Article

A Deep Learning Sentiment Analyser for Social Media Comments in Low-Resource Languages

Zenun Kastrati [1] [iD]**, Lule Ahmedi** [2]**, Arianit Kurti** [1],*** [iD]**, Fatbardh Kadriu** [2] [iD]**, Doruntina Murtezaj** [2] **and Fatbardh Gashi** [2]

[1] Department of Informatics, Linnaeus University, 351 95 Växjö, Sweden; zenun.kastrati@lnu.se
[2] Department of Computer Engineering, University of Prishtina, 10000 Prishtine, Kosovo;
lule.ahmedi@uni-pr.edu (L.A.); fatbardh.kadriu@student.uni-pr.edu (F.K.);
doruntina.murtezaj@student.uni-pr.edu (D.M.); fatbardh.gashi4@student.uni-pr.edu (F.G.)
* Correspondence: arianit.kurti@lnu.se

Abstract: During the pandemic, when people needed to physically distance, social media platforms have been one of the outlets where people expressed their opinions, thoughts, sentiments, and emotions regarding the pandemic situation. The core object of this research study is the sentiment analysis of peoples' opinions expressed on Facebook regarding the current pandemic situation in low-resource languages. To do this, we have created a large-scale dataset comprising of 10,742 manually classified comments in the Albanian language. Furthermore, in this paper we report our efforts on the design and development of a sentiment analyser that relies on deep learning. As a result, we report the experimental findings obtained from our proposed sentiment analyser using various classifier models with static and contextualized word embeddings, that is, fastText and BERT, trained and validated on our collected and curated dataset. Specifically, the findings reveal that combining the BiLSTM with an attention mechanism achieved the highest performance on our sentiment analysis task, with an F1 score of 72.09%.

Keywords: sentiment analysis; machine learning; deep neural network; 1D-CNN; BiLSTM; attention mechanism; Facebook comments; COVID-19

1. Introduction

Currently, the world is facing the challenges posed by the COVID-19 pandemic [1]. In the last few months, due to these changes, almost the entire population of the world has been affected in terms of their day to day operations. Nowadays, people are working, studying, shopping and socializing from a distance. The need for physical distancing has also affected peoples' emotions and their expression. Social media platforms became one of the main outlets on which people express, among others, their thoughts, sentiments, emotions and so forth, regarding the pandemic situation. Recent studies also show that social media has been one of the main channels for misinformation [2], especially during the ongoing pandemic crisis. Besides this, social media channels were considered and used by relevant Public Health Authorities for the distribution of information to the wider public [3]. Kosova, as a young country, has been following these trends. The National Institute of Public Health of Kosova has been utilizing their Facebook page to disseminate information and recommendations daily regarding the pandemic situation. These posts have created much engagement with local populations in terms of impressions and comments where general public shared their thoughts and emotions, as well as their sentiments regarding the ongoing pandemic that the country was/is going through. The social engagement on Facebook around the Public Health Institute posts created a rich and diverse set of data that captured quite well the overall public discourse and sentiments around the COVID-19 pandemic in the Albanian language.

Sentiment analysis within academia is defined as a computational examination of end user opinions, attitudes and emotions expressed towards a particular topic or event [4]. Sentiment analysis systems use various learning approaches to detect sentiment from textual data including lexicon-based [5], machine/deep learning [6,7], a combination of lexicon and machine learning [8], concept-based learning approaches [9,10], and so forth. Sentiment analysis became an important field of research for machine learning applications. Predominately, social media sentiment analysis has been one of the main field of research especially during the current COVID-19 pandemic [1]. In these studies, the prime focus has been on assessing public sentiment analysis in order to gain insights for making appropriate public health responses. Besides this, other areas where sentiment analysis is applied include, among others, election predictions [11], financial markets [12], students' reviews [13,14], and so forth, just to name a few. A common denominator of all these cases across diverse application areas shows that sentiment analysis is a valuable tool to provide accurate insights into general public opinions about particular topics of interest. The application of the sentiment analysis is closely related to the availability of the data sets that are usually related to high-resource languages. In our case of analysis, we were dealing with a relatively low-resource language, the Albanian language. Having in mind the possibility that sentiment analysis could provide insights into peoples' opinions during the pandemic and the measures taken for its prevention. Motivated by this, we designed this study, in which we created a data set and evaluated different deep learning models and conventional machine learning algorithms in a low-resource language such as the Albanian language. The main contributions of this article are as follows:

- The collection of a large-scale dataset composed of 10,742 manually classified Facebook comments related to the COVID-19 pandemic. To the best of our knowledge, this is the first study that performed sentiment analysis of Facebook comments in a low-resource language such as the Albanian language.
- A deep learning based sentiment analyser called ALBANA is proposed and validated on the collected and curated COVID-19 dataset.
- An attention mechanism is used to characterize the word level interactions within a local and global context to capture the semantic meaning of words.

The rest of the paper is structured as follows: Section 2 presents some related work, especially from the context of sentiment analysis in the Albanian language. In Section 3, we describe the methodology used to conduct the research. A description of the dataset and classifier models used to conduct the experiments is provided in Section 4. Section 5 depicts experimental results followed by their discussion presented in Section 6. The paper concludes with some future directions presented in Section 7.

2. Related Work

Albanian is spoken by over 10 million speakers, as the official language of Kosovo and Albania, but is also one of two official languages in North Macedonia, and is spoken by the Albanian community in the Balkans and the region, as well as among the large Albanians' migrants community residing mainly in European countries, America and Oceania. Albanian is an Indo-European language, an independent branch of its own, featuring lexical peculiarities distinguishable from other languages [15].

Considering the advancement of research in Natural Language Processing (NLP) and in sentiment analysis in particular for many world-spoken languages, NLP and sentiment analysis research on the Albanian language stands behind even some other low resource languages.

Sentiment analysis research for the English language has achieved significant results already, and advanced to not only adapting the latest theories in the areas of lexical analysis and machine learning (ML), but also at application level [6,7,16–19].

In [16], the literature on sentiment analysis using different ML techniques over social media data to predict epidemics and outbreaks, or for other application domains is surveyed. ML, linguistic-based and hybrid approaches for sentiment analysis are compared.

ML approaches take precedence over linguistic-based except for short sentences. Classical ML techniques such as SVM, Naive Bayes, Logistic Regression, Random Forest and Decision Trees are shown as most accurate, each for certain dataset and domain.

In [17], in an analysis about COVID-19 on 85.04M tweets from 182 countries during March to June 2020, the distribution of sentiments was found to vary over time and country, uncovering thus the public perception of emerging policies such as social distancing and remote work. Authors conclude that social media analysis for other platforms and languages is critical towards identifying misinformation and online discourse.

In [18], Facebook pages of Public Health Authorities (PHAs) and the public response in Singapore, the United States, and England from January 2019 to March 2020 are analyzed in terms of the outreach effects. Among metrics measured are mean posts per day ranging 1.4 to 5, mean comments per post ranging 12.5 to 255.3, mean sentiment polarity, positive to negative sentiments ratio ranging 0.55 to 0.94, and toxicity in comments which turned to be rare across all PHAs.

In [6], the authors seek to understand the usefulness/harm of tweets by identifying sentiments and opinions in themes of serious concerns like pandemics. The proposed model for sentiment analysis uses deep learning classifiers with accuracy up to 81%. Another proposed model bases on fuzzy logic and is implemented by SVM with an accuracy of 79%.

In [19], sentiment analysis of Tweets about Coronavirus using Naive Bayes and Logistic Regression is presented. Tweets of varying lengths, that is, less than 77 characters (small to medium) and less than 120 characters (longer) are analyzed separately. Naive Bayes performed better on classifying small to medium size Coronavirus Tweets sentiments with an accuracy of 91%. For longer Tweets, both methods showed weak performance with an accuracy not over 57%.

