

Article

A Survey on Sentiment Analysis and Opinion Mining in Greek Social Media

Georgios Alexandridis ^{1,2,*} , Iraklis Varlamis ² , Konstantinos Korovesis ³, George Caridakis ¹ 
and Panagiotis Tsantilas ³

¹ Department of Cultural Technology and Communication, University of the Aegean, 81100 Mitilini, Greece; gcari@aegean.gr

² Department of Informatics and Telematics, Harokopio University of Athens, 17671 Kallithea, Greece; varlamis@hua.gr

³ Palo Services Ltd., 10562 Athens, Greece; kk@paloservices.com (K.K.); pt@paloservices.com (P.T.)

* Correspondence: gealexandri@aegean.gr

Abstract: As the amount of content that is created on social media is constantly increasing, more and more opinions and sentiments are expressed by people in various subjects. In this respect, sentiment analysis and opinion mining techniques can be valuable for the automatic analysis of huge textual corpora (comments, reviews, tweets etc.). Despite the advances in text mining algorithms, deep learning techniques, and text representation models, the results in such tasks are very good for only a few high-density languages (e.g., English) that possess large training corpora and rich linguistic resources; nevertheless, there is still room for improvement for the other lower-density languages as well. In this direction, the current work employs various language models for representing social media texts and text classifiers in the Greek language, for detecting the polarity of opinions expressed on social media. The experimental results on a related dataset collected by the authors of the current work are promising, since various classifiers based on the language models (naive bayesian, random forests, support vector machines, logistic regression, deep feed-forward neural networks) outperform those of word or sentence-based embeddings (word2vec, GloVe), achieving a classification accuracy of more than 80%. Additionally, a new language model for Greek social media has also been trained on the aforementioned dataset, proving that language models based on domain specific corpora can improve the performance of generic language models by a margin of 2%. Finally, the resulting models are made freely available to the research community.

Keywords: sentiment analysis; opinion mining; Bidirectional Encoder Representations from Transformers; text embeddings; transformers; Greek social media



Citation: Alexandridis, G.; Varlamis, I.; Korovesis, K.; Caridakis, G.; Tsantilas, P. A Survey on Sentiment Analysis and Opinion Mining in Greek Social Media. *Information* **2021**, *12*, 331. <https://doi.org/10.3390/info12080331>

Academic Editor: Willy Susilo

Received: 28 July 2021

Accepted: 13 August 2021

Published: 18 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays social media are the biggest repository of public opinion about everything, from places, companies, and persons to products and ideas. This is mainly due to the fact that people prefer to express and share on social media their opinions on daily, local, or global issues and tend to comment on other people's views, which in turn creates a huge amount of opinionated data. The proper management and analysis of such data may uncover interests, beliefs, trends, risks, and opportunities that are valuable for marketing companies that design campaigns and promote products, persons, and concepts [1].

Sentiment analysis tasks involve the detection of emotions, their classification (binary or multi-class), and the mining of opinion polarity and subjectivity [2]. Argument extraction and opinion summarization can also be applied in order to describe public opinion on a more detailed level. The analysis can be performed at different levels of granularity, ranging from the sentiment of the whole document or each sentence to an opinion of a certain entity or aspect. The associated text mining tasks are mainly handled as classification problems using supervised or semi-supervised learning techniques [3]. However, some

approaches exist that employ unsupervised learning and a big group of techniques that include knowledge bases (lexicons, corpora, dictionaries) in order to improve results [4]. A taxonomy of the techniques and algorithms employed for sentiment analysis and opinion mining is depicted in Figure 1 and more details are given in the following paragraphs.

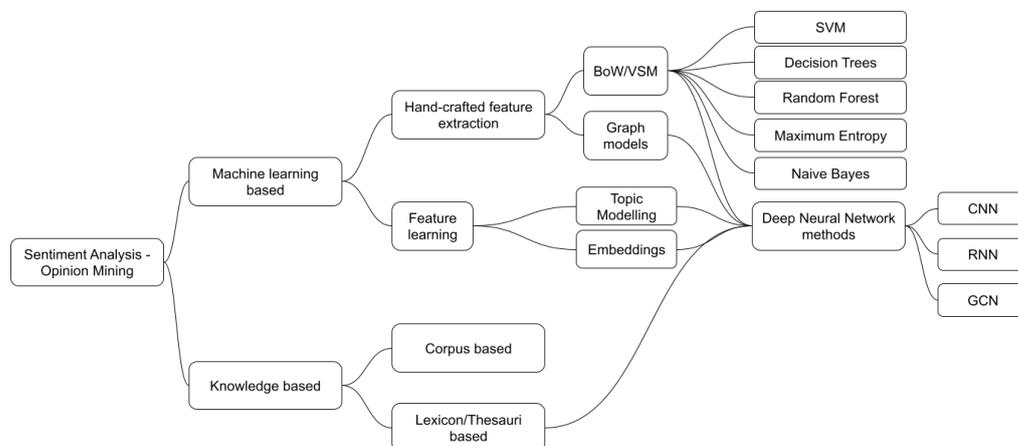


Figure 1. An overview of the techniques, algorithms, and models used for sentiment analysis and opinion mining.

A general pipeline for sentiment or opinion analysis is as follows: (i) a step pre-processing the corpus in order to remove noise and prepare the textual items for analysis, (ii) the extraction of features that will assist in the task and the representation of the items using a comprehensive model, (iii) the classification of the represented items using machine learning or knowledge-based techniques, and (iv) the visualization of results [5].

Natural language processing (NLP) techniques can be employed in the pre-processing step for sentence splitting, tokenization, Part-of-speech (POS) tagging, stopword removal, stemming, or even lemmatization. Most of these techniques require pre-trained language models and other linguistic resources, which limit their applicability in lower-density languages [6]. In the absence of linguistic resources, researchers rely on pre-trained models in a high-density language (e.g., English) and translation or parallel corpora in other languages, in order to perform cross-language sentiment classification [7,8].

As far as feature extraction/construction and representation of documents is concerned, the simplest solution is to represent documents as vectors in the vector space of all words used in the collection. For example, each document may be handled in a bag-of-words fashion or a vector of word features, where each word is weighted according to its importance to the document (term frequency), to the overall collection (inverse document frequency), or to the task (e.g., sentiment words from a lexicon are boosted). Each feature weighting technique calculates the usefulness of each feature and then weighs their values using the appropriate weights.

Despite its simplicity and its ability to model rich lexical meanings, the Bag of words (BoW) model fails to capture sentiment information. For this purpose, the list of words is expanded with symbols (emoticons) and other terms of positive or negative meaning, while words that can modify the polarity of sentiment (e.g., negation words) are considered to improve results [9]. In order to increase the number of features, some techniques employ different representations of the text, using n -grams or graphs (at the word or character level) and statistics in order to perform supervised tasks (e.g., text classification) [10–13].

Feature learning techniques have been proven to be better than vector-based approaches, since they are able in capturing semantic term–document information, as well as rich sentiment content [14]. Topic modeling is a form of unsupervised learning that has been extensively employed for the extraction and categorization of aspect from online product reviews [15] and for sentiment analysis [16]. Topic modeling techniques (e.g., Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (PLSA), Linear

Discriminant Analysis (LDA)) jointly model sentiment and semantic information of terms, but they do not generalize to new, unseen documents or words and they are prone to overfitting. Other works employ large-scale training corpora, which are usually automatically annotated (e.g., using emoticons or review ratings), in order to learn sentiment-specific word embeddings [17], which capture the meaning and semantic relations between words by exploiting their co-occurrence in documents belonging to a given corpus (e.g., Word2Vec, GloVe, etc.). As it holds with topic modeling techniques, embedding techniques are unable to handle unknown or out-of-vocabulary (OOV) words. In addition, they do not share representation at sub-word levels (e.g., prefixes or suffixes) which, however, can be captured using character embeddings.

