



Tweet-based Target Market Classification Using Ensemble Method

Muhammad Adi Khairul Anshary* & Bambang Riyanto Trilaksono

School of Electrical Engineering and Informatics,
Institut Teknologi Bandung, Jalan Ganesha No. 10, Bandung 40132, Indonesia

*E-mail: adikhairul@unsil.ac.id

Abstract. Target market classification is aimed at focusing marketing activities on the right targets. Classification of target markets can be done through data mining and by utilizing data from social media, e.g. Twitter. The end result of data mining are learning models that can classify new data. Ensemble methods can improve the accuracy of the models and therefore provide better results. In this study, classification of target markets was conducted on a dataset of 3000 tweets in order to extract features. Classification models were constructed to manipulate the training data using two ensemble methods (bagging and boosting). To investigate the effectiveness of the ensemble methods, this study used the CART (classification and regression tree) algorithm for comparison. Three categories of consumer goods (computers, mobile phones and cameras) and three categories of sentiments (positive, negative and neutral) were classified towards three target-market categories. Machine learning was performed using Weka 3.6.9. The results of the test data showed that the bagging method improved the accuracy of CART with 1.9% (to 85.20%). On the other hand, for sentiment classification, the ensemble methods were not successful in increasing the accuracy of CART. The results of this study may be taken into consideration by companies who approach their customers through social media, especially Twitter.

Keywords: *bagging; boosting; classification; ensemble methods; target market.*

1 Introduction

A company can't satisfy the needs of all consumers with uniform product offerings because their needs differ. Each consumer is influenced by personal desires and preferences. Therefore, a company must be able to understand its consumers' needs according to their wishes and tastes [1]. Also, the company should be able to classify its customers according to their needs [2]. By selecting the right target markets, a company can maintain a balance between human resources, financial capacity and material resources, and hence uses its resources appropriately [3].

Formerly, companies had to conduct interviews or surveys to be able to know and understand the needs of consumers. In the era of social media, a company

Received January 20th, 2015, 1st Revision March 11th, 2015, 2nd Revision March 8th, 2016, 3rd Revision March 29th, 2016, 4th Revision April 21st, 2016 Accepted for publication May 24th, 2016.

Copyright © 2016 Published by ITB Journal Publisher, ISSN: 2337-5787, DOI: 10.5614/itbj.ict.res.appl.2016.10.2.3

can also take advantage of social media to sort out consumers' needs based on their desires and favorite products. Companies can look for information about consumers in social media accounts and analyze consumer trends. The process of analyzing consumer trends can be done with data mining [4].

Social media contain data pertaining to individuals [5]. Such data may include age, gender, place of residence, frequently discussed topics, frequently visited places, and so forth. There are also data that describe relationships with other users, such as with whom the user is friends, to whom they comment, and so on. These data can reveal the type or character of the individual. As data about individuals are required for determining target markets, data obtained from social media can be used for target market analysis. Twitter as a social media application has advantages over other social media. For instance, it provides developers with easy access to user data needed for processing into a more useful form through classification. So far, classification was conducted manually according to predefined criteria.

Data mining produces models that can perform consumer trend analysis. A simple data mining process will produce models quickly, but its accuracy will not be quite sufficient. A complex process will produce models that take a long time to execute but will provide results with higher accuracy. However, what we want is a method that can produce models quickly and still give results with high accuracy. Ensemble methods can be used to improve the accuracy of the results of classification algorithms. One way to build an ensemble method is by manipulating training data.

In this study, an experiment on classification using ensemble methods was carried out on a company that is engaged in e-commerce and has been using social media, in order to obtain higher accuracy in target market classification. We used learning algorithms contained in Weka by Machine Learning Group (version 3.6.9), i.e. CART, AdaBoost (boosting) CART, bagging CART and adaBoostM1 CART. CART was used to study the effect of the ensemble methods on the learning algorithms.

The purpose of this study was to classify followers of the @ptxyz account as the target market and then extract the features/attributes needed for target market classification based on their tweets. Next, a model was built to classify the target market using ensemble methods, followed by testing of the classification results. Lastly, an evaluation was conducted to find out whether implementation of the ensemble methods improved the accuracy of the learning algorithms, especially in the case of tweet-based target market classification. The model of the target market described in this paper can be applied on many industries and

business activities, but in this case the model was used on a retail company in Indonesia.

