



17 November 2021
Iran Institute of Industrial Engineering

Prediction of marketing strategies performance based on clickstream data

Maryam Nezhad Afrasiabi¹, Akbar Esfahanipour^{1*}, Alimohammad Kimiagari¹

¹*Department of Industrial Engineering and Management Systems, Amirkabir University of Technology, Tehran, Iran*

mafrasiabi@aut.ac.ir, esfahaa@aut.ac.ir, kimiagar@aut.ac.ir

Abstract

Today, Internet-based businesses are one of the most useful tools to make gain in the economies of developing and developed countries. It can even said that the expansion of the World Wide Web caused other businesses to seek customers in the virtual advertising and online world to increase their sales. This study presents a data-driven approach to predict the success of the marketing strategies performance of an online shopping store. The data has been collected by a Poland online shopping website in the year 2008, which has extracted in the UCI datasets. In the data preparation phase, a decision tree (DT) is developed and 13 features of customers are selected for modeling phase. In the proposed method in this research, the rminer package of R software is used. In which three classification models including neural network(NN), support vector machine (SVM), and logistic regression(LR) are developed. Then, two criteria of AUC and ROC curves are used to compare these three models. By comparing the models, it is determined that the NN technique works better than the other three models in prediction. This result can be helpful for marketing managers to plan effectively in website design to attract new visitors and shoppers.

Keywords: Classification, sales forecasting, machine learning, clickstream Data, marketing plan, neural network

1-Introduction

In the past, due to limitation of variety of products and the number of customers, firms were not concerned about selling their products and marketing plans. But Over time and the prevalence of customer-oriented systems, the variety of products and the exchanges increase lead to a wide range of information that can be used as a suitable marketing plan and sale strategies. One of the most important source of producing this large amount of information is the web pages of online stores. Which the use of data mining techniques in these databases can be helpful to know deeply about the current customers and to attract new customers.

In this study, after reviewing the empirical and theoretical studies, a suitable model is developed to evaluate the characteristics of customers in the website of an online shopping clothing store, which is taken from the UCI dataset.

*Corresponding author

The data is actually extracted from the website of an online women's clothing store in Poland in the year 2008, which customers are from 42 different countries according to their IP addresses. In the next step, four data mining techniques are implemented on the current data. Finally, these models are compared using appropriate criteria to determine which model can work better in predicting marketing performance and customer behavior.

2-Literature review

Commercial companies to attract more customers, are constantly looking to increase their products variety and categories. In a day huge commercial organizations often have millions of website visitors. These potential customers visit the catalogs and categories of products. As a result, the main challenge of e-commerce companies is how to convert these potential customers into active customers and finally increase the product sales. A lot of research has been done in this area and in discovering potential customers. Some of studies have focused on customer interest rating (Cleger-Tamayo, Fernández-Luna, & Huete, 2012; Zhao, Niu, & Chen, 2013). However, the analysis of the customer interest rating determines which type of products by what characteristics has been desired by the customers but this information is limited and it's not useful in website navigation processes. In fact, the customer interest rating is related to various features of e-business such as price, service and logistics so the product and interest rating of new customers cannot be used as a valid reference. On the other hand, experienced customers are often reluctant to rate every time they visit the website.

Several researchers have analyzed user interests in websites and social media (Li, 2009; Zeng, Zhang, & Wu, 2008). According to their studies the user comments and interests is frequently reflected in the webpage they visited and also the posts they shared. This idea can be implemented similarly in e-commerce companies and websites. Different users, according to their idea, interest and goal, reach the desired pages on the website through different paths. Then they will look for the desired products and services for a while on the desired page. In other words, such information, shows the users' interests more accurately.