The last step of the sentiment or opinion mining process is to use a classifier, which is fed with the feature vectors and decides on the subjectivity or polarity of the text. A large variety of classifiers has been employed in these tasks, with deep neural network architectures having been proven more efficient, especially when they are fed with sentiment-specific embeddings [17]. Convolutional and recurrent neural networks, bidirectional Long Short-Term Memory units (LSTMs), and gated recurrent units are some of the architectures that have been employed in such classification tasks and outperform simple feed-forward Neural Networks (NNs).

A major disadvantage of the aforementioned deep NN architectures is their difficulty in parallelizing and memorizing long dependencies within text sequences. Although the retroactive nature of recurrent NNs is suitable for sentence processing tasks (e.g., translation) when the size of the sentence is relatively small, it is insufficient for longer texts. Google [18] changed the way we perceive the processing of sequences in NLP tasks by introducing sequence-to-sequence (seq2seq) models that capture the structure of a language using encoders and decoders along with an *attention mechanism*; a context vector that attains the hidden states of the encoder and allows the decoder to selectively (weight-based) combine information from multiple hidden states. The transformer model [19] is based on the same encoder-decoder structure and employs a multi-head attention mechanism in order to model the semantic relation of terms that are distant in text, thus further improving performance in NLP tasks.

Pre-trained language models and transformers have recently demonstrated an improved ability in several NLP tasks and are now considered to be the state-of-the-art in creating contextualized word representations [20]. Bidirectional Encoder Representations from Transformers (BERT), Embeddings from Language Models (ELMo), and Generative Pre-trained Transformer 2 (GPT-2) are gradually replacing static word embeddings in task-oriented applications [21] and using the proper transformer, they can be fine-tuned for different NLP tasks. This generative and unsupervised pre-training of a language model can process all of the available text corpora in the first step, and the resulting model can be fine-tuned using a small labeled dataset [22], thus allowing few-shot learning to take place.

This work adopts the latter approach for analyzing the sentiment of social media content in the Greek language. The limited linguistic resources for Greek NLP tasks make this a promising solution, as validated by the experiments that follow. The main contributions of this work can be summarized as follows:

- A comprehensive survey of methods, tools, linguistic corpora and models that can be employed for sentiment analysis and opinion mining in Greek texts.
- A comparative evaluation of the state-of-the-art sentiment classification methods that employ pre-trained language models, both in the binary and three-class sentiment classification that also considers a highly represented neutral class.
- A new pre-trained language model for Greek social media texts, that has been trained on a large corpus collected from the Greek-speaking social media accounts.

In Section 2 that follows, we provide a survey of existing approaches in sentiment analysis and opinion mining from Greek social media content and of the resources that can be employed for this task. Section 3 illustrates the proposed method and Section 4 describes our experimental evaluation setup and the results we achieve using different

approaches. Finally, Section 5 summarizes the main findings of this work and discusses the next steps.

2. Related Work

Greek, like many other European languages, is a lower-density language with few linguistic resources and knowledge bases for basic text mining and natural language processing tasks. Recent works on Greek NLP [23,24] survey the work of Greek researchers on a wide categorization of NLP tasks, which range from syntactic and morphological analysis to generic and domain-specific applications. It is evident in these surveys that research efforts focus on the following: (i) developing NLP software and training specific models for each task, or extending popular NLP software to support the Greek language, (ii) annotating training corpora that are used for the tasks (usually for machine learning, supervised, tasks such as Named-Entity Recognition (NER), POS tagging, or sentiment classification), (iii) training embeddings in large Greek corpora, in order to capture the morphology and semantics of the language.

2.1. NLP Models and Software

One of the basic morphological tasks of NLP is stemming. All the available Greek stemmer implementations (e.g., the popular GreekStemmer [25]) originate from the Greek stemming rules introduced by Ntais [26]. Another work that focuses on the morphological and syntactic annotation of the Greek text is the Greek Dependency Treebank [27].

A publicly available model for Greek POS tagging has been developed by the NLP Team of the Department of Informatics, Athens University of Economics and Business (AUEB) [28], which, apart from POS tags, also assigns information, such as the gender, number, and case of each noun, the voice, tense, and form of each verb, etc. Recently, a supervised POS tagger for the Greek Language of the Social Web has been presented [29], showing that Greek social media content analysis is even more demanding than the analysis of general Greek text (e.g., news articles).

Regarding NER, the Named-entity recognizer for Greek text [30] identifies temporal expressions, person, and organization names, using an ensemble of Support Vector Machines (SVM) and active learning for person and organization names and patterns for the temporal expressions. Another NER module for Greek that distinguishes between persons, locations, organizations, and products has been trained on manually annotated social media content and semi-automatically annotated Wikipedia corpora, and achieved an 80% precision in the detection of named entities in tweets, Facebook posts, and news articles, with a recall rate ranging between 50% and 60% [31].

The semantic analysis of Greek text is another NLP task with many applications, including sentiment analysis and opinion mining. It includes several sub-tasks such as word sense disambiguation, named entity recognition, and semantic tagging of terms. A major restriction for the semantic processing of Greek text is that it usually requires rich linguistic resources and knowledge bases, which are in general unavailable for the Greek language. For example, WordNet and Senti-Wordnet are two principal resources employed for opinion mining and sentiment analysis tasks in many languages [32], but are not available for the Greek language (the Greek Wordnet [33] has been developed in the framework of the BalkaNet project [34], but it was never been finalized or made publicly available).

Because of their importance at the global level, sentiment analysis [35] and opinion mining tasks have also attracted the interest of Greek language researchers. The approaches recorded so far are supervised and either employ feature extraction techniques (including text-based and POS-based features) and sentiment lexicons, or feature learning techniques and Greek word or sentence embeddings as a basis. A brief survey of sentiment analysis from Greek text was made available by Papantoniou et al. [23]. A large-scale approach to sentiment analysis for brand monitoring in Greek social media has been presented by Petasis et al. [36] in the form of an automated Software as a Service (SaaS) application.

The “OpinionBuster” service was responsible for the extraction of named entities and the polarity associated with their mentions in texts, and exploited a large set of knowledge sources, such as open data from the Greek Wikipedia and Greek governmental APIs (e.g., data.gov.gr, geodata.gov.gr, etc.), privately owned entity lexicons, a NER extraction grammar that was specifically designed for the extraction of named-entities, a domain identifier and a NER component based on Conditional Random Fields, a rule-based co-reference resolution component, etc.

The extension of popular NLP software libraries to support the Greek language has been attempted by several researchers in order to facilitate general NLP tasks. Ellogon is one of the oldest toolkits for Greek NLP [37]. It supports several basic NLP tasks such as tokenization, splitting, POS tagging, lexical, and morphological analysis, etc., and has been used for solving complex tasks such as NER, sentiment analysis, etc. The neural NLP toolkit for Greek [38], integrates modules for POS tagging, lemmatization, dependency parsing, and text classification. It is based on language resources including web crawled corpora, word embeddings, large lexica, and corpora manually annotated at different levels of linguistic analysis. Among the various platforms for generic NLP tasks (e.g., NLTK [39], Spacy [40], OpenNLP [41], etc.), Spacy has been extended with Greek models for sentence splitting, tokenization, and POS tagging. The Greek version of spacy has been developed as an open source project [42] under the auspices of Google Summer of Code and Greek Free and Open Source Society. The CLARIN:EL European project has developed a set of NLP tools for various European languages including Greek, for which are available a POS tagger, a lemmatizer, and a syntactic tree tagger [43].

Similar attempts have been made for sentiment analysis in most European languages that are underrepresented, and have few annotated corpora for training machine learning models. For example, Wolk [44] uses machine translation as pivot in order to analyse sentiment in Polish dialog systems. Strimaitis et al. [45] tackle the complexity of the Lithuanian language using a supervised machine learning model to assign sentiment on financial context news. Pecar et al. [46] propose the use of a model ensemble for improving sentiment analysis in customer reviews written in the Slovak language and provide references to more recent works on sentiment classification in Slavic languages.

