Sentiment Analysis from Students' Feedback

A Romanian High School Case Study

Daniela MARCU, Mirela DANUBIANU
Stefan cel Mare University of Suceava
Suceava 720229, Romania
mdaniela.marcu@yahoo.ro, mdanub@eed.usv.ro

***Abstract*—The education system is a source that generates significant amounts of data, daily, in various formats and, often, hiding valuable information. Finding a good way to unravel those hidden gems, represents one of the most challenging problems of natural language processing, namely sentiment analysis. This involves, applying NLP and text analysis techniques to identify and classify subjective opinions in different materials such as documents or sentences. In our work, we used as raw data, the opinions of students from eleven high schools in Suceava, related to various aspects of the educational process. They were collected through a Google Docs form, and analyzed through the Orange environment (an open source tool for machine learning and data visualization). In this paper, we make a comparative study of the obtained results using the Ekman and Plutchik models. Each model extracts from the analyzed texts, a different emotion, based on which the students' sentiments towards the educational process will be analyzed.**

Keywords*—*sentiment analysis, data mining, text mining

I. INTRODUCTION

In recent years, there has been an increase in popularity of data science and analytics. In addition the ubiquity of Internet offers the opportunity to access huge volumes of data, most often in written text form. For example, on social networks, many users have the ability to post comments on different educational services, or fill in questionnaires that measure their satisfaction, items that would require free answers, etc. All of this data is subject to interpretation.

Sentiment analysis is a relative new research area. Also known as opinion mining, it involves applying data mining, NLP and text analysis techniques in order to identify and classify subjective opinions from different materials such as written documents [1].

Cultural and social students' foundation may strongly influence the educational process, which raises an interesting challenge, getting to know their opinions about this process in order to continuously adjust it.

This article represents the first step towards analyzing sentiments of high school students related to aspects of educational system. More specifically, we will be using text mining methods implemented in the Orange tool, to try and classify answers regarding school, into classes representing possible sentiments, as defined by the Ekman and Plutchik models.

The objective is to find the best solution for sentiment analysis which will then be used as part of a larger project targeting the pre-university educational system in Romania.

The structure of this article is as follows: Section 2 presents some basic notions about sentiment analysis, and two models (Ekman and Plutchik) for emotional states classification, Section 3 goes through some of the concepts about knowledge discovery in databases, data mining and text mining, Section 4 walks through a case study for sentiment analysis in a school survey, Section 5 collates and analyzes the obtained results and finally, Section 6 shows the conclusions and future work.

II. SENTIMENT ANALYSIS. EKMAN AND PLUTCHIK MODELS

Human emotion is difficult to classify. Sentiment analysis attempts to provide information about the emotion that describes with the greatest accuracy the author's intention in a text message.

The sentiment analysis can be performed at the document, sentence or word level, determining the polarity of the message as positive or negative, or the expression of an emotion. Because it is a topical issue, a lot of algorithms have been developed, starting with those based on the bag-of-word model and ending with the modern ones, based on neural networks [2][3].

Since the 1960s, various theories about the detection and classification of emotions have begun to be developed. In our work, two models built on the theory of basic emotions were used: the Plutchik model and the Ekman model.

Ekman believes that there are common triggers for similar emotions. Emotions are physically expressed through a specific facial mimic, which is not socially learned, but native, that means it may be cultural differences in expression of basic emotions. According to him, the basic emotions form six discrete categories, which do not mix with each other: anger, disgust, fear, joy, sadness and surprise. These belong to one of two states: positive or negative [4].

American researcher Robert Plutchik has classified emotions based on several postulates:

This work was partially supported from the project "Integrated Center for research, development and innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for fabrication and control", Contract No. 671/09.04.2015, Sectoral Operational Program for Increase of the Economic Competitiveness co-funded from the European Regional Development Fund.

978-1-7281-6870-8/20/$31.00 ©2020 IEEE

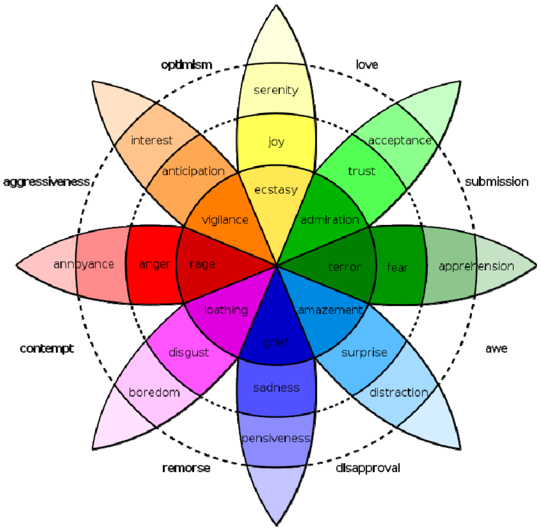

Fig. 1. Plutchik's wheel of emotions [6]

- emotion is specific to both humans and animals;
- emotion has an evolutionary history;
- emotion plays an important role in adaptation and survival;
- emotion may have different degrees of intensity;
- there is a limited number of basic emotions from which all the others derive [5].