The reaction of people from different cultures to the Coronavirus expressed on social media and their attitudes about the actions taken by different countries is analyzed in [7]. Tweets related to COVID-19 were collected for six neighboring countries with similar cultures and circumstances including Pakistan, India, Norway, Sweden, USA and Canada. Three different deep learning models including DNN, LSTM, and CNN along with three general-purpose word embedding techniques, namely fastText, GloVE and GloVe for twitter, were employed for sentiment analysis. The best performance with an F1-score of 82.4% was achieved by LSTM with fastText.

Work on other languages concerning sentiment analysis is also growing, such as on German [20,21], Swedish [22,23], or multilingual social media posts [24]. A detailed description of the past and recent advancements on multilingual sentiment analysis conducted on both formal and informal languages used on online social platforms is explored in the survey conducted by Lo et al. in [25].

There are only a few works on sentiment analysis (opinion mining) in the Albanian language [26–28], as well as few related to sentiment analysis on emotion detection in the Albanian language [29,30].

In [26], an ML model is developed to classify documents as having positive or negative sentiment. The corpora built to develop the model consists of 400 documents covering five different topics, each topic represented by 80 documents tagged evenly with positive and negative sentiment. Six different ML algorithms, namely Bayesian Logistic Regression, Logistic Regression, SVM, Voted Perceptron, Naive Bayes and Hyper Pipes are used for classification, performing with 86% to 92% accuracy depending on the topic. The whole corpora being political news articles is characterized by a complex language and very rich technical vocabulary. The paper concludes that a larger dataset in the Albanian language is needed to achieve a high performance sentiment classifier.

In [27], a comprehensive selection of ML algorithms is evaluated for opinion mining in the Albanian language, resulting to five best performing algorithms, Logistic and Multi Class Classifier, Hyper Pipes, RBF Classifier, and RBF Network with 79% to 94% of correctly classified instances. The opinions are classified as positive or negative. The classification model is developed over a corpus of 500 newspaper articles in Albanian covering 5 different

subjects, each with a balanced set of articles with positive and negative opinions. The results varied also from subject to subject. This research is later extended from an in-domain corpus to multi-domains corpuses combining opinions from 5 different topics [28]. All the corpuses are used to train and test for opinion mining the performance of 50 classification algorithms implemented in Weka. Algorithms perform better in in-domain corpus than in multi-domain corpus. As authors state, a bigger corpus in the Albanian language could provide a clearer picture on the performance of classification algorithms for opinion mining.

In [29], a CNN sentence-based classifier is developed to classify a given text fragment into one out of six pre-defined emotion classes based on Ekman' model: joy, fear, disgust, anger, shame and sadness. Experimental evaluation shows that a deep learning model (CNN) with classification accuracy of emotions ranging from 67% to 92.4% in overall outperforms three classical classification algorithms, Naive Bayes (NB), Instance-based learner (IBK), and Support Vector Machines (SMO). Findings related to the impact of the length of text on classification are also presented. The stemming of text prior to classification improves the accuracy. Another contribution is the corpus built—some 6000 posts by politicians on Facebook in the Albanian language—to develop the model. Further, in [30], the authors extend their framework with clustering to extract representative sets of generic words for a given emotion class. The authors list deep neural network architectures that take the sequential nature of text data into account, such as LSTM, as worth considering for an emotion detection model in the future.

Our approach follows the rationale that a larger dataset is a prerequisite for developing a model that is not prone to overfitting. Recalling the classification results in the related work mentioned above, there is a huge variation in accuracy between distinct sub-datasets of a size of merely some hundreds of tagged data. Moreover, the sequential ordering of text is to be learned from, and hence considering the usage of deep neural networks for the model as well as NLP based representation techniques, that is, static and contextualized word embeddings, is unavoidable. Multi-class classification into positive, neutral, and negative sentiments is also of interest to validate the applicability of our approach not only to sentiment analysis, but also to other domains like detecting emotions or other multi-class text mining tasks (e.g., review of items in the scale 1 to 10) in the Albanian Language.

3. Methodology

The research was carried out using a quantitative research method and it was comprised of five phases including the first two that constitute human-related tasks and the remaining three phases in which a machine is involved. More specifically, the first phase entailed collecting users' posts on Facebook from the day when the first few cases were reported (13 March until 15 August). The second phase was constituted of the labeling of collected posts. A manual labeling process has taken place where three human annotators assessed the attitudes and opinions of users expressed in Facebook posts and properly classified them to either positive, neutral or negative categories.

In the third phase, a text pre-processing was performed to remove punctuation, words with length less or equal to two characters, and words that are not purely constructed of alphabetical characters from users' comments. Additionally, all text comments were converted to lowercase.

The fourth phase involved a representation model to prepare and transform the posts to an appropriate numerical format to be fed into the sentiment classifiers. A bag of word representation model with its implementations, term frequency inverse document frequency—$tf * idf$ was employed. Furthermore, we used a representation model that generated dense vector representations for words occurring in comments known as word embeddings. A static pre-trained word embedding method called fastText along with a contextualized word embeddings model—BERT, were used to learn and generate word vectors.

The final phase is constituted of the sentiment analyser, which aims to predict the sentiment of each users' comment into one of the three categories, namely positive, neutral

or negative. The analyser involves several classifiers including deep neural networks as well as conventional machine learning algorithms for sentiment orientation detection.

A high level architecture of the proposed ALBANA analyser involving all the phases elaborated above is illustrated in Figure 1.
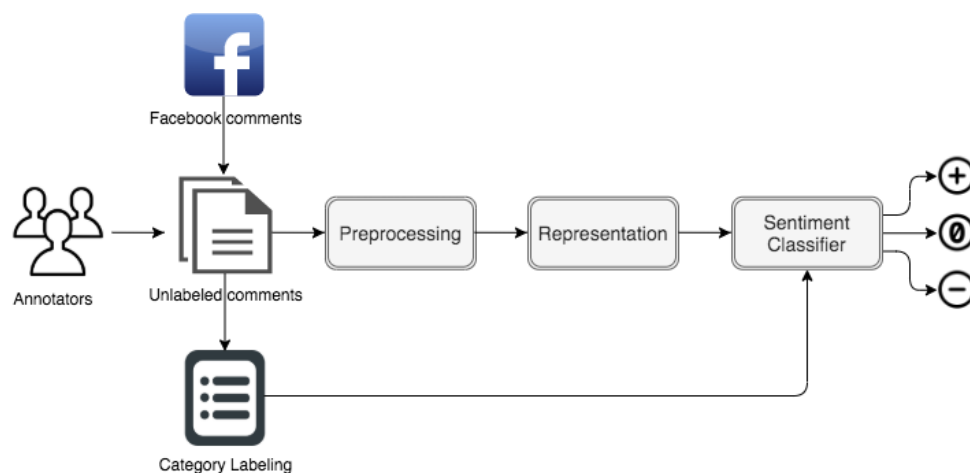

Figure 1. High-level architecture diagram of the proposed ALBANA analyser.

4. Experiments

This section describes the data collection and annotation procedure applied to creating the dataset as well as the classifier models used to conduct the sentiment classification task.

4.1. Dataset

The dataset consists of people's opinions expressed towards daily Facebook posts of the National Institute of Public Health of Kosova (NIPHK) (https://www.facebook.com/IKSHPK, accessed on 5 April 2021) regarding the spread of the COVID-19 virus in the Republic of Kosova. Dataset creation involved data collection and annotation, which are described in the following subsections.

4.1.1. Dataset Collection

We collected comments from the official Facebook page of the NIPHK Institute for a period of 6 months, from 13 March till 15 August 2020. 13th of March marked the confirmation of first cases of COVID-19 in Kosova. To retrieve comments, we used CommentExporter (http://www.commentexporter.com, accessed on 10 December 2020). This is a tool that allows to export the original comments to an Excel file. The open source version of this tool is limited to a maximum of 300 comments to export in one usage excluding replied comments. Due to this limitation, there are few days (e.g., 27 July) during this 6 months time period that are not included as the number of comments was over 300. Additionally, there are few days (e.g., 15 March) that are also missing because there was no official announcement (no post) from the NIPHK Institute. The total number of collected comments is 10,742 and this constitute the first version of the dataset referred as version 1.0. Dataset 1.0 contains the unlabeled comments and some metadata information as illustrated in Figure 2.