This paper is divided into five sections. Section 1 explains the background of the research. Section 2 contains a literature review. Section 3 describes the analysis and design of the study. The results of the study are discussed in Section 4. The last section consists of conclusions and suggestions for further research.

2 Related Work

Several researches have been conducted on data mining for marketing strategy and target market classification. Guoxiang and Zhiheng [6] describe what can be done with data mining in the field of marketing strategy. The most important objective of applying data mining to consumer data is to expand the market, i.e. develop a marketing strategy and attract more customers. For example by answering questions such as: What are the characteristics of customers who prefer a certain commodity? What kinds of customers have the same characteristics? What do high-value customers have in common?

Liu, *et al.* [7] conducted research to find the most accurate and efficient data mining algorithms to determine a target market. In their study 14 algorithms were selected to compare the cost of processing, accuracy, error rate Type I and error rate Type II. The dataset used in the experiment was collected from a Portuguese direct-marketing campaign related to bank-deposit subscription. The outcome was that Multinomial Naïve Bayes is still the ideal algorithm in the domain of bank-deposit term prediction.

Khan and Khan [8] discuss classification mechanisms for data streams. Data streams contain a high volume of multi-dimensional, unlabeled data generated in environments such as stock markets, astronomical applications, weblogs, clickstreams, flood, fire and police corps monitoring. The proposed method took a novel approach towards classification of data streams through applying unsupervised classification techniques such as clustering followed by a supervised classifier such as Support Vector Machine.

Shahrokhi, *et al.* [4] investigated classification of aircraft passenger candidates to assist them in selecting an airline suitable for them through the use of survey data and classification algorithm C5.0. Another research [9] investigated the classification of consumers in a health club. This research developed a method of information retrieval to identify which goods have potential market value. The data used were from consumers who were already members of the club or who were potential new members of the club.

Ahn, *et al.* [10] have conducted research on consumer classification to facilitate cross-selling in a mobile telecommunication company by combining five learning algorithms, i.e. SVM, K-NN, LR, ANN, and C4.5. A customer classification model was constructed using a collection of consumer data, including demographic data and patterns of use of products and services. The data were used to define new products and services with high sales potential. This research used more than one data-mining technique. First, the data were processed using logistic regression, an artificial neural network and a decision tree. Then, the models produced by each algorithm were combined using a genetic algorithm. This phase produced the final result to determine what products would be appropriate to offer to consumers.

Y. Wan [11] has done research on sentiment classification of Twitter data for airline service analysis. In this study, an ensemble sentiment classification strategy was applied, based on the majority-vote principle of multi-classification methods, including Naïve Bayes, SVM, Bayesian Network, C4.5 Decision Tree, and Random Forest algorithms. The result showed that the proposed ensemble approach outperformed the individual classifiers.

Martínez-Cámara, *et al.* [12] performed a combination of different types of sentiment analysis in order to improve their individual performance. These consisted of: (i) a ranking algorithm for scoring sentiments featured as bi-grams and skip-grams extracted from annotated corpora; (ii) a polarity classifier based on a deep learning algorithm; and (iii) a semi-supervised system founded on combining sentiment resources.

In the present study, classification of followers was performed according to product category (computer goods, mobile phones and cameras). The three product categories were based on the types of goods with the highest sales for the investigated company [13]. Sentiment classification was also conducted in this study, using three categories. The objective was for the company to give special treatment to individual consumers based on their sentiments.

The issues investigated were how to collect data, determine the right features/attributes of Twitter users in terms of issues related to tweet-based target marketing and the building of models through training and testing of classification models through the use of ensemble methods.

We used data extracted from users' tweets derived from the followers of the Twitter account of a company engaged in e-commerce. The tweets were classified into three product categories: computers, mobile phones and cameras. In addition, an analysis of positive, negative, and neutral sentiments was also performed. The ensemble methods used in this study were a way to manipulate

the bagging CART and adaBoostingM1 CART training data. To show the effectiveness of bagging and adaBoostingM1, their accuracy was compared with that of the CART algorithm. One thing that made this study different from other studies is that it only processed tweets that were expressed in the Indonesian language. It is well known that Indonesian has different sentence patterns compared to other languages, with different difficulty levels.

