ECOMMERCE: AN EFFICIENT DIGITAL MARKETING DATA MINING FRAMEWORK TO PREDICT CUSTOMER PERFORMANCE

Mansoureh Zare, Sabzevar Branch, Islamic Azad University, Sabzevar, Iran
Hassan Shakeri, Mashhad Branch, Islamic Azad University, Sabzevar, Iran
Roya Mahmoudi, Sabzevar Branch, Islamic Azad University, Sabzevar, Iran

ABSTRACT

The connection of people's lives with digital space has created a new and attractive concept in marketing called "digital marketing". Data is one of the most important parts of any organization. The framework of data extraction establishes close relationships with the customer and manages the relationships between organizations. Data mining has become very popular in various digital marketing applications in recent years, and the classification model is an important data mining technique, this model is used to predict customer behavior to enhance digital marketing decision-making processes to retain valuable customers. This article provides an effective framework for exploring and classifying the Na Bayve Bayes classification model.

Keywords: Data mining, Digital marketing, Classification

INTRODUCTION

Data mining means analyzing visible data sets to find reliable relationships between data. Data mining is defined as a method that uses mathematical, statistical, artificial intelligence, and machining techniques to extract and identify useful information and subsequently acquire knowledge from databases. Information technology tools, advanced Internet technologies, and explosions in customer data have improved marketing opportunities and changed the way relationships are organized between organizations and their customers Alnoukari (2015). Data mining helps you to know your business behavior in the past and predict the future with a high approximation. Data mining makes the business environment transparent and forces you to make realistic decisions so DM is characterized as a set of methods created on the Web to influence clients to purchase an item or benefit Saura (2020) & Avery (2012).

Digital marketing, on the other hand, is a marketing activity that starts with market research, compels and examines markets that use horizontal media. Digital marketing is currently a very popular strategy and is used by almost all marketers around the world. The rise of the tech world has made the Internet a platform for a very forward-looking market. Digital marketing includes many of the techniques and techniques available in the Internet marketing category. Digital marketing also includes psychological, anthropological, anthropological, and technological factors that become new media with great capacity, interactivity, and multimedia Samran et al. (2019).

In digital marketing, it is very important to have a long-term, gradual, step-by-step look at the site. It had to become a buyer quickly and in one step. By "attracting his attention", you turn him into a clue, and by "getting closer" to the clue, you turn him into your customer. If this "*new customer*" is satisfied with your product and service, he becomes the customer will be your "*loyal and loyal customer*" and will recommend you to others in the guise of a free fan.

The larger the volume of data and the more complex the relationships between them, the more difficult it becomes to access the information hidden between the data, and the clearer the role of data mining as a method of knowledge discovery. Basic data mining concepts usually refer to the discovery of useful patterns among the data. A useful template is a data model that describes the relationship between a subset of data and is valid, simple, understandable, and new. Data mining involves the use of sophisticated data analysis tools to discover patterns and previously unknown relationships, interesting in large data sets. These tools can include statistical models, mathematical algorithms, and machine learning methods Krishnaiah et al. (2014); Jadhav et al. (2016).

The main purpose of data mining is to extract useful information from big raw data and convert it. It is understandable for its effective and efficient use. In general, data mining tasks can be divided into two categories: descriptive and predictive classification techniques Patankar et al. (2014); Nikam (2015).

Data classification is the process of organizing data into categories/groups so that the data objects of a group are more similar and the data objects are very different from different groups. The classification algorithm assigns each sample to a specific class so that the classification error is minimal Jadhav (2016) & Nikam (2015).

Classification is the evaluation of the characteristics of a set of data and then assigning them to a set of predefined groups. This is the most common data mining feature. Data mining can be used by using historical data to produce a model or view of a group based on data characteristics. This defined model can then be used to classify new data sets. It can also be used to determine prospects by determining which view is appropriate Jadhav (2016).

To demonstrate the performance of classification models, digital marketing programs such as focusing on audience behavior, security, risk, communication, maintaining different businesses, audience satisfaction, site visitor analysis, Internet advertising, data management platforms and targeting We consider the audience in the field of online sales. Online sales websites use existing data to retain their best customers and identify future sales opportunities. In this study, we consider a Naïve Bayes (NB) classification model as a possible simple classification based on the application of the Bayes theorem.

This paper presents an efficient data mining framework and examines the effectiveness of a data mining classification model in predicting customer behavior in digital marketing. The data set is known as digital marketing data from two online sales sites. Classification criteria such as speed, sensitivity analysis, and characteristics can be used to evaluate classification performance. The tool used for this study is Weka.

