BRAIN. Broad Research in Artificial Intelligence and Neuroscience
ISSN: 2068-0473 | e-ISSN: 2067-3957
Covered in: Web of Science (WOS); PubMed.gov; IndexCopernicus; The Linguist List; Google Academic; Ulrichs; getCITED; Genamics JournalSeek; J-Gate; SHERPA/RoMEO; Dayang Journal System; Public Knowledge Project; BIUM; NewJour; ArticleReach Direct; Link+; CSB; CiteSeerX; Socolar; KVK; WorldCat; CrossRef; Ideas RePeC; Econpapers; Socionet.

2024, Volume 15, Issue 2, pages: 320-343 | https://doi.org/10.18662/brain/15.2/578
Submitted: November 29th, 2023 | *Accepted for publication:* May 7th, 2024

Toward Better Education Quality through Students' Sentiment Analysis Using AutoML

Corina SIMIONESCU [1*]
Daniela MARCU [2]
Marius Silviu MĂCIUCĂ [3]

[1,*] Doctoral School of Applied Sciences and Engineering, Faculty of Electrical Engineering and Computer Science, Stefan cel Mare University of Suceava, Romania, corina.simionescu@usm.ro
[2] Doctoral School of Applied Sciences and Engineering, Faculty of Electrical Engineering and Computer Science, Stefan cel Mare University of Suceava, Romania, mdaniela.marcu@yahoo.ro
[3] Doctoral School of Applied Sciences and Engineering, Faculty of Electrical Engineering and Computer Science, Stefan cel Mare University of Suceava, Romania, marius.maciuca@student.usv.ro

Abstract: *Sentiment analysis from students' interactions with learning environments is a topic of interest for researchers in the field of education because it can make important contributions to improving the quality of instructional processes through recommendation systems integrated into learning applications, or by improving the quality of courses, by grouping students according to their common interests and providing feedback on school progress. There are two approaches to sentiment analysis: one lexicon-based and another that uses machine learning.*
In this study, we present a sentiment analysis from two own data sets that represent students' opinions about school. Our goal is to create a model that helps us to automatically label students' opinions, assigning sentiment scores between 0 and 4 (0 for an extremely negative opinion). To train and evaluate the performance of the model, we used opinions collected from 1443 Romanian high school students. The novelty that we propose is the manual labeling system.
Our current research which uses a machine learning approach to classify students' opinions obtains an accuracy of 86.507%.

Keywords: *AutoML; classification models; cloud services; education; sentiment analysis.*

How to cite: Simionescu, C., Marcu, D., & Măciucă, M. S. (2024). Toward better education quality through students' sentiment analysis using AutoML. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 15*(2), 320-343. https://doi.org/10.18662/brain/15.2/578

Introduction

Data mining to discover patterns based on which valuable knowledge can be extracted is successfully used in fields such: as medicine, media, or marketing (Oludare et al., 2018). At the level of the international community, we notice a special interest in the application of Knowledge Discovery in Databases (KDD) in the educational field. Our concerns aim for methods, techniques, tools, frameworks, and architectures that can be used, possibly on a large scale, to increase the quality of educational processes. That can mean: creating recommendation systems and delivering bibliography or content, grouping students according to their level of performance or interests to more efficient curricular adaptation, and using Natural Language Processing (NLP) to perform sentiment analysis from text and images.

To understand to what extent NLP is applied for the analysis of data from the Romanian educational system, which are the current concerns in fields such as sentiment analysis and using AutoML, we performed a search in the IEEE Xplore database. After refining the search by trying different terms, and synonyms, we obtained the expression: (("All Metadata": sentiment analysis) OR ("All Metadata": opinion mining) OR ("All Metadata": Natural Language Processing)) AND ("All Metadata": Romanian) AND (("All Metadata": education) OR ("All Metadata": NLP) OR ("All Metadata": AutoML))). For the period 2011-2022, we obtained 18 results, 17 being materials presented at conferences and 1 article from more domains (see Table 1).

Table 1. Distribution of results by domains

No	Domain	Number of results
1	Natural Language Processing	13
2	Text Analysis	7
3	Learning (Artificial Intelligence)	5
4	Internet	3
5	Data Mining	3
6	Speech processing	3
7	Speech synthesis	3
8	Pattern classification	2
9	Public domain software	2
10	Sentiment analysis	2
11	Statistical analysis	2
12	Vocabulary	2
13	SQL, Application program interfaces, Audio signal processing, Authorization, Biocomputing, Biological tissues, Biometrics, Call-centers, Cloud computing, CNN, Computerized monitoring, Cellular Biophysics	One result at a time

Following the results provided by the search, the articles and proceedings in Table 2 were selected.

Table 2. Description of the articles and proceedings returned by the search

Reference	The purpose of the research	Algorithms, methods, resources, and tools used
Buzea et. al, 2019	Evaluating opinions and emotions to identify the extent to which the results obtained can be used in a real system that provides relevant information about the opinions expressed online.	Romanian news corpus, Romanian dictionary Classification algorithms: Naive Bayes, Decision Trees, Support Vector Machine
Marcu & Danubianu, 2020	Application of NLP and text analysis techniques to the analysis of students' sentiments about the school, based on the models of psychologists Ekman and Plutchik.	Classification algorithms, Ekman and Plutchik models, Orange tool
Russu et al., 2014	Sentiment analysis from documents written in Romanian, using unstructured data.	Rule-based approach, features extracted
Schuszter, 2018	Classification of sentences from psychological surveys to find out whether respondents express emotions or behaviors.	Wikipedia and Common Crawl data sets Deep Learning Convolutional and Recurrent Neural Networks (CNN, RNN)
Ungurean and Burileanu, 2011	Construction of a Text-to-Speech system for the Romanian language and integration in the Speech Synthesis Markup Language system.	Speech Synthesis Markup Language Text-to-Speech system
Vasiu and Potolea, 2020	A tokenization strategy of the Romanian text with the approach of three text formats coming from sources: standard, archaic, and medical.	NLP - Romanian tokenization

Although in recent years, much of the research has focused on topics such as detecting irony in online news (Buzea et al., 2020), voice-based biometric authentication (Cocioceanu et al., 2016), human-computer interaction (Boros & Dumitrescu, 2017; Ilie et al., 2013; Păiş et al., 2021) and others, we notice an increase in interest in the field of education, especially after 2018. Some research is not exclusively aimed at education (Buzea et al., 2019; Ungurean & Burileanu, 2011; Vasiu & Potolea, 2020), but it can be a valuable direction for future studies. The sentiment analysis that uses unstructured data in the Romanian language (Russu & Dinsoreanu, 2014) is a rather little-researched field, unlike the classification of emotions based on texts written in English (Marcu & Danubianu, 2020). The use of deep learning for CNN or RNN architectures to classify the answers from psychological surveys (Schuszter, 2018) is another direction of research identified in the search we conducted.

