Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring

Roman Gramyak[a], Hrystyna Lipyanina-Goncharenko[a], Anatoliy Sachenko[a,b], Taras Lendyuk[a] and Diana Zahorodnia[a]

[a] *West Ukrainian National University, Lvivska str., 11, Ternopil, 46000, Ukraine*
[b] *Kazimierz Pulaski University of Technology and Humanities in Radom, Department of Informatics, Jacek Malczewski str., 29, Radom, 26 600, Poland*

Abstract

Finding the best products for sale is one of the most important steps in the process of a profitable company creating. That is why the choice of goods for the online store must be made carefully, taking into account both the opportunities and analysis of prospects in the niche, and a number of other important parameters. One of the methods of competitive product choosing can be products analysis in marketplaces based on the emotional feedbacks coloring. Research on product feedbacks is an extremely popular topic, as confirmed by research analysis. Feedbacks can be constantly re-read, but when there are many products in one segment, because there are more and more manufacturers, it is time consuming. Therefore, the development of technology that can automate this process is necessary for the sales business. Paper develops the intelligent method of a competitive product choosing based on the emotional feedbacks coloring, which is divided into three blocks: parser of feedbacks, emotional coloring determination and feedbacks classification. The data will help retailers manage their websites wisely and help customers make purchasing decisions. The implementation of the method was carried out on the data of the Ukrainian site Rozetka, where 4477 feedbacks were used. The classification was tested by eight classical machine-based classification methods, namely Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression.

Keywords

Product, feedback, parser, text emotional coloring, classification, machine learning.

1. Introduction

Other people's opinions have always been an important piece of information for most of us in the decision-making process. The interest shown by users in online feedback and comments, as well as the potential impact of these comments on issues in discourse and decision-making, make us pay attention to this aspect of online activity. The purpose of tone analysis is to find ideas in the text and determine their properties. This method finds its practical application in such fields as sociology, political science, marketing, medicine and many others.

The main approaches that can be used to analyze emotional mood are machine learning and a vocabulary-based approach. Currently, the analysis of attitudes is an interesting topic and direction of development, as it has many practical applications. Companies use it to automatically analyze survey feedbacks, product feedbacks, and social media comments to gain valuable information about their brands, products, and services. However, today seller can use the analysis of emotional mood and to

select the potentially most profitable product that is in good demand in the market, it will allow online sellers to reduce the risks of choosing a product for trade, and thus increase revenue.

In this regard, it can be considered that the development of a method of selection by machine learning algorithms of the intelligent method of a competitive product choosing based on the emotional feedbacks coloring is one of the most promising areas in online commerce.

This paper is structured as follows. A Section 2 discusses the analysis of related work, and the Section 3 presents the proposed intelligent method of a competitive product choosing based on the emotional feedbacks coloring. The Section 4 presents a case study and discussion, and the Section 5 is summarizing the obtained results.

2. Related works

Reviewing customer feedbacks in online stores is attracting more and more attention from practitioners and scientists. Many studies have analyzed the impact of online feedbacks on sales of goods and services: Amazon products [1], technical characteristics of products from social networks [8], Flipkart, Snapdeal and Amazon India brands [34] and Amazon, Sony and Pocketbook on Facebook, reviews of restaurants from Yelp.com [49], reviews of books (book.dangdang.com). Also, there are works that analyze research on predicting the usefulness of the review [40], identifying data sources, ML methods [23].

Important after data retrieval is the ranking of these data, [18] presents the structure of merging information to rank product feedbacks. Also, ranking approaches are proposed: method on different aspects of alternative products [11]; model [19] of regression prediction based on their quality; hierarchical approach [20], word level and review level; classifier [15] Ensemble, which uses linguistic features to praise or classify complaints; method [21], which returns the ranking lists of useful feedbacks according to their usefulness in relation to the product, a model of multiple linear regression using the method of elastic network regularization.

After structured feedback receiving, seller need to conduct a detailed analysis. In [3] there is developed the system for predicting the usefulness of the feedback for multilingual online feedbacks, which displays the best results in terms of forecasting the usefulness and classification. A [13] model for detecting product defects based on feedback on social networks is proposed. In [14] the method of analysis of competitive advantages of the product is offered, which provides an important basis for quality management and development of marketing strategy by UGC mining. An empirical study of the choice of functions for predicting the usefulness of online product feedbacks was conducted in [22]. The relationship between moods and usefulness for feedback was studied [26], a method was developed for full use of mood features in assessing usefulness for feedback and it was investigated whether the type of product affects the assessment of usefulness for feedback. Based on the bias of negativity and the theory of summation, [33] proposed a theoretical model that explains the usefulness of feedbacks on the Internet based on the specific characteristics of these feedbacks (i.e., duration, evaluation, arguments frame). [43] developed a system that takes all feedbacks containing Hindi, as well as texts in English, and determines the mood expressed in this feedback for each attribute of the product, as well as the final feedback of the product. Several utility predictions models have been studied [48] using multidimensional adaptive regression, classification and regression trees, random forest approaches, neural networks, and deep neural networks using two real Amazon product feedback datasets.

