

Customer Feedback Text Analysis for Online Stores Reviews in Bulgarian

Tsvetanka L. Georgieva-Trifonova, Milena E. Stefanova, and Stefan D. Kalchev

Abstract—In the present paper, applying a text analysis of customer feedback received from online stores reviews in Bulgarian is considered. For this purpose, a freely available dataset with customer reviews in Bulgarian is created. Besides, an approach to automatically association of the reviews with predefined categories reflecting the customer satisfaction is proposed. An enrichment of the vector space model for document presentations based on computing a pointwise mutual information measure of the terms in regard to the categories is performed. Experiments are conducted on the collected dataset by using different classifiers. The results of measures to assess the classification performance are provided.

Index Terms—text classification, customer reviews, pointwise mutual information

I. INTRODUCTION

THE growth of e-commerce leads to a significant increment in the volume of the user reviews about the offered products, as well as about the online stores themselves on the Web in the form of free text. Collecting and analyzing the customer feedback is important, because it allows discovering useful information about the various ways of improving and adapting the customer preferences. On the other hand, the large number of the user reviews complicates producers or managers, since it is impossible for them to read and analyze all reviews that are unstructured type of data. This process can be significantly supported by applying methods for the automatic classification of collected user reviews in order to associate them into predefined categories such as positive, negative, neutral or mixed, suggestions, request for information, etc.

The purpose of the present research is applying text analysis on customer feedback received from the user reviews in Bulgarian language. Our study indicates a lack of a dataset with user reviews in Bulgarian. This necessitated the creation of a new dataset with customer feedback on online stores.

Manuscript received May 22, 2018; revised September 1, 2018. This work was supported in part by the Department of Computer Systems and Technologies, Faculty of Mathematics and Informatics at “St. Cyril and St. Methodius” University of Veliko Tarnovo.

T. L. Georgieva-Trifonova is with the Department of Information Technologies, University of Veliko Tarnovo, Bulgaria (corresponding author; e-mail: cv.georgieva@uni-vt.bg; cvetygeorg@gmail.com).

M. E. Stefanova is with the Department of Computer Systems and Technologies, University of Veliko Tarnovo, Bulgaria, (e-mail: m_stefanova@abv.bg).

St. D. Kalchev is with the Department of Computer Systems and Technologies, University of Veliko Tarnovo, Bulgaria, (e-mail: stkalchev@mail.bg).

It is freely accessible [1] and it is available in different file formats – xls, txt, xml (with specified DTD (Document Type Definition) and XML Schema), rdf. A data model is defined based on the vector space model (VSM), enriched with pointwise mutual information (PMI). Text classification is applied through different classifiers and the results on the classification performance are summarized and analyzed.

The rest of the paper is organized as follows. In section 2, existing approaches to text analysis of consumer feedback are examined. In Section 3, a dataset with customer reviews in Bulgarian is described, as well as the proposed approach to analyzing feedback from users. The results from measures for evaluation of the classification performance are presented and discussed.

II. RELATED WORKS ON EXISTING APPROACHES TO TEXT ANALYSIS OF USER REVIEWS

In [2], an algorithm for classifying user reviews in two categories (recommended, not recommended) is proposed. For this purpose, a semantic orientation of the review phrases is defined – positive or negative, based on calculated pointwise mutual information between the word pairs.

In [3], a topic model for the text analysis of customer reviews, taking into account the numerical rating given by users is proposed and is tested on datasets for hotels and restaurants. It is built on the basis of LDA (*Latent Dirichlet Allocation*) model that is extended to relate the topic probabilities to the product ratings.

An analysis of customer opinion is applied in [4] by performing the text classification through Naive Bayes and K-NN classifiers into three categories (good, bad, mixed) by using RapidMiner (<http://rapidminer.com>) without modifying the data model.

Satisfying customer personalized requirements by partitioning service modules is examined in [5]. For this purpose, a quantitative method is proposed. This method is based on Kano model, but successfully overcomes both shortcomings – the subjectivity and ambiguity. The Kano model is applied to analyze customer requirements [6] based on collected text customer reviews classified into two categories.

For the purposes of the present work, we have examined the studies analyzing customer feedback by applying methods for sentiment analysis that are reviewed and classified based on used techniques in [7].

In [8], the aspect identification is examined by sentence

clustering using a bag of nouns rather than a bag of words. An approach to sentiment analysis is proposed, which is based on text classification into three categories (positive, neutral, negative).

A system that performs the classification of customer reviews about hotels by sentiment analysis is presented in [9]. An approach to classifying customer feedback into three categories (good, bad, neutral) on the basis of existing domain-specific corpus by applying the lexicon-based sentiment analysis is proposed. The lexicon generation consists of extracting the words in the training dataset and identifying them as parts of speech. The association of words into categories and setting the weights used for classification is performed by taking into account the occurrence frequency of words in reviews with specific user-defined numeric ratings (out of five possible).

In [10], the advantages of the rule-based classification algorithms for the complaint detection are analyzed by using RapidMiner.

Clustering of documents that contain customer reviews written in Turkish is implemented in [11]. The word weights are obtained by calculating TF-IDF (*Term Frequency – Inverse Document Frequency*) – numerical characteristics defined by the frequency of word occurrence in the document and the inverse document frequency. The cosine similarity is selected as a measure of similarity between documents.

In [12], the customer reviews received from feedback about Microsoft Office's users, written in four languages (English, French, Spanish and Japanese) are analyzed by applying classification into the following categories: comment, request, bug, complaint, meaningless, undetermined. For this purpose, a multinomial naive Bayes approach is used for polynomial classification and its combination with a score obtained by sentiment analysis.

The classification of user reviews in the categories (positive and negative) is considered in [13] by identifying similar semantic features from different domains using Word2Vec.

Besides, the existing related works about texts in Bulgarian have studied. The research in [14] is devoted to the emotional meaning of the adjectives in Bulgarian language and their classification in positivity and emotional axes built by groups of manually selected and relatively commonly used words.