By development of information technology and social media, it becomes possible for the owners of e-commerce companies to provide accurate information about all the activities of individuals with details and the lowest cost at the moment. All information about users' activities on the Internet is known as clickstream data (Bucklin & Sismeiro, 2009). Clickstream data actually provides a wide range of information to the marketing team of companies in a specific and accurate time, which consists of visiting a webpage, click on ads-banner, stopping time in a specific webpage and finally the products which bought by costumers. By analyzing all these data, a deep recognition of customers can be achieved and then an appropriate strategic decision can be made. Therefore, it can be said that in today's business world, clickstream data has become an important source of information for understanding customer's behavior. Extracting information from these dataset is so different from traditional and previous datasets (Babcock, Babu, Datar, Motwani, & Widom, 2002). In fact, the connection between the generated data and the large amount of data in the information flow, leads to the extraction of appropriate knowledge. Currently, there are various programs to generate instantaneous data flow, which can be referred to the flow of transactions in retail, recording all customer's clickstreams on web pages and also recording voice and conversations in telecommunications (Li, 2009). Li (2009) in his research, points to the use of web pages and the clickstreams. In his research, he selects a web page with a moving window, and examines the k-top pattern of users' clickstream. Finally he proposes an efficient pattern for flow of information in the moving window mode. Nasraoui et al. (2003) use a scalable clustering technique inspired by the immune system to identify clickstream and customer profiles.

In terms of customer recognition, a multi-level review of the dataset is very important. Because it actually helps to discover the correlation between the data and the hierarchical classification of the data with a certain amount of reliability. Chompaisal et al. (2014) propose a method to discover the most popular repetitive pattern for users, followed by an algorithm called compatible FP-growth. An important application of customer recognition is credit scoring in financial institutions. Xia et al. (2018) examined four classifiers namely SVM, RF, XGBoost, GPC and then proposed a novel

classifier. The comparison results demonstrate the superiority of the new ensemble credit scoring model.

One of the benefits of digital marketing is to track visitors on websites and proposing them special offers. Any interactions that the visitors have on a website is like a fingerprint to identify useful information such as the visitors' interest and requirements. Kohen et al.(2020) study an online shopping store and predict customer behavior through the clicks they make by using the recursive neural network method. Their data is related to an online shopping store from May 20 to July 20, 2018. Kawaf et al.(2019) study online fashion stores and examine the role of marketing plans through Facebook and online analysis of visitors. In their research, they used the individual construction theory proposed by (Kelly, 1955) and rely on how individuals and visitors are influenced by their past decisions. Most previous works have simply exploited and applied existing sequential pattern algorithms to the mining of clickstream patterns, and a few studies have weighed the clickstreams, which also have a wide range of applications (Huynh et al. 2020). Huynh et al. (2020) proposed an approach based on the average weight measure for clickstream pattern mining. Results show that their improved method enhance both the efficiency and memory consumption. Table 1 summarizes characteristics of the related studies in the literature.

Table 1. Literature review summary

Customer feature	Key studies	Data type and sample size	Method
Customer interest rating	Cleger-Tamayo, Fernández-Luna, & Huete, 2012	Digital newspapers	News recommendation, Classification
Customer interest rating	Zhao, Niu, & Chen, 2013	MovieLens dataset	Matching users' interests first, and High quality items that users will like
User comments and interests	Zeng, Zhang, & Wu, 2008	Interactive website	Hidden Markov model (HMM), Classification
User comments and interests	Li, 2009	Web usage mining	Sliding window-based Web data mining algorithm
User profiles	Nasraoui, Cardona, Rojas, & Gonz'alez , 2003	Web usage mining	Scalable clustering methodology
Shopping behavior on the Internet	Bucklin, & Sismeiro, 2009	Clickstream data	Prediction, strengths and limitations
N-most interesting patterns	Chompaisal, Amphawan, & Surarerks, 2014	Two datasets, T10I4D100K and T20I6D100K with 4 hierarchical levels	Compatible FP-growth, Association Rules
Customer reviews	Kawaf, & Istanbuluoglu, 2019	Facebook	Qualitative study
Online shopping behavior	Koehn, Lessmann, & Schaal, 2020	Digital marketing	Recurrent neural networks (RNNs)
User profiles	Xia, Liu, Da, Xie, 2018	UCI credit dataset	Credit scoring, Novel classifiers
Customer reviews	Huynha, Nguyenb, Voc, Nguyend, Tseng, 2020	Clickstream mining, BIBLE, FIFA, SIGN, Chainstore, and KDD	Weighted clickstream pattern mining

3-Methodology

3-1- Data

This study focuses on which customers by what behavioral patterns decide to buy a product. Therefore in this study, the UCI dataset of an online shopping clothing for women is used. The details of this dataset has been summarized in table 2:

Table 2. General characteristics of the online shopping dataset

Characteristics	Definition
Data Set Characteristics	Multivariate, Sequential
Attribute Characteristics	Integer, Real
Associated Tasks	Classification, Regression, Clustering
Number of Instances	32957
Number of Attributes	14
Missing Values?	N/A
Area	Business
Date Donated	2019-12-09

Initially, in the data preparation phase, 80% of the data is used as training data, which is used in feature selection and modeling. Then the rest of the data used as test data to predict the desired target variable.