The sections in front of the article are organized in this way, Section 2 includes the data used in the research, Section 3 identifies the problem, Section 4 describes the digital marketing data framework.

The work is explained, in Section 6 the logic of the execution and in Section 7 the analysis of the test results are presented, and finally in Section 8 the conclusion of the article is presented.

Data Collection

The data set used for testing in this paper contains the results of customer data from three online sales websites through digital marketing, which was examined between Aug 2019 and Feb 2020.

The data set includes 4521 samples with two possible items (customer satisfied or not). In total, the data includes 16 input variables. Eight variables are related to the customer and eight variables are the latest digital marketing activities:

1. Customer age.
2. Marital status and education.
3. What are the customer's preferences by default?
4. The type of communication the audience has.
5. Last customer call.
6. The result of past digital marketing campaigns.
7. The number of audiences who have made more purchases.
8. The number of days that the customer makes the most purchases.
9. The result of digital marketing on customer preferences so far.
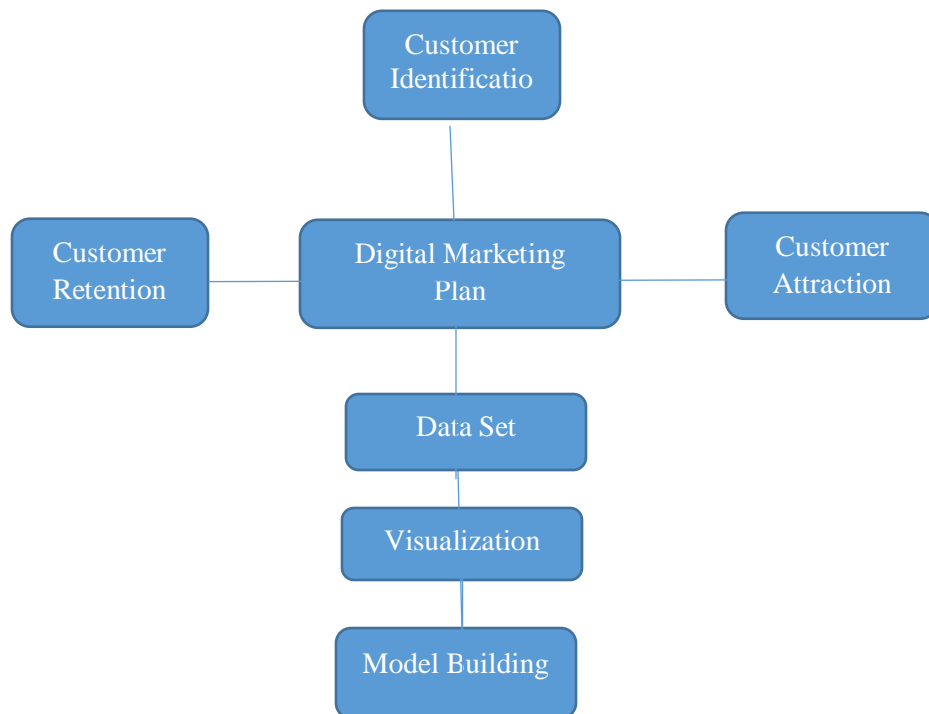10. Statement of the problem.

FIGURE1

DIGITAL MARKETING-DATA MINING FRAMEWORKFIFF

An efficient digital marketing data mining framework is proposed to predict customer behavior on online sales sites, and a classification model is studied and evaluated in the proposed framework in Figure 1.

The proposed framework for digital marketing data mining is shown in Figure 1, identifying business objectives for building a business plan includes the initial stage, and a careful study of digital marketing will help improve customer engagement and retention. It helps to prepare and process data by cleaning processes, selecting features, converting data, etc. to build more models and evaluate them. Building a model in the context of digital marketing is an important step towards achieving business goals.

CONCEPTS USED

Naive Bayes: Naive Bayes Classifier is the basic Statistical Bayesian Classifier Jadhav (2016) Duda et al. (2017).

It is called Naive as it expects that all factors contribute towards order and are commonly related. This supposition that is called class contingent autonomy Jadhav (2016), Friedman et al. (1997).

It is additionally called Idiot's Bayes, Simple Bayes, and Independence Bayes. They can anticipate class enrollment probabilities, for example, the likelihood that a given information thing has a place with a specific class name. A Naive Bayes classifier thinks about that the nearness (or nonappearance) of a specific feature(attribute) of a class is inconsequential to the nearness (or nonattendance) of whatever other component when the class variable is given Jadhav (2016).