2.2. Annotated Corpora

A sentiment lexicon for Greek has been introduced by Kalamatianos et al. [47], while the same authors released a collection of tweets, which are manually rated for their sentiment intensity. The Greek Sentiment Lexicon consisted of 2,315 entries along with metadata about the tone, POS, objectivity, and emotional content of each word. Nikiforos et al. [24] released a POS tagged set of Greek tweets in order to support NLP tasks in social media content, and especially on Twitter. A detailed sentiment lexicon for Greek has been released by Tsakalidis et al. [48], containing for each term the POS tag, the subjectivity, the polarity, as well as the intensity of the six basic emotions.

2.3. Greek Embeddings

In an effort of creating reusable resources for Greek NLP research, several works have been published the last few years, which focus on developing Greek word embeddings, after extensive training on large Greek corpora. Greek Word2Vec embeddings have been trained by Outsios [49] using a large web corpus, crawled from about 20 million Greek-language URLs, with a raw text size of approximately 10 TB and a final corpus, in text form, of about 50 GB, containing 120 million sentences. The authors have created an online demo [50] that provides term synonyms, word embeddings for any crawled Greek term, etc. Nevertheless, the said model has not yet been employed in opinion mining or sentiment analysis tasks.

In order to improve the performance of word2vec embeddings trained in a Greek version of Wikipedia, Giatsoglou et al. [51] have extended the feature set with lexicon-based sentiment features that are extracted from the text. The NLPL word embeddings

repository [52] also provides word2vec embeddings, trained on the CoNLL17 corpus using the ELMo and the continuous skip-gram models. Pre-trained Greek word embeddings are also available for fastText, trained on Common Crawl and Wikipedia [53].

Very recently, researchers employed transformer-based language models, such as BERT, and achieved state-of-the-art performance in several NLP tasks. The GreekBERT model [54] has been pre-trained on 29 GB of text from the Greek part of (i) Wikipedia, (ii) the European Parliament Proceedings Parallel Corpus (Europarl), and (iii) of OSCAR [55], a clean version of Common Crawl [56]. Finally, a GPT-2 model for Greek trained on a large text corpus (almost 5GB) from Wikipedia, is also available online [57].

The following section explains how the Greek embeddings have been employed for extracting vector representations of text from Greek social media. We comparatively evaluate their performance against the traditional vector space model and bag-of-words representation.

3. Proposed Method

Sentiment analysis and opinion mining are usually handled as text classification tasks that consist of (Figure 2) (i) gathering of data samples, (ii) annotation of the samples, (iii) selection of a representation model, (iv) selection and training of the classification model, and (v) model deployment. Following the same process, we examine all the available contextualized word embeddings for the Greek language and we test different classification models in order to comparatively evaluate the different alternatives.

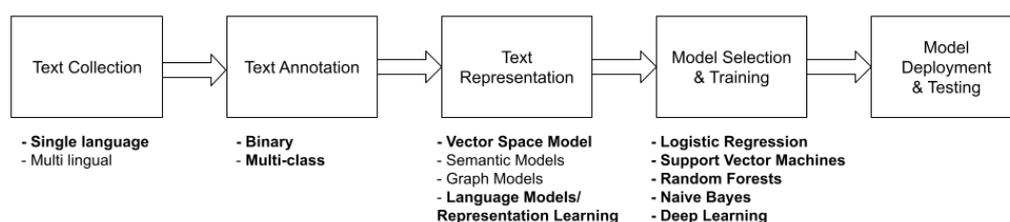


Figure 2. The basic steps for text classification and the proposed choices.

Regarding the first two steps, it is important to choose a dataset that is representative of the task at hand and the real situation that we are training the model for. In the case of Greek social media, we are interested in classifying Greek content only, so we compile a corpus of Greek texts. Since opinion in social media can be positive, negative, or neutral, with neutral being the majority, and the other two classes being comparable in size, we suggest examining text subjectivity before examining text polarity [58]. For completeness, we decide to examine both the binary classification task (using only positive and negative texts for training and testing) and the multi-class task using samples from all three classes.

In the case of text representation for NLP, the following alternatives may be considered: (i) the Vector Space Model (VSM) and its variations [59], which either capture the occurrence or absence of each term (or n -gram) in the document (one-hot vectors or BoW) or weight each word (or n -gram) according to its importance in the text (Term Frequency, Inverse Document Frequency, TF-IDF); (ii) semantic models that capture the inherent semantic relationships between terms, such as Latent Semantic Indexing (LSI), Probabilistic Latent Semantic Indexing (PLSI), and LDA [60]; (iii) graph representations [61], which map words to nodes and draw edges between related (by positional proximity, semantic similarity, etc.) words; and (iv) language models which are based on learning the representation [62] of words either individually (Word2Vec, GloVe, or FastText) or in their context (ELMo, BERT, GPT2, etc.). The latter category of methods has demonstrated better effectiveness than the others and constitutes the state-of-the-art in text representation for many NLP tasks, including text classification. For this reason, they are the main choice for our proposed approach and they are compared for their performance in the sentiment analysis of Greek text from social media. We compare several pre-trained language models and then train our own model directly on a large collection of Greek social media text.

As far as text classification algorithms are concerned, the related surveys in the field [63–65] highlight many classifiers that can be trained using vectors as input and produce a binary or multiclass classification. However, due to the high dimensionality of the induced vector space, random forests, SVMs, and deep neural networks are among the most popular approaches [66]. Logistic regression and naive Bayes classification are also quite popular because of their simplicity and low algorithmic complexity [67]. In this respect, all available text representations are evaluated on the same set of algorithms.

3.1. Datasets, Algorithms, and Models Employed

In order to evaluate our approach, a collection of Greek social media text has been collected and manually annotated. Annotation conforms to criteria defined by end-customers, which are transformed in well-defined and non-conflicting rules that dictate the correlation between a text excerpt and its sentiment value. Correlation rules can be analyzed by their scope and unit of focus. Scope can be either *global* or *customer-specific*; the former refers to commonly agreed criteria, such as association of negative sentiment and lists of insulting words or phrases. The latter refers to specific customer rules, which may contradict with the global rules and, in those cases, will take precedence. Unit of focus is the aspect of the given text excerpt being examined by the rule. Aspects include the text and its metadata, which in turn consist in the respective data source information (e.g., news site and related category), entity type (e.g., facebook post or comment), generation time (e.g., twitter comments posted after midnight), author, etc. Indicatively, a rule can examine the text for occurrence of specific words as well as its origin site.

The aforementioned process is a manual task that involves human annotators (Palo Ltd. employees), specifically trained for this work. The annotators study the defined rules and apply them by assigning sentiment values to the training sample. Sample selection follows the stratified random sampling methodology, with subgroups defined by the respective data sources (sites, blogs, social media, etc.). Annotation is software-assisted and forms the model training data set which is then employed by the training procedure.

The dataset is comprised of 59,810 texts of varying lengths, coming from different social media. The distribution of length (in words) is depicted in Figure 3, whereas more details on document polarity and social media channel distribution are outlined in Table 1.

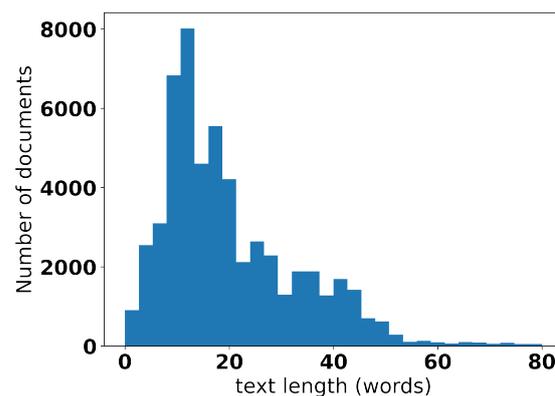


Figure 3. The distribution of text lengths.