3 Target Market Classification

3.1 Target Market

A target market is a group of people who are the target of the marketing of a certain product because they are considered to have relatively similar needs. The target market serves as the basis for preparing a marketing strategy in terms of product strategies, pricing, distribution and promotion (marketing mix) [2]. Formulation of the target market for market segmentation can be done on the basis of the target market area characteristics such as density of the area (geographic criteria), population characteristics such as age, gender, education, income and so forth (demographic criteria), psychological characteristics such as lifestyle and personality (psychographic criteria), and behavioral criteria of the product's users such as benefits consumers seek in choosing/consuming the product (benefit segmentation).

3.2 Target Market Classification Process

3.2.1 Preprocessing

3.2.1.1 Manual Labeling

In this study, category labeling was performed manually based on keywords related to the product categories. The keywords were in the form of brands, components that make up the product, activities related to the product, and other products that complement the product. Each tweet was labeled according to its category (computers, mobile phones or cameras). The selection of these labels was based on the types of goods sold by the company.

The labeling of the sentiments was done automatically, based on existing collections of sentiment attributes as a result of previous research [14]. This automatic labeling was based on values obtained from tweets. When the tweets yielded a final value of 0, they would be labeled neutral, when the value was positive the label would be positive, and when the value was negative they would be labeled negative.

3.2.1.2 Tokenization and Stemming

In conducting tokenization, sentences in a tweet were broken down according to the gaps between words (spaces). Each word would be ensured to contain only the characters [A-Za-z]. Therefore, it was necessary to eliminate words that contained characters other than those allowed. URLs, mention (@) and hashtag (#) signs in tweets would be removed. After this, stemming was performed. Stemming is the process of determining root words referring to the *Kamus Besar Bahasa Indonesia* (Great Indonesian Dictionary). The stemming analysis follows the format of Indonesian words with affixes, prefixes, suffixes, infixes and confixes. Words with maximum affixations in Indonesian have the following pattern:

DP + DP + root word + DS + PP + P

DP: derivation prefix

DS: derivation suffix

PP: possessive pronoun

P: particle

The process of stemming was done using an algorithm developed by Arifin and Setiono, which is simpler compared to the algorithm of Nazief and Adriani. Both algorithms use dictionaries as a reference and provide a recoding function [15].

The first step is to get rid of the prefixes and then remove the suffixes. Next, those words whose prefixes and suffixes have been omitted are matched to words in the dictionary. When a word is not found in the dictionary, the affixes removed earlier will be placed back into the words already undergoing stemming in order to reduce the occurrence of overstemming. Then, the words are referred back to the dictionary. This process is repeated until a word has been matched in the dictionary.

3.2.1.3 Word Normalization

Words used in social media do not always follow the rules of standard language because social media, especially Twitter, limit the number of characters that are displayed in the system. Therefore, many users of Twitter or other social media shorten often-used words. Thus, a process is required to normalize abbreviated words. For example 'dgn' = 'dengan' (with).

3.2.1.4 Cleaning Stop Words

Stop words are words commonly found in large numbers and are considered meaningless. For example, stop words in the Indonesian language are ‘yang’, ‘dan’, ‘di’, ‘ke’ and so forth. The cleaning process of stop words is done so only meaningful words are included in the next process, i.e. feature selection. The result of the preprocessing stage are shown in Table 1.

Table 1 Data before and after data preprocessing.

Data before <i>preprocessing</i>	Data after <i>preprocessing</i>
Kecepatan Internet Indonesia	cepat internet Indonesia rendah
Terendah di Asia Pasifik	asia pasifik
http://t.co/G7XuHRTJlg	
AMD Hadirkan Prosesor ARM	amd hadir prosesor arm
Pertamanya: Opteron A1100 8	pertamanya opteron a core
Core http://t.co/lfjk0yYq2P	

3.2.2 Feature Extraction

Word ranking and feature extraction were conducted after tokenization and stemming. The data used were the results of the tokenization and stemming processes.

Feature extraction of products is different from feature extraction of sentiments. In the first case, the feature extraction process is done automatically using three methods of extraction, i.e. document frequency, information gain, and chi-squared. In the second case, the extraction process is done manually by utilizing a list of words that contain positive and negative sentiments, available from the findings of previous research.

In tests on feature extraction methods it was found that the document frequency (DF) method has the highest accuracy compared to other extraction methods. Therefore, in this study document frequency was used as the method for feature extraction. In DF, a word ranking is calculated according to the frequency of occurrence of the words in a document. The more often a word appears in the document, the higher its ranking. DF is a general measurement of the level of a word (the log of the calculation result of the number of documents divided by the number of documents containing the specific word).