Unlike Ekman, Plutchik believes that all emotions are the result of combining some of the eight basic emotions: anger, disgust, fear, joy, sadness, surprise, trust and anticipation. He proposed a structural model known as the wheel of emotions (Fig. 1). The similar emotions are described with nuances of the same colors and are placed at 180° of the opposite emotions. Their intensity is correlated with the distance from the center of the wheel [6].

Emotions are bipolar: joy-sadness, anger-fear, trust-disgust and surprise-anticipation.

III. KNOWLEDGE DISCOVERY IN DATABASES, DATA MINING AND TEXT MINING

Data Mining is a modern research field aiming to discover hidden patterns in data, by applying specific algorithms. In a broader context, data mining is only a stage in a complex process, namely ***knowledge discovery in data*** (KDD). This process aims to identify "valid, novelty and potentially useful" patterns from large volumes of data [7].

By its specificity, KDD lies at the intersection of the following domains: databases, statistics, machine learning, artificial intelligence and visualization, as shown in Fig. 2.

The knowledge discovery in data has aroused a great interest both in the academic world and in the economic environment, each of them proposing its own models.

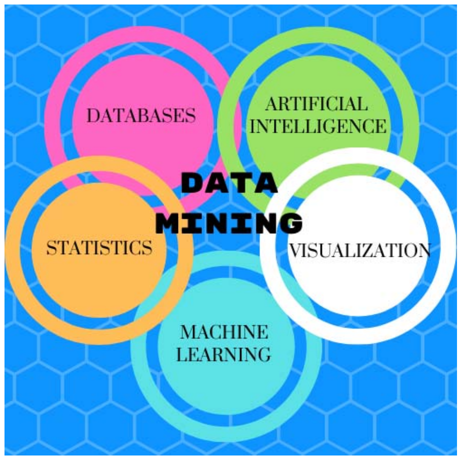

Fig. 2. The relationship between data mining and other research areas

Regardless of the approach and the number of steps considered, in all the variants on can find the phases presented in Fig. 3.

No system, even less the educational one, does not provide data in an acceptable form for immediate processing.

These contain noise, anomalies, redundancies, which must be cleaned and removed in order to assure a proper quality of data for further processing. In the preprocessing stage, the data are also transformed to be brought to the right form for the application of modeling algorithms.

The central point of the KDD process is **data mining** or data modeling, when based on specific models and algorithms, data patterns are determined.

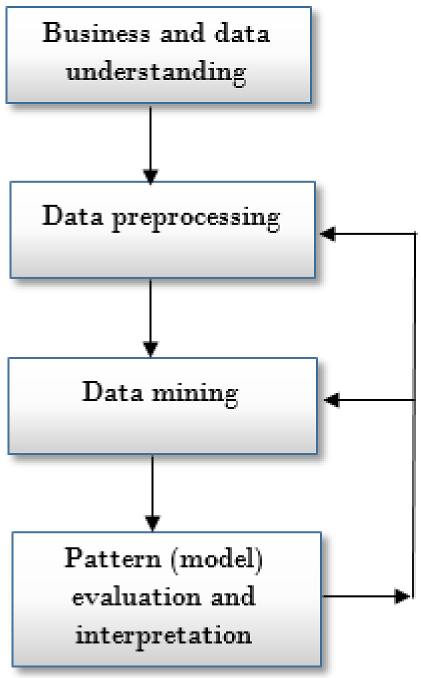

Fig. 3. Main steps in KDD models

These patterns are validated and interpreted by experts in the field [8].

By appropriate methods, data mining aims to solve two broad categories of problems: prediction and description. For each of these problems, specific methods are used. Prediction can be achieved by classification and regression, whereas for description it can be used clustering, deviation detection or association rules [9].

As we noted above, one of the data mining methods is ***classification***, that involves the building a predictive model. To obtain a such model it must be used a training data set which is characterized by attributes that will help to identify the data belonging to a basic class.