The obtained text has been subsequently mapped on a vector space, using pre-trained Greek language embeddings and more specifically *fastText* [68], *BERT* [69] and *GPT-2* [57]. Finally, the models considered for sentiment classifications are the naive bayesian classifier (NB), random forests (RF), support vector machines with an Radial Basis Function (RBF) kernel, logistic regression (LR), and deep feed-forward neural network (DL) consisting of two hidden layers and an intermediate dropout layer (to avoid overfitting). For the baseline VSM, we used a count vectorizer and a TF transformer.

Table 1. The distribution of texts by polarity and channel.

	Positive	Negative	Neutral
Blogs	79	99	433
Facebook	1271	553	4575
Forums	1	3	14
Instagram	257	23	2128
News	697	372	3201
Twitter	4295	6869	33,297
YouTube	442	58	1143
Total	7042	7977	44,791

3.2. Training a Language Model for the Greek Social Media

In order to train a language model for Greek social media, we have crawled a much larger dataset without sentiment annotation, which is comprised of 458,293 documents from the same set of social media channels as the labeled one. Their distribution is depicted in Figure 4. As the source of text originates from social media, it contains certain artifacts that are very common in these media, like URLs, emoticons, hashtags, mentions, etc. Prior to using the collected data for the language model training, the following basic preprocessing rules have been applied: (i) make all characters lowercase, (ii) replace accented vowels with unaccented ones, (iii) remove the retweet characters, and (iv) remove URLs.

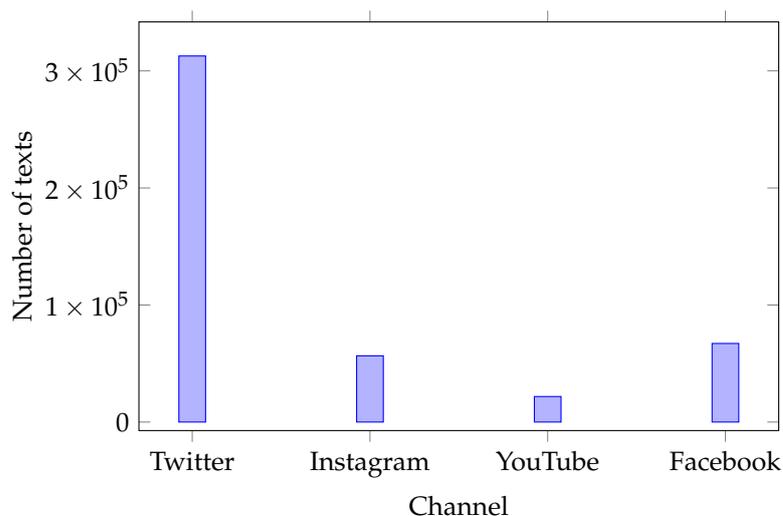


Figure 4. The distribution of unlabeled texts per channel.

The next step was to specify which language model to train on the collected dataset. In this respect, two different approaches were followed; the first was to train a new language model from scratch and the second was to further train an existing Greek language model. Regarding the first approach, the new model has been based on RoBERTa [70], where the encoder contains 6 hidden layers (of 768 neurons each) and 12 attention heads, thereby having $6 \times 12 = 72$ attention mechanisms. In this case, a new GTP-2 tokenizer has also been trained from scratch on the textual part of the collected dataset, after applying the basic preprocessing rules discussed above. The vocabulary size and the minimum word frequency were set to the default values of 52,000 and 2, respectively. The training objective was set to masked language modelling, where, according to BERT standards, 15% of tokens were selected for possible replacement. The maximum sequence length was set to 512, while the input sequences were trained in batches of 16. The warm-up steps were set to 500 and the weight decay to 0.01. The new language model, which we further mention as *PaloBERT*, has been trained for 10 epochs, as illustrated in Figure 5.

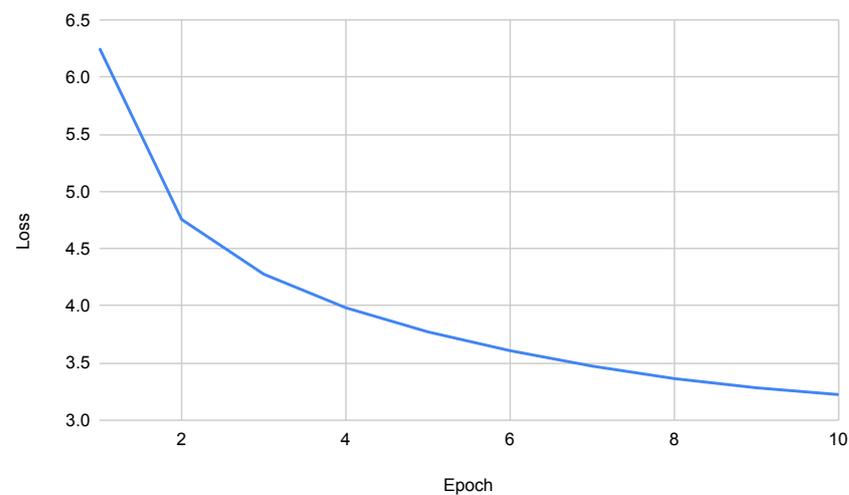


Figure 5. PaloBERT language model training loss.

PaloBERT training has been performed on the ARIS High Performance Computing Infrastructure [71], deployed and operated by GRNET (National Infrastructures for Research and Technology of Greece), on a GPU node containing two NVIDIA Tesla k40m GPU cards. The resulting model and tokenizer are available for download at HuggingFace’s model repository [72].

In addition to PaloBERT, we decided to further train an existing language model on the collected dataset. More specifically, we selected GreekBERT [54] (Section 2.3) and we trained it for an additional 10 epochs (Figure 6) on our social media corpus, on an AWS p2.xlarge instance running the Deep Learning AMI (Amazon Linux) Version 48.0. The resulting model, which is further mentioned as *GreekSocialBERT*, is also available for download at HuggingFace’s model repository [73].

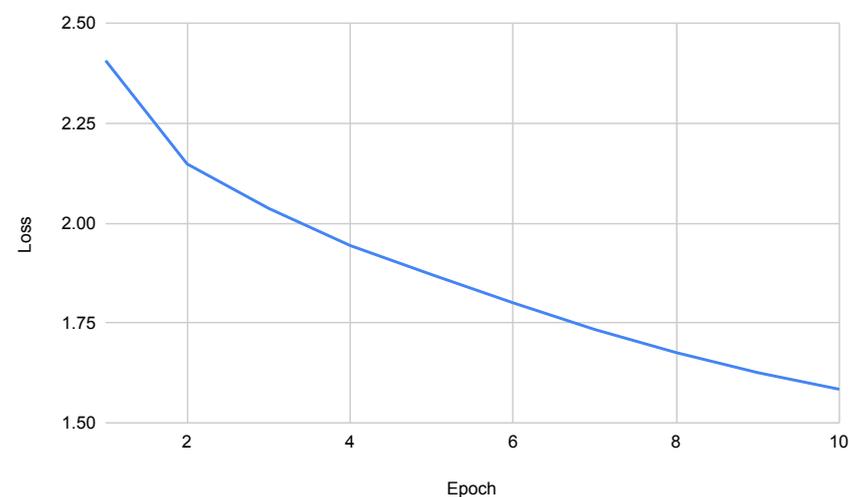


Figure 6. GreekSocialBERT language model training loss.

