Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Implementation of Dynamic Mutual Information and Support Vector Machine for Customer Loyalty Classification

IOP Conf. Series: Journal of Physics: Conf. Series **1338** (2019) 012050 doi:10.1088/1742-6596/1338/1/012050

Implementation of Dynamic Mutual Information and Support Vector Machine for Customer Loyalty Classification

H Sulistiani[1,2,a], K Muludi[1] and A Syarif[1]

[1]Departement of Computer Science, Faculty of Mathematics and Sciences, Lampung University, Jl. Sumantri Brojonegoro No.1 Bandar Lampung, Indonesia, 35145
[2]Departement of Accounting Information System, Faculty of Computer Science and Engineering , Teknokrat Indonesia University ,Jl. ZA. Pagar Alam No. 9-11, Bandar Lampung, Indonesia, 35145

[a]henisulistiani@teknokrat.ac.id

Abstract. Fast Moving Consumer Goods (FMCG) is known one of the important industrial sectors worldwide. It includes household and personal care products as well as processed foods and beverages. Because of the tight competition company must develop good marketing strategies. So, it is important for the company to know customer loyalty and also to predict the income as reference in company development planning. Data mining now is becoming popular technique for predicting customer loyalty. One of the well known data mining strategies is retaining customer's strategy. In this paper, we would present a new model for predicting customer loyalty. The model is based on Dynamic Mutual Information and Support Vector Machine (DMI-SVM) to identify the relevant factors that affect the performance of the classification of customer loyalty. The comparison of two classification methods and several selected features is given to show the effectiveness of the methods. We validated the model by 10-fold cross validation method. Classification accuracy, precision, recall and f-measure are used to evaluate classifier performance on a test/hold-out sample. A result in this paper is shown that SVM method gives better performance accuracy than Naïve Bayes.

1. Introduction

Fast Moving Consumer Goods (FMCG) is known as one of the important industrial sectors worldwide. It is includes household and personal care products as well as processed foods and beverages [1]. FMCG entities are growing rapidly and are very active organizations with produce various goods. Company must develop a good marketing strategies because of the intense business competition. Hence, the company must know customer loyalty and revenue predictions as references in the company's development planning. Customer loyalty is the asset of the longest surviving company [2].

The technique used to predict customer loyalty is data mining, which is currently very popular. One of the well-known data mining strategies is retaining strategy. Research field of customer loyalty analysis using

IOP Conf. Series: Journal of Physics: Conf. Series **1338** (2019) 012050 doi:10.1088/1742-6596/1338/1/012050

data mining has been done by many researchers. To predict the customer loyalty, several researchers have mainly presented the following methods decision tree [3], Naïve Bayes and nearest neighbour [4], logistic regression [5], Bayesian network [6], artificial neural network [7]. This paper proposes support vector machine (SVM) method to predict customer loyalty. This method has been show to be an accurate method to separate customer data using hyperplane generated by SVM. It is show to know closest the data vector plane, optimal separating hyperplane and classification decision function can be guaranteed [5].

To perform predictive analysis required data with a large number. A lot of data will have excessive or irrelevant attributes, so it must be done deletion or attribute reduction before further processing of data. Technique used to reduce the number of attributes is to perform the selection of attributes or features. There are many benefits gained from the feature selection process, among other things facilitating data visualization and understanding of data. Otherwise, it can reduce storage, low training time and reduce data dimension to improve performance of prediction analysis [8]. Several methods are for selecting feature are gain ratio, chi square and information gain [9]. However, the feature selection method in that research is not suitable for dataset with missing value. Other method for feature selection is mutual information. This method measures the dependency value of two variables [10]. The advantages from mutual information are can estimate the dependency value between two variables that not only measure linearly (Pearson's correlation) or monotonic relationship (Spearman's correlation) and can measure complicated non-linear relationships [11]. One of the methods developed from mutual information is dynamic mutual information (DMI). This method divides the feature groups into two types: labelled and unlabelled [12]. DMI method can minimize information distortion and measuring different types of relationships.