For our work such attributes could be: the marks obtained in certain baccalaureate disciplines, the optional disciplines chosen by students, the scholars' opinions about a certain subject, the score given regarding the usefulness of some study disciplines, the descriptive characteristics of the pupils.

Another method is ***clustering*** that involves building models that group the data into clusters according to certain similarities [10].

Text mining, is considered as the new frontier between data mining and predictive analytics [11] and it aims to identify facts or relationships that could remain hidden in textual data by extracting keywords from huge volumes of words and using them for building meaningful patterns or for making predictions.

The process of knowledge discovery in text respects the model presented in Fig. 3. The pre-processing stage, however, has as a central objective the conversion of unstructured text into structured data on which on can apply analytical techniques to classify or cluster [12].

IV. SENTIMENT ANALYSIS IN A SCHOOL SURVEY

Cultural and social foundation of peoples plays a crucial role in sentiment expression and analysis [13]. That is the reason we intend to start a such analyse using an original data set reflecting the characteristics of Romanian high-school students.

The data we used in our study were collected using a Google Docs form. It contains several questions that refer, among other things, to: preferred discipline, the assessment concerning the school's contribution to career choice, the evaluation of additional support, received from outside the school, for the fulfillment of school obligations.

Data used in this research have been collected as text paragraph form, as open answer to the question: ***"Please describe in a few words, honestly, how you feel about your school"***. Pupils with age between 16 and 18 years from 4 localities and 11 high schools from the surrounding area or from the city of Suceava were invited. A sample of the received answers is presented in Table I.

Once collected, data was saved to an Excel file. It was necessary to select only relevant data, so that, after removing the incomplete or irrelevant answers for our subject, a number

TABLE I. SAMPLE OF ANSWERS CONSIDERED FOR SENTIMENT ANALYSIS

# Doc	Content
9	I like school, I wouldn't give up on her. I spend a lot of time with her and this gives me the opportunity to stay with colleagues.
10	It's good at school, except when you have to do homework
11	A way to go through life
12	My mother forces me to go to school
13	I am not happy to go to school because I am stressed with irelevant subjects. I think I learning much better in a few hours on the internet
14	There are feelings of happiness, although some teachers destroy these feelings.
15	The school should be restructured, rethought because it does not help at all. it is totally wrong what happens in the so-called Romanian "education".
16	I like to go to school to meet friends, but sometimes it's boring.
17	Sometimes interesting and useful things are done, but generally not. Teachers are often bored. Competition is almost non-existent, both colleagues and teachers are eagerly waiting for the time to finish, doing nothing. Everything has a very slow rhythm and you have to get used to not thinking so as not to give in psychically. There are many irelevant hours, and the number of hours is oddly distributed (eg: in the mathematics-informatics profile, phisics has 3 hours per week, while for programming it is allocated only one hour per week). The most interesting thing I do in school is to walk through the classrooms, socializing. Otherwise, I try not to cry during the hours because I do almost nothing.
18	School is a place where you gather knowledge and interact with people
19	Long time invested for too few results
20	The school of today does not seem at all balanced because I am obliged to learn and to grasp all the objects even if they some of these not interested to me.
21	It's not my favorite place
22	I feel I don't have a purpose or a passion, a pleasure, and really scares me the thought of choosing the subject of future learning

of 191 records were saved. The children's answers were in Romanian, so it was necessary to translate them into English. We used only these answers for data workflow. In the next step, the Excel File was converted as a tab file.

Our application is built with Orange tool [14].

We used the Tweet Profiler widget. It connects to a corpus containing 191 documents.

The data flow first goes through a pre-processing step consisting of: dividing the text into tokens and creating

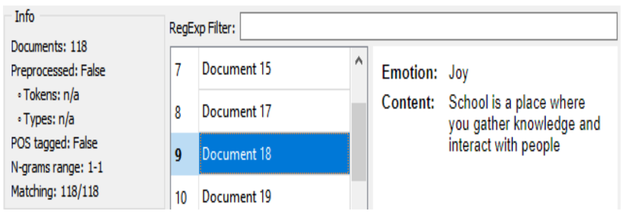

Fig. 4. The document corpus associated with Joy emotion classified with the Ekman model