4. Results

4.1. Binary Sentiment Prediction with Pre-Trained Language Models

The first experiment comparatively evaluates the performance of the different pre-trained Greek language models in a binary sentiment prediction task, using only the positive and negative documents from the annotated dataset. In order to evaluate the performance of the training dataset size and find the language models that perform better with fewer and more samples, respectively, we kept a held-out set of 2000 annotated documents as a test set and used a size-increasing training dataset for training. We performed random stratified sampling for both test and training datasets in order to keep the class ratio in the

original annotated dataset, and repeated the experiment 5 times using a different seed each time. The classification algorithms discussed in Section 3.1 have been tested with the BERT language model, whereas GPT-2 has been tested using the GPT-2 model transformer with a sequence classification head (linear layer) on top. Finally, the fastText embeddings and the baseline VSM have been evaluated using the same deep neural network (DL) as the BERT model and logistic regression (VSM-LR).

Figure 7 demonstrates the performance of the different language model and classification algorithm combinations (Section 3.1). Due to space limitations, we plot the average weighted F-measure score in the 5 runs. We omit the detailed results of the 5 repetitions of the experiment, but it is worth mentioning that the standard deviation in the experiments with less than 2000 training samples in all cases ranges below 0.02, whereas in the experiments with 5000 or more training samples it ranges below 0.005. These numbers denote the statistical significance of any difference that is more than 0.02 (i.e., 2%) in the first and more than 0.005 (i.e., 0.5%) in the second case, at the 0.01 level of significance. Several methods have been tested and omitted from the plot since they perform worse than the results displayed in the plot.

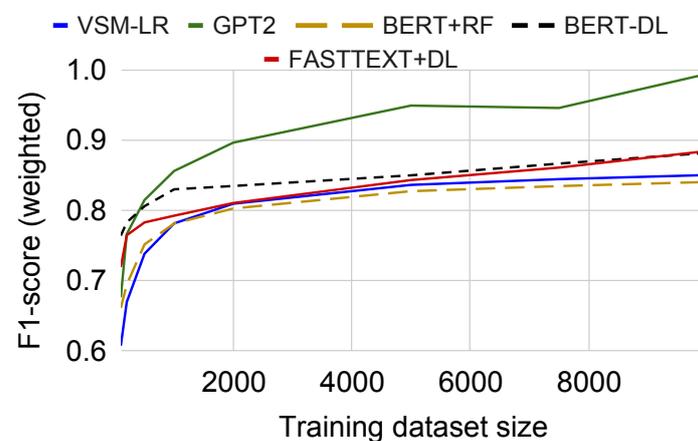


Figure 7. Classification performance (weighted F1-score) for the different combinations of language model and classifiers.

The performance of the GPT-2 language model is significantly better than that of BERT or fastText. Especially when 10,000 training samples are used for training the classifier, its performance reaches 99% in the unseen test samples. All methods that employ pre-trained models outperform the baseline VSM, although the choice of the classification algorithm is important and the differences, especially when a larger training dataset is employed, are not significant.

It is worth mentioning that all methods that employ language models perform better than the VSM, as expected. As shown in Figure 8 the pre-trained GreekBERT language model, combined with deep neural networks, outperforms all other methods and reaches a classification accuracy (in most experiments Recall, Precision, and accuracy were comparable to the displayed F-measure of 80%).

4.2. Three-Class Sentiment Prediction with Custom Language Models

In order to evaluate the performance of the different language models and classification techniques, we consider a real case where, apart from positive and negative documents, we also have a large amount of sentiment-neutral text, which significantly affects the balance and consequently the performance of sentiment prediction [74]. As shown in Table 1, the neutral class is the majority in the dataset (75% of the samples).

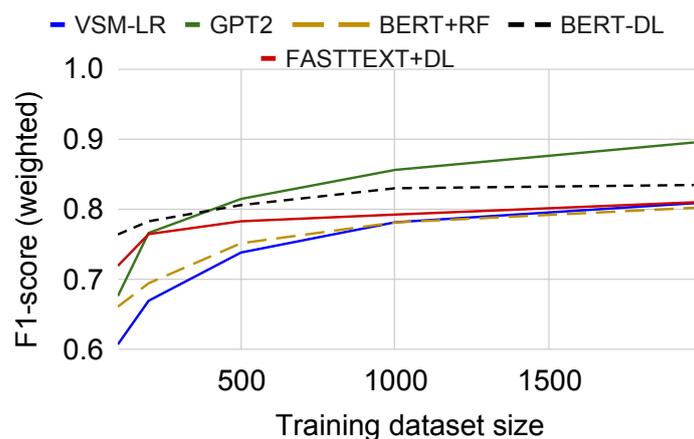


Figure 8. Classification performance (weighted F1-score) for the different combinations of language model and classifiers (few training samples).

The dataset employed in the following experiments has been described in Section 3.2, along with the two new language models (PaloBERT and GreekSocialBERT) introduced in this work. Moreover, we considered two levels of text preprocessing for the sentiment prediction task. The first ($p1$) is the basic preprocessing described in Section 3.2, while the second ($p2$) includes all the steps of $p1$ and additionally emoticons are converted to English language terms, the hashtag (#) and mention (@) symbols are removed, along with any hashtags or mentions appearing in the begging or the end of the social media text.

The resulting corpora ($p1$ or $p2$ preprocessed) are further divided into training, validation, and testing subsets, using a 60%:20%:20% stratified split, which results in 35,886 documents for training, 11,962 for validation, and 11,962 for testing. On top of each language model, we add a fully connected layer and we train each network on the training dataset for 10 epochs. We validate the performance of each classifier using the validation set, displaying the respective losses in Figure 9 for the $p1$ preprocessing case. It is obvious from the results that in all models the training loss decreases, though the generalization loss increases immediately. This is an indication of overfitting, which could be avoided if a dropout layer is added on-top of each BERT model, as we did in the previous set of experiments. However, in this case we employed early stopping, retaining the parameters of each respective model at the epoch exhibiting the smallest generalization loss.

Following, the trained sentiment classifiers were evaluated on the test subset. Table 2 summarizes their performance. The detailed results for each class show higher Precision for the positive and negative classes but much lower Recall. The Precision and Recall for the neutral class is much higher (i.e., around 90%). In order to allow a straight comparison of the three models, we also report the F1-scores (the harmonic mean of Precision and Recall) for each class and the micro-averaged F1-score for all classes. The last column of Table 2 reports the micro-averaged F1-score and shows that the GreekSocialBERT improves the performance of PaloBERT, which in turn outperforms GreekBERT in the sentiment classification task with the three classes. The improvements of $\sim 1\%$ and $\sim 1.5\%$, respectively, need to be further experimentally validated for their significance but are a clear indication that the GreekSocialBERT language model that we trained can boost the sentiment prediction task when social media text is considered.

When further pre-processing is applied (i.e., strategy $p2$) to the labeled dataset the performance of the classifier improves, using the same language models, as shown in the second part of Table 2. The improvement is between 1% and 1.5% compared to the basic pre-processing strategy ($p1$). These results are an indication that the proper pre-processing of the corpus, especially in the case of social media texts that contain a lot of special symbols and notations, can improve the performance of the sentiment prediction task. A similar pre-processing strategy for the unlabeled corpus could probably further improve the results, but this would require retraining of the language model and will be part of our future

work in the field. The future work will include experimentation with more pre-processing strategies in order to find the best combination before retraining the language model.

