Designing a Customer Relationship Management System in Online Business

Negin Ziaei[1], Hamed Fazlollahtabar[2]

[1]Department of Information Technology Engineering, Mazandaran University of Science and Technology, Babol, Iran

[2]Department of Industrial Engineering, School of Engineering, Damghan University, Damghan, Iran

hfazl@du.ac.ir

Abstract

With the advancement of online shopping technology, it has become the first choice for most consumers. The activity of online stores in this competitive business space should be in line with the expectations of their customers. Understanding, collecting, maintaining and organize data in online stores makes it easier for managers to decide. So, in this research, we examine the textual and non-textual of user opinions and reviews. We use rapid miner software and text mining. In this research, the processes are aimed at finding active users, analyze the user type and their suggestions, analyzing the strengths and weaknesses of the products, and categorizing them with the K-NN and Naïve Bayes algorithms. Finally, suggestions were made to increase loyalty and improve business using the results obtained from the processes.

Keywords: Business intelligence (BI); Customer relationship management (CRM); Online shopping

1 . Introduction

Interest in customer relationship management (CRM) began to grow in 1990s(Ling & Yen, 2001; Y. Xu, Yen, Lin, & Chou, 2002).Regardless of the size of an organization, businesses are still motivated to adopt CRM to create and manage the relationships with their customers more effectively. An enhanced relationship with one's customers can ultimately lead to greater customer loyalty and retention and, also, profitability (Ngai, 2005). CRM has been one of the greatest Technological contributions to enterprises in the twenty-first century (Chao, Jen, Chi, & Lin, 2007).CRM consists of guidelines, procedures, processes and strategies which provide organizations the ability to merge customer interactions and also keep track of all customer related information. Technologies are utilized to attract new and profitable customers, retain and strengthen ties with current ones. CRM revolves around the concept of maintaining long-lasting, valuable relationships with customers(Khan, Ehsan, Mirza, & Sarwar, 2012).

Online shopping is growing faster than any moment in history of commerce and attracting many consumers to choose shopping online instead of traditional shopping channels. This new type of shopping mode, also known as e-shopping, online shopping, network shopping, Internet shopping, or Web-based shopping, featuring in freeing consumers from having to personally visit physical stores, is anticipated to greatly change people's everyday lives (Hsiao, 2009).The importance of effective CRM implementation is intensified in the e-business environment since customer loyalty is much more difficult to establish in this domain (Kımıloğlu & Zaralı, 2009).

The simple definition of e-CRM is customer relationship management on the web; however, e-CRM also includes the use of e-mail, e-commerce activity, and any other Internet-based customer touch points. Electronic customer relationship management (e-CRM) enables retailers to better meet the needs of their customers across retail formats and, at the same time, to maximize the strategic benefits of a multichannel strategy(Warrington, Hagen, & Feinberg, 2009).

Till now, the concepts of satisfaction and loyalty for website which involved in providing services on the website and transacting online, is a central concern of marketers. In recent years, electronic commerce has entered a phase of exponential growth and the use of the Internet in the consumer decision making process ensures that traders to make greater use of this tool. While consumer behavior in ecommerce seems to be a complex subject, the consumer expectations are changing, challenging traditional patterns of supply of commercial websites (Bashar & Wasiq, 2013).

In this research, we examine the textual and non-textual of user opinions and reviews about products they buy. We use rapid miner software and text mining and Data mining techniques to turn these data into understandable knowledge. The processes are aimed at finding active users, analyze the user type and their suggestions, analyzing the strengths and weaknesses of the products, and categorizing them. Finally, suggestions were made to increase loyalty and improve business using the results obtained from the processes.

The rest of this paper is organized as follows. Next, the related literature is reviewed. In Section 3, the proposed problem is described. Section 4, illustrates the algorithms required for implementation. Section 5 conducts a case study. Finally, we conclude in Section 6.

2 . Related works

Chen and Tseng (2011) propose a method for evaluating the quality of information in product reviews. They treat the quality evaluation of product reviews as a classification problem and employ a multiclass support vector machine (multiclass SVM) model to categorize reviews. In addition, they adopt a mature information quality (IQ) framework.Ghoseand Ipeirotis(2011) study, integrates econometric, text mining, and predictive modeling techniques By using Random Forest-based classifiers toward a more complete analysis of the information captured by user-generated online reviews in order to estimate their helpfulness and economic impact. Many tourism companies now actively use Internet sites as a key marketing and sales vehicle for their products and services. To be successful, tourism e-commerce services must be trustworthy. Zheng et al. (2013) developed a semi-supervised system called Online Review Quality Mining (ORQM). Embedded with independent component analysis and semi-supervised ensemble learning, ORQM exploits two opportunities: the improvement of classification performance through the use of a few labeled instances and numerous unlabeled instances, and the effectiveness of the social characteristics of e-commerce communities as identifiers of influential reviewers who write high-quality reviews.

