Customer Relations Management using J48 Tree, Ranking Algorithm, and Chi-Square

Marzieh mohammadi
Faculty of Computer Engineering,
Najafabad branch, Islamic Azad
University, Isfahan, Iran

Hamid Rastegari
Faculty of Computer Engineering,
Najafabad branch, Islamic Azad
University, Isfahan, Iran

Abstract— **Customer relations management (CRM) integrates all business activities to identify and manage customers in order to increase sales in the long run and thus raise the value of companies. Efficient administration of CRM requires the recognition of appropriate patterns within the existing datasets. Analysis of such patterns will then enable managers and analysts to make the best possible decisions in critical situations. Pattern recognition is one of the fundamental goals of data mining techniques. Decision trees are popular data mining approaches commonly used as prediction models. The present study proposed a model which utilized both classification (based on J48 tree) and feature selection for the accurate prediction of marketing performance. The efficacy of the proposed model was evaluated in three datasets and the results were compared with other widely used data mining algorithms including the reduced-error pruning (REP) tree, random decision tree, support vector machine (SVM), and J48 tree. The results confirmed the higher precision, accuracy, and recall and lower error rate of the proposed model compared to the other four methods.**

Keywords- Customer Relations Management (CRM); Feature Selection; Data mining; Classification; J48 Tree; Ranker Algorithm; Chi-square

I. INTRODUCTION

Customer relationship management (CRM) involves a series of processes and technologies utilized by companies to identify and attract potential customers, expand their customer base, provide higher quality services, and ultimately retain customers. CRM principles discuss the methods of successful establishment, implementation, and maintenance of an efficient customer relationship system [1]. As an integrated information system, a CRM system seeks to improve customers' experiences of interaction with a corporate by using an assortment of channels (e.g. corporate's website, email, and telephone) to organize, schedule, and control pre- and post-sale activities. In other words, in its attempt to increase profitability, revenue, and customer satisfaction, a CRM (as a grand strategy) organizes business activities based on different groups of customers and tries to create and promote particular behaviors which can eventually promote customer satisfaction [2]. Such a system will help businesses not only to retain current customers, but also to acquire new groups of customers. Organizations may adopt various methods, including CRM, customer value analysis, and organizational and service provision strategies, to enhance the efficiency of customer relationships and thus attract and retain new customers. CRM can be applied to monitor all activities related to direct customers (e.g. firms), shorten the sales cycle, improve customer loyalty, and increase profit [3].

Data mining is a commonly used approach to manage, organize, and discover predictive patterns existing in a large dataset without the need to involve users. It employs several techniques, such as preprocessing, classification, clustering, and feature selection, and algorithms to provide managers and analysts with the information required for decision-making in critical organizational situations [4]. Identification

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

of customer needs using data mining methods will enable businesses to tailor their services based on the obtained patterns and available information about products, customers, and their interests. Therefore, the relationship between data mining and CRM can play a major role in the success of organizations. A previous study proposed the application of decision trees for the identification and prediction of changes in marketing and sales trends. After evaluating the approach using a dataset extracted from an online shopping website in Korea, the model was found to be capable of helping managers to make appropriate decisions based on customer needs and requests [5].

Decisions trees are one of the oldest and most popular data mining techniques. They generally perform classification by predicting an analog output based on categorical or actual inputs. Owing to their ability to comprehensibly describe the relations within a dataset, decision trees are widely used for classification and prediction purposes in a variety of fields (e.g. marketing) [6]. Since the decision tree (algorithm output) is generated based on the values of a set of selected features, the accuracy of a pattern recognition system largely depends on appropriate feature selection. On the other hand, as higher numbers of features impose additional computational costs on the system, design and implementation of systems with the minimum possible number of features seems essential. Feature selection, i.e. choosing the optimal (or suboptimal) subset of features in the dataset, aims to reduce the decision tree size without decreasing classification accuracy. It is hence crucial to pattern recognition, machine learning, and data mining [7].

The present study applied J48 and ranking algorithms to develop a decision tree for predicting customer behavior and selecting the set of activities that can potentially increase profitability. Chi-square tests were performed for data analysis. The proposed model was then compared with several other data mining algorithms including reduced-error pruning (REP) decision trees, random decision trees (RDT), support vector machines (SVM), and J48 tree. Ross Quinlan introduced C4.5 as an algorithm to create decision trees from a set of labeled training data [8]. Weka open source data mining software incorporates J48 algorithm as an implementation of C4.5.