Determining the emotional state is the next step in analyzing user feedback. To analyze the emotional mood of feedbacks, seller can use several approaches, namely: models [30, 32] of multiple regression; ensemble method [9] with several revisions; clustering of text [6] focused on the business sector; model [4] consumer decision-making in terms of risk; method [16] to dynamically assign a coefficient of influence to each basic student in the ensemble; SentiCon by LSA consolidation [25]; model [29] prediction based on a combination of neural network (CNN) and TransE; method [38] of classification of action verbs in the feedback text; integrated model [35] of decision support for finding products on the Internet; approach [44] is based on language modeling (LM); model [7] of product recommendations based on product rating, filtering and hybrid classification of decision trees; method [10] of hybrid classification of moods on the basis of attributes for the purpose of receiving orientation on moods, approach [12] on the basis of fuzzy numbers of an interval of type 2; model [31] predicting the usefulness of feedback is presented using a truncated decision tree C4.5.

The most accurate in determining the emotional mood of the feedback are models [5, 36, 37] based on machine learning, deep learning [17, 39] and convolutional neural network [2]. An uncontrolled thematic model of emotion extraction has been proposed [24], which adds appropriate relationships between aspect words and thought words to express comments as a bag of aspect-thought pairs. A model is presented [27] for determining the usefulness of comments using text functions removed from feedbacks, based on different types of functions, including emotional, linguistic and textual features, values of valence, excitation and dominance, feedback duration and polarity of comments. An in-depth model has been proposed [28] for using the features of feedbacks based on content, semantics, attitudes, and metadata to predict the usefulness of the feedback. [41] presents a general structure that uses natural language processing (NLP) methods, including mood analysis, text data analysis, and clustering methods, to obtain new estimates based on consumer sentiment for different product characteristics. It is proposed [42] to use the plot of the film Wikipedia to identify genre factions using methods of text extraction, based on the model of word reasoning, where (frequency) of occurrence of each word is used as a feature to prepare a classifier. [45] various model architectures have been proposed that can be used to predict the genre and rating of a film in the different languages available in our abstract-based data set. Proposed [46] approach to the classification of moods and tweet language, the architecture includes a convolutional neural network (ConvNet) with two different outputs, each of which is designed to minimize the error of classification or assignment of distribution, or language identification. [47] The classification of the film genre is proposed only on the basis of poster images, on the basis of an in-depth neural network.

As it confirmed by the analysis above a research on product feedbacks is an extremely popular topic, the data help retailers manage their websites wisely and help them make purchasing decisions. Analysis of the emotional state of the feedbacks is an important point in solving the problem of determining the most popular product, which allows the seller to choose the most profitable product for sale.

In this regard, a goal of this paper is to develop the intelligent method of a competitive product choosing based on the emotional feedbacks coloring.

3. Proposed method

To indicate further the novelty of the proposed method authors conducted a detailed analysis of analogues where data is extracted, the emotional mood of the feedbacks is determined and classified (Table 1).

Table 1
Analysis of research analogues

Source	Formulation of the problem	Data mining	Emotional coloring	Classification		Data
				Method	Accuracy	
6	A model of product recommendations based on mood assessments; real-time product data is proposed	DOM Parser	Normalized Product Feedback Score	Random Tree Hoeffding Tree Adaboost +RT	0.9691 0.978 0.967	www.amazon.com
11	A new ranking method is offered through online feedbacks based on various aspects of alternative products, which combines both objective and subjective values.	-	LDA	PageRank	-	-
14	A method of analysis of competitive advantages	Parser	UGC	NTUSD Hownet	0.7425 0.6425	Social networks

Source	Formulation of the problem	Data mining	Emotional coloring	Classification		Data
				Method	Accuracy	
	of the product is proposed, which provides an important basis for quality management and development of marketing strategy through the extraction of UGC.			Trained domain-specific sentiment lexicon with H=9	0.8625	
18	The approach of networks to the detection of social networks based on the method of fuzzy clustering.	Parser	LSGDM	SVM	0.63	www.twitter.com
				SVM-DS	0.97	
21	A method with an extended list of functions for presenting the overview and a model of multiple linear regression using the method of elastic network regularization is proposed.	DOM	LS2F	Decision tree with LS2F	0.77	www.amazon.com
26	The data set used methods to obtain information and assess moods. Gradient amplification and F-estimation methods were used to classify the data set with these characteristics.	-	Python library	Random forest	0.5742	www.amazon.com
			SentiWordNet	Gradient boosting	0.5721	
33	ANOVA assays are used to identify the levels of each of these characteristics that maximize the tangible usefulness of online consumer feedbacks, and an artificial neural network approach is used to predict the usefulness of this feedback based on its characteristics.	-	ANOVA	ANN	0.84	-
37	A new approach called value-based neighbor selection (VNS) is	-	-	Classification threshold	0.706	JD.com