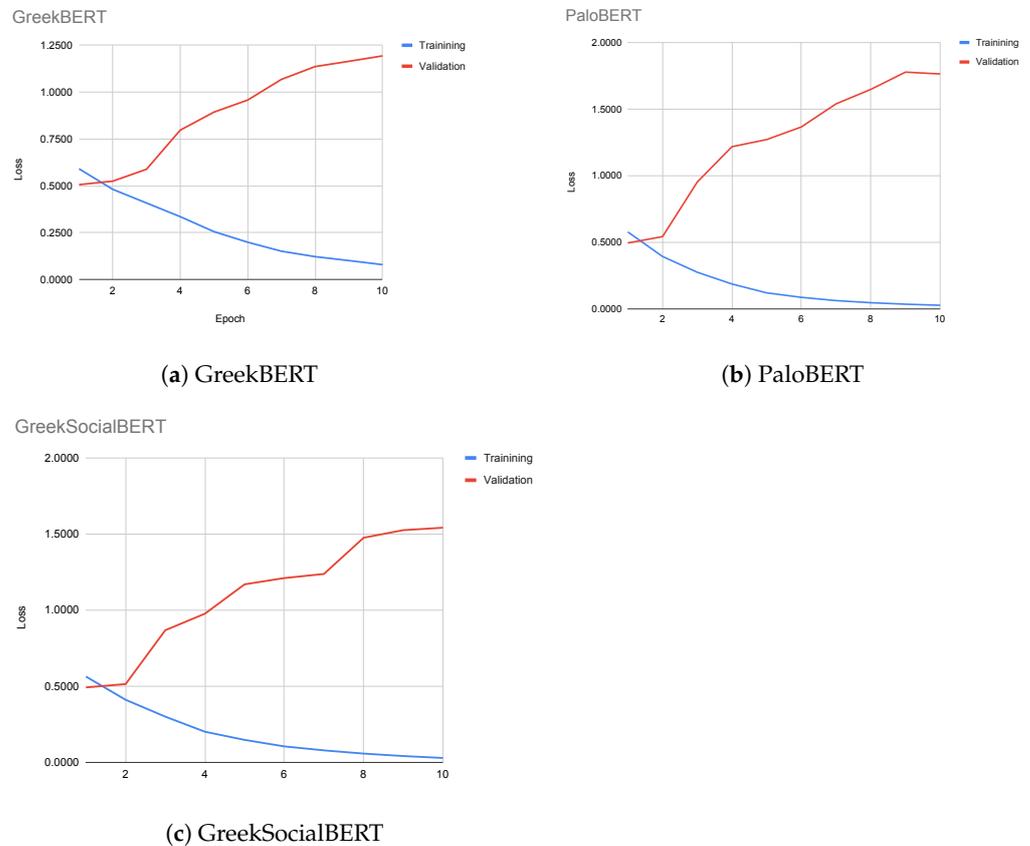


Figure 9. Classification performance (error loss) of the three language models on the training and validation datasets.

Table 2. Classification performance using the different models (*P* stands for *precision*, *R* for *recall*, *F1* for *F1 – score*, and *Avg.F1* for the micro-averaged *F1-score*) in the three classes (i.e., Negative, Positive, Neutral). The first table assumes minimum pre-processing of the texts of the labeled corpus (i.e., using pre-processing strategy *p1*). The second table assumes additional pre-processing (i.e., using pre-processing strategy *p2*).

<i>p1</i>	Negative			Positive			Neutral			Avg.F1
	P	R	F1	P	R	F1	P	R	F1	
GreekBERT	67.88	26.10	37.70	72.04	38.86	50.49	81.46	96.02	88.14	77.14
PaloBERT	62.70	42.58	50.72	76.39	35.56	48.53	82.66	94.13	88.03	78.53
GreekSocialBERT	61.35	41.63	49.60	62.82	55.10	58.71	84.73	91.10	87.80	79.40
<i>p2</i>	Negative			Positive			Neutral			Avg.F1
	P	R	F1	P	R	F1	P	R	F1	
GreekBERT	61.93	38.83	47.73	70.25	41.74	52.37	82.88	93.53	87.88	78.45
PaloBERT	89.86	13.11	22.89	71.27	42.24	53.04	80.23	97.32	87.95	79.80
GreekSocialBERT	71.43	23.55	35.42	67.77	45.76	54.63	81.71	95.44	88.04	80.17

5. Discussion & Conclusions

The experimental evaluation that has been performed in this work has validated the claim that language models can perform better than traditional representation models, such as the vector space model. Although it is impossible to check all the combinations of representation models and classification algorithms, the overall performance is definitely affected by the choice of the language model and the choice of the classification method.

The abundance of neutral texts also creates an imbalance in the class distribution, which makes the task of identifying positive or negative statements in social media even more difficult.

Additionally, the existence of a proper language model, which has been trained on a corpus that is related to the domain and task at hand (as in the case of Greek social media), can improve the performance of a generic language model that has been trained on a much larger corpus. As shown in the second experiment, the PaloBERT model that has been trained on a ten-times smaller dataset has outperformed GreekBERT. The improvement was even bigger when the generic model was used as a basis and was further trained with task-specific corpora.

Concerning the various pre-trained language models for the Greek language, it seems that the generic GPT-2 for Greek has the best performance, especially when more labeled data are used for training the classifier, whereas the GreekBERT performs better even with a small sample of labeled data. This difference in behavior is probably due to the classifiers employed in each case, since GPT-2 used a sequence classification head (linear layer) that performs better on text sequences, whereas BERT has been combined with a deep learning network with two fully connected layers separated by a dropout layer. The dropout layer may have a negative effect on the performance of the model on data similar to the ones that have been used for training, but increases the ability of the model to generalize. More experiments will be needed to examine the various aspects that affect the classification performance, and lead to a generalized sentiment classification model.

One of the limitations of this study is that the statistical significance of the results has not been evaluated in both cases. The binary sentiment prediction task has been repeated several times and resulted in a statistical significance of 2% (at the 0.01 level of significance), for the methods that outperform the others. However, by repeating the experiments with the three classes using the newly trained language models, we will be able to validate the significance of the new results.

The use of two different pre-processing strategies has shown an improvement of at least 2% when special characters and symbols have been replaced by their equivalent text or were removed. The second pre-processing strategy demonstrates better results than the first by at least 1.5%, although it has only been employed in the labeled data. The improvement would be even bigger if the same strategy were to be applied in the unlabeled corpus too, before training the language model on it. A second set of language models, trained using the unlabeled corpus that has been pre-processed using a different strategy could also be beneficial for our study. The retraining of the models requires additional resources and it is among the next plans of our work in this field.

The great performance of OpenAI GPT-3 in complex NLP tasks [75], such as text comprehension and generation, has increased the expectation of researchers for all NLP tasks. Although the sentiment prediction task can be formatted in such a way that it can be solved using GPT-3, this approach still has many limitations. For example, GPT-3 is only available via an API and only to people who apply for access. The size of the model makes it hard to run locally on any reasonable hardware, and this also limits the options for fine-tuning the model using domain-specific corpora.

In conclusion, the current study applied text representation and text classification algorithms in Greek social media documents and has comparatively evaluated the performance of various combinations language models and classifiers. The results show that pre-trained language models outperform traditional vector space representation models and that their performance can be further improved using language models trained on smaller, but domain-specific, corpora. The next steps of our work include the fine-tuning of the language models using different pre-processing strategies and the development of a custom language model using GPT-2.

Author Contributions: Conceptualization, G.A. and P.T.; methodology, I.V.; software, G.A. and K.K.; validation, I.V., G.C. and G.A.; formal analysis, G.C.; investigation, I.V.; resources, P.T.; data curation, K.K.; writing—original draft preparation, G.A. and I.V.; writing—review and editing, I.V., G.A. and

G.C.; visualization, I.V.; supervision, G.C. and I.V.; project administration, G.A.; funding acquisition, P.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH–CREATE–INNOVATE (project code: T1EDK-03470). This work was also supported by computational time granted from the National Infrastructures for Research and Technology S.A. (GRNET S.A.) in the National HPC facility–ARIS–under project ID 1003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The language models introduced in this work are available at the HuggingFace repository. PaloBERT may be obtained from [72], while GreekSocialBERT may be obtained from [73].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BERT	Bidirectional Encoder Representations from Transformers
BoW	Bag of words
DL	Deep Learning
ELMo	Embeddings from Language Models
GPT-2	Generative Pre-trained Transformer 2
GPT-3	Generative Pre-trained Transformer 3
LDA	Linear Discriminant Analysis
LR	Logistic Regression
LSA	Latent Semantic Analysis
LSI	Latent Semantic Indexing
LSTM	Long Short-Term Memory
NB	Naive bayesian classifier
NER	Named-Entity Recognition
NLP	Natural Language Processing
NN	Neural Network
VSM	Vector Space Model
OOV	Out of Vocabulary
PLSA	Probabilistic Latent Semantic Analysis
PLSI	Probabilistic Latent Semantic Indexing
POS	Part of Speech
RBF	Radial Basis Function
RF	Random Forest
TF-IDF	Term Frequency, Inverse Document Frequency
SVM	Support Vector Machines
VSM	Vector Space Model