Faed et al. (2014) proposed a conceptual framework including mathematical models, hypothesised relationships, perceived value and interactivity between customer, business and the system, as well as customer satisfaction and loyalty analytics. Based on nonlinear modelling and using a fuzzy inference system, namely the Takagi–Sugeno-type approach, they defined fuzzy rules, by means of which they ascertain the relationship between customer satisfaction and the main relevant variables. Jack and Tsai (2015) presented a framework for using text mining and *R*, a statistical software to gather customer feedback from Amazon website. A case study comparison of three devices compares and contrasts positive and negative aspects mentioned by the users, which is useful to improve future generations of products.Dixit and Kr (2016) analyzed feedback of customer on three different mobile brand for this two different classifiers has built to extract the feedback of customers, those are shared in e-commerce website and classify them broadly into 3 categories – good, bad and mixed. Xu and Li (2016)used latent semantic analysis (LSA), which is a text mining approach, they analyze online customer reviews of hotels. Their study provides a clue for hoteliers to enhance customer satisfaction and alleviate customer dissatisfaction by improving service and satisfying the customers' needs for the different types of hotels the hoteliers own.Małeckiet al. (2017) proposed the methodological approach of the decision support system for identifying Internet customer typology. Online users clustering was performed by cluster analysis. Additionally, by introducing a prediction mechanism based on the mathematical model of the Graph Cellular Automaton, the new customer can be quickly adjusted to the defined group of shop customers.

3 . Proposed problem

With the advancement of technology, people are looking for the easiest way to do their daily routines that online shopping is one of them. There are many reasons to shop online instead of going to the stores. It's easier to shop online, which means that it's always available 24 hours a day, 7 days a week, and in any weather conditions without having to look for a parking space and suffering in traffic these days. One of the other reasons is ease of use for searching the Intended products and selecting between various brands and the ability to compare products and their prices for reaching a good decision to purchase.So, given the benefits of online shopping, it can be concluded that there is a lot of competition between online stopes. In this business, customers play an essential role. In fact, choosing products online is consumers and customers job. They have the power to choose and the world of online retailers at their fingertips. If online shops do not meet their needs, they quickly go to the rival store. This is where the issue of customer loyalty is raised, which is also the goal of our research to increase loyalty in online shopping.

Increasing customer loyalty plays an important role in the organization to be successful, especially when customer purchases are not long-term success. To measure customer loyalty in a reviewing mechanism, we analyze the opinions, reviews, and ratings of customers and consumers who have shared their experiences on the website. Comments and ratings of products on the website are an important source of information for customer decision making. Here, our data is mostly made up of comments and customer reviews that are unstructured texts. So, we use text mining techniques in this research and analyze consumer's reviews by creating processes in rapid miner. The steps involved in this research to reach the goal and to analyze the review of the users are briefly illustrated as a model in **Figure 1**.

In the process of buying from online shops, users search on different parts of a website to find the product they wants. In this process, the user will be faced with a variety of options, and they may Giving up a lot of high quality products with small number of comments. In this research, finding the desired product for analysis is considered as an early and important step. At this stage, we have tried to use the best-selling, most viewed and most popular products that also have a large number of comment.