There were two other following versions until the final dataset was created. The second version was updated by adding two new manually extracted features related to the post: post's timestamp and the URL of the post. In the third version, we added three more post-related features including number of deaths, number of infected persons and number of healed persons for the day when the post was published. These three features were also manually extracted from the content of each post.

Id	Profile name	Comment	Like	Timestamp
238		IKSHP se pari ju faleminderojm per punen e vazhdus	0	Saturday, March 21, 2020
239		Faleminderit shumë për punën tuaj!	0	Saturday, March 21, 2020
240		zoti ju ndihmoft ju dhe ne tgjithve !	0	Saturday, March 21, 2020
241		IKSHP ju lutem ju lutem siq edhe ju keni lutje ndaj ne	1	Saturday, March 21, 2020
242		Pasi indentifikohen rastet pozitiv si veproni me indiv	0	Saturday, March 21, 2020
243		Ju lutem tregojeni lokacionin e dy grave ne Prishtine	1	Saturday, March 21, 2020

Figure 2. Dataset 1.0—Unlabeled comments.

The third version of the dataset evolved to the final dataset by labeling the comments from human annotators. In order to avoid the bias and make the labeling process more objective, all comments were annotated by three human annotators who were third year bachelor students at Computer Engineering department in University of Prishtina. Then, a majority voting was applied to get the final sentiment label for each comment. The inter-annotator agreement determined by computing the Pearson's correlation between the scores given by each annotator is depicted in Table 1. The correlation coefficients indicate a strong agreement between annotator 1 and annotators 2 and 3, whereas a moderate agreement between annotator 2 and annotator 3.

Table 1. The Pearson's correlation among three annotators.

	Annotator 1	Annotator 2	Annotator 3
Annotator 1	1.00	0.57	0.62
Annotator 2	0.57	1.00	0.46
Annotator 3	0.62	0.46	1.00

Table 2 shows few examples labeled as neutral, positive and negative for which a perfect agreement among the three annotators is achieved.

Table 2. Examples of comments annotated with a perfect agreement among annotators.

Comment (English Translation)	Sentiment
Do te thot Peja edhe sonte spaska asnje rast (It means that even tonight Peja does not have any case)	Neutral
Bravo ekipet e IKShP per punen e shkelqyeshme dhe perkushtimin! (Well done the NIPHK teams for the great job and dedication!)	Positive
Keni kalu tash ne monotoni, te pa arsyshem jeni tash. (You have now passed into monotony, you are now unreasonable.)	Negative

The most challenging part when it comes to comment labeling was the assignment of the sentiment to comments expressing people's opinions on various topics/entities. The sentiment was assigned to the entire comment and no analysis of entities/sentences in the comment was carried out. For example, the comment

> *"Comment No 894: Juve stafit mjeksor respekt ndersa ktyre qe jane raste kontakti qe nuk kane nejt ne shtepi po kane shku musafir e jone infektu turp! (Respect to the medical staff while shame on contact cases who have not stayed at home but have visited relatives and got infected!)"*

has a positive initial sentence and expresses positive sentiment towards the medical staff whereas the second part of the comment expresses negative sentiment towards contact cases who are not isolated and got infected. This comment contains both positive and negative sentiments and it was typically annotated differently by human annotators.

Another challenging aspect for the annotators was the labeling of comments comprising figurative language such as sarcasm and irony. Figurative language is very contextual, environmental, and topic reliant, and this caused difficulties for the annotators to find

people's actual sentiment expressed in the comment, and as a result, the given comment might have been annotated differently by annotators. For instance, the comment

> "*Comment No 5570: Bile diqka me pas pozitive në këtë Kosovë. (At least to have something positive in Kosova.)*"

is a sarcastic expression that without any contextual clues might have been understood and annotated differently by annotators.

The final dataset, whose screen shot is illustrated in Figure 3, contains 13 attributes that are described in the following:

- **ID**—unique identification number for each comment
- **Comment**—the content of the comment where the sentiment analysis is performed
- **Like**—the number of Facebook reactions to the relevant comment
- **CMNT's TS**—comment's timestamp shows the day, date and time of the comment
- **Post's TS**—post's timestamp indicates the day, date and time of the post to which the comment belongs
- **Post's URL**—post's link of the relevant comment
- **#Deaths**—number of persons who have died due to the pandemic in that day
- **#Infected**—number of infected persons with COVID-19
- **#Healed**—number of people who are recovered from COVID-19
- **Annot 1**—annotator 1
- **Annot 2**—annotator 2
- **Annot 3**—annotator 3
- **Final annotation**—the final sentiment of the comment derived through a majority voting strategy.

The final dataset is available on GitHub (https://github.com/lule-ahmedi/AlbAna, accessed on 8 April 2021).

Id	Comment	Like	CMNT's TS	Post's TS	Post's URL	#Deaths	#Infected	#Healed	Annot 1	Annot 2	Annot 3	Final annotation
238	IKSHP se pari ju faleminderojm per pune	0	Saturday	Saturd:	https:/	0	4	0	1	1	1	1
239	Faleminderit shumë për punën tuaj!	0	Saturday	Saturd:	https:/	0	4	0	1	1	1	1
240	zoti ju ndihmoft ju dhe ne tgjithve !	0	Saturday	Saturd:	https:/	0	4	0	1	1	1	1
241	IKSHP ju lutem ju lutem siq edhe ju keni	1	Saturday	Saturd:	https:/	0	4	0	1	1	0	1
242	Pasi indentifikohen rastet pozitiv si vepr	0	Saturday	Saturd:	https:/	0	4	0	2	0	0	0
243	Ju lutem tregojeni lokacionin e dy grave	1	Saturday	Saturd:	https:/	0	4	0	0	0	0	0
244	Jeni vonu per ORE POLICORE shqiptari c	1	Saturday	Saturd:	https:/	0	4	0	2	0	2	2
245	Ju lutem na tregoni ne cillen pjese te Pri	1	Saturday	Saturd:	https:/	0	4	0	0	0	0	0

Figure 3. Final dataset.

4.1.2. Dataset Statistics

As described in the previous section, the curated dataset contains three classes that, along with the users' comments, assigned to them are depicted in Table 3. The distribution statics show that the dataset is highly imbalanced with the neutral sentiment class comprising more than half of the comments (56.4%), followed by negative comments and positive comments with 28.0% and 15.6%, respectively.

Table 3. Number of users' comments among sentiment classes.

Class	# of Comments	Percent (%)
Positive	1677	15.6
Negative	3010	28.0
Neutral	6055	56.4
Total	10,742	100.0