References

1. Zhang, W.; Xu, M.; Jiang, Q. Opinion mining and sentiment analysis in social media: Challenges and applications. In Proceedings of the International Conference on HCI in Business, Government, and Organizations, Las Vegas, NV, USA, 15–20 July 2018; Springer: Cham, Switzerland, 2018; pp. 536–548.
2. Soong, H.C.; Jalil, N.B.A.; Ayyasamy, R.K.; Akbar, R. The essential of sentiment analysis and opinion mining in social media: Introduction and survey of the recent approaches and techniques. In Proceedings of the 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE), Kota Kinabalu, Malaysia, 27–28 April 2019; pp. 272–277.
3. Samal, B.; Behera, A.K.; Panda, M. Performance analysis of supervised machine learning techniques for sentiment analysis. In Proceedings of the 2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS), Chennai, India, 4–5 May 2017; pp. 128–133.

4. Katakis, I.M.; Varlamis, I.; Tsatsaronis, G. Pythia: Employing lexical and semantic features for sentiment analysis. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Nancy, France, 14–18 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 448–451.
5. Penalver-Martinez, I.; Garcia-Sanchez, F.; Valencia-Garcia, R.; Rodriguez-Garcia, M.A.; Moreno, V.; Fraga, A.; Sanchez-Cervantes, J.L. Feature-based opinion mining through ontologies. *Expert Syst. Appl.* **2014**, *41*, 5995–6008. [[CrossRef](#)]
6. Maxwell, M.; Hughes, B. Frontiers in linguistic annotation for lower-density languages. In Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006, Sydney, Australia, 15–16 July 2006; Association for Computational Linguistics: Stroudsburg, PA, USA, 2006; pp. 29–37.
7. Zhou, H.; Chen, L.; Shi, F.; Huang, D. Learning bilingual sentiment word embeddings for cross-language sentiment classification. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 430–440.
8. Xu, K.; Wan, X. Towards a universal sentiment classifier in multiple languages. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; pp. 511–520.
9. Balazs, J.A.; Velásquez, J.D. Opinion mining and information fusion: A survey. *Inf. Fusion* **2016**, *27*, 95–110. [[CrossRef](#)]
10. Dey, A.; Jenamani, M.; Thakkar, J.J. Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Syst. Appl.* **2018**, *103*, 92–105. [[CrossRef](#)]
11. Taher, S.A.; Akhter, K.A.; Hasan, K.A. N-gram based sentiment mining for bangla text using support vector machine. In Proceedings of the 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 21–22 September 2018; pp. 1–5.
12. Violos, J.; Tserpes, K.; Varlamis, I.; Varvarigou, T. Text classification using the n-gram graph representation model over high frequency data streams. *Front. Appl. Math. Stat.* **2018**, *4*, 41. [[CrossRef](#)]
13. Skianis, K.; Malliaros, F.; Vazirgiannis, M. Fusing document, collection and label graph-based representations with word embeddings for text classification. In Proceedings of the Twelfth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-12), New Orleans, LA, USA, 6 June 2018; pp. 49–58.
14. Maas, A.; Daly, R.E.; Pham, P.T.; Huang, D.; Ng, A.Y.; Potts, C. Learning word vectors for sentiment analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 142–150.
15. Kwon, H.J.; Ban, H.J.; Jun, J.K.; Kim, H.S. Topic modeling and sentiment analysis of online review for airlines. *Information* **2021**, *12*, 78. [[CrossRef](#)]
16. Rana, T.A.; Cheah, Y.N.; Letchmunan, S. Topic Modeling in Sentiment Analysis: A Systematic Review. *J. ICT Res. Appl.* **2016**, *10*, 76–93. [[CrossRef](#)]
17. Tang, D.; Wei, F.; Yang, N.; Zhou, M.; Liu, T.; Qin, B. Learning sentiment-specific word embedding for twitter sentiment classification. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, MD, USA, 23–25 June 2014; pp. 1555–1565.
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, U.; Polosukhin, I. Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA, 4–9 December 2017; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp. 6000–6010.
19. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the NAACL-HLT, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186
20. Ethayarajh, K. How contextual are contextualized word representations? comparing the geometry of BERT, ELMo, and GPT-2 embeddings. *arXiv* **2019**, arXiv:1909.00512.
21. Budzianowski, P.; Vulić, I. Hello, it's GPT-2—how can I help you? towards the use of pretrained language models for task-oriented dialogue systems. *arXiv* **2019**, arXiv:1907.05774.
22. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. Available online: <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf> (accessed on 1 July 2021)
23. Papantoniou, K.; Tzitzikas, Y. NLP for the Greek Language: A Brief Survey. In Proceedings of the 11th Hellenic Conference on Artificial Intelligence, Athens, Greece, 2–4 September 2020; pp. 101–109.
24. Nikiforos, M.N.; Voutos, Y.; Drougani, A.; Mylonas, P.; Kermanidis, K.L. The Modern Greek Language on the Social Web: A Survey of Data Sets and Mining Applications. *Data* **2021**, *6*, 52. [[CrossRef](#)]
25. GitHub. Skroutz/Greek_Stemmer: A Simple Greek Stemming Library. 2014. Available online: https://github.com/skroutz/greek_stemmer (accessed on 1 July 2021).
26. Ntais, G. Development of a Stemmer for the Greek Language. Master's Thesis, Department of Computer and Systems Sciences, Stockholm University/Royal Institute of Technology, Stockholm, Sweden, 2006; pp. 1–40.
27. Prokopidis, P.; Desipri, E.; Koutsombogera, M.; Papageorgiou, H.; Piperidis, S. Theoretical and practical issues in the construction of a Greek dependency treebank. In Proceedings of the 4th Workshop on Treebanks and Linguistic Theories (TLT 2005), Barcelona, Spain, 9–10 December 2005.
28. AUEB. NLP Group. 2021 Available online: <http://nlp.cs.aueb.gr/software.html> (accessed on 1 July 2021).
29. Nikiforos, M.N.; Kermanidis, K.L. A Supervised Part-Of-Speech Tagger for the Greek Language of the Social Web. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; pp. 3861–3867.