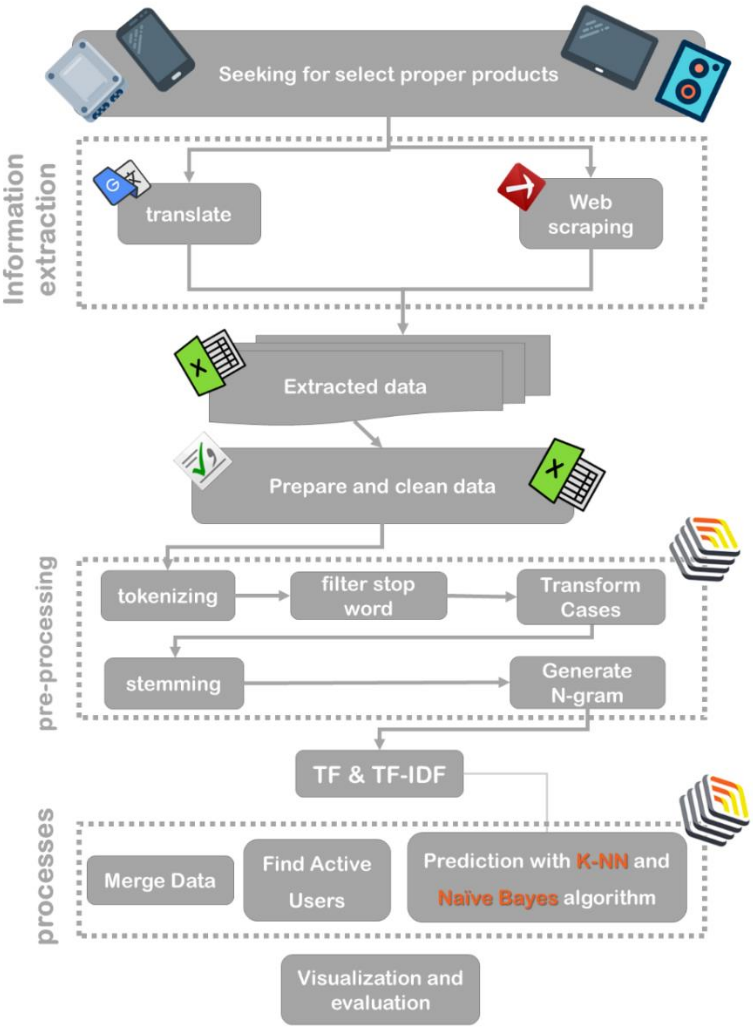

Figure 1 Research model

After selecting the desired products, we will extract and translate them for the next step. In this research, we use the web content mining method to extract data from comments. Web Scraping (also termed Screen Scraping, Web Data Extraction and Web Harvesting etc.) is a technique employed to extract large amounts of data from websites whereby the data is extracted and saved to a local file in your computer or to a database in table (spreadsheet) format. We use data miner and google translate extension in Google Chrome, for web scraping and translate comment (Persian to English) at the same time.

4 . Required algorithms

In data mining processes for prediction, we need to convert text data to vector and numeric data. This is done by the TF-IDF method. After the data are pre-processed then they are ready to use them in RapidMiner processes and data mining algorithms. Some processes include classifications predicting some features of products. In these processes, we use K-NN and Naive Bayes algorithms that described below.

4.1 . TF-IDF

The technique of calculating text weighting is called TF–IDF, which stands for Term Frequency–Inverse Document Frequency.

Calculating TF is very easy: it is simply the ratio of the number of times a keyword appears in a given document, $\mathbf{n}_k$ (where k is the keyword), to the total number of terms in the document, **n**:

$$TF = {}^{n_k}/_{n}$$

Considering the above example, a common English word such as "that" will have a fairly high TF score and a word such as "RapidMiner" will have a much lower TF score.IDF is defined as follows:

$$IDF = \log_2 {}^{N}/_{N_k}$$

Where **N** is the number of documents under consideration. For most text mining problems, **N** is the number of documents that we are trying to mine, and $\mathbf{N}_k$ is the number of documents that contain the keyword, k. Again, a word such as "that" would arguably appear in every document and thus the ratio ($\mathbf{N/N}_k$) would be close to 1, and the IDF score would be close to zero for "that." However, a word like "RapidMiner" would possibly appear in a relatively fewer number of documents and so the ratio ($\mathbf{N/N}_k$) would be much greater than 1. Thus the IDF score would be high for this less common keyword (Kotu & Deshpande, 2014).

Finally, TF–IDF is expressed as the simple product as shown below:

$$TF - IDF = n_k/n * \log_2 {}^{N}/_{N_k}$$

In RapidMiner studio user has four different options, each of which represents the relationship between the words/terms and the documents with different numbers:

1. **Binary Term Occurrence** places 1 in the intersection cell between a document (row) and a word/term (column) if the word/term occurs at least once in that document and places 0 otherwise. The number of occurrences in the document is ignored in this measure.

2. **Term Occurrence** places the exact number of occurrences of a word/term in the intersection cell between the document (row) and the word/term (column). If the word/term does not occur in that document, 0 is placed in the intersection cell.

3. **Term Frequency** places the relative frequency of the word/term in the document in the intersection cell. This measure is calculated by dividing the number of occurrences of a word/term into the number of total words in that document.

4. **TF-IDF** stands for *Term Frequency-Inverse Document Frequency*. It is arguably the most commonly used numerical representation in text mining. It calculates a numerical value that emphasizes both the frequency of the term in a document (more is better) and the rareness of the same term in the collection of all documents (less is better) (Miner, Elder IV, & Hill, 2012).

In this study, we use **TF-IDF** and **Term Frequency** for text mining operations.

4.2 . Naïve Bayes

The naïve Bayesian algorithm is built on Bayes' theorem, named after Reverend Thomas Bayes. We make use of Bayesian algorithm for the conditional stochastic text occurrence. Consider **X** as an evidence and Y as an output. Then, the probability of output P(Y) is called *prior probability*. P(Y|**X**) is called the *conditional probability,* which provides the probability of an outcome given the evidence when we know the value of **X**. Bayes' theorem states that

$$P(Y|\boldsymbol{X}) = \frac{p(Y) * P(\boldsymbol{X}|Y)}{P(\boldsymbol{X})}$$

More generally, in an example set with n attributes **X** = {X_1, X_2, X_3... X_n},

$$P(Y|X) = \frac{p(Y) * \prod_{i=1}^{n} P(\boldsymbol{X}_i|Y)}{P(\boldsymbol{X})}$$

Since P(X) is constant for every value of Y, it is enough to calculate the numerator of the equation $\mathrm{p(Y)} * \prod_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{P}(\mathbf{X}_{\mathrm{i}}|\mathrm{Y})$ for every class value (Kotu & Deshpande, 2014) .