II. Concepts and Related work

A. Concepts

1) Rep tree

Rep Tree are simple and fast decision trees. After building a tree by using information gain as the splitting criterion, REP is applied to prune the created tree [9,10]. This algorithm sorts values for numeric features only once and handles missing values according to C4.5 for fractional instances [10].

2) Random tree

RDTs are increasingly used as an efficient method in the field of data mining. An RDT is basically created by selecting a random feature at each node. These trees are usually left unpruned and are considered as control trees [9,10]. The general belief is that producing a large group of RDTs can ensure accuracy while preventing the problem of overfitting. Due to random selection of the features, all trees of the group will have equal chance of being sampled [10].

3) SVM

SVM is a kernel method originally developed by Vapnik (1992) based on the probability theory in machine learning [11]. SVM owes its popularity to its accuracy in handwriting recognition (which is equal to that of neural networks). While SVM algorithms have been widely used in various applications, only two instances of them have been applied in direct marketing.

SVM algorithms use clustering, classification, ranking, and data cleaning to recognize complex patterns in a dataset. While these algorithms are highly efficient in some applications, computational complications decrease their efficiency in dealing with large datasets [12,13]. Therefore, SVM classifiers are suitable when the training data have limited features.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

4) Feature Selection

Feature selection techniques are commonly applied in data mining practices to identify and eliminate irrelevant or redundant data and thus reduce dataset size while maintaining the accuracy of predictive models [14]. Such a reduction in the number of features will decrease the complexity and improve the cost-effectiveness, speed, and performance of predictive models [15]. Feature selection techniques can be categorized as filter methods (e.g. chi-square test, information gain, and correlation coefficient scores), wrapper methods (e.g. random hill-climbing algorithm and forward selection and backward elimination), and embedded methods [e.g. regularization techniques such as the least absolute shrinkage and selection operator (LASSO), elastic net, and ridge regression].

B. Related Work

Numerous studies have assessed the applicability of various data mining methods in different fields of business [16]. Previous research has also focused on appropriate intelligent data mining methods to improve CRM [17-18]. Several techniques including association rules, clustering, classification, and sequence discovery have been adopted to deal with customer complaint and loyalty issues (as components of CRM) and enhance customer retention. Any classification method in data mining comprises a learning phase (during which a classifier is produced by using a predefined set of data classes) and the actual classification phase [19].

In business studies, decision trees are widely used to predict future customer behaviors and describe sequences of interrelated decisions based on patterns extracted from available customer data [1,20,21]. Kim et al. proposed a decision tree-based method and evaluated its efficacy using a Korean online shopping center. They concluded that the method could detect changes in customer behaviors not only in structured situations (where the manager was interested in a specific research matter), but also under dynamic conditions [21]. Bin et al. tried to develop an accurate and efficient model based on decision trees to predict customer churn among Personal Handyphone System Service users. They performed three experiments to construct an accurate and effective customer churn model and reported the optimal model to use random sampling for training set selection, a sub-period time of 10 days, and a misclassification rate of 1:5 [22].

III. THE PROPOSED MODEL

Data mining techniques are widely employed to create frameworks covering all aspects of CRM [16]. Such a process comprises four stages. First, the problem has to be analyzed. The identification of hidden patterns in customer behaviors during this stage will facilitate decision making for business managers. In the next stage, i.e. data preparation stage, the required data should be collected from different sources and issues related to missing data and outliers should be carefully dealt with. The prepared data will be used in the third stage to build an appropriate model for CRM. In the final stage, the developed model is validated to ensure its applicability in CRM [10]. In the present study, a model based on J48 decision tree is proposed. We believe that the model should have a high accuracy in making predictions in the field of CRM. The proposed model will involve three phases including preprocessing, feature selection, and classification with J48 tree (Figure 1).