It is also interesting to note that the dataset contains comments of various length. The shortest comment is composed of 1 word and the longest comprises 212 words. The average length of comments from the entire corpus is 16.01 words. The length variation of

comments constituting our dataset is depicted in Figure 4, where the histogram diagram illustrates the number of words per comment distributed by sentiments. As can been seen in Figure 4, the negative comments are generally the longest, with an average length of 22.68 words per comment. More specifically, it is only one comment in negative class which is below the average length of comments in the entire corpus. On the other side, the neutral and positive comments seem to be shorter, with an average length of 13.65 and 12.59 words per comment, respectively.
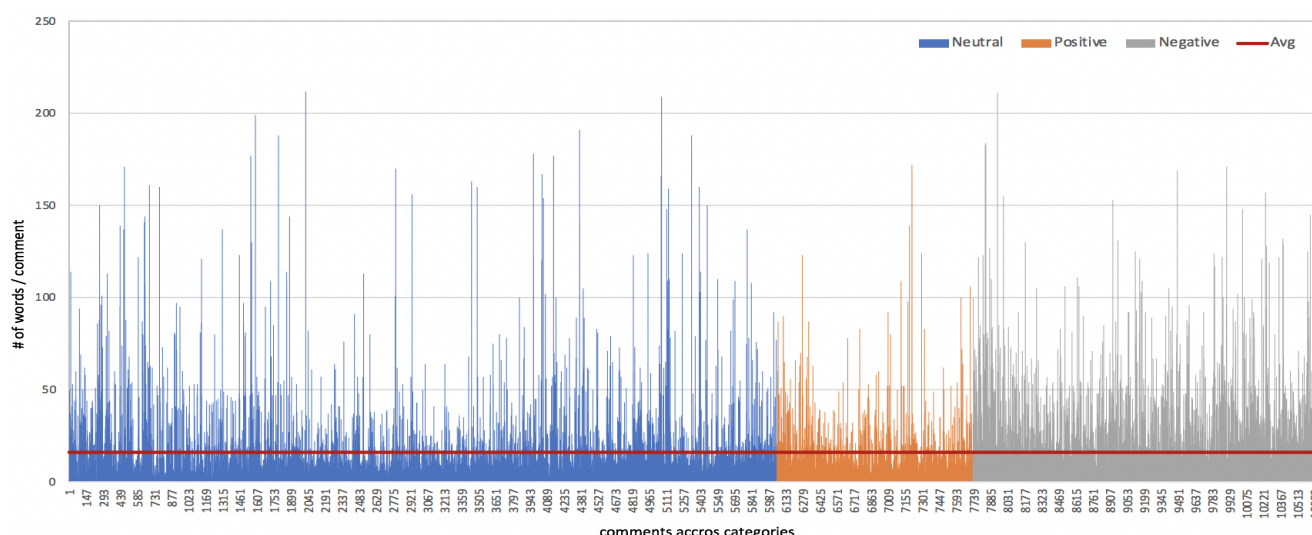

Figure 4. Length of comments among sentiment classes.

Figure 5 depicts the number of comments distributed across the months. As can be seen from the graph, there was an increasing trend of comments related to COVID-19 disease during April and July. This trend is seen for all three types of comments (neutral, positive and negative). It is also very interesting to note that during June and July there is a significant increase in negative comments. An explanation for this is that the first wave of the COVID-19 pandemic hit Kosova during this time with new cases and the death toll growing rapidly.
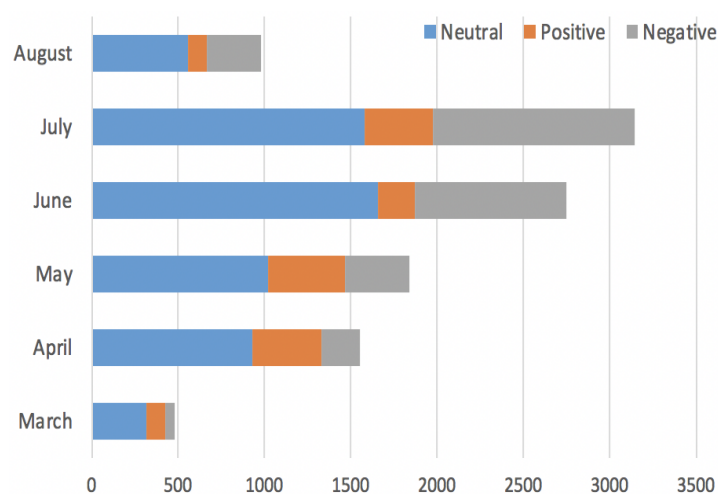

Figure 5. Number of comments across months.

4.2. Deep Neural Networks

To identify the opinion orientation of users towards the COVID-19 pandemic expressed on Facebook comments, we employed three different deep neural networks, namely 1D-CNN, BiLSTM, and a hybrid 1D-CNN + BiLSTM model, as depicted in Figure 6. We

have chosen these networks due to the different nature in their text modeling capabilities. More specifically, 1D-CNN has the ability to extract local features from the comment, BiLSTM is good at capturing contextual information from both direction as well as the long-range dependencies, and the hybrid model that takes the advantages of both complementary 1D-CNN and BiLSTM architectures.
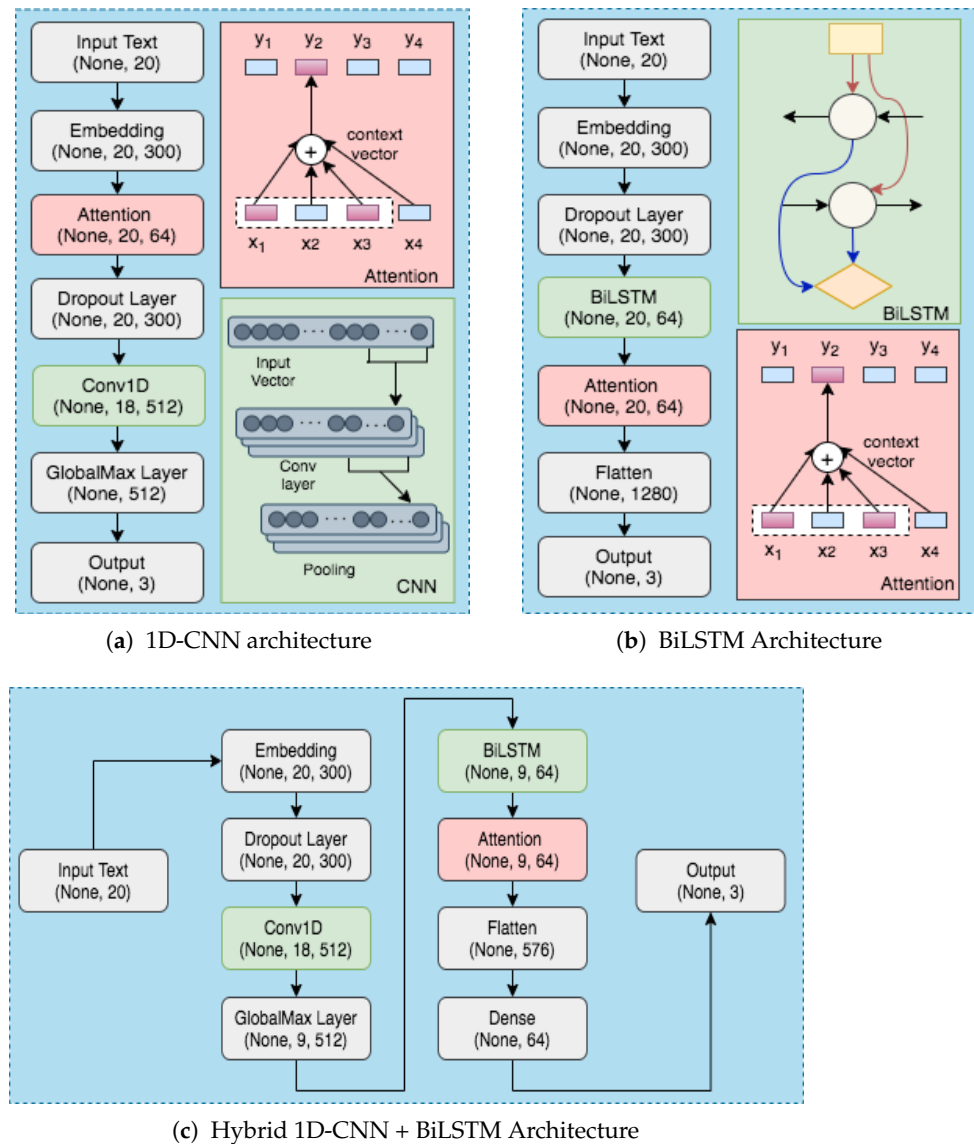

(**a**) 1D-CNN architecture

(**b**) BiLSTM Architecture

(**c**) Hybrid 1D-CNN + BiLSTM Architecture

Figure 6. Deep neural networks used for sentiment classification.