4.3. K-NN

The *k*-nearest-neighbor method was first described in the early 1950s. The method is labor intensive when given large training sets and did not gain popularity until the 1960s when increased computing power became available. It has since been widely used in the area of pattern recognition. When given an unknown tuple, a ***k*-nearest-neighbor classifier** searches the pattern space for the *k* training tuples that are closest to the unknown tuple. These *k* training tuples are the *k* "nearest neighbors" of the unknown tuple.

"Closeness" is defined in terms of a distance metric, such as Euclidean distance. The Euclidean distance between two points or tuples, say, $\mathbf{X}_1$ =(x_{11}, x_{12}, ... , x_{1n})and $\mathbf{X}_2$ =(x_{21}, x_{22}, ... , x_{2n}), is

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (X_{1i} - X_{2i})^2}$$

In other words, for each numeric attribute, we take the difference between the corresponding values of that attribute in tuple $\mathbf{X}_1$ and in tuple $\mathbf{X}_2$, square this difference, and accumulate it. The square root is taken of the total accumulated distance count. Typically, we normalize the values of each attribute before using above Equation. This helps prevent attributes with initially large ranges (e.g., *income*) from outweighing attributes with initially smaller ranges (e.g., binary attributes).Min-max normalization, for example, can be used to transform a value *v* of a numeric attribute *A* to *v*' 0 in the range [0, 1] by computing

$$v' = \frac{v - min_A}{max_A - min_A}$$

where min_a and max_A are the minimum and maximum values of attribute *A*. For *k*-nearest-neighbor classification, the unknown tuple is assigned the most common class among its *k*-nearest neighbors. When *k*=1, the unknown tuple is assigned the class of the training tuple that is closest to it in pattern space. Nearest-neighbor classifiers can also be used for numeric prediction, that is, to return a real-valued prediction for a given unknown tuple. In this case, the classifier returns the average value of the real-valued labels associated with the *k*-nearest neighbors of the unknown tuple (Han, Pei, & Kamber, 2011) .

5 . Case study

The purpose of this research is to analyze the texts of customers' comments with the help of text mining and provide strategic Approach for online shops to enhance customer loyalty. Digikala is the online shop that we investigated for turn their comments into knowledge. The segmentation goods on the digikala website includes 8 parts: digital, fashion and clothing, cosmetics, books, culture and art, sports and travel, mother and child, and vehicles and industrial. According to the reviews, we are select products from the digital category. Some reviews examples include:

- In the fashion and clothing section, most of the products belonged to another site called Digi style. We noticed that users did not comment on Digi style website, even though there was a definite panel for the review section. In order to increase the trust and attraction of customers, to improve customer relationship management, the client should be encouraged to interact with the site. Also, by reviewing this section, we found that bestsellers, as well as the most popular products are smart watches and gadgets, could be considered as part of digital products.

- In the home and kitchen section, all the products have small amount of comment except audio and video equipment and game consoles.

- In the vehicle and industrial section, there were few products, which there have small amount of a comments and those a lot of comments, have nobuyers.

As discussed above, digital products have more comments than other section products so for this study we select 4 products from this section. Categories of the products we intend to mind their comments are: smartphones, tablets, external hard and speakers that shown in **Table 1**. Comments were sorted in the order of their usefulness, not according to the date.

After selecting the products, we are at the stage of extracting customer data and comments about the four products that mentioned before. With the web scraping technique and using the data miner extension, 4 recipes were created for selected products.

Table 1 selected products

products	Name	section	Number of all comments	Number of chosen comment
	Samsung Galaxy Note 8 SM-N950FD Dual SIM 64GB Mobile Phone	Digital product > cellphone	825	500
	ASUS Zen Pad 8.0 Z380KNL 4G 16GB Tablet	Digital Product > Tablet & E-reader > Tablet	314	All
	Western Digital Elements External Hard Drive - 2TB	Digital Products > Computers and Accessories > Data Storage Equipment > External Hard Disk	254	All
	JBL Go Portable Bluetooth Speaker	Digital Products > Computers and Accessories > Speaker	214	All

After initial preparation and cleaning, data are entered into the RapidMiner software or RM. RM consists of two important section view, design view which processes are formed and displays and result view that show the results of that processes.

In this research, five processes are constructed using RM that shown in **Figure** , which we will explain about them in further detail. Process 1 and 2 are just generate inputs for other process so we are not discuss about them.