Following this trend, we want to extend the scope of NLP in the Romanian educational field with research aimed at Sentiment Analysis from text data. As a part of a project within the doctoral studies, started in 2018, which aims to improve the quality of educational services in Romania by using new data exploration technologies, our research aims to create a model for classifying students' opinions about the school using sentiment analysis.

Continuing our previous research, which aimed to analyze the students' opinions using the classifications of basic emotions, we present in this article another approach, which consists of the use of cloud services to classify the opinions expressed by students in relation to the school.

We use an approach based on supervised machine learning, so the data are labeled with sentiment scores between 0 and 4 (0 expressing a deeply negative sentiment about the school, 1 expresses a negative sentiment, but not so vehemently like an opinion labeled with the score 0, 2 is associated with an opinion that expresses negative and positive sentiments in the same proportion, 3 is for a generally good opinion, while 4 is the score of sentiment associated with enthusiasm about the school). The choice of 5 sentiment classes is motivated by previous research, within the doctoral studies program, carried out to train, validate and evaluate a model on a data set labeled with scores from 0 to 4, available online, to make a comparative study between its performances and those obtained by the model trained on own dataset, manually labeled, presented in this article.

The novelty of our paper is that we used original data, collected by involving many students who completed sets of questionnaires designed according to the characteristics of Romanian education. The analysis of these data has two objectives: on the one hand, it reveals those aspects that displease the students, aspects extracted from the questionnaire's responses, and on the other hand, to use the experience of the model for the automatic labeling of another data set, of the same type.

Further, the paper is structured as follows: II. Related work, III. Related terminologies, IV. Working Methodology, V. Results and Discussion, VI. Conclusions and VII. Limitations of the study and implications for educators.

Related work

Sentiment analysis from educational data can lead to the improvement of educational processes with a direct consequence of increasing student performance and reducing school dropout. Knowing and understanding the students' sentiments by analyzing the feedback provided by them is essential for making the learning process more efficient. Some authors (Dolianiti et al., 2019) identify 5 types of tasks that sentiment analysis could solve: evaluation of training and decisions at the level of educational institutions, improvement of training systems, evaluation of learning tasks, and creation of new research perspectives. A study on the tools, techniques, and methods of sentiment analysis, but also the identification of the benefits brought to the educational field led to the results presented in the current section.

The increase in the number of educational platforms and students, and the diversification of technologies used in education have imposed new ways of data analysis, some of which have already proven their usefulness in other fields. In a literature review, Mite-Baidal et al. (Mite-Baidal et al., 2018) show that MOOC forums and social networks are the most used sources for educational data collection and Naive Bayes is one of the most used sentiment analysis techniques. Promising results in the evaluation of universities are obtained by Toçoğlu and Onan (Toçoğlu & Onan, 2021) by analyzing a corpus made up of 700 student reviews. The approach used is based on machine learning. Along with Naive Bayes, the authors also use classifiers: SVM (Support Vector Machine), Logistic Regression, and Random Forest. Using machine learning and lexicon-based approaches, Chauhan and colleagues (Chauhan et al., 2019) perform a sentiment analysis using opinions expressed by students and teachers on different social media platforms.

A Deep Learning-based approach achieves very good results in sentiment analysis from educational data with SVM and MLP (Multilayer Perceptron) models using the backpropagation algorithm. The authors (Sultana et al., 2018) use WEKA as a working tool and perform a comparative analysis with other methods used such as Decision Tree, Bayes Net, Simple Logistics, Multi-class Classifier, K-star, and Random Forest. The performances of the models are evaluated with the Accuracy, RMSE (Root Mean Square Error), Specificity, Sensitivity, and ROC curve area metrics. Onan (Onan, 2021) shows that architectures based on deep learning are more efficient. He uses a dataset of 66000 MOOC reviews, achieving a classification accuracy of about 96% using long short-term memory networks in conjunction with GloVe (Global Vectors for Word Representation). Osmanoğlu and collaborators (Osmanoğlu et al., 2020) use the triple Likert method to scale over 6000 feedbacks provided by online learners of Anadolu University, data that will be used by the Logistic Regression algorithm for sentiment analysis.

Mostafa (Mostafa, 2020) analyzes the sentiments of two groups of students, 700 using games created for educational purposes (Gamification) and 300 using classic courses. The best performances are obtained with the Naive Bayes algorithm for which the accuracy indicator has a value of 83%, followed by SVM with 79%, and respectively Decision Tree with 76%. The results of the research show that the use of Gamification tools positively influences the school performance of students, with them obtaining the minimum qualification of 78, compared to others who obtain 47.

Another approach to sentiment analysis is the lexicon-based one. This generally involves computing the positive, negative, or neutral sentiment orientation at the word or phrase level based on a dictionary (Gupta & Agrawal, 2020). There are two approaches, one based on the dictionary and another based on a corpus of words specific to the domain or context under analysis (Aung & Myo, 2017). Aung and Myo evaluate the academic performance of teachers based on the opinions expressed by students. For this, they built a sentiment word database to identify the polarity of words. Synonyms are considered positive opinions and antonyms are negative. The sentiment score assigned to the words ranges from -3 to +3. Negative scores are associated with negative opinions, those between 1 and 3 with positive opinions, and a score of 0 is a neutral opinion. Based on the dictionary, each word of the sentence is given a polarity score. Next, the semantic orientation score of the sentence and the average polarity score at the level of all comments are calculated.

Other authors (Ardianto et al., 2020) use the RapidMiner tool and the Naive Bayes and SVM algorithms together with the Synthetic Minority Oversampling Technique (SMOTE) to classify 15,000 tweets about e-sports into two sentiment categories, positive and negative. The comparative analysis reveals that the Naive Bayes algorithm together with SMOTE obtains the highest scores for accuracy and AUC.