A sentiment analysis of movie reviews written in Bulgarian language and associated with an 11-scale star rating is performed in [15]. For this purpose, a dataset with movie reviews in Bulgarian language is created and a sentiment polarity lexicon of words extracted from the collected feedbacks is generated. The sentiment polarity score of each word is calculated as the difference between the pointwise mutual information of the word in regard to

positive and negative category. The score lexicon of the user feedback is obtained by summing the scores of words in the text. A 3-way, 5-way and 11-way classification and a regression analysis are implemented; each review is represented by the binary weights of words and emoticons, calculated sentiment score and contextual characteristics.

The pointwise mutual information used in the present study is applicable in language researches, since it provides ranking n -grams by comparison of the frequency of the multiword expressions candidate to the frequency of the components of the multiword expressions. With this measure, automatically identification of bi-gram multiword expressions in parallel Latvian and Lithuanian corpora is performed [16].

The main purpose of the present paper is implementing the text analysis of customer reviews about online stores written in Bulgarian. A dataset is created and a text classification of user reviews into predefined categories – compliments, complaints, mixed, suggestions is applied. The presentation of text reviews is performed by the vector space model enriched based on pointwise mutual information of the terms in regard to each of the four categories. The results from the measures for validity and reliability evaluation of the classification model are summarized.

III. APPLYING CUSTOMER FEEDBACK TEXT ANALYSIS FOR ONLINE STORES REVIEWS IN BULGARIAN

The implementation of customer feedback text analysis for online stores reviews in Bulgarian includes the following tasks, which are described in detail in separate subsections: creation of a dataset; preliminary process of the collected text data; definition of the data model; application of text classification.

A. Creating a dataset for customer feedback analysis in Bulgarian

The lack of a dataset about customer reviews in the Bulgarian language necessitates the creation of the dataset *Customer_feedback_bg*.

TABLE I
SUMMARY INFORMATION ABOUT THE NUMBER OF REVIEWS FOR THE FIVE ONLINE STORES WITH THE MOST REVIEWS IN THE DATASET *CUSTOMER_FEEDBACK_BG*

Online store	Number of reviews in <i>Customer_feedback_bg</i>
sapirshop.com	192
bgtelefon.com	164
smartfoni.bg	56
olx.bg	38
sms.bg	32

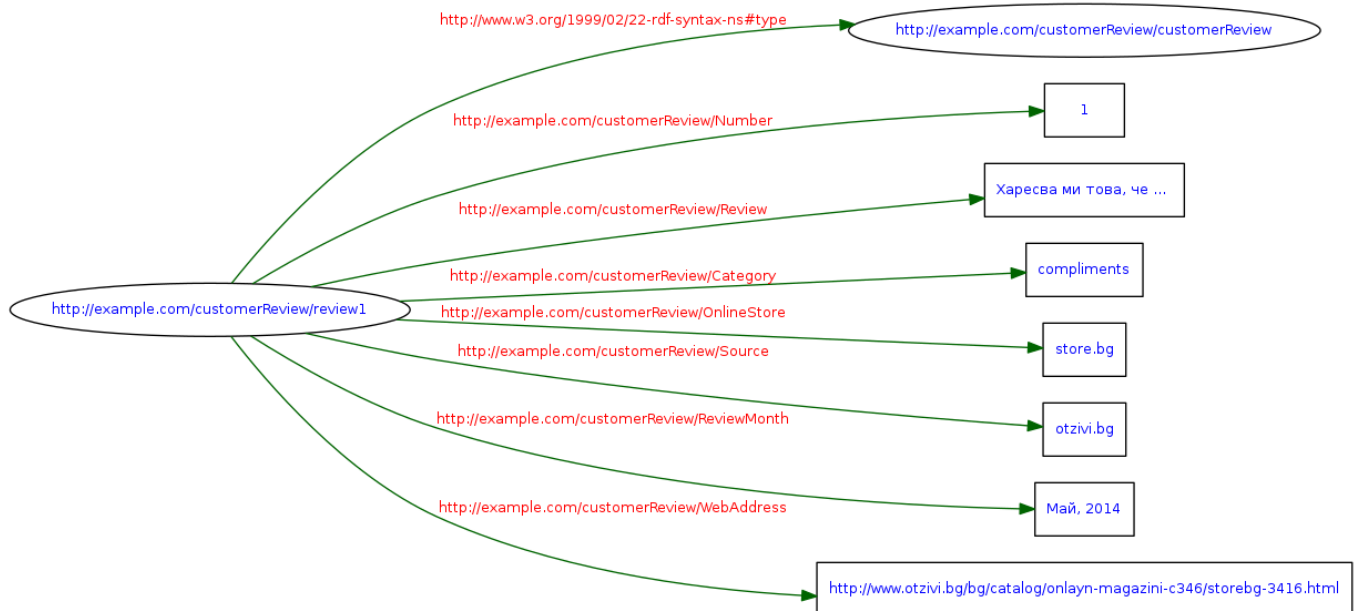


Fig. 2. Graphical representation of an instance, stored in the dataset *Customer_feedback_bg* in RDF format.

The data are retrieved from *otzivi.bg* and *pazaruvaj.com*, and represent user reviews in Bulgarian language about 87 online stores. 906 customer reviews were collected in free text and manually associated with the following categories: compliments, complaints, mixed, suggestions. Table I and Figure 1 contain a summary of the number of reviews in the dataset *Customer_feedback_bg* by online stores, Table II – by categories.