Source	Formulation of the problem	Data mining	Emotional coloring	Classification		Data
				Method	Accuracy	
	proposed to address the above constraints.					
39	The OCC model and the CNN-based generalization method for Chinese microblogging systems are proposed.	-	OCC	CNN	0.84	Chinese social networks
41	A general structure that uses natural language processing techniques, including mood analysis, text data analysis, and clustering techniques, to obtain new estimates based on consumer sentiment for different product characteristics.	-	-NLP	RSS	-	www.amazon.com
48	Several utility prediction models are built using multidimensional adaptive regression, classification and regression tree, random forest approaches, neural network, and deep neural network using two real-world Amazon product feedback datasets.	-	-	MAR	0.90	www.amazon.com
				CART	0.97	
				RandF	0.82	
				Neural Net	0.75	
				Deep NN	0.60	

To reduce the time spent on choosing a popular new product, the authors have developed the intelligent method of a competitive product choosing based on the emotional feedbacks coloring. The proposed method is illustrated schematically (Fig.1) and is represented by the following steps.

Step 1. Connect the necessary libraries. They provide general mathematical and numerical operations in the form of pre-compiled, fast functions. They are combined into high-level packages (Block 1).

Step 2. Select the desired site for further work (Block 2).

Step 3. Enter the link to the product categories that user needs (Block 3).

Step 4. Parsing of user feedbacks to goods is performed. Divide them by product id and add to the text file. A separate feedback for each line (Block 4).

Step 5. Creating a database of feedbacks collected together (Block 5).

Step 6. Connection of feedbacks from a DB (Block 5) for the further work (Block 6).

Step 7. The raw text is quite messy for these feedbacks, so seller need to clean the text to make it easier for the algorithm to recognize the mood (Block 7).

Step 8. Vectorization is performed (Block 8). Vectorization is a type of program parallelization in which single-threaded applications that perform one operation at a time are modified to perform several operations of the same type at the same time. To perform the vectorization process seller need to: tokenization (Block 8.1), ignoring single characters (Block 8.3), converting text into numerical vectors and n-grams (Block 8.5) and then form an array of values (Block 8.6).
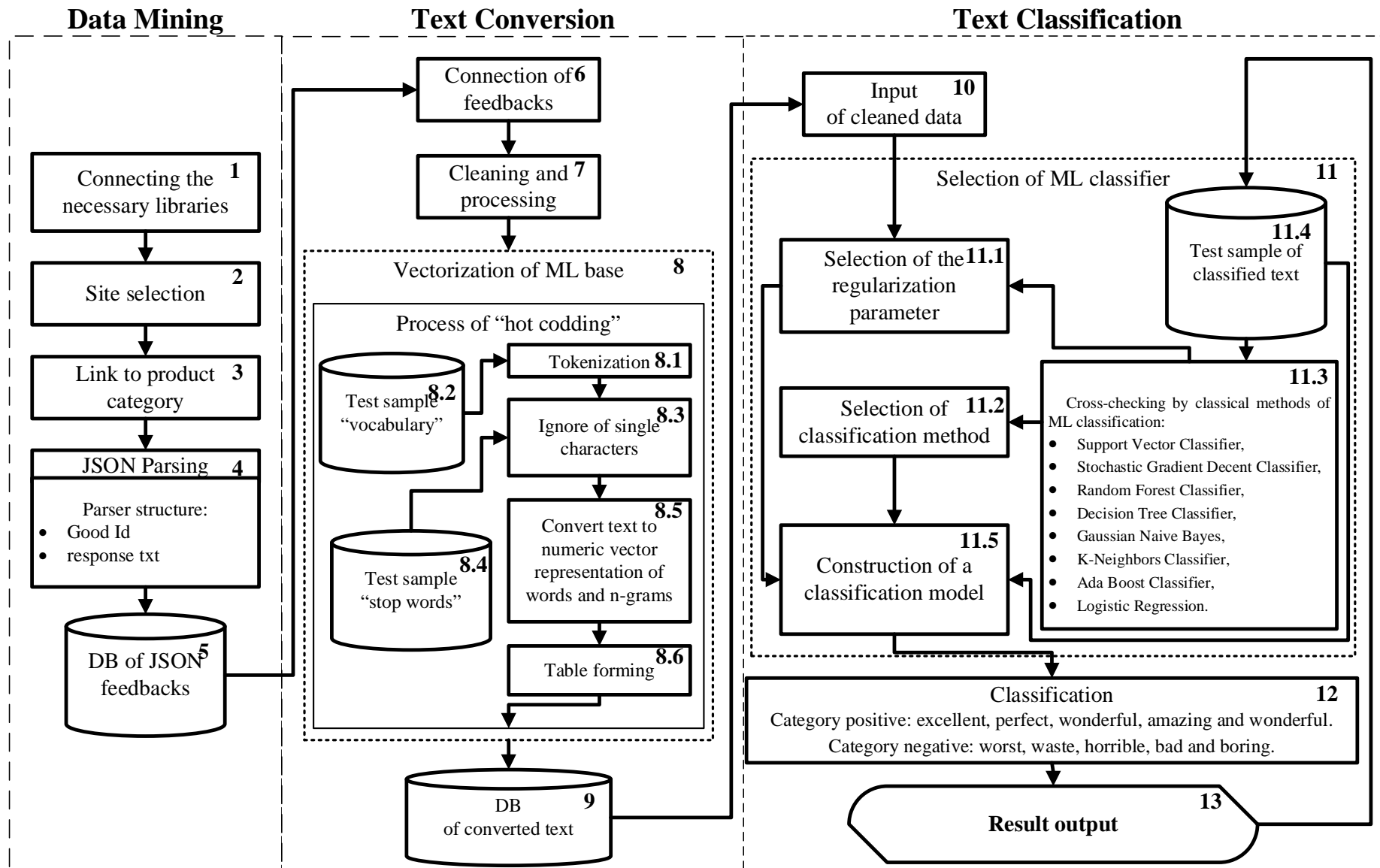
Data Mining