Kastrati and collaborators (Kastrati et al., 2020) create a weakly supervised framework for aspect-based sentiment analysis consisting of 4 basic components with roles in information input, aspect category learning, weak label propagation, and aspect category polarity. The authors use two datasets, one containing over 100,000 reviews collected from Coursera and another consisting of feedback from 6,000 students in the traditional system. A multi-layer Convolutional Neural Network (CNN) architecture and ReLU and softmax as activation functions are used to make predictions regarding appearance categories and label propagation.

Tools and frameworks capable of automating machine learning have the advantage that they can be used not only by data scientists. This is one of the reasons why AutoML (Automated Machine Learning) has become increasingly used. An analysis of searches conducted in the IEEE Xplore, ACM Digital Library, and Google Trends databases from 2012 to 2021 (Siriborvornratanakul, 2022) reveals that Google Cloud AutoML is more popular than AutoKeras, Auto-Sklearn or Auto-WEKA, especially after 2017. The study refers to specific aspects of the field of health.

In a study of log files resulting from interactions with three MOODLE courses held at the Aristotle University of Thessaloniki in 2017-2018 (Tsiakmaki et al., 2020), the authors use AutoML to make predictions on the final performances of 591 students. The data collection was carried out by customizing a plugin of the platform that calculates for each student: the number of views and posts, the grades obtained in various activities such as workshops or quizzes, and the final grades, aggregated event counters. The tool used in algorithm selection, feature selection, and preprocessing was WEKA. To determine the classifier with the highest performance, the authors use Auto-WEKA and SMAC. The classification and regression algorithms (Naive Bayes, Rule-based classifiers - JRip, One Rule (OneR), PART and M5Rules -, Random Forest, Function-cased classifiers, kNN, and Meta classifiers) were 10 times executed using a cross-validation technique. The authors followed three aspects: predicting students' academic performance, predicting Pass/Fail, and predicting dropout students. The obtained results reveal that AutoML provides better performance than classical learning algorithms with default hyperparameter settings. Tuning the hyperparameters and identifying the best model are very difficult tasks for building learning models, but Auto-WEKA provides solutions it can be used by educators who are not data scientists' specialists.

Most of the research identified in the searches used data from the social network Twitter. One such research, using Google Cloud's AutoML Natural Language Sentiment Analysis achieves 80.89% accuracy in classifying tweets shared with the hashtag "5G" into three sentiment classes: positive, negative, and neutral (Herrera-Contreras et al., 2020).

To the best of our knowledge, our approach using Google Cloud's AutoML Language service to perform sentiment analysis on high school students' opinions about school appears to be unique at this time. In this article, we will present the classification of opinions into 5 sentiment classes (from 0 to 4). We achieved a classification accuracy of 86,507%, which encourages us to continue research in this direction, on large data sets, collected from learning applications that will be built in Google Cloud.

Related terminologies

Natural Language Processing (NLP) can be defined as a set of techniques and tools that gives machines or applications the ability to understand human language, written or spoken, in the same way, that humans do, to perform a series of tasks (Sethunya et al., 2016). NLP is encountered daily in the interaction with virtual assistants such as chatbox or those such as Alexa, spell checking, and spam filters.

NLP is from the 1950s at the crossroads of two areas: artificial intelligence and linguistics (Nadkarni et al., 2011). The algorithms used by NLP combine two techniques related to the semantics and syntax of the sentence (Chapman, 2006) and they are projected to solve specific problems such as speech recognition, entity recognition, sentiment analysis, and speech analysis (Nadkarni et al., 2011). Among the most used analysis techniques are Support Vector Machine (SVM), Bayesian Networks, Decision Trees, and those based on deep and convolutional neural networks (Young et al., 2018).

A list of low-level NLP tasks includes tokenization, part-of-speech identification, sentence boundary detection, morphological decomposition, and segmentation of text. High-level tasks are based on them to solve specific problems such as: named entity recognition or spelling / grammatical error identification and recovery (Nadkarni et al., 2011).

Sentiment analysis uses specific NLP techniques to identify the opinion expressed by the author of a text (Medhat et al., 2014). In general, they can be classified into positive, negative, or neutral opinions, but there are also other classifications such as those based on emotions (Marcu & Danubianu, 2020) or by assigning a sentiment score, as described in this article. Some specific NLP tasks used for sentiment analysis can be: aspect extraction, entity recognition (Poria et al., 2016), or sarcasm detection (Poria et al., 2017).

There are two approaches to text classification: based on statistics (Lexicon Based Approach) and based on machine learning (Thangaraj & Sivakami, 2018).

Machine learning supervised algorithms are based on labeled data that will be used to train the models and evaluate the performances provided by them. The experience gained through training can be used on new data sets (Sumit & Santanu, 2017).

The sentiment analysis lexicon-based approach requires a predefined lexicon that represents a set of terms, a dictionary specific to a subject or a certain language. This lexicon, usually manually created, contains words or phrases that are labeled according to their sentiment orientation: positive or negative (Bonta et al., 2019).

At present, there is a growing interest in using Automated Machine Learning (AutoML) which is based on the experience of pre-trained models on huge data sets. This type of service allows researchers to automatically build scalable machine learning models (Karmaker et al., 2021).

We used a specialized service from Google Cloud, named AutoML Language. Before that, it was necessary: to configure cloud services and obtain and preprocess the data used, operations that were described in the next section.

Training data was used to identify patterns in data. Each iteration of the training process aims to reduce classification errors by changing the values of the weights of the connections between the neurons of the neural network. A part of this data is used during the model training process to modify its hyperparameters to adjust the model structure. In this way, the risk that the model will encounter difficulties in generalizing some examples that do not exactly match the training data is lower (AutoML, 2022). The evaluation data was used to evaluate the performance of the model.

For sentiment analysis, deep learning is currently used as a powerful machine learning technique with very good classification results (Zhang et al., 2018).

AutoML services have the advantage of facilitating the use of deep learning and involve the automated construction of a machine learning pipeline that consists generally of data preparation, model generation, and evaluation (He et al., 2021).

AutoML automatically identifies and uses the best type of machine learning algorithm to increase performance using the concept of Neural Architecture Search (NAS) which automates the design of neural networks (AutoML, 2022).