To select which words will be used as a feature, a threshold is required. The threshold in DF is established by looking at the frequency point where the co-occurrence of a word and the number of a feature become more constant.

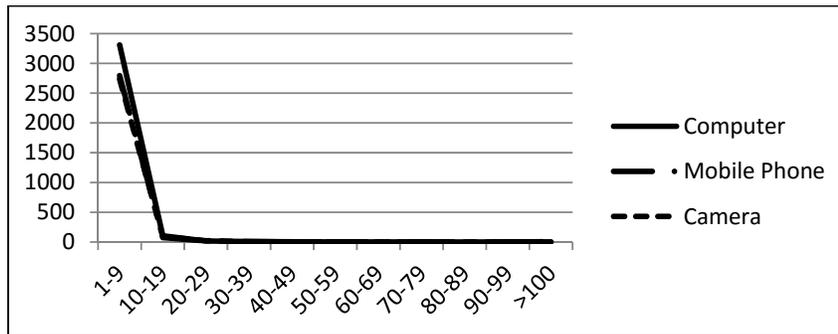


Figure 1 Feature distribution chart with DF method.

Based on Figure 1, it can be seen that the continuous line begins to look relatively constant at the value interval between 20 and 39. In this case we decided to use 30 as threshold value.

3.2.3 Developing Vector Data

After the features to be used in the classification process were obtained, the next step was to present the documents (tweets) in a vector-space model. All words contained in the tweets were registered and described in $\{f_1, f_2, f_3, \dots, f_m\}$. This study used 7 kinds of values in building the vector data: binary, df, idf, tf, log tf, tfidf and tfidf log. After that, which kinds of value determination gave the highest accuracy was determined.

3.2.4 Developing a Learning Model

After the vector data were successfully developed, the next step was to perform trainings on the vector data to produce a learning model. The formation of the learning model was done using Weka 3.6.9 involving three learning algorithms, i.e. bagging CART, AdaBoost CART, and CART. The CART algorithm represented the conventional method and therefore performance comparison with the two other methods (with ensemble method and without ensemble method) could be done.

The datasets used in developing the learning model consisted of three groups of features and 7 kinds of data values. The 3 groups of features were: document frequency, information gain and chi square, while the 7 kinds of data values were: binary, df, IDF, tf, log tf, tfidf and tfidf log.

Data processing by passing the vector data files for category training and sentiment training through Weka was carried out twice. The first stage was done

to produce a product category classification model (category model), while the second was done to produce a sentiment classification model (sentiment model).

After the models were obtained from the training yields, they were used for further processing, i.e. the testing process. In the testing stage, the product category model obtained was evaluated using the vector data files of the product category, while the sentiment model was evaluated using the vector data files of the sentiment category. The results of this testing stage were a product category model and a sentiment model whose levels of accuracy were better than those of the models at the training stage.

3.3 Ensemble Methods

Ensemble methods use more than one classifier (model) to solve the classification problem [16]. Unlike the usual learning approach, which uses a single classifier on training data, this method builds a collection of classifiers and then combines them. The ensemble approach is also called ‘committee-based learning’ or ‘multiple classifier system’.

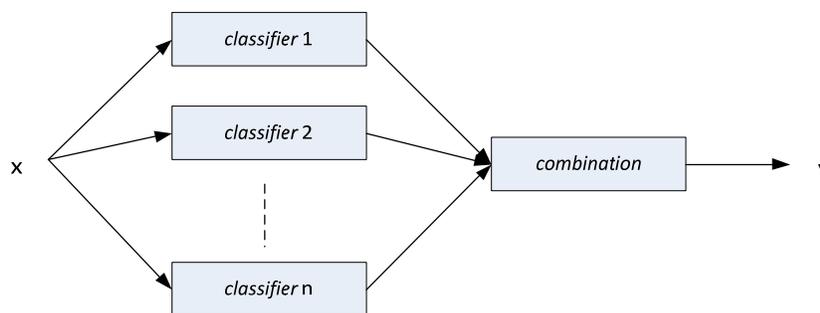


Figure 2 Ensemble method architecture.