The architecture of 1D-CNN consists of an input, an output and 5 hidden layers, as shown in Figure 6a. The input layer takes a textual comment padded to a fixed length of 20 words, followed by an embedding layer comprising word embeddings of size 300D. Next comes an attention layer that aims to extract high level feature vectors. The attention layer is a sub-unit comprised of context vectors that align the source input denoted by $x_1, x_2,...,x_n$, and target output indicated by $y_1, y_2,...,y_n$. An illustration of the attention mechanism is shown on the top right corner in Figure 6a. Feature vectors extracted from the attention layer serve as inputs to the SpatialDropout1D layer. A Conv1D layer (bottom right) with 512 1D convolution filters of size 3 and a *ReLU* activation function is applied on top of the dropout layer. Finally, a fully-connected dense layer composed of a *softmax* function and 3 units is used to compute the probability distribution over three sentiment orientations (positive, neutral, negative).

The second network applied to detect the opinion orientation of Facebook users is a BiLSTM architecture, as illustrated in Figure 6b. This network architecture is slightly different from the one shown in Figure 6a, where Conv1D and GlobalMax layers are replaced with BiLSTM and Flatten layers, respectively. Similar to 1D-CNN, an illustration of BiLSTM architecture and the attention mechanism is shown in the right side of Figure 6b.

The third network architecture, illustrated in Figure 6c, constitutes a hybrid model that takes the advantages of the two complementary deep neural models, 1D-CNN and BiLSTM, and the attention mechanism by combining them into a single unified architecture. Specifically, 1D-CNN layer will be applied on top of the embedding layer to capture local features such as n-grams. These features will serve as the inputs to the BiLSTM layer which will be used to model the contextual information from both directions (backward and forward), and capture the long-range dependencies in the comment. Then, an attention layer was applied to the outputs of BiLSTM to capture important information by assigning different weights to different words in the comment. Finally, an output layer proceeded by a dense layer used to map extracted features to a more separable space is applied.

4.3. Conventional Machine Learning Models

This section briefly discusses the conventional machine learning models employed in this study for sentiment classification. The models include Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), and Random Forest (RF).

SVM is a classifier, which can be either a parametric or non-parametric model depending on the linearity property. Linear SVM is parametric as it contains a fixed size of parameters derived by the weight coefficient whereas non-linear SVM can be considered as non-parametric due to the kernel matrix that is created by calculating the pair-wise distances of two feature vectors.

NB is a parametric classifier that applies a statistical-based model to learn the underling function from the training data. The learning function is characterized by a fixed number of parameters defined by the Bayes rule. As indicated by the name, the 'naive' part of this classifier comes from the assumption of strong independence of features and the class variables.

DT is a non-parametric classifier that does not use any parameters to learn the underlying probability density function. It employs a tree structure model to perform the classification and it uses the information provided by training samples alone. Overfitting is the major limitation of DT and to overcome this limitation, multi-classifiers systems such as Random Forest have emerged.

RF belongs to the family of multi-classifier systems as it combines multiple decision trees into a single unique architecture. In contrast to single decision tree classifiers which can not handle the noise properly, RF is robust to noise and outliers due to its randomness property which reduces the variance between the different decision trees.

5. Results

This section provides the experimental results obtained from various sentiment classifiers trained and validated on our collected dataset.

5.1. Parameter Settings

All deep neural networks employed for sentiment classification were implemented using Python open-source software library called Keras (https://keras.io, accessed on 5 April 2021). Scikit-learn (https://scikit-learn.org/stable/, accessed on 2 April 2021), a simple and efficient tool in Python, is used for developing conventional machine learning algorithms. BERT model is implemented in the Peltarion (https://peltarion.com/platform, accessed on 7 April 2021) operational AI platform. The maximum number of words to be used in the tokenizer model was set to 20,000 and the input comment sequence is padded to 20 words.

The following hyper-parameters were used to conduct the experiments. The training batch size of 256, Adam stochastic optimizer with the learning of 0.001, categorical cross-entropy was used as the loss function, and an accuracy metric to detect the convergence of the models. The number of epochs used to train and validate the model was set to 15.

In order to avoid overfitting in our deep neural networks, we used a dropout strategy where certain units (neurons) along with their incoming and outgoing connections were temporarily removed from the network models. Dropout prevents model units from co-adapting too much on the training data and thus it led to better generalization on the testing set as well [31] . In our case, the dropout rate was set to 0.3.

The dataset was split into three sets: 70% for training, 15% for validation and the remaining 15% for testing.

5.2. Our Baseline Model

First, we established a baseline model which is a simple deep neural network (DNN) with an architecture similar to the one reported in the research work in [7]. Specifically, the DNN architecture shown in Figure 7 consists of an embedding layer with 300 dimensions, a GlobalMaxPooling layer, three dense layers with 128, 64, 32 units and *ReLU* as an activation function, and the output layer using 3 units and a *softmax* function.

```
Layer (type)                    Output Shape              Param #
=================================================================
embedding_1 (Embedding)         (None, 20, 300)           6776700

global_max_pooling1d_1 (Glob    (None, 300)               0

dense_1 (Dense)                 (None, 128)               38528

dense_2 (Dense)                 (None, 64)                8256

dense_3 (Dense)                 (None, 32)                2080

dense_4 (Dense)                 (None, 3)                 99
=================================================================
Total params: 6,825,663
Trainable params: 6,825,663
Non-trainable params: 0
```

Figure 7. Baseline model summary.

The sentiment classification performance obtained from the baseline model in terms of precision, recall, and F1 score, is given in Table 4.

Table 4. Performance of the Baseline model.

Class	P	R	F1
Neutral	78.92	70.58	74.52
Positive	51.77	69.27	59.26
Negative	63.92	65.24	64.57
Weighted avg	70.43	68.88	69.32

5.3. Deep Neural Networks

In the next set of experiments, we have conducted sentiment analysis using the deep neural networks described in Section 4.2.

5.3.1. Attention Mechanism

In this section, we examine the effect of the attention mechanism on capturing the long range dependencies in the collected comments. For this purpose, an attention layer considering a global and local context is used on top of BiLSTM to extract the high-level features. Global context characterizes the entire comments and it is too broad. Local context

is defined from a small window of different sizes. In our case, we tested local context of various sizes, from 2 up to 10 words, as illustrated in Figure 8. It is worth noting that the classification performance increases by increasing the context width. The increasing trend continues up to a window size of 8 when the highest performance is achieved. The performance gradually degrades as we continue to increase the window size more than 8 words. Therefore, we chose the window size of 8 words as the optimal context to extract semantic features using the attention mechanism when we conducted the rest of the experiments on sentiment classification task.
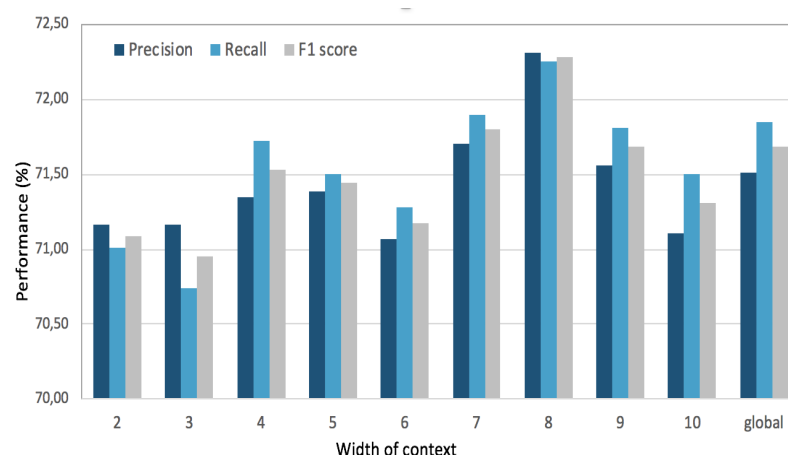

Figure 8. Performance of BiLSTM with respect to various sizes of context.