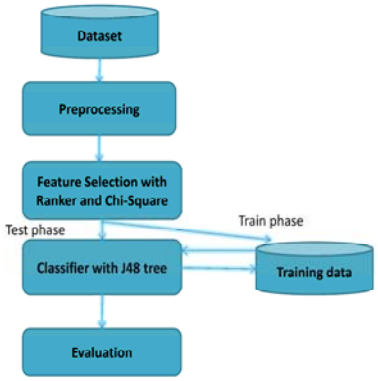

Figure 1: The proposed model

A. Preprocessing phase

Since data preparation and filtering can potentially be very time-consuming,data pre-processing, including normalization, cleaning, transformation procedures [23], is identified as a critical stage of all data mining processes. Normalization (e.g. rescaling) is an essential data preprocessing approach [14] with well-known benefits for prediction purposes [24]. The existing normalization methods, such as min-max, z-score, and decimal scaling, are commonly applied to scale different parameters into the same range and thus eliminate the variations in predictions. Our proposed model uses z-scores to normalize unstructured data during the preprocessing phase:

$$V_i' = \frac{V_i - E}{std(E)} \quad (1)$$

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

where *Vi'* is the z-score normalized value and *Vi* is the value in row E and the i^{th} column. If there are five rows (V-Z) each containing *n* columns (or variables), then:

$$\text{Std(E)} = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(vi - \hat{E})^2} \qquad (2)$$

$$\hat{E} = \frac{1}{n}\sum_{i=1}^{n} vi \text{ or mean value}$$

The above-mentioned equations can be used to calculate normalized values in each row. When all values in a particular row are identical, the standard deviation of that row, and thus all the normalized values in that row, will be equal to zero.

B. *Feature selection with a Ranker and Chi Square Test*

Among the various feature selection methods, ranking approached are used to categorize disorganized groups of data based on ranks allocated to each item according to one or more of its features. Rankings are applied for prioritizing tasks or comparing the performance of different products. Although rankings can be easily visualized, the interpretation of the obtained visualizations might be tricky since the rank of a particular object may not fully describe the complicated relations between its features and attributes of other objects. In fact, since numerous rankings can be developed within a specific setting, their comparison is essential to determine the effects of heterogeneous features on each ranking. Moreover, the efficiency of this process needs to be enhanced through the application of advanced visual examinations [25]. Ranking methods are usually in used in combination with techniques to determine feature significance (e.g. Relief-F, gain ratio, and entropy). In order to perform the ranking, a specific threshold is determined for the features and measures above this threshold are selected. A general algorithm will also be required to select one of the best ranking measures with the most suitable features from the classification point of view.

Chi-square test is a non-parametric (distribution-free) test widely employed to compare ratios, frequencies, and percentages in univariate or multivariate problems. This test was incorporated into our proposed model since its non-parametric nature allowed for undetermined parameters, null values, and missing, heterogeneous, and dispersed data.

In models developed by previous research [10, 25], a ranking algorithm is first applied to sort each feature. The ChangesSameValue function is then used to identify similar and different values for a particular feature and compute the maximum number of changed labels. In other words, this function calculates the ranking of each feature as the number of changes in class labels. In our proposed model, the probability table for each feature is first developed based on the frequency of each value in each class. The expected values are then determined based on the probability of the incidence of each value in a specific class. In the next step, equation (3) is used to calculate chi-square from the values in the obtained probability table. Chi-square values for each feature are stored in the variable named NLC and the algorithm is repeated for all features. Ultimately, the values in the NLC are sorted. Figure 2 shows a pseudocode for the proposed method.

The Proposed Algorithm()

Input: E training(N instances, M features)

Output: E reduced(N instances, M ranker features)

For each feature $A_i \in 1..M$ {

(A) Expand the contingency table to have both row totals and column totals and an overall total.

(B) Calculate the "expected values" for each cell in the table based on the probabilities using the totals of each row and column.

(C) $NLC_i \leftarrow$ Organize the frequencies into a new table to calculate X^2.

That $X^2(\text{features}) = \sum_{j=1}^{m} \frac{(fo-fe)^2}{fe}$

}

Sort NLC Feature Ranking

Select the ranked feature

Figure 2: The pseudocode for the proposed method

Chi-square tests are applied to evaluate the features of each node. In fact, chi-square test is a statistical test which compares the occurrence of each feature with its expected occurrence in each class [26]. The independent and dependent variables in chi-square tests are features and classes, respectively. Chi-square value for a particular feature can be calculated as [27]:

$$X^2(\text{features}) = \sum_{j=1}^{m} \frac{(fo-fe)^2}{fe} \qquad (3)$$

where f_o and f_e are respectively the observed and expected occurrences of the feature in a particular class. M is the number of features. Figure 3. The flowchart of the pseudocode for the proposed feature selection approach