Figure 2 depicts the architecture of ensemble methods. The ensemble consists of multiple classifiers, called ‘base learners’. Base learners are usually generated from the training data that are fed into a base of learning algorithms, such as decision trees, neural networks, and so forth. Ensemble methods use only one base-learning algorithm to build a homogenous classifier, i.e. by using only algorithms of the same type.

Ensemble methods can provide better results of accuracy compared to stand-alone classifiers because they can solve the problem posed by a limited number of data and avoid the problem of learning algorithms trapped in a local minimum error function [17]. With a limited number of data, a learning algorithm can only generate hypotheses (classifiers) that have almost the same

values. By using an ensemble method, the algorithm will ‘set an average’ for the hypotheses generated, which reduces the risk of choosing the wrong classifier. The problem of a learning algorithm trapped in a local minimum error function can be solved by running the algorithm from different starting points. Thus, some local minimum error functions will be obtained. From these local minimum error functions generated there is the expectation to find the global minimum error function.

In general, ensemble methods consist of two steps. The first is to produce a classification of each classifier and the second is to combine them.

3.3.1 Bagging

Bagging (bootstrap aggregating) is one of the ensemble methods. It works by creating new samples from the original training data [16]. Pseudo code of the Bagging algorithm is shown in the Figure 3. The following is the sequence of steps that make bagging work:

1. Out of the available training data, prepare t bootstrap samples. The bootstrap samples are made by sampling with replacement, i.e. some records in the training data already taken as sample data can be retrieved again for the data sample; in other words, there can be several records with equal values in the sample data.
2. Subsequently, each bootstrap sample is trained using a classification algorithm to generate t classifiers.
3. Testing data are inserted into the classifiers so that each classifier will yield a class prediction.
4. To produce the final prediction, voting on classes, mostly generated by the classifier, is performed.

```

Input : Dataset  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ 
        Base Learning Algorithm  $\mathcal{L}$ ;
        Number of base learners  $T$ ;

Process :
    1. for  $t = 1, 2, \dots, T$ :
    2.    $h_t = L(D, D_{bs})$  //  $D_{bs}$  is the bootstrap distribution
    3. end

Output :  $H(x) = \underset{y \in Y}{\operatorname{argmax}} \sum_{t=1}^T I(h_t(x) = y)$  //  $I$  is a indicator function
        // which have 1 or 0 value.

```

Figure 3 Bagging algorithm [16].

3.3.2 Boosting

The most widely known boosting method is called AdaBoost, introduced by Freund and subsequently renamed AdaBoostM1 [18]. AdaBoostM1 produces a collection of classifiers using a base learner (the first classifier produced). Each classifier depends on the previous classifiers and focuses on errors they have generated. Training data incorrectly classified into one of the previous classifiers will be given greater values.

The initial training data D are in the form $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where x_i is the input vector and y_i is the class label, and $y_i \in Y$ (set of class label). By setting values to the training data, we get $\{(x_1, y_1, w_1), (x_2, y_2, w_2), \dots, (x_n, y_n, w_n)\}$ and $\sum_i w_i = 1$

Pseudo code of the AdaBoostM1 algorithm is shown in Figure 4:

```

1. Initialize  $D(w_i) \leftarrow \frac{1}{n}$  for all  $i$ ; //initialize the weight
2. for  $t = 1$  to  $k$  do
3.    $f_t \leftarrow \text{BaseLearner}(D_t)$ ; //build a new classifier  $f_t$ 
4.    $e_t \leftarrow \sum_{i: f_t(D_t(x_i)) \neq y_i} D_t(w_i)$ ; //compute the error of  $f_t$ 
5.   if  $e_t > \frac{1}{2}$  then //if error is too large,
6.      $k \leftarrow k - 1$ ; //remove the iteration and
7.     exit-loop //exit
8.   else
9.      $\beta_t \leftarrow \frac{e_t}{(1-e_t)}$ ;
10.     $D_{t+1}(w_i) \leftarrow D_t(w_i) \times \begin{cases} \beta_t, & \text{if } f_t(D_t(x_i)) = y_i \\ 1 & \text{otherwise} \end{cases}$  //update weight
11.     $D_{t+1}(w_i) \leftarrow \frac{D_{t+1}(w_i)}{\sum_{i=1}^n D_{t+1}(w_i)}$  //Normalize the weights
12.  endif
13. endfor
14.  $f_{\text{final}}(x) \leftarrow \text{argmax}_{\sum_{t: f_t(x)=y} \log \frac{1}{\beta_t}}$  //the final output classifier
    
```

Figure 4 Boosting algorithm [18].