TABLE II
SUMMARY INFORMATION ABOUT THE NUMBER OF REVIEWS BY CATEGORIES IN THE DATASET *CUSTOMER_FEEDBACK_BG*

Category	Number of reviews in <i>Customer_feedback_bg</i>
Compliments	540
Complaints	184
Mixed	138
Suggestions	44

In order to ensure free accessibility of the dataset *Customer_feedback_bg* for future research, it is published in Dataverse repository [1]. The dataset is available in different file formats:

- XLS format – Microsoft Excel Spreadsheet;
- TXT – text file with separator (.txt), as a separator is used the symbol Tab;
- XML (*eXtensible Markup Language*) – XML data with related document type definition in .dtd file and XML scheme in .xsd file;
- RDF (Resource Description Framework) in RDF/XML syntax.

A graphical representation of an instance, stored in the dataset *Customer_feedback_bg* in RDF format has the form shown in Figure 2. The visualization is done by online validator <http://www.w3.org/RDF/Validator/>.

B. Text preprocessing

The preliminary processing of the collected texts includes implementation of the following steps:

- Tokenization;
As a result of the tokenization, 46930 words are retrieved, of which 6098 are different.
- Stop words filtering;
Filtering the stop words which are prepositions, adverbs and other common words in Bulgarian language, as well as filtering according to the word length (less than 2) is performed.
- Stemming.
Stemming is implemented on the basis of a dictionary that is generated by applying the rules on extracted words in Bulgarian. The rules defined in [17] are used.

After preprocessing by stop words filtering and stemming, the number of remaining words is 23444, of which 2373 are different and are used to build a data model for the conducted experiments.

C. Definition of a data model

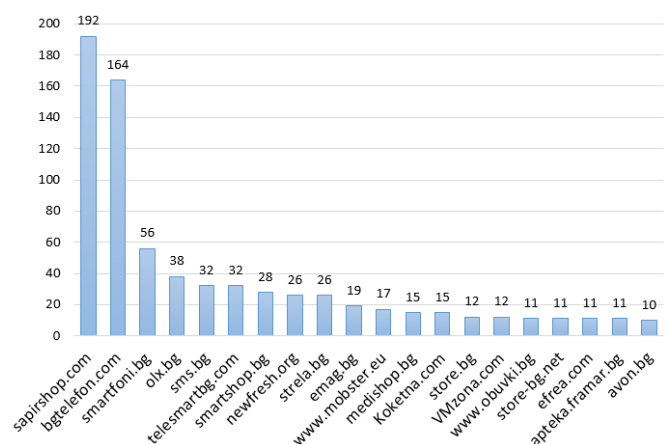


Fig. 1. Summary information about the number of reviews for 20 online stores with the most reviews in the dataset *Customer_feedback_bg*.

The representation of a document *d* in the vector space is defined as a vector of real numbers $w(d) = (w(d, term_1), \dots, w(d, term_m))$, where each component $w(d, term_j)$ is a word weight and is obtained on the base of frequency of occurrence of *term_j* in a document. The simplest vector

representation is Boolean, i.e. $w(d, term_j) \in \{0, 1\}$ and it indicates the presence or absence of the word $term_j$ in the document d . When using Boolean presentation, the importance of all words is the same. In order to improve the performance of text mining tasks, the words are associated with weights, reflecting the importance of words in a given document of the considered collection. One of the possible ways to calculate the term weight, is based on term occurrence frequency $tf(d, term)$ in document d .

Another commonly used approach assigning higher values of the weights to words that occur frequently in the relevant documents, but rarely in the entire collection of documents. For this purpose, the computation of the word weights, is based on the so-called TF-IDF measure, which is widely applicable and its efficiency is theoretically justified [18]. In particular, the weight $w(d, term)$ of $term$ in document d is calculated as the multiplication of its frequency $tf(d, term)$ and so-called *invert document frequency* (IDF) $idf(term)$, which describes the word in the collection of documents, i.e.

$$w(d, term) = tf(d, term) \cdot idf(term) \quad (1)$$

The invert document frequency is defined as $idf(term) = \log(n/n_{term})$, where n is a size of the document collection D and n_{term} is number of documents in D , that contain the word $term$.

In the present paper, the new approach to enrich the vector space model for documents presentation is proposed, which is based on the calculation of the pointwise mutual information of the words in regard to the categories.

Let denote the predefined categories with C_1, \dots, C_k . For each word $term_i$ we calculate the pointwise mutual information in regard to the categories $pmi(term_i, C_j)$ as follows ($i = 1, \dots, m; j = 1, \dots, k$):

$$pmi(term_i, C_j) = \log(p(term_i, C_j)/(p(term_i) \cdot p(C_j))),$$

where

- $p(term_i, C_j)$ is computed as the number of occurrences of $term_i$ in documents, which are associated with category C_j , divided by the total number of occurrences of all words in all documents;
- $p(term_i)$ is the number of occurrences of $term_i$, divided by the total number of occurrences of all words in all documents;
- $p(C_j)$ is calculated as the number of occurrences of the words of the documents associated with the category C_j , divided by the total number of occurrences of all words in all documents.

By this way, we find $m \times k$ matrix PMI:

$$PMI = \begin{bmatrix} & \mathbf{C_1} & \mathbf{C_2} & \dots & \mathbf{C_k} \\ \mathbf{term_1} & pmi_{11} & pmi_{12} & \dots & pmi_{1k} \\ \mathbf{term_2} & pmi_{21} & pmi_{22} & \dots & pmi_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{term_m} & pmi_{m1} & pmi_{m2} & \dots & pmi_{mk} \end{bmatrix}$$

The resulting matrix PMI is used for modification of the weights from a vector space model for $j = 1, \dots, k$ and for each document in collection as follows:

$$wp(d, C_j) = (w(d, term_1), \dots, w(d, term_m)) \cdot \begin{pmatrix} pmi_{1j} \\ \vdots \\ pmi_{mj} \end{pmatrix}, \quad (2)$$

where $w(d, term_1), \dots, w(d, term_m)$ are calculated in accordance with equality (1).