5.3.2. 1D-CNN w/o Attention Mechanism

Next, we investigated the effect of integrating attention mechanism into a one dimensional convolution neural network (1D-CNN). The network integrates the attention layer to obtain high-level features of the reviews to train the sentiment classification model. To show the benefit of this mechanism we report side by side the results exhibited by 1D-CNN with and without attention in Table 5. The results show that 1D-CNN with attention (1D-CNN + Att) generally outperforms the 1D-CNN model in sentiment classification achieving an F1 score of 71.56%. It is interesting to notice that a more substantial improvement is achieved by 1D-CNN + Att model on a positive class where the F1 score is increased from 63.85% to 67.16%.

Table 5. Performance of 1D-CNN w/o attention.

Class	1D-CNN			1D-CNN + Att		
	P	**R**	**F1**	**P**	**R**	**F1**
Neutral	74.89	80.44	77.57	75.92	79.57	77.71
Positive	72.60	56.98	63.85	72.12	62.85	67.16
Negative	61.83	60.16	60.98	62.44	60.95	61.69
Weighted avg	70.88	71.05	70.76	71.55	71.72	71.56

5.3.3. BiLSTM w/o Attention Mechanism

This section focuses on examining the performance of BiLSTM in sentiment classification task. Specifically, we conducted experiments with two different classification settings in terms of the network architecture used. The first setting consists of an BiLSTM architecture with an embedding layer and a dense layer. This network architecture extended with an attention layer integrated on top of BiLSTM constitutes the second classification setting. The obtained results of both architectures with respect to precision, recall and F1 score are summarized in Table 6.

Table 6. Performance of BiLSTM w/o attention.

Class	BiLSTM			BiLSTM + Att		
	P	**R**	**F1**	**P**	**R**	**F1**
Neutral	74.74	80.76	77.63	76.01	80.21	78.05
Positive	76.92	55.87	64.72	76.51	60.06	67.29
Negative	60.38	60.00	60.19	62.48	63.17	62.83
Weighted avg	71.08	71.01	70.71	72.31	72.25	72.09

5.3.4. Hybrid Model w/o Attention Mechanism

Next, we investigated the effect of a hybrid model where two complementary deep networks, 1D-CNN and BiLSTM, were combined into one unified architecture for sentiment classification. Like two architectures described in Sections 5.3.2 and 5.3.3, the same classification settings with respect to using the attention mechanism were explored. Performance of the hybrid model achieved by exploiting or not the attention mechanism is summarized in Table 7. Results show that the best overall and class-wise performance was achieved by the model on sentiment classification task when the attention mechanism was applied.

Table 7. Performance of hybrid model w/o attention.

Class	Hybrid			Hybrid + Att		
	P	**R**	**F1**	**P**	**R**	**F1**
Neutral	78.32	72.95	75.54	75.43	79.89	77.59
Positive	65.63	59.22	62.26	71.15	60.61	65.46
Negative	58.11	69.37	63.24	63.16	60.95	62.04
Weighted avg	70.67	69.77	70.00	71.32	71.54	71.32

### 5.3.5. Static Word Embeddings

In this section, we analyze the effect of general-purpose pre-trained word embeddings on the sentiment classification task. More specifically, we used 300d pre-trained word vectors trained on the free online encyclopedia Wikipedia and data from the common crawl projects in the Albanian language using fastText model. These vectors were fed to the four different neural networks, namely DNN, 1D-CNN, BiLSTM, and Hybrid model. The results summarized in Table 8 show that 1D-CNN with an F1 score of 70.45% achieved the best classification performance followed by BiLSTM and Hybrid model with an F1 score of 68.95% and 68.18%, respectively. Overall, the results are slightly worse compared to the ones shown in Tables 5–7 and one possible explanation for this is the out of vocabulary issue. Even though fastText handles this problem to some extent by taking a character n-gram level representation, we still have a high number of null word embeddings, 7852 out of 22,859 unique tokens. FastText is trained on documents that are generally written using a standard language whereas our dataset consists of Facebook posts which are written without any regard to the general-rules and standards of the Albanian language, that is, spelling mistakes, and also contain non-standard words and phrases such as slang abbreviations, emojis, and so forth.

Table 8. Performance of deep neural models and the baseline using fastText.

Class	DNN			1D-CNN			BiLSTM			Hybrid		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Neutral	73.32	66.96	69.99	72.52	85.33	78.41	71.97	82.02	76.67	72.23	80.21	76.01
Positive	58.19	48.60	52.97	69.49	57.26	62.79	75.32	48.60	59.08	73.17	50.28	59.60
Negative	50.06	63.49	55.98	68.87	51.27	58.78	61.55	56.67	59.01	58.64	56.03	57.31
Weight avg	64.42	63.08	63.38	71.02	71.37	70.45	69.59	69.64	68.95	68.58	68.71	68.18

5.3.6. Contextualized Word Embeddings

In the same way as in Section 5.3.5, we investigated the effect of contextualized word embeddings on the sentiment classification task. In particular, we employed mBERT model as illustrated in Figure 9. The model architecture (https://tinyurl.com/3c2vk2zf, accessed on 7 April 2021) comprises an input layer representing textual comments coming into the BERT tokenizer layer that converted them into tokens. A sequence of 128 tokens was then fed to the Multilingual BERT—mBERT encoder layer. The encoder in mBERT is an attention-based architecture composed of 12 successive transformer layers trained on Wikipedia pages with shared vocabulary across 104 languages including the Albanian language. The output of mBERT layer is a vector that was passed to a dense layer with *softmax* function to predict the comment into one of the three opinion classes, that is, positive, neutral and negative.
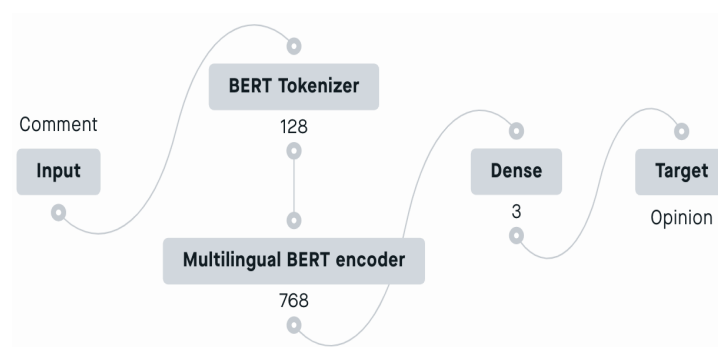
Figure 9. Sentiment classifier using mBERT.

The class-wise and weighted average performance of the mBERT model with respect to precision, recall and F1-score obtained from our dataset is summarized in Table 9.

Table 9. Performance of sentiment classifier using mBERT.

Class	P	R	F1
Neutral	81.13	70.38	75.37
Positive	64.35	79.87	71.27
Negative	57.77	69.16	62.95
Weighted avg	68.43	75.58	71.35

Figure 10 depicts a confusion matrix arising out of the sentiment classifier using mBERT. A quick glimpse at the confusion matrix shows that a better class-wise performance is achieved when using mBERT model. Specifically, it is the negative sentiment class (minority class), in which 69.20% of the comments are correctly classified using mBERT model compared to 63.9%, 62.44%, 62.48%, and 63.16% of comments that are correctly classified by Baseline, 1D-CNN + Att, BiLSTM + Att, and Hybrid + Att, respectively.
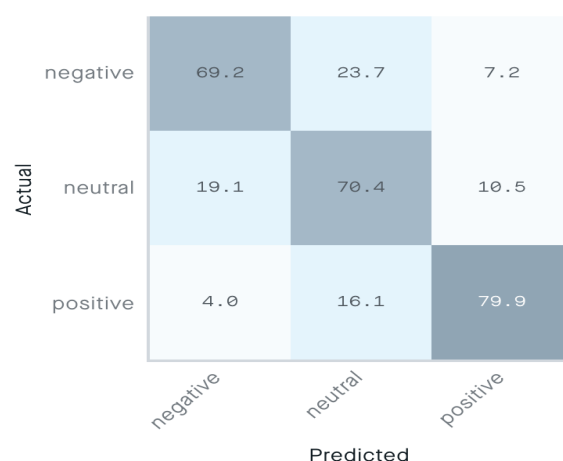
Figure 10. Confusion matrix of sentiment classifier using mBERT.