Connecting the necessary libraries **1**

Site selection **2**

Link to product category **3**

JSON Parsing **4**

Parser structure:
- Good Id
- response txt

DB of JSON feedbacks **5**

Text Conversion

Connection of feedbacks **6**

Cleaning and processing **7**

Vectorization of ML base **8**

Process of "hot codding"

Tokenization 8.1

Test sample "vocabulary" 8.2

Ignore of single characters 8.3

Test sample "stop words" 8.4

Convert text to numeric vector representation of words and n-grams 8.5

Table forming 8.6

DB of converted text **9**

Text Classification

Input of cleaned data **10**

Selection of ML classifier **11**

Test sample of classified text 11.4

Selection of the regularization parameter 11.1

Selection of classification method 11.2

Cross-checking by classical methods of ML classification: 11.3
- Support Vector Classifier,
- Stochastic Gradient Decent Classifier,
- Random Forest Classifier,
- Decision Tree Classifier,
- Gaussian Naive Bayes,
- K-Neighbors Classifier,
- Ada Boost Classifier,
- Logistic Regression.

Construction of a classification model 11.5

Classification 12
Category positive: excellent, perfect, wonderful, amazing and wonderful.
Category negative: worst, waste, horrible, bad and boring.

Result output **13**

Figure 1: Structure of the intelligent method of a competitive product choosing based on the emotional feedbacks coloring

Step 8.1. Tokenization (Block 8.1) is required to replace a confidential data item with a non-confidential equivalent called a token that has no independent meaning / significance for external or internal use. Sample, which evaluates the quality of the constructed model. The quality assessment made from the test sample can be used to select the best model (Block 8.2).

Step 8.2. The process of ignoring single characters that interfere with the program (Block 8.3) is carried out on the basis of a ready set of "stop words" (Block 8.4). Stop words are very common words such as "if", "but", "we", "he", "she" and "they". Seller can usually delete these words without changing the text semantics (but not always), improving the model performance.

Step 8.3. In this step, the text is converted into numerical vectors and add n-grams (Block 8.5). N-grams are simply a sequence of n elements (sounds, syllables, words or symbols) that go in a row in a text. If seller divide the text into several small fragments represented by N-grams, they are easy to compare with each other and thus obtain a degree of similarity of the analyzed documents. N-grams are often used successfully to categorize text and language. In addition, they can be used to create functions that allow seller to gain knowledge from text data.

Step 8.4. Forming an array of vectorized text values for further storage in the database (Block 8.6).

Step 9. Forming a database of the converted text to select the classifier model (Block 9).

Step 10. Enter the cleared data (Block 10).

Step 11. The next step is to select the most optimal classifier by machine learning algorithms (Block 11).

Step 11.1. Search for the best regularization parameter (Block 11.1). The regularization parameter reduces retraining, which reduces the variance of the estimated regression parameters. Seller need to use it. Cross-checking by eight classical classification methods: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression (Block 11.3) [50, 52-54]. A ready-made training sample (Block 11.4) was used to check the algorithms.

Step 11.2. Based on the obtained results, the best classification method is chosen (Block 11.2) of the classification with the highest regularization parameters (Block 11.1) and form a classification model (Block 11.5) based on the test sample (Block 11.4).

Step 12. Next is the data classification (Block 12) for positive and negative comments. Namely, positive feedbacks are divided into categories: excellent, perfect, wonderful, amazing and wonderful. Negatives are divided into categories: worst, waste, horrible, bad and boring.

Step 13. The last step is to derive the result on the basis of which seller can decide on a profitable investment in a new product.

The results of experimental studies confirmed the correctness of the developed method, which will be described in more detail in the camera-ready and the presentation.

4. Experimental results and Discussion

Python was chosen to conduct an intelligent method of choosing a competitive product based on the feedback emotional coloring. The following libraries were used: pandas, numpy, train_test_split, SVC, GridSearchCV, SGDClassifier, RandomForestClassifier, DecisionTreeClassifier, GaussianNB KNeighborsClassifier, AdaBoostClassifier, LogisticRegression.