In this paper, we present a new model for predicting customer loyalty. The model is based on Dynamic Mutual Information and Support Vector Machine (DMI-SVM) to identify the relevant factors that affect the performance of the classification of customer loyalty. The comparison of two classification methods and several selected features are given to show the effectiveness of the methods. From the literatures review, the proposed method has never been applied to the predictions of FMCG customer loyalty.

The structure of the rest is organized as follows. In section 2 presents some basic concepts about DMI method and SVM. Section 3 provides implementation and result. Finally, conclusions and future works are given in the end.

2. Basic Concept

2.1. Dynamic Mutual Information

One of the stage in processes data mining is feature selection conducted before the training [13]. In feature selection problem, the relevant features will have important information related to the output (class label), while irrelevant features will have little or no information related to the output [14]. Therefore necessary designing an evaluation function to measure the quality of candidate features so that classification performance increases [15]. One of the methods can measure the quality information is Dynamic Mutual Information method. The method is for feature selection based on information theory. Entropy is a key for measure of information in information theory [16] which capable to measure quantifies the uncertainty of random variables and to scale the amount of information shared by them effectively. Other measurement metrics can only consider linear correlation between features or variables [17]. The uncertainty value of a random variable *X* with discrete values can be calculated using entropy
function *H(X)*, which is defined as:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{1}$$

IOP Conf. Series: Journal of Physics: Conf. Series **1338** (2019) 012050 doi:10.1088/1742-6596/1338/1/012050

Where *p(x)* is the marginal probability distribution of *X*. For measure the joint entropy from variable *X* and *Y* can use the joint *p(x, y)*, which is defined as:

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y) \tag{2}$$

In calculation of entropy or mutual information, the probability density distribution of the random variable must be estimated [18]. Mutual information is the amount of information that both variables share, and it defines as:

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y) log\frac{p(x,y)}{p(x)p(y)} \tag{3}$$

If value of $I(X, Y) > 0$ then variables *X* and *Y* are closely related with each other. Otherwise, $I(X, Y) \leq 0$ then two variables are unrelated. This method should pick out the optimal feature performing best in prediction, without losing characteristics of much sampling data greatly. This method divides the feature in dataset groups into two types: labelled and unlabelled. Before identifying feature, first stage in feature selection to estimate the value of mutual information from candidate feature. Algorithm of the dynamic mutual information can be seen in figure 1.

```
Procedure: Dynamic Mutual Information
input: Training dataset T=D(F, C)
output: Selected Feature S
 (1)    initialize relative parameters: F = F;
        S = Ø; D_u = D; D_l = Ø;
 (2)    Repeat
 (3)      For each feature f ϵ F do
 (4)        calculate mutual information I(C; f) on D_u;
 (5)        If I(C; f) = 0 then F = F – {f};
 (6)      choose feature with highest value I(C; f);
 (7)      S = S U {f}; F = F\{f};
 (8)      obtain new labelled instances D_l from D_u induced by f;
 (9)      remove the feature from D_u,
 (10)   Until F= Ø or |D_u| = I_T
```

Figure 1. Procedure dynamic mutuali information

2.2. *Support Vector Machine*

Solving classification problems based on statistical learning theory can use the non-parametric method, such as Support Vector Machine (SVM) [19]. However, SVM has several main problems [20], including the problem of large amount of data. The second is noise and interaction: data distribution will be difficult to explain and the margin of separation between classes will be unclear. The third is unbalanced data: difference in the number of samples in a class with another class.

Then it needs a model that can handle the problems. SVM is a model derived from statistical learning theory that will provide better results than other methods. In Linear SVM, each training data is known as

(x_i, y_i), where $i = 1,2,\ldots,N$ and $x_i=\{xi_1,xi_2,\ldots,xi_q\}^T$ are set attribute for training data I, $y_i \in \{-1, +1\}$ is a class label.