Each row of this dataset includes the target variable whether the purchase was made or not, and the input variables includes: country, month, day, order, main category, clothing model, color, location of the image on the page, photography model, price and the page number of the website. In fact, the considering of these variables together shows behavior patterns of each customer. As one of the consequences of this research, it is examined whether by analyzing the behavior of customers in the website, it is possible to predict the marketing performance or not.

3-2-Data mining models

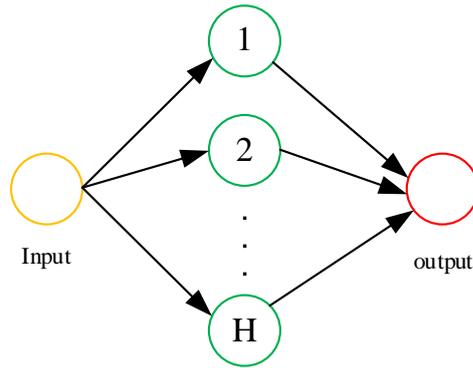
In this research, four data mining techniques as rminer package of R software is used (Cortez, 2010). These techniques are neural network (NN), support vector machine (SVM), logistic regression (LR) and decision tree (DT).

The LR is used to analyze the relationship between variables. It is practical especially in the fields of medicine, psychology, social sciences and economics. For example, examining and modeling of the relationship between daily activity and diabetes is an example of analysis in which the LR model is used. In fact, over this study, by using LR, the class probability of $p(c|X_k)$ can be estimated(Venables, 2003):

$$p(c|X_k) = \frac{1}{1 + \exp(w_b + \sum_{i=1}^M w_i X_{k,i})} \quad (1)$$

In equation 1 $p(c|X_k)$ is the class probability of c by k th input variable, with M features and w_i denotes a weight factor, adjusted by a learning algorithm.

One of the most widely used data mining algorithm is DT. The DT is actually a tree structure that represents a set of laws and causal relationships. The NN includes machine learning algorithms, which lead to the classification of inputs and the optimal output. The NN receives the data and analyzes it in their hidden layers to finally provide an output. In this study, one hidden layer with H hidden nodes and an output that is actually buying or not buying decision is considered as shown in figure 1.



One hidden layer with H nodes

Fig 1. The neural network under investigation with one hidden layer and H hidden nodes

The value of H in the hidden layer has been actually set the complexity of the NN. It means that the higher value of H, the greater complexity of nonlinear relationships. In the NNs where the value of H parameter is considered to zero, it will be the same as the LR model. For a given input x_k the state of i-th neuron s_i is computed as equation 2:

$$s_i = f(w_{i,0} + \sum_{j \in P_i} w_{i,j} \times s_j) \quad (2)$$

In equation 2, P_i represents the set of nodes reaching node i; f is the logistic function; $w_{i,j}$ denotes the weight of the connection between nodes j and i; and the output node estimates the probability of buying decision which is between zero and one. The ultimate solution that the neural network offers is actually influenced by the choice of starting point weight. Therefore, in this research, according to the suggestion of Hastie et al. (Hastie, 2008) the rminer package is used. In this package, a group of different neural networks are examined and finally, the average of the individual predictions presented as output.

But what if the data is completely inseparable with a linear boundary? This is where the SVM algorithm would be useful. The main idea of the SVM is mapping to a higher-dimensional space using the kernel function, so that better classification of data happens (Cortes & Vapnik, 1995). According to the rminer package, for the kernel function, the general Gaussian function, which uses less parameters, is used instead of other kernel functions, such as polynomial functions (Hastie, 2008).