The construction of the model using the NAS technique involves search space, search strategy, and performance estimation strategy (Elsken et al., 2019). Different search strategies can be applied to explore the space of neural architectures such as Bayesian optimization or reinforcement learning (RL). Bayesian optimization is one of the most popular methods of optimizing hyperparameters that contributes to the generation of the model through optimization methods (Shahriari et al., 2016). Two characteristics of reinforcement learning are the search by trial and error and the delayed reward that define the action of learning through a reward system (Sutton, 1992).

We used this technique to find the best neural network architecture for model generation using our dataset. This involves hyperparameters optimization (such as learning rate, a parameter used in the training of the model), and architecture optimization (such as the number of layers of the neural network).

Working methodology

To obtain the results, we went through several stages. These are preprocessing, modeling, and model use (see Figure 1). The preprocessing consists of the manual labeling of the data that will be used for the training and evaluation of the model, following their collection through a questionnaire. Next, the dataset was split into several files that were grouped into an archive and uploaded to the cloud. The second stage is to build, train and evaluate the model. For this, we prepared the cloud infrastructure necessary to achieve the proposed goal, then divided the data set into data for training and evaluation, and, after the training process was done, calculated the metrics necessary to assess the performance obtained by the created model. Finally, we used the model's performance to classify students' opinions about the school, opinions collected through another questionnaire.

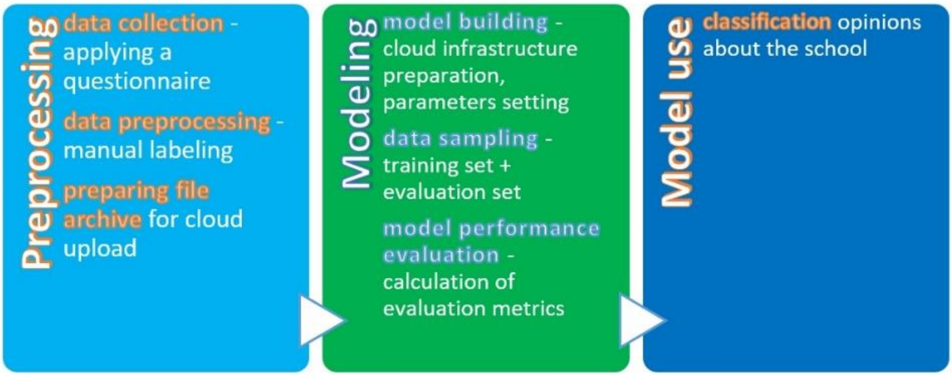

Figure 1. Stages of working methodology

A. Datasets description

In our research, we used two data sets. The first set was used to train and evaluate the model. We named this set *opinions-about-school*. The second set, consisting of 191 opinions about the school, was used to classify the opinions expressed by the students. For this purpose, we used the created model. We named this set *opinions-used*.

To students can express their opinions similarly, we used indicators, formulated based on the general objectives of UNESCO (The United Nations Educational, Scientific and Cultural Organization) (UNESCO, 2018) which support the formulation of an opinion and aims at various aspects regarding education such as gender and opportunity equality, the participation in non-formal activities or extracurricular projects in the high school years, and others.

The questionnaire "Opinions about the school" was applied, between October 26 and November 5, 2021, to a number of 1443 respondents. They were selected, voluntarily, from high school students in Romania, from the 9th to 13th grades. The participation was anonymous, without accessing personal identification data such as e-mail, telephone, address, name, and school. Table 3 presents the structure, questions, and answer types in this questionnaire.

Table 3. *Opinions about the school* questionnaire

ID	Question	Answer type
Section 1 – Data about students		
I1	What grade are you in in the 2021-2022 school year?	Multiple answers with one choice from 9, 10, 11, 12, 13
I2	What is the profile of the class in which you study?	List of options which includes 11 existing profiles and specializations in current state education (2021)
I3	In which county do you study high school?	List of options comprising 42 counties from Romania
Section 2 – Students' opinions about the school		
I4	To what extent does the education in your school promote equality of opportunity and gender?	Linear scale with values from 0 to 4 (with 0 for very small measures)
I5	Did you participate in non-formal activities or extracurricular projects during high school?	Multiple answers with one choice: Yes or No
I6	Did you receive information on human rights, health education, or financial education in compulsory, optional, or leadership classes?	
I7	To what extent do you have access, in school, to the internet, drinking water, heat, and infrastructure adapted to people with disabilities?	Linear scale with values from 0 to 4 (with 0 for very small measures)
I8	Do you feel safe at school?	

I9	To what extent do you think school achievement contributes to the development of skills needed for your career, professional or academic?	
I10	To what extent do teachers provide feedback for your school activities?	
I11	Please express, in a few words, your sincere opinion about the school. For formulating that, you can be guided by the answers given to questions I4 to I10 or refer to other aspects related to school.	Open answer

254 of the respondents are in the 9th grade, and 408 are in the 10th grade. Most of the respondents are in the 11th grade (415), while the fewest are in the 13th grade (37). 329 students are in the 12th grade.

For our study, it was interesting to analyze the free answers associated with the last question, answers in text form, to model the opinions of the students regarding the school.

The second data set called *opinions-used* consists of 191 opinions expressed by 11th and 12th-grade students, from the Real profile. Data were classified through labeling with sentiment scores from 0 to 4 using the Opinions model experience.

The data has been collected as text, using a Google Docs form, as an answer to the question: "*Please describe in a few words, honestly, how you feel about your school*". Students between 16 and 18 years old from 4 localities and 11 high schools from the surrounding area or the city of Suceava were invited. A sample of the received answers is presented in Table 4.

Table 4. A sample of the *opinion-used* dataset

#Doc	Content
9	I like school. I wouldn't give up on her. I spend a lot of time with her and allows 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 non-relevant subjects. I think I learn much better in a few hours on the internet.
14	There are feelings of happiness, although some teachers destroy these feelings.

Training the model on its data set makes the performances obtained in the classification of opinions to be good. The second set of data is very similar to the first, which determines a classification of students' sentiments about the school that corresponds to the reality expressed by them.

B. Description of the infrastructure used

To perform sentiment analysis about the school, we used the AutoML Language service from Google Cloud. It is built to solve NLP problems such as document classification, entity extraction, and sentiment analysis. A series of service configurations were performed for the research: Identity and Access Management (IAM) to manage permissions granted to cloud-logged members to access resources such as buckets, virtual machine instances, and GKE clusters. Google Cloud SDK for resource management from the Google Cloud platform has been installed on an Intel (R) Core (TM) i7 - 8565U CPU @ 1.80 GHz 2.0 GHz system with 8 GB RAM, 64-bit Operating System, x64 - based processor, OS Windows 10 Pro, 2019. From the Management Tools product range we used Google Cloud Console for the visual administration of Compute Engine resources. We used Google Cloud Shell to manage infrastructure and write commands.