3. Result and Discussion

The numbers of instances used in this study were 386, consisting of two classes: loyalty and no loyalty. Feature consists of demographic, psychographic, product and promotional data totalling 26 features. List of features used is given the table 1.

Table 1. List of feature

Category	Feature Name
Demographic Data	Age, Gender, Address, Marital Status, Job, Education, Residence Status
Psychographic Data	Switch Brand, Price Satisfaction, Consumption of Other Brands, Consumption Time, Reason to Consume, Reason Switch Brand, Return Behaviour, Average Consumption, Amount of Consumption, Brand Satisfaction
Transaction Data	Total Purchase, Total Expenses, Distance of Purchase, Place of Purchase
Product Data	Product Display, Brand
Promotional Data	Promotion Media, Recommendation, Comment

Three test scenarios were proposed in this paper. The first scenario discuses the feature selection used dynamic mutual information. The second scenario, to verify the validity of a selected feature, this paper carries out classifications using two classifiers i.e. the SVM and Naïve Bayes. Features used for classification are also extracted into two types, namely:

a. Use all the features;
b. Use the selected feature by dynamic mutual information method.

The last scenario is comparing the performance from two classification algorithms. Those scenarios can be seen in Figure 2.

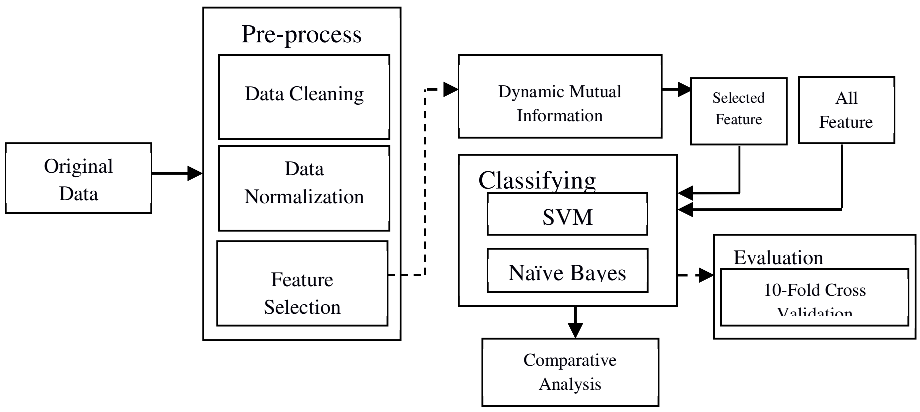

Figure 2. Research methodology

The evaluation criterias are accuracy (Acc), precision (P), recall (R) and F-Measure (F) [21]. To test the accuracy of required classification models we used a method 10-fold cross validation. Calculation of evaluation criteria can use the following equation.

IOP Conf. Series: Journal of Physics: Conf. Series **1338** (2019) 012050 doi:10.1088/1742-6596/1338/1/012050

$$Acc = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$P = \frac{TP}{TP+FP} \tag{5}$$

$$R = \frac{TP}{TP+FN} \tag{6}$$

$$F = 2\frac{P*R}{P+R} \tag{7}$$

Table 2. Result and feature ranking

Feature	MI Value
Age	1.22
Comment	0.02
Total Expenses	0.02
Average Consumption	0.02
Address	0.02
Recommendation	0.01
Return Behavior	0.01
Brand Satisfaction	0.01
Product Display	0.01
Reason Switch Brand	0.01
Distance of Purchase	0.01
Reason to Consume	0.01
Consumption of Other Brands	0.01
Education	0
Place of Purchase	0
Total Purchase	0
Brand	0
Job	0
Consumption Time	0
Price Satisfaction	0
Amount of Consumption	0
Residence Status	0
Gender	0
Promotion Media	0
Switch Brand	0
Marital Status	0

Denotes:
TP (true positive): the number of positive examples correctly classified as positive,
TN (true negative): the number of negative examples correctly classified as negative,

FP (false positive): the number of negative examples falsely classified as positive,
FN (false negative): the number of positive examples falsely classified as negative.