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

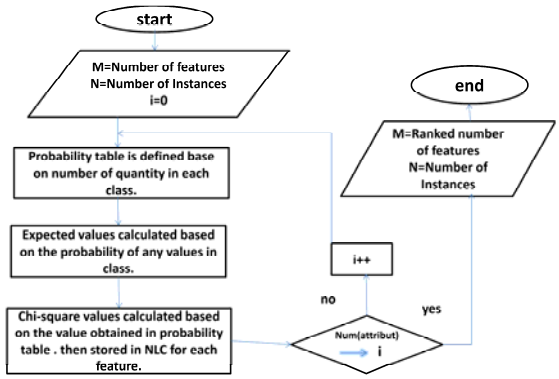

Figure 2: The pseudocode for the proposed feature selection approach

C. *Classifier phase with J48 tree*

As an implementation of C4.5, the J48 decision tree is popular in data mining applications [14]. Since decision trees created by C4.5 can be employed for classification purposes, they are also known as statistical classifiers [9]. The J48 is a pruned tree which considers all possible tests for the classification of features and selects the best test outputs based on entropy and chi-square techniques.
Three steps are generally followed to produce a J48 tree:

a) When all instances belong to the same class, the tree will have a leaf labeled with the mentioned class.

b) A test is performed on the feature and the possible information for each feature is obtained. The information gain is then calculated accordingly. During this stage, entropy is used as an indicator of data disorder:

$$\text{Entropy}(\bar{y}) = -\sum_{j=1}^{n} \frac{|x_{ij}|}{|\bar{y}|} \log\left(\frac{|x_{ij}|}{|\bar{y}|}\right)$$
$$\text{Entropy}(j|\bar{y}) = \frac{|x_{ij}|}{|\bar{y}|} \log\left(\frac{|x_{ij}|}{|\bar{y}|}\right) \quad (4)$$

Information gain can thus be computed as:

$$\text{Gain}(\bar{y}, j) = \text{Entropy}(\bar{y} - \text{Entropy}(j|\bar{y})) \quad (5)$$

c) Finally , the best feature is selected based on the present selection measure and that feature selected for branching [26].

IV. Results and Discussion

The proposed model was validated by its administration on three experimental datasets including Bank-Data.csv [1] , Car.arff[2], and Bank-full.csv[3] (Table 1). Precision, accuracy, recall, and error rate were computed with 10-fold cross-validation (equations 6-9).

$$\text{Precision} = TP/(TP+FP) \quad (6)$$

$$\text{Accuracy} = (TP+TN)/N \quad (7)$$

$$\text{Recall} = TP/(TP+FN) \quad (8)$$

$$\text{Error rate} = (FN+FP)/N \quad (9)$$

where TP is the number of features correctly identified by the classifier; FP is the number of features incorrectly identified by the classifier; TN is the numbers of features correctly rejected by the classifier; FN is the number of features incorrectly rejected by the classifier; and N is the number of features.

Table 1 The used datasets:

TABLE I. DATASET IN USE

Dataset	Number of Instances	Number of Features
Bank-Data.arff	600	16
Bank-full.arff	41189	21
Car.arff	1727	7

The results of the proposed model were compared with those of REP tree, RDT, SVM, and J48 tree. The proposed model was implemented by java Net Beans and the .jar files from WEKA were imported into the source code. All experiments were executed on a computer system with a Core i7 CPU and 8 GB of RAM.

Tables 2 and 3 compare the precision and recall of the proposed model with that of the REP tree, RDT, SVM, and J48 tree. As seen, in all datasets, the use of data normalization and feature ranking by chi-square tests increased the precision and recall of the proposed model compared to the other three algorithms. The second highest precision and recall in all datasets belonged to the J48 tree algorithm which used entropy and all possible tests to identify the best branches and produce a pruned tree. Moreover, in Bank-full and Car datasets, the REP tree

[1] http://facweb.cs.depaul.edu/mobasher/classes/ect584/WEKA/preprocess.html
[2] http://repository.seasr.org/Datasets/UCI/arff/
[3] http://mlr.cs.umass.edu/ml/machine-learning-databases/00222/

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

algorithm had higher precision and recall than the SVM and RDT algorithms. The RDT and REP tree algorithms had the lowest precision in the Bank-full and Bank-data datasets, respectively. Due to the absence of pruning in the RDT algorithm, this method had the lowest recall in both Bank-full and Car datasets. The REP tree algorithm had the lowest recall in the Bank-data dataset.