4477 feedbacks from the Rozetka site were used as input [51]. Feedbacks are collected by parsing and formed into a JSON file, which presents the product id and feedback on it.

Next, the data was cleaned for the presence of characters that do not affect the content of the text: ".", ";", ":", "!", "", "?", ",", """", "()", "[]". The next step is removing the "stop words", which are in the Russian language, because most feedbacks are in Russian, further research will consider the possibility of feedback distributing by language, for better results.

After vectorization and lemmatization, the classification of the most optimal classifiers of machine learning algorithms was performed. Therefore, results will be cross-checked (Table 2) and evaluated based on 8 different methods: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression.

Table 2

The results of cross-evaluation

#	Method	Prediction assessment
1	SupportVectorClassifier	0.73
2	StochasticGradientDecentC	0.74
3	RandomForestClassifier	0.78
4	DecisionTreeClassifier	0.67
5	GaussianNB	0.77
6	KNeighborsClassifier	0.76
7	AdaBoostClassifier	0.76
8	LogisticRegression	0.73

Table 2 shows that all methods showed poor results, due to the fact that the feedbacks include both Ukrainian and Russian languages. However, the best is RandomForestClassifier, with a forecast score of 0.78, which is a valid score for evaluation. Probably, the model accuracy can be increased by increasing the dataset size, as the four thousand dataset is quite modest. Also, it would be possible to reduce the problem to a binary classification of feedbacks to positive and negative, which would also increase accuracy.

Next, the feedbacks will be arranged into positive and negative based on the classification method RandomForestClassifier (Fig. 4). Arranging is based on the inverted n-gram index and relative to the selected product with id – 122360970 (one of the most popular headphones).

Table 3

Arranged feedbacks by product with id – 122360970

ID	poswords	pos_importance	negwords	neg_importance
122360970	top	22.57	the worst	-0.77
122360970	good	1.82	horror	-0.76
122360970	super	0.85	I do not recommend	-0.76
122360970	cool	0.71	badly	-0.69
122360970	best	0.71	nonsense	-0.62
122360970	super-duper	0.71	broke	-0.62
122360970	cool	0.58	bad	-0.58
122360970	best	0.56	stopped working	-0.56
122360970	excellent	0.55	negative	-0.56
122360970	reliable	0.53	disappointed	-0.54

When arranging (see table 3), 10 positive and negative words were separated relative to the n-gram indices. Accordingly, for the product with id – 122360970, the following positive words with the largest n-gram index were identified: "top" (2.57) and "good" (1.82). Also, negative words with the lowest n-gram index – "worst" (-0.77), "horror" (-0.76). From the numerical characteristics of n-grams, show that the positive words in the feedbacks more than negative, respectively, we can conclude that the product with id – 122360970, is of good quality and worth choosing for resale.

Therefore, the developed method differs from analogues (see table 1) in that it will allow to parse the relevant data from the target site, and the classification is tested by eight classical classification methods, namely Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression. That gives the chance to make administrative decisions concerning profitable investment in the new goods.

5. Conclusions

The intelligent method of a competitive product choosing based on the emotional feedbacks coloring, based on which the seller can make management decisions regarding profitable investments in a new product, and thus reduce the risks of non-profit sales. Also, the proposed method reduces the time spent searching for popular and quality products based on user feedback.

The method was implemented on feedbacks (4477 feedbacks) from the Rozetka website. Feedback is collected by parsing and formed into a JSON file and cleaned of unnecessary characters and "stop words". After vectorization, lemmatization and classification of the most optimal classifiers of machine learning algorithms: Support Vector Classifier, Stochastic Gradient Decent Classifier, Random Forest Classifier, Decision Tree Classifier, Gaussian Naive Bayes, K-Neighbors Classifier, Ada Boost Classifier, Logistic Regression. All methods showed poor result, due to the fact that the feedbacks include both Ukrainian and Russian feedbacks. However, the best is RandomForestClassifier, with a forecast score of 0.78. Next, the feedbacks were sorted into positive and negative based on the RandomForestClassifier classification. The words with the largest n-gram index are highlighted: positive ("top" (2.57) and "good" (1.82)) and negative ("worst" (-0.77), "horror" (-0.76)). The positive words in the feedbacks are more than the negative ones, based on the n-gram indices, respectively, it can be concluded that they are of good quality and worth choosing for resale.

Areas of further research include an in-depth study of the effectiveness of the developed method in expanding the range of goods and their geography, as well as a significant increase in the number of user feedbacks and the ability to distribute feedback by languages. In addition, it is worth trying to reduce the problem to a binary classification of feedbacks to positive and negative, which would also increase accuracy. Moreover, authors are going to explore ontology models [55] and Deep Learning [56] in domains above.