Feature selection technique applied in this paper uses feature ranking with threshold > 0.01 [22]. Feature with a value of less then 0.01 should be excluded from analyzed dataset. The results of assessment and ranking feature based on their mutual information individual values are shown in the table 2.

MI values are calculated using an entropy formula that measures the proximity value between two variables. From table 2, we can see that there are five features with the highest ranking, namely Age, Comments, Total Expenditures, Average Consumption, and Addresses. Finally, we use these five features to be classified using SVM and Naive Bayes, because they have a value of MI > 0.01. In the second scenario, we conducted experiments to evaluate the performance and usefulness of different classification algorithms. Implementation results using all the selected features and features that are classified using two methods can be seen in table 3.

Table 3. Implementation results

Classification Methods	All Feature	Selected Feature using DMI
SVM	76.42%	73.57%
Naive Bayes	72.54%	70.46%

Based on table 3 SVM method has a higher accuracy value when compared to the Naive Bayes method. The accuracy value of the SVM method is 76.42% when using all features, while the Naive Bayes method is 2.54%. When applying the DMI feature selection method, the SVM accuracy value is 3.57% and Naive Bayes is only 70.46%. Figure 3 shows a comparison graph of accuracy, precision, recall and f-measure levels before and after implementing the DMI feature selection technique.

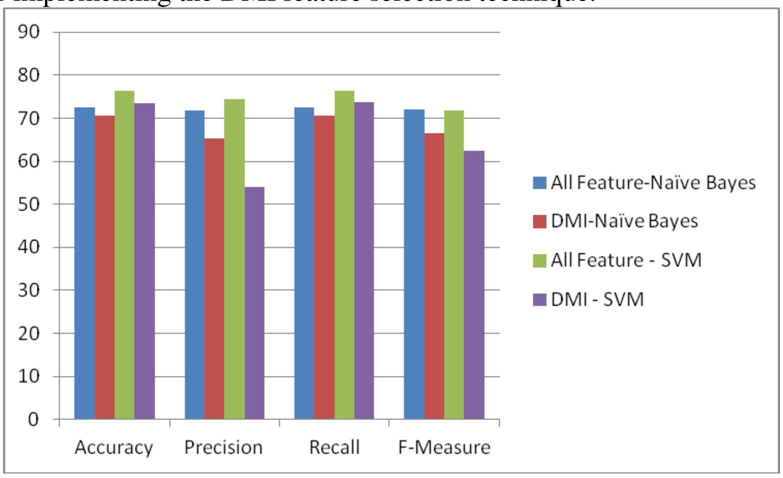

Figure 3. Experimental results value

Figure 3 shows that SVM method has higher values of accuracy, precision, recall and f-measure when compared to Naive Bayes. However, the implementation results in this paper show a decrease in the accuracy of the classifiers. Data reduction with feature selection usually improves the performance of predictive models because features that are not relevant to the classification target have been reduced.

4. Conclusion

We have experiment by comparing the proposed with SVM and Naive Bayes. It is show that SVM method has higher values of accuracy, precision, recall and f-measure. By applying DMI as a feature selection method, obtained there are five selected features, namely Age, Comment, Total Expenses, Average Consumption and Address. The accuracy value of the SVM method is 76.42% when using all features, while the Naive Bayes method only has a value of 72.54%. When applying the DMI feature selection method, the SVM accuracy value was 73.57% and Naive Bayes was only 70.46%. However, the implementation results in this paper show a decrease in the accuracy of the classifiers.