This classification technique is based on the Bayes theorem and the assumption of independence in predictors. Simply put, the Bayes Naive classification assumes that the presence of a particular attribute in a class has nothing to do with any other attribute. The Naive Bayes model is easy to build and is especially useful for very large data sets. Along with simplicity, Naive Bayes performs better in very complex classification methods. The Bayes proposition provides a posterior probability method in the form of Equation (1).

$$P(c \mid x) = \frac{P(x \mid c) P(c)}{P(x)}$$

$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c) \tag{1}$$

Bayesian classifiers are useful in predicting the probability that a sample of a particular class will belong. This technique is used in large databases due to its high accuracy and sufficient speed for training with simple models. To estimate the parameters (mean and variance of variables) required for classification, the classifier requires only a small amount of training data.

It also uses real and discrete data. Naive Bayes requires short computational time for training, and also improves classification performance by eliminating irrelevant features.

WEKA

In recent years, due to the development of information and communication technology and with the increase of digital data and the growth of computing power of computers, many commercial and educational software for data mining have been offered in various fields. The use of data mining tools and algorithms can model the complex nature of data and the intangible relationships between these data. The data mining tools take the data and create an image of reality in the form of a model. This model describes the relationships between the data. WEKA is a collection of machine learning algorithms and data processing tools.

The software was introduced by Waikato University in New Zealand. It is also open-source software written in Java. It provides valuable support for entire experimental data mining processes. This software was first written in 1997 in its modern form and was released under the GPL license Bahari (2015).

WEKA software is a set of machine learning algorithms for performing data mining tasks. This environment includes methods for initial data processing, classification, regression, clustering, and association rules and proper forecasting. We have used this software to test in this research.

Implementation Logic

Data processing: Initially, the data set included 7925 contacts and 38 attributes. Contacts were deleted with missing data or unspecified results, resulting in 4511 samples and 17 attributes without missing values. Features of contact information and customers the online sales database report was used. The output feature is whether the customer is satisfied or not. Of the 38 attributes, several unrelated features negatively affect the data mining learning process.

One method of selecting the manual feature using Rattle analysis was used to reduce the properties to 29 input features and 1 output feature. This was further reduced in various CRISP-DM iterations, and finally 17 features were selected, including the output feature for test data.

Modeling

All tests were performed using the Weka tool in Windows 7. In this experiment, we use the NB classification model in data mining, For the model, the tenfold validation test method was used. Nodes are all sigmoid except when the class is numerical. We set the number of hidden layers using heuristic a = round (M / 2), where M is the number of attributes and classes. In the modeling step, we successfully tested the NB model using the weka tool.

Test Result

Table 1 summarizes the results of our experiment to automatically identify a given dataset, That shows values for classification. The classification accuracy (amount of correctly classified instances), true positive rate (the proportion of actual positive ones correctly defined as such), false positive rate (incorrectly classified positive), ROC area (area under the ROC curve) and the time taken to construct the classifier model are shown for each form.

Table 1 RESULTS OF CLASSIFIERS, AVERAGE OVER 10 RUNS					
Classifier	Classification Accuracy (%)	True Positive Rate(TPR)	False Positive Rate(FPR)	ROC Area	Time taken to build models (s)
NB	87.97	0.47	0.067	0.858	**0.08**

The NB classification model shows accuracy (87.97) in the tested model. NB shows values of TPR (0.47), FPR (0.067) and ROC region (0.858). The time required to build the model for NB (0.08) is very good. In our research three statistical tests are used to test the efficiency of the classification models, namely classification precision, sensitivity, and specificity. In terms of instances classified as true positive (TP), true negative (TN), false positive (FP) and false negative (FN), this information is given. This classification information describes a confusion matrix and is given in Table 2.

Table 2 CONFUSION MATRIX FOR CLASSIFIER			
Classifier	Confusion Matrix a	Confusion Matrix b	← **Classified as**
NB	246 269	275 3731	**a = yes** **b = no**

The accuracy of classification as shown in equation (2) is equal to the sum of TP and TN divided by the total number of cases N, and refers to the ratio of the number of cases correctly classified Bahari (2015).

$$\text{Accuracy} = \frac{TP+TN}{N} \qquad (2)$$

Sensitivity in equation (3) is proportional to the TP-to-TP-to-FN ratio, and corresponds to the correctly defined positive (True Positive Rate) Bahari (2015).

