

[8] Wong, W.K. and Guo, Z.X., 2010. A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. International Journal of Production Economics, 128(2), pp.614-624.

[9] Au, K.F., Choi, T.M. and Yu, Y., 2008. Fashion retail forecasting by evolutionary neural networks. International Journal of Production Economics, 114(2), pp.615-630.

[10] Ahn, H., Choi, E. and Han, I., 2007. Extracting underlying meaningful features and canceling noise using independent component analysis for direct marketing. Expert Systems with Applications, 33(1), pp.181-191.

[11] Kang, P., Cho, S. and MacLachlan, D.L., 2012. Improved response modeling based on clustering, under-sampling, and ensemble. Expert Systems with Applications, 39(8), pp.6738-6753.

[12] Lu, C.J., Lee, T.S. and Lian, C.M., 2012. Sales forecasting for computer wholesalers: A comparison of multivariate adaptive regression splines and artificial neural networks. Decision Support Systems, 54(1), pp.584-596.

[13] Wu, Q., 2009. The forecasting model based on wavelet -support vector machine. Expert Systems with Applications, 36(4), pp.7604-7610.

[14] Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2(2), pp.121-167.

[15] Shim, B., Choi, K. and Suh, Y., 2012. CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns. Expert Systems with Applications, 39(9), pp.7736-7742.

[16] De Andrs, J., Lorca, P., de Cos Juez, F.J. and Snchez-Lasheras, F., 2011. Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). Expert Systems with Applications, 38(3), pp.1866-1875.

[17] Molina, L.C., Belanche, L. and Nebot, ., 2002. Feature selection algorithms: A survey and experimental evaluation. In Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on (pp. 306-313). IEEE.

[18] Kumar, Vipin, and Sonajharia Minz. "Feature Selection." SmartCR 4, no. 3 (2014): 211-229.

[19] Tsai, C.F. and Chen, M.Y., 2010. Variable selection by association rules for customer churn prediction of multimedia on demand. Expert Systems with Applications, 37(3), pp.2006-2015.

[20] Lee, T.S. and Chen, I.F., 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, 28(4), pp.743-752.

[21] Chen, W., Ma, C. and Ma, L., 2009. Mining the customer credit using hybrid support vector machine technique. Expert systems with applications, 36(4), pp.7611-7616.

[22] Sahin, S., Tolun, M.R. and Hassanpour, R., 2012. Hybrid expert systems: A survey of current approaches and applications. Expert Systems with Applications, 39(4), pp.4609-4617.

[23] Chiang, D.A., Wang, Y.F., Lee, S.L. and Lin, C.J., 2003. Goal-oriented sequential pattern for network banking churn analysis. Expert Systems with Applications, 25(3), pp.293-302.

[24] Tsumoto, S. and Tanaka, H., Automated Extraction of Expert System Rules from Databases based on Rough Set Theory. Multistrategy Learning, p.313.

[25] Yoon, H., Park, C.S., Kim, J.S. and Baek, J.G., 2013. Algorithm learning based neural network integrating feature selection and classification. Expert Systems with Applications, 40(1), pp.231-241.

[26] Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A., 1984. Classification and regression trees. CRC press.

[27] Breiman, L., 2001. Random forests. Machine learning, 45(1), pp.5-32.

[28] Marsland, S., 2015. Machine learning: an algorithmic perspective. CRC press.