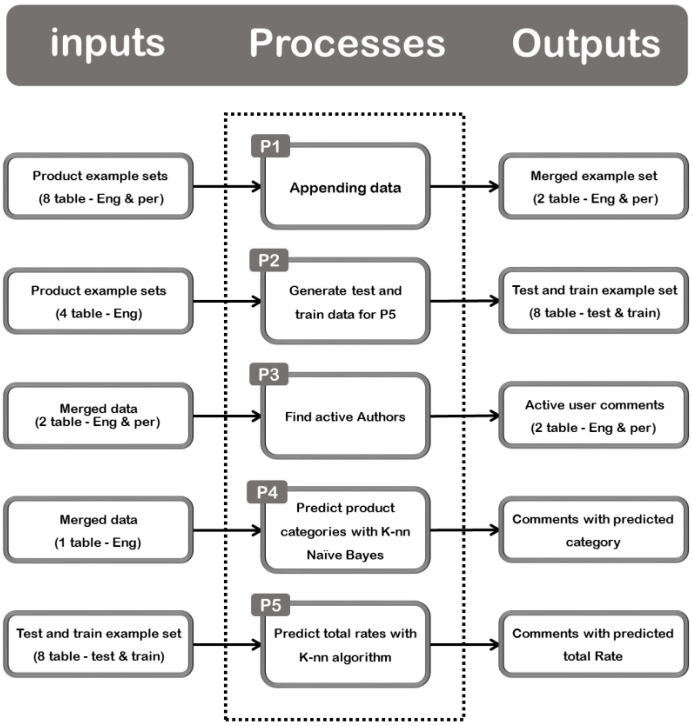

Figure 2 Research process

Figure shows the results view and the chart section of a scatter diagram of output data from p3. The x-axis specifies the number of times each user comment, including 4 values, but because of the jitter data are in decimal form. The Y-axis shows the date. The colors in this chart are the authors. The comments in the shape of the circle are all green, which means the authors are identical. The number of times the author comment is in x-axis and the dates can also be distinguished from the y-axis.

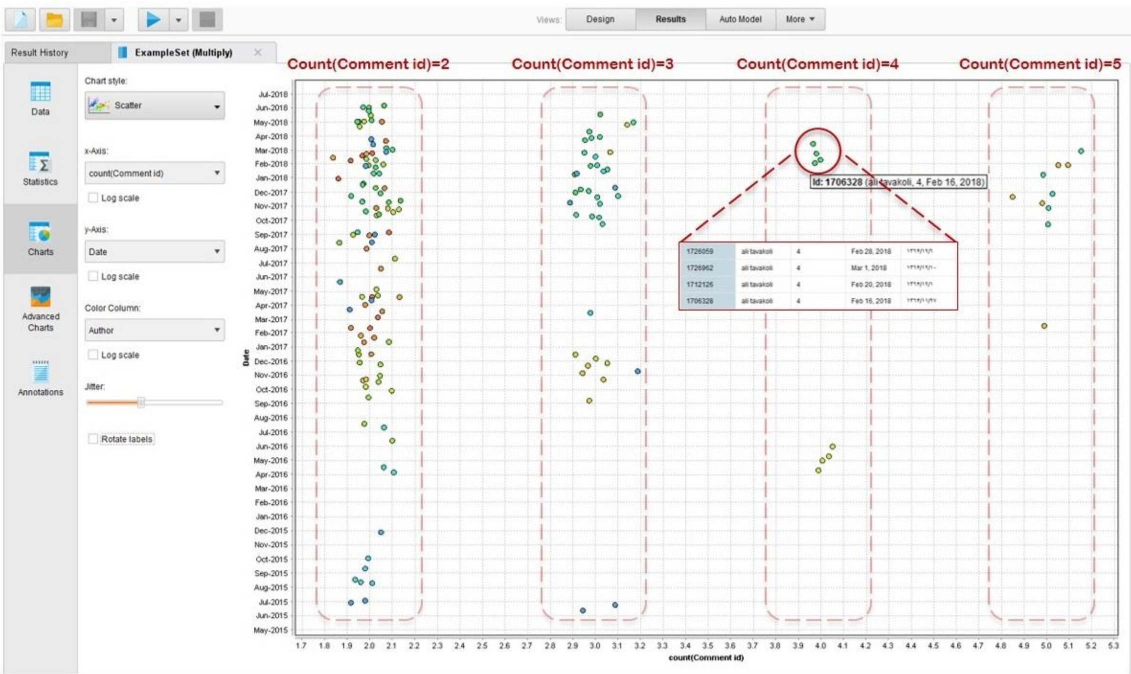

Figure 3 Scatter chart of output data (P3)

The results of this process include 2 (Persian and English) example set. The output of p3 is table consists of 157 rows, from 1278 review 175 reviews were made by guest users and 67 of them were known as active user. Now, in **Table 2**, we discuss about these 67 people, which comments 157 times.