Table2. Comparison of the proposed model and the reduced-error pruning (REP) decision tree, random decision tree (RDT), support vector machine (SVM), and J48 tree in terms of precision.

TABLE II . THE COMPARISON OF PRECISION WITH REP TREE, RDT,SVM, J48 TREE AND PROPOSED MODEL

dataset	Rep tree	RDT	SVM	J48 tree	The Proposed model
Bank-full.arff	90.5	88.4	88.7	90.7	95.5
Bank-Data.arff	29.5	60.6	61.1	89.9	92.9
Car.arff	88.0	82.7	80.0	92.4	96.9

As it can be seen in table 2, the precision of proposed model is higher than the other algorithm. J48 tree is more precise than Rep tree and Random tree and Rep tree has the precise than Random tree because random tree is not pruning.

Table 3. Comparison of the proposed model and the reduced-error pruning (REP) decision tree, random decision tree (RDT), support vector machine (SVM), and J48 tree in terms of recall

TABLE III. THE COMPARISON OF RECALL WITH REP TREE, RDT, SVM, J48 TREE AND PROPOSED MODEL .

dataset	REP tree	RDT	SVM	J48 tree	The Proposed model
Bank-full.arff	91.1	88.6	90.3	91.2	95.7
Bank-Data.arff	54.3	58.0	61.3	89.8	92.7
Car.arff	87.7	83.2	84.7	92.4	96.8

Table 4 compares the accuracy of the proposed model with that of the REP tree, RDT, SVM, and J48 tree. Apparently, due to the above-mentioned reasons, the proposed model had the highest accuracy in all datasets. Furthermore, the J48 tree algorithm was the second most accurate method in all datasets. In the Bank-full and Car datasets, the REP tree had higher accuracy compared to the SVM and RDT algorithms.

Table 4. Comparison of the proposed model and the reduced-error pruning (REP) decision tree, random decision tree (RDT), support vector machine (SVM), and J48 tree in terms of accuracy

TABLE IV. THE COMPARISON OF ACCURACY WITH REP TREE, RDT, SVM, J48 TREE AND PROPOSED MODEL .

Dataset	REP tree	RDT	SVM	J48 tree	The proposed model
Bank-full.arff	91.1	88.6	90.3	91.2	95.7
Bank-Data.arff	54.3	58,0	61.3	89.8	92.7
Car.arff	87.7	83.2	84.7	92.4	96.8

Table 5 compares the four models in terms of error rate. As expected based on the obtained precision values, the lowest error rate (and thus the highest precision) in all datasets belonged to the proposed model. Data normalization during the preprocessing phase, feature selection, and elimination of redundant and irrelevant features can justify this finding. The second lowest error rate was detected in case of the SVM algorithm. Finally, the REP tree algorithm had a lower error rate than the RDT algorithm.

Table 5. Comparison of the proposed model and the reduced-error pruning (REP) decision tree, random decision tree (RDT), support vector machine (SVM), and J48 tree in terms of error rate

TABLE V. THE COMPARISON OF ERROR RATE WITH REP TREE, RDT, SVM, J48 TREE AND PROPOSED MODEL .

Dataset	REP tree	RDT	SVM	J48 tree	The Proposed model
Bank-full.arff	0.114	0.115	0.097	0.111	0.071
Bank-Data.arff	0.496	0.469	0.386	0.200	0.132
Car.arff	0.063	0.077	0.269	0.058	0.0174

V. CONCLUSION

Nowadays, many of companies spent a lot of money and time on the decisions for marketing their services and products and decision making by discovering interesting patterns in the amount of data has become a key task in today's business background. Generally, companies make a customer prediction model to discovery the prospects for a particular product. Data mining uses the algorithms to find useful patterns and trends from the extracted data so that it able to yield main insights including prediction models and associations can support companies understand their customer well. Analyzing and Examining data able to turn raw data into valuable information about customer's requires. Classifying and features selection are two main techniques of data mining. In this paper, a model base on the classification of J48 tree and feature selection with Ranker is proposed to predict precise marketing performance. The propose model is evaluated conducted 3datasets and the

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

results are compared with other algorithms such as Rep tree, Random tree and J48 tree. The experimental results show that the proposed model has higher precision and lower error rate in comparison of J48 tree, Rep tree and Random tree.

REFERENCES

[1] Perreault Jr, W. D., E. J. McCarthy and J. P. Cannon ,Basic marketing: A marketing strategy planning approach, McGraw-Hill/Irwin, (2006).