Using Google Cloud Storage, the *sentiment-analysis-1* bucket was created where the data will be uploaded. It uses a multi-region location, the Google-managed key encryption type, and contains 3 folders: *Data-opinions*, *Results-Sentiment-Analysis,* and *Sentiment-Analysis* (see Figure 2).

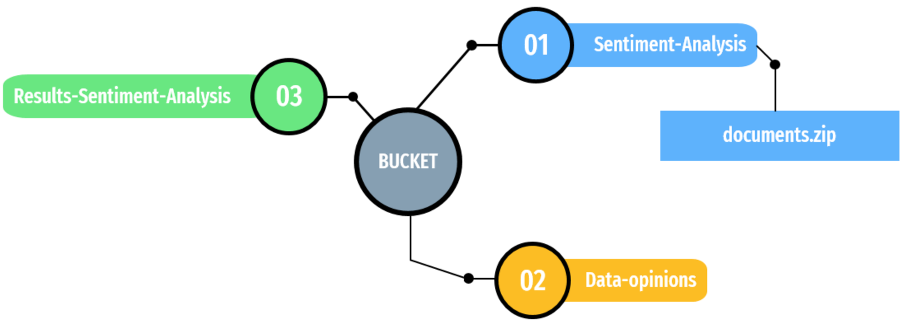

Figure 2. Structure of sentiment-analysis-1 bucket.

The *documents.zip* archive, which contains the data needed to train and evaluate the model, was uploaded to the *Sentiment-Analysis* folder in the cloud. Obtaining this archive is detailed in the next section which describes the entire operational flow.

The *Data-Opinions* folder contains the PDF *opinions-about-school* file, obtained by concatenating the 191 opinions in the second data set to be used to assign a sentiment score by the model.

The results obtained by the model were saved in a CSV format file, named *Results_Sentiment-Analysis_export_data.csv*, in the *Results-Sentiment-Analysis* folder. This file contains records related to the data type (training, evaluation), details of each text file used in the *documents.zip* archive, and the sentiment score associated with it. A record from this file is: "*TRAIN gs://sentiment-analysis-1/Results-Sentiment-Analysis/export_data-sentiment_analysis-2021-12-02T19:38:31.502609Z/files/fe870cea-90d0-4565-b6bc-e7b9b1c074a2.txt 1*".

We created the model Opinions for which we used data labeled with sentiment scores in the range of 0 to 4, 0 corresponds to an extremely bad opinion.

Before using cloud services for sentiment analysis, a step of data preprocessing was required to obtain a data set suitable for modeling.

C. Data preprocessing. Operational flow

The labeling of the data necessary for training and evaluation of the model was a complex process that involved: 2 professors from the high school system in Romania, a student in Human Psychology from Tilburg University in the Netherlands, and two students in the 11th grade, Real profile, at a high school in Romania. 4 team members independently assigned students' messages a score of 0 to 4. The results were confronted and a new stage of relabeling opinions followed which not received the same sentiment score from all team members. Messages with different scores resulting from the second labeling were retrieved for a new labeled process by the team's second teacher. Finally, opinions that were labeled differently received the label of the majority sentiment score (Table 5).

Table 5. Examples of opinions expressed by students

Sentiment score	Opinion
0	Nothing is learned at school for use in the future.
1	My sincere opinion of the school is that much more should be invested in education.
2	For me, a school is a place of training and development, of knowledge and understanding of some essential things. I do not consider that this information received in high school will help me in my future career.
3	It is a high school with good teachers, who are close to the students. Instead, be necessary some important facilities (heat, a modernized gym) and more extracurricular activities (excursions, activities, possibly, on important days, dress codes, as they are at other high schools).
4	My opinion about the school is a good one because the classes go very well and the teachers have a super friendly attitude with the students.

An example of an opinion labeled with different scores is: "There are no cases of inequality of opportunity, and the feedback of the teachers does not appear much because I am not very good at learning." She received 1 from one student, 2 from a second, 3 from a teacher and 2 from the other, and 2 from a psychology student. Final score: 2.

Table 6. *documents.zip* archive structure

Folder from documents.zip archive	Number of TXT files
0	153
1	155
2	248
3	222
4	494

We obtained 1272 records associated with sentiment scores as follows: 153 with a score of 0, 155 with a score of 1, 248 with a score of 2, 222 with a score of 3, and 494 with a score of 4.

Each opinion in the Excel file that contains all the messages along with the associated sentiment scores has been saved in a TXT file. The splitting process was automated by creating a program using the C ++

programming language, the Code::Blocks 20.03 programming environment, and the GNU GCC Compiler.

We grouped these files into folders for each sentiment score. The 5 folders obtained (0, 1, 2, 3, 4) were grouped in the *documents.zip* archive (Table 6).

This archive was uploaded to the bucket in the *Sentiment-Analysis* folder (see Figure 2).

The dataset *opinions-about-school* was divided into training and evaluation data in the proportions indicated in Table 7.

Table 7. Dividing *opinion-about-school* dataset into training and evaluation data

Sentiment score	Training data (90%)s	Evaluation data (10%)
0	138	15
1	140	15
2	223	25
3	200	22
4	445	49

Results and discussion

A. Model assessment

At the level of all sentiments, equal values were obtained for Precision and Recall, namely 86.51%.

Precision values range from 77.27% to 100% for each class, and those for the Recall metric range from 77.27% to 93.33% (Table 8). For opinions labeled with a sentiment score of 0, there is no False Positive case because Precision is 100%. Recall higher values for sentiments labeled with scores 1 and 4 produced fewer False Negative results. A value of 93.33% for Recall produces a single False Negative case for the label score 1 (student opinion: "In my opinion, the school is outdated, because in 2021 I still don't have the heat in the school, the hot water and I have a classroom unrenovated in the last 30 years. Also, the school has not changed much in the last 100 years. We are still learning from books and lessons taught by teachers, although we now have a multitude of materials online at a single click away.").