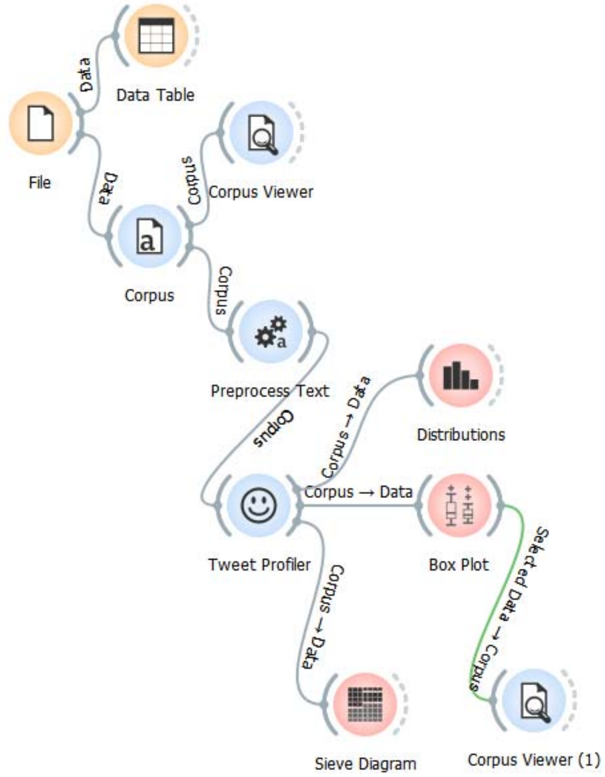

Fig. 5. Data flow design using the Orange tool

n-grams, filtering (stopwords), normalization (stemming and word lemming). Then we select one of the two models of classification of emotions.

The data will be sent to a server where the probabilities of emotion specifique for each document will be calculated.

The data flow is connected to a tool for visualizing the distribution of emotions (BoxPlot).

The documents classified with the selected emotion will be identified in a corpus of documents as Fig 4 shows.

The diagram of the whole process is presented in Fig. 5

V. RESULTS AND DISCUSSION

After the execution of two text mining processes, through which a sentiment analysis based on the two Ekman and Plutchik models was performed, a classification of the received answers was obtained, in classes representing the emotions described by each of them.

Each answer was considered a document, having as index its position in the set of received answers (column #Doc in Table I).

Table II presents a detailed view of results obtained in classifying documents with each of the models. We highlighted the documents' numbers that were same classified, for each emotion.

Fig. 6 presents a comparative study regarding the documents that reflect the different sentiments

TABLE II. RESULTS OBTAINED BY CLASSIFYING THE EMOTIONS DETECTED IN THE STUDENT TEXTS WITH THE PLUTCHIK AND EKMAN MODELS

Emotion	Plutchik Model		Ekman Model	
	Number of classified documents	*Documents reflecting this emotion*	*Number of classified documents*	*Documents reflecting this emotion*
Anger	1	**130**	22	6, 24, 26, 37, 42, 59, 95, 118, 119, 123, **130**, 142, 161, 163, 168, 176, 178, 179, 181, 184, 185, 190
Disgust	0	-	0	-
Fear	12	**22**, 32, 68, 73, **74**, **88**, **90**, 110, **112**, **120**, 170, **173**	17	3, **22**, 33, 34, 66, **74**, **88**, **90**, 96, **112**, **120**, 129, 154, 155, 162, 166, **173**
Joy	72	**7**, **10**, **11**, **12**, **13**, 14, **18**, **30**, **36**, **38**, **39**, **43**, **44**, **45**, **47**, **48**, **49**, 50, 53, **55**, **56**, **57**, **69**, **70**, **71**, **72**, **75**, **77**, 78, **81**, **82**, **84**, **89**, **93**, **102**, **106**, **108**, **109**, **114**, 118, 119, 123, 124, **126**, **127**, **131**, **132**, **135**, **136**, **137**, **138**, **141**, **144**, **146**, **152**, **153**, **157**, **160**, 161, 163, **165**, **167**, 168, 176, 178, 179, 181, **183**, 184, **188**, 190, **191**	112	**7**, 9, **10**, **11**, **12**, **13**, 15, 17, **18**, 19, 20, 25, 28, 29, **30**, 32, 35, **36**, **38**, **39**, **43**, **44**, **45**, **47**, **48**, **49**, 51, 52, 54, **55**, **56**, **57**, 58, 60, 61, 62, 63, 65, 68, **69**, **70**, **71**, **72**, 73, **75**, 76, **77**, 79, 80, **81**, **82**, 83, **84**, 86, 87, **89**, 92, **93**, 98, 99, 100, 101, **102**, 104, **106**, **108**, **109**, 110, 111, **114**, 115, 116, 121, 122, 125, **126**, **127**, **131**, **132**, **135**, **136**, **137**, **138**, 140, **141**, 143, **144**, 145, **146**, 147, 150, 151, **152**, **153**, 156, **157**, 158, **160**, **165**, **167**, 169, 170, 171, 174, 175, 177, 180, 182, **183**, 186, **188**, **191**
Sadness	41	**8**, 15, **16**, 17, 20, **21**, **23**, 25, 28, 35, 40, 51, 54, 58, 60, 62, **64**, 79, 80, **85**, 87, 92, 98, 99, 100, 101, 104, 105, 111, 113, 115, 116, 122, 143, 145, 147, 154, 171, 174, 175, 182	26	1, 2, 5, **8**, 14, **16**, **21**, **23**, 27, 31, **64**, 67, 78, **85**, 94, 97, 117, 124, 133, 134, 149, 159, 164, 172, 187, 189
Surprise	2	86, **128**	14	4, 40, 41, 46, 50, 53, 91, 103, 105, 107, 113, **128**, 139, 148
Trust	62	1, 2, 3, 4, 5, 6, 9, 24, 26, 27, 29, 31, 33, 34, 37, 41, 42, 46, 52, 59, 61, 63, 65, 66, 67, 76, 83, 91, 94, 95, 96, 97, 103, 107, 117, 121, 125, 129, 133, 134, 139, 140, 142, 148, 149, 150, 151, 155, 156, 158, 159, 162, 164, 166, 169, 172, 177, 180, 185, 186, 187, 189	-	Plutchik model specific emotion
Anticipation	1	19	-	Plutchik model specific emotion