AdaBoostM1 generates as many as k classifiers, where k is determined by the users taken from the first yield (the base learner). The initialization of the initial value for each data is $\frac{1}{n}$. Each iteration yields a D_t derived from the same data but with a different value. Each iteration generates a new classifier.

3.4 Classification and Regression Tree (CART)

CART is a method of recursive partitioning used for regression and classification. CART is constructed by splitting subsets of a dataset using predictor variables to create two child nodes repeatedly, starting from the entire dataset. The goal is to produce a subset of the data that is as homogeneous as possible in order to classify the target variables.

At the beginning of the process, there has to be a training set consisting of the available classified records. The training set is used to construct a tree that allows the placement of a class into a new record target variable based on the values of other variables or independent variables.

CART builds binary trees by splitting the records at every node based on the function of a single input variable. The first task is to determine which independent variable serves as the best splitter. The best splitter is the one that can reduce node diversity. Nodes that undergo no further breakdown are called leaf nodes.

Splitting the records at each node causes the number of records to decrease from the root node to the child node, up to the leaf node. The smaller the number of records, the fewer representative nodes there are. The result is that a model tree can accurately predict only records that belong to the training set but cannot accurately predict new records from outside the training set, which is called overtraining. To reduce overtraining, tree pruning can be done. Pruning produces candidate sub-trees.

Some sub-tree candidates are selected based on their ability to predict new records. The selection requires a new data set, which is a test set that contains new records that are different from the records in the training set. Each sub-tree candidate is used to predict existing records in the test set. The sub-tree that provides the smallest error value is selected as the tree model.

The final step is to evaluate the sub-tree selected by applying it to a new data set, i.e. a validation set. The values of error obtained from the validation set can be used to predict the expected performance of the predictive models.

4 Evaluation

4.1 Dataset

The data used in this study were tweets derived from the followers of the account @ptxyz. The first thing to do was to gather user IDs from the followers of @ptxyz. 5000 user IDs were collected from a total of 200,000 followers. This was due to Twitter's policy of limiting the intake of follower IDs of an account. From all the IDs obtained, 200 tweets were taken from their most recently posted tweets. Not all of the data obtained were used in the dataset in this study. The data to be used in this study were manually labeled at the preprocessing stage (See Table 2).

Table 2 Datasets.

Dataset type	Number of datasets
Training data (80% of total dataset)	2400
Test data (20% of total dataset)	600
The total dataset	3000

4.2 Category

One classification selected for this study was based on product category (computer goods, mobile phones and cameras (see Table 3)). These three categories were selected based on the type of goods with the highest sales in the company. Category labeling was done manually with a balanced composition for each category of a total of 1000 labels. Next to product category classification, automatic classification of sentiments into three categories (positive, negative, neutral) was performed.

Table 3 Product and sentiment categories.

Products	Computer goods
	Mobile phones
	Cameras
Sentiments	Positive
	Negative
	Neutral

4.3 Experiment and Results

This study emphasized the application of ensemble methods for target market classification based on tweets that were divided into 3 product categories (computer goods, mobile phones and cameras) and 3 sentiment categories (positive, negative and neutral). There were 3000 testing data with a balanced composition (1000 computer goods, 1000 mobile phone, and 1000 camera).

The development of a classification model in this study used the bagging and adaBoostingM1, while the CART was used for comparison. Accuracy was measured by comparing the experimental data predicted to be correct relative to the test data. The accuracy of the test results for each learning algorithm can be seen in the Figure 5. Based on Figure 5, it can be seen that the highest accuracy in product classification was generated by the bagging algorithm (85.20%). This suggests that the ensemble method improved the accuracy level of the learning algorithm. For instance, the accuracy of CART prior to the use of the ensemble method was around 83.30%, so with the use of the ensemble method its accuracy increased with 1.9%.

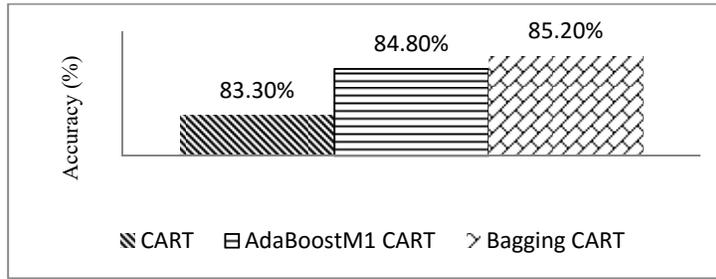


Figure 5 Product category classification accuracy chart.