Table 8. Precision and Recall calculated for each sentiment score

Sentiment score	Precision (%)	Recall (%)
Sentiment score 0	100%	86.67%
Sentiment score 1	87.5%	93.33%
Sentiment score 2	80%	80%

Sentiment score 3	77.27%	77.27%
Sentiment score 4	90%	91.84%

The Opinions model makes correct classifications in a balanced way at the level of sentiment classes. The correct associations are in the range of 77% - 93%. Opinions associated with sentiment score 1 are 93% correctly labeled. The opinions associated with the sentiment score 4 are correctly classified in the proportion of 92%. 77% of the data used for evaluation are correctly labeled with a score of 3 (see Figure 3).

Most misclassifications are made to sentiments with a score of 3 and those with a score of 0. 14% of opinions with an actual score of 3 receive a score of 4, and 13% of opinions that have been labeled with a score 0, are associated by the model to class with score 3.

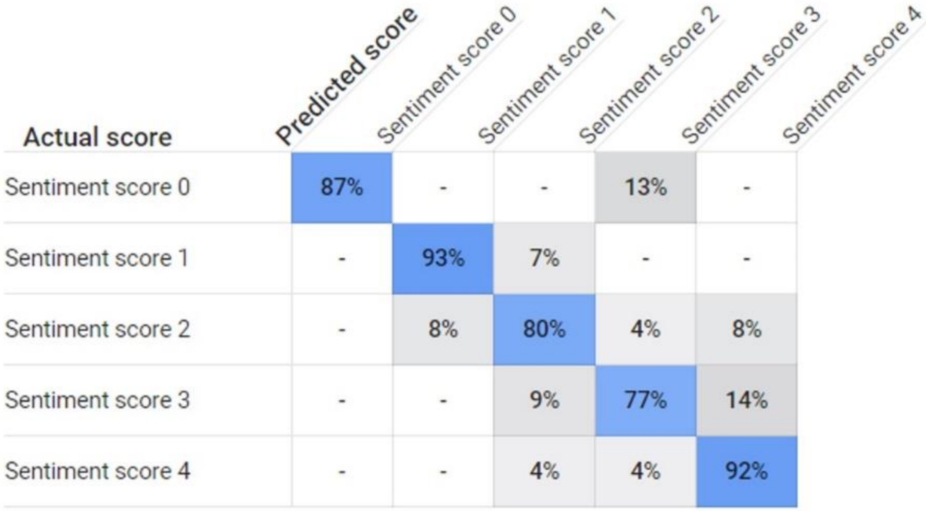

Figure 3. The confusion matrix – percentage format.

B. Student's sentiment analysis through using the Opinions model

Following the use of the Opinions model for the *opinions-used* dataset, 125 opinions were labeled with a score of 0, 6 with a score of 1, an opinion with a score of 2, 52 with a score of 3, and 7 with a score of 4 (Table 9).

Table 9. Number of opinions corresponding to the use of the *Opinions model* for each sentiment score

Sentiment score	Number of opinions classified with the Opinions model
Sentiment score 0	125
Sentiment score 1	6
Sentiment score 2	1

Sentiment score 3	52
Sentiment score 4	7

Using the model performance to label document *opinions-about-school.PDF* loaded into the *sentiment-analysis-1* bucket, in the *Data-opinions* folder we obtained a sentiment score equal to 1.

Given that approximately 69% of the 191 opinions are labeled with sentiment scores 0 and 1, at the level of each opinion, the score 1 at the level of the entire document containing the 191 opinions is a realistic one.

Table 10 presents opinions labeled with different sentiment scores.

Table 10. Examples of opinions labeled using the *Opinions model*

Opinion	**Sentiment score associated with the Opinions model**
An ok environment, relatively quiet. It creates an indifferent state in me	1
School is a way of socializing and learning for me	2
It does not make me happy, but I like it	3
I like sports hours	4
Long time invested for too few results	0
Destroy creativity	0

Conclusions

Cloud computing is the new paradigm of IT solutions developed especially for the online environment. Companies such as Google, Amazon, Microsoft, and Oracle offer the latest, strongest, and most secure computing architecture that integrates the latest software solutions in areas of interest such as deep learning, Big Data, NLP, and computer vision. Cloud services offer scalability and flexibility as well as processing speed, change and adapt to the latest market trends and discoveries in algorithms and software products for modeling, analyzing, and visualizing data and information needed for further business development, benefits which also turn them into a viable solution for their use by the pre-university education system.

Our concerns are related to natural language processing. We created a model for sentiment analysis of the opinions expressed by the students about school. We used the AutoML Language service from Google Cloud.

The main novelty we bring is the use of our data set for training and evaluating the Opinions model.

The *opinion-about-school* dataset was obtained following the application of a questionnaire to a number of 1443 respondents, students from grades 9-

13 in state pre-university education in Romania. To use it for our supervised machine learning model, each opinion went through a difficult process of manual labeling.

The Opinions model experience was used to label a second dataset called *opinions-used*.

The classification performance of the model is quite good.

Knowing students' opinions about the school can lead to resolving some problems and improving system performance.

We believe that training on a larger dataset could lead to even better results, and this could be a future direction for research.

Limitations of the study and implications for educators

One of the limitations refers to the quality of the data that can influence the results obtained. The labeling process can be subjective. With the increase in the amount of data, the duration of preprocessing also increases. To train the model, we used 494 opinions labeled with a maximum score, of 4 and only 153 with a minimum score, of 0. There is an imbalance between these classes. Future research, on another data set, more voluminous and more balanced regarding the division of opinions by sentiment scores, could lead to comparative analysis and the confirmation of the hypothesis that the obtained model would have a higher classification accuracy.

Another limitation refers to the fact that choosing the best algorithm by the AutoML service is not a transparent process, which is why it is difficult to find correlations between the results and the training process, the only conclusions being limited to the data. One of the reasons for using this service is that it does not require advanced knowledge about machine learning, therefore it can be used by teachers who do not have special skills in this field. Further research could lead to the creation of a learning application in the cloud and the use of such a service to measure the degree of satisfaction of students about a certain subject of study.

The development of machine learning models for sentiment analysis from text or image data is a difficult and time-consuming task, even for experts, that's why AutoML becomes a useful alternative. The implementation of sentiment analysis can be done by using libraries such as HyperOpt SkLearn, and TPot (Mahima et al., 2021). Other authors (Lopes et al., 2021) combine text analysis with image analysis to achieve a data classification using AutoML.