References

[1] Angeles-martinez L, Theodoropoulos C and Lopez-quiroga E 2018 The Honeycomb Model: A Platform for Systematic Analysis of Different Manufacturing Scenarios for Fast-Moving Consumer Goods *Journal of Cleaner Production*.
[2] Pan Y, Sheng S and Xie F T 2012 Antecedents of Customer Loyalty: An Empirical Synthesis and Reexamination," *Journal of Retailing and Consumer Services*, Vol. 19, No. 1, pp. 150–158, 2012.
[3] Moedjiono S, Isak Y R and Kusdaryono A 2016 Customer Loyalty Prediction in Multimedia Service Provider Company with K-Means Segmentation and C4.5 Algorithm *International Conference on Informatics and Computing, ICIC* pp 210–215
[4] Wijaya A and Girsang A S 2016 The Use of Data Mining for Prediction of Customer Loyalty *CommIT (Communication & Information Technology) Journal* **10** 1 pp 41–47
[5] Guo-en X I Aand Wei-dong J I N 2008 Model of Customer Churn Prediction on Support Vector Machine *Systems Engineering - Theory & Practice* **28** 1 pp 71–77
[6] Goyal P and Chanda U 2016 A Bayesian Network Model on the Association Between CSR, Perceived Service Quality and Customer Loyalty in Indian Banking Industry," *Sustainable Production and Consumption* **10** pp 50–65
[7] Ansari A and Riasi A 2016 Modelling and evaluating Customer Loyalty using Neural Networks: Evidence from Startup Insurance Companies *Future Business Journal* **2** 1 pp 15–30
[8] Guyon I and Elisseeff A 2003 An Introduction to Variable and Feature Selection *Journal of Machine Learning Research (JMLR)* **3** 3 pp 1157–82
[9] Sulistiani H and Tjahyanto A 2017 Comparative Analysis of Feature Selection Method to Predict Customer Loyalty *IPTEK, Journal of Engginering* **3** 1 pp 1–5
[10] Antonelli M, Ducange P and Marcelloni F 2013 Feature Selection Based on Fuzzy Mutual Information," *Springer International Publishing Switzerland* pp 36–43
[11] Tsai Y S, Yang U C, Chung I F, and Der Huang C 2013 A Comparison of Mutual and Fuzzy-Mutual Information-based Feature Selection Strategies *IEEE International Conference on Fuzzy Systems* 1
[12] Liu H, Sun J, Liu L and Zhang H 2009 Feature Selection with Dynamic Mutual Information *Pattern Recognition* **42** pp 1330–39
[13] Bennasar M, Hicks Y and Setchi R 2015 Feature Selection using Joint Mutual Information Maximisation *Expert Systems with Applications* **42** 22 pp 8520–32
[14] Li Y F, Xie M, and Goh T N 2009 A Study of Mutual Information Based Feature Selection for Case Based Reasoning in Software Cost Estimation *Expert Systems with Applications* **36** 3 pp 5921–31
[15] Qian W and Shu W2015 Mutual information criterion for feature selection from incomplete data

Neurocomputing **168** pp 210–220

[16] Sulistiani H and Tjahyanto A 2016 Heterogeneous Feature Selection for Classification of Customer Loyalty Fast Moving Consumer Goods (Case Study: Instant Noodle) *Journal of Theoretical and Applied Information Technology* **94** 1 pp 77–83

[17] Liu H 2016 Conditional Dynamic Mutual Information-Based Feature Selction *Computing and Informatics* **31** pp 1193–1216

[18] Liping W 2015 Feature Selection Algorithm Based on Conditional Dynamic Mutual Information I*nternational Journal on Smart Sensing and Intelligent Systems* **8** 1 pp 316–337

[19] Chen H and Chen L 2017 Support Vector Machine Classification of Drunk Driving Behaviour *International Journal of Environmental Research and Public Health* **14** 1

[20] Huang C H, Yang K C and Kao H Y 2014 Analyzing Big Data with the Hybrid Interval Regression Methods *Scientific World Journal*

[21] Zheng K and Wang X 2018 Feature Selection Method with Joint Maximal Information Entropy Between Features and Class *Pattern Recognition* **77** pp 20–29

[22] Osmanbegović E, Suljić M and Agić H 2014 Determining Dominant Factor for Students Performance Prediction by Using Data Mining *Original scientific paper Izvorni znanstveni rad,* **17** pp 147–158