Table 2 Analyze active user comments

num	observation	analysis
1	The number of dispersion of buyer and non-purchasers, except mobile, which was only one buyers that comment twice, was regular in other products.	Despite the fact that the comments in the mobile category are also much more in terms of number, detail and accuracy of product reviews than other categories, this is a sign of uncertainty and trust of users in expensive items. So, there is a need to attract customer loyalty and trust in these types of products.
2	[1] Some people commented on several different products. [2] Many comments have been posted during a period of time for months or even years, including complete analysis and reviews. [3] Some other comments were also written in one day, sometimes containing similar content or repetitive texts.	The presence of people in Group [2] is a good sign that users are constantly sharing their opinions when they buy and after using the product at different period of times, which has a significant impact on users who are going to buy that products. Somehow it will attract and increase the range of new customers. In fact, these users as well as the users of Group [1] are active users who need more attention and more services with the aim of maintaining them and increasing their loyalty. Users of the group [3] have little or no effect on business improvement, it is possible to increase this impact by giving them more awareness.
3	• Despite having a separate section called Q&A, some questions have been expressed in some comments! • Some comments also include requests for the availability of goods, the existence of a particular color, and a special offers.	• Changing the UI/UX website to help guide the user towards the Q&A section or add the ability to reply to comments in the same section of the comments may help to solve the issue. • By adding a part to express user requests, this weakness can turn out to be a strong point, which will greatly increase sales.

The output of the *k-nn* algorithm model in **Figure** shows that the learning model consists of 895 comments and 50133 columns with n = 8, and hard, mobile, speaker and tablet classes.

Figure 4 The model output of the *k-nn* algorithm (P4)

Example Set output of this algorithm contains 383 comments that have our test data, 50133 regular attributes and 8 special attributes. These eight attributes are, id column with the role id, a category column that represents the actual commented class, and another prediction (category) column that displays the predicted classes, we also have 4 confidence columns for each label class, which indicates the likelihood of each label class value.

The output of the model in the Naïve Bayes algorithm has 3 tabs, as shown in **Figure** . The first one is *description* shows that the output of the model has 23832 columns, and the numbers generated for each class represent the probability of data in that class. Given that the number of comments in the mobile class is greater than the other 3 classes, the probability is also higher. The next tab is *chart*, contains probability density functions for the specified attributes. In these charts, the x-axis is the value or in other words, the weight of the word, and the y-axis represents the density of that value in different classes with the color of that class. The third tab is *distribution table* with all attribute values and corresponding probability measures. In this table, the mean parameters and standard deviations of each word are calculated in different classes. For example, in distribution table the mobile column is set to ascending and the note word is in the highest row, so the number of occurrences is also high in this class. In the chart section, we also set attr to note, in the specified curves it is known that the part where the note is zero is the highest density for 3 classes: hard, speakers and tablet, and the green curve showing the mobile class has densities in non-zero sections indicates The existence of this word is in the mobile class and it is almost non-existent in the other 3 classes.

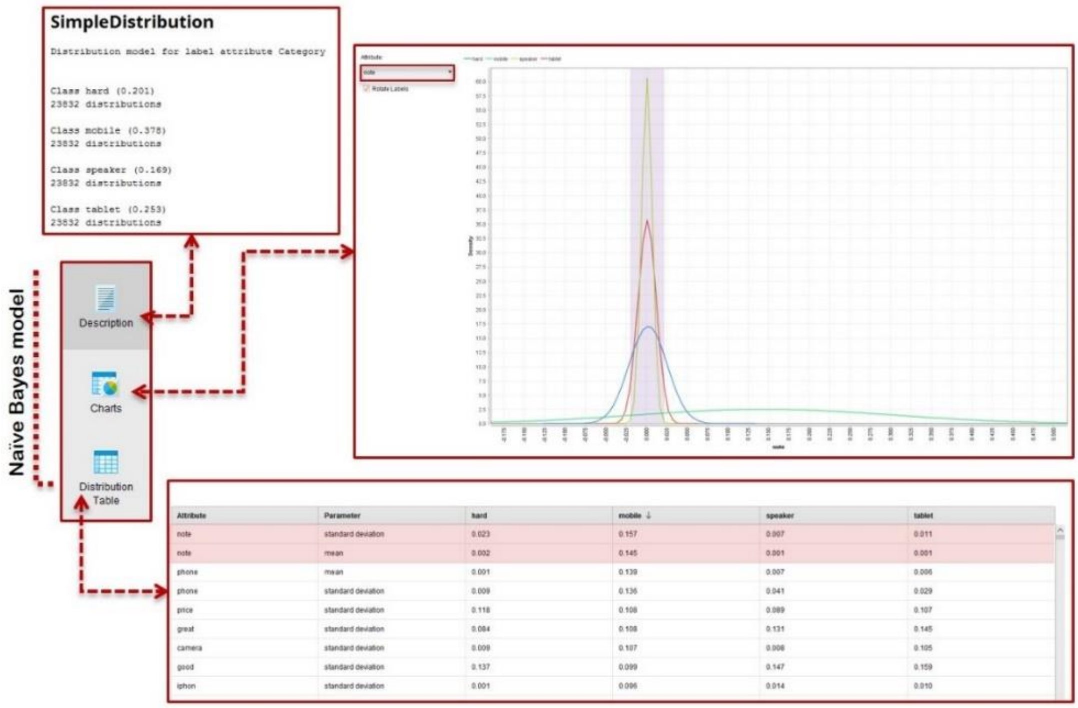

Figure 5 The model output of the Naïve Bayes algorithm that include 3 tabs (P4)