Since the use of AutoML for sentiment analysis from data from the educational system is a rather underexplored topic, it can constitute a valuable direction of research for doctoral students, data scientists, and high school teachers.

References

Ardianto, R., Rivanie, T., Alkhalifi, Y., Nugraha, F. S., Gata, W. (2020). Sentiment analysis on e-sports for education curriculum using Naive Bayes and Support Vector Machine. Journal of Computer Science and Information, 13(2), 109-122, https://doi.org/10.21609/jiki.v13i2.885

Aung, K. Z., & Myo, N. N. (2017, May). Sentiment analysis of students' comment using lexicon based approach. In 2017 IEEE/ACIS 16th international conference on computer and information science (ICIS) (pp. 149-154). IEEE. https://ieeexplore.ieee.org/document/7959985

AutoML (2022). AutoML. AutoML. https://www.automl.org/automl/. Accessed 12 Oct 2022.

Bonta, V., Kumaresh, N., & Janardhan, N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. Asian Journal of Computer Science and Technology, 8(S2), 1–6. https://doi.org/10.51983/ajcst-2019.8.S2.2037

Boros, T., Dumitrescu, S. D. (2017). A "small-data"-driven approach to dialogue systems for natural language human-computer interaction, International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucharest, Romania, pp. 1-6. DOI: 10.1109/SPED.2017.7990441

Buzea, M., Trăuşan-Matu, Ş., Rebedea, T. (2019). A Three Word-Level approach used in machine learning for Romanian sentiment analysis. In Proceedings of the 18th RoEduNet Conference: Networking in Education and Research (RoEduNet) (pp. 1-6). Galati, Romania: Institute of Electrical and Electronics Engineers [IEEE]. https://doi.org/10.1109/ROEDUNET.2019.8909458

Buzea, M.-C., Trausan-Matu, S., Rebedea, T. (2020). Automatic irony detection for Romanian online news. In 2020 24th International Conference on System Theory, Control and Computing (ICSTCC) (pp. 72-77). Institute of Electrical and Electronics Engineers [IEEE]. https://doi.org/10.1109/ICSTCC50638.2020.9259715

Chapman, W. (2006). Natural Language Processing for Biosurveillance. In M., M., Wagner, A., W., Moore, R., M., Aryel. (Eds.). Handbook of Biosurveillance (pp. 255-271). Academic Press. https://doi.org/10.1016/B978-012369378-5/50019-3.

Chauhan, G.S., Agrawal, P., Meena, Y.K. (2019). Aspect-based sentiment analysis of students' feedback to improve teaching–learning process. In S., Satapathy, A. Joshi, (Eds.) Information and Communication Technology for Intelligent Systems. Smart Innovation, Systems and Technologies (vol.

107, pp: 259–266). Springer, Singapore. https://doi.org/10.1007/978-981-13-1747-7_25

Cocioceanu, A., Barbulescu, M., Ivanoaica, T., Raportaru, M., Nicolin, A. I. (2016). Testing voice-based biometrics authentication platforms for Romanian utterances through infrequent consonant clusters, 15th RoEduNet Conference: Networking in Education and Research, Bucharest, Romania, pp. 1-4, DOI: 10.1109/RoEduNet.2016.7753205

Dolianiti, F.S., Iakovakis, D., Dias, S.B., Hadjileontiadou, S., Diniz, J.A., Hadjileontiadis, L. (2019). Sentiment analysis techniques and applications in education: A survey. In M. A. Tsitouridou, J., Diniz, T., Mikropoulos (Eds.). Technology and Innovation in Learning, Teaching and Education. TECH-EDU 2018. Communications in Computer and Information Science (vol. 993, pp: 412–427). Springer, Cham. https://doi.org/10.1007/978-3-030-20954-4_31

Elsken, T., Metzen, J. H., Hutter, F. (2019). Neural architecture search: A survey. The Journal of Machine Learning Research, 20 (55), 1-21. https://jmlr.org/papers/v20/18-598.html

Gupta, N., Agrawal, R. (2020). Chapter 1 - Application and techniques of opinion mining. In S., Bhattacharyya, V., Snášel, D., Gupta, A., Khanna (Eds.). Hybrid Computational Intelligence - Hybrid Computational Intelligence for Pattern Analysis and Understanding. Academic Press, pp. 1-23. https://doi.org/10.1016/B978-0-12-818699-2.00001-9.

He, X., Zhao, K., Chu, X. (2021). AutoML: A survey of the state-of-the-art. Knowledge-Based Systems, 212, 106622.

https://doi.org/10.48550/arXiv.1908.00709

Herrera-Contreras, A. A., Sánchez-Delacruz, E., & Meza-Ruiz, I. V. (2020). Twitter opinion analysis about topic 5G technology. In M. Botto-Tobar, M. Zambrano Vizuete, P. Torres-Carrión, S. Montes León, G. Pizarro Vásquez, & B. Durakovic (Eds.), Applied Technologies (pp. 191-203). Springer, Cham. https://doi.org/10.1007/978-3-030-42517-3_15

Ilie, M. D., Ciobanu, A., Negrescu, C., Stanomir, D. (2013). Towards expressive Romanian speaking 3D avatars for multimedia interfaces, 20th International Conference on Systems, Signals and Image Processing (IWSSIP), Bucharest, Romania, pp. 47-50. DOI:10.1109/IWSSIP.2013.6623446

Karmaker, S. K., Hassan, M. M., Smith, M. J., Xu, L., Zhai, C., & Veeramachaneni, K. (2021). Automl to date and beyond: Challenges and opportunities. ACM Computing Surveys (CSUR), 54(8), 1-36.https://doi.org/10.1145/3470918.

Kastrati, Z., Imran, A. S., Kurti, A. (2020). Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs. In IEEE

Access (vol. 8, pp. 106799-106810) https://ieeexplore.ieee.org/document/9110884

Lopes, V., Gaspar, A., Alexandre, L. A. and Cordeiro, J. (2021). An AutoML-based Approach to Multimodal Image Sentiment Analysis, International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, pp. 1-9, doi: 10.1109/IJCNN52387.2021.9533552

Mahima, K. T.Y., Ginige, T.N.D.S., De Zoysa, K. (2021). Evaluation of sentiment analysis based on AutoML and traditional approaches. International Journal of Advanced Computer Science and Applications, 12 (2). http://dx.doi.org/10.14569/IJACSA.2021.0120277

Marcu, D., & Danubianu, M. (2020). Sentiment analysis from students' feedback: A Romanian high school case study. In Proceedings of the 2020 International Conference on Development and Application Systems (DAS) (pp. 204-209). Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/DAS49615.2020.9108927. http://www.dasconference.ro/cd2020/data/papers/D39-paper.pdf

Medhat, W., Hassan, A., Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011.