5.4. Conventional Machine Learning Models

In the second round of the experiments, we analyzed the performance of conventional machine learning (CML) models using Bag-of-Words (BoW) representation on sentiment classification task. The CML models include four classifiers described in Section 4.3. Two BoW implementations, namely count occurrence tf and term frequency inverse document frequency $tf * idf$, are employed as feature representations to feed the CML classifiers. Parameter values for all CML classifiers were set to default besides for RF where maximum depth of the tree was set to 200. The obtained results with respect to weighted precision, recall, and F1 score, are summarized in Tables 10 and 11.

Table 10. Performance of CML models using tf representations.

Class	SVM			NB			DT			RF		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Neutral	72.09	82.89	77.11	76.40	76.65	76.52	71.40	72.56	71.98	72.59	86.34	78.87
Positive	71.22	55.31	62.26	75.99	60.55	67.40	54.47	55.23	54.85	77.19	60.75	67.99
Negative	64.42	53.37	58.26	58.09	64.51	61.13	50.95	48.81	49.85	64.63	46.99	54.41
Weight avg	69.81	70.21	69.49	71.34	70.80	70.89	63.16	63.36	63.25	71.14	71.58	70.49

Table 11. Performance of CML models using $tf * idf$ representations.

Class	SVM			NB			DT			RF		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Neutral	72.15	87.85	79.23	70.02	92.71	79.78	70.82	75.56	73.11	71.81	90.15	79.94
Positive	83.26	52.79	64.62	90.51	45.17	60.26	61.63	53.85	57.47	84.52	56.02	67.38
Negative	67.84	52.22	59.01	72.86	44.60	55.33	52.68	49.15	50.85	70.40	46.53	56.03
Weight avg	72.71	72.34	71.27	74.02	72.11	70.04	64.43	64.94	64.58	73.43	72.88	71.44

As can be seen from Tables 10 and 11, a better classification performance is achieved when $tf * idf$ vectorizer was used to generate textual features to be fed to CML classifiers compared to the performance of CML classifiers using features extracted from the count vectorizer. It is also interesting to note that RF outperformed all the CML classifier models in both classification settings, achieving an F1 score of 70.49% and 71.44% using tf and $tf * idf$, respectively.

A summary of the results for all the classifier models using various embeddings including domain, static and contextualized embeddings, that is, FastText, mBERT, as well as distribution embeddings, that is, tf, $tf * idf$, is depicted in Table 12. As highlighted in

Table 12, BiLSTM with an attention mechanism and word embeddings generated from our collected dataset outperforms the other classifier models, achieving an F1 score of 72.09%.

Table 12. Summary of the results for all the classifier models trained and validated on our dataset.

Model	Embeddings	F1-Score
DNN	Domain	69.32
	FastText	63.38
1D-CNN	Domain	70.76
	FastText	70.45
1D-CNN + Att	Domain	71.56
BiLSTM	Domain	70.71
	FastText	68.95
BiLSTM + Att	**Domain**	**72.09**
Hybrid	Domain	70.00
	FastText	68.18
Hybrid + Att	Domain	71.32
BERT	mBERT	71.35
SVM	tf	69.49
	$tf*idf$	71.27
NB	tf	70.89
	$tf*idf$	70.04
DT	tf	63.25
	$tf*idf$	64.58
RF	tf	70.49
	$tf*idf$	71.44

6. Discussion

Based on the experimental results provided in Section 5, deep learning models (1D-CNN, BiLSTM, Hybrid, and BERT) generally perform better than conventional machine learning models (SVM, NB, DT, RF). This can be attributed to the capabilities of deep neural networks on modeling textual comments. 1D-CNN is a classifier model which is good at identifying local features in the comments regardless of their position. The model also applies pooling to reduce the output dimensionality and extract the most salient features. On the other hand, BiLSTM is a classification model that has the ability to capture contextual information in both forward and backward directions and to learn long-range dependencies from the comments. Hybrid classifier model takes the advantages of both complementary 1D-CNN and BiLSTM architectures whereas BERT classifier is capable to understand the meaning of each word using a bidirectional strategy and attention mechanism.

It is also interesting to note that the performance of all the deep learning models is improved using the attention mechanism. This mechanism is used to explicitly make the classifiers more robust for understanding the semantic meaning of each word within a local or global context. The empirical data (Figure 8) showed that the local context works better than the global one in our case.

Another interesting fact that can be observed from the experimental results is a better class-wise performance achieved from deep learning classifiers compared to conventional machine learning models. A significant improvement is evident in classes with small numbers of comments. More specifically, the neutral class registered an average F1 score of 62.38% when deep learning classifiers with attention mechanism and domain embeddings are applied for sentiment classification compared to an average F1 score of 55.91% obtained from conventional classifiers with $tf*idf$ distribution embeddings.

Despite better performance of deep learning classifiers on our sentiment classification task, there are still a few advantages to using conventional machine learning models. One advantage is that these models are financially and computationally cheap as they can run on decent CPU and do not require very expensive hardware such as GPU and TPU. Another advantage of conventional classifier models is the interpretability. These models are easy to interpret and understand as they involve direct feature engineering in contrast to deep learning models which extract features automatically.

In general, the results are inspiring given the fact that the Albanian language is considered a resource-constrained language and it faces many challenges when it comes to natural language processing tasks in general, and in sentiment analysis in particular. These challenges involve both technical and linguistic related aspects. From a technical point of view, systems for sentiment analysis of Albanian text face a scarcity of NLP tools and techniques such as tools for text stemming and lemmatization, list of stop words etc. From a linguistic perspective, there are various aspects which affect the performance of sentiment analysis systems for the Albanian language including negations (explicit and implicit) slang words/acronyms, figurative language (sarcasm, irony), etc.

7. Conclusions and Future Work

This article presented a sentiment analyser for extracting opinions, thoughts and attitudes of people expressed on social media related to the COVID-19 pandemic. Three deep neural networks utilizing an attention mechanism and a pre-trained embedding model, that is, fastText, are trained and validated on a real-life large-scale dataset collected for this purpose. The dataset consisted of users' comments in the Albanian language posted on NIPHK Facebook page during the period of March to August 2020. Our findings showed that our proposed sentiment analyser performed pretty well, even outperforming the baseline classifier on the collected dataset. Specifically, an F1 score of 72.09% is achieved by integrating a local attention mechanism with BiLSTM. These results are very promising considering the fact that the dataset is composed of social media user-generated reviews which are typically written in an informal manner—without any regard to standards of the Albanian language and also consisting of informal words and phrases like slang words, emoticons, acronyms, etc. The findings validated the usefulness of our proposed approach as an effective solution for handling users' sentiment expressed in social media in low-resource languages.

In future work, we will focus on studying more colloquial textual data on social platforms like Twitter, Instagram, and so forth, and propose deep learning models that can be enriched with semantically rich representations [32] for effectively extracting peoples' opinions and attitudes. Another interesting aspect that will be investigated in the future is using emojis (emoticons) as an input data because they are also an effective way to express people's emotions and attitudes towards a certain event. Furthermore, the collected dataset is highly imbalanced with the neutral class having more than half of the comments, thus future work will concentrate on applying data balancing strategies including synthetic data generation and oversampling techniques, that is, SMOTE, as well as text generation models such as GPT-2.