$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad (3)$$

Specificity in equation (4) is equal to TN divided by sum of TN and FP and refers to the correctly defined negative rate (True Negative Rate) Bahari (2015).

$$\text{Specificity} = \frac{TN}{TN+FP} \qquad (4)$$

The NB model shows the best samples for accuracy (87.97 and feature (93.28) for training samples, as well as the best sensitivity (47.2%). The classification function is compared in terms

of accuracy, sensitivity and characteristics, and the comparison results are shown as shown in Table 3.

TABLE 3 PERFORMANCE COMPARISON OF CLASSIFIER IN TERMS OF ACCURACY, SENSITIVITY AND SPECIFICITY			
Classifier	Accuracy(%)	Sensitivity(%)	**Specificity(%)**
NB	87.97	47.2	**93.28**

Of the 4511 samples in the data set, the ones that are classified correctly and the ones that are incorrectly classified for each model, as shown in Table 4.

TABLE 4 CLASSIFICATION OF 4511 INSTANCES IN THE DATASET.		
Classifier	**Correctly classified instances**	**Incorrectly classified instances**
NB	3970	**541**

CONCLUSIONS AND FUTURE RESEARCH

Data mining means finding the dominant pattern between elements within your large data set. Simply put, recruitment process data mining can be used using larger volumes of raw data. In data mining, the pattern between data is analyzed by several experimental software. The data mining algorithm is a series of exploratory and computational methods that aim to create a model of the desired data. To create a model, the data is first analyzed by the algorithm to find a pattern or approach. The algorithm then finds the most optimal parameters by applying the result of this analysis on the samples and creates a model. These parameters are then applied to the data set and an application pattern is obtained. On the other hand digital marketing is the introduction and promotion of sales of products and brands through one or more electronic media. One of the most important features of digital marketing is the ability to measure the effectiveness and efficiency of advertising campaigns and its cheapness. Digital marketing is also compatible with a variety of business scales from small to large. In this paper, we have proposed an efficient digital data mining framework for predicting customer performance. A classification model was used to predict. Classifier performance was also compared in terms of accuracy, sensitivity, and feature. And we found that by using digital marketing data mining, we can achieve the effectiveness of digital marketing in achieving the goals of the organization in attracting and retaining customers and thus profitability.

For future research, the results can be extended to other new models, and other classification models can be compared with this model, and the same research results can be used for other online sales sites. Research limitations are the type of classification used in this research, as results may differ from other categories.

REFERENCES

M. Alnoukari, (2015) "ASD-BI: An agile methodology for effective integration of data mining in business intelligence systems," in *Integration of Data Mining in Business Intelligence Systems*: IGI Global, 2015, pp. 61-82.

J.R. Saura, (2020) "Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics," *Journal of Innovation & Knowledge,* 2020.

J. Avery, T.J. Steenburgh, J. Deighton, & M. Caravella, (2012) "Adding bricks to clicks: Predicting the patterns of cross-channel elasticities over time," *Journal of Marketing,* vol. 76, no. 3, pp. 96-111.

Z. Samran, S. Wahyuni, M. Misril, R. Nabila, & A. Putri, (2019) "Determination of Digital Marketing Strategies As Effective Communication Techniques For GoOntravel Brand Awareness," *Journal of Research in Marketing,* vol. 9, no. 3, pp. 752-757.

V. Krishnaiah, G. Narsimha, & N.S. Chandra, (2014) "Survey of classification techniques in data mining," *International Journal of Computer Sciences and Engineering,* vol. 2, no. 9, pp. 65-74.

S.D., Jadhav & H. Channe, (2016) "Comparative study of K-NN, naive Bayes and decision tree classification techniques," *International Journal of Science and Research (IJSR),* vol. 5, no. 1, pp. 1842-1845.

B. Patankar & V. Chavda, (2014) "A comparative study of decision tree, naive Bayesian and k-nn classifiers in data mining," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 4, no. 12, pp. 776-779.

S.S. Nikam, (2015) "A comparative study of classification techniques in data mining algorithms," *Oriental journal of computer science & technology,* vol. 8, no. 1, pp. 13-19.

R.O. Duda, P.E. Hart, & D.G. Stork, (1973) *Pattern classification and scene analysis*. Wiley New York, 1973.

N. Friedman, D. Geiger, & M. Goldszmidt, (1997) "Bayesian network classifiers," *Machine learning,* vol. 29, no. 2-3, pp. 131-163.

T.F. Bahari & M.S. Elayidom, (2015) "An efficient CRM-data mining framework for the prediction of customer behaviour," *Procedia computer science,* vol. 46, pp. 725-731.