Mite-Baidal, K., Delgado-Vera, C., Solís-Avilés, E., Espinoza, A.H., Ortiz-Zambrano, J., Varela-Tapia, E. (2018). Sentiment Analysis in Education Domain: A Systematic Literature Review. In R., Valencia-García, G., Alcaraz-Mármol, J., Del Cioppo-Morstadt, N., Vera-Lucio, M., Bucaram-Leverone (Eds.), Technologies and Innovation. CITI 2018. Communications in Computer and Information Science, (vol 883, pp: 285–297). Springer, Cham. https://doi.org/10.1007/978-3-030-00940-3_21

Mostafa, L. (2020). Student sentiment analysis using gamification for education context. In A., Hassanien, K., Shaalan, M., Tolba (Eds.). Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2019. AISI 2019. Advances in Intelligent Systems and Computing (vol. 1058, pp: 329–339). Springer, Cham. https://doi.org/10.1007/978-3-030-31129-2_30

Nadkarni, P. M., Ohno-Machado, L., Chapman, W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18 (5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

Oludare, I. A., Aman, J., Abiodun, E. O., Kemi, V. D., Nachaat, A. M., Humaira, A. (2018). State-of-the-art in artificial neural network applications: A survey. Heliyon, 4 (11). e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Onan, A. (2021). Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach. Computer Applications in Engineering Education, 29(3), 572-589.https://doi.org/10.1002/cae.22253

Osmanoğlu, U. Ö., Atak, O. N. , Çağlar, K. , Kayhan, H., Can, T. (2020). Sentiment analysis for distance education course materials: A machine learning approach. Journal of Educational Technology and Online Learning, 3(1), 31-48. https://doi.org/10.31681/jetol.663733

Păiş, V., Ion, R., Avram, A. M., Irimia, E., Mititelu, V. B., Mitrofan, M. (2021). Human-Machine Interaction Speech Corpus from the ROBIN project, International Conference on Speech Technology and Human-Computer Dialogue (SpeD), Bucharest, Romania, 91-96. DOI: 10.1109/SpeD53181.2021.9587355

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. Knowledge-Based Systems, 108, 42-49.https://doi.org/10.1016/j.knosys.2016.06.009

Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2017). A deeper look into sarcastic tweets using deep convolutional neural networks [ArXiv]. arXiv:1610.08815. https://doi.org/10.48550/arXiv.1610.08815

Russu, R. M., Dinsoreanu, M., Vlad, O. L., & Potolea, R. (2014). An opinion mining approach for Romanian language. In 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 43-46). Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/ICCP.2014.6936978

Schuszter, I. C. (2018). Integrating deep learning for NLP in Romanian psychology. In Proceedings of the 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC) (pp. 237-244). Institute of Electrical and Electronics Engineers (IEEE). DOI: 10.1109/SYNASC.2018.00045

Sethunya, J., Kutlwano, S., Freeson, K., Hlomani, H., Keletso, L. (2016). Natural language processing: A review. International Journal of Research in Engineering and Applied Sciences, 6 (3), 207-210

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2015). Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE, 104(1), 148-175. https://ieeexplore.ieee.org/document/7352306

Siriborvornratanakul, T. (2022). Human behavior in image-based road health inspection systems despite the emerging AutoML. Journal of Big Data, 9 (96). https://doi.org/10.1186/s40537-022-00646-8

Sultana, J., Sultana, N., Yadav, K., Alfayez, F. (2018). Prediction of sentiment analysis on educational data based on deep learning approach, 21st Saudi

Computer Society National Computer Conference (NCC), Riyadh, Saudi Arabia, pp. 1-5, DOI: 10.1109/NCG.2018.8593108

Sumit, G., Santanu, M. (2017). Supervised machine learning vs. Lexicon-based text classification for sentiment analysis: A comparative study. In D. Guha, B. Chakraborty, H. S. Dutta (Eds.) Computer, Communication and Electrical Technology (pp.55-59), CRC Press, https://doi.org/10.1201/9781315400624

Sutton, R.S. (1992). Introduction: The challenge of reinforcement learning. In R.S., Sutton (Ed.) Reinforcement Learning. The Springer International Series in Engineering and Computer Science (pp. 1-3), 173. Springer. https://doi.org/10.1007/978-1-4615-3618-5_1

Thangaraj, M., Sivakami, M. (2018). Text classification techniques: A literature review. Interdisciplinary Journal of Information, Knowledge, and Management, 13, 117-135. https://doi.org/10.28945/4066

Toçoğlu, M.A., Onan, A. (2021). Sentiment analysis on students' evaluation of higher educational institutions. In: Kahraman C., Cevik Onar S., Oztaysi B., Sari I., Cebi S., Tolga A. (Eds.). Intelligent and Fuzzy Techniques: Smart and Innovative Solutions. INFUS 2020. Advances in Intelligent Systems and Computing (vol. 1197, pp: 1693–1700). Springer, Cham. https://doi.org/10.1007/978-3-030-51156-2_197

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., Ragos, O. (2020). Implementing AutoML in Educational Data Mining for Prediction Tasks. Applied Sciences. 10(1):90. https://doi.org/10.3390/app10010090

UNESCO (2018). Quick Guide to Education Indicators for SDG 4. UNESCO Institute for Statistics, Canada. http://uis.unesco.org/sites/default/files/documents/quick-guide-education-indicators-sdg4-2018-en.pdf. Accessed 18 Sep 2022.

Ungurean, C., Burileanu, D. (2011). An advanced NLP framework for high-quality Text-to-Speech synthesis, 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2011, Brasov, Romania, 1-6. DOI: 10.1109/SPED.2011.5940733

Vasiu, M. A., Potolea, R. (2020). Enhancing tokenization by embedding Romanian language specific morphology, IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 243-250. DOI: 10.1109/ICCP51029.2020.9266140.

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review article]. IEEE Computational Intelligence Magazine, 13(3), 55-75. https://doi.org/10.1109/MCI.2018.2840738

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253. https://doi.org/10.1002/widm.1253