6. References

[1] K. Kaushik, R. Mishra, N. Rana, Y. K. Dwivedi, Exploring reviews and review sequences on e-commerce platform: A study of helpful reviews on Amazon.in. Journal of Retailing and Consumer Services, 45 (2018) 21–32. https://doi:10.1016/j.jretconser.2018.08.002

[2] S. Saumya, J.P. Singh, & Y.K. Dwivedi, Predicting the helpfulness score of online reviews using convolutional neural network. Soft Comput 24 (2020) 10989–11005. https://doi.org/10.1007/s00500-019-03851-5

[3] Y. Zhang, Z. Lin, Predicting the helpfulness of online product reviews: A multilingual approach. Electronic Commerce Research and Applications, 27 (2018) 1–10. doi: 10.1016/j.elerap.2017.10.008

[4] J. Wu, Y. Wu, J. Sun, & Z. Yang, User reviews and uncertainty assessment: A two stage model of consumers' willingness-to-pay in online markets. Decision Support Systems, 55.1 (2013) 175-185.

[5] J. P. Singh, S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, & P. K. Roy, Predicting the "helpfulness" of online consumer reviews. Journal of Business Research, 70 (2017) 346-355.

[6] L. Celardo, & M. G. Everett, Network text analysis: A two-way classification approach. International Journal of Information Management, 51 (2020) 102009.

[7] M. Syamala, & N. J. Nalini, A Filter Based Improved Decision Tree Sentiment Classification Model for Real-Time Amazon Product Review Data. International Journal of Intelligent Engineering and Systems, 13.1 (2020) 191-202.

[8] M. A. Mirtalaie, & O. K. Hussain, Sentiment aggregation of targeted features by capturing their dependencies: Making sense from customer reviews. International Journal of Information Management, 53 (2020) 102097.

[9] Y. Liu, C. Jiang, & H. Zhao, Using contextual features and multi-view ensemble learning in product defect identification from online discussion forums. Decision Support Systems, 105 (2018) 1-12.

[10] B. Bansal, S. Srivastava, Hybrid attribute based sentiment classification of online reviews for consumer intelligence. Appl Intell, 49 (2019) 137–149. https://doi.org/10.1007/s10489-018-1299-7

[11] C. Guo, Z. Du, & X. Kou, Products Ranking Through Aspect-Based Sentiment Analysis of Online Heterogeneous Reviews. J. Syst. Sci. Syst. Eng. 27 (2018) 542–558. https://doi.org/10.1007/s11518-018-5388-2

[12] J.-W. Bi, Y. Liu, Z.-P. Fan, Representing sentiment analysis results of online reviews using interval type-2 fuzzy numbers and its application to product ranking. Information Sciences, 504 (2019) 293–307. doi:10.1016/j.ins.2019.07.025

[13] L. Zheng, Z. He, & S. He, A novel probabilistic graphic model to detect product defects from social media data. Decision Support Systems, 137 (2020) 113369. doi:10.1016/j.dss.2020.113369

[14] Y. Liu, C. Jiang, & H. Zhao, Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media. Decision Support Systems, 123 (2019) 113079. https://doi.org/10.1016/j.dss.2019.113079

[15] S. Khedkar, S. Shinde, Ensemble Classifier for Praise or Complaint Classification and Visualization from Big Data. In: Dey N., Mahalle P., Shafi P., Kimabahune V., Hassanien A. (eds) Internet of Things, Smart Computing and Technology: A Roadmap Ahead. Studies in Systems, Decision and Control, vol. 266, (2020) 97-118. https://doi.org/10.1007/978-3-030-39047-1_5

[16] M. Savargiv, B. Masoumi, & M.R. Keyvanpour, A new ensemble learning method based on learning automata. J Ambient Intell Human Comput. (2020) https://doi.org/10.1007/s12652-020-01882-7

[17] S. Khedkar, S. Shinde, Deep Learning-Based Approach to Classify Praises or Complaints from Customer Reviews. In: Bhalla S., Kwan P., Bedekar M., Phalnikar R., Sirsikar S. (eds) Proceeding of International Conference on Computational Science and Applications. Algorithms for Intelligent Systems. Springer, Singapore, (2020) 391-402. https://doi.org/10.1007/978-981-15-0790-8_38

[18] Z.-P. Fan, G.-M. Li, Y. Liu, Processes and methods of information fusion for ranking products based on online reviews: An overview. Information Fusion, (2020) 1-27. doi:10.1016/j.inffus.2020.02.007.

[19] S. Saumya, J. P. Singh, A. M. Baabdullah, N. P. Rana, & Y. K. Dwivedi, Ranking online consumer reviews. Electronic Commerce Research and Applications, 29 (2018) 78-89.

[20] H. C. Lee, H. C. Rim, & D. G. Lee, Learning to rank products based on online product reviews using a hierarchical deep neural network. Electronic Commerce Research and Applications, 36 (2019) 100874.