The highest value of precision and recall in product classification was generated by bagging CART, i.e. 86.9% for precision and 85.2% for recall against 0.86 and against 0.85 respectively for CART (Figure 6). The highest F-measure was generated by AdaBoostM1 CART, i.e. 93.8% against 0.93 for CART. These results show that the ensemble methods improved the precision, recall and F-measure of the learning algorithm.

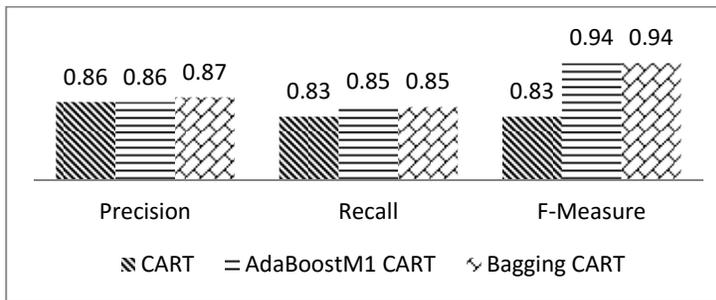


Figure 6 Precision, recall, and F-measure for product classification.

For the accuracy of sentiment classification, the following can be seen:

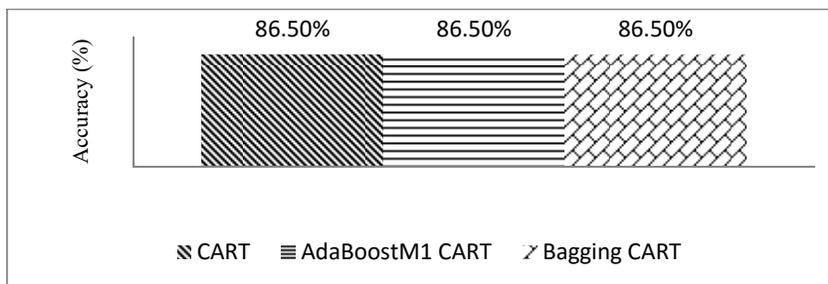


Figure 7 Sentiment classification chart.

Based on the Figure 7 it can be seen that the accuracy in sentiment classification of the three algorithms had an equal value of 86.5%. In sentiment classification, the ensemble methods were not successful in improving the accuracy of the learning algorithm. The sentiment feature still needs to be combined with automatic feature extraction, as a number of manually extracted features cannot represent the features optimally.

Similar to the value of sentiment classification accuracy, the values for precision, recall and F-measure generated by the three algorithms were the same: 0.75 or 75.7% for precision, 0.86 or 86.5 % for recall, and 0.80 or 80.5% for F-measure (Figure 8).

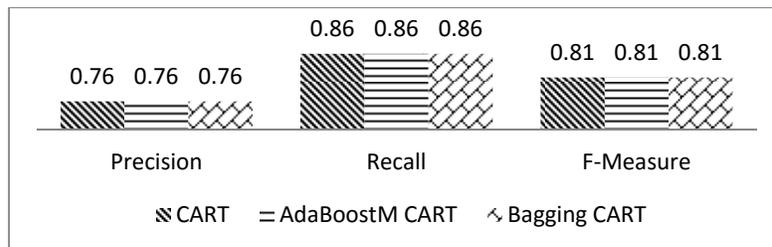


Figure 8 Precision, recall and F-measure in sentiment classification.

5 Conclusion and Recommendations

In this study, classification of target markets based on tweets was developed using ensemble methods with 2400 training data to construct models and 600 testing data to test the constructed models. Features of the extraction results using the document frequency method gave the highest accuracy. From the testing data, the obtained result was that the bagging ensemble method increased the accuracy of the CART algorithm in product classification by 1.9% (to 85.20%). Bagging successfully increased the value of precision to 86.9%, of recall to 85.2% and of F-measure to 93.6%. In sentiment classification, the ensemble methods were not successful in improving the accuracy, precision, recall or F-measure of the CART algorithm. This is because the sentiment feature still needs to be combined with automatic feature extraction because a number of manually extracted features do not adequately represent the features. The results of this study may be taken into consideration by companies who approach their customers through social media, especially Twitter.