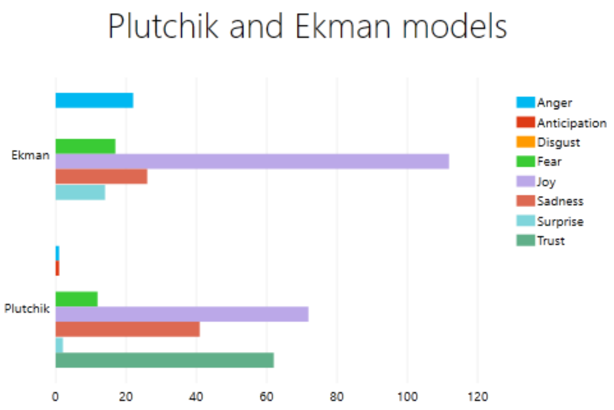

Fig. 6. Comparative study of Ekman and Plutchik models results

We notice that the Plutchik model classifies two emotions in addition to Ekman: *Trust and Anticipation*.

What documents classified with the other model correspond to them? In the table below (Table III), the numerical values represent the number of documents classified with the specified emotion. It is easy to see that Trust and Anticipation are predominantly assimilated by Ekman's classification as emotions that probabilistically describe *Joy*.

TABLE III. HOW MANY DOCUMENTS CLASSIFIED WITH THE EKMAN MODEL FOR EACH BASE EMOTION, CORRESPOND WITH CLASSIFICATION FOR TRUST AND ANTICIPATION EMOTIONS (PLUTCHIK)

Plutchik model		Ekman model					
		Anger	Disgust	Fear	Joy	Sadness	Surprise
Trust	62	9	0	9	19	17	8
Anticipation	1	0	0	0	1	0	0

Perhaps it would be interesting to illustrate with a few texts the correlations that exist, or on the contrary, do not exist, between the classifications of emotions with the two models. Not all document classifications correspond to the emotions described (Table IV).

Related to these examples, the following observations are appropriate.

TABLE IV. INPUT TEXTS SAMPLE DIFFERENTLY CLASSIFIED WITH THE TWO MODELS

Emotion (Plutchik)	**Emotion (Ekman)**	**Source text from the tested document corpus**
Trust	Joy	Document 186: ”I think the school helps us a lot, both to form our general knowledge and to make our first friends.”
Trust	Fear	Document 155: ”In vain, useless, inhibiting horizons, it limits me and does not motivate me to learn what I want more”
Trust	Sadness	Document 164: ” A waste of time”
Trust	Surprise	Document 103: ”good”
Anticipation	Joy	Document 19: ”Long time invested for too few results.”

Document 186 expresses both confidence and joy, being classified as Joy with Ekman model.

Document 155 expresses disappointment and could be considered correctly classified by the Ekman model, but in no way expresses a sense of confidence.