[2] Kostojohn, S., M. Johnson, and B. Paulen, Maintaining and Evolving CRM, in CRM Fundamentals. 2011, Springer. p. 197-224.

[3] Jutla, D., J. Craig and P. Bodorik , Enabling and measuring electronic customer relationship management,(2001). Proceeding of the 34th Annual Hawaii International Conference on.2001.IEEE.

[4] Linoff, G.S. and M.J. Berry, Data mining techniques: for marketing, sales, and customer relationship management. 2011: John Wiley & Sons.

[5] Quinlan, J.R., Induction of decision trees. Machine learning, 1986. **1**(1): p. 81-106.

[6] Kikuti, D., F. G. Cozman and R. Shirota Filho , Sequential decision making with partially ordered preferences, Artificial Intelligence (2011).175(7): p. 1346-1365.

[7] Dash, M. and H. Liu, Feature selection for classification. Intelligent data analysis, 1997. **1**(1): p. 131-156.

[8] Kim, J.K., et al., Detecting the change of customer behavior based on decision tree analysis. Expert Systems, 2005. **22**(4): p. 193-205.

[9] Witten, I.H. and E. Frank, Data Mining: Practical machine learning tools and techniques. 2005: Morgan Kaufmann.

[10] Mohammadi .M, H. Rastegari, Using J48 tree for value-based customer ralations management (CRM),2015vol.13 ,no.7 :p. 49-54.

[11] Drucker, H.,S. Wu, and V.N. Vapnik, Support vector machines for spam categorization.Neural Network,IEEE Transactions on, 1999. 10(5): p. 1048-1054.

[12] FENG, Y. and H. ZHOU, An Effective and Efficient Two-stage Dimensionality Reduction Algorithm for Content-based Spam Filtering* . Journal of Computational Information System, 2013. 9(4): p. 1407-1420.

[13] Liu, H., J.Li, and L. Wong, A comparative study on feature selection and classification methods usinggene expression profiles and proteomic patterns. Genome informatics, 2002. 13:p. 51-60

[14] Gholap, J., Performance tuning of J48 Algorithm for prediction of soil fertility. arXiv preprint arXiv:1208.3943, 2012.

[15] Hall, M.A., Correlation-based feature selection for machine learning. 1999, The University of Waikato.

[16] Ranjan, J. and V. Bhatnagar, Data mining tools: a CRM perspective. International Journal of Electronic Customer Relationship Management, 2008. **2**(4): p. 315-331.

[17] Wong, K.W., et al. Intelligent data mining and personalisation for customer relationship management. in Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th. 2004. IEEE.

[18] Han, J., M. Kamber, and J. Pei, Data mining: concepts and techniques: concepts and techniques. 2011: Elsevier.

[19] Ahmed, S.R. Applications of data mining in retail business. in Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on. 2004. IEEE.

[20] Chen, Y.-L., C.-L. Hsu, and S.-C. Chou, Constructing a multi-valued and multi-labeled decision tree. Expert Systems with Applications, 2003. **25**(2): p. 199-209.

[21] Kim, J.K., et al., Detecting the change of customer behavior based on decision tree analysis. Expert Systems, 2005. **22**(4): p. 193-205.

[22] Bin, L., S. Peiji, and L. Juan. Customer churn prediction based on the decision tree in personal handyphone system service. in Service Systems and Service Management, 2007 International Conference on. 2007. IEEE.

[23] Patro, S., et al., Technical Analysis on Financial Forecasting. arXiv preprint arXiv:1503.03011, 2015.

[24] Gratzl, S., et al., Lineup: Visual analysis of multi-attribute rankings. Visualization and Computer Graphics, IEEE Transactions on, 2013. **19**(12): p. 2277-2286.

[25] Ruiz, R., J.C. Riquelme, and J.S. Aguilar-Ruiz. Fast feature ranking algorithm. in Knowledge-Based Intelligent Information and Engineering Systems. 2003. Springer.

https://sites.google.com/site/ijcsis/
ISSN 1947-5500

[26] Yerazunis, W.S., et al. A unified model of spam filtration. in Proceedings of the MIT Spam Conference, Cambridge, MA, USA. 2005.
[27] Subramaniam, T., H.A. Jalab, and A.Y. Taqa, Overview of textual anti-spam filtering techniques.