[21] C. Vo, D. Duong, D. Nguyen, and T. Cao, From Helpfulness Prediction to Helpful Review Retrieval for Online Product Reviews. In Proceedings of the Ninth International Symposium on Information and Communication Technology (SoICT 2018). Association for Computing Machinery, New York, NY, USA, (2018) 38–45. https://doi.org/10.1145/3287921.3287931

[22] J. Du, J. Rong, S. Michalska, H. Wang, Y. Zhang, Feature selection for helpfulness prediction of online product reviews: An empirical study. PLoS ONE, 14.12 (2019) e0226902. https://doi.org/10.1371/journal.pone.0226902

[23] M. Bilal, M. Marjani, I. A. T. Hashem, A. M. Abdullahi, M. Tayyab and A. Gani, Predicting Helpfulness of Crowd-Sourced Reviews: A Survey. In Proceedings of the 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), Karachi, Pakistan, (2019) pp. 1-8, doi: 10.1109/MACS48846.2019.9024814.

[24] X. Luo, & Y. Yi, Topic-Specific Emotion Mining Model for Online Comments. Future Internet, 11.3 (2019) 79.

[25] S. Mitra, & M. Jenamani, SentiCon: A Concept Based Feature Set For Sentiment Analysis. In Proceedings of the 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS) (2018) 246-250.

[26] Z. Zeng, Z. Zhou, and X. Mu, User review helpfulness assessment based on sentiment analysis, The Electronic Library, 38.2 (2020) 337-351. https://doi.org/10.1108/EL-08-2019-0200

[27] F. Fouladfar, M. N. Dehkordi and M. E. Basiri, Predicting the Helpfulness Score of Product Reviews Using an Evidential Score Fusion Method, IEEE Access, 8 (2020) 82662-82687. doi: 10.1109/ACCESS.2020.2988872.

[28] M. E. Basiri, & S. Habibi, Review Helpfulness Prediction Using Convolutional Neural Networks and Gated Recurrent Units. Proceedings of the 2020 6th International Conference on Web Research (ICWR) (2020) 191-196. doi: 10.1109/icwr49608.2020.9122297

[29] L. Kong, C. Li, J. Ge, V. Ng, & B. Luo, Predicting Product Review Helpfulness A Hybrid Method. IEEE Transactions on Services Computing, (2020). doi: 10.1109/TSC.2020.3041095.

[30] H. Dong, Y. Hou, M. Hao, J. Wang, & S. Li, Method for Ranking the Helpfulness of Online Reviews Based on SO-ILES TODIM. IEEE Access, 9 (2020) 1723-1736. Doi: 10.1109/ACCESS.2020.3040151

[31] M.S.I. Malik, A. Hussain, Helpfulness of product reviews as a function of discrete positive and negative emotions. Computers in Human Behavior, 73 (2017) 290–302. doi: 10.1016/j.chb.2017.03.053

[32] L. Li, T.T. Goh, & D. Jin, How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. Neural Comput & Applic, 32 (2020) 4387–4415. https://doi.org/10.1007/s00521-018-3865-7

[33] S. P. Eslami, M. Ghasemaghaei, & K. Hassanein, Which online reviews do consumers find most helpful? A multi-method investigation. Decision Support Systems, 113 (2018) 32-42.

[34] U. Chakraborty, and S. Bhat, Credibility of online reviews and its impact on brand image, Management Research Review, 41.1 (2018) 148-164. https://doi.org/10.1108/MRR-06-2017-0173

[35] R. Liang, Jq. Wang, A Linguistic Intuitionistic Cloud Decision Support Model with Sentiment Analysis for Product Selection in E-commerce. Int. J. Fuzzy Syst., 21 (2019) 963–977. https://doi.org/10.1007/s40815-019-00606-0

[36] B. Samal, A. K. Behera, & M. Panda, Performance analysis of supervised machine learning techniques for sentiment analysis. On Proceedings of the 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS) (2017) 128-133. doi:10.1109/ssps.2017.8071579

[37] X. Sun, M. Han, J. Feng, Helpfulness of online reviews: Examining review informativeness and classification thresholds by search products and experience products. Decision Support Systems, (2019) 113099. doi: 10.1016/j.dss.2019.113099

[38] M. Akbarabadi, & M. Hosseini, Predicting the helpfulness of online customer reviews: The role of title features. International Journal of Market Research, 62.3 (2020) 272-287. https://doi.org/10.1177/1470785318819979

[39] P. Wu, X. Li, S. Shen, & D. He, Social media opinion summarization using emotion cognition and convolutional neural networks. International Journal of Information Management, 51 (2020) 101978.

[40] M. Arif, U. Qamar, F.H. Khan, S. Bashir, A Survey of Customer Review Helpfulness Prediction Techniques. In: Arai K., Kapoor S., Bhatia R. (eds) Intelligent Systems and Applications. IntelliSys 2018. Advances in Intelligent Systems and Computing, vol 868. Springer, Cham, (2019) 215-226. https://doi.org/10.1007/978-3-030-01054-6_15

[41] E. Kauffmann, J. Peral, D. Gil, A. Ferrández, R. Sellers, H. Mora, Managing Marketing Decision-Making with Sentiment Analysis: An Evaluation of the Main Product Features Using Text Data Mining. Sustainability, 11.15 (2019) 4235. doi:10.3390/su11154235

[42] S. Saumya, J. Kumar, J.P. Singh, Genre Fraction Detection of a Movie Using Text Mining. In: Chaki R., Cortesi A., Saeed K., Chaki N. (eds) Advanced Computing and Systems for Security. Advances in Intelligent Systems and Computing, vol 666 (2018) 167-177. Springer, Singapore. https://doi.org/10.1007/978-981-10-8180-4_11.

[43] J.P. Singh, N.P. Rana, W. Alkhowaiter, Sentiment Analysis of Products' Reviews Containing English and Hindi Texts. In: Janssen M. et al. (eds) Open and Big Data Management and Innovation. I3E 2015. Lecture Notes in Computer Science, vol 9373. Springer, Cham (2015) 416-422. https://doi.org/10.1007/978-3-319-25013-7_33

[44] V. Kumar, S. Pasari, V. P. Patil, S. Seniaray, Machine Learning based Language Modelling of Code Switched Data. In Proceedings of the 2020 International Conference on Electronics and

Sustainable Communication Systems (ICESC) (2020) 552–557. doi:10.1109/ICESC48915.2020.9155695

[45] V. Battu, V. Batchu, R. R. R. Gangula, M. M. K. R. Dakannagari, & R. Mamidi, Predicting the Genre and Rating of a Movie Based on its Synopsis. In PACLIC. (2018) 52-62.

[46] J. Wehrmann, W. E. Becker, R. C. Barros, A multi-task neural network for multilingual sentiment classification and language detection on Twitter. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing – SAC'18 (2018) 1805–1812. doi:10.1145/3167132.3167325

[47] W.-T. Chu and H.-J. Guo, Movie Genre Classification based on Poster Images with Deep Neural Networks. In Proceedings of the Workshop on Multimodal Understanding of Social, Affective and Subjective Attributes (MUSA2 '17). Association for Computing Machinery, New York, NY, USA, (2017) 39–45. https://doi.org/10.1145/3132515.3132516

[48] M.S.I. Malik, Predicting users' review helpfulness: the role of significant review and reviewer characteristics. Soft Comput, 24 (2020) 13913–13928. https://doi.org/10.1007/s00500-020-04767-1

[49] S. Zhou, & B. Guo, The order effect on online review helpfulness: A social influence perspective. Decision Support Systems, 93 (2017) 77-87.

[50] H. Lipyanina, V. Maksymovych, A. Sachenko, T. Lendyuk, A. Fomenko, I. Kit, Assessing the Investment Risk of Virtual IT Company Based on Machine Learning. In: Babichev S., Peleshko D., Vynokurova O. (eds) Data Stream Mining & Processing. DSMP 2020. Communications in Computer and Information Science, vol 1158 (2020) 167-187 Springer, Cham. https://doi.org/10.1007/978-3-030-61656-4_11

[51] Internet-Shop "Rozetka™" – https://rozetka.com.ua/ua/

[52] C. Wang, N. Shakhovska, A. Sachenko, M. Komar, A New Approach for Missing Data Imputation in Big Data Interface. Information Technology and Control, 49.4, (2020) 541-555. https://doi.org/10.5755/j01.itc.49.4.27386.

[53] H. Lipyanina, S. Sachenko, T. Lendyuk, A. Sachenko, Targeting Model of HEI Video Marketing based on Classification Tree. In: Proceedings of the 16th International Conference on ICT in Education, Research and Industrial Applications. Kharkiv, Ukraine, October 6-10, 2020, pp. 487-498. ISSN: 1613-0073. http://ceur-ws.org/Vol-2732/20200487.pdf.

[54] H. Lipyanina, A. Sachenko, T. Lendyuk, S. Nadvynychny, S. Grodskyi, Decision Tree Based Targeting Model of Customer Interaction with Business Page. In: Proceedings of the third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), CEUR Workshop Proceedings, vol. 2608, 2020, pp. 1001-1012. ISSN: 1613-0073. Electronic copy at: http://ceur-ws.org/Vol-2608/paper75.pdf.

[55] C. Shu, D. Dosyn, V. Lytvyn, V. Vysotska, A. Sachenko, S. Jun, Building of the Predicate Recognition System for the NLP Ontology Learning Module. In: Proceedings of the 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2019), 18-21 September, 2019, Metz, France, vol. 2, (2019) 802-808. https://doi.org/10.1109/IDAACS.2019.8924410.

[56] V. Golovko, S. Bezobrazov, A. Kroshchanka, A. Sachenko, M. Komar, A. Karachka, Convolutional Neural Network Based Solar Photovoltaic Panel Detection in Satellite Photos. In: Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2017), September 21-23, 2017, Bucharest, Romania, (2017) 14-19. https://doi.org/10.1109/IDAACS.2017.8094501.