A suggestion for further research is to expand this study by adding other ensemble methods in addition to manipulating the training data, such as using a method of input feature or output target manipulation, combining manual and

automatic sentiment feature extraction, and increasing the collection of the training data so as to enhance its accuracy.

References

- [1] Wells, W., Burnett, J. & Moriarty, S., *Advertising: Principles and Practice*. Fifth Edition, Prentice-Hall, New Jersey, United States 2000.
- [2] Aaker, D.A., Myers, J.G. & Batra, R., *Advertising Management*, Prentice-Hall, New Jersey, United States, 1996.
- [3] Xiang li, S.Y. & Niu, Y-P *Targeting Market Selection Based-On Fuzzy Preference Relation*, Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004, no. August, pp. 2390-2392, 2004.
- [4] Shahrokhi, N., Dehzad, R. & Sahami, S., *Targeting Customers with Data Mining Techniques: Classification*, International Conference on User Science and Engineering (i-USER), pp. 212-215, 2011.
- [5] Canali, C., Casolari, S. & Lancellotti, R., *A Quantitative Methodology to Identify Relevant Users in Social Networks*, IEEE Online Journal., <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5730307&tag=1> (22 April 2014).
- [6] Guoxiang, L. & Zhiheng, Q., *Data Mining Applications in Marketing Strategy*, Third International Conference on Intelligent System Design and Engineering Applications, Hong Kong, China, pp. 518-520, Jan. 2013.
- [7] Liu, L., Yang, Z. & Benslimane, Y., *Conducting Efficient and Cost-Effective Targeted Marketing Using Data Mining Techniques*, 2013 Fourth Global Congress on Intelligent Systems, pp. 102-106, Dec. 2013.
- [8] Khan, M.A. & Khan, A., *A Novel Learning Method to Classify Data Streams in the Internet of Things*, National Software Engineering Conference (NSEC), Islamabad, Pakistan, pp. 61-66, 2014.
- [9] Yao, Y.Y. & Zhong, N., *Mining Market Value Functions for Targeted Marketing*, 25th Annual International Computer Software and Application Conference, Chicago, Illinois, United States, pp. 517-522, 2001.
- [10] Ahn, H., Joon, J., Joo, K. & Ha, D., *Expert Systems with Applications Facilitating Cross-Selling in a Mobile Telecom Market to Develop Customer Classification Model Based on Hybrid Data Mining Techniques*, Expert Systems With Applications, **38**(5), pp. 5005-5012, 2011.
- [11] Wan, Y., *An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis*, Thesis, Dalhousie University Halifax, Nova Scotia, Canada, March, 2015.
- [12] Martínez-Cámara, E., Gutiérrez-Vázquez, Y. & Fernández, J., *Ensemble Classifier for Twitter Sentiment Analysis*, <http://wordpress.let.vupr.nl/>

- nlpapplications/files/2015/06/WNACP-2015_submission_6.pdf (4 December 2015).
- [13] J. Manalu, “*Bhinneka.com targeting Turnover Grows 40%*,” *Bisnis Indonesia*, 2014, <http://industri.bisnis.com/read/20140213/12/202921/bhinneka.com-targetkan-omzet-tumbuh-40>. (Text in Indonesian, accessed on 13 February 2014).
 - [14] Naradhipa, A.R. *Sentiment Classification for Indonesian Messages on Social Media use Noisy Text for Preprocessing*, Thesis, School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia, 2012.
 - [15] Asian, J., Williams, H.E. & Tahaghoghi, S.M.M., *Stemming Indonesian*, 28th Australasian Computer Science Conference (ACSC2005), The University of Newcastle, Australia, 2005.
 - [16] Zhou, Z-H., *Ensemble Methods Foundations and Algorithms*, CRC Press, Boca Raton, Florida, United States, 2012.
 - [17] Dietterich, T.G., *Ensemble Methods in Machine Learning*, In First International Workshop on Multiple Classifier Systems, J. KITTLER AND F. ROLI, Eds. Lecture Notes in Computer Science, New York, United States, Springer Verlag, pp. 1-15, 2000.
 - [18] Liu, B., *Web Data Mining, Exploring Hyperlinks, Contents and Usage Data*, Second Edition, Springer, New York, United States, 2011.