References

1.   Wang, T.; Lu, K.; Chow, K.P.; Zhu, Q. COVID-19 Sensing: Negative sentiment analysis on social media in China via Bert Model. *IEEE Access* **2020**, *8*, 138162–138169. [CrossRef]
2.   Allington, D.; Duffy, B.; Wessely, S.; Dhavan, N.; Rubin, J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychol. Med.* **2020**, 1–7. doi:10.1017/S003329172000224X. [CrossRef] [PubMed]
3.   Chan, A.K.; Nickson, C.; Rudolph, J.; Lee, A.; Joynt, G. Social Media for Rapid Knowledge Dissemination: Early Experience from the COVID-19 Pandemic. *Anaesthesia* **2020**, *75*, 1579–1582. [CrossRef]
4.   Singh, N.K.; Tomar, D.S.; Sangaiah, A.K. Sentiment analysis: A review and comparative analysis over social media. *J. Ambient Intell. Humaniz. Comput.* **2020**, *11*, 97–117. [CrossRef]
5.   Rajput, Q.; Haider, S.; Ghani, S. Lexicon-based sentiment analysis of teachers' evaluation. *Appl. Comput. Intell. Soft Comput.* **2016**, *2016*, 2385429. [CrossRef]
6.   Chakraborty, K.; Bhatia, S.; Bhattacharyya, S.; Platos, J.; Bag, R.; Hassanien, A.E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Appl. Soft Comput.* **2020**, *97*, 106754. [CrossRef]
7.   Imran, A.S.; Daudpota, S.M.; Kastrati, Z.; Batra, R. Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets. *IEEE Access* **2020**, *8*, 181074–181090. [CrossRef]
8.   Chauhan, G.S.; Agrawal, P.; Meena, Y.K. Aspect-based sentiment analysis of students' feedback to improve teaching–learning process. *Inf. Commun. Technol. Intell. Syst.* **2019**, *12*, 259–266.
9.   Vilares, D.; Peng, H.; Satapathy, R.; Cambria, E. BabelSenticNet: A Commonsense Reasoning Framework for Multilingual Sentiment Analysis In Proceedings of the IEEE Symposium Series on Computational Intelligence, Bangalore, India, 18–21 November, 2018; pp. 1292–1298.
10.   Stappen, L.; Baird, A.; Cambria, E.; Schuller, B. Sentiment Analysis and Topic Recognition in Video Transcriptions. *IEEE Intell. Syst.* **2021**, *36*, 1–8.
11.   Chauhan, P.; Sharma, N.; Sikka, G. The emergence of social media data and sentiment analysis in election prediction. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 2601–2627. [CrossRef]
12.   Carosia, A.; Coelho, G.P.; Silva, A. Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media. *Appl. Artif. Intell.* **2020**, *34*, 1–19. [CrossRef]
13.   Kastrati, Z.; Imran, A.S.; Kurti, A. Weakly supervised framework for aspect-based sentiment analysis on students' reviews of moocs. *IEEE Access* **2020**, *8*, 106799–106810. [CrossRef]
14.   Kastrati, Z.; Arifaj, B.; Lubishtani, A.; Gashi, F.; Nishliu, E. Aspect-Based Opinion Mining of Students' Reviews on Online Courses. In Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence, ICCAI '20; Tianjin, China, 23–26 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 510–514.
15.   Beci, B. *Gramatika e Gjuhes Shqipe*; Logos-A: Shkup, North Macedonia, 2005.
16.   Singh, R.; Singh, R.; Bhatia, A. Sentiment analysis using Machine Learning technique to predict outbreaks and epidemics. *Int. J. Adv. Sci. Res.* **2018**, *3*, 19–24.
17.   Sharma, K.; Seo, S.; Meng, C.; Rambhatla, S.; Liu, Y. COVID-19 on Social Media: Analyzing Misinformation in Twitter Conversations. *arXiv* **2020**, arXiv:2003.12309.
18.   Sesagiri Raamkumar, A.; Tan, S.G.; Wee, H.L. Measuring the Outreach Efforts of Public Health Authorities and the Public Response on Facebook During the COVID-19 Pandemic in Early 2020: Cross-Country Comparison. *J. Med. Internet Res.* **2020**, *22*, e19334. [CrossRef]
19.   Samuel, J.; Ali, G.G.M.N.; Rahman, M.M.; Esawi, E.; Samuel, Y. COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. *Information* **2020**, *11*, 314. [CrossRef]
20.   Steinbauer, F.; Kröll, M. Sentiment analysis for German Facebook pages. In *International Conference on Applications of Natural Language to Information Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 427–432.
21.   Sidorenko, W. Sentiment Analysis of German Twitter. *arXiv* **2019**, arXiv:1911.13062.
22.   Palm, N. Sentiment Classification of Swedish Twitter Data. Master's Thesis, Uppsala University, Uppsala, Sweden, 2019.
23.   Dadoun, M.; Olssson, D. Sentiment Classification Techniques Applied to Swedish Tweets Investigating the Effects of translation on Sentiments from Swedish into English, 2016. Available online: https://www.diva-portal.org/smash/get/diva2:926472/FULLTEXT01.pdf (accessed on 5 April 2021).
24.   Mozetič, I.; Grčar, M.; Smailović, J. Multilingual Twitter sentiment classification: The role of human annotators. *PLoS ONE* **2016**, *11*, e0155036. [CrossRef]
25.   Lo, S.L.; Cambria, E.; Chiong, R.; Cornforth, D. Multilingual sentiment analysis: From formal to informal and scarce resource languages. *Artif. Intell. Rev.* **2017**, *48*, 499–527. [CrossRef]
26.   Biba, M.; Mane, M. Sentiment Analysis through Machine Learning: An Experimental Evaluation for Albanian. In *Recent Advances in Intelligent Informatics—Proceedings of the Second International Symposium on Intelligent Informatics, ISI 2013, Mysore, India, 23–24 August 2013*; Thampi, S.M., Abraham, A., Pal, S.K., Rodríguez, J.M.C., Eds.; Advances in Intelligent Systems and Computing; Springer: Berlin/Heidelberg, Germany, 2013; Volume 235, pp. 195–203.

27.   Kote, N.; Biba, M.; Trandafili, E. A Thorough Experimental Evaluation of Algorithms for Opinion Mining in Albanian. In Proceedings of the Advances in Internet, Data & Web Technologies, the 6th International Conference on Emerging Internet, Data & Web Technologies, EIDWT-2018, Tirana, Albania, 15–17 March 2018; Barolli, L., Xhafa, F., Javaid, N., Spaho, E., Kolici, V., Eds.; Lecture Notes on Data Engineering and Communications Technologies; Springer: Berlin/Heidelberg, Germany, 2018; Volume 17, pp. 525–536.

28.   Kote, N.; Biba, M.; Trendafili, E. An Experimental Evaluation of Algorithms for Opinion Mining in Multi-domain Corpus in Albanian. In Proceedings of the Foundations of Intelligent Systems—24th International Symposium, ISMIS 2018, Limassol, Cyprus, 29–31 October 2018; Ceci, M., Japkowicz, N., Liu, J., Papadopoulos, G.A., Ras, Z.W., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11177, pp. 439–447.

29.   Skenduli, M.P.; Biba, M.; Loglisci, C.; Ceci, M.; Malerba, D. User-Emotion Detection Through Sentence-Based Classification Using Deep Learning: A Case-Study with Microblogs in Albanian. In Proceedings of the Foundations of Intelligent Systems—24th International Symposium, ISMIS 2018, Limassol, Cyprus, 29–31 October 2018; Ceci, M., Japkowicz, N., Liu, J., Papadopoulos, G.A., Ras, Z.W., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11177, pp. 258–267.

30.   Skenduli, M.P.; Biba, M. Classification and Clustering of Emotive Microblogs in Albanian: Two User-Oriented Tasks. In *Complex Pattern Mining—New Challenges, Methods and Applications*; Appice, A., Ceci, M., Loglisci, C., Manco, G., Masciari, E., Ras, Z.W., Eds.; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2020; Volume 880, pp. 153–171.

31.   Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

32.   Kastrati, Z.; Imran, A.S.; Yayilgan, S.Y. The impact of deep learning on document classification using semantically rich representations. *Inf. Process. Manag.* **2019**, *56*, 1618–1632. [CrossRef]