Document 164 expresses sadness, but not confidence, being correctly classified by the Ekman model.

Document 103 can express confidence and possibly surprise. Document 19 did not express joy, but expresses anticipation about the future.

Finally, we can note that 70 documents out of 191 were classified equally by the two models, considering only their common emotions: Anger, Disgust, Fear, Joy, Sadness and Surprise. Only for these we calculated the model accuracy, according to expression (1).

$$\boldsymbol{acc=dcc/td *100} \qquad (1)$$

where:

dcc is the total number of documents correct classified by both models and *td* means total number of analyzed documents.

We obtained 36.64% for model accuracy.

VI. CONCLUSIONS

In this paper, we will compare the results of two different models for sentiment analysis, on the students' feedback.

The pool of users contained students from 11 high schools in or around Suceava.

The data was collected using a Google docs form, including open answers to the following question: "Please describe in a few words, honestly, how you feel about your school".

Then, we removed noise such as incomplete or irrelevant answers, and ended up with a total of 191 answers.

We performed sentiment analysis using the two models, Ekman and Plutchik, and classified the output according to the emotions generated by each model.

For the purpose of this test, we will only be considering 5 emotions: anger, disgust, fear, joy, sadness and surprise. This yielded that almost 37% of the documents were marked identically by the two models.

In addition to those feelings, the Plutchik model classifies two extra basic ones: trust and anticipation.

Results showed that the 30% of the data that was marked with these feelings, were also associated with joy by the Ekman model.

Going forwards, a larger experiment will be run, targeting the pre-university educational system in Romania.

One of this project goals is to find the best solution for sentiment analysis.

This will help us to understand better what the students' immediate expectations and needs are in relation to their education.

REFERENCES

[1] Luo, T., Chen, S., Xu, G., Zhou, J, "Sentiment analysis. Trust-Based collective view prediction", 2013, pp.53–68.

[2] Erik Tromp, Mykola Pechenizkiy, "Pattern – Based emotion classification on social media. Advance in social media analysis", Springer International Publishing Switzerland. 2015

[3] R. A. Calvo and S. D'Mello, "Affect detection: an interdisciplinary review of models, methods, and their applications", IEEE Trans. Affect. Comput., vol. 1, no. 1, pp. 18–37, Jan. 2010

[4] A. M. Cadayona, N. M. S. Cerilla, D. M. M. Jurilla, A. K. D. Balan and J. C. Goma, "Emotional state classification: an additional step in emotion classification through face detection", IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA), Tokyo, Japan, 2019, pp. 667-671.

[5] Plutchik, R., "Emotions and life: perspectives from psychology, biology, and evolution", American Psychological Association, 2003

[6] R. Plutchik, "The emotions", University Press of America, 1991.

[7] FPS96 Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996), "The KDD process for extracting useful knowledge from volumes of data", Communications of the ACM, 39(11), pp.27-34, 1996

[8] Danubianu, Mirela. (2015), "Step by step data preprocessing for data mining. A case study", Proceedings of the International Conference on Information Technologies (InfoTech-2015) Sept. 2014, Bulgaria

[9] M. Danubianu, S.G. Pentiuc, I. Tobolcea, O.A. Schipor, "Advanced information technology - support of improved personalized therapy of speech disorders", International Journal of Computers Communications & Control , Oradea, ISSN: 1841- 9836, Vol: 5, Nr: 5, 2010, p. 684-692

[10] Marcu, D., Danubianu, M., "Learning Analytics or Educational Data Mining? This is the question....", BRAIN, Broad Research In Artificial Intelligence And Neuroscience, Vol.10, 2019, pp.1-14

[11] Kotu, V., & Deshpande, B., "Text mining. Predictive analytics and data mining", 2015, pp. 275–303

[12] Chauhan, G. S., Agrawal, P., & Meena, Y. K. (2018), "Aspect-Based sentiment analysis of students' feedback to improve Teaching–Learning Process", Smart Innovation, Systems and Technologies, 259–266.

[13] H. Hamdi, P. Richard, A. Suteau, and P. Allain, "Emotion assessment for affective computing based on physiological responses," in IEEE International Conference on Fuzzy Systems, 2012.

[14] Martin Stražar, Lan Žagar, Jaka Kokošar, Vesna Tanko, Aleš Erjavec, Pavlin G Poličar, et al., "scOrange—a tool for hands-on training of concepts from single-cell data analytics", Bioinformatics, Volume 35, Issue 14, July 2019, Pages i4–i12