As a result, we receive PMI-enriched VSM model, which is applied for classification of customer reviews about online stores and its form is:

$$\begin{matrix} & \mathbf{C_1} & \mathbf{C_2} & \dots & \mathbf{C_k} \\ \mathbf{d_1} & wp_{11} & wp_{12} & \dots & wp_{1k} \\ \mathbf{d_2} & wp_{21} & wp_{22} & \dots & wp_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ \mathbf{d_n} & wp_{n1} & wp_{n2} & \dots & wp_{nk} \end{matrix}$$

The document presentation through k -dimensional vectors is used for execution of algorithms for text classification with purpose of association of each document, i.e. each user review with a category which is described in the next section.

From the way the proposed PMI-enriched model is constructed, it becomes clear that its building requires additional calculations compared to the model of the vector space, as it is necessary to retrieve the pointwise mutual information of the words with respect to each of the categories. In this regard, it is important to note that as a result a k -dimensional representation is obtained instead of the m -dimensional of each document. During the text classifying, the number of categories k (in the present case is 4) is significantly smaller than the number of words m , which leads to an increment in the performance of text classification algorithms on already built model.

D. Applying of algorithms for text classification through different classifiers

The application of algorithms for text classification through different classifiers is made by means of RapidMiner [19]. The classifiers used are:

- SVM (*Support Vector Machine*);
The application of support vector machines for text classification is proposed initially in [20]. For the experiments, presented in this research LibSVM implementation [21] is used, available in RapidMiner.
- K-NN (*K-Nearest Neighbors*);
K-NN [22] is based on a measure of distance or similarity between two documents as Euclidean distance or cosine measure and the selection of K (neighboring) documents, which are closest (or most similar) to a given document that is subject to classification, is performed. For K-NN, a new document d is associated with the category that contains the most adjacent to d documents.
- Naive Bayes;
Naive Bayes classifier represents probabilistic classifier, based on the application of the Bayes theorem and the assumption of attribute independence. This kind of classifiers are widely explored and applied for classifying text [23].
- H2O's Deep Learning;
H2O's Deep Learning [24] is based on multi-layer feed forward artificial neural network, which is being trained by using a back-propagation algorithm.
- Rule-based classifiers Ridor, Jrip, PART.
Rule-based classifiers are included in the conducted experiments using WEKA extension for RapidMiner.

The classification is performed for the text reviews collected in the created dataset *Customer_feedback_bg* by their presentation through the vector space model and through PMI-enriched model described in the previous section. Measures are used to assess the validity and reliability of the classification model. The calculations are made at different approaches to obtaining the term weights in VSM:

- Boolean weights (binary);
- Term frequencies (TF) in documents;
- Term frequency – inverse document frequency (TF-IDF), computed in accordance with equality (1).

Validity evaluation

Measures that are particularly useful for assessing the validity of classification models and classifiers are applied, providing quantified understanding of the error. The results of the following measures for the performance evaluation of the classification model [25] are calculated and summarized:

- Accuracy is defined by the ratio of the number of correctly classified documents to the total number of documents;
- The *F*-measure is defined as the mean harmonic value between the precision *P* and the recall *R*:

$$F = \frac{2 \cdot P \cdot R}{P + R},$$

where

TABLE III
F-MEASURE WHEN APPLYING VSM AND PMI-ENRICHED VSM FOR SVM, K-NN, NAÏVE BAYES, H2O'S DEEP LEARNING CLASSIFIERS

Classifier	Data model	Category			
		Compliments	Complaints	Mixed	Suggestions
SVM	VSM (binary)	75.94%	31.09%	6.45%	0.00%
	PMI-enriched VSM (binary)	92.79%	83.43%	84.38%	73.24%
	VSM (TF)	75.75%	32.23%	1.32%	0.00%
	PMI-enriched VSM (TF)	92.14%	81.27%	81.02%	73.24%
	VSM (TF-IDF)	75.75%	32.23%	1.32%	0.00%
	PMI-enriched VSM (TF-IDF)	94.31%	87.32%	87.06%	81.58%
K-NN	VSM (binary)	74.68%	32.34%	13.82%	0.00%
	PMI-enriched VSM (binary)	91.57%	79.15%	79.53%	78.95%
	VSM (TF)	75.62%	39.88%	17.09%	0.00%
	PMI-enriched VSM (TF)	90.99%	79.67%	77.47%	83.55%
	VSM (TF-IDF)	75.75%	39.50%	19.10%	0.00%
	PMI-enriched VSM (TF-IDF)	94.09%	87.32%	85.28%	81.58%
Naive Bayes	VSM (binary)	54.53%	39.78%	17.69%	8.44%
	PMI-enriched VSM (binary)	90.04%	73.45%	79.53%	75.95%
	VSM (TF)	56.68%	41.47%	17.17%	8.53%
	PMI-enriched VSM (TF)	93.29%	81.89%	81.48%	83.95%
	VSM (TF-IDF)	56.19%	41.96%	15.91%	8.53%
	PMI-enriched VSM (TF-IDF)	94.53%	87.60%	85.50%	82.92%
H2O's Deep Learning	VSM (binary)	81.04%	54.23%	23.43%	3.12%
	PMI-enriched VSM (binary)	92.29%	80.98%	83.65%	79.06%
	VSM (TF)	77.12%	52.27%	20.90%	5.80%
	PMI-enriched VSM (TF)	93.06%	82.04%	82.48%	82.50%
	VSM (TF-IDF)	76.78%	46.88%	22.78%	9.09%
	PMI-enriched VSM (TF-IDF)	94.61%	85.63%	86.99%	86.42%