3-3- Data preparation

The importance of data preparation is due to the fact that the lack of proper data equals bad inputs in modeling section and ultimately brings inadequate outputs. Before fitting the DM models the input data is first standardized to a zero mean and one standard deviation (Hastie, 2008). The data is then divided into two parts: training data and test data. In this section, 80% of the data is selected as training data. Then, by using the DT, the features have been selected. For the DT, rminer adopts the default parameters of the rpart R package, which implements the popular CART algorithm (Breiman, 1984). In this algorithm, by combining different features, the price is considered as the root node, which is shown in figure 2. The DT in figure 2 shows that if the average price of a particular product is higher or equal than the 46\$ for the entire product category, be chosen by the customer with the probability of 0.52. Otherwise the main category of product is effective. After that, there is a 78% probability of a sale (one of the main category modes) in the product, in which case it will not be selected by the customer. For more details, see figure 2.

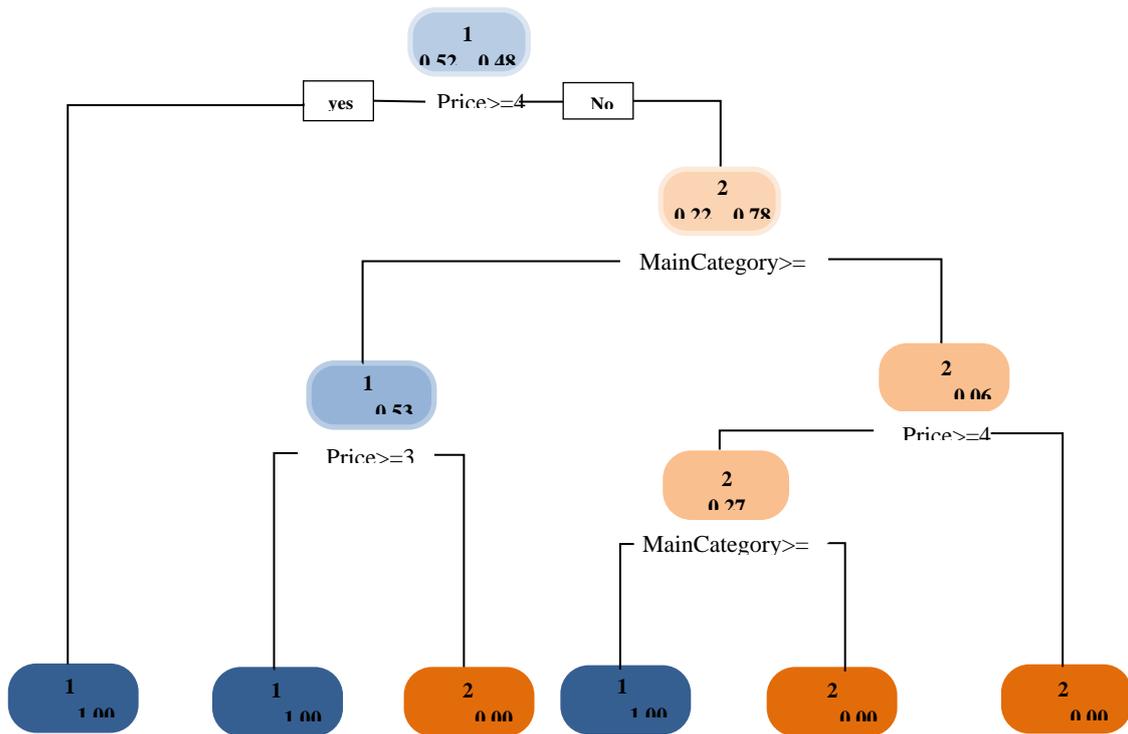


Fig 2. The DT which is used in feature selection phase

According to the output of the DT, the figure 2 and the confusion matrix calculated the DT performance is computed as table 3 which indicates that the DT is satisfied and the selected features provide a good estimate.

Table 3. The Measured the performance of DT

Metrics	Measure
Accuracy	86%
Precision	74%
Recall	99.7%

As the result of the DT, the features in table 4 are considered as the main features in the modeling.