In **Figure** , the accuracy and classification errors of the comments are also expressed in terms of mean precision, mean recall and F-score ratings for comments on 4 products. The mean precision for the external hard drive was higher than the rest, and 3 and 4 rates were well predicted. The highest performance is in the various metrics except for the precision is for the mobile. The reason for the higher percentage of mobile was the number of comments first, compared with the other products, and secondly, the number of classes was 3, which resulted in a better prediction accuracy, with most of the comments being attributed to one of the three classes, which has led to two classes There is no other right sorting.

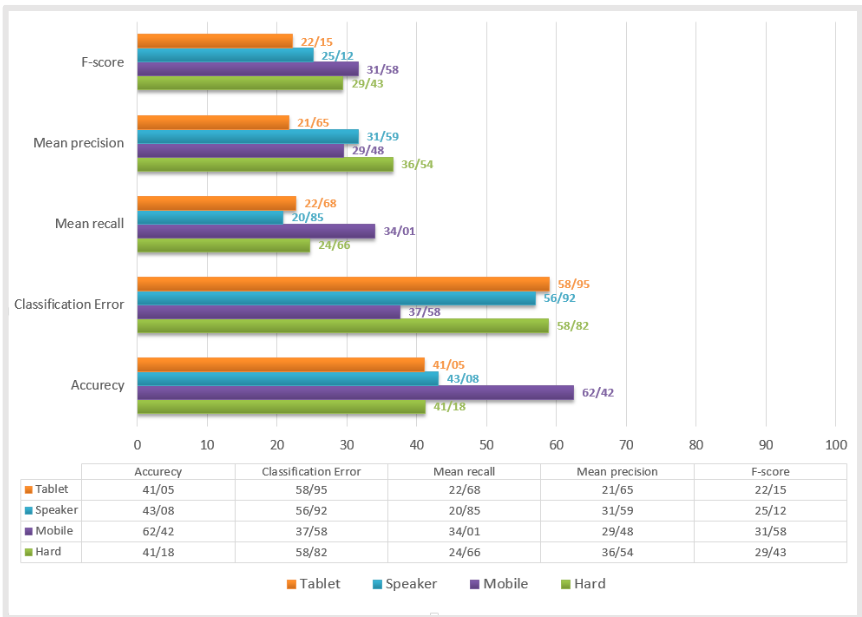

Figure 6 Final result of classification (P5)

In this process, given the accuracy of predictions, these scores need to be improved. Regarding these results, a number of users do not comment correctly and scrupulously. The accuracy of users in the ratings and reviews is very effective in the output. To be more precise, it's better to increase the rating periods, instead of 1 to 5 users have the ability to rate from 1 to 100. Also, if there are more comments on the various products like mobile, we will have more precise predictions, so there is a need to encourage users to more comment the experience of working with products to increase comments and to have a more accurate classification.

6 . Conclusion

As mentioned, online shopping has a greater advantage than physical shopping stores, so competition among online shops is very high. Considering that customers play an important role in these businesses, the need for customer relationship management is increasingly felt. User comments need to be taken into consideration in order to keep past customers and attract new customers, which in our research we analyzed reviews from digikala website. At first, various product groups were selected for product selection for analysis. The result of the observation was digital products for having better feedbacks. To improve this with advertising, trust, and better service, you can increase the number of non-digital consumer comments, which will increase the number of buyers by reading comments in the face of high-quality services. In the following, four products were selected: External hard, smartphone, speaker and tablet from the Digital Products section, then extracted, translated and stored in Excel format. Analysis on comments was created in the form of processes in the RM software. This study can be used to support management decisions in web design, products introductions, and to utilize the suppliers in the customers' interests. Below are some suggestions for future research in this area:

- Using Python for Persian Text mining by the hazm library and using stemming techniques in Persian language
- Text mining and sentiment analyzing of tweets containing the hashtag of Digikala
- Use more products with the same categories as well as analysis on non-digital products
- Sentiment Analysis for different product categories

References

1. Bashar, A., & Wasiq, M. (2013). E-satisfaction and E-loyalty of Consumers Shopping Online. Global Sci-Tech, 5(1), 6-19.

2. Chao, C., Jen, W., Chi, Y., & Lin, B. (2007). Determining technology trends and forecasts of CRM through a historical review and bibliometric analysis of data from 1991 to 2005. International Journal of Management and Enterprise Development, 4(4), 415-427.

3. Chen, C. C., & Tseng, Y.-D. (2011). Quality evaluation of product reviews using an information quality framework. Decision Support Systems, 50(4), 755-768.