30. Lucarelli, G.; Vasilakos, X.; Androutopoulos, I. Named entity recognition in greek texts with an ensemble of svms and active learning. *Int. J. Artif. Intell. Tools* **2007**, *16*, 1015–1045. [CrossRef]
31. Makrynioti, N.; Grivas, A.; Sardianos, C.; Tsirakis, N.; Varlamis, I.; Vassalos, V.; Pouloupoulos, V.; Tsantilas, P. PaloPro: A platform for knowledge extraction from big social data and the news. *Int. J. Big Data Intell.* **2017**, *4*, 3–22. [CrossRef]
32. Sadegh, M.; Ibrahim, R.; Othman, Z.A. Opinion mining and sentiment analysis: A survey. *Int. J. Comput. Technol.* **2012**, *2*, 171–178.
33. Grigoriadou, M.; Kornilakis, H.; Galiotou, E.; Stamou, S.; Papakitsos, E. The software infrastructure for the development and validation of the Greek WordNet. *Rom. J. Inf. Sci. Technol.* **2004**, *7*, 89–105.
34. BalkaNet. Project Home Page. 2000 Available online: <http://www.dblab.upatras.gr/balkanet/> (accessed on 1 July 2021).
35. Guo, X.; Li, J. A Novel Twitter Sentiment Analysis Model with Baseline Correlation for Financial Market Prediction with Improved Efficiency. In Proceedings of the 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS), Granada, Spain, 22–25 October 2019; pp. 472–477. [CrossRef]
36. Petasis, G.; Spiliotopoulos, D.; Tsirakis, N.; Tsantilas, P. Sentiment analysis for reputation management: Mining the greek web. In Proceedings of the Hellenic Conference on Artificial Intelligence, Ioannina, Greece, 15–17 May 2014; Springer: Cham, Switzerland, 2014; pp. 327–340.
37. Petasis, G.; Karkaletsis, V.; Paliouras, G.; Androutopoulos, I.; Spyropoulos, C.D. Ellogon: A new text engineering platform. *arXiv* **2002**, arXiv:cs/0205017.
38. Prokopidis, P.; Piperidis, S. A Neural NLP toolkit for Greek. In Proceedings of the 11th Hellenic Conference on Artificial Intelligence, Athens, Greece, 2–4 September 2020; pp. 125–128.
39. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2009.
40. Honnibal, M.; Montani, I.; Van Landeghem, S.; Boyd, A. SpaCy: Industrial-Strength Natural Language Processing in Python; 2020, Available online: <https://zenodo.org/record/5115698#.YRnUSEQzZPY> (accessed on 1 July 2021).
41. Apache Software Foundation. OpenNLP Natural Language Processing Library. 2014. Available online: <http://opennlp.apache.org/> (accessed on 1 July 2021).
42. GitHub. Eellak/Gsoc2018-Spacy: [GSOC] Greek Language Support for Spacy.io Python NLP Software. 2018 Available online: <https://github.com/eellak/gsoc2018-spacy> (accessed on 1 July 2021).
43. CLARIN ERIC. Part-of-Speech Taggers and Lemmatizers. 2021. Available online: <https://www.clarin.eu/resource-families/tools-part-speech-tagging-and-lemmatization> (accessed on 1 July 2021).
44. Wołk, K. Real-Time Sentiment Analysis for Polish Dialog Systems Using MT as Pivot. *Electronics* **2021**, *10*, 1813. [CrossRef]
45. Štrimaitis, R.; Stefanovič, P.; Ramanauskaitė, S.; Slotkienė, A. Financial Context News Sentiment Analysis for the Lithuanian Language. *Appl. Sci.* **2021**, *11*, 4443. [CrossRef]
46. Pecar, S.; Šimko, M.; Bielikova, M. Improving Sentiment Classification in Slovak Language. In Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing, Florence, Italy, 2 August 2019; pp. 114–119.
47. Kalamatianos, G.; Mallis, D.; Symeonidis, S.; Arampatzis, A. Sentiment analysis of greek tweets and hashtags using a sentiment lexicon. In Proceedings of the 19th Panhellenic Conference on Informatics, Athens, Greece, 1–3 October 2015; pp. 63–68.
48. Tsakalidis, A.; Papadopoulos, S.; Voskaki, R.; Ioannidou, K.; Boididou, C.; Cristea, A.I.; Liakata, M.; Kompatsiaris, Y. Building and evaluating resources for sentiment analysis in the Greek language. *Lang. Resour. Eval.* **2018**, *52*, 1021–1044. [CrossRef]
49. Outsios, S.; Karatsalos, C.; Skianis, K.; Vazirgiannis, M. Evaluation of Greek Word Embeddings. In Proceedings of the 12th Language Resources and Evaluation Conference, Marseille, France, 11–16 May 2020; European Language Resources Association: Marseille, France, 2020; pp. 2543–2551.
50. Greek Word2Vec. 2018 Available online: <http://archive.aueb.gr:7000/> (accessed on 1 July 2021).
51. Giatsoglou, M.; Vozalis, M.G.; Diamantaras, K.; Vakali, A.; Sarigiannidis, G.; Chatzisavvas, K.C. Sentiment analysis leveraging emotions and word embeddings. *Expert Syst. Appl.* **2017**, *69*, 214–224. [CrossRef]
52. Fares, M.; Kutuzov, A.; Oepen, S.; Veldal, E. Word vectors, reuse, and replicability: Towards a community repository of large-text resources. In Proceedings of the 21st Nordic Conference on Computational Linguistics, Gothenburg, Sweden, 22–24 May 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 271–276.
53. Grave, E.; Bojanowski, P.; Gupta, P.; Joulin, A.; Mikolov, T. Learning Word Vectors for 157 Languages. In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May 2018.
54. Koutsikakis, J.; Chalkidis, I.; Malakasiotis, P.; Androutopoulos, I. Greek-bert: The greeks visiting sesame street. In Proceedings of the 11th Hellenic Conference on Artificial Intelligence, Athens, Greece, 2–4 September 2020; pp. 110–117.
55. Suárez, P.J.O.; Sagot, B.; Romary, L. Asynchronous pipeline for processing huge corpora on medium to low resource infrastructures. In Proceedings of the 7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7), Cardiff, UK, 22 July 2019; Leibniz-Institut für Deutsche Sprache: Mannheim, Germany, 2019.
56. Common Crawl. 2021 Available online: <http://commoncrawl.org/> (accessed on 1 July 2021).
57. Hugging Face. 2021. Available online: <https://huggingface.co/nikokons/gpt2-greek> (accessed on 1 July 2021).
58. Esuli, A.; Sebastiani, F. Determining term subjectivity and term orientation for opinion mining. In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, Italy, 5–6 April 2006.
59. Salton, G.; Wong, A.; Yang, C.S. A vector space model for automatic indexing. *Commun. ACM* **1975**, *18*, 613–620. [CrossRef]

60. Hofmann, T. Probabilistic latent semantic indexing. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA, USA, 15–19 August 1999; pp. 50–57.
61. Sonawane, S.S.; Kulkarni, P.A. Graph based representation and analysis of text document: A survey of techniques. *Int. J. Comput. Appl.* **2014**, *96*, 19. [CrossRef]
62. Liu, Z.; Lin, Y.; Sun, M. Representation Learning and NLP. In *Representation Learning for Natural Language Processing*; Springer: Singapore, 2020; pp. 1–11. [CrossRef]
63. Aggarwal, C.C.; Zhai, C. A survey of text classification algorithms. In *Mining Text Data*; Springer: Boston, MA, USA, 2012; pp. 163–222.
64. Vijayan, V.K.; Bindu, K.; Parameswaran, L. A comprehensive study of text classification algorithms. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017; pp. 1109–1113.
65. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150. [CrossRef]
66. Hartmann, J.; Huppertz, J.; Schamp, C.; Heitmann, M. Comparing automated text classification methods. *Int. J. Res. Mark.* **2019**, *36*, 20–38. [CrossRef]
67. Pranckevičius, T.; Marcinkevičius, V. Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Balt. J. Mod. Comput.* **2017**, *5*, 221. [CrossRef]
68. fastText. Word Vectors for 157 Languages. 2020. Available online: <https://fasttext.cc/docs/en/crawl-vectors.html> (accessed on 1 July 2021).
69. GitHub. Nlpaueb/Greek-Bert: A Greek Edition of BERT Pre-Trained Language Model. 2021 Available online: <https://github.com/nlpaueb/greek-bert> (accessed on 1 July 2021).
70. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
71. HPC. National HPC Infrastructure. 2016. Available online: <https://hpc.grnet.gr/en/> (accessed on 1 July 2021).
72. Hugging Face. 2021 Available online: <https://huggingface.co/gealexandri/palobert-base-greek-uncased-v1> (accessed on 1 July 2021).
73. Hugging Face. 2021. Available online: <https://huggingface.co/gealexandri/greeksocialbert-base-greek-uncased-v1> (accessed on 1 July 2021).
74. Tsytsarau, M.; Palpanas, T. Managing diverse sentiments at large scale. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 3028–3040. [CrossRef]
75. Edwards, C. The best of NLP. *Commun. ACM* **2021**, *64*, 9–11. [CrossRef]