Table 4. Selected features

Number	Feature	Definition	Type
1	MONTH	from April (4) to August (8)	Numeric
2	DAY	day number of the month	Numeric
3	ORDER	sequence of clicks during one session	Numeric
4	COUNTRY	the IP address	Numeric
5	SESSION ID	session id	Numeric
6	MAIN CATEGORY	the main product category(trousers, skirts, blouses, sale)	Numeric
7	CLOTHING MODEL	217 products	Character
8	COLOUR	colour of product (14 colours)	Numeric
9	LOCATION	photo location on the page	Numeric
10	MODEL PHOTOGRAPHY	two categories(en face, profile)	Numeric
11	PRICE	price in US dollars	Numeric
12	PRICE 2	whether the price of a particular product is higher than the average price	Numeric
13	PAGE	page number within the e-store website (from 1 to 5)	Numeric

3-4- Modeling

For all the techniques performed in this research, the rminer package of R software is used. Also, for data mining models implemented in R software, at least 20 outputs are taken to consider different states. The rminer package uses the efficient BFGS algorithm for the LR and the NN models (Møller, 1993), Which is the family of quasi-Newton methods, while the SVM method is trained using the sequential Minimal Optimization (SOM) (Platt, 1998).

According to the description in the previous section, the R software runs on the windows system. Table 5 compares the results of using three data mining techniques by using the AUC (Area Under Curve) metric. The AUC metric actually states that the higher the level below the graph, the greater accuracy of the applied model. According to this metric, a model with a higher AUC will be better for prediction. In the following step, to compare the performance of the three used models, the ROC curve is plotted according to figure 3. This curve actually shows how each model accurately classified the dataset, and how much the prediction rate of true positive rate(TPR) is higher than the false positive rate(FPR). According to this definition, any curve that is higher than other curves have the better performance. According to figure 3, the NN model has the best performance, and then the SVM and LR models have better performance in classification and predicting new customers, respectively.

Table 5. Comparison of the AUC criteria of three data mining techniques used in the research

Metrics	LR	SVM	NN
AUC	0.84	0.89	0.992*

(* denotes the best value)

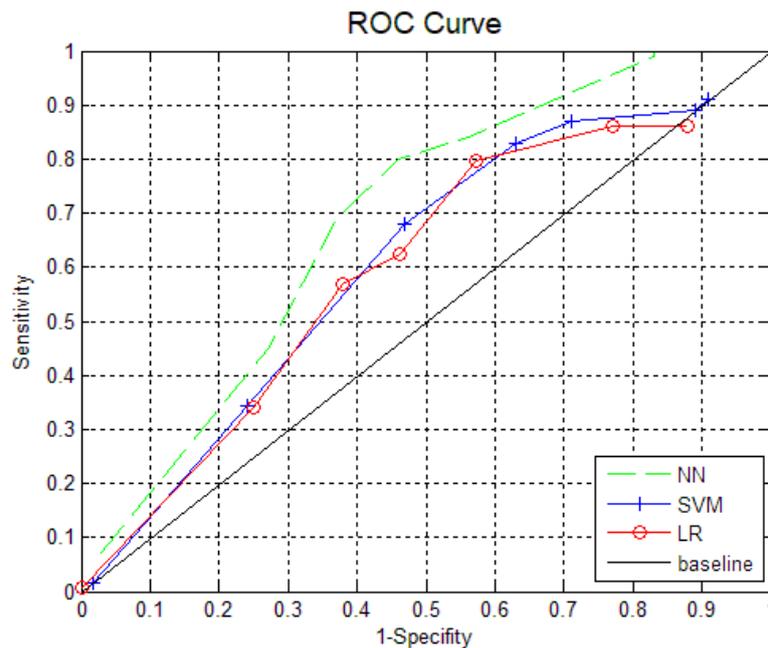


Fig 3. The ROC curve of data mining models used in the research

4-Conclusion

Today world have greatly affected by customer orientation, increasing competition, and rapid technological changes. Therefore, in today's highly competitive world, most organizations have found their survival moving towards to customer orientation and satisfaction. The use of marketing techniques makes organizations to be aware of customer's favorites properly, which consequences to better prediction of demands. Since demand forecasting is one of the most important factors in operations management. In this research, a data mining technique are developed for analysis of marketing strategies performance. The dataset under investigation is for an online shopping clothing

store for women. The collected data is for the year 2008, which includes 14 features and 32957 records. First, in the data preparation phase, the DT of the rminer package in R software is developed. Then, based on the implemented DT, important features are selected for our modeling. The three data mining models including: the LR, the SVM, and the NN models in rminer package of R software ran and compared. The three data mining models, are compared by two AUC and ROC curve criteria. The evidence exhibits that the NN model works better in prediction of the success of marketing performance. As future research suggested to use other variables such as reiterative IPs, as it makes the dataset more rich and the prediction more accurate.