4. Dixit, S., & Kr, S. (2016). Collaborative Analysis of Customer Feedbacks using Rapid Miner. International Journal of Computer Applications, 142.(2)

5. Faed, A., Hussain, O. K., & Chang, E. (2014). A methodology to map customer complaints and measure customer satisfaction and loyalty. Service Oriented Computing and Applications, 8(1), 33-53.

6. Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. IEEE Transactions on Knowledge and Data Engineering, 23(10), 1498-1512.

7. Gohary, A., Hamzelu, B., & Alizadeh, H. (2016). Please explain why it happened! How perceived justice and customer involvement affect post co-recovery evaluations: A study of Iranian online shoppers. Journal of retailing and Consumer services, 31, 127-142.

8. Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques: Elsevier.

9. Hsiao, M.-H. (2009). Shopping mode choice: Physical store shopping versus e-shopping. Transportation Research Part E: Logistics and Transportation Review, 45(1), 86-95.

10. Jack, L., & Tsai, Y. (2015). Using Text Mining of Amazon Reviews to Explore User-Defined Product Highlights and Issues. Paper presented at the Proceedings of the International Conference on Data Mining (DMIN).

11. Khan, A., Ehsan, N., Mirza, E., & Sarwar, S. Z. (2012). Integration between customer relationship management (CRM) and data warehousing. Procedia Technology, 1, 239-249.

12. Kim, J., Jin, B., & Swinney, J. L. (2009). Therole of etail quality, e-satisfaction and e-trust in online loyalty development process. Journal of retailing and Consumer services, 16(4), 239-247.

13. Kim, M.-J., Chung, N., & Lee, C.-K. (2011). The effect of perceived trust on electronic commerce: Shopping online for tourism products and services in South Korea. Tourism Management, 32(2), 256-265.

14. Kımıloğlu, H., & Zaralı, H. (2009). What signifies success in e-CRM? Marketing intelligence & planning, 27(2), 246-267.

15. Kotu, V., & Deshpande, B. (2014). Predictive analytics and data mining: concepts and practice with rapidminer: Morgan Kaufmann.

16. Ling, R., & Yen, D. C. (2001). Customer relationship management: An analysis framework and implementation strategies. Journal of computer information systems, 41.97-82 ,(3)

17. Małecki, K., & Wątróbski, J. (2017). The classification of internet shop customers based on the cluster analysis and graph cellular automata. Procedia Computer Science, 112, 2280-2289.

18. Miner, G., Elder IV, J., & Hill, T. (2012). Practical textmining and statistical analysis for non-structured text data applications: Academic Press.

19. Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. Marketing Science, 31(3), 5.543-21

20. Ngai, E. W. (2005). Customer relationship management research (1992-2002) An academic literature review and classification. Marketing intelligence & planning, 23(6), 582-605.

21. Park, C.-H., & Kim, Y.-G. (2003). Identifying key factors affecting consumer purchase behavior in an online shopping context. International journal of retail & distribution management, 31(1), 16-29.

22. Reichheld, F. F., Markey Jr, R. G., & Hopton, C. (2000). E-customer loyalty-applying the traditional rules of business for online success. European Business Journal, 12(4), 173.

23. Srinivasan, S. S., Anderson, R., & Ponnavolu, K. (2002). Customer loyalty in e-commerce: an exploration of its antecedents and consequences. Journal of retailing, 78(1), 41-50.

24. Van Riel, A. C., Liljander, V., & Jurriens, P. (2001). Exploring consumer evaluations of e-services: a portal site. International Journal of Service Industry Management, 12(4), 359-377.

25. Warrington, P. T., Hagen, A., & Feinberg, R. (2009). Multi-channel retailing and customer satisfaction Comparison-Shopping Services and Agent Designs (pp. 165-177): IGI Global.

26. Wei, J.-T., Lee, M.-C., Chen, H.-K., & Wu, H.-H. (2013). Customer relationship management in the hairdressing industry: An application of data mining techniques. ExpertSystems with Applications, 40(18), 7513-7518.

27. Xu, X., & Li, Y. (2016). The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. International journal of hospitality management, 55, 57-69.

28. Xu, Y ,.Yen, D. C., Lin, B., & Chou, D. C. (2002). Adopting customer relationship management technology. Industrial management & data systems, 102(8), 442-452.

29. Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty: The role of switching costs. Psychology & Marketing, 21(10), 799-822.

30. Yee Liau, B., & Pei Tan, P. (2014). Gaining customer knowledge in low cost airlines through text mining. Industrial management & data systems, 114(9), 1344-1359.

31. Zhan, J., Loh, H. T., & Liu, Y .(2009) .Gather customer concerns from online product reviews–A text summarization approach. Expert Systems with Applications, 36(2), 2107-2115.

32. Zheng, X., Zhu, S., & Lin, Z. (2013). Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach. Decision Support Systems, 56, 211-222.