TABLE IV
F-MEASURE WHEN APPLYING VSM AND PMI-ENRICHED VSM FOR RULE-BASED CLASSIFIERS RIDOR, JRIP, PART

Classifier	Data model	Category			
		Compliments	Complaints	Mixed	Suggestions
Ridor	VSM (binary)	72.91%	40.00%	6.71%	10.53%
	PMI-enriched VSM (binary)	92.30%	80.00%	78.26%	69.05%
	VSM (TF)	74.86%	45.23%	11.24%	5.13%
	PMI-enriched VSM (TF)	91.93%	79.56%	76.12%	78.16%
	VSM (TF-IDF)	74.79%	38.61%	15.96%	12.66%
	PMI-enriched VSM (TF-IDF)	93.62%	86.57%	83.15%	80.49%
Jrip	VSM (binary)	76.56%	39.01%	0.00%	0.00%
	PMI-enriched VSM (binary)	90.61%	80.32%	76.61%	65.88%
	VSM (TF)	76.92%	38.49%	0.00%	0.00%
	PMI-enriched VSM (TF)	92.58%	83.01%	78.79%	79.52%
	VSM (TF-IDF)	76.26%	36.03%	0.00%	12.00%
	PMI-enriched VSM (TF-IDF)	93.97%	86.74%	81.10%	77.65%
PART	VSM (binary)	72.09%	44.50%	23.35%	2.74%
	PMI-enriched VSM (binary)	92.21%	80.65%	80.44%	69.88%
	VSM (TF)	72.76%	46.81%	22.06%	2.86%
	PMI-enriched VSM (TF)	93.07%	81.92%	79.23%	80.00%
	VSM (TF-IDF)	75.78%	50.00%	19.33%	7.79%
	PMI-enriched VSM (TF-IDF)	94.20%	85.56%	83.21%	78.57%

- The precision P is calculated as the ratio of the number of correctly classified documents in a given category to the number of all documents classified in this category;
- The recall R is calculated as the ratio of the number of correctly classified documents from a given category to the number of all documents that are actually in this category.

Table III represents the results of the F -measure when applying the vector space model (VSM) and the PMI-enriched VSM for the classifiers SVM, K-NN, Naïve Bayes, H2O's Deep Learning. The obtained values of F -measure show that the quality of the classification of texts reviews remains above 73% for all classifiers and term weights, even for the category *Suggestions*, which are associated with the least number of reviews.

Table IV summarizes the results of the F -measure when applying vector space model (VSM) and the PMI-enriched VSM for the rule-based classifiers Ridor, Jrip, PART. For these classifiers, the minimum value of the F -measure (65.88%) with the proposed model is obtained again for the category *Suggestions*; the corresponding F -measure of the VSM (binary) for the same category and the same classifier is 0.

Figures 3 and 4 illustrate the results from precision computation and figures 5 and 6 – the results from recall computation when applying VSM and PMI-enriched VSM for different classifiers and term weights.

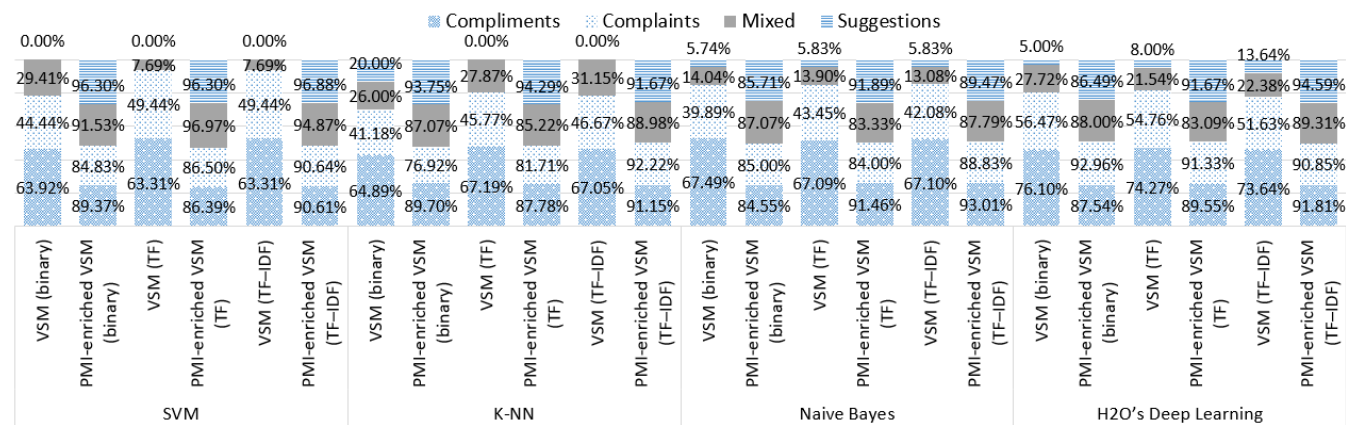


Fig. 3. Precision measure when applying VSM and PMI-enriched VSM for SVM, K-NN, Naïve Bayes, H2O's Deep Learning classifiers.

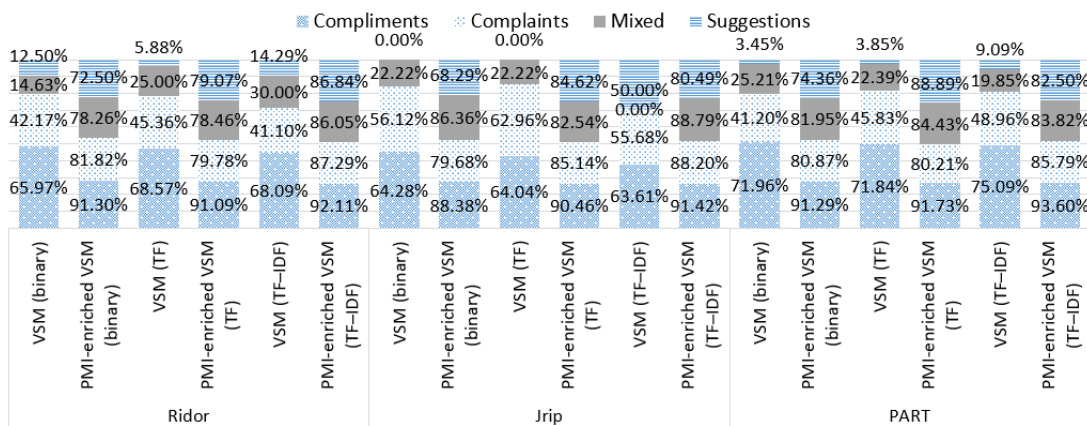


Fig. 4. Precision measure when applying VSM and PMI-enriched VSM for rule-based classifiers Ridor, Jrip, PART.

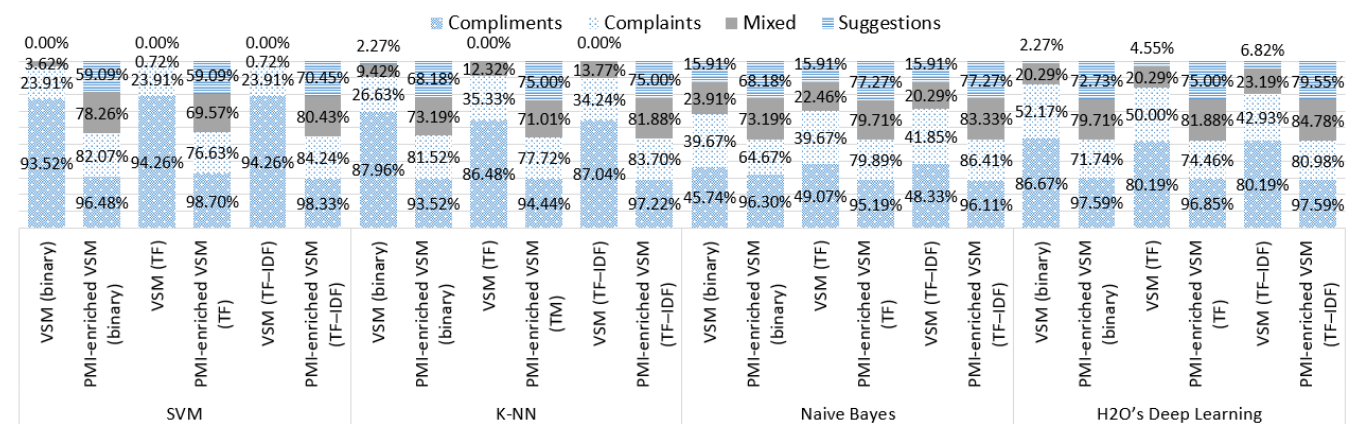


Fig. 5. Recall measure when applying VSM and PMI-enriched VSM for SVM, K-NN, Naïve Bayes, H2O's Deep Learning classifiers.

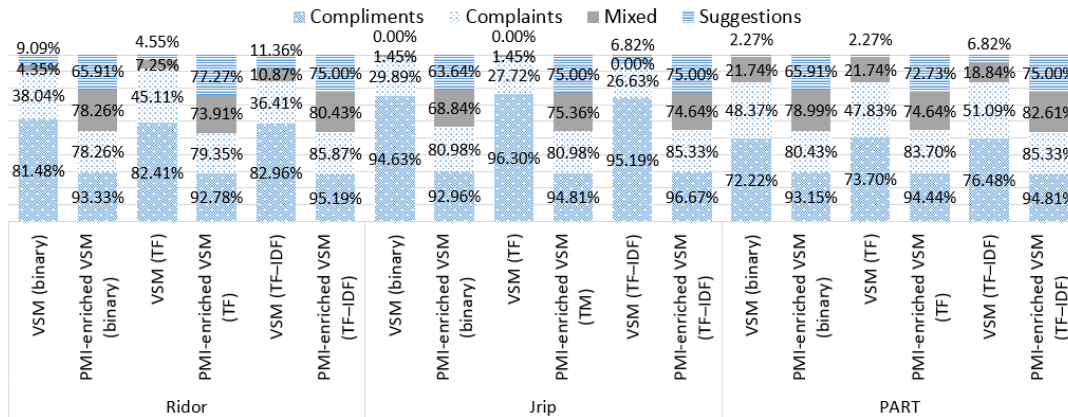


Fig. 6. Recall measure when applying VSM and PMI-enriched VSM for rule-based classifiers Ridor, Jrip, PART.

TABLE V
ACCURACY OF TEXT CLASSIFICATION WHEN APPLYING VSM AND PMI-ENRICHED VSM FOR THE RELEVANT CLASSIFIERS

	VSM (binary)	PMI-enriched VSM (binary)	VSM (TF)	PMI-enriched VSM (TF)	VSM (TF-IDF)	PMI-enriched VSM (TF-IDF)
SVM	61.13%	88.98%	61.13%	87.91%	61.13%	91.41%
K-NN	59.38%	86.78%	60.57%	86.57%	60.91%	91.09%
Naive Bayes	39.73%	85.02%	41.50%	88.87%	41.19%	91.28%
H2O's Deep Learning	65.46%	88.44%	61.23%	89.00%	60.32%	91.40%
Ridor	57.38%	86.66%	59.58%	86.41%	59.04%	90.06%
Jrip	62.70%	85.44%	63.24%	88.13%	62.49%	89.97%
PART	56.30%	87.08%	57.08%	88.18%	59.22%	90.07%

TABLE VI
MEASURES FOR RELIABILITY ASSESSMENT WHEN APPLYING VSM AND PMI-ENRICHED VSM FOR SVM, K-NN, NAÏVE BAYES, H2O'S DEEP LEARNING CLASSIFIERS

Classifier	Data model	Measure			
		Kappa	Pearson	Spearman	Kendall tau
SVM	VSM (binary)	0.135	0.175	0.167	0.158
	PMI-enriched VSM (binary)	0.792	0.802	0.818	0.802
	VSM (TF)	0.124	0.133	0.130	0.123
	PMI-enriched VSM (TF)	0.755	0.766	0.786	0.769
	VSM (TF-IDF)	0.124	0.133	0.130	0.123
	PMI-enriched VSM (TF-IDF)	0.835	0.838	0.854	0.842
K-NN	VSM (binary)	0.153	0.230	0.193	0.182
	PMI-enriched VSM (binary)	0.757	0.788	0.796	0.778
	VSM (TF)	0.207	0.252	0.240	0.226
	PMI-enriched VSM (TF)	0.747	0.780	0.776	0.760
	VSM (TF-IDF)	0.206	0.260	0.240	0.226
	PMI-enriched VSM (TF-IDF)	0.832	0.829	0.841	0.829
Naive Bayes	VSM (binary)	0.103	0.127	0.001	-0.001
	PMI-enriched VSM (binary)	0.706	0.762	0.759	0.741
	VSM (TF)	0.109	0.127	0.006	0.005
	PMI-enriched VSM (TF)	0.793	0.825	0.834	0.817
	VSM (TF-IDF)	0.110	0.122	-0.002	-0.004
	PMI-enriched VSM (TF-IDF)	0.842	0.847	0.859	0.846
H2O's Deep Learning	VSM (binary)	0.325	0.349	0.399	0.371
	PMI-enriched VSM (binary)	0.771	0.792	0.800	0.785
	VSM (TF)	0.285	0.334	0.347	0.324
	PMI-enriched VSM (TF)	0.785	0.812	0.820	0.806
	VSM (TF-IDF)	0.311	0.329	0.348	0.329
	PMI-enriched VSM (TF-IDF)	0.833	0.863	0.875	0.861

TABLE VII
MEASURES FOR RELIABILITY ASSESSMENT WHEN APPLYING VSM AND PMI-ENRICHED VSM FOR RULE-BASED CLASSIFIERS RIDOR, JRIP, PART

Classifier	Data model	Measure			
		Kappa	Pearson	Spearman	Kendall tau
Ridor	VSM (binary)	0.172	0.197	0.174	0.163
	PMI-enriched VSM (binary)	0.758	0.769	0.795	0.776
	VSM (TF)	0.221	0.232	0.248	0.233
	PMI-enriched VSM (TF)	0.756	0.799	0.807	0.789
	VSM (TF-IDF)	0.202	0.250	0.242	0.225
	PMI-enriched VSM (TF-IDF)	0.821	0.820	0.835	0.821
Jrip	VSM (binary)	0.170	0.159	0.181	0.172
	PMI-enriched VSM (binary)	0.733	0.731	0.755	0.737
	VSM (TF)	0.164	0.144	0.170	0.163
	PMI-enriched VSM (TF)	0.778	0.796	0.812	0.796
	VSM (TF-IDF)	0.150	0.160	0.159	0.151
	PMI-enriched VSM (TF-IDF)	0.815	0.823	0.842	0.828
PART	VSM (binary)	0.225	0.292	0.260	0.241
	PMI-enriched VSM (binary)	0.767	0.783	0.802	0.783
	VSM (TF)	0.243	0.250	0.242	0.225
	PMI-enriched VSM (TF)	0.784	0.803	0.819	0.804
	VSM (TF-IDF)	0.286	0.298	0.320	0.297
	PMI-enriched VSM (TF-IDF)	0.823	0.846	0.855	0.839

The figures clearly show the lack of a fall in the measure values for assessment of the classification performance in the categories that are associated with the least number of customer reviews (i.e. *Mixed*, *Suggestions*).

Table V presents the results of the accuracy of the text classification when applying VSM and PMI-enriched VSM for the listed classifiers. The average value of the accuracy for different classifiers and term weights with the proposed model is 88.51%, which is a 53.53-percent increment in comparison to the average of the accuracy with VSM.

Reliability evaluation

Measures that are particularly useful for assessing the reliability of classification models and classifiers are calculated. The presence of a high degree of reliability implies a consistency, i.e. high probability of obtaining similar results when repeating the tests. For this purpose, the following measures are computed:

- Kappa statistic;
The advantage of Kappa measure in comparison with the accuracy consists of its robust because it takes into account the possibility of accurate classification resulting from chance.
- Pearson correlation;
Pearson correlation is a measure of the linear relationship between two variables, i.e. the actual and predicted category.
- Spearman correlation;
It represents a rank correlation between the actual and predicted categories. Unlike Pearson, Spearman's correlation is not limited to a linear relationship. It measures the monotone association (only strictly increasing or decreasing but not mixed) between two variables and relies on the order of values.

- Kendall tau correlation.

The Kendall tau correlation coefficient measures the strength of the relationship between two variables. Unlike the Spearman coefficient, Kendall's tau does not take into account the difference between ranks, only a directional agreement.

The calculation of these measures is described in detail and applied to assess the classification performance in [26, 27, 28]. Their values fall within the range of -1 (negative correlation) and +1 (positive correlation, i.e. a high degree of consistency between the actual and predicted categories).

Table VI represents the results of the above-listed measures when applying VSM and PMI-enriched VSM for the classifiers SVM, K-NN, Naïve Bayes, H2O's Deep Learning. The resulted values are in the range [0.004, 0.399] for the VSM model and in the range [0.706, 0.875] for the PMI-enriched VSM model.

Table VII represents the results of the above-mentioned measures when applying VSM and PMI-enriched VSM for the classifiers Ridor, Jrip, PART. The obtained values are in the range [0.144, 0.320] for the VSM model and in the range [0.731, 0.855] for the PMI-enriched VSM model.

IV. CONCLUSION

The text classification and its application for customer feedback about online stores are studied in this paper. For this purpose, a dataset consisting of user reviews in Bulgarian is created and published. Besides, a model that enriches vector space model with further extracting the pointwise mutual information of words in regard to categories, is proposed. The results of the experiments confirm the usefulness of the proposed model in the application of different classifiers.

Our future work includes considering emoticons, as well as taking into account the words, written entirely in capital letters in text analysis of user reviews.

REFERENCES

- [1] T. Georgieva-Trifonova, M. Stefanova, St. Kalchev, "Dataset for: Customer Feedback Text Analysis for Online Stores Reviews in Bulgarian", Available: <https://doi.org/10.7910/DVN/TXIK9P>, Harvard Dataverse, 2018.
- [2] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews", *In Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 417-424, 2002.
- [3] J. Büschken, G. M. Allenby, "Sentence-Based Text Analysis for Customer Reviews", *Marketing Science*, vol. 35, no. 6, pp. 953-975, 2015.
- [4] S. Dixit, S. Kr, "Collaborative Analysis of Customer Feedbacks using Rapid Miner", *International Journal of Computer Applications*, vol. 142, no. 2, pp. 29-36, 2016.
- [5] Lu Li, Yun Lin, Xu Wang, Tian Guo, Jie Zhang, Hua Lin, Fuqian Nan, "A Clustering-Classification Two-Phase Model on Service Module Partition Oriented to Customer Satisfaction", *Engineering Letters*, vol. 26, no.1, pp. 76-83, 2018.
- [6] H. Min, J. Yun, Y. Geum, "Analyzing Dynamic Change in Customer Requirements: An Approach Using Review-Based Kano Analysis", *Sustainability*, vol. 10, no. 746, 2018, doi:10.3390/su10030746.
- [7] V. V. Chaudhari, C. A. Dhawale and S. Misra, "Sentiment Analysis Classification: A Brief Review", *International Journal of Control Theory and Applications*, International Science Press, vol. 9, no. 23, pp. 447-454, 2016.
- [8] M. Farhadloo, E. Rolland, "Multi-Class Sentiment Analysis with Clustering and Score Representation", *IEEE 13th International Conference on Data Mining Workshops (ICDMW)*, 2013.
- [9] D. Gräbner, M. Zanker, G. Fliedl and M. Fuchs, "Classification of Customer Reviews based on Sentiment Analysis", *19th Conference on Information and Communication Technologies in Tourism (ENTER)*, Springer, Helsingborg, Sweden, 2012.
- [10] S. Tayel, M. Reif, A. Dengel, "Rule-based Complaint Detection using RapidMiner", *RapidMiner Community Meeting and Conference (RCOMM)*, pp. 141-149, 2013.
- [11] E. A. Stoica and E. K. Özyirmidokuz, "Mining Customer Feedback Documents", *International Journal of Knowledge Engineering*, vol. 1, no. 1, pp. 68-71, 2015.
- [12] P. Lohar, K. D. Chowdhury, H. Afli, Mohammed Hasanuzzaman, Andy Way, "A Multinomial Naive Bayes Classification Approach for Customer Feedback Analysis task", *Proceedings of the 8th International Joint Conference on Natural Language Processing*, pp. 161-169, 2017.
- [13] X. Wei, H. Lin, L. Yang, "Cross-domain Sentiment Classification via Constructing Semantic Correlation", *IAENG International Journal of Computer Science*, vol. 44, no.2, pp. 172-179, 2017.
- [14] B. Kraychev, *Retrieving and analyzing comments and sentiments from online text sources*, PhD Thesis, Sofia University "St. Kliment Ohridski", 2014 (in Bulgarian).
- [15] B. Kapukaranov, P. Nakov, "Fine-Grained Sentiment Analysis for Movie Reviews in Bulgarian", *Proceedings of Recent Advances in Natural Language Processing*, Hissar, Bulgaria, pp. 266-274, 2015.
- [16] J. Mandravickaite, T. Krilavicius, K. Lok Man, "A Combined Approach for Automatic Identification of Multi-Word Expressions for Latvian and Lithuanian", *IAENG International Journal of Computer Science*, vol. 44, no.4, pp. 598-606, 2017.
- [17] P. Nakov, "BulStem: Design and Evaluation of Inflectional Stemmer for Bulgarian", *In Proceedings of Workshop on Balkan Language Resources and Tools*, 2003.
- [18] St. Robertson, "Understanding Inverse Document Frequency: On theoretical arguments for IDF", *Journal of Documentation*, vol. 60, no. 5, pp. 503-520, 2004.
- [19] M. Hofmann, R. Klinkenberg, *RapidMiner: Data Mining Use Cases and Business Analytics Applications*, CRC Press/Taylor & Francis Group, 2014.
- [20] T. Joachims, "Text categorization with support vector machines: learning with many relevant features", *In Proceedings of the 10th European Conference on Machine Learning*, pp. 137-142, 1998.
- [21] Ch. Chih-Chung; L. Chih-Jen, "LIBSVM: A library for support vector machines", *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, 2011.
- [22] F. Lu, Q. Bai, "A Refined Weighted K-Nearest Neighbours Algorithm for Text Categorization", *In Proceedings of International Conference on Intelligent Systems and Knowledge Engendering*, IEEE, pp. 326-330, 2010.
- [23] W. Zhanga, F. Gao, "An Improvement to Naive Bayes for Text Classification", *Procedia Engineering*, Elsevier, vol. 15, pp. 2160-2164, 2011.
- [24] A. Candell, V. Parmar, *Deep Learning with H2O*, H2O.ai, Inc., 2015.
- [25] F. Sebastiani, "Machine learning in automated text categorization", *ACM Computing Surveys*, vol. 34, no. 1, pp. 1-47, 2002.
- [26] C. Ferri, J. Hernández-Orallo, R. Modroiu, "An experimental comparison of performance measures for classification", *Pattern Recognition Letters*, vol. 30, pp. 27-38, 2009.
- [27] J. Jenness, J.J. Wynne, *Cohen's Kappa and classification table metrics 2.0: an ArcView 3x extension for accuracy assessment of spatially explicit models*, U.S. Geological Survey Open-File Report OF 2005-1363. U.S. Geological Survey, Southwest Biological Science Center, Flagstaff, AZ, 2005.
- [28] D. Powers, *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation*, Technical Report SIE-07-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia, 2007.