References

Babcock, B., Babu, S., Datar, M., Motwani, R., & Widom, J. (2002). Models and Issues in Data Stream Systems" Proceedings of the twenty-first symposium on Principles of database systems, 1–16, <https://doi.org/10.1145/543613.54361>.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). Classification And Regression Trees (Routledge Ed), 1st ed.

Bucklin, R. E & Sismeiro, C. (2009). Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing. *Journal of Interactive Marketing*, 23(1), 35-48. doi:<https://doi.org/10.1016/j.intmar.2008.10.004>.

Chompaisal, S., Amphawan, K., & Surarerks, A.(2014) .Mining N-most Interesting Multi-level Frequent Itemsets without Support Threshold. Paper presented at the Recent Advances in Information and Communication Technology, Cham.

Cleger-Tamayo, S., Fernández-Luna, J. M., & Huete, J. F. (2012). Top-N news recommendations in digital newspapers. *Knowledge-Based Systems*, 27, 180-189. doi:<https://doi.org/10.1016/j.knosys.2011.11.017>.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. doi:10.1007/BF00994018.

Cortez, P. (2010). Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool. Paper presented at the Advances in Data Mining. Applications and Theoretical Aspects, Berlin, Heidelberg.

Hastie, T., Tibshirani, R., Friedman, J. (2008) .The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.): Springer-Verlag, NY, USA.

Huynha, H. M., Nguyenb, L. T. T., Voc, B., Nguyend. A., Tseng, V. S. (2020), Efficient methods for mining weighted clickstream patterns. *Expert Systems with Applications*, 142, 112993. doi: <https://doi.org/10.1016/j.eswa.2019.112993>.

Kawaf, F., & Istanbuluoglu, D. (2019). Online fashion shopping paradox: The role of customer reviews and facebook marketing. *Journal of Retailing and Consumer Services*, 48, 144-153. doi:<https://doi.org/10.1016/j.jretconser.2019.02.017>.

Kelly, G. A. (1955). The psychology of personal constructs, (Vol. 1).

- Koehn, D., Lessmann, S., & Schaal, M. (2020). Predicting online shopping behaviour from clickstream data using deep learning. *Expert Systems with Applications*, 150, 113-342. doi:<https://doi.org/10.1016/j.eswa.2020.113342>
- Li, H.-F. (2009). A sliding window method for finding top-k path traversal patterns over streaming Web click-sequences. *Expert Systems with Applications*, 36(3, Part 1), 4382-4386. doi:<https://doi.org/10.1016/j.eswa.2008.05.025>.
- Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks*, 6(4), 525-533. doi:[https://doi.org/10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5).
- Nasraoui, O., Cardona, C., Rojas, C., & Gonz'alez, F. (2003). Mining Evolving User Profiles in Noisy Web Clickstream Data with a Scalable Immune System Clustering Algorithm.
- Platt, J. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. *Advances in Kernel Methods-Support Vector Learning*, 208.
- Venables, W. N., Ripley, B.D. (2003). *Modern Applied Statistics with S* (4th ed.).
- Xia, Y., Liu, C., Da, B., Xie, F. (2018). A novel heterogeneous ensemble credit scoring model based on bstacking approach. *Expert Systems With Applications*, 93, 182-199. doi: [10.1016/j.eswa.2017.10.022](https://doi.org/10.1016/j.eswa.2017.10.022)
- Zeng, J., Zhang, S., & Wu, C. (2008). A framework for WWW user activity analysis based on user interest. *Knowledge-Based Systems*, 21(8), 905-910. doi:<https://doi.org/10.1016/j.knosys.2008.03.049>.
- Zhao, X., Niu, Z., & Chen, W. (2013). Interest before liking: Two-step recommendation approaches. *Knowledge-Based Systems* ,48 ,46-56 .doi:<https://doi.org/10.1016/j.